

The Systematic Analysis of lncRNA Expression in Ebola-infected Human Macrophages

Mathavan Muthaiyan

Pondicherry University

Leimarembi Devi Naorem

Pondicherry University

VigneshwarSuriya Prakash Sinnarasan

Pondicherry University

Dahrii Paul

Pondicherry University

Amouda Venkatesan (✉ amouda@bicpu.edu.in)

Pondicherry University <https://orcid.org/0000-0002-9586-1887>

Research Article

Keywords: lncRNAs, DEGs, Ebola infection, DE-lncRNA, novel lncRNA, RESTV

Posted Date: February 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-213151/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Ebola is a dangerous virus which causes severe headache, muscle pain, hemorrhagic fever, and multi-organ failure. The underlying mechanism of Ebola-human interaction at the molecular level remains unexplored. The recently reported research revealed that lncRNA are known to be important in the pathogenesis of viruses. The study carried out the comprehensive analysis of lncRNA identification in EBOLA, LPS and RESTV infected MDMs. In EBOLA, the total number of known lncRNA is EBOLA 555 in, in RESTV 476, in LPS 550 and the novel lncRNAs in EBOLA 142, RESTV 127, in LPS 136 identified. Also, differentially expressed lncRNAs and neighbourhood genes of each lncRNA about 100kb identified for analysis. The study reported the known and novel lncRNA in EBOLA, LPS and RESTV infected cells of MDMs. The results of the study may help to understand the immune response further.

Introduction

Viral infections are a deadly threat to the worldwide community. Though many viral infections get treated with advanced treatment, there is still a need for an effective treatment for some virus diseases. Ebola belongs to such a group of virus with no effective and adequate treatment. The Ebola virus causes a deadly infection named Ebola Virus Disease (EVD), which produces a fatal outcome in many patients. The major outbreak of Ebola reported in 2013–2016 [1].

Zaire ebolavirus (EBOV) and *Reston ebolavirus (RESTV)* differ in the characteristic nature of pathogenesis though belong to the same Filovirus family. EBOV causes devastating effects in humans such as dysregulation of immune response and induces the cytokine storm. RESTV considered as non-pathogenic as there are no reported cases against humans. The differences in pathogenicity among EBOV and RESTV are not thoroughly understood. The current study uses RNA Seq dataset (PRJNA328248) to identify the known, novel and differentially expressed lncRNAs. Primary human monocyte-derived macrophages (MDMs) cells are treated with EBOV, RESTV and control with Lippo Poly Saccharide (LPS) [2], [3] to understand the host immune response.

Long non-coding RNAs (lncRNAs) are regulatory molecules controlling the various biological processes. However, the functions and characterization of identified lncRNA remain unexplored. lncRNAs of length > 200 nt in size lacks the coding potential and ubiquitously expressed in a mammalian system. lncRNAs have undergone post-transcriptional modification as similar to coding mRNA such as capping polyadenylation. lncRNAs are important regulator to control various biological processes in cells and organ systems such as regulating protein complexes, trafficking process of genes and chromosomes to their specific locations. Many reports have shown that differentially expressed cellular lncRNA of virus-infected cells involved in immune response [4]–[7] and some of them favour the viral replication by inhibiting the immune response [8].

The classified lncRNAs under different categories are based on the genomic locations, proximity to the respective protein-coding genes including exon sense overlapping, intron sense overlapping, bidirectional,

and intergenic lncRNAs [9]. Recent findings explored that lncRNA are novel players in antiviral immune response [10]. Thus, the study aims to i) identify the lncRNA ii) differentiate the known and novel lncRNA. iii) identify the differentially expressed lncRNAs with the neighborhood genes in a range of 100km distance. RNA sequencing is more benefited to capture RNA expressions when compared to microarray system such as analysis of novel transcript identification, allele-specific expression and splice junctions. This study analyzed the ebola infected human macrophages to identify the known, novel lncRNA and differentially expressed lncRNA. Many animal viruses like Epstein-Barr virus [11], herpes virus [12], Marek's disease virus [13], severe acute respiratory syndrome coronavirus (SARS-CoV) [4], human immune deficiency virus [14] reported reveals that the cellular lncRNAs expresses during the infection and favour the virus replication.

Materials And Methods

Data Collection and Pre-processing

Publically available data from Sequence Reads Archive database with the accession number PRJNA328248 is considered for the study. The fastq files directly downloaded from European Nucleotide Archive browser (<https://www.ebi.ac.uk/ena/browser/home>) and checked the quality of reads using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) [15]. Fig 1 depicted the workflow.

Read Alignment and Transcript Assembly

After the quality check, the reads aligned with the human reference genome (GRCh38) using the hierarchical indexing for spliced alignment of transcript (HISAT2 v.2.1.0) with the default settings. HISAT2, a splice aware alignment, it is worked based on the Burrows-Wheeler transform and the Ferragina-Manzini (FM) index (<http://www.ccb.jhu.edu/software/hisat/>) [16]. The output of HISAT2 (SAM format) is converted into BAM file. The sorted BAM file serves as an input for assembling using StringTie. It produces an accurate, complete reconstructs of genes and estimate the expression level of the transcripts [17].

Identification of Novel lncRNA

The merged assembled transcript from StringTie used to identify the novel lncRNAs. To start with the transcript length >200 nucleotides with strand information are considered for further steps and the subset of a filtered transcript is compared with hg38 annotation file using Gffcompare . The class code representing non-coding regions "i", "u", and "x" are retained for subsequent steps [18]. The transcripts with Open Reading Frame (ORF) identified using TransDecoder are discarded. The remaining reads checked for coding potential using the tool CPAT(Coding Potential Assessment Tool) and PLEK, (Predictor of Long Non-coding RNAs and messenger RNAs based on an improved k-mer scheme <https://sourceforge.net/projects/plek/files/> [19]. CPAT is an alignment-free logistic model to identify non-coding regions of the transcripts [20]. The tool performs better in terms of sensitivity, specificity and accuracy compared to other non-coding region prediction tool, CPC, PhyloCSF. The lncRNA transcripts

with a score of less than zero are filtered out for further analysis. The transcripts are processed using standalone BLASTX against Swissprot database to check the false positives. The transcript with an alignment score, E-value $>10^{-5}$ were removed. The outputs of BLASTX are subject to BLASTN against LNCipedia and NONCODE database to get the novel lncRNAs.

GC Content Analysis

Emboss geecee is an online tool used to calculate the content G+C bases of the nucleic acid sequence(s). It sums the number of G and C bases and reports the result to file in fractions in the interval 0.0 to 1.0 [21].

Differential Gene Expression Analyses

The BAM file is input to Sub read package to generate the expression gene counts matrix [22]. The count matrix process to identify the differentially expressed genes of EBOV, RESTV, LPS treated cells using DESeq2 [23]. The genes with threshold $\log_{2}FC \pm 1.5$ and adjusted p-value < 0.05 are considered as significant.

DE-lncRNA target prediction and functional annotations

To understand the functional role of DE-lncRNA better nearby genes are identified with the distance of 100kb of upstream and downstream regions for further investigation. The nearby genes extracted using BEDOPS [24] and BED TOOLS [25].

Gene Enrichment Analysis

Gene enrichment analysis would help to identify the interested genes and proteins generated through high-throughput studies. WebGestalt (WEB-based GENESeTAnaLysis Toolkit) is most widely used tool for gene enrichment analysis. The significant gene terms are filtered out based on the p-value < 0.05 .

Results

Data Collection and Quality Checking

The publically available dataset PRJNA328248 downloaded from NCBI SRA. The quality of the reads is checked by FastQC tool and the reads with Phred score above 20 are retained for the analysis.

Identification of Novel lncRNA

The protocol to identify the lncRNA is designed based on previous studies[26]–[30]. On checking of the quality, each sample was mapped to human reference genome hg38 using HISAT2. The overall alignment score greater than 75 per cent has taken for further steps as listed in Supplementary file S1. The generated HISAT2 alignment files of each sample were further processed to transcript assembly using StringTie. Subsequently, the Gffcompare is used to identify the non-coding reads and lncRNAs identified with the use of TranDecoder, CPAT and BLASTX, BLASTN. The lncRNAs without the strand information

were removed. It is evident from the research reported that the lncRNAs regulating the coding genes are nearby in the upstream and downstream regions (26, 27). The current study adopted the pipeline suggested in the research to identify the Ebola known lncRNA (555) and novel lncRNA (142), LPS Ebola known lncRNA (550) and novel lncRNA (136), RESTV known lncRNA (476) and novel lncRNA (127). Fig 2 depicted the identified novel lncRNA and their counts as a pie diagram in Supplementary file S2-S7.

GC Content Analysis

The GC content of lncRNAs identified through Emboss geecee tool. The value of GC contents of novel lncRNAs ranges from 0.78 to 0.31 (Fig 3) & known lncRNA from 0.8 to 0.29 (Fig 4).

Differential Expression Analysis of lncRNA

The gene expression count matrix file generated using Subread package. Differentially expressed lncRNA identified using R package DESeq2 for EBOV, LPS and RESTV. The differentially expressed transcripts are shown in Fig 5 and Table 1. The results showed that EBOV, LPS has strongly induced the immune response than RESTV. The differentially expressed transcripts lncRNA counts showed Supplementary file S8, Table 1 and Fig 6, the overlapping of DE-lncRNA between time points as shown in and Fig 7.

Gene Enrichment Analysis

The gene enrichment analysis of identified DE-lncRNAs neighborhood genes was performed in WebGestalt. The highly enriched terms of EBOV, LPS, and RESTV cells identified the significant GO terms as in (Table 2) and the Supplementary file S9.

Discussion

lncRNA research has become a fascinating field of biological research like cancer research, genetic disorders, and infectious diseases. High throughput sequencing and bioinformatics methods make researchers possible to uncover the functions and characterization of lncRNAs in many species. lncRNA, play a important role in regulating the coding genes around 10 to 100 kb distance of up and downstream regions [31], [32]. Also, play a vital role in the regulation and control of multiple biological processes. Different classes of lncRNAs reported inducing cytokine production during viral infections. The scientific literature evidence proved that lncRNA involved in the host-virus interactions, such as activate the pathogen recognition receptors, epigenetic modulation, controlling transcriptional and post-transcriptional process [33].

The current study, whole transcriptome analysis of EBOV, LPS and RESTV infected MDM cells performed. High-throughput techniques with a bioinformatics approach allow the scientist to uncover the role and characterization of lncRNA. The known and the unknown lncRNA expression in EBOV, LPS and RESTV cells identified. Cellular lncRNAs which actively expressed during viral infections may help to promote the virus replications, suppress the antiviral immune response [34]. In total, 1581 known and 405 novel

lncRNA identified. Further characterization of identified novel lncRNAs may give the functional role in the activation of the immune response.

As a result, in total, 1278 DE-lncRNAs in EBOV, LPS and RESTV were identified. Some of this lncRNA overlapped with different time points which may play an important role in the immune response. The neighborhood genes of each DE-lncRNAs in the range of 100 kb distance in upstream and downstream positions identified. The reported research on the gene enriched ontology terms of nearby genes and its associated functions of other viruses support and evident the current study of ebola.

The important enrichment terms of EBOV (i) GO: 0061676, the importin protein α hijacked by Ebola VP24 protein to block STAT mediated IFN- α/β and IFN- γ synthesis. Further, importin $\alpha 7$ involved in the formation of inclusion bodies. Also, it involves potentially in pathogenesis (37–39). ii) GO: 0032036 Myosin heavy chain binding. The micropinocytosis is uptake EBOV with the initiation of an external stimulus to activate the receptor tyrosine kinases. Several regulators involved to carry out this process for an example Arp2/3 and myosin [35]. iii) GO:0046875 ephrin receptor binding are known to be involved in cell-to-cell interactions [36].

LPS enriched terms: i) GO: 0005229 intracellular calcium-activated chloride channel activity. Many bacterial LPS reported that it has the potential to induce calcium signalling and chloride signalling. ii) GO: 0001614 purinergic nucleotide receptor activity. LPS causes the dysregulated ATP release that intervenes with autocrine purinergic signalling mechanism important for the antimicrobial host defence process [37].

RESTV enriched terms: i) GO:0015038 glutathione disulfide proteins involved in the immune and inflammatory responses to infection[38]. ii) Oxidoreductase activity reported in many viral infection[39]. iii) GO:0008276 protein methyltransferase activity are essential for the epigenetic regulation like methylation, histone and non-histone proteins[40]. The identified lncRNAs may regulate the gene enrichment terms. Further studies are required to understand the lncRNA regulation of the neighborhood genes.

Conclusion

Limited research on lncRNA expression and characterization in Ebola virus infection need to be investigated further. The study identified the novel and known lncRNA in Ebola, LPS, and RESTV treated MDM cells. The results suggested that RESTV lacks immune activation by comparing EBOV and LPS. The DE-lncRNAs along with the neighbourhood genes with a distance of 100 kb up and downstream positions reported. The information out the research helps to understand the function of the immune responses. The future directions are to study the functional and structural characteristics of novel lncRNAs. Further investigation is required to understand the role of DE-lncRNAs and neighbourhood genes.

Declarations

Acknowledgement

The author thanks the Centre for Bioinformatics, Pondicherry University for computational facilities to carry out the research work. Mathavan Muthaiyan acknowledges a Senior Research Fellowship from Rajiv Gandhi National Fellowship (RGNF).

Conflict of interest

The authors declare no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Coltart CEM, Lindsey B, Ghinai I, Johnson AM, Heymann DL (2017) The Ebola outbreak, 2013–2016 : old lessons for new epidemics. *Philos Trans R Soc B Biol Sci* 372:2013–2016
2. Miranda ME et al (1999) Epidemiology of Ebola (Subtype Reston) Virus in the Philippines, 1996. *J Infect Dis* 179(Suppl 1):115–119
3. Olejnik J et al (2017) Ebolaviruses Associated with Differential Pathogenicity Induce Distinct Host Responses in Human Macrophages. *J Virol* 91(11):1–22
4. Peng X et al (2010) Unique Signatures of Long Noncoding RNA Expression in Response to Virus Infection and Altered Innate Immune Signaling. *MBio* 1(5):1–9
5. Josset L et al (2014) Annotation of long non-coding RNAs expressed in Collaborative Cross founder mice in response to interferon-stimulated transcripts Annotation of long non-coding RNAs expressed in Collaborative Cross founder mice in response to respiratory virus infection r. *RNA Biol* 11(7):875–890
6. More S, Zhu Z, Lin K, Huang C, Pushparaj S (2019) Long non-coding RNA PSMB8-AS1 regulates influenza virus replication. *RNA Biol* 16(3):340–353
7. Valadkhan S, Gunawardane LS, “lncRNA-mediated regulation of the interferon response,” *Virus*, vol. 212, no. January, pp. 127–136, 2020
8. Ma Y, Ouyang J, Wei J, Maarouf M, Chen J, “Involvement of Host Non-Coding RNAs in the Pathogenesis of the Influenza Virus,” 2017
9. Rnas L, Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10:155–159
10. Batista PJ, Chang HY (2013) Review Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell* 152(6):1298–1307

11. Zhang J et al., "Long noncoding RNAs involvement in Epstein-Barr virus infection and tumorigenesis," pp. 1–8, 2020
12. Sonkoly E et al (2005) Identification and Characterization of a Novel, Psoriasis Susceptibility-related Noncoding RNA gene, PRINS *. *J Biol Chem* 280(25):24159–24167
13. Ahanda ME et al (2009) Non-coding RNAs revealed during identification of genes involved in chicken immune responses. *Immunogenetics* 61:55–70
14. Quan Zhang K-TJ, Chen C-Y, Venkat SRK, Yedavalli (2013) NEAT1 Long Noncoding RNA and Paraspeckle Bodies Modulate HIV-1 Posttranscriptional Expression. *MBio* 4(1):1–9
15. "Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data."
16. Kim D, Langmead B, Salzberg SL (2015) HISAT: A fast spliced aligner with low memory requirements. *Nat Methods* 12(4):357–360
17. Pertea M et al (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads HHS Public Access Author manuscript. *Nat Biotechnol* 33(3):290–295
18. Pertea G, Pertea M, Love MI, "Open Peer Review GFF Utilities: GffRead and GffCompare [version 1; peer review: 2 approved] 1 2 report report," 2020
19. Li A, Zhang J, Zhou Z (2014) PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics* 15:1–10
20. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W, "CPAT: Coding-potential assessment tool using an alignment-free logistic regression model," *Nucleic Acids Res.*, vol. 41, no. 6, pp. 1–7, 2013
21. Bruskiwich R, "Emboss geecee," (1999) [Online]. Available: <http://www.bioinformatics.nl/cgi-bin/emboss/help/geecee>
22. Liao Y, Smyth GK, Shi W (2014) Sequence analysis featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923–930
23. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Nucleic Acids Res* 15:1–21
24. Neph S et al (2012) BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28(14):1919–1920
25. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842
26. Azali A, Obeidat SM, Yunus MA, Ghows A, "Systematic identification and characterization of *Aedes aegypti* long noncoding RNAs (lncRNAs)," *Sci. Rep.*, no. January, pp. 1–9, 2019
27. Etebari K, Asad S, Zhang G, Asgari S (2016) Identification of *Aedes aegypti* Long Intergenic Non-coding RNAs and Their Association with Wolbachia and Dengue Virus Infection. *PLoS Negl Trop Dis* 10(10):1–18
28. Wu Y, Cheng T, Liu C, Liu D, Zhang Q, Long R (2016) Systematic Identification and Characterization of Long Non-Coding RNAs in the Silkworm, *Bombyx mori*. *PLoS One* 11(1):1–25

29. Chen B, Zhang Y, Zhang X, Jia S, Chen S, Kang L, "Genome-wide identification and developmental expression profiling of long noncoding RNAs during *Drosophila* metamorphosis," *Nat. Publ. Gr.*, vol. 6, no. February, pp. 4–11, 2016
30. Zhao P, Liu S, Zhong Z, Jiang T, "Analysis of expression profiles of long noncoding RNAs and mRNAs in brains of mice infected by rabies virus by RNA sequencing," *Sci. Rep.*, no. July, pp. 1–10, 2018
31. Yang L et al (2013) lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* 500(7464):598–602
32. Bonasio R, Shiekhattar R, "Regulation of Transcription by Long Noncoding RNAs."
33. Liu S, Shi W, Li J, Carr MJ, "Long noncoding RNAs: Novel regulators of virus - host interactions," *Wiley Public Heal. Emerg. Collect.*, vol. 29, no. January, pp. 1–12, 2019
34. Wang J, Cen S (2020) Roles of lncRNAs in influenza virus infection. *Emerg Microbes Infect* 9(1):1407–1414
35. Mulherkar N, Raaben M, Carlos J, Torre D, Whelan SP, Chandran K (2011) The Ebola virus glycoprotein mediates entry via a non-classical dynamin-dependent macropinocytic pathway. *Virology* 419(2):72–83
36. Bonaparte MI et al (2005) Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. *PNAS* 102(30):10352–10657
37. Kondo Y et al., "Frontline Science: *Escherichia coli* use LPS as decoy to impair neutrophil chemotaxis and defeat antimicrobial host defense," *J. Leucoc. Biol.*, no. March, pp. 1–9, 2019
38. Ciriolo MR et al., "Loss of GSH, Oxidative Stress, and Decrease of Intracellular pH as Sequential Steps in Viral Infection *," vol. 272, no. 5, pp. 2700–2708, 1997
39. Mathys L, Balzarini J (2015) The role of cellular oxidoreductases in viral entry and virus infection-associated oxidative stress: potential therapeutic applications. *Expert Opin* 20:1–21
40. Zeng H, Xu W, Chap. 16. *Enzymatic Assays of Histone Methyltransferase Enzymes*. Elsevier Inc., 2015

Tables

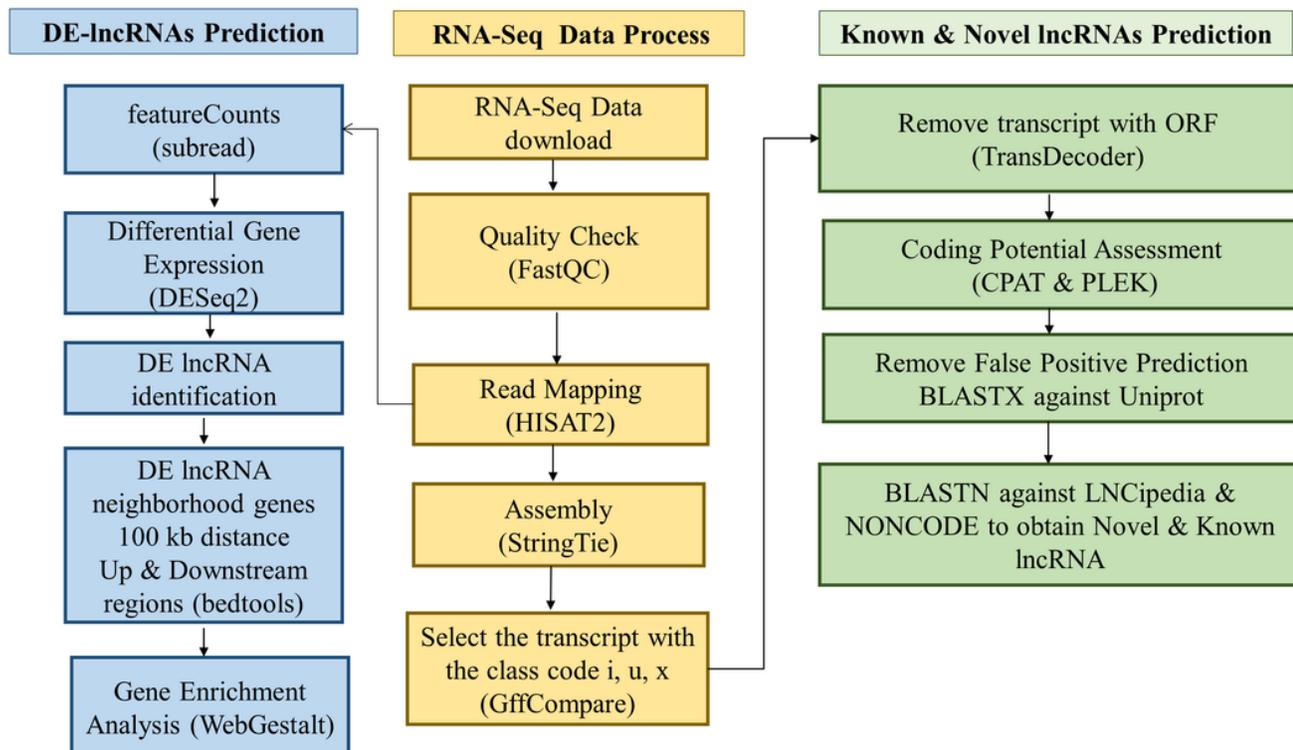
Treatment of Cells in hours	6	6	24	24	48	48
Differential Gene Expression	UP	DOWN	UP	DOWN	UP	DOWN
EBOV	17	18	121	141	128	96
LPS	64	44	78	142	105	196
RESTV	15	24	25	16	21	39

Table 1: The differentially expressed lncRNA transcripts in 6, 24 and 48 hrs treated cells of EBOV, LPS and RESTV.

Treatment	Treatment in hrs	UP/Down-Regulation	Significant Enriched terms of DE-lncRNA neighbourhood genes in 100kb
ZEBOV	6	Down	GO:0061676 importin-alpha family protein binding
LPS	6	Down	GO:0005229 intracellular calcium-activated chloride channel activity
RESTV	6	Down	GO:0008276 protein methyltransferase activity
ZEBOV	6	Up	GO:0004028 3-chloroallyl aldehyde dehydrogenase activity
LPS	6	Up	GO:0001614 purinergic nucleotide receptor activity
RESTV	6	Up	GO:0031720 haptoglobin binding
ZEBOV	24	Down	GO:0016886 ligase activity, forming phosphoric ester bonds
LPS	24	Down	GO:0042802 identical protein binding
RESTV	24	Down	GO:0015248 sterol transporter activity
ZEBOV	24	Up	GO:0032036 myosin heavy chain binding
LPS	24	Up	GO:0015318 inorganic molecular entity transmembrane transporter activity
RESTV	24	Up	GO:0001517 N-acetylglucosamine 6-O-sulfotransferase activity
ZEBOV	48	Down	GO:0019864 IgG binding
LPS	48	Down	GO:0008821 crossover junction endo deoxy ribonuclease activity
RESTV	48	Down	GO:0015038 glutathione disulfide oxidoreductase activity
ZEBOV	48	Up	GO:0046875 ephrin receptor binding
LPS	48	Up	GO:0008092 cytoskeletal protein binding
RESTV	48	Up	GO:0001784: phosphotyrosine residue binding

Table 2: Gene enrichment analysis of neighbourhood genes of differentially expressed lncRNAs

Figures



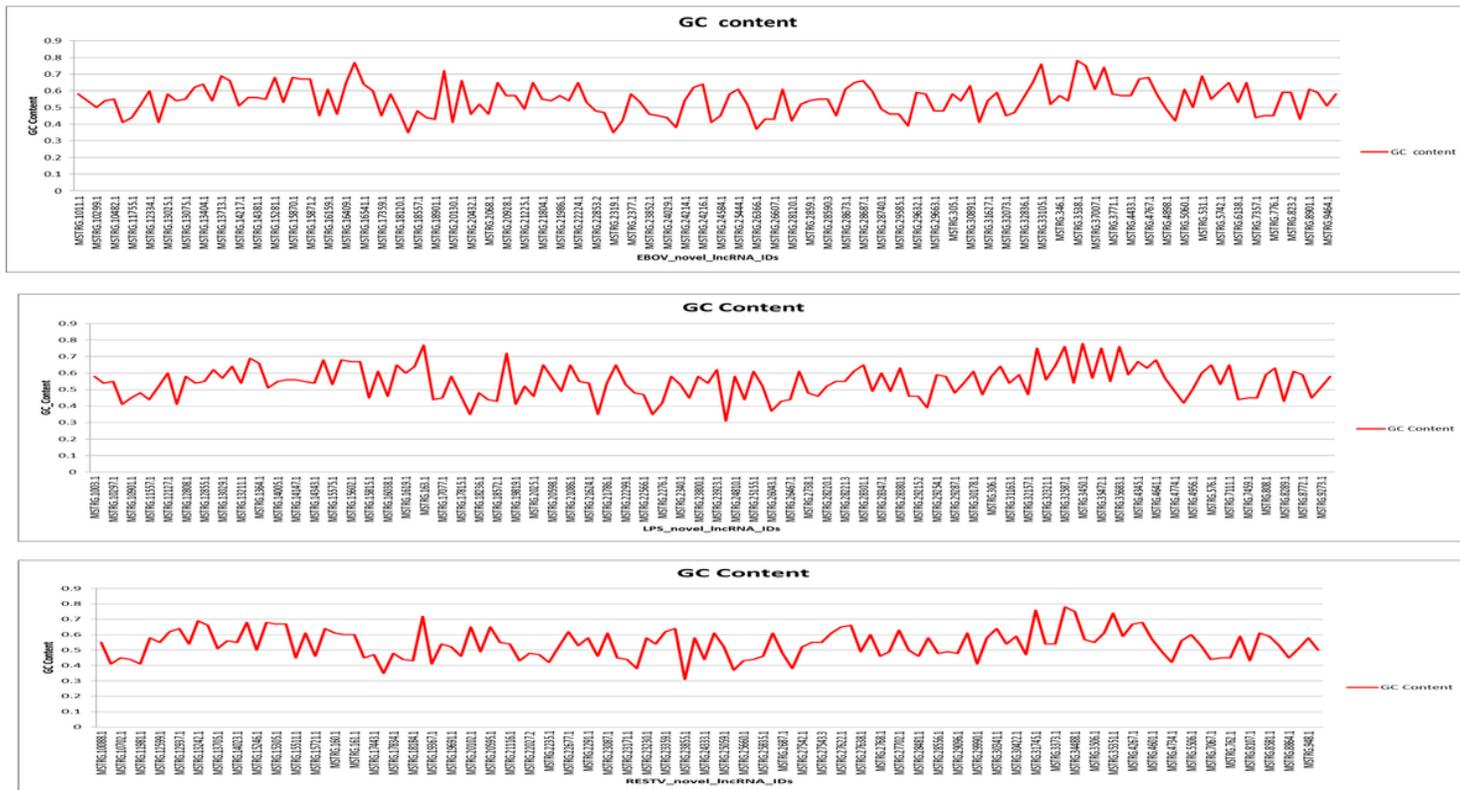


Figure 4

The line plot shows the distribution of novel lncRNAs GC content percentage

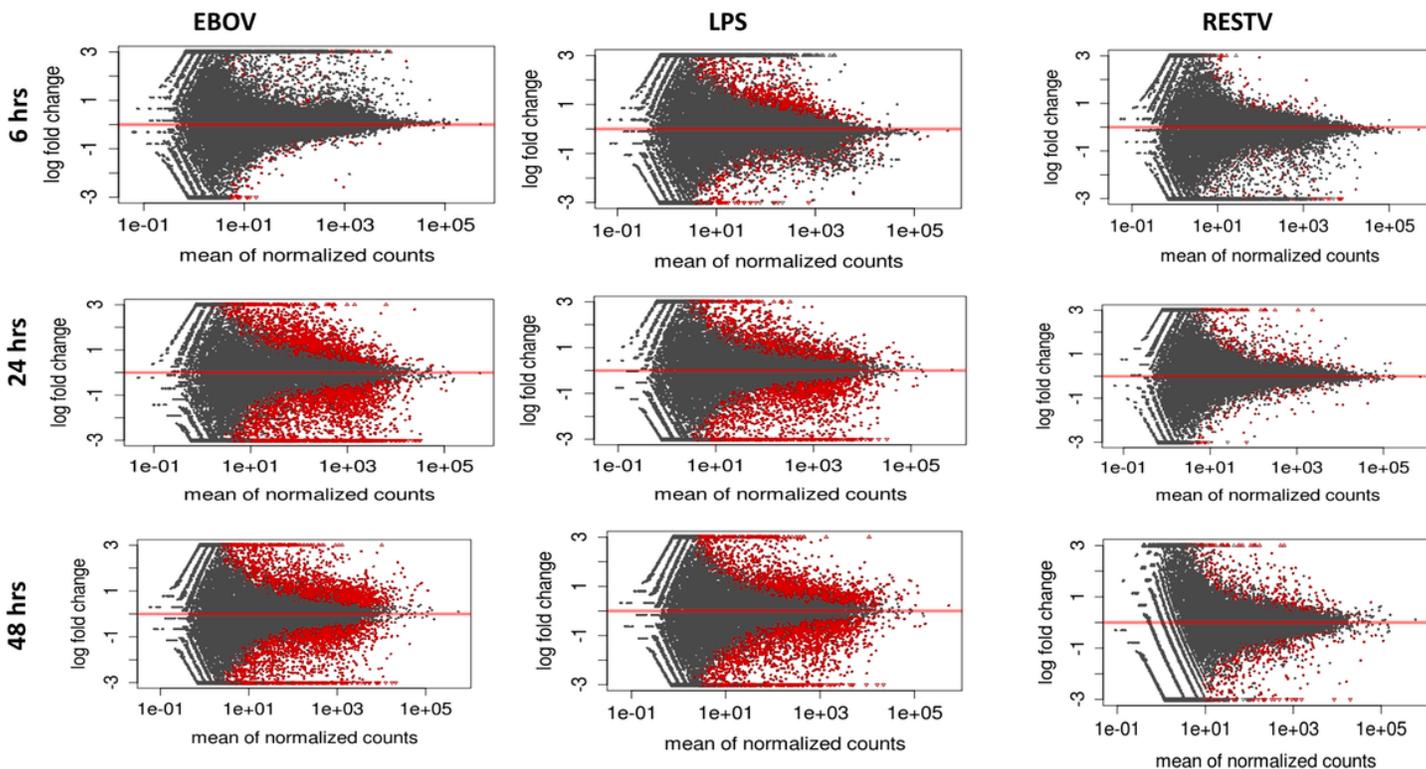


Figure 5

MA Plots showing the mean of the normalized transcripts against log₂ fold changes. The significantly expressed transcripts are shown in red colour.

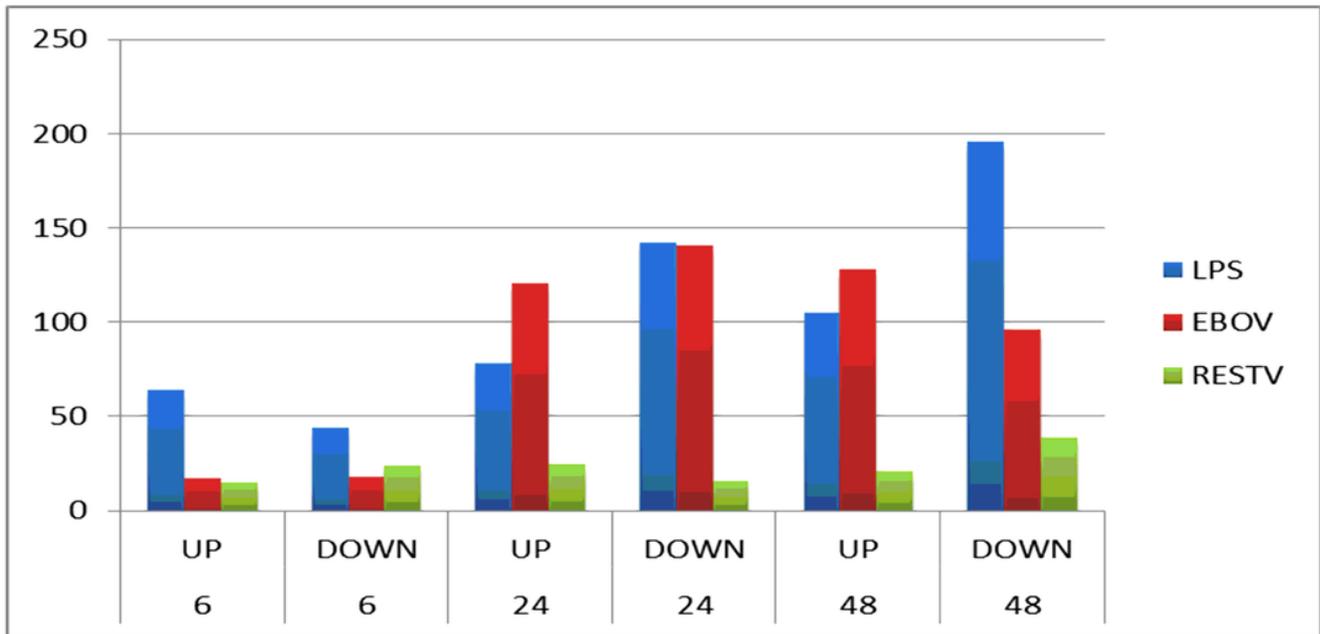


Figure 6

The bar chart depicts differentially expressed lncRNA transcripts in 6, 24 and 48 hrs treated cells of EBOV, LPS and RESTV.

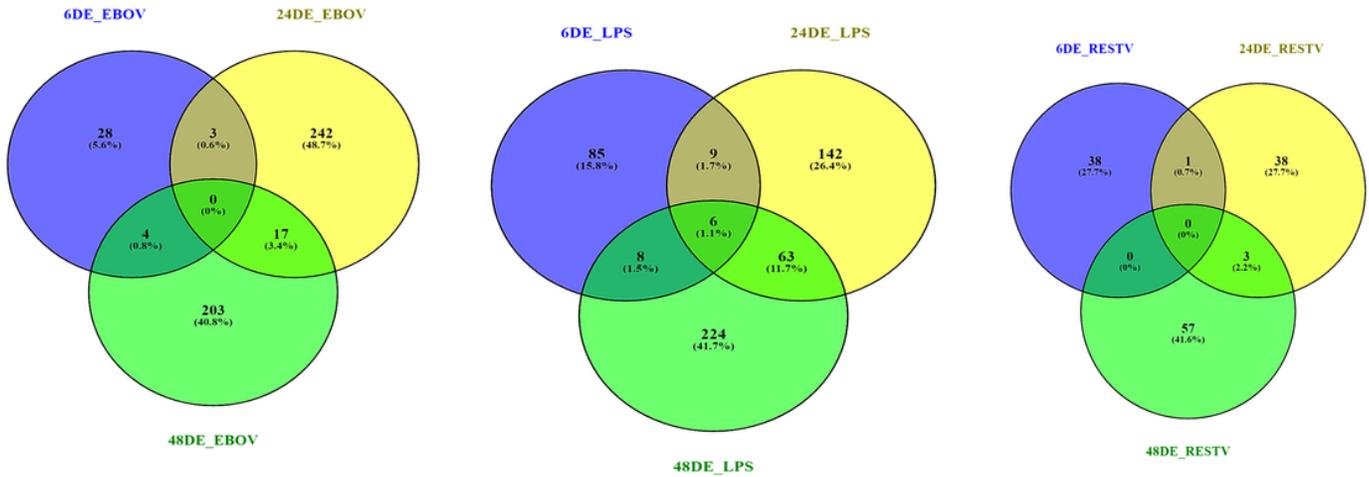


Figure 7

DE-lncRNA expressions between groups of EBOLA, LPS and RESTV, Venn diagram showing the overlap and non-overlap of the DE-lncRNAs in infected macrophages

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S1Hisatalignmentscores.docx](#)
- [S2EBOVKnownlncRNAs.fasta](#)
- [S3EBOVNovellncRNAs.fasta](#)
- [S4LPSKnownlncRNAs.fasta](#)
- [S5LPSNovellncRNAs.fasta](#)
- [S6RESTVknownlncRNAs.fasta](#)
- [S7RESTVNovellncRNAs.fasta](#)
- [S8DElncRNASlist.xlsx](#)
- [S9.docx](#)