ORIGINAL ARTICLE

Transcriptomic Profiling of LPD-3 Mutant in *C. elegans* for Bridging Differential Expression with Human Orthologs

Samiksha Garse¹, Vaishnavi Thakur¹, Shine Devarajan¹, Mrunal Gokhale^{1*}

¹ School of Biotechnology and Bioinformatics, D Y Patil Deemed to be University, CBD Belapur, Navi Mumbai, Maharashtra, India- 400 614.

* Corresponding Author's Email: mrunal.gokhale@dypatil.edu, mrunal09gokhale@gmail.com

ABSTRACT

Caenorhabditis elegans is a model organism for studying various human functionalities, diseases and their pathways. The correlation of its phenotypic and genotypic traits with human orthologs has been strikingly similar due to its evolutionary conserved genetic makeup. LPD-3 has a bridge domain that accommodates lipids and facilitates their trafficking from the endoplasmic reticulum to the membrane. Leveraging on this information, wherein complex gene networks are integrated with transcriptomic profiling, curious minds have now discovered the impact of lipid metabolism with Alkuraya-Kucinskas Syndrome and cancer-specific signaling pathways. Mutations in LPD-3 have shown an abnormal distribution of these lipids, affecting various signaling pathways that trigger some differentially expressed genes to either over or under-express. Our research focuses on RNA-Seq data of LPD3 mutants within C. elegans retrieved from open resources for transcriptomic profiling of raw reads. A comparative analysis of the wild type and LPD-3 mutants in C. elegans identified a few differentially expressed genes, annotated with human orthologs indicating cancer-specific gene functions and potential molecular targets. Moreover, the results uncovered a stronger linkage of these human orthologs with distinct types of cancer and their signaling pathways, leading to the discovery of novel biomarkers, and highlighting future research for prognosis and potential therapeutics.

Keywords: Biomarkers, C. elegans, Cancer, Human Ortholog, Lipid Metabolism, Mutation, Transcriptomic Profiling

Differential Expression with Human Orthologs. Adv. Biores. Vol 16 [2] March 2025. 266-274

Received 19.01.2025Revised 21.03.2025Accepted 30.03.2025How to cite this article:
Samiksha G, Vaishnavi T, Shine D, Mrunal G. Transcriptomic Profiling of LPD-3 Mutant in *C. elegans* for Bridging

INTRODUCTION

Caenorhabditis elegans with its short lifespan and a genome that is substantially homologous to mammals, is recognized as a remarkable model organism for researching various human diseases and gene functions [1]. This worm-based model replicates the relationships between genotypic and phenotypic traits and a range of human diseases, demonstrating evolutionary conserved genetic makeup [2]. The homology has allowed researchers to get an insight into lipid metabolism to comprehend the complex gene networks associated with various signaling pathways [3]. Lipid transport in eukaryotic cells usually occurs at the adjacent organelle membranes known as Membrane contact sites (MCS) [3, 4]. These contact sites facilitate the transport of lipids, especially phospholipids, thus playing an important role in understanding various processes like lipid metabolism, its trafficking and association with diverse signaling pathways [4]. Within *C elegans*, a very large and long protein consisting of 4424 amino acid residues, known as LPD-3 is involved in lipid metabolism [4, 5]. Similarly, the gene BLTP1 (Bridge-like lipid transfer proteins (LTPs)) is a human ortholog of LPD-3. BLTP1 enhances the bulk transfer of lipids via their repeating β -groove (RBG) domains that appear as tunnel-like structures creating the impression of a bridge [6]. Therefore, these lipid-accommodating hydrophobic domains bridge the lipid transport from the endoplasmic reticulum (ER) to the plasma membrane (PM) localized at the ER-PM junction [4]. Mutations in LPD-3 lead to aberrant phospholipid distribution, further reducing cold resistance in C. elegans by altering the unsaturated phospholipids and membrane fluidity through a compromised homeostatic mechanism [7]. Likewise, a mutation in BLTP1 is associated with Alkuraya-Kucinskas Syndrome (AKS) and other signaling pathways directly or indirectly linked with various types of cancer

[4, 5, 7]. We aim to investigate the impact of transcriptomic profiling to identify differentially expressed genes (DEGs) in LPD-3 mutant samples of *C. elegans* and its correlation with Human orthologous genes to state involvement in cancer-signalling pathways for suggesting future prognosis and drug therapies.

MATERIAL AND METHODS Sample Collection:

The raw sequence reads were retrieved from EMBL nucleotide archive (ENA) with BioProject accession number: PRJNA827259 (https://www.ebi.ac.uk /ena/browser/view/PRJNA827259) consisting of RNA-Seq data of wild type against LPD-3 mutant within *C. elegans* (Table 1). NCBI dataset requires an SRA toolkit to extract the data whereas the nucleotide archive supports data extraction from the FTP file format through the command line on the LINUX platform.

Accession Number:	Sample	Experiment	Run	Generated FASTQ				
PRJNA827259	Accession	Accession	Accession	Files				
LPD-3 Wild Type <i>C. elegans</i>	SAMN27607360	SRX14879184	SRR18779825	SRR18779825_1				
				SRR18779825_2				
	SAMN27607361	SRX14879185	SRR18779824	SRR18779824_1				
				SRR18779824_2				
	SAMN27607362	SRX14879186	SRR18779823	SRR18779823_1				
				SRR18779823_2				
LPD-3 Mutant <i>C. elegans</i>	SAMN27607363	SRX14879187	SRR18779822	SRR18779822_1				
				SRR18779822_2				
	SAMN27607364	SRX14879188	SRR18779821	SRR18779821_1				
				SRR18779821_2				
	SAMN27607365	SRX14879189	SRR18779820	SRR18779820_1				
				SRR18779820_2				

Pre-Processing and Quality Analysis:

Figure 1. illustrates and overall workflow for the quality analysis and filtration that was performed on a LINUX platform, followed by the identification of DEGs. The samples had forward read and reverse reads in FASTQ format. After these reads were imported successfully, their Quality Control (QC) and Filtration were done using the FASTQC Tool [8] and fastp Tool [9]. The QC clarifies the distribution of quality scores across all the positions of all the reads [8, 9]. The Phred quality scores >30 which signifies 99.9% of the accuracy of the base calling during sequencing [9].



Figure 1: Flowchart for the Transcriptomics Analysis

Alignment and Mapping:

Filtered reads were locally aligned to determine the origin of reads in the genome. STAR Splice-aware aligner was used to map against the reference genome [10]. The reference genome was obtained through iGenome database's [11] UCSC genome browser [12] with ce10 genome build.

(https://s3.amazonaws.com/igenomes.illumina.com/Caenorhabditis_elegans/UCSC/ce10/Caenorhabditi s_elegans_UCSC_ce10.tar.gz) The reference genome is indexed first for the generation of index files and later the quality filtered reads were mapped across this indexed reference genome [13]. The quality mapped filtered reads were further analyzed using Qualimap [14], with the filter criteria of mapping quality score cutoff as <=20, to generate the quality filtered BAM files. Annotation for *C. elegans* and HT-Seq Tool [15] were used to generate the functional raw reads data (read counts).

Identification of DEGs and Survival Analysis:

Metadata was prepared for all six samples and was coupled with functional raw read counts for the identification of upregulated and downregulated genes using R programming code with the DESeq2 package [16]. The differential genes were filtered with adjusted Pvalue <0.05 and \log_2 fc ± 2, further annotated with the WormBase database [17] to retrieve human Orthologs, their HGNC locations and associated diseases. The identified genes were further validated using the UALCAN database [18].

RESULTS

Quality Analysis and identification of DEGS

The fastp tool is the one-stop solution to check adapter contamination, bad ligation and low Phred quality. It results in the percentage of filtered data or discarded data within the range of 1.08-5.7%, which is acceptable for RNA-Seq data.

There was a total of 39 up and 20 downregulated genes identified out of which overall 10 genes (**Table 2**) were selected for studying the human orthologs and their association with cancer.

C. elegans	Log2fc	Human Orthologs	Cancer Type in Human	References		
UPREGULATED GENES						
ROM-3	5.794331	RHBDF1	High levels are known to be correlated with several cancer types including breast cancer and melanoma.	[19, 20]		
GRK-2	5.395345	GRK2	High levels are indicators of diabetes, heart failure, hypertension and breast tumorigenesis.	[21]		
F47B8.10	4.9799	MMP16	High expression levels are linked with poor prognosis in gastric cancer patients as they enhance the growth and metastasis of gastric cancer cells.	[22]		
CPR-3 CPR-4	3.591209 3.199197	CSTB	Over-expressed in many cancer types like lung squamous cell carcinoma (SCC), hepatocellular carcinoma (HCC), gastric adenocarcinoma, breast cancer and esophageal squamous cell carcinoma.	[23]		
LET-756	3.510689	FGF3	Overexpression has led to poor prognosis in breast cancer and demonstrated cellular aggression in ovarian and esophageal cancer.	[24]		
RNT-1	3.160406	RUNX2 RUNX3	High expression levels of RUNX2 in breast cancer have led to metastasis by activating PI3K/Akt pathway. Whereas, high levels of RUNX3 led to basal cell carcinoma cells.	[25, 26]		
DOWNREGULATED GENES						
Y116A8C.20 CCCH-2	-2.97005 -2.056	ZFP36	Encodes for Tristetraprolin and are highly expressed in the normal mammary epithelium, while downregulation frequently accompanies tumorigenesis.	[27]		
E04F6.15 DHS-7	-2.89807 -2.56227	WWOX	Frequent downregulation and loss of WWOX gene expression in human HCC.	[28]		
SOD3	-2.02358	SOD2	Decreased SOD2 expression links with the mortality rate due to HCC.	[29]		

 Table 2: List of up and downregulated genes

 Table 2: List of up and downregulated genes

The results generated from the UALCAN validated these 10 hub genes against the TCGA datasets [30] and generated boxplots. From **Figure 2a**, **2b** and **2c** boxplots, we can conclude that in most of the genes, primary tumor expression is higher. This difference in expression levels could be important for understanding the role of these genes in the development and progression of different types of cancer. Higher expression in tumor samples suggests that these genes play a role in tumor growth or its maintenance, making it a potential target for therapeutic interventions.



Figure 2a : The boxplots generated compare expression levels of ZFP36, WWOX, SOD2, RUNX2, RHBDF1, genes in two types of samples from the TCGA database: Normal and Primary tumors for different types of cancer. The X-axis represents TCGA samples and the Y-axis represents Transcript per million with different scales for different genes. 'n' in the boxplots represents the number of samples present.







Figure 2c : The boxplots generated compare expression levels of MMP16, CTSB genes in two types of samples from the TCGA database: Normal and Primary tumors for different types of cancer. The X-axis represents TCGA samples and the Y-axis represents Transcript per million with different scales for different genes. 'n' in the boxplots represents the number of samples present.

Figure 3a and 3b boxplots demonstrate the expression levels of these hub genes in different types of cancer across different stages of nodal metastasis. The Normal Group displayed by the blue box plot represents the expression levels in normal samples. N0 to N3 Groups show the expression levels in samples with varying degrees of nodal metastasis, from N0 (no metastasis) to N3 (advanced metastasis). The graph helps in understanding how the expression of genes changes with the progression of nodal metastasis. This information may benefit significantly in the development of potential cancer treatment strategies.



Figure 3a: The boxplots constructed compare expression levels of the genes ZFP36, WWOX, SOD2, RUNX2, RHBDF1, FGF3 in various cancer types based on their nodal metastasis status. The X-axis represents various regions for nodal metastasis like normal, N0, N1, N2 and N3. Here, N0 = no metastasis in regional lymph nodes, N1 = present in 1 to 3 axillary lymph nodes (ALN), N2 = present in 4 to 9 ALN and N3 = present in 10 or more ALN. The Y-axis represents Transcript per million with different scales for different genes.



Figure 3b: The boxplots constructed compare expression levels of the genes FGF3, CSTB, MMP16, in various cancer types, based on their nodal metastasis status. The X-axis represents various regions for nodal metastasis like normal, N0, N1, N2 and N3. Here, N0 = no metastasis in regional lymph nodes, N1 = present in 1 to 3 axillary lymph nodes (ALN), N2 = present in 4 to 9 ALN and N3 = present in 10 or more ALN. The Y-axis represents Transcript per million with different scales for different genes.

The survival analysis in **Figure 4a**, **4b** and **4c** provides insights into the effects of gene expression levels on the survival probability of the patient. High gene expression in patients corresponds to lower survival probability over time compared to patients with low- medium gene expression. P-values less than 0.5 suggests that the difference in survival probabilities between the two groups is statistically significant. This means that, based on this data, there is strong evidence to conclude that gene expression levels significantly affect survival outcomes in cancer patients. This information can be useful for researchers and clinicians in understanding the role of gene expressions in different types of cancers and potentially leading future research and treatment strategies.



Figure 4a: Constructed graphs known as Kaplan-Meier survival plot displays the effect of levels of expression of genes CSTB and FGF3 on the survival probabilities of patients in different types of cancer. The red line in the graph represents high expression, the blue line represents low - medium expression and 'n' represents the number of patients. The x-axis is labelled Time in days and the y-axis is labelled Survival Probability. A p-value of plots represents statistical significance i.e. difference between two or more survival curves. It is used to determine whether the null hypothesis is true for survival tests. The null hypothesis states that survival curves in each group are identical for the overall population. A P-value less than 0.5 indicates that there is evidence to reject the null hypothesis and a P-value more than 0.5 states that there is no evidence to suggest a difference.



Figure 4b: Constructed graphs known as Kaplan-Meier survival plot displays the effect of levels of expression of genes MMP16, RHBDF1, RUNX2 and SOD2 on the survival probabilities of patients in different types of cancer. The red line in the graph represents high expression, the blue line represents low - medium expression and 'n' represents the number of patients. The x-axis is labelled Time in days and the y-axis is labelled Survival Probability. A p-value of plots represents statistical significance i.e. difference between two or more survival curves. It is used to determine whether the null hypothesis is true for survival tests. The null hypothesis states that survival curves in each group are identical for the overall population. A P-value less than 0.5 indicates that there is evidence to reject the null hypothesis and a P-value more than 0.5 states that there is no evidence to suggest a difference.



Figure 4c: Constructed graphs known as Kaplan-Meier survival plot displays the effect of levels of expression of genes WWOX, CSTB and ZFP36 on the survival probabilities of patients in different types of cancer. The red line in the graph represents high expression, the blue line represents low - medium expression and 'n' represents the number of patients. The x-axis is labelled Time in days and the y-axis is labelled Survival Probability. A p-value of plots represents statistical significance i.e. difference between two or more survival curves. It is used to determine whether the null hypothesis is true for survival tests. The null hypothesis states that survival curves in each group are identical for the overall population. A P-value less than 0.5 indicates that there is evidence to reject the null hypothesis and a P-value more than 0.5 states that there is no evidence to suggest a difference.

Moreover, some of the up and downregulated genes were orthologous in humans which were associated with features/symptoms of AKS. Over-expression of human orthologs of ace-2 and grk-2 were observed in certain cardiac anomalies which were marked as the features of AKS, whereas under-expressed genes involved human orthologs of dgn-2, acs-2, gba-4, T05E7.1, and lipl-4.

Apart from this, a few other downregulated genes also searched against WormBase annotation, yielding findings related to a variety of other disease categories associated with their matched human orthologs. The diseases range from neurological disorders including, major depressive disorder, Schizophrenia, Autism, and Bipolar disorder to various types of cancers like breast cancer, SCC, HCC, pancreatic cancer, gastric, glioblastoma, etc. The downregulation of all these genes indicates that BTLP1 mutations and/ or the downregulation may not be associated only with AKS but also with a wide array of diseases. The precise mechanism responsible for these effects, whether it be at the level of gene expression, protein-protein interactions, protein-ligand interactions, or another mechanism entirely, needs to be thoroughly examined and researched.

DISCUSSION

LPD-3 plays a major role in lipid metabolism within *C. elegans* suggesting that the bridge-like structure promotes lipid trafficking and its associated mechanisms. This confers with cold tolerance within the worm and maintains a balance between phospholipid and membrane fluidity. Mutation within this gene can cause various disruptions in this metabolism leading to various abnormalities. BLTP1 is a human ortholog of LPD-3. BTLP1 mutations and/ or its downregulation may not be associated only with AKS but also with different types of malignancies. Impaired lipid transportation due to mutations/ downregulation of BLTP1 might be having some effect on the levels of expression of various genes.

Our research aimed to study the RNA-Seq data for LPD-3 mutant versus wild type within *C. elegans*. The transcriptomic profiling resulted in filtered quality raw read counts within the set cutoff for QC analysis, with an accuracy rate of 99.9% base calling. These read counts were analyzed using the DESeq2 package of R programming to identify DEGs. Genes with log₂fc ±2 were annotated using the WormBase Database to retrieve human orthologs, their HGNC locations and their associated diseases.

The resultant genes indicated up and down-regulation, depicting that they can be novel biomarkers for the prognosis or early diagnosis of different cancer types. There is strong literature support that builds the foundation of this study and is validated using the UALCAN database. Some of the highly upregulated genes have corresponding human orthologs which are known to be overexpressed in some types of cancers while some moderately downregulated genes have corresponding human orthologs that are under-expressed in certain types of cancers. The types of cancers associated with these human orthologs are breast cancer, HCC, melanoma, gastric cancer, SCC, skin cancer, esophageal cancer, stomach cancer, etc. The box plots conclude the high expression levels of each of these orthologous genes within their respective cancer type. It also indicated the growth rate and progression within different stages of cancer of each of the cancer types. Moreover, the survival analysis showed significant results for each of these orthologous genes.

This research also identified down-regulation of DGN-2, ACS-2, GBA-4, T05E7.1 & LIPL-4 giving a clue that low levels of expression in their corresponding human orthologs are responsible for the development of various symptoms/features of AKS. Although mutation in BLTP1 is highly related to AKS, the human orthologs of some upregulated and downregulated genes in the presence of LPD-3 mutation, are associated with different cardiac anomalies and neurological disorders also.

CONCLUSION

The transcriptomic analysis can be a significant pipeline to bridge the DEGs from *C. elegans* with the human orthologs to find the correlation between molecular mechanisms and the functional abnormalities of the genes. The findings of this research lay a foundation for understanding the exact correlation between levels of expression of these genes and BLTP1 dys-functioning in various types of cancer. Future molecular investigations and additional research can be conducted using bioinformatics strategies to evidently map the mechanism of BLTP1 with each cancer type, which may enables targeting specific-gene receptors for the establishment of novel therapeutics and drug therapies.

REFERENCES

- 1. Markaki, M. and N. (2020). Tavernarakis, *Caenorhabditis elegans* as a model system for human diseases. Current opinion in biotechnology, **63**: p. 118-125.
- 2. Markaki, M. and N. (2010). Tavernarakis, Modeling human diseases in *Caenorhabditis elegans*. Biotechnology journal, **5**(12): p. 1261-1276.
- 3. An, L., X. Fu, J. Chen, and J. Ma, (2023). Application of Caenorhabditis elegans in lipid metabolism research. International journal of molecular sciences, **24**(2): p. 1173.
- 4. Pandey, T., et al., (2024). LPD-3 as a megaprotein brake for aging and insulin-mTOR signaling in C. elegans. Cell reports, **43**(3).
- 5. Pandey, T., et al., (2023). A megaprotein brake for aging and insulin-mTOR signaling. bioRxiv.
- 6. Neuman, S.D., T.P. Levine, and A. Bashirullah, (2022). A novel superfamily of bridge-like lipid transfer proteins. Trends in cell biology, **32**(11): p. 962-974.
- 7. Wang, C., et al., A conserved megaprotein-based molecular bridge critical for lipid trafficking and cold resilience. Nature communications, 2022. **13**(1): p. 6805.
- 8. Andrews, S., (2010). FastQC: a quality control tool for high throughput sequence data. 2017.
- 9. Chen, S., Y. Zhou, Y. Chen, and J. Gu, (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, **34**(17): p. i884-i890.
- 10. Križanović, K., A. Echchiki, J. Roux, and M. Šikić, (2018). Evaluation of tools for long read RNA-seq splice-aware alignment. Bioinformatics, **34**(5): p. 748-754.
- 11. Lin, F.-M., H.-D. Huang, Y.-C. Chang, and J.-T. Horng, (2004). i-Genome: A database to summarize oligonucleotide data in genomes. BMC genomics, **5**: p. 1-6.

- 12. Navarro Gonzalez, J., et al., (2021). The UCSC genome browser database: 2021 update. Nucleic Acids Research. **49**(D1): p. D1046-D1057.
- 13. Sheng, Q., et al., (2017). Multi-perspective quality control of Illumina RNA sequencing data analysis. Briefings in functional genomics. **16**(4): p. 194-204.
- 14. Okonechnikov, K., A. Conesa, and F. García-Alcalde, (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. **32**(2): p. 292-294.
- 15. Anders, S., P.T. Pyl, and W. Huber, (2015). HTSeq—a Python framework to work with high-throughput sequencing data. bioinformatics, **31**(2): p. 166-169.
- 16. Wang, L., et al., (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. **26**(1): p. 136-138.
- 17. Sternberg, P.W., et al., (2024). WormBase 2024: status and transitioning to Alliance infrastructure. Genetics. **227**(1): p. iyae050.
- 18. Chandrashekar, D.S., et al., (2022). UALCAN: An update to the integrated cancer data analysis platform. Neoplasia, **25**: p. 18-27.
- 19. Updates to the Alliance of Genome Resources central infrastructure. Genetics, 2024. 227(1): p. iyae049.
- 20. Yan, Z., et al., (2008). Human rhomboid family-1 gene silencing causes apoptosis or autophagy to epithelial cancer cells and inhibits xenograft tumor growth. Molecular cancer therapeutics, **7**(6): p. 1355-1364.
- 21. Chaudhary, P.K. and S. Kim, (2021). The GRKs reactome: role in cell biology and pathology. International Journal of Molecular Sciences, **22**(7): p. 3375.
- 22. Shen, Z., et al., (2017). MMP16 promotes tumor metastasis and indicates poor prognosis in hepatocellular carcinoma. Oncotarget, **8**(42): p. 72197.
- 23. Huhtaniemi, I., Encyclopedia of endocrine diseases. 2018: Academic Press.
- 24. Kornmann, M. and M. Korc, (2013). Fibroblast growth factor receptors and cancer-associated perturbations.
- 25. Sun, S.-S., L. Zhang, J. Yang, and X. Zhou, (2015). Role of runt-related transcription factor 2 in signal network of tumors as an inter-mediator. Cancer Letters. **361**(1): p. 1-7.
- 26. Lotem, J., et al., (2015). Runx3 at the interface of immunity, inflammation and cancer. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer. **1855**(2): p. 131-143.
- 27. Zhang, D., et al., (2021). Tristetraprolin, a potential safeguard against carcinoma: role in the tumor microenvironment. Frontiers in Oncology. **11**: p. 632189.
- 28. Yang, T., et al., (2021). WWOX activation by toosendanin suppresses hepatocellular carcinoma metastasis through JAK2/Stat3 and Wnt/β-catenin signaling. Cancer Letters, **513**: p. 50-62.
- 29. Wang, R., et al., (2016). Reduced SOD2 expression is associated with mortality of hepatocellular carcinoma patients in a mutant p53-dependent manner. Aging (Albany NY), **8**(6): p. 1184.
- 30. Tomczak, K., P. Czerwińska, and M. Wiznerowicz, (2015). Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemporary Oncology/Współczesna Onkologia, (1): p. 68-77.

Copyright: © **2025 Author.** This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.