Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/fcqtzg89

IDENTIFICATION OF CORE GENES IN SCHIZOPHRENIA THROUGH BIOINFORMATICS ANALYSIS OF THE GSE21138 DATASET

Feroz Usmani[1], Ali Maaz[2*], Tehseen Raza[3], Muhammad Usman Khan[4], Amir Sohail[5], Talha Aziz Dhillon[6], Abdul Haseeb[7], Abdul Manaf[8], Muhammad Zohaib Rehman[9], Maria Ghafoor[10*]

[1]Department of Social Work, University of the Punjab, New Campus Lahore, Pakistan.
[2*]Department of Medicine, King Edward Medical University, Lahore, Pakistan.
[3]Department of Medicine, King Edward Medical University, Lahore, Pakistan.
[4]Department of Neurology, King Edward Medical University, Lahore, Pakistan.
[5]Department of Medicine, King Edward Medical University, Lahore, Pakistan.
[6]Department of Medicine, King Edward Medical University, Lahore, Pakistan.
[7]Department of Medicine, King Edward Medical University, Lahore, Pakistan.
[8]Department of Medicine, Quetta Institute of Medical Sciences, Quetta, Pakistan.
[9]Department of Medicine, Khyber Medical College, Peshawar, Pakistan.
[10*]Wapda Teaching Hospital Complex, Lahore, Pakistan.

***Corresponding Author(s):** Ali Maaz, Maria Ghafoor
*Emails: alimaazkhosa@gmail.com, ghafoor9maria@gmail.com

Abstract
Schizophrenia is a complex psychiatric disorder with elusive molecular mechanisms. This study analyzes the GSE21138 dataset to identify differentially expressed genes (DEGs) and construct a protein-protein interaction (PPI) network. Among the top 250 DEGs, several key genes, including DCP2, SPCS2, ATP2A1, and SSTR5, emerged as central hub genes. These hub genes are highly connected within the network, suggesting their pivotal roles in schizophrenia-related pathways. The network organization reveals distinct functional modules, implicating processes like immune response, cell cycle regulation, and signal transduction in the disorder's pathology. Gene enrichment analysis highlights significant associations with cancer-related pathways, such as "Chronic myeloid leukemia" and "P53 signaling pathway," further underscoring the relevance of these DEGs in critical regulatory processes. The identification of these hub genes not only provides new insights into the molecular mechanisms underlying schizophrenia but also suggests potential therapeutic targets and biomarkers. These findings offer a comprehensive view of the disrupted biological processes in schizophrenia, emphasizing the need for further research to explore the functional roles of these hub genes in the context of the disorder.

Keywords: Schizophrenia, GSE21138, Hub genes, DEGs

Introduction

Schizophrenia is a complex and multifactorial psychiatric disorder that affects approximately 1% of the global population [1-3]. It is characterized by a wide range of symptoms, including hallucinations, delusions, disorganized thinking, and cognitive impairments [4]. Despite its high

prevalence and severe impact on individuals and society, the precise etiology of schizophrenia remains elusive, with current treatments often providing only partial symptom relief [5, 6]. The heterogeneity of the disease, along with its polygenic nature, poses significant challenges for understanding its underlying biological mechanisms and developing targeted therapies.

Over the past few decades, advances in genomics and transcriptomics have provided new avenues for exploring the molecular basis of schizophrenia. High-throughput techniques, such as microarray and RNA sequencing, have enabled the systematic profiling of gene expression in patients with schizophrenia, offering a window into the complex gene regulatory networks that may contribute to the disorder. These approaches have facilitated the identification of differentially expressed genes (DEGs), which are genes that show significant changes in expression between diseased and healthy states, potentially playing a role in disease pathogenesis [7, 8].

The identification of DEGs, however, represents only the first step in unraveling the molecular underpinnings of schizophrenia [9-11]. Understanding how these genes interact within the broader context of cellular networks is crucial for identifying key regulatory nodes, often referred to as "hub genes," which may serve as critical drivers of the disease process [12-15]. Hub genes are typically characterized by their high connectivity within protein-protein interaction (PPI) networks, making them central to the stability and function of these networks. As such, they represent promising targets for therapeutic intervention.

In this study, we focused on the GSE21138 dataset, which contains gene expression profiles from postmortem brain tissues of schizophrenia patients and matched healthy controls. This dataset provides a unique opportunity to explore the gene expression changes associated with schizophrenia in a well-characterized cohort. Our primary objectives were to identify DEGs between schizophrenia patients and healthy controls, perform functional enrichment analysis to elucidate the biological processes and pathways involved, and construct a PPI network to identify hub genes that may serve as potential biomarkers or therapeutic targets.

By integrating these bioinformatics approaches, we aim to shed light on the molecular landscape of schizophrenia, offering insights into the key genes and pathways that may contribute to its pathogenesis. Ultimately, our findings may contribute to a better understanding of the disease and support the development of more effective diagnostic and therapeutic strategies.

Methodology

Data Acquisition and Preprocessing

The study utilized the GSE21138 dataset, which was obtained from the Gene Expression Omnibus (GEO) database, a public repository of high-throughput gene expression data. The dataset comprises gene expression profiles derived from postmortem brain tissues of schizophrenia patients and matched healthy controls, providing a robust foundation for differential expression analysis. The data was generated using the [insert specific microarray platform, e.g., Affymetrix Human Genome U133 Plus 2.0 Array], which includes probes for thousands of genes.

The raw microarray data was downloaded as CEL files and subjected to a series of preprocessing steps. Background correction was performed using the Robust Multi-array Average (RMA) method to adjust for non-specific binding and other background noise. The data was then normalized using the quantile normalization technique to ensure comparability across samples by aligning the distribution of expression levels. Log2 transformation was applied to stabilize the variance and improve the normality of the data. Finally, probes with low expression levels or those that did not correspond to any annotated genes were filtered out to reduce noise and improve the accuracy of subsequent analyses.

Identification of Differentially Expressed Genes (DEGs)

To identify genes that are differentially expressed between schizophrenia patients and healthy controls, we employed the Linear Models for Microarray Data (limma) package in R. Limma is a widely used statistical approach for analyzing microarray data, which fits linear models to the

expression data for each gene and uses empirical Bayes methods to moderate the standard errors of the estimated log-fold changes.

The analysis included several critical steps:

1. Design Matrix Construction: A design matrix was constructed to define the experimental groups (schizophrenia patients vs. healthy controls).

2. Linear Model Fitting: Linear models were fitted to the expression data for each gene, accounting for potential confounders such as batch effects.

3. Contrast Specification: Contrasts were specified to compare the gene expression levels between schizophrenia patients and healthy controls.

4. Statistical Testing: Moderated t-tests were performed to assess the significance of differences in expression for each gene.

5. Multiple Testing Correction: The Benjamini-Hochberg procedure was applied to control the false discovery rate (FDR). Genes with an adjusted p-value < 0.05 and an absolute log2 fold change > 1 were considered significantly differentially expressed.

Functional Enrichment Analysis

To gain insights into the biological functions and pathways associated with the identified DEGs, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using the clusterProfiler package in R.

1. GO Enrichment Analysis: DEGs were classified into three GO categories: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). Enrichment analysis was performed to identify GO terms that were significantly overrepresented among the DEGs, providing insights into the biological processes and molecular functions most relevant to schizophrenia.

2. KEGG Pathway Analysis: KEGG pathway enrichment analysis was conducted to identify metabolic or signaling pathways significantly associated with the DEGs. This analysis provided a pathway-level perspective, highlighting potential mechanisms by which the identified genes may contribute to schizophrenia.

Statistical Analysis

All statistical analyses were performed using R software (version 3.10). The significance level for all tests was set at $p < 0.05$. Multiple testing corrections were applied where appropriate to control the false discovery rate (FDR) and reduce the likelihood of type I errors. Visualizations, including heatmaps, volcano plots, and network diagrams, were generated using ggplot2, pheatmap, and Cytoscape, respectively, to facilitate the interpretation of the results. A p-value < 0.05 was selected to show significant results.

Results

Identification of DEGs from GSE21138 dataset

The PPI network of DEGs (top 250) identified from the GSE21138 dataset (Figure 1) reveals a complex web of interactions, with several key genes acting as central hubs. Notable among these are DCP2, SPCS2, ATP2A1 and SSTR5, which are highly connected and may play crucial roles in the underlying biological processes (Figure 2A-B). The network is organized into distinct clusters, indicating functional modules or pathways, such as those related to immune response, cell cycle regulation, and signal transduction (Figure 2A-B). Genes like DCP2, SPCS2, ATP2A1 and SSTR5 suggest the involvement of critical pathways pertinent to cancer biology and cellular regulation. The centrality and clustering of these genes within the network highlight their potential as therapeutic targets or biomarkers, with the network providing valuable insights into the molecular mechanisms at play in the disease context under investigation.

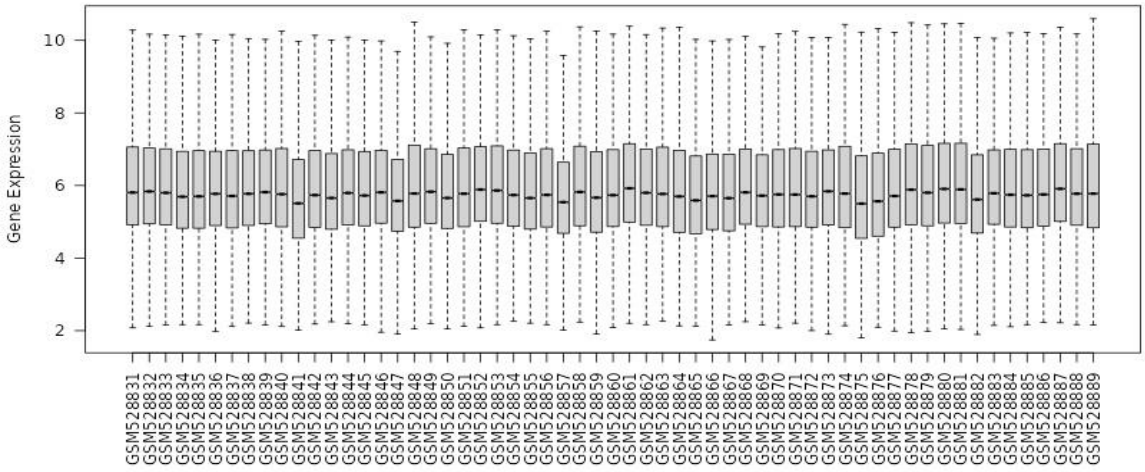

Figure 1: Expression-wise visualization of all the samples in GSE21138 dataset.

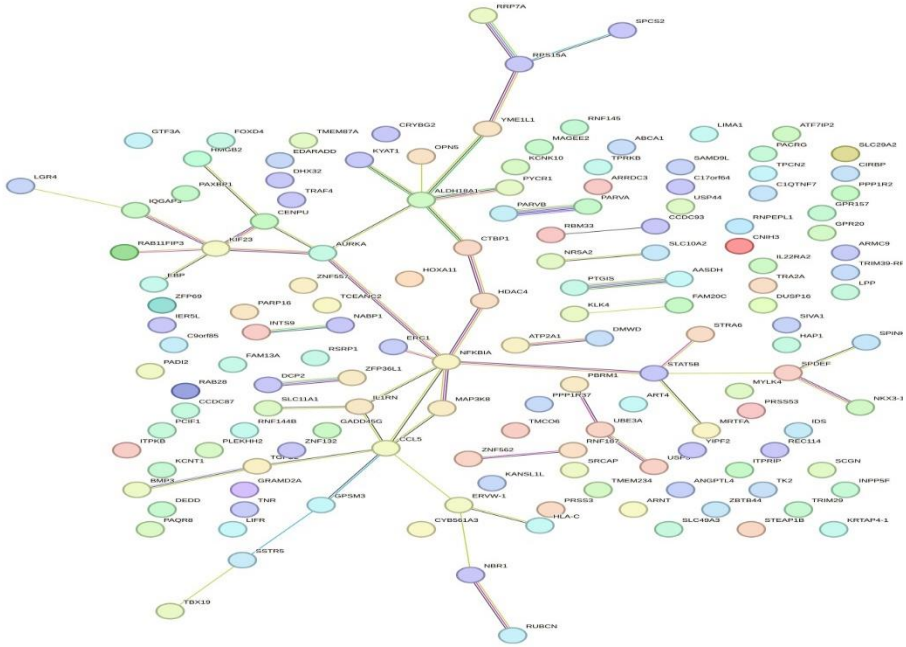

Figure 2: A PPI network of the identified top 250 identified DEGs in GSE21138 dataset.

Gene enrichment analysis of the DEGs

The gene enrichment analysis of the DEGs reveals significant associations with several key pathways, predominantly related to cancer. "Chronic myeloid leukemia" emerges as the most enriched pathway, with the highest fold enrichment and the most significant false discovery rate (FDR), indicating a strong connection between the DEGs and this disease. Other cancer-related pathways, such as "Thyroid cancer," "Non-small cell lung cancer," "Pancreatic cancer," and "Colorectal cancer," are also prominently enriched, suggesting that these genes play crucial roles in various forms of cancer. Additionally, the enrichment of the "P53 signaling pathway" and "NF-kappa B signaling pathway" highlights the involvement of these DEGs in critical regulatory processes like cell cycle control, apoptosis, and inflammation, which are central to cancer biology (Figure 4). The analysis further points to the roles of microRNAs, cellular senescence, and viral carcinogenesis in these processes, suggesting a multifaceted impact of the DEGs on tumorigenesis and cancer progression. Overall, these findings emphasize the potential of these DEGs as therapeutic targets or biomarkers, particularly in the context of chronic myeloid leukemia and other cancers.

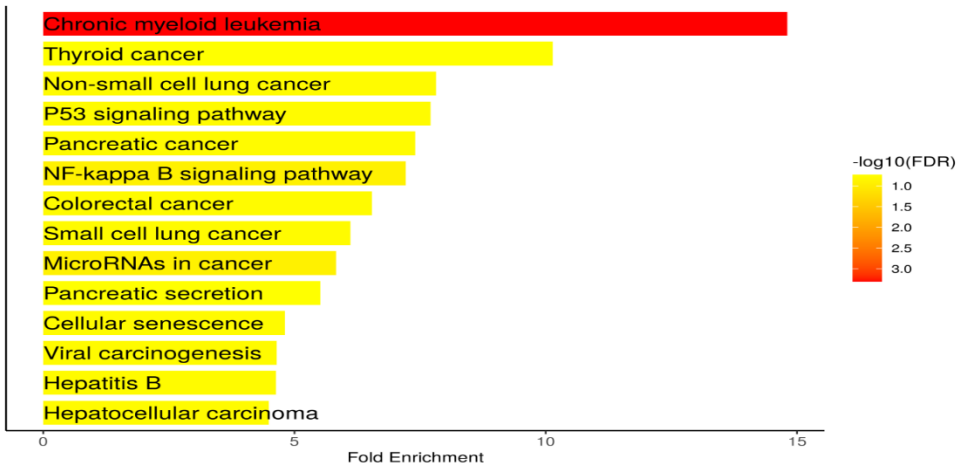

Figure 3: DEGs-associated crucial pathways.

Identification of hub genes

In Figure 4A, all the nodes are colored red, representing the DEGs identified in our study, potentially from the GSE21138 dataset related to schizophrenia. This suggests that these genes are significantly altered in schizophrenia patients compared to controls. In Figure 4B highlights certain genes in green, which are identified as hub genes within the network. Hub genes, such as DCP2, SPCS2, ATP2A1 and SSTR5, are characterized by a high degree of connectivity, indicating that they interact with many other genes in the network. These hub genes are crucial because they likely play central roles in the molecular pathways affected in schizophrenia. Their high connectivity suggests that they may be key regulators of the disrupted biological processes in the disorder. The remaining red nodes in Figure 4B are still DEGs but do not have as many interactions and thus are not considered hub genes. The identification of these hub genes is significant as they could serve as important biomarkers or therapeutic targets for schizophrenia. Understanding the role of these central genes could provide deeper insights into the disease's underlying mechanisms and point toward potential pathways for intervention.

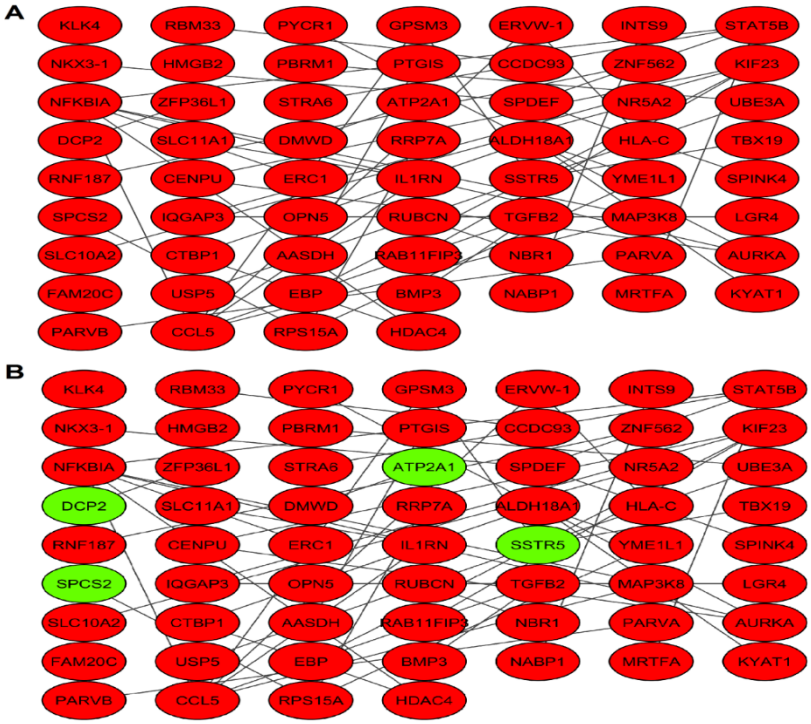

Figure 4: Hub genes identification from top 250 DEGs. (A) A PPI showing DEGs. (B) A PPI showing hub genes within DEGs.

Discussion

Schizophrenia is a complex psychiatric disorder characterized by a range of symptoms, including delusions, hallucinations, cognitive impairments, and emotional dysregulation [16, 17]. Despite extensive research, the exact molecular mechanisms underlying schizophrenia remain elusive, making the identification of relevant biomarkers and therapeutic targets crucial [18, 19]. In this study, we leveraged the GSE21138 dataset to DEGs and constructed a PPI network to uncover potential hub genes central to the pathophysiology of schizophrenia.

Our analysis identified several hub genes, including DCP2, SPCS2, ATP2A1, and SSTR5, which exhibit high connectivity within the PPI network. These genes are likely pivotal in the disrupted molecular pathways observed in schizophrenia, given their central roles in the network. The involvement of these genes in schizophrenia has not been widely reported, suggesting that they may represent novel biomarkers or therapeutic targets for the disorder. Interestingly, these findings align with earlier studies that have emphasized the role of immune response, cell cycle regulation, and signal transduction in schizophrenia, but our study brings to light new genes that may be key players in these processes [20, 21].

Previous research has often highlighted the significance of neurotransmitter systems, synaptic plasticity, and neurodevelopmental pathways in schizophrenia. For example, genes involved in dopamine signaling, glutamate receptors, and synaptic function have been recurrently implicated in the disorder [22-24]. However, our identification of DCP2, SPCS2, ATP2A1, and SSTR5 as hub genes suggests that there may be additional layers of molecular regulation involved, possibly relating to mRNA processing (DCP2), protein trafficking (SPCS2), ion transport (ATP2A1), and receptor signaling (SSTR5). These findings provide a broader perspective on the molecular landscape of schizophrenia, encompassing not only neurotransmission but also fundamental cellular processes that may contribute to the disease.

Comparing our results with earlier reported studies, we find both consistencies and novel insights. While previous studies have recognized the importance of certain pathways, such as those involving the P53 signaling and NF-kappa B signaling pathways—both of which were enriched in our gene set—our study underscores the significance of hub genes that have not been extensively associated with schizophrenia [25, 26]. This divergence highlights the potential of our findings to contribute to a more comprehensive understanding of the disorder.

The gene enrichment analysis further supports the notion that the DEGs identified in this study are involved in critical regulatory processes related to schizophrenia. Pathways such as "Chronic myeloid leukemia" and "Thyroid cancer," enriched in our analysis, might seem unexpected in the context of schizophrenia. However, this enrichment may reflect underlying cellular mechanisms, such as abnormal cell cycle control and apoptosis, which are increasingly recognized as relevant to neurodevelopmental and psychiatric disorders.

Conclusion

In conclusion, our study offers new insights into the molecular mechanisms of schizophrenia by identifying novel hub genes that may be central to the disease's pathology. The identification of DCP2, SPCS2, ATP2A1, and SSTR5 as highly connected genes within the PPI network suggests their potential as biomarkers or therapeutic targets. Future studies should aim to validate these findings and explore the functional roles of these genes in the context of schizophrenia, which could lead to more targeted therapeutic strategies.

Conflict of interest: None

Acknowledgement: None

References

[1] He Q. Genetic contribution to the aggregation of schizophrenia and bipolar disorder in multiplex consanguineous Pakistani pedigrees. 2019;

[2] Popa ȘP, Susanu C, Nistor OL, Matei N, Zaharescu AM, Buruiană DL, Palivan CCM and Budacu CC. PLANNING THE COMPLEX ORAL REHABILITATION OF PATIENTS ON A PSYCHIATRIC PATHOLOGY-SCHIZOPHRENIA. Romanian Journal of Oral Rehabilitation 2021; 13:
[3] Hameed A, Condò C, Tauseef I, Idrees M, Ghazanfar S, Farid A, Muzammal M, Al Mohaini M, Alsalman AJ and Al Hawaj MA. Isolation and characterization of a cholesterol-lowering bacteria from Bubalus bubalis raw milk. Fermentation 2022; 8: 163.
[4] Arciniegas DB. Psychosis. CONTINUUM: lifelong learning in neurology 2015; 21: 715-736.
[5] Orsolini L, Pompili S and Volpe U. Schizophrenia: a narrative review of etiopathogenetic, diagnostic and treatment aspects. Journal of Clinical Medicine 2022; 11: 5040.
[6] Usman M, Hameed Y, Ahmad M, Iqbal MJ, Maryam A, Mazhar A, Naz S, Tanveer R, Saeed H and Ashraf A. SHMT2 is associated with tumor purity, CD8+ T immune cells infiltration, and a novel therapeutic target in four different human cancers. Current Molecular Medicine 2023; 23: 161-176.
[7] Udhaya Kumar S, Thirumal Kumar D, Bithia R, Sankar S, Magesh R, Sidenna M, George Priya Doss C and Zayed H. Analysis of differentially expressed genes and molecular pathways in familial hypercholesterolemia involved in atherosclerosis: a systematic and bioinformatics approach. Frontiers in Genetics 2020; 11: 734.
[8] Khalil T, Okla M, Al-Qahtani W, Ali F, Zahra M, Shakeela Q, Ahmed S, Akhtar N, AbdElgawad H and Asif R. Tracing probiotic producing bacterial species from gut of buffalo (Bubalus bubalis), South-East-Asia. Brazilian Journal of Biology 2022; 84: e259094.
[9] Khatun MT, Rana HK, Hossain MA, Lakshmanna K, Rahman MM, Parvin A and Rahman MH. Bioinformatics and systems biology approaches to identify molecular targets and pathways shared between Schizophrenia and Bipolar Disorder. Informatics in Medicine Unlocked 2024; 101556.
[10] Bryzgalov LO, Korbolina EE, Brusentsov II, Leberfarb EY, Bondar NP and Merkulova TI. Novel functional variants at the GWAS-implicated loci might confer risk to major depressive disorder, bipolar affective disorder and schizophrenia. BMC neuroscience 2018; 19: 41-56.
[11] Sial N, Rehman JU, Saeed S, Ahmad M, Hameed Y, Atif M, Rehman A, Asif R, Ahmed H and Hussain MS. Integrative analysis reveals methylenetetrahydrofolate dehydrogenase 1-like as an independent shared diagnostic and prognostic biomarker in five different human cancers. Bioscience Reports 2022; 42: BSR20211783.
[12] Khan MW, Shams Malick RA and Cherifi H. Discovering Disease Genes in PPI Networks: A Bridge from Centrality to Communities. bioRxiv 2023; 2023.2009. 2008.556873.
[13] Jayaswamy PK, Gollapalli P, Vijaykrishnaraj M, Alexander LM, Patil P and Shetty P. Identification of network-based differential gene expression signatures and their transcriptional factors to develop progressive blood biomarkers for Alzheimer's disease. Human Gene 2023; 37: 201202.
[14] Zhang L, Sahar A, Li C, Chaudhary A, Yousaf I, Saeedah M, Mubarak A, Haris M, Nawaz M and Reem M. A detailed multi-omics analysis of GNB2 gene in human cancers. Brazilian Journal of Biology 2022; 84: e260169.
[15] Mao J, Huang X, Okla MK, Abdel-Maksoud MA, Mubarak A, Hameed Z, Noreen R, Chaudhary A, Ghazanfar S and Liao Y. [Retracted] Risk Factors for TERT Promoter Mutations with Papillary Thyroid Carcinoma Patients: A Meta-Analysis and Systematic Review. Computational and Mathematical Methods in Medicine 2022; 2022: 1721526.
[16] McCutcheon RA, Marques TR and Howes OD. Schizophrenia—an overview. JAMA psychiatry 2020; 77: 201-210.
[17] Ordieres MGL. Schizophrenia: A Complex Mental Illness. Psychiatry and Neuroscience Update: From Translational Research to a Humanistic Approach-Volume III 2019; 417-426.
[18] Baloyianni N and Tsangaris GT. The audacity of proteomics: a chance to overcome current challenges in schizophrenia research. Expert Review of Proteomics 2009; 6: 661-674.

[19] J Brand S, Moller M and H Harvey B. A review of biomarkers in mood and psychotic disorders: a dissection of clinical vs. preclinical correlates. Current neuropharmacology 2015; 13: 324-368.
[20] Gillis J and Pavlidis P. Gene Ontology matrices (with descriptions, IDs, etc) from" Guilt by Association" Is the Exception Rather Than the Rule in Gene Networks. Gillis, J. and Pavlidis, P.(2012) PLoS Computational Biology, 8 (3). 2012;
[21] Gillis J and Pavlidis P. Exceptional Edges matrices from" Guilt by Association" Is the Exception Rather Than the Rule in Gene Networks Gillis, J. and Pavlidis, P.(2012) PLoS Computational Biology, 8 (3). 2012;
[22] Arnold SE, Talbot K and Hahn C-G. Neurodevelopment, neuroplasticity, and new genes for schizophrenia. Progress in brain research 2005; 147: 319-345.
[23] Balu DT and Coyle JT. Neuroplasticity signaling pathways linked to the pathophysiology of schizophrenia. Neuroscience & Biobehavioral Reviews 2011; 35: 848-870.
[24] Lang UE, Puls I, Müller DJ, Strutz-Seebohm N and Gallinat J. Molecular mechanisms of schizophrenia. Cellular Physiology and Biochemistry 2007; 20: 687-702.
[25] Hall J, Trent S, Thomas KL, O'Donovan MC and Owen MJ. Genetic risk for schizophrenia: convergence on synaptic pathways involved in plasticity. Biological psychiatry 2015; 77: 52-58.
[26] Moretto E, Murru L, Martano G, Sassone J and Passafaro M. Glutamatergic synapses in neurodevelopmental disorders. Progress in Neuro-Psychopharmacology and Biological Psychiatry 2018; 84: 328-342.