

Raw Data

Raw data should be stored and compressed on the server in your working directory. This data will be 'archived' with appropriate permissions or moved to your new working directory after the course ends. Please request a Xanadu user account if you need substantial storage for your analysis.

Process and Directory Structure

Define a directory structure for your project within your working directory. Organize the data with informative names.

The entire process of the applied workflow should be documented and organized in Github, including the code itself and appropriate visualizations. This can also reference server paths as appropriate.

- **Overview and Motivation:** Provide an overview of the project goals and the motivation. Background of the research: Amyotrophic lateral sclerosis (ALS) is a non-demyelinating neurodegenerative disease in adults characterized by progressive muscle paralysis. Muscle paralysis is determined by the degeneration of motoneurons in the motor cortex brainstem and spinal cord. The ALS pathogenetic mechanisms are still unclear. The OXTR, a 7-transmembrane G protein-coupled receptor capable of binding to either Gi or Gq proteins, activates a set of signaling cascades, such as the MAPK, PKC, PLC, or CaMK pathways. The cellular response to OXT include of neurite outgrowth, cellular viability, and increased survival. RNA-sequencing (RNA-Seq) technology is a high-throughput sequencing technology which has developed rapidly in recent years. It can study gene function, gene structure and gene expression changes at the level of gene transcription and explore the genetic evidence supporting ALS treatment of peripheral nerve injury.
- **Related Work/Citations:** If you are using public data, you would want to include relevant descriptions of how the data was generated and cite this information. Your analysis should be different from anything published on the data. If you are using data in your lab, ensure this data can be shared with your partner. A full description of how this data was generated (study design, libraries, sequencing, etc) should be included.
- <https://pubmed.ncbi.nlm.nih.gov/27487029/> Brohawn, D. G., O'Brien, L. C., & Bennett, J. P., Jr (2016). RNAseq Analyses Identify Tumor Necrosis Factor-Mediated Inflammation as a Major Abnormality in ALS Spinal Cord. PloS one, 11(8), e0160520. <https://doi.org/10.1371/journal.pone.0160520> "We isolated total RNA using postmortem cervical spinal section homogenates from 7 sporadic ALS and 8 neurologically healthy control samples. We next prepared individual Illumina Total Stranded RNA-sequencing libraries for each sample. We sequenced each library with >50 million 2X150 sequencing reads on the NextSeq500. Our goal was to generate gene expression data for use in systems biology analyses to elucidate differences between the sALS and neurologically healthy control groups that may contribute to disease pathology."
- **Initial Questions:** What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

- **Research goal:** To compare OXTR gene expression level to elucidate genetic differences between the ALS and neurologically healthy control groups that may contribute to disease pathology. Initial questions: What is the association between OXTR gene expression and ALS incidence?

- **Data:** Location, Organization, Source

Illumina Total Stranded RNA sequencing reads from 15 postmortem cervical spinal sections (7 ALS and 8 healthy controls) 1 ILLUMINA (NextSeq 500) run: 71.8M spots, 21.5G bases, 9.7Gb

downloads [https://www.ncbi.nlm.nih.gov/sra/SRX1304782\[accn\]](https://www.ncbi.nlm.nih.gov/sra/SRX1304782[accn]) NCBI SRA RNAseq

- **Workflow/Analysis:** What visualizations did you use to look at your data in different ways? What are the different statistical methods you considered? Why did you select the packages you did?

What visualizations did you use to look at your data in different ways? I am going to check the new gene, OXTR, to look at the association between OXTR and ALS risk by comparing the gene expression level in ALS patients and healthy control group. What are the different statistical methods you considered? I am going to use R, Linux, python which did not mention in the study. Why did you select the packages you did? The OXTR is a pain risk gene which related to etiology of neuropathy. By assessing and comparison of the gene expression level between ALS and healthy control group, we can understand the potential pathological mechanisms of ALS.

- **Discussion:** What did you learn? What additional sequencing and/or analysis is necessary? What questions remain? What new insights were gained?

- What did you learn? During this project preparation, I have learned that how to find the high throughput data and how to create file in github. Also I have learned how to download the RNAseq data from NCBI website. What additional sequencing and/or analysis is necessary? I am going to find the OXTR gene expression level from the data source of RNAseq. What questions remain? The author did not explore novel transcripts, indels, or selective exon usage specific to our sALS group in this dataset. And we may need to examine other gene expression level to see whether OXTR alone or together with other genes to influence the incidence of ALS. What new insights were gained? Our next step is to investigate expression of smaller non-coding RNA's, particularly microRNA's, that regulate mRNA stability in the future study.