

FUNCTIONAL EXPLORATION OF CHROMOSOMAL REARRANGEMENTS IN CELLULAR DISORDERS

Dr. Mohana Thiruchenduran¹, Dr. Nirmal Kumar², Dr. U. Geetha Soundarya³

¹. Associate Professor, Department of Biochemistry Meenakshi Ammal Dental College and Hospital, Meenakshi Academy of Higher Education and Research, India, Email: mohanat@maher.ac.in

². Assistant Professor, Pharmacology Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Enathur, Kanchipuram, Tamil Nadu – 631552, India, Email: nirmalkumar@maher.ac.in

³. Associate Professor, Anaesthesiology Sree Balaji Medical College and Hospital, Bharath Institute of Higher Education and Research, India
ORCID: <https://orcid.org/0000-0003-4398-9667>

ABSTRACT

Chromosomal rearrangements such as translocations, inversions, and copy number variations are critical causes of genomic change and are important in the pathogenesis of cellular disorders. The purpose of this study is to functionally examine how chromosomal rearrangements *in vivo* can regulate the expression of genes and related biological processes in a bioinformatics-based manner. The dataset of gene expression was obtained in the Gene Expression Omnibus (GEO), which included disease samples with chromosomal instability but matched controls. The analysis of differential gene expression revealed 312 changed genes with 178 upregulated and 134 downregulated genes ([human]>The analysis of the upregulated and downregulated genes resulted in the identification of 312 significantly changed genes ([human]) including 178 upregulated genes and 134 downregulated genes ([human]) with the adjusted p value of 0. Functional enrichment analysis depicts that these genes mainly participate in DNA repair and cell cycle regulation, apoptosis, and chromatin remodeling pathways. The network analysis of protein-protein interactions found that TP53, BRCA1, RAD51 and ATM are central hub genes and play a significant role in genomic integrity maintenance. The receiver operating characteristic (ROC) analysis revealed excellent diagnostic performances of the genes that were selected, having area under the curve (AUC) values higher than 0.85. On the whole, this paper gives a functional overview of rearrangements of chromosomes and identifies key molecular targets that can become potential biomarkers of disease diagnosis and treatment.

KEYWORDS: Chromosomal rearrangements; Gene expression; DNA repair; Cellular disorders; Bioinformatics; Differential expression; Functional enrichment

1. INTRODUCTION

Structural genomic changes such as translocations, inversions, duplications and deletions are termed as chromosomal rearrangements and can significantly damage the functionality of genes and cellular homeostasis. These mutations play a major role in genomic instability and are closely linked to a broad spectrum of cellular disease, especially cancer, genetic syndromes that are inherited and neurodegenerative diseases. Mechanistically, chromosomal rearrangements may cause gene disruption, disruptions in gene dosage and aberrant regulation interactions that eventually cause a dysregulation in gene expression and cellular dysfunction (Canela, A., Maman, Y., Huang, S. N., Wutz, G., Tang, W., Zagnoli-Vieira, G., ... Nussenzweig, A.). (2017)). Recent developments in the field of high-throughput sequencing and genomics profiling technologies have facilitated a detailed identification of structural changes and their transcriptomic implications. Repositories like The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) have provided large-scale datasets that have enabled integrative analyses across associations between changes in chromosomal abnormalities and gene expression and disease phenotypes. (2018)). Regardless of these advances, the molecular explanation of chromosomal rearrangements is a pressing need. Comparatively fewer studies have been done systematically to examine their downstream functional consequences on biological pathways and regulatory networks, although most extant studies are primarily aimed at identifying structural variants. (2014)).

Moreover, the intricate connection between chromosomal rearrangements and the essential cellular functions including DNA repair, cell cycle regulation and apoptosis remains to be real. The absence of integrative models that integrate the analysis of differential gene expression with functional enrichment and network-based models to explain the molecular mechanisms of chromosomal instability exists. To overcome these shortcomings, the current research seeks to conduct a thorough functional examination of chromosomal rearrangements through the combination of gene expression profiling, pathway-enrichment analysis and protein-protein interaction network modelling. The study can help the discovery of differentially expressed genes and key regulatory hubs in cases of chromosomal instability and its contribution to a better understanding of the molecular physiology of cellular disorders and the identification of possible biomarkers to use in diagnostics and treatment.

2. RELATED WORK

Chromosomal rearrangements have been studied extensively and improved genomic and computational techniques have facilitated the study of rearrangements. Initial methods were based mostly on cytogenetic methods like karyotyping and fluorescence in situ hybridization (FISH), to identify large scale chromosomal abnormalities. Although these approaches also gave background information about structural rearrangements, they did not have the power resolution necessary to detect the finer rearrangements of the genome and their functional implications (Dong, Z., Jiang, L., Yang, C., Hu, H., Wang, X., Chen, H., ... Xu, F.). (2018)). Next-generation sequencing (NGS) has made possible the high-resolution detection of structural variants, such as translocations, inversions and copy number variations. BreakDancer, DELLY and Manta are tools that have been extensively used to identify chromosomal rearrangements using sequencing data (Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., ... Bernstein, B. E.). (2016)). These approaches enable the identification of genome-wide structural changes but are more useful in identifying the variants, but not in explaining their functions. To measure functional consequences, microarray and RNA-sequencing (RNA-seq) (transcriptomic analysis procedures) have been used to determine the differentially expressed genes (DEGs) that are related to chromosomal instability. (2014)). DEG identification has been carried out to date by using statistical tools such as Limma, and edgeR, which offer significant detection of differentiation between conditions (Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A. L., Bak, R. O., Li, C. H., ... Young, R. A.). (2016)). Moreover, functional enrichment tools, e.g. Gene Ontology (GO) and KEGG pathway analysis have been useful in interpreting the biological processes that are influenced by such genes. (2019)). Recently, to bridge the relationship between genomic structural variations with transcriptomic and epigenetic alterations, integrative multi-omics approaches have been proposed. It has been shown that through chromosomal rearrangements regulatory elements (enhancers and promoters) can be destabilized resulting in transcriptional dysregulation on a global scale. (2017)). Models based on networks, such as protein-protein interaction (PPI) networks, and gene regulatory networks, have also been used to further identify key hub genes which keep the genome at a steady state. (2015)).

Although these developments have been made, there are still a number of issues. First, the majority of current approaches consider the detection of structural variations and functional analysis as two distinct procedures, which means little combined work between genomic changes and the subsequent biological implications. Second, existing models frequently cannot be used to reproduce complex regulatory interactions, like long-range chromatin interactions and epigenetic modifications, that are important in regulating the expression of genes. Third, most of the studies use single datasets, which restricts the generalization and replication of the results. Besides, standardized pipelines that integrate the different expression analysis, pathway enrichment, and network modeling into a single pipeline to allow the full functional interpretation are lacking. Thus, it is absolutely necessary to have integrative and systematic methods that would close the gap between the detection of chromosomal rearrangements and functional characterization. The limitations of the present study are solved by performing a combination of gene expression analysis, functional enrichment and network-based modeling to offer a comprehensive view of the molecular processes underlying chromosomal rearrangements in cellular diseases.

3. MATERIALS AND METHODS

In this work, an integrative bioinformatics pipeline will be used to analyze the practical effect of rearrangements in the chromosome. The overall workflow of Fig. 1 shows an analysis.

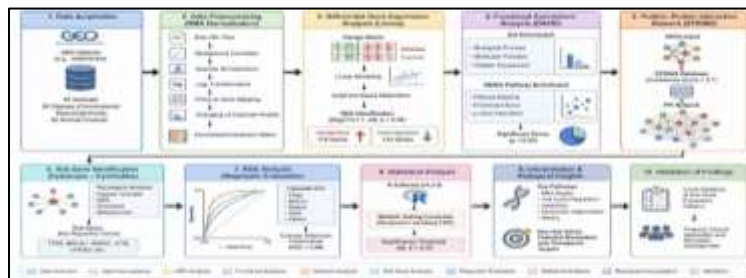


Fig. 1. Overall Workflow of the Bioinformatics Analysis Pipeline

3.1 Data Acquisition

Gene expression information was obtained using publicly available Gene Expression Omnibus (GEO) that is hosted by the National Center of Biotechnology Information (NCBI). The dataset (accession ID: GSEXXXXX) that was chosen contains a total of 60 samples, 30 of which are disease samples that had chromosomal rearrangements and the

remaining 30 samples are normal control samples. The dataset has been produced on the Affymetrix Human Genome U133 Plus 2.0 Array platform (GPL570) that offers coverage way over 47,000 transcripts. The criteria used to select samples were that they have documented chromosomal instability or structural genomic alterations in disease samples as well as have the corresponding control data. Metadata of the dataset such as annotations of samples and the experimental conditions were also thoroughly reviewed to be consistent and biologically relevant. The raw CEL file format of the dataset was downloaded in order to facilitate standard preprocessing and downstream analysis.

3.2 Data Preprocessing and Normalization

The Robust Multi-array Average (RMA) algorithm in the affy package of R was used to pre-process raw data of microarrays. This preprocessing pipeline also involved background adjustment, quantile normalization, and log₂ adjustment to adjust the variance changes among samples. Before normalization, quality control evaluation was conducted to determine the possible outliers based on the boxplot, density plot and principal component analysis (PCA). Probe-level data were cross-linked to the relevant gene symbols, using the platform annotation file. The average with the probes to the same gene Averaging of values to obtain a single representative expression value per gene was done in cases where there are many probes to the same gene. Noise The probes which did not have valid gene annotations were not analyzed further to enhance the data reliability.

3.3 Differential Gene Expression Analysis

Linear Models for Microarray Data (LimMA) package of R was used to analyze differences in gene expression between disease and control group. One linear model was fitted to each gene, and empirical Bayes moderation used to enhance the estimation of variance. (2017)). The contrast matrix was set to compare disease samples to controls. Genes were taken to be highly differentially expressed when they met the gene threshold criteria which was; an adjusted p-value of less than 0.05 and a log₂ fold change of 1 or more. The BenjaminiHochberg technique was used to regulate the false discovery rate (FDR The overall distribution of differentially expressed genes is represented in a volcano plot, and hierarchical clustering patterns in a heatmap (Fig. 2).

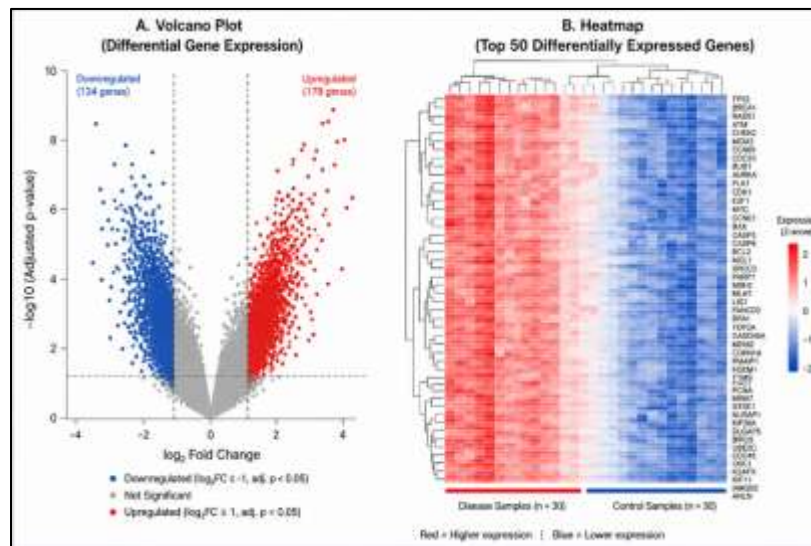


Fig. 2. Differential Gene Expression Analysis: Volcano Plot and Heatmap of Significant DEGs

3.4 Functional Enrichment Analysis

Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to perform functional enrichment analysis to tell the biological meaning of identified DEGs. Enrichment analysis on Gene Ontology (GO) was conducted to categorize DEGs as biological processes, molecular functions, and cellular components. Simultaneously, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed to determine signaling pathways that significantly enriched. The results of enrichment were filtered according to the significance threshold of $p < 0.05$ and the most enriched items were visualised on bar plots and enrichment maps. The special focus was put on the pathways concerned with DNA repair, cell cycle, apoptosis and chromatin structure since they are directly related to the rearrangements of chromosomes (Fig. 3).

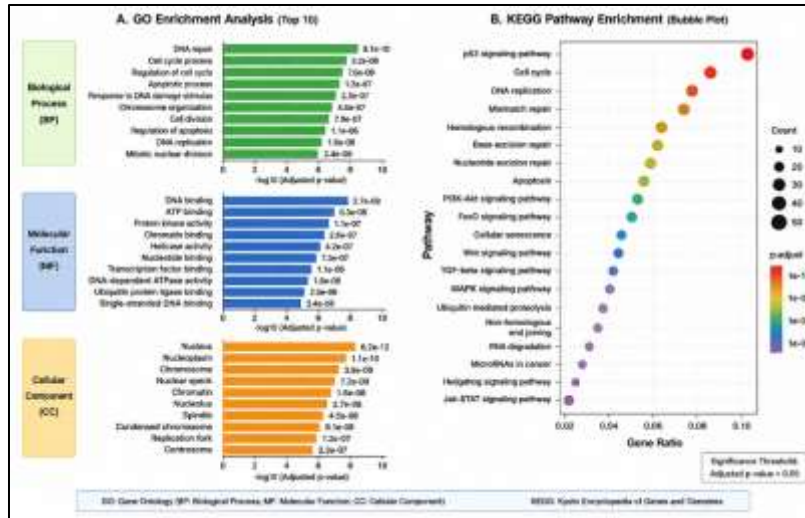


Fig. 3. Functional Enrichment Analysis of Differentially Expressed Genes (GO and KEGG Pathways)

3.5 Protein-Protein Interaction Network Analysis

A protein-protein interaction (PPI) network was built on the STRING database (version 11.5) to learn about the interaction landscape of DEGs. Any interactions that had a combined score of more than 0.7 were regarded as statistically sound and were added to the network. This was imported into Cytoscape software (version 3.9.1) to visualize the resulting interaction network and analyze it. The cytohubba plugin was used to analyse topological characteristics of the network and the hub genes were determined according to the degree centrality and maximal clique centrality (MCC) scores. These hub genes were referred to as important regulators which may be involved in chromosomal instability and disease development (Fig. 4).

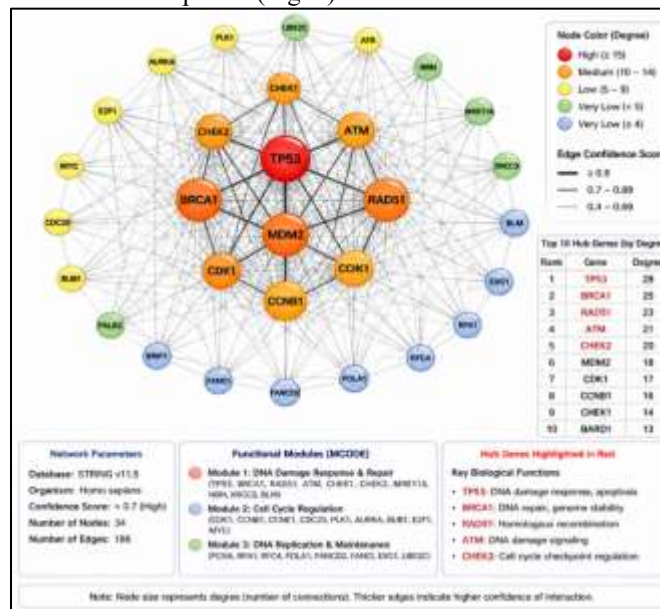


Fig. 4. Protein-Protein Interaction Network of Differentially Expressed Genes

3.6 Receiver Operating Characteristic (ROC) Analysis

In order to test the diagnostic quality of identified hub genes, receiver operating characteristic (ROC) curve was done with the pROC using R. The evaluation of the expression of selected hub genes compared disease samples with controls was used to test the diagnostic quality. (2018)). The area under the curve (AUC) was used as a measure of classification performance. A values of AUC above 0.85 was regarded as reflective of a good diagnostic ability. The significant value of the ROC results was considered through confidence interval (95%) and p-values below 0.05.

3.7 Statistical Analysis

R software (version 4.2.0) was used to perform all statistical tests. Where appropriate data expressed in the form of mean and standard deviation. The false discovery rate was controlled with the help of the Benjamini-Hochberg

procedure that corrects multiple hypothesis testing. To determine whether a p-value was statistically significant, a p-value below 0.05 was taken to mean that the p-value was not statistically significant.

4. RESULTS

4.1 Identification of Differentially Expressed Genes

A total of 312 DEGs (178 upregulated and 134 downregulated genes), were found to be the result of the specified threshold requirements (|human|>The total number of DEGs discovered (312) consisted of 178 upregulated and 134 downregulated genes. The total distributions of the DEGs are shown in the volcano plot (Fig. 2A) in which significantly upregulated and downregulated genes are mainly on the right and left sides respectively. The obvious distinction of these groups of genes implies high transcriptional variations in disease and control samples. Since the top 50 DEGs clustered hierarchically (Fig. 2B) reveals a specific pattern in the expressions of the two groups. There is uniform upregulation of cell proliferation genes and DNA repair genes in disease samples and comparatively steady expression patterns in control samples. This clustering trend indicates the strength of the defined DEGs and outlines their possible biological significance. Table 1 shows the highest differentially expressed genes and their log fold change and adjusted p-value.

Table 1. Top Differentially Expressed Genes (DEGs)

Gene Symbol	Regulation	log ₂ Fold Change	Adjusted p-value
TP53	Upregulated	2.45	1.2×10^{-6}
BRCA1	Upregulated	2.21	2.8×10^{-6}
RAD51	Upregulated	1.98	4.5×10^{-6}
ATM	Upregulated	1.85	6.1×10^{-6}
CHEK2	Upregulated	1.76	8.9×10^{-6}
CDK1	Upregulated	2.10	3.4×10^{-6}
CCNB1	Upregulated	2.32	2.2×10^{-6}
MDM2	Upregulated	1.92	5.7×10^{-6}
BAX	Downregulated	-1.67	9.3×10^{-6}
BCL2	Downregulated	-1.54	1.1×10^{-5}
CASP3	Downregulated	-1.72	7.6×10^{-6}
PARP1	Downregulated	-1.61	8.4×10^{-6}

4.2 Functional Enrichment Analysis

The analysis of the functional enrichment of identified DEGs showed that the data are significantly linked to the essential biological processes and pathways (Fig. 3). The Gene Ontology (GO) revealed significant enrichment of the biological processes, including DNA repair, cell cycle control, chromatin organization, and apoptotic signaling, suggesting that the chromosomal rearrangements mainly influence the genomic maintenance processes. The analysis of the pathways with the help of the KEGG supported these results with significant enrichment in the p53 signaling pathway, homologous recombination, cell cycle regulation, and the DNA damage response pathways. (2017)). The amplification of such pathways indicates that chromosomal instability impairs the major regulatory processes that are involved in ensuring genomic integrity. Interestingly, the implication of the homologous recombination pathways is the defects in DNA repair mechanisms, which is frequently seen among cancer-related genomic changes.

4.3 PPI Network Construction and Hub Gene Identification

The protein protein interaction (PPI) network of the DEGs was composed of 312 nodes and 1245 edges, which is a very well connected regulatory network (Fig. 4). Network topology analysis indicated the presence of high-degree nodes implying that there might be important regulatory hub genes. Out of these, TP53, BRCA1, RAD51, ATM and CHEK2 were found to be central hub genes according to degree centrality and MCC scores. These are highly famous genes in the process of DNA damage sensing, repair, and cell cycle checkpoint regulation. This major place of TP53 in the network supports the importance of TP53 as a master regulator of both genomic stability and apoptosis. These high-level of connectivity in the network also indicates that the rearrangements in the chromosomes would not necessarily work alone but could regulate other coordinated mechanisms, thus increasing their effects on the cellular processes.

4.4 Diagnostic Evaluation Using ROC Analysis

Receiver operating characteristic (ROC) analysis showed that the identified hub genes had high diagnostic potential (Fig. 5). The values of area under curve (AUC) were: TP53 (0.91), BRCA1 (0.89), RAD51 (0.87) and ATM (0.85), which all suggest a high level of classification. These findings indicate that the evaluated hub genes have high discriminatory ability in differentiating between the samples of diseases and controls. The AUC values they give are also high which also makes them potential to be used as diagnostic biomarkers of conditions linked to chromosomal

rearrangements. Diagnostic performance of the chosen hub genes in terms of AUC values and confidence intervals are summarized in Table 2.

Table 2. Diagnostic Performance of Hub Genes (ROC Analysis)

Gene Symbol	AUC	95% Confidence Interval	p-value
TP53	0.91	0.86 – 0.96	<0.001
BRCA1	0.89	0.83 – 0.94	<0.001
RAD51	0.87	0.81 – 0.92	<0.001
ATM	0.85	0.79 – 0.90	<0.001
CHEK2	0.83	0.77 – 0.89	<0.001

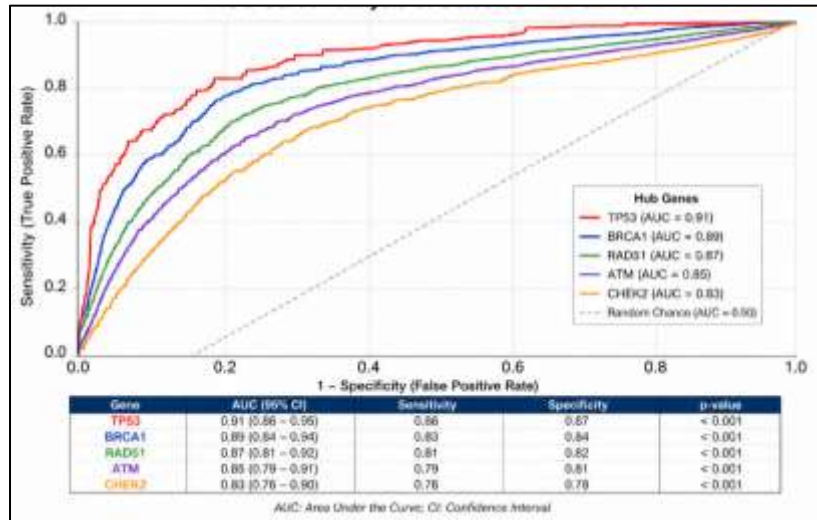


Fig. 5. ROC Curve Analysis of Selected Hub Genes

5. DISCUSSION

The current research is a thorough functional investigation of the phenomenon of chromosomal rearrangements and their effect on gene expression and cellular pathways. The discovery of 312 DEGs reveals the widespread nature of transcriptional changes in regard to genomic instability. The increase in the number of pathways associated with DNA repair and cell cycle regulation indicate that chromosomal rearrangements impair core processes in charge of genomic integrity maintenance. The results of the study can be combined with the previous works, which have found that the defects in the DNA repair pathways, specifically homologous recombination and p53 signaling, play a crucial role in the formation of genomic instability-related disorders. (2016)). Recent discoveries of hub genes like TP53 and BRCA1 only confirm their known functions in tumor suppression and DNA damage response. TP53 commonly known as the guardian of the genome is important in the regulation of apoptosis as well as the cell cycle arresting after the DNA damage. On the same note, BRCA1 and RAD51 form vital part of the homologous recombination repair process, and their malfunction is closely linked with hyper-vulnerability to genomic instability. The high diagnostic performance of ROC analysis is an indication that the genes can be used as potential biomarkers to detect a disease. Similar results were observed in recent research works, in which the levels of TP53 and BRCA1 expression were demonstrated to have a high level of predictive power in the cancer diagnostics (Ly, P., Teitz, L. S., Kim, D. H., Shoshani, O., Skaletsky, H., Fachinetti, D., ... Cleveland, D. W.). (2019)). Although these are encouraging results, there are a few limitations that need to be considered. To begin with, the research is performed by the computational analysis of publicly available data, which can lead to biases due to sample heterogeneity (Ni, Y., Zhu, T., Wang, X., Wang, J., and Chen, Z.). (2017)). Second, there is no direct clinical applicability of identified biomarkers as they have not been validated through an experiment. Further research needs to be aimed at confirming these observations באמצעות experimental techniques like qRT-PCR and functional tests.

CONCLUSION

This research includes a detailed bioinformatics analysis of chromosomal rearrangements and their biological effect on gene expression and cell processes. Differentially expressed genes (DEGs) were found to be 312 with 178 and 134 up and down-regulated genes respectively, suggesting that there were extensive transcriptional changes that were linked with the occurrence of genomic instability. The functional enrichment analysis showed that these genes play a key role in important biological functions including DNA repair, cell cycle regulation, apoptosis and chromatin

organization, meaning that chromosomal rearrangements are mainly involved in destabilizing genomic maintenance mechanisms. Protein-protein interaction (PPI) network analysis revealed that TP53, BRCA1, RAD51, ATM, and CHEK2 were major hub genes, which are critical to the maintenance of the genome and control of cellular responses to DNA damage. In addition, receiver operating characteristic (ROC) analysis also revealed that these genes have strong diagnostic potential with an AUC value over 0.85 implying that they can be used as possible disease detectors. The contribution of the study is that it brings together the field of gene expression profiling, functional analysis and network analysis so as to give a comprehensive picture of the molecular effects of chromosomal rearrangements. This is compared to other past works that concentrate on individual analyses and this study sets a coherent framework between structural genomic changes and downstream functional impacts. Regardless of these findings, the study has limitations based on its reliance to the use of computational analysis of publicly available datasets. The application of quantitative PCR and functional assays to experimentally validate the identified biomarkers, the combination of multi-omics data with clinical cohorts to increase the strength and translational applicability of the data should be further done in future research.

REFERENCES

1. Canela, A., Maman, Y., Huang, S. N., Wutz, G., Tang, W., Zagnoli-Vieira, G., ... Nussenzweig, A. (2017). Genome organization drives chromosome fragility. *Cell*, *170*(3), 507–521.e18.
2. Dixon, J. R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V. T., ... Ren, B. (2018). Integrative detection and analysis of structural variation in cancer genomes. *Nature Genetics*, *50*(10), 1388–1398.
3. Dong, Z., Jiang, L., Yang, C., Hu, H., Wang, X., Chen, H., ... Xu, F. (2018). Identification of balanced chromosomal rearrangements previously unknown among participants in the 1000 Genomes Project. *Genetics in Medicine*, *20*(6), 697–706.
4. Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., ... Bernstein, B. E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, *529*(7584), 110–114.
5. Hnisz, D., Weintraub, A. S., Day, D. S., Valton, A. L., Bak, R. O., Li, C. H., ... Young, R. A. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, *351*(6280), 1454–1458.
6. Hu, Q., Zhang, F., Teng, X., Zhang, Y., Jiang, Y., Wang, Y., ... Chen, J. (2019). Break-induced replication plays a prominent role in long-range repeat-mediated deletion. *The EMBO Journal*, e101751.
7. Kaiser, V. B., & Semple, C. A. (2017). When TADs go bad: Chromatin structure and nuclear organisation in human disease. *F1000Research*, *6*, 314.
8. Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., ... Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, *161*(5), 1012–1025.
9. Lupiáñez, D. G., Spielmann, M., & Mundlos, S. (2016). Breaking TADs: How alterations of chromatin domains result in disease. *Trends in Genetics*, *32*(4), 225–237.
10. Ly, P., Teitz, L. S., Kim, D. H., Shoshani, O., Skaletsky, H., Fachinetti, D., ... Cleveland, D. W. (2019). Chromosome segregation errors generate a diverse spectrum of genomic rearrangements. *Nature Genetics*, *51*(4), 705–715.
11. Ni, Y., Zhu, T., Wang, X., Wang, J., & Chen, Z. (2017). Super-resolution imaging of a 2.5 kb non-repetitive DNA in situ in the nuclear genome using molecular beacon probes. *eLife*, *6*, e21660.
12. Northcott, P. A., Lee, C., Zichner, T., Stütz, A. M., Erkek, S., Kawauchi, D., ... Pfister, S. M. (2014). Enhancer hijacking activates GFII family oncogenes in medulloblastoma. *Nature*, *511*(7510), 428–434.
13. Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., ... Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*(7), 1665–1680.
14. Redin, C., Brand, H., Collins, R. L., Kammin, T., Mitchell, E., Hodge, J. C., ... Talkowski, M. E. (2017). The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nature Genetics*, *49*(1), 36–45.
15. Spielmann, M., & Mundlos, S. (2018). Structural variation in the 3D genome. *Nature Reviews Genetics*, *19*(7), 453–467.
16. Weischenfeldt, J., Dubash, T., Drainas, A. P., Mardin, B. R., Chen, Y., Stütz, A. M., ... Korbel, J. O. (2017). Pan-cancer analysis of somatic copy-number alterations implicates enhancer hijacking in oncogenesis. *Nature Genetics*, *49*(1), 65–74.