

Original Article

RNA Sequencing of Neurodegenerative Disease Postmortem CNS Tissues from Persons with Amyotrophic Lateral Sclerosis, or Alzheimer's or Parkinson's Diseases Reveal Heterogeneous Dysregulation of Gene Expression

James P Bennett^{1,2*}, Paula M. Keeney^{1,2} and David G Brohawn^{2,3}

¹Neurodegeneration Therapeutics, Inc. 3050A Berkmar Drive, Charlottesville, VA 22901, USA

²Parkinson's Center, Virginia Commonwealth University, Richmond VA, USA

³Department of Medical Genetics, Virginia Commonwealth University, Richmond, VA, USA

***Corresponding author**

James P Bennett, President and CSO, Neurodegeneration Therapeutics, Inc. 3050A Berkmar Drive Charlottesville, VA 22901, USA, Tel: 4345296457; Fax: 4345296458; Email: jpb8u@me.com /jpb8u@icloud.com

Submitted: 25 March 2020

Accepted: 24 October 2020

Published: 15 November 2020

ISSN: 2476-2032

Copyright

© 2020 Bennett JP, et al.

OPEN ACCESS

Abstract

RNA sequencing (RNAseq) of postmortem tissues may provide insights into disease pathogenesis. We carried out paired-end RNA sequencing (RNA-seq) and single-end small RNA-seq of total tissue RNA in postmortem nervous tissues of neurodegenerative disease (NDD) subjects (or controls, CTL) from individuals who died with amyotrophic lateral sclerosis (ALS, cervical spinal cord, mRNA only), Alzheimer's disease (AD, frontal cortex, mRNA and miRNA), or Parkinson's disease (PD, ventral midbrain, mRNA and miRNA). Gene expression levels and miRNA counts were averaged and standard deviations (s.d.) calculated to assay NDD/CTL mean ratios and gene or miRNA variances (variance = standard deviation²), respectively.

We observed a wide spectrum of gene expression variances across our populations. Median, 25th, and 75th percentile gene expression variances in all three NDD populations were several fold greater than in CTL samples. For all three NDD's, mRNA and miRNA variances in the NDD and CTL groups demonstrated power law relationships (log-log plots were ~linear).

Our findings support the hypothesis that NDDs result in heterogeneous gene expression (mRNA) patterns. Variances of mRNA's and miRNA's in both NDD and CTL tissues appeared to have a power law relationship with each other, suggesting that chaos mathematics influences gene and miRNA expressions.

INTRODUCTION

In spite of decades of research, processes underlying sporadic neurodegenerative disease (NDD) pathogenesis remain unclear. Rare, autosomal variants of each NDD exist, but account for a minority of cases and still require reaching adulthood for clinical expression [1-3].

RNA sequencing (RNAseq) is a powerful technological component of "next generation sequencing" (NGS) that has become accessible in pricing with publicly available alignment (<https://ccb.jhu.edu/software/tophat>; <https://ccb.jhu.edu/software/hisat2>) and quantitation (<http://cole-trapnell-lab.github.io/cufflinks>) (<https://sourceforge.net/projects/mirdeepstar/>) algo-

ritms and is increasingly reported in the scientific literature. A major unsolved dilemma relates to the biological interpretation of RNAseq data that can be easily acquired with the necessary economic, technological and computational resources.

Much effort to this point has involved the separation of "differentially expressed genes" (DEG) from the larger group of non-differentially expressed genes, taking into account the problems of defining differential expression in the setting of multiple sampling [4]. Various corrections for multiple sampling have been used, with a currently popular approach being the use of "False Discovery Rate" cutoffs. However, these statistical manipulations, while important, do not necessarily help the

biologist or clinician concerned about disease pathogenesis and progression.

Various artificial intelligence clustering programs have been developed that utilize both gene family over-representation approaches [5-8] or “biologically blind” mathematical clustering [9] with post-hoc biological analysis of clustered groups. Both approaches ultimately require curating of genes into biological families, with all its attendant biases and imperfections.

The above problems are increased in postmortem studies of human CNS tissues from individuals who died from/with NDD phenotypes. By the time such persons die, typically several years have passed with the individual exhibiting progressive phenotype deficiencies that can reflect neuronal losses of the disease-vulnerable populations. One also does not know to what degree the gene expression changes observed represent adaptive changes in “survivor” neurons or “survivor” changes in surrounding cells (ie., astrocytes). Should one focus on changes in individual neurons (or other cell types), now that RNA amplification strategies allow “single cell RNAseq” [10]? Or should the total cellular milieu be sampled, as in whole tissue RNAseq?

Further compounding the problem are the relative inaccessibility of the CNS, compared to peripheral sites, and the presence of advanced disease in those who donate tissues for postmortem storage. Work-around to these problems include induced pluripotential stem cells (iPSC’s) and their neuronal derivatives [11-13], generated from living persons with disease phenotypes at early presentations. A more contemporary approach involves the isolation and RNAseq analysis of small pieces of cells, called “exosomes”, that break off from living and dying cells, are small enough to pass across the blood-brain barrier and appear in blood and other bodily fluids [14,15].

Research involving iPSC’s and their derivatives suffer from the problem of “youngness”, with recent attempts to “age” these cells for more adult genotypes [16]. Research into utility of exosomes is still in early stages.

We have previously published papers describing our work on postmortem samples of cervical spinal cords from persons dying with amyotrophic lateral sclerosis (ALS) [17], frontal cortex from persons dying with Alzheimer’s disease (AD) [18] and ventral midbrain from persons dying with Parkinson’s disease (PD) [18]. In each of these analyses, we found gene expression changes involving many different genes and biological groups. Qualitatively, these results also characterize the experiences of others.

In the present study we have examined the hypothesis that heterogeneity of gene expression is present in human postmortem CNS tissues and is of greater magnitude in samples from persons dying with a NDD compared to CTL tissue samples.

METHODS

To test this hypothesis, we aligned our sequencing files to the most current version of the human genome (hg38, released in 2014). Tissue demographics have previously been reported [17,18]. Our sequencing files were generated from tissue total RNA, using the Illumina® paired-end, multiplex approach and

>60 million reads/sample. ALS (and CTL) cervical spinal cord samples were aligned with TOPHAT2, AD (and CTL) frontal cortex, PD (and CTL) ventral midbrain samples were aligned using HISAT2. Cufflinks was used for quantitation of gene expression. miRDEEP* was used for miRNA identification and quantitation (<https://sourceforge.net/projects/mirdeepstar/>). All graph construction, curve fitting and statistical analyses used GraphPad Prism® (v 8.0.2).

RESULTS

Gene expression variances in NDD show a normal (Gaussian) distribution around Disease/CTL ratios

Figure 1 (ALS, CTL (ALS)); 2 (AD, CTL(AD)); and 3 (PD, CTL(PD)) show plots of individual gene expression variances (=standard deviation²) VS NDD mean FPKMs/CTL mean FPKMs gene expression ratios (graphs on left), while graphs on right show Gaussian fits of expression variances VS gene expression ratios. The vertical lines on left-hand graphs indicate NDD/CTL ratios of 1.0 (ie, equivalent expression). In all NDD and CTL tissues, gene variances VS. NDD/CTL expression ratios showed distributions that fit normal curves of Gaussian distribution (Figures 1-3), right-hand graphs).

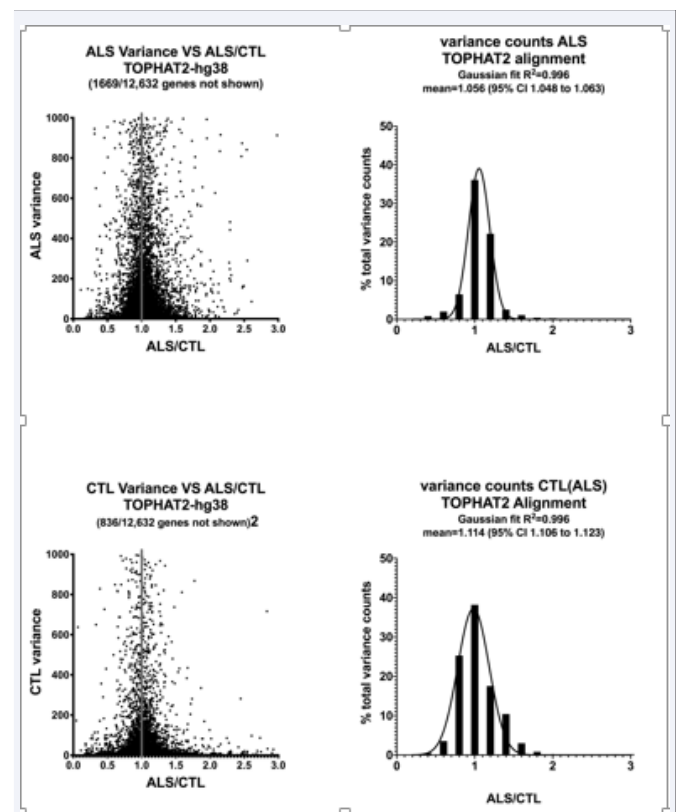


Figure 1 Gene expression variances VS gene expression ratios are normally distributed. (left) Plots of gene expression variance (= standard deviation²) VS gene expression ratios, based on FPKM values (FPKM = fragments per kilobase of exon per million reads) obtained from RNA sequencing of total RNA derived from cervical spinal cord sections of ALS (upper) or CTL (lower) subjects. The vertical line in each plot represents expression ratio of 1.0. (right) Plots of distribution of gene expression variances VS expression ratios for ALS (top) and CTL (bottom) samples, fit to Gaussian distribution using Prism®.

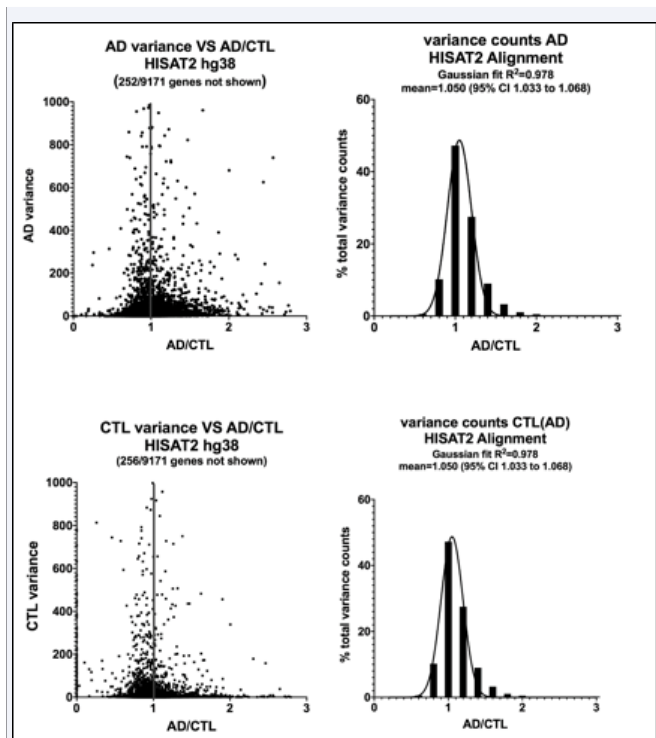


Figure 2 Same as for Figure 1 except for AD and CTL frontal cortex.

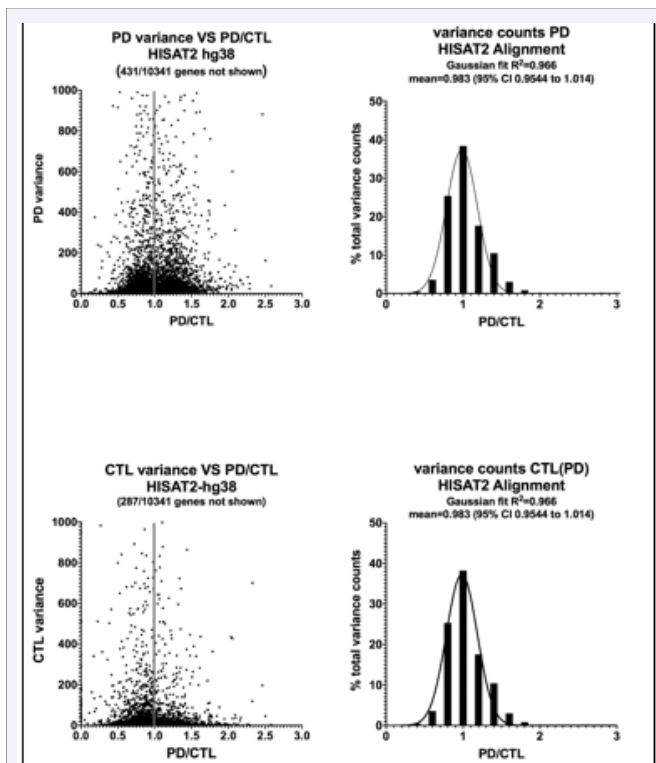


Figure 3 Same as for Figure 1 except for PD and CTL ventral midbrain.

We next defined the distributions of gene variances. Table 1 showed that for all NDDs, gene variance medians and 25th and 75th percentile limits were several fold higher than variances for CTL tissues analyzed at the same time.

We then examined relationships among individual gene variances in NDD and respective CTL groups. We plotted \log_{10} of variances (NDD) vs \log_{10} of variances (respective CTL) for individual genes (or miRNAs) of samples obtained from ALS, AD or PD subjects. We found (Figures 4,5) ~linear log-log relationships that suggest a power law relationship of the variances.

DISCUSSION

There are at least two major findings of our work

First, gene variances VS NDD/CTL expression ratios (based on FPKMs) appear to be distributed normally across human postmortem CNS tissues in both disease (ALS, AD, PD) and CTL cases. The peaks of variances occur near Disease/CTL ratios of ~1.0, suggesting that apparently equal expression of genes in these tissues is an artifact of large variances in gene expressions, at least for some genes.

Second, in disease (ALS, AD, PD) cases, we observed potential power law relationships among variances (NDD vs CTL) on an individual gene basis. This finding suggests that variances (and underlying standard deviations that in turn reflect individual gene or miRNA expressions) are determined by chaotic process (es).

Our analysis of gene variance distribution showed that the median values for gene variances were much higher in NDDs than in CTL samples. Taken together, our gene expression data in three adult NDDs demonstrated heterogeneity much greater than in CTL samples.

Aging itself appears to increase heterogeneity of gene expression [19,20]. Although multiple studies have shown the same qualitative phenomenon, underlying mechanisms remain unclear. Possibilities include factors that alter gene transcription (transcription factor availability, gene modification (ie, methylation), blockade of transcription site availability (chromatin changes)), and post-transcriptional factors (miRNA's).

These findings beg the question of whether these heterogeneities were in any way causal to the NDD processes, or instead simply correlated with the NDD processes. That question cannot be answered by our data, as we only had access to advanced disease cases and not to intermediate or early cases.

We are also limited in our interpretations by the following:

1. All of our NDD cases are from individuals with advanced NDD. The resulting gene expression values we obtained could represent "survival" responses of remaining neurons and/or "survival adaptations" of other CNS cells, such as astrocytes, that comprise the bulk of the tissues studied.
2. Some of the gene expression changes found, particularly under-expression in Disease, may simply represent loss of neurons in each NDD.
3. There are no reliable animal or cellular models of sporadic NDD's. iPSC technology represents "young" neural cells (at best), and exosomal analysis, while hopeful, is at too early a stage of development.

Table 1: Distributions of gene expression variances for NDD and CTL samples.

	AD	CTL (AD)	PD	CTL (PD)	ALS	CTL (ALS)
25 th percentile	1.03	0.43	2.12	0.42	6.74	1.6
median	2.96	1.1	6.68	1.16	28.05	4.79
75 th percentile	11.61	4.15	28.21	5.03	153.1	20.07

behavior. From a systems biology/network perspective, small changes in expression of a regulatory “driver” gene could have major consequences in expression of downstream “passenger” genes. If this approach has merit, then one should be seeking to identify these regulatory genes that could be present in the “low variance” group of each NDD. It remains to be seen whether existing clustering algorithms can identify these regulatory genes, and it should be considered that the identity of these hypothesized regulatory genes might vary across subjects.

ETHICS STATEMENT

All ALS tissues were acquired from National Disease Research Interchange (<https://ndriresource.org/>) under the auspices of their informed consent procedures. Most AD and most PD (and CTLs) tissues were collected under IRB oversight; other tissues were considered autopsy sourced and not subject to IRB oversight.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of Virginia Commonwealth University (VCU) Parkinson’s Center for the initial ALS studies, Department of Medical Genetics at Virginia Commonwealth University for fellowship support of DGB, ALS Worldwide for support of the ALS, AD and PD studies, and Neurodegeneration Therapeutics for support of the ALS, AD and PD studies.

DATA SHARING

Trimmed (Trimmomatic®) sequencing files are the property

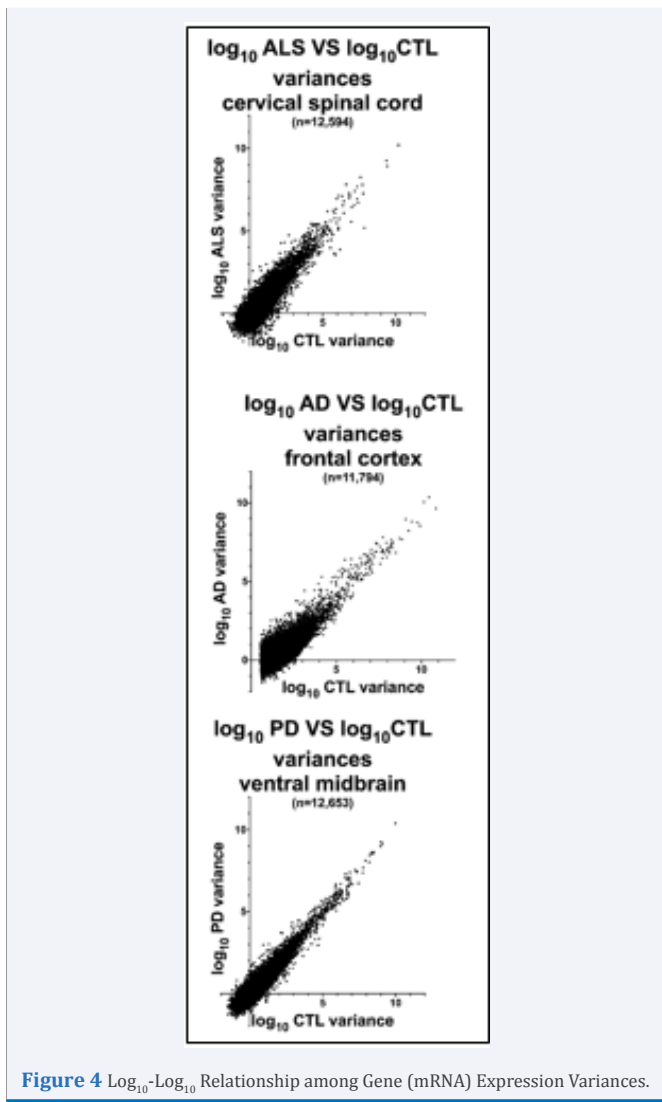


Figure 4 Log₁₀-Log₁₀ Relationship among Gene (mRNA) Expression Variances.

In spite of these significant limitations, our findings demonstrate that regulation of gene expression is altered in NDD tissues, at least at death. Variances are more “scattered” relative to CTL tissues, but it is not clear whether this is due to transcriptional (ie. gene methylation, histone post-translational modifications) or post-transcriptional (ie., miRNA effects, other mRNA catabolism) mechanisms. Both types of changes could be operative and can be sought in NDD tissues.

Is gene expression a chaotic process (in the mathematical sense) that increases with aging and is made more chaotic by NDD? One simplistic view of chaotic processes is that small changes in input variables can lead to large changes in output

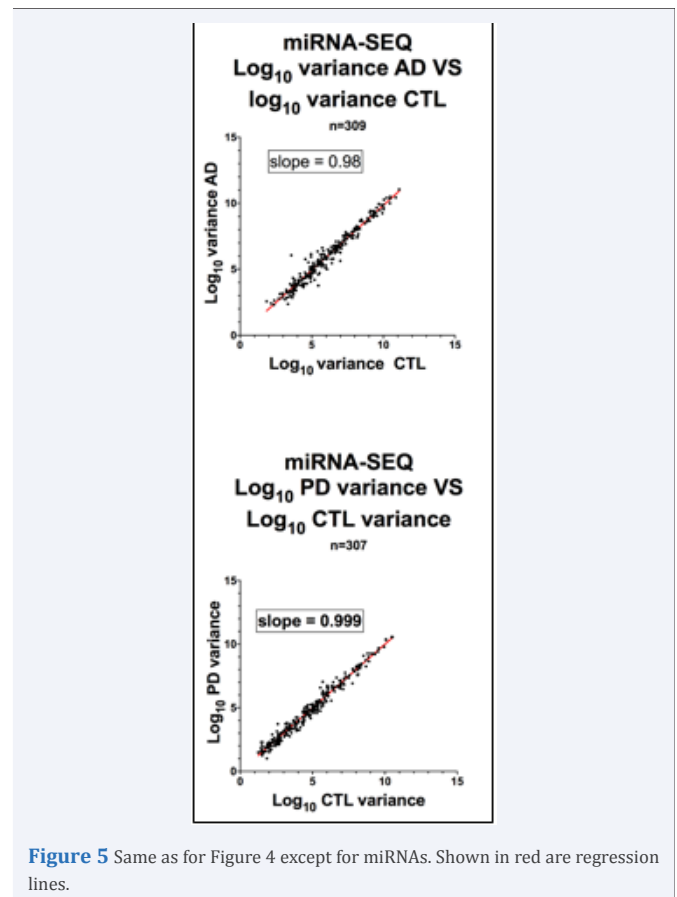


Figure 5 Same as for Figure 4 except for miRNAs. Shown in red are regression lines.

of Neurodegeneration Therapeutics and will be shared without cost to any legitimate non-profit investigators following contact with the Corresponding author (JPB).

REFERENCES

- Castellani RJ, Villa PG, Perry G. The amyloid cascade and Alzheimer's disease therapeutics: theory versus observation. *Lab Invest.* 2019; 99: 958-970.
- Del Rey NL, Varela AQ, Garbayo E, Carbajal IC, Santiago RF, Monje MHG, et al. Advances in Parkinson's Disease: 200 Years Later. *Front Neuroanat.* 2018; 12: 113.
- Taylor JP, Brown RH, Cleveland DW. Decoding ALS: from genes to mechanism. *Nature.* 2016; 539: 197-206.
- Su W, Sun J, Shimizu K, Kadota K. TCC-GUI: a Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC Res Notes.* 2019; 12: 133.
- Huang DW, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R, et al. Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics.* 2009; 13: 11.
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4: 44-57.
- Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009; 37: 1-13.
- Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A, et al. GOATOOLS: A Python library for Gene Ontology analyses. *Sci Rep.* 2018; 8: 10872.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9: 559.
- Picelli S. Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biol.* 2017; 14: 637-650.
- Farkhondeh A, Li R, Gorshkov K, Chen KG, Might M, Rodems S, et al. Induced pluripotent stem cells for neural drug discovery. *Drug Discov Today.* 2019; 24: 992-999.
- Teng YD. Functional multipotency of stem cells: Biological traits gleaned from neural progeny studies. *Semin Cell Dev Biol.* 2019; 95: 74-83.
- Mukherjee O, Acharya S, Rao M. Making NSC and Neurons from Patient-Derived Tissue Samples Methods. *Mol Biol.* 2019; 1919: 9-24.
- Zhang ZG, Buller B, Chopp M. Exosomes-beyond stem cells for restorative therapy in stroke and neurological injury. *Nat Rev Neurol.* 2019; 15: 193-203.
- Shi M, Sheng L, Stewart T, Zabetian CP, Zhang J. New windows into the brain: Central nervous system-derived extracellular vesicles in blood. *Prog Neurobiol.* 2019; 175: 96-106.
- Silva MC, Haggarty SJ. Human pluripotent stem cell-derived models and drug screening in CNS precision medicine. *Ann NY Acad Sci.* 2020; 1471: 18-56.
- Brohawn DG, O'Brien LC, Bennett JP Jr. RNAseq Analyses Identify Tumor Necrosis Factor-Mediated Inflammation as a Major Abnormality in ALS Spinal Cord. *PLoS One.* 2016; 11: e0160520.
- Bennett JP, Keeney PM. RNA-Sequencing Reveals Similarities and Differences in Gene Expression in Vulnerable Brain Tissues of Alzheimer's and Parkinson's Diseases. *J Alzheimers Dis Rep.* 2018; 2: 129-137.
- Işıldak U. 2019.
- Somel M, Khaitovich P, Bahn S, Pääbo S, Lachmann M. Gene expression becomes heterogeneous with age. *Curr Biol.* 2006; 16: R359- R360.

Cite this article

Bennett JP, Keeney PM, Brohawn DG (2020) RNA Sequencing of Neurodegenerative Disease Postmortem CNS Tissues from Persons with Amyotrophic Lateral Sclerosis, or Alzheimer's or Parkinson's Diseases Reveal Heterogeneous Dysregulation of Gene Expression. *Ann Neurodegener Dis* 4(1): 1029.