Analysis of Differential Cell Type Composition and Gene Expression in Amyotrophic Lateral Sclerosis with Spinal Cord and Cortex Bulk RNA-seq

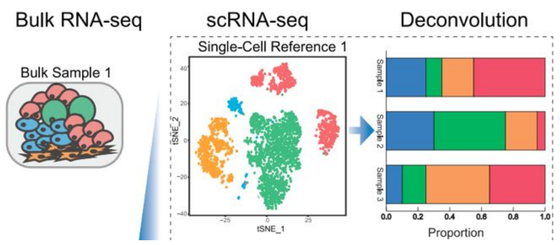
Bhandari, Nischal; Geiger, Heather; Fu, Rui

Computational Biology, New York Genome Center, 101 Avenue of the Americas, New York 10013

Background

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease that impacts the nerves in the brain and spinal cord, leading to progressive loss of voluntary muscle movements. Previous studies have shown differential cell composition in ALS samples and thus inferred the role of certain cell types in the progression of the fatal neurodegenerative disease. Some hallmark changes during the disease were found to be changes in proportion of astrocytes, microglia, and oligodendrocytes. It has been found that the degeneration of motor neurons leads to microglial stimulation which in turn activates astrocytes in previous studies (Town, Nikolic, & Tan 2005; Liddelow et al. 2017). In order to further investigate the composition of cell types in ALS and healthy samples, bulk RNA-seq data from New York Genome Center ALS Consortium were studied.

To further understand the nature of ALS, this research project studied the differential cell type composition and gene expression in ALS samples compared to healthy controls in both spinal cord and cortex. Five different deconvolution methods were employed to examine the cell type composition of bulk samples. Two single cell references – one from cortex and another from spinal cord – were used for deconvolution. Cortical single-cell reference was from Mathys et al. (2019), while spinal single-cell reference was obtained from internal, unpublished source. A general overview of the deconvolution is shown below, while a brief introduction to deconvolution methods are outlined in *figure 2* (*left*).



Materials and Methods

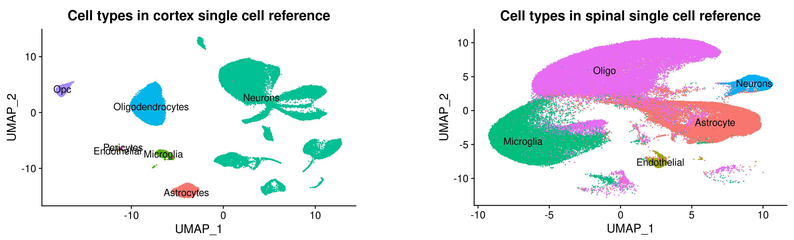
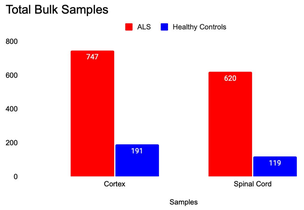


Fig 1: UMAP of Single Cell References, clusters colored by cell types. *Left* : Cortex Single Cell Reference (*total 66,111 cells*). *Right*: Spinal Single Cell Reference (*total cells 113,785*)



Deconvolution Method	Inputs	Principals	Platform + Efficiency
SCDC	Expression sets	Sparse Non-negative matrix	R; Computationally Demanding; Slow
NNLS	Expression set and Single Cell Experiment (SCE) Objects	Non-negative Least Squares regression	R; Computationally demanding; faster than SCDC.
MuSiC	Expression set and SCE Objects	Weighted NNLS	R; Similar to NNLS.
MuSiC2	Control and Case Expression sets and SCE objects	Weighted NNLS + removal of cell-specific DEGs.	R; Computationally heavy; supposedly more accurate.
CIBERSORTx	scSignature Matrix and Bulk Gene Expression Mixture	Support Vector Regression	Web; fast

Fig 2: *Left* : ALS and Control Distribution in Bulk RNA-seq Spinal Cord and Cortex Samples. *Right*: Five different deconvolution methods used in this study. All deconvolution method requires at least single-cell RNA-seq and Bulk RNA-seq data.

Results

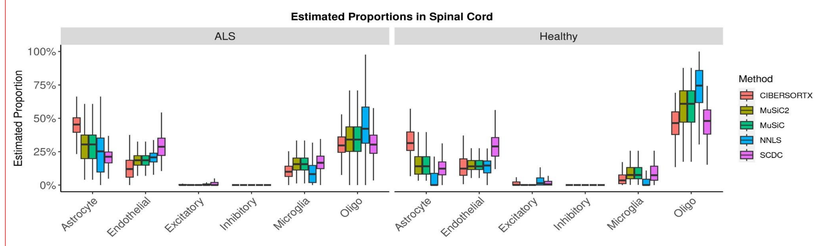


Fig 3: Cell type estimations by several deconvolution methods in Spinal Cord Samples from Spinal Cord Single Cell Reference. Kruskal-Wallis p-value < 2.2e-16 suggested significant difference in estimates across at least two methods. Estimated median proportion for MuSiC and MuSiC2 tend to be similar across cell types whereas CIBERSORTX and SCDC tend to different from all other methods.

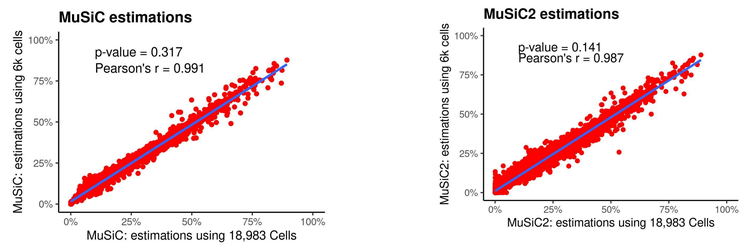


Fig 4: Deconvolution using MuSiC and MuSiC2 with 18k+ cells vs downsampled 6k cells. MuSiC makes similar estimations with the subset of single cell reference.

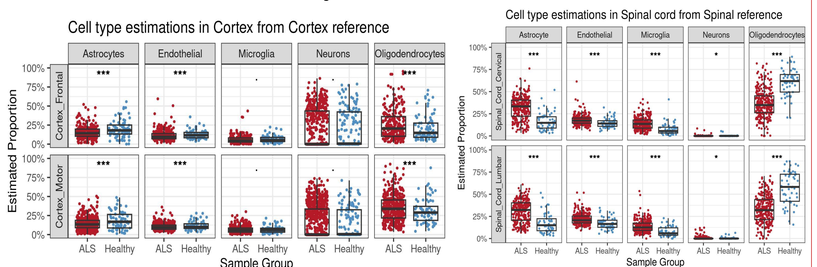


Fig 5: Deconvolution Results in Bulk Cortex (*left*) and Bulk Spinal Cord Samples (*right*) from cortex single cell and spinal single cell reference respectively. Aestricks (*) denotes the level of significance in estimation between ALS and Control Samples. P-values were calculate using two-sided Wilcoxon nonparametric test after regressing for technical covariates – sites of sample collection and Sequencing platform – and then the p-values were adjusted using Bonferroni correction. *** := < 0.0001, ** := < 0.001, * := < 0.05, . := > 0.05.

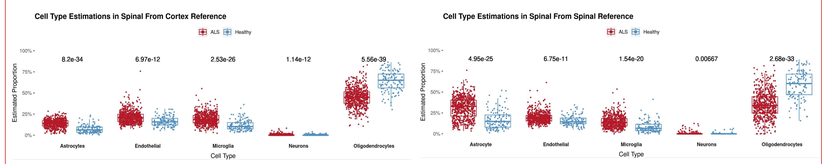


Fig 6: Overall Cell Type Estimation in Spinal Bulk Samples with adjusted p-values. Deconvolution with Cortex Single Reference (*left*) and Spinal Single Cell Reference (*right*).

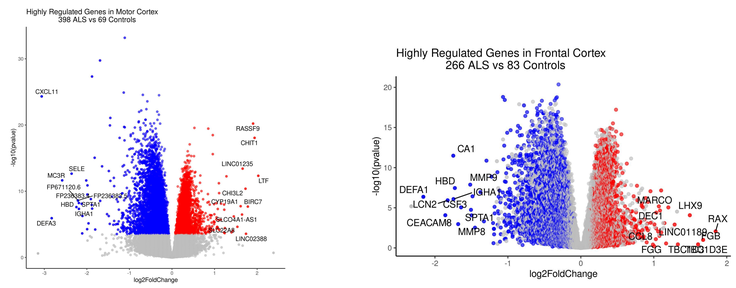


Fig 7: *Left*: Differentially Regulated Genes in motor cortex samples; *Right*: Regulated genes in frontal cortex. Upregulated genes (adjusted p-value < 0.001) are colored in Red, downregulated genes (adjusted p-value < 0.001) are colored in blue, and non-significant genes are colored in gray.

Results

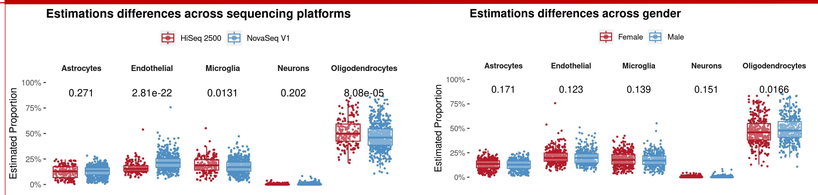


Fig 8: Estimation differences across platform (*left*) and gender (*right*). Estimations differences in cell types across genders were not found to be significantly different except oligodendrocytes. In contrast, there were significant differences in estimation in cell types across platforms. To account for such technical differences, linear regression was performed on technical covariates (sequencing platforms and site of specimen collection) to calculate p-values.

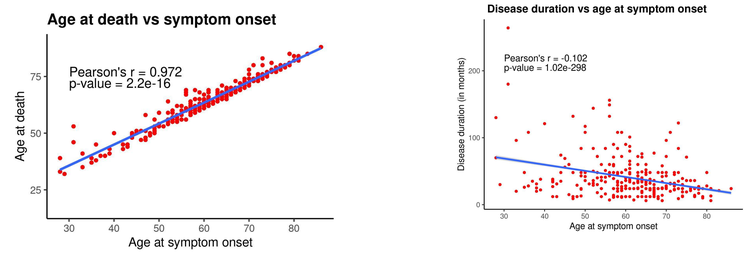


Fig 9: Correlation between age at death and symptom onset (*left*) and disease duration and symptom onset (*right*) in spinal cord samples.

Conclusion and Future Directions

- Different deconvolution methods tend to make similar conclusions. All methods predicted Astrocytes and Microglia to be higher in ALS cases in spinal cord. MuSiC was selected to perform the rest of the studies due to its ease of use and ability to make consistent estimations from the subset of single cell reference.
- Significant differences were found between ALS and control cases across cell type estimations. Astrocytes and Microglia were higher both in the cervical and lumbar ALS spinal cord samples compared to control samples. However, Oligodendrocytes were estimated to be significantly higher in control samples. In contrast, cortex was estimated to have more Astrocytes in control samples and more Oligodendrocytes in ALS samples.
- CCL18 was upregulated and CXCL11 was downregulated in frontal cortex. CHIT1 was upregulated in motor cortex. All of these genes are involved in inflammatory and immune responses.
- Significant positive correlation was found between age at symptom onset and age at death whereas significant negative correlation was observed between age at symptom onset and disease duration.
- Other cell types like fibroblasts and macrophages found to play role in ALS pathogenesis were not studied in this project and may be a subject of further research.
- While differences between ALS cases and controls were found to be significant, the findings might have been limited by the fewer control samples. In the future, increasing the sample size might add to the validity and reliability of the findings.

References

1. Humphrey *et al.* Integrative transcriptomic analysis of the amyotrophic lateral sclerosis spinal cord implicates glial activation and suggests new risk genes. (2023).
2. Meichen Dong and others, SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references, *Briefings in Bioinformatics*. (2021)
3. Newman, A.M., Steen, C.B., Liu, C.L. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 37, 773–782 (2019).
4. Wang, X., Park, J., Susztak, K. *et al.* Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* 10, 380 (2019).
5. Town, T., Nikolic, V., & Tan, J. (2005). The microglial "activation" continuum: from innate to adaptive responses. *Journal of neuroinflammation*, *2*, 24.
6. Liddelow, S. A. *et al.*(2017). Neurotoxic reactive astrocytes are induced by activated microglia. *Nature*, *541*(7638), 481–487.
7. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. Nature 570, 332–337 (2019).

Acknowledgments

We are grateful for the opportunity granted by the NYGC Summer Internship Program and assistance provided by Computational Biology Lab throughout to pursue this research. We also want to thank the ALS Consortium and Synapse Data Access System for providing us with RNA-seq data used to conduct the study.