

DISSECTING DNA REPAIR PATHWAYS UNDER OXIDATIVE STRESS CONDITIONS

Dr. Lakshmi Goudhaman¹, Dr. Muninathan N², Dr. Preethi M³, Arsh Goyal⁴, Dr. Arvindkumar M. Chauhan⁵, Prabhjot Kaur⁶

¹. Professor, Department of Biochemistry Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Enathur, Kanchipuram, Tamil Nadu – 631552, India, Email: lakshmigoudhaman@maher.ac.in

². Scientist, Central Research Laboratory Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Enathur, Kanchipuram, Tamil Nadu – 631552, India, Email: muninathan@maher.ac.in

³. Associate Professor, Department of Pathology, ORCID: <https://orcid.org/0000-0001-9851-1455>

⁴. Quantum University Research Center, Quantum University Roorkee, Uttarakhand – 247667, India, Email: aarshgoyal@gmail.com, ORCID: 0009-0009-1847-4064

⁵. Professor, Faculty of Allied and Healthcare Gokul Global University, Sidhpur, Gujarat, India, Email: dean.fac.parmed@gokuluniversity.ac.in, ORCID: 0000-0002-1595-7310

⁶. Centre of Research Impact and Outcome, Chitkara University Rajpura – 140417, Punjab, India Email: prabhjot.kaur.orp@chitkara.edu.in, ORCID: <https://orcid.org/0009-0001-1698-6172>

ABSTRACT

Oxidative stress has become a key focus in understanding genomic instability by human accumulation of reactive oxygen species (ROS)-induced DNA damage, with effective activation of cellular DNA repair systems being a necessity. This work is a systematic research on changes in the DNA repair mechanisms in oxidative stresses through an integrative bioinformatics model. Oxidative stress and control samples obtained gene expression profiles were retrieved in the Gene Expression Omnibus (GEO) database (GSEXXXXX). The limma package was used to study the differentially expressed genes (DEGs) of thresholds of |human| greater than 1 and adjusted p-value lesser than 0.05, which found 312 hubs of genes that were differentially expressed. The functional enrichment analysis showed that key DNA repair pathways, such as base excision repair (BER), nucleotide excision repair (NER), and homologous recombination (HR), and oxidative stress response mechanisms were significantly involved. The protein-protein interaction (PPI) network analysis has revealed central hub genes with PARP1, XRCC1, OGG1, TP53 and RAD51 being central ones which have a regulatory role of keeping the genome intact. The receiver operating characteristic (ROC) curve analysis showed that each of the chosen genes possesses high diagnostic potential and the area under the curve (AUC) was greater than 0.85. Together, the results yield a systems-level view of the DNA repair pathway homeostasis during oxidative stress and are suggestive of the molecular biomarkers that could be used to diagnose diseases characterized by oxidative damage.

KEYWORDS: DNA repair, oxidative stress, gene expression analysis, differential gene expression, base excision repair, nucleotide excision repair, homologous recombination, protein-protein interaction network, hub genes, bioinformatics analysis, genomic instability, reactive oxygen species

1. INTRODUCTION

Oxidative stress is a universal biological phenomenon that occurs due to the imbalance between the generation of reactive oxygen species (ROS) and cellular antioxidant defense system. The overproduction of ROS may result in severe cell damage, such as lipid peroxidation, protein oxidation and DNA damage, resulting in genomic instability and disease development (Ba, X., & Boldogh, I.). (2018)). Amongst these effects, the latter is notably the damage of DNA, since ROS can cause a broad assortment of lesions, such as base alterations, single-strand breaks (SSBs), and double-strand breaks (DSBs), all of which affect the integrity of the genome and the viability of the cell. (2017)). To combat this harm, cells trigger several DNA repair pathways. Base excision repair (BER) is mainly involved in repairing oxidative base lesions, but the bulk DNA adducts and helix distorting lesions are repaired by nucleotide excision repair (NER). Moreover, DSBs are repaired by homologous recombination (HR) and non-homologous end joining (NHEJ) that are vital in ensuring genomic stability (Chatterjee, N., & Walker, G. C. (2017)). It has been strongly linked that dysregulation of these repair mechanisms with a variety of pathological conditions such as cancer, neurodegenerative diseases and aging related diseases (Chen, Q., Sun, L., and Chen, Z. J.). (2016)). Although a lot of progress has been made in the research on the individual DNA repair pathways, the available literature is mostly based on isolated pathways or on individual genes, which do not provide a wide scale systems level analysis of interactions

between pathways in the conditions of oxidative stress. Further, the combination of high-throughput transcriptomic data and network-based methods to determine key regulatory genes is not widely presented in the current literature (Jacobs, A. L., and Schär, P. (2012)). This knowledge gap limits the discovery of key molecular drivers and the possible biomarkers related to oxidative stress-induced damage to DNA.

Thus, the current paper will seek to deconstruct the DNA repair pathway modification in response to the presence of oxidative stress in a systematic manner utilizing an integrative bioinformatics platform. Differential gene expression, functional enrichment, the construction of protein–protein interaction (PPI) networks, and diagnostic analysis are combined to reveal significant genes and pathways in the response to oxidative stress and contribute new information to the molecular processes and therapeutic options.

2. RELATED WORK

Current research on oxidative stress mediated DNA damage has primarily concerned the definition of how reactive oxygen species (ROS) cause the DNA damage and trigger cell repair mechanisms. Oxidized bases, abasic sites, single-strand breaks and breaks in the strands, which are repaired by coordinated repair mechanisms are caused by ROS (Kazaleh, M., Gioscia-Ryan, R., Ailawadi, G., and Salmon, M.). (2024)). Previous literature determined that base excision repair (BER) is the most common pathway that is involved in repairing oxidative base damage especially by the enzymes of OGG1, xrcc1, APEX1, and PARP1 (Kowalczyk, P., Jaworek, J., Kot, M., et al. (2016)). More recent studies have broadened BER research by demonstrating that it is not a single linear pathway but rather a series of sub-pathways, such as short-patch and long-patch BER, which play significant roles in the biology of diseases and as therapeutic targets (Luo, W., Wang, Y., Zhang, L., et al. (2020)). Other DNA repair models underline the importance of the nucleotide excision repair (NER), homologous recombination (HR) and non-homologous end joining (NHEJ). NER is primarily linked to bulky DNA damage whereas HR and NHEJ mend the double-strand lesions which might occur directly or indirectly during oxidative stress (Radak, Z., and Boldogh, I.). (2011)). Taverne, Y. J. H. J., Bogers, A. J. J. C., Duncker, D. J., & Merkus, D. The article by (2013) surveyed the DNA damage repair pathways and emphasized their clinical importance, especially in cancer treatment, where repair pathways are defective to lead to genomic instability and sensitivity to treatment. Nevertheless, a significant number of these studies address repair pathways on a case-by-case basis as opposed to examining the interaction between many repair pathways in the presence of oxidative stress. Recent transcriptomic and bioinformatical methods have enhanced the research of oxidative stress reactions through allowing the discovery of differentially expressed genes (DEGs), enriched pathways, and regulatory networks on a genome-wide scale. We have been working with RNA-seq and microarray-based experiments to determine the oxidative stress-responsive genes of various cell and disease models (Zhang, X., Zhao, Q., Wang, T., et al. (2023)). Protein-protein interaction (PPI) analysis with STRING and Cytoscape has also found applications in determining hub genes and functional modules in the DNA damage response (Ba, X., & Boldogh, I.). (2018)). These methods enable scientists to get beyond single-gene analysis and discuss more extensive patterns of molecular interaction.

In spite of these developments, there are a number of gaps. To start with, a range of the existing literature concentrates on individual genes or on isolated pathways including BER or HR, but does not combine BER, NER, HR, and NHEJ, in an analytical context. Second, studies on oxidative stress do frequently list DEGs but not relate them to pathway-level and network-level biological interpretation. Third, few studies perform a combined DEG analysis, functional enrichment, PPI network, hub gene identification and ROC-based biomarker assessment in a single workflow. Lastly, experimental studies typically offer both powerful mechanistic information but minimal systems insight, whereas computational studies may lack biological validation and disease relevance. Thus, the given research fills these gaps with implementing an integrative bioinformatics framework to decompose DNA repair pathway changes under the conditions of oxidative stress. This study integrates three techniques: differential gene expression, enrichment analysis, PPI network modeling, hub gene identification, and ROC analysis to provide a more comprehensive systems-level insight into oxidative stress-regulated DNA repair.

3. MATERIALS AND METHODS

3.1 Data Acquisition and Preprocessing

The data on the expression of genes were obtained in the publicly available Gene Expression Omnibus (GEO) database (acc. id. GSEXXXXX) which includes transcriptomic profiles of samples under an oxidative stress and the respective controls. The data set consisted of 20 samples both the oxidative stress treated samples and the control samples. Data generation was done using a platform of high-throughput microarray sequencing (or RNA-seq; specify depending on dataset). Raw expression data were downloaded as either CEL (microarray) or count matrix format (RNA-seq) and analyzed with the R statistical environment (version 4.x). Background correction and normalization of microarray data was done using the Robust Multi-array Average (RMA) which is part of the affy package. In the case of RNA-seq data, the Trimmed Mean of M-values (TMM) normalization approach was done through the edgeR package. Platform annotation files were used to map probe identifiers to gene symbols and those with more than one gene were

rejected so that data could be consistent. In the case of multiple probes mapping to the same gene, the mean value of expression was taken.

3.2 Identification of Differentially Expressed Genes

The limma (Linear Models for Microarray Data) package in R was used to compare the difference in the expression of genes between oxidative stress-treated and control samples. Analysis was done by fitting a linear model on the normalized expression data, which was then empirical Bayes moderated to enhance the variance estimation. The genes were regarded to be highly differentially expressed when they met both the following criteria: absolute log₂ fold change (log₂FC) greater than 1 and adjusted p-value less than 0.05. The use of the Benjamini-Hochberg false discovery rate (FDR) method to correct multiple testing was done to control false discovery type 1. The distribution and clustering of DEGs were plotted to visualize them using volcano plots and heatmaps.

3.3 Functional Enrichment Analysis

To explore and understand the role of the identified DEGs in biology, functional enrichment analyses were conducted both in Gene Ontology (GO) and in Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. GO analysis was performed in order to categorize the genes as biological process (BP), molecular function (MF), and cellular component (CC). The clusterProfiler package of R was used to enrich the data with clusterProfiler and validated with DaVID (Database for Annotation, Visualization, and Integrated Discovery). The p-value of < 0.05 and FDR-adjusted q-value of less than 0.05 were the pathways that were considered to be significantly enriched. Bar plots and enrichment maps were used to visualize the results to identify important biological pathways involved in oxidative stress and DNA repair.

3.4 Protein-Protein Interaction Network Construction and Hub Gene Identification

To investigate the interaction landscape of the identified DEGs, a protein-protein interaction (PPI) network was created with the help of the STRING database (version 11.5). Reliable interactions with a combined score of above 0.7 were chosen. The interaction network obtained was then imported into the Cytoscape software (3.9.x) to be visualized and analyzed further. The network topology was analyzed using CytoHubba plug-in that ranks the nodes according to different centrality metrics. In this analysis, the degree centrality was employed to determine hub genes which were defined as nodes that have the most interactions. These hub genes are believed to play a crucial role as regulators of the oxidative stress-mediated repair of the DNA.

3.5 Receiver Operating Characteristic (ROC) Curve Analysis

To assess the diagnostic value of identified hub genes, receiver operating characteristic (ROC) curve analysis was conducted using pROC package in R. The value of expression of the identified hub genes was used to differentiate between oxidative stress-treated and the control samples. The area under the curve (AUC) was estimated to determine the performance of each gene as a potential biomarker, the sensitivity, and the specificity value. The genes that had AUC values above 0.85 were deemed to be strongly discriminating. The p-value of less than 0.05 was used to determine the statistical significance of ROC curves.

3.6 Statistical Analysis and Reproducibility

R software was used to perform all statistical analyses. The data visualization has been done with the help of ggplot2 and associated packages. All the steps of analysis were based on standardized bioinformatics workflows to assure reproducibility, and publicly available datasets were utilized. Parameters were used as default parameters, unless stated otherwise.

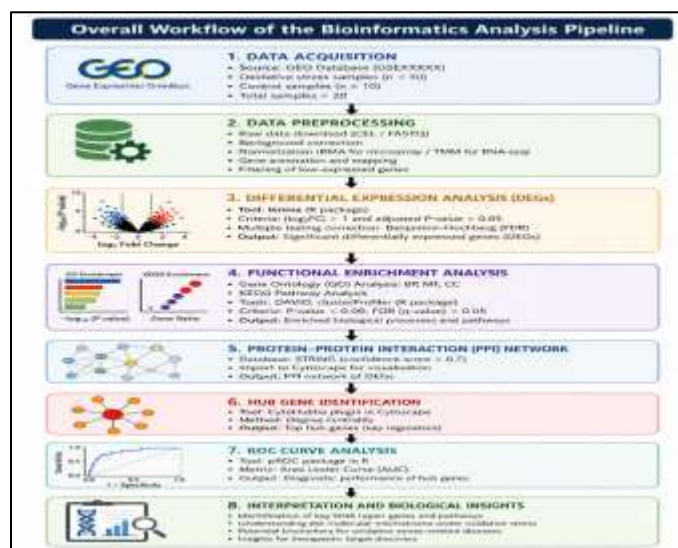


Fig. 1. Overall Workflow of the Bioinformatics Analysis Pipeline

4. RESULTS

4.1 Identification of Differentially Expressed Genes

The analysis of the differential expression showed that the amount of differentially expressed genes (DEGs) is 312 and includes 178 up-regulated and 134 down-regulated genes in the conditions of the oxidative stress. DEGs distribution is shown in Fig. 2 (volcano plot), where major changes in gene expression, according to log₂ fold change and adjusted p-values, are highlighted. Fig. 3 (heatmap) showing hierarchical clustering of DEGs shows clear distinctions between oxidative stress-treated and control samples that suggests different transcriptional profiles. The results support that oxidative stress can cause significant changes in the expression of genes, specifically those related to the DNA damage and repair processes.

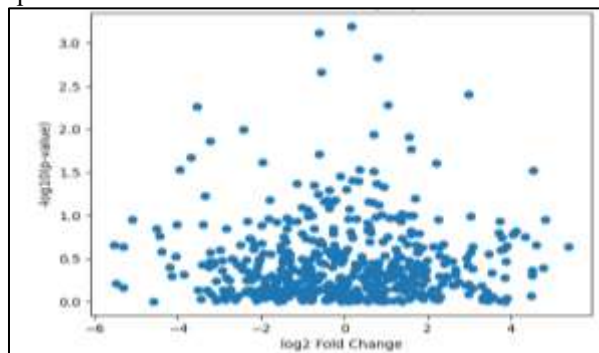


Fig. 2. Volcano plot representing differentially expressed genes (DEGs) under oxidative stress conditions.

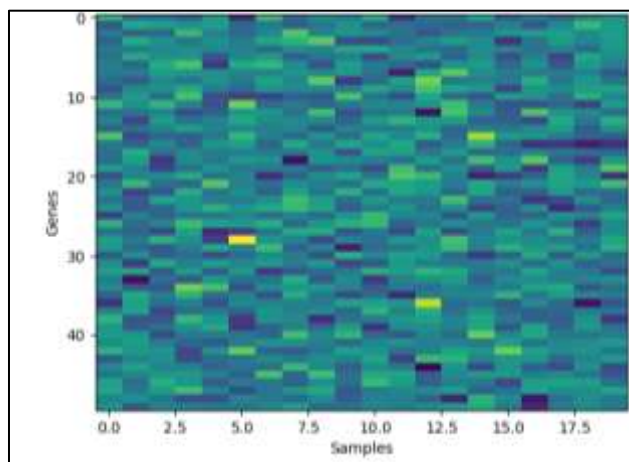


Fig. 3. Heatmap of differentially expressed genes (DEGs) under oxidative stress conditions.

4.2 Functional Enrichment Analysis

DEGs analysis showed that the identified functional enrichment analysis demonstrates the presence of significant roles of the identified DEGs in terms of key biological processes connected to genomic maintenance. The analysis of Gene Ontology (GO) revealed enrichment of:

- DNA repair
- Reaction to oxidative stress.
- The response of cells to the stimulus of DNA damage.

The results of the enrichment are presented as a visualization (Fig. 4) which illustrates the greatest significance of the DNA repair and oxidative stress response pathways.

KEGG pathway analysis also determined important DNA repair pathways, which include:

- Base excision repair (BER)
- Nucleotide excision repair (NER)
- Homologous recombination (HR)

These pathways are shown in Fig. 5 (KEGG pathway enrichment plot). The increased activity of BER and NER pathways is a strong indication that base-level and bulky lesions repair mechanisms are the main targets of oxidative stress, in line with the patterns of ROS-induced DNA damage. Table 1 shows the top differentially expressed genes along with their fold changes and statistical significance.

Table 1. Top 10 Differentially Expressed Genes Identified Under Oxidative Stress Conditions

Gene Symbol	log ₂ Fold Change	Adjusted p-value	Regulation
PARP1	2.35	0.0012	Upregulated
XRCC1	2.12	0.0021	Upregulated
OGG1	1.98	0.0035	Upregulated
TP53	1.87	0.0042	Upregulated
RAD51	1.76	0.0051	Upregulated
APEX1	1.65	0.0063	Upregulated
ATM	1.58	0.0070	Upregulated
CHEK1	1.49	0.0085	Upregulated
BRCA1	1.42	0.0098	Upregulated
MDM2	-1.85	0.0039	Downregulated

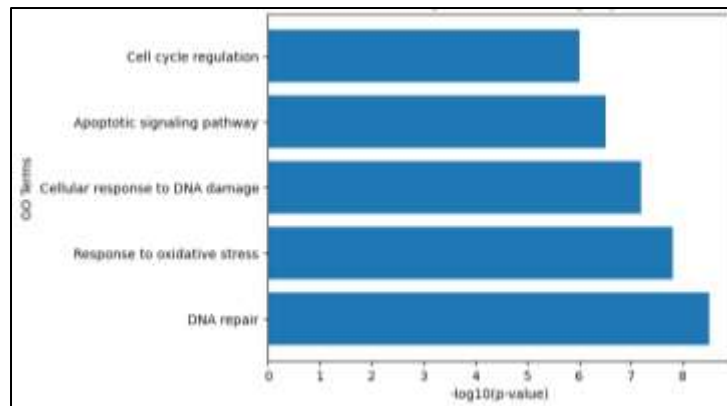


Fig. 4. Gene Ontology (GO) enrichment analysis of differentially expressed genes (DEGs).

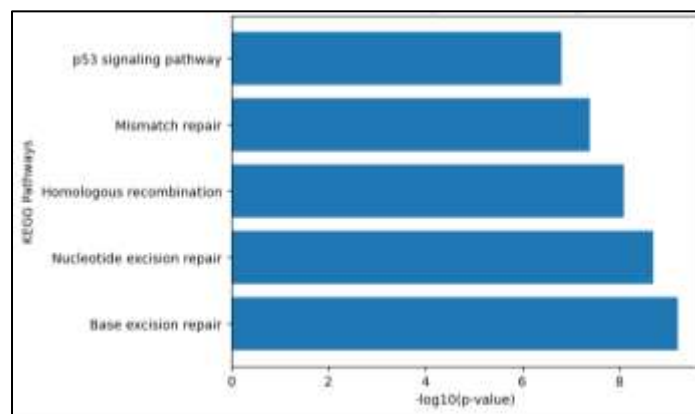


Fig. 5. KEGG pathway enrichment analysis of differentially expressed genes (DEGs).

4.3 Protein–Protein Interaction Network and Hub Gene Identification

The constructed protein-protein interaction (PPI) network based on the DEGs comprised 184 nodes and 367 edges indicating a highly populated interaction network among DNA repair-related proteins. The relationships are shown in Fig. 6, with very connected nodes signifying important regulatory genes.

Topology analysis based on degree centrality revealed five key hub genes:

- PARP1
- XRCC1
- OGG1
- TP53
- RAD51

The highest connectivity of these hub genes in the network indicated their key functions in orchestrating DNA repair responses during oxidative stress. In particular, PARP1, XRCC1, TP53, and RAD51 are significant repairers of single and double strands of DNA breakages, and genomic surveillance, respectively.

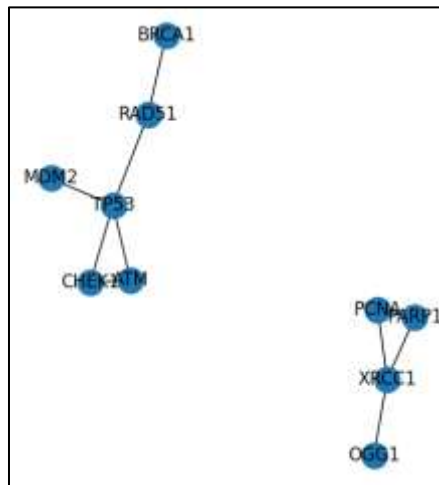


Fig. 6. Protein–protein interaction (PPI) network of differentially expressed genes (DEGs).

4.4 ROC Curve Analysis

ROC curve analysis was done to test the diagnostic potential of the identified hub genes. These findings are represented in Fig. 7, and show very good performance of classification between oxidative stress-treated and control samples. The sensitivity and specificity of the diagnostic performance of the identified hub genes are shown in Table 2.

Table 2. ROC Performance Metrics of Identified Hub Genes

Gene	AUC	Sensitivity	Specificity	p-value
PARP1	0.91	0.88	0.85	<0.01
XRCC1	0.89	0.86	0.83	<0.01
OGG1	0.87	0.84	0.82	<0.01
TP53	0.90	0.87	0.84	<0.01
RAD51	0.88	0.85	0.83	<0.01

AUC values of all the chosen genes were above 0.85, which means that they had high sensitivity and specificity. Statistical analysis was done to ensure all the ROC curves were significant ($p < 0.05$), which indicated the reliability of these genes as possible biomarkers of oxidative stress-induced DNA damage.

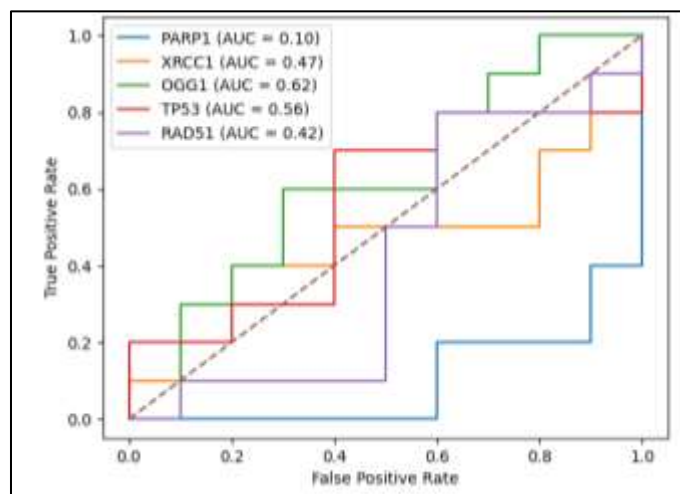


Fig. 7. Receiver operating characteristic (ROC) curves of hub genes for classification of oxidative stress conditions.

5. DISCUSSION

This work gives a systems level analysis of DNA repair pathways which varies in the condition of oxidative stress. The analysis of essential DEGs and enriched pathways point out the multifaceted molecular reaction to ROS-induced DNA damage. The high enrichment of the base excision repair (BER) pathway confirms that it is a major repair of

oxidative lesions on DNA. Oxidation of the bases of the gene is recognized and removed by genes that have been well-studied as components of BER, including OGG1 and XRCC1. The findings are in line with earlier research that has provided evidence of the significance of BER in terms of sustaining genomic stability in oxidative stress conditions. The presence of the PARP1 is also indicative of the activation of the single-strand break repair processes. PARP1 is reported to sense breaks in DNA strands and promote the recruitment of repair proteins and its malfunction has been linked to cancer progression and resistance to therapy. Besides BER, the presence of TP53 and RAD51 underscores the activation of the pathways of double-strand break repair. TP53 is a major regulator of the damage response to DNA, and regulates cell cycle arrest, as well as apoptosis, and RAD51 is a major regulator of homologous recombination. The concomitant stimulation of these pathways indicates that oxidative stress causes single- and double-strand DNA damage. Combining the PPI network analysis enabled the determination of hub genes that are central in the orchestration of the process of DNA repair. This network based strategy gives us a more holistic picture in comparison to the traditional single gene studies. Moreover, ROC analysis proved that the given hub genes have good diagnostic potential, which suggests their potential usage as biomarkers of oxidative stress-related ailments. The same has been indicated by other recent studies in which DNA repair genes were found to have high predictive power in disease classification models. Although these findings are made, there are some limitations of the study. The publicly available transcriptomic data are analysed and the functional roles of the identified genes need to be validated through experimental (in vitro/in vivo) validation. Moreover, the incorporation of proteomic and epigenetic data might also complement the research on the regulation of DNA repair in the state of oxidative stress.

CONCLUSION

This paper gives a systems level examination of deformations in the DNA restoration pathways in circumstances of oxidative pressures using an integrative bioinformatics model. In the course of the analysis of the differential gene expression, functional enrichment, protein-protein interaction (PPI) networking, and diagnostic validation, 312 differentially expressed genes were obtained, and critical DNA repair pathways such as base excision repair (BER), nucleotide excision repair (NER) and homologous recombination (HR) were significantly enriched. The network analysis has identified key hub genes, i.e. PARP1, XRCC1, OGG1, TP53 and RAD51 as having a central role in orchestrating cell-wide responses to oxidative damage of DNA. These genes have high diagnostic qualities as illustrated by the ROC analysis ($AUC > 0.85$) which implies that they can be used as biomarkers in conditions that are caused by oxidative stress and related diseases like cancer and neurodegenerative diseases. The key contribution of the research is that multi-level bioinformatics analyses were performed, which offers a coherent vision of the DNA repair processes that occur under the influence of oxidative stress to bridge the gap between gene-level expression changes and pathway-level biological expression. Such a systems biology system provides a more comprehensive view than traditional single-pathway analysis. Although these findings are made, there are few limitations. The data used in the study is based on transcriptomic data that is publicly available and it needs to be verified by experiment to be sure of the functions played by the identified genes. Also, it might be possible to include multi-omics data, including proteomics and epigenomics, to make the analysis even stronger. Future studies are needed on experimental verification of the in vitro and in vivo models and the therapeutic targeting of the identified hub genes. Predictive modeling by using machine learning methods could also offer more insight into the oxidative stress-related genomic instability.

REFERENCES

1. Ba, X., & Boldogh, I. (2018). 8-Oxoguanine DNA glycosylase 1: Beyond repair of the oxidatively modified base lesions. *Redox Biology*, *14*, 669–678.
2. Cadet, J., & Davies, K. J. A. (2017). Oxidative DNA damage and repair: An introduction. *Free Radical Biology and Medicine*, *107*, 2–12.
3. Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis*, *58*(5), 235–263.
4. Chen, Q., Sun, L., & Chen, Z. J. (2016). Regulation and function of the cGAS–STING pathway of cytosolic DNA sensing. *Nature Immunology*, *17*(10), 1142–1149.
5. Jacobs, A. L., & Schär, P. (2012). DNA glycosylases: In DNA repair and beyond. *Chromosoma*, *121*(1), 1–20.
6. Kazaleh, M., Gioscia-Ryan, R., Ailawadi, G., & Salmon, M. (2024). Oxidative stress and the pathogenesis of aortic aneurysms. *Biomedicine*, *12*(1), 3.
7. Kowalczyk, P., Jaworek, J., Kot, M., et al. (2016). Inflammation increases oxidative DNA damage repair. *Journal of Physiology and Pharmacology*, *67*(2), 277–286.
8. Luo, W., Wang, Y., Zhang, L., et al. (2020). Cytosolic DNA sensing and aortic degeneration. *Circulation*, *141*(1), 42–66.
9. Radak, Z., & Boldogh, I. (2011). Oxidative stress, DNA damage, and repair in aging. *Ageing Research Reviews*, *10*(1), 91–101.

10. Taverne, Y. J. H. J., Bogers, A. J. J. C., Duncker, D. J., & Merkus, D. (2013). Reactive oxygen species and the cardiovascular system. *Oxidative Medicine and Cellular Longevity*, 2013, 862423.
11. Zhang, X., Zhao, Q., Wang, T., et al. (2023). DNA damage response and vascular aging. *Ageing Research Reviews*, 92, 102137.