Transcriptomic Analysis of the Receptors Implicated in Schizophenia

Anjali Jha[1(✉)], I. A. Shylesh Murthy[2], and Preenon Bagchi[1,2,3]

[1] Padmashree Institute of Management and Sciences, Bengaluru, India
meanjalijha1@gmail.com
[2] Vasishth Academy of Advanced Studies and Research (Sarvasumana Association), Bengaluru, India
[3] MGM Institute of Biosciences and Technology, Aurangabad, India

Abstract. Schizophrenia is a chronic, severe mental disorder that affects the way a person thinks, acts, expresses emotions, perceives reality, and relates to others. Metatranscriptomics analysis enables understanding of how the micro biome responds to the environment by studying the functional analysis of genes expressed by the microbiome. Galaxy was used for Metatranscriptomics analysis. Schizophrenia RNA-seq data SRR15629349, SRR15629350, SRR15629351 and SRR15629352 were retrieved from SRA database. The quality of the reads using the FASTQC tool before and after trimming was assessed. The results of FASTQC tool of all the sequences are combined to MultiQC. As per MultiQC results we see that bad, duplicate and overlapping reads are removed by Trimmomatic tool. De novo transcriptome reconstruction with RNA-Seq pipeline was used to identify the expressed genes in our genome. Further Transcriptome assembly was done. The differential gene expression analysis of our sequence was using FeatureCounts to count reads per transcript which was followed by DESeq2 to generate normalized transcript counts.

Keywords: Schizophrenia · Metatranscriptomics analysis · RNA-seq · data · SRA · Krona

1 Introduction

Next-generation sequencing (NGS) is an advanced version of non-Sanger-based sequencing technology that offers ultra-high throughput, scalability, and speed. Galaxy is an open source, web-based platform for next generation computational biomedical research. Metatranscriptomics analysis enables understanding of how the microbiome responds to the environment by studying the functional analysis of genes expressed by the microbiome. The genes from the Metagenomic analysis were transcribed from functional data, active metabolic pathways can be identified in our selected microbiome community. Due to arrival of Next generation sequencing the detection of these pathogenic variants genes became possible. Next generation sequencing technology initially was concerned with studying genomes that were tractable from the standpoint size and repetitive content

R. Somashekhar et al. (Eds.): ICBDS 2022, AHSR 58, pp. 31–49, 2023.
https://doi.org/10.2991/978-94-6463-164-7_4

and with characterization of multiple genes associated with the disease. The technology used to determine the order of nucleotides or targeted regions of DNA or RNA. Here raw data generation is no longer a rate limiting factor in genome scale studies. Galaxy an open source platform for NGS data analysis. The pipeline used here is metatranscriptomics analysis which enables us to understand how the microbiome respond to the host by studying functional analysis of genes expressed in schizophrenia which is a chronic, severe mental disorder that affects the way a person thinks, acts, expresses emotions, perceives reality, and relates to others [1–5]. The disorder results in some combination of hallucinations in combination with delusions with extremely disordered thinking and behavior that impairs disabling daily functioning [1, 2, 6–8]. People affected with schizophrenia require lifelong treatment, psychotherapy and coordinated speciality care services. The exact causal factor of schizophrenia is unknown, but a combination of genetics, environment and altered brain chemistry and structure plays vital role [9–15].

2 Materials and Methods

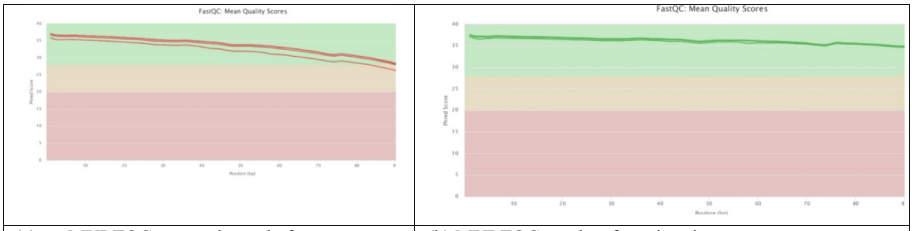
2.1 Transcriptome Reconstruction with RNA-Seq

Galaxy tutorial by Mallory Freeberg, Mo Heydarian, titled "De novo transcriptome reconstruction with RNA-Seq" available at Galaxy Training Materials is used for analysis of neuroaids' RNA-seq. Schizophrenia RNA-seq data SRR15629349, SRR15629350, SRR15629351 and SRR15629352 were retrieved from SRA database. We assessed the quality of the reads using the FASTQC tool before and after trimming. The sequences were trimmed using Trimmomatic tool for the low quality bases from the reads and repeating & overlapping reads, if any, increase mapping efficiency.

To make sense of the reads, their positions within human genome we perform aligning or 'mapping' the reads to with the reference human genome using HISAT2 tool. Further, we perform *de novo* transcriptome reconstruction wherein we determine the transcript structures that are represented by the aligned reads. This unbiased approach permits the comprehensive identification of all transcripts present in our sequence. Though, common gene/transcript databases are quite large, they are not comprehensive, hence, the *de novo* transcriptome reconstruction approach ensures complete transcriptome(s) identification from the experimental samples. This is done using the tool Stringtie.

Next, to compare the abundance of transcripts between different cellular states, we need to quantify the number of reads per transcript, i.e., counting reads in genomic features. This is done by tool FeatureCounts.

Finally we perform differential gene expression (DGE) testing or Transcript expression which is estimated from read counts. This is absolutely essential to obtaining accurate results. We use DESeq2 tool for DGE where we give the output of FeatureCounts as input and the tool applies size factor normalization, i.e., computation for each gene of the geometric mean of read counts across all samples, division of every gene count by the geometric mean and the usage of the median of these ratios as our sequences' size factor for normalization. Further we do volcano plot (is a type of scatterplot) to identify our expressed genes.



(a) MULTQC results before trimming

(b) MULTQC results after trimming

Fig. 1. MultiQC results before and after trimming

3 Results and Discussion

Fastq sequences of the genome identified in Schizophrenia are SRR15629349, SRR15629350, SRR15629351 and SRR15629352. Sequence quality before and after trimming are noted using FASTQC tool and results of FASTQC are merged using MULTIQC tool for better visualization as given in Fig. 1.

As per Fig. 1(b) we can see that the low quality bases from the reads, repeating & overlapping reads were trimmed or removed by the Trimmomatic tool as results seen in 1(b) is better than 1(a). Further the sequences were mapped with human reference genome using HISAT2 tool which is an accurate and fast tool for mapping spliced reads to a genome.

Next, we generated the transcriptomes with Stringtie representing each of the neuroAIDS's RNA-seq libraries in the absence of a reference transcriptome. Further, we made a transcriptome database by using the tool Stringtie - Merge to combine redundant transcript structures across our sequences with our human RefSeq reference as seen in Table 1. Next, GFFcompare annotated the transcripts of our newly created transcriptome to make us known the relationship of each transcript to the human RefSeq reference as seen in Table 2.

Table 1. StringTie merge results (1st few lines of the output)

Seqname	Source	Feature	Start	End	Score	Strand	Frame	Attributes
# StringTie version 2.1.7								
chr1	StringTie	transcript	11874	15876	1000	+	.	gene_id "MSTRG.1"; transcript_id "MSTRG.1.1";
chr1	StringTie	exon	11874	12227	1000	+	.	gene_id "MSTRG.1"; transcript_id "MSTRG.1.1"; exon_number "1";
chr1	StringTie	exon	12613	12721	1000	+	.	gene_id "MSTRG.1"; transcript_id "MSTRG.1.1"; exon_number "2";
chr1	StringTie	exon	13221	14829	1000	+	.	gene_id "MSTRG.1"; transcript_id "MSTRG.1.1"; exon_number "3";
chr1	StringTie	exon	14970	15876	1000	+	.	gene_id "MSTRG.1"; transcript_id "MSTRG.1.1"; exon_number "4";
chr1	StringTie	transcript	11874	14829	1000	+	.	gene_id "MSTRG.1"; transcript_id "NR_046018.2"; ref_gene_id "NR_046018.2";

(*continued*)

Table 1. (*continued*)

Seqname	Source	Feature	Start	End	Score	Strand	Frame	Attributes
chr1	StringTie	exon	11874	12227	1000	+	.	gene_id "MSTRG.1"; transcript_id "NR_046018.2"; exon_number "1"; ref_gene_id "NR_046018.2";
chr1	StringTie	exon	12613	12721	1000	+	.	gene_id "MSTRG.1"; transcript_id "NR_046018.2"; exon_number "2"; ref_gene_id "NR_046018.2";
chr1	StringTie	exon	13221	14829	1000	+	.	gene_id "MSTRG.1"; transcript_id "NR_046018.2"; exon_number "3"; ref_gene_id "NR_046018.2";
chr1	StringTie	transcript	14362	29442	1000	−	.	gene_id "MSTRG.2"; transcript_id "NR_024540.1"; ref_gene_id "NR_024540.1";
chr1	StringTie	exon	14362	14829	1000	−	.	gene_id "MSTRG.2"; transcript_id "NR_024540.1"; exon_number "1"; ref_gene_id "NR_024540.1";
chr1	StringTie	exon	14970	15038	1000	−	.	gene_id "MSTRG.2"; transcript_id "NR_024540.1"; exon_number "2"; ref_gene_id "NR_024540.1";

(*continued*)

Table 1. (*continued*)

Seqname	Source	Feature	Start	End	Score	Strand	Frame	Attributes
chr1	StringTie	exon	15796	15947	1000	–	.	gene_id "MSTRG.2"; transcript_id "NR_024540.1"; exon_number "3"; ref_gene_id "NR_024540.1";
chr1	StringTie	exon	16607	16765	1000	–	.	gene_id "MSTRG.2"; transcript_id "NR_024540.1"; exon_number "4"; ref_gene_id "NR_024540.1";
chr1	StringTie	exon	16858	17055	1000	–	.	gene_id "MSTRG.2"; transcript_id "NR_024540.1"; exon_number "5"; ref_gene_id "NR_024540.1";
chr1	StringTie	exon	17233	17368	1000	–	.	gene_id "MSTRG.2"; transcript_id "NR_024540.1"; exon_number "6"; ref_gene_id "NR_024540.1";
chr1	StringTie	exon	17606	17742	1000	–	.	gene_id "MSTRG.2"; transcript_id "NR_024540.1"; exon_number "7"; ref_gene_id "NR_024540.1";

(*continued*)

Table 1. (*continued*)

Seqname	Source	Feature	Start	End	Score	Strand	Frame	Attributes
chr1	StringTie	exon	17915	18061	1000	–	.	gene_id “MSTRG.2”; transcript_id “NR_024540.1”; exon_number “8”; ref_gene_id “NR_024540.1”;
chr1	StringTie	exon	18268	18366	1000	–	.	gene_id “MSTRG.2”; transcript_id “NR_024540.1”; exon_number “9”; ref_gene_id “NR_024540.1”;
chr1	StringTie	exon	24738	24891	1000	–	.	gene_id “MSTRG.2”; transcript_id “NR_024540.1”; exon_number “10”; ref_gene_id “NR_024540.1”;
chr1	StringTie	exon	29321	29442	1000	–	.	gene_id “MSTRG.2”; transcript_id “NR_024540.1”; exon_number “11”; ref_gene_id “NR_024540.1”;
chr1	StringTie	transcript	14407	29442	1000	–	.	gene_id “MSTRG.2”; transcript_id “MSTRG.2.2”;

(*continued*)

Table 1. (*continued*)

Seqname	Source	Feature	Start	End	Score	Strand	Frame	Attributes
chr1	StringTie	exon	14407	14829	1000	–	.	gene_id “MSTRG.2”; transcript_id “MSTRG.2.2”; exon_number “1”;
chr1	StringTie	exon	14970	15038	1000	–	.	gene_id “MSTRG.2”; transcript_id “MSTRG.2.2”; exon_number “2”;
chr1	StringTie	exon	15796	16765	1000	–	.	gene_id “MSTRG.2”; transcript_id “MSTRG.2.2”; exon_number “3”;
chr1	StringTie	exon	16858	17055	1000	–	.	gene_id “MSTRG.2”; transcript_id “MSTRG.2.2”; exon_number “4”;
chr1	StringTie	exon	17233	17368	1000	–	.	gene_id “MSTRG.2”; transcript_id “MSTRG.2.2”; exon_number “5”;
chr1	StringTie	exon	17606	17742	1000	–	.	gene_id “MSTRG.2”; transcript_id “MSTRG.2.2”; exon_number “6”;
chr1	StringTie	exon	17915	18061	1000	–	.	gene_id “MSTRG.2”; transcript_id “MSTRG.2.2”; exon_number “7”;

Table 2. GffCompare annotated transcripts (1st few lines of the output)

Seqname	Source	Feature	Start	End	Score	Strand	Frame	Attributes
chr1	StringTie	transcript	11874	14829	.	+	.	transcript_id "NR_046018.2"; gene_id "MSTRG.1"; gene_name "NR_046018.2"; xloc "XLOC_000001"; ref_gene_id "NR_046018.2"; contained_in "MSTRG.1.1"; cmp_ref "NR_046018.2"; class_code " = "; tss_id "TSS1";
chr1	StringTie	exon	11874	12227	.	+	.	transcript_id "NR_046018.2"; gene_id "MSTRG.1"; exon_number "1";
chr1	StringTie	exon	12613	12721	.	+	.	transcript_id "NR_046018.2"; gene_id "MSTRG.1"; exon_number "2";
chr1	StringTie	exon	13221	14829	.	+	.	transcript_id "NR_046018.2"; gene_id "MSTRG.1"; exon_number "3";
chr1	StringTie	transcript	11874	15876	.	+	.	transcript_id "MSTRG.1.1"; gene_id "MSTRG.1"; gene_name "NR_046018.2"; xloc "XLOC_000001"; cmp_ref "NR_046018.2"; class_code "k"; tss_id "TSS1";
chr1	StringTie	exon	11874	12227	.	+	.	transcript_id "MSTRG.1.1"; gene_id "MSTRG.1"; exon_number "1";

(*continued*)

Table 2. (*continued*)

Seqname	Source	Feature	Start	End	Score	Strand	Frame	Attributes
chr1	StringTie	exon	12613	12721	.	+	.	transcript_id "MSTRG.1.1"; gene_id "MSTRG.1"; exon_number "2";
chr1	StringTie	exon	13221	14829	.	+	.	transcript_id "MSTRG.1.1"; gene_id "MSTRG.1"; exon_number "3";
chr1	StringTie	exon	14970	15876	.	+	.	transcript_id "MSTRG.1.1"; gene_id "MSTRG.1"; exon_number "4";
chr1	StringTie	transcript	16018	29425	.	+	.	transcript_id "MSTRG.3.1"; gene_id "MSTRG.3"; gene_name "NR_024540.1"; xloc "XLOC_000002"; cmp_ref "NR_024540.1"; class_code "s"; tss_id "TSS2";
chr1	StringTie	exon	16018	16765	.	+	.	transcript_id "MSTRG.3.1"; gene_id "MSTRG.3"; exon_number "1";
chr1	StringTie	exon	16858	17055	.	+	.	transcript_id "MSTRG.3.1"; gene_id "MSTRG.3"; exon_number "2";
chr1	StringTie	exon	17606	17742	.	+	.	transcript_id "MSTRG.3.1"; gene_id "MSTRG.3"; exon_number "3";
chr1	StringTie	exon	17915	18061	.	+	.	transcript_id "MSTRG.3.1"; gene_id "MSTRG.3"; exon_number "4";
chr1	StringTie	exon	18268	18369	.	+	.	transcript_id "MSTRG.3.1"; gene_id "MSTRG.3"; exon_number "5";

(*continued*)

Table 2. (*continued*)

Seqname	Source	Feature	Start	End	Score	Strand	Frame	Attributes
chr1	StringTie	exon	29321	29425	.	+	.	transcript_id "MSTRG.3.1"; gene_id "MSTRG.3"; exon_number "6";
chr1	StringTie	transcript	17320	29425	.	+	.	transcript_id "MSTRG.3.5"; gene_id "MSTRG.3"; gene_name "NR_024540.1"; xloc "XLOC_000002"; cmp_ref "NR_024540.1"; class_code "s"; tss_id "TSS3";
chr1	StringTie	exon	17320	17368	.	+	.	transcript_id "MSTRG.3.5"; gene_id "MSTRG.3"; exon_number "1";
chr1	StringTie	exon	17606	17742	.	+	.	transcript_id "MSTRG.3.5"; gene_id "MSTRG.3"; exon_number "2";
chr1	StringTie	exon	17915	18061	.	+	.	transcript_id "MSTRG.3.5"; gene_id "MSTRG.3"; exon_number "3";
chr1	StringTie	exon	18268	18366	.	+	.	transcript_id "MSTRG.3.5"; gene_id "MSTRG.3"; exon_number "4";
chr1	StringTie	exon	24738	24891	.	+	.	transcript_id "MSTRG.3.5"; gene_id "MSTRG.3"; exon_number "5";
chr1	StringTie	exon	29321	29425	.	+	.	transcript_id "MSTRG.3.5"; gene_id "MSTRG.3"; exon_number "6";

Next, we used FeatureCounts tool which counted the reads aligning in exons of our GFFCompare generated transcriptome database (Table 3).

Table 3. FeatureCounts output for sequence SRR15629352

Status	**HISAT2 on data 10 and data 9: aligned reads (BAM)**
Status	HISAT2 on data 10 and data 9: aligned reads (BAM)
Assigned	1251
Unassigned_Unmapped	3003267
Unassigned_Read_Type	0
Unassigned_Singleton	0
Unassigned_MappingQuality	0
Unassigned_Chimera	47341809
Unassigned_FragmentLength	0
Unassigned_Duplicate	0
Unassigned_MultiMapping	747243
Unassigned_Secondary	0
Unassigned_NonSplit	0
Unassigned_NoFeatures	12302
Unassigned_Overlapping_Length	0
Unassigned_Ambiguity	1331

Table 4. FeatureCounts output for sequence SRR15629351

Status	**HISAT2 on data 14 and data 13: aligned reads (BAM)**
Status	HISAT2 on data 14 and data 13: aligned reads (BAM)
Assigned	1023
Unassigned_Unmapped	2190827
Unassigned_Read_Type	0
Unassigned_Singleton	0
Unassigned_MappingQuality	0
Unassigned_Chimera	46807736
Unassigned_FragmentLength	0
Unassigned_Duplicate	0
Unassigned_MultiMapping	737672
Unassigned_Secondary	0
Unassigned_NonSplit	0

(*continued*)

Table 4. (*continued*)

Status	HISAT2 on data 14 and data 13: aligned reads (BAM)
Unassigned_NoFeatures	9806
Unassigned_Overlapping_Length	0
Unassigned_Ambiguity	1021

Table 5. FeatureCounts output for sequence SRR15629350

Status	HISAT2 on data 18 and data 17: aligned reads (BAM)
Status	HISAT2 on data 18 and data 17: aligned reads (BAM)
Assigned	1137
Unassigned_Unmapped	3577510
Unassigned_Read_Type	0
Unassigned_Singleton	0
Unassigned_MappingQuality	0
Unassigned_Chimera	44632568
Unassigned_FragmentLength	0
Unassigned_Duplicate	0
Unassigned_MultiMapping	645854
Unassigned_Secondary	0
Unassigned_NonSplit	0
Unassigned_NoFeatures	8733
Unassigned_Overlapping_Length	0
Unassigned_Ambiguity	1263

Table 6. FeatureCounts output for sequence SRR15629349

Status	HISAT2 on data 53 and data 52: aligned reads (BAM)
Status	HISAT2 on data 53 and data 52: aligned reads (BAM)
Assigned	1472
Unassigned_Unmapped	3991746
Unassigned_Read_Type	0
Unassigned_Singleton	0

(*continued*)

Table 6. (*continued*)

Status	HISAT2 on data 53 and data 52: aligned reads (BAM)
Unassigned_MappingQuality	0
Unassigned_Chimera	56908058
Unassigned_FragmentLength	0
Unassigned_Duplicate	0
Unassigned_MultiMapping	811032
Unassigned_Secondary	0
Unassigned_NonSplit	0
Unassigned_NoFeatures	10316
Unassigned_Overlapping_Length	0
Unassigned_Ambiguity	1438

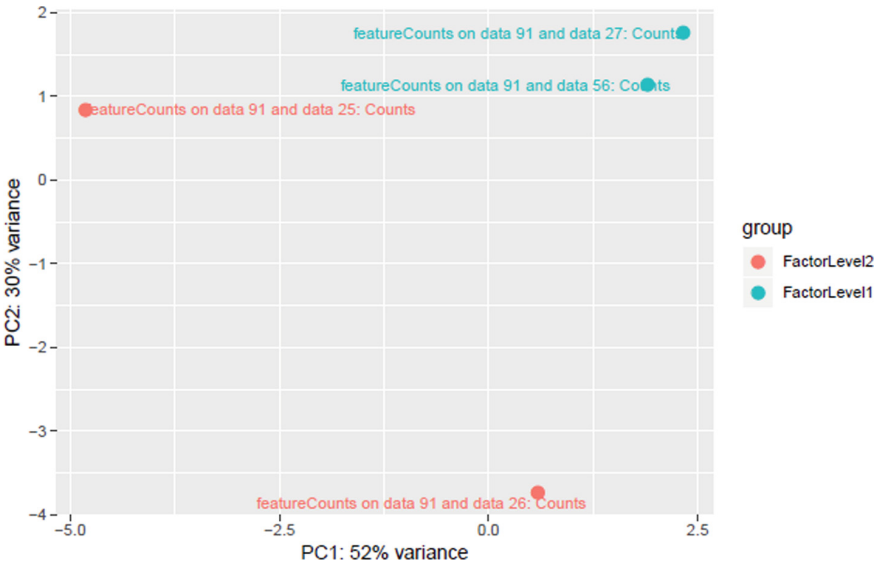


Fig. 2. PCA analysis of sequences

Further, by using the DESeq2 tool we performed differential gene expression analysis using the read counts produced by FeatureCounts (Figs. 2, 3, 4, 5, and 6).

Since, principal components are created in order of the amount of variation they cover: PC1 captures the most variation and PC2 — the second most. In out plot (Fig. 2) we see PC1 has 52% variance. The points in the top (data 27 corresponding to SRR15629350 and data 56 SRR15629348) may account for **largest possible variance** in our data set. Our PC2 has 30% variance may account for data 25 corresponding to SRR15629352 and data 26 corresponding to SRR15629351 (Tables 4, 5, and 6).

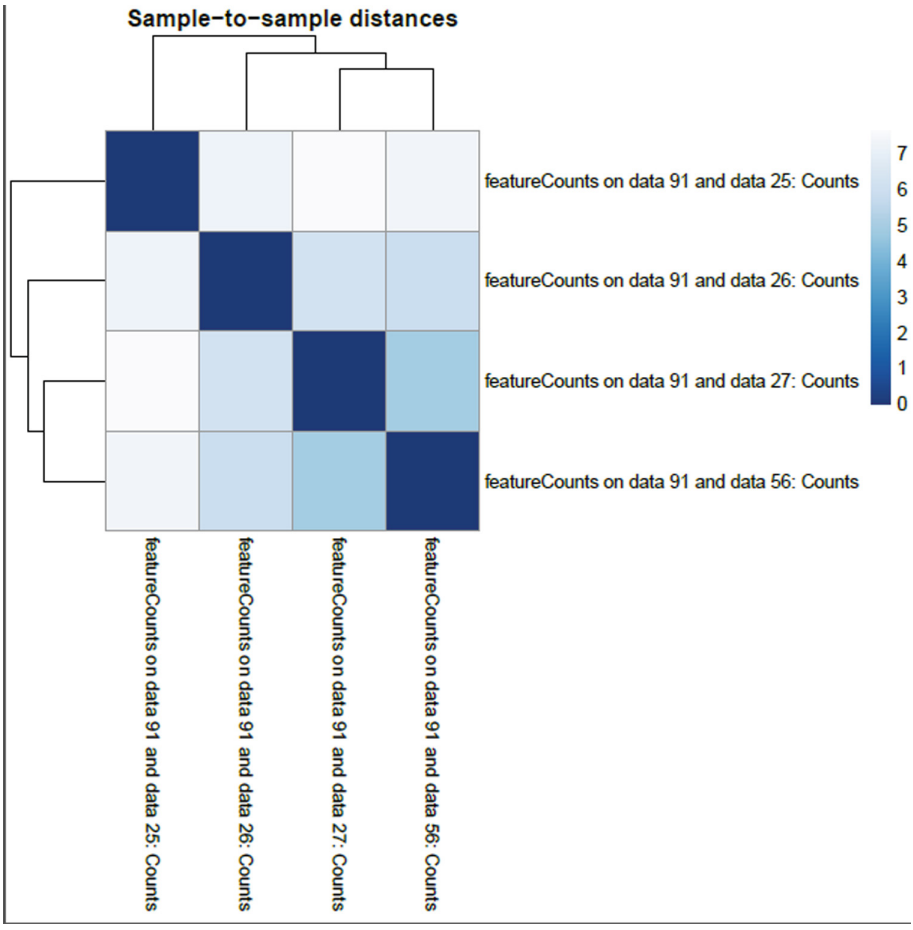


Fig. 3. Heatmap plot of sequences

The Heatmap (Fig. 3) of the sample-to-sample distance matrix (with clustering) gives an overview of similarities and dissimilarities between samples. The colour represents the distance between the samples. Dark blue means shorter distance, i.e. closer samples given the normalized counts.

In the Dispersion estimates plot (Fig. 4) we see the gene-wise estimates in black, the fitted values in red, and the final maximum a posteriori estimates used in testing (blue).

Histogram (Fig. 5) of *p*-values for the genes in the comparison between the 2 levels of the 1st factor.

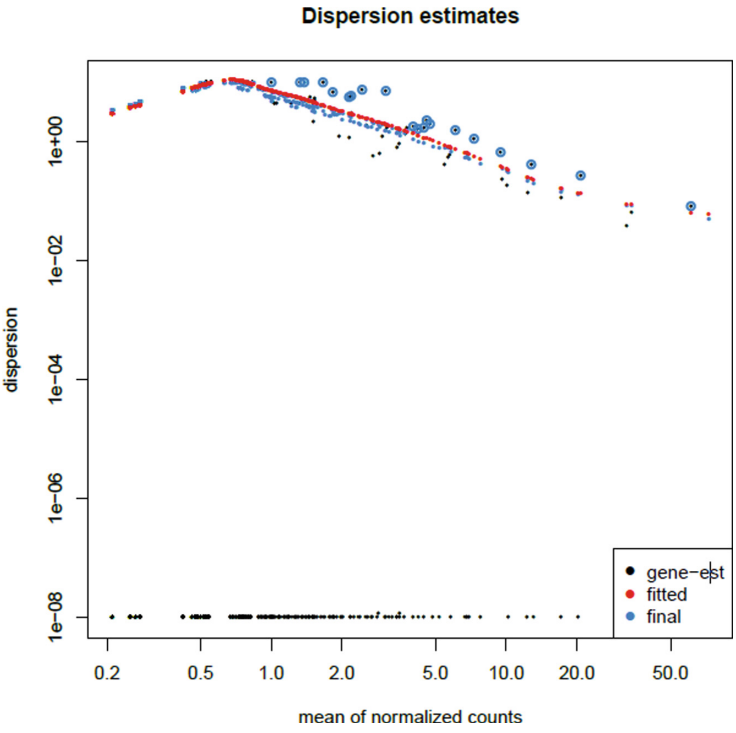


Fig. 4. Dispersion estimates plot of sequences

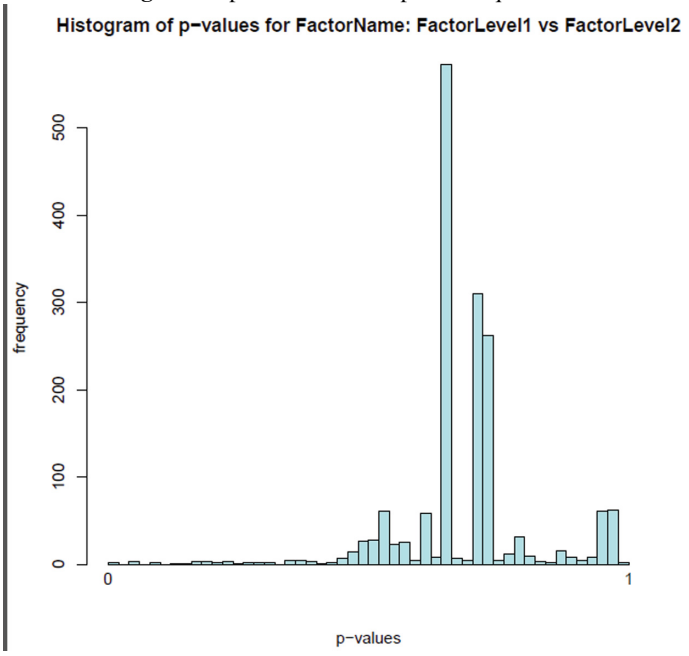


Fig. 5. Histogram plot of sequences

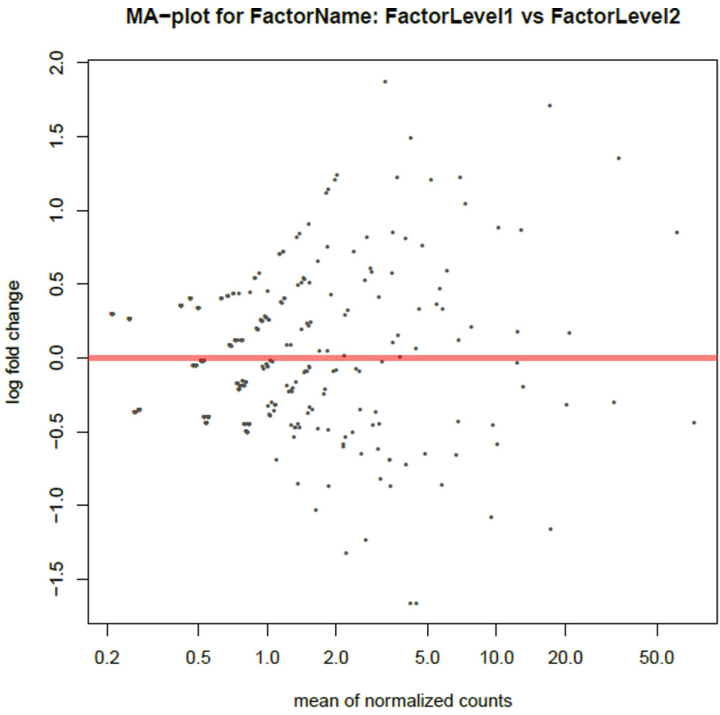


Fig. 6. MA plot of sequences

Our MA plot (Fig. 6) represents the global view of the relationship between the expression change of conditions (log ratios, M), the average expression strength of the genes (average mean, A), and the ability of the algorithm to detect differential gene expression. We can see that our genes that passed the significance threshold (adjusted p-value < 0.1) are coloured in red.

Further, we have used volcano plot (Fig. 7) to find the statistical significance (P value) versus magnitude of change (fold change) to identify genes with large fold changes that are also statistically significant. These statistically significant genes may be the most biologically significant. In this plot, the most upregulated genes are towards the right, the most downregulated genes are towards the left, and the most statistically significant genes are towards the top. We have selected the 1st 10 genes from DESeq2 results (NM_001039211.3, NR_038351.1, NR_111945.1, NM_015001.3, XM_011541571.2, XR_947031.2, NM_013943.3, NR_034112.2, XM_017001139.2 and NM_001394062.1) to be displayed in the volcano plot as expressed and significant genes. NR_111945.1, NR_034112.2, XR_947031.2 and NM_015001.3 are the most upregulated genes which are towards the right and NM_001039211.3, NR_038351.1, XM_011541571.2, NM_013943.3, XM_017001139.2 and NM_001394062.1 are the most downregulated genes which are towards the left. From the genes selected, NR_038351.1 is the most significant gene since it is on the topmost amongst the genes.

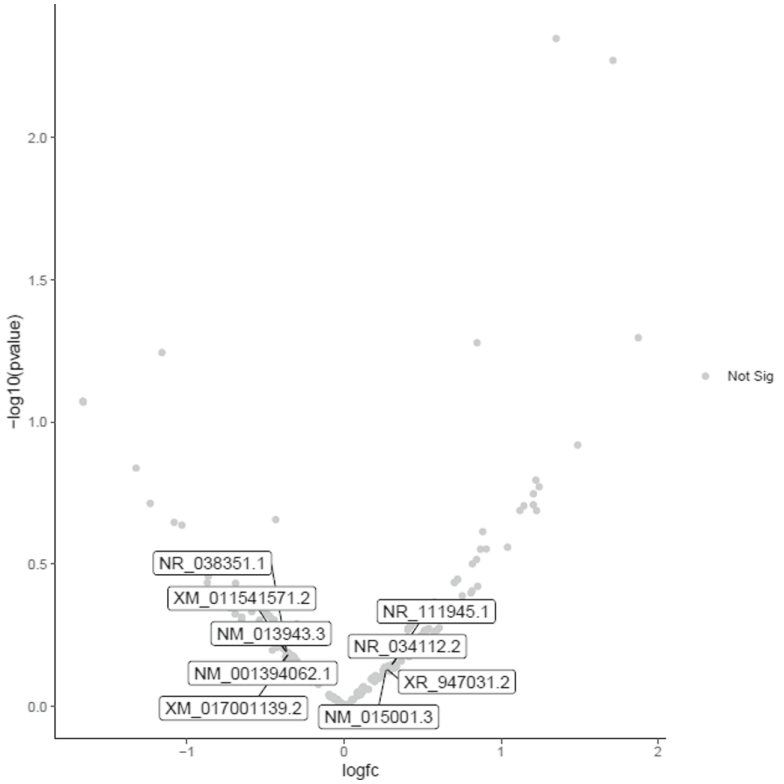


Fig. 7. Volcano Plot

4 Conclusion

As per MultiQC results we see that bad, duplicate and overlapping reads are removed by Trimmomatic tool. Further, with the DESeq tool we identify the expressed genes. NR_111945.1, NR_034112.2, XR_947031.2 and NM_015001.3 which are towards the right of the volcano plot are the most upregulated genes and NM_001039211.3, NR_038351.1, XM_011541571.2, NM_013943.3, XM_017001139.2 and NM_001394062.1 are towards the left are the most downregulated genes. From the genes above, NR_038351.1 is the most significant gene since it is on the topmost amongst the genes.

References

1. Andreasen, N. C., & Flaum, M. (1991). Schizophrenia: the characteristic symptoms. *Schizophrenia bulletin*, *17*(1), 27-49.
2. Andreasen, N. C. (1982). Negative symptoms in schizophrenia: definition and reliability. *Archives of general psychiatry*, *39*(7), 784-788.
3. Chapman, J. (1966). The early symptoms of schizophrenia. *The British Journal of Psychiatry*, *112*(484), 225-251.

4. Crow, T. J. (1981). Positive and negative schizophrenia symptoms and the role of dopamine. *The British Journal of Psychiatry*, *139*(3), 251-254.
5. Kane, J. M., & Marder, S. R. (1993). Psychopharmacologic treatment of schizophrenia. *Schizophrenia bulletin*, *19*(2), 287-302.
6. Meltzer, H. Y. (2001). Treatment of suicidality in schizophrenia. *Annals of the New York Academy of Sciences*, *932*(1), 44-60.
7. Tandon, R., Nasrallah, H. A., & Keshavan, M. S. (2010). Schizophrenia,"Just the Facts" 5. Treatment and prevention Past, present, and future. *Schizophrenia research*, *122*(1–3), 1–23.
8. Kim, J., Kim, M. S., Koh, A. Y., Xie, Y., & Zhan, X. (2016). FMAP: functional mapping and analysis pipeline for metagenomics and metatranscriptomics studies. *BMC bioinformatics*, *17*(1), 1-8.
9. Ghorbani, M., Rajandas, H., Parimannan, S., Joseph, G. B. S., Tew, M. M., Ramly, S. S., ... & Lee, S. Y. (2021). Understanding the role of gut microbiota in the pathogenesis of schizophrenia. *Psychiatric Genetics*, *31*(2), 39-49.
10. Nasrallah, H. A., Targum, S. D., Tandon, R., McCombs, J. S., & Ross, R. (2005). Defining and measuring clinical effectiveness in the treatment of schizophrenia. *Psychiatric Services*, *56*(3), 273-282.
11. Mimmack, M. L., Ryan, M., Baba, H., Navarro-Ruiz, J., Iritani, S., Faull, R. L., ... & Bahn, S. (2002). Gene expression analysis in schizophrenia: reproducible up-regulation of several members of the apolipoprotein L family located in a high-susceptibility locus for schizophrenia on chromosome 22. *Proceedings of the National Academy of Sciences*, *99*(7), 4680-4685.
12. Sundararajan, T., Manzardo, A. M., & Butler, M. G. (2018). Functional analysis of schizophrenia genes using GeneAnalytics program and integrated databases. *Gene*, *641*, 25-34.
13. Norton, N., Williams, H. J., Williams, N. M., Spurlock, G., Zammit, S., Jones, G., ... & Owen, M. J. (2003). Mutation screening of the Homer gene family and association analysis in schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *120*(1), 18-21.
14. O'Donovan, M. C., & Owen, M. J. (1999). Candidate-gene association studies of schizophrenia. *American journal of human genetics*, *65*(3), 587.
15. Sundararajan, T., Manzardo, A. M., & Butler, M. G. (2018). Functional analysis of schizophrenia genes using GeneAnalytics program and integrated databases. *Gene*, *641*, 25-34.