

Transcriptomics Analyses of ALS Post-mortem Motor Cortex highlight alteration and potential biomarkers in the Neuropeptide Signalling pathway

Renata Kabiljo*¹, Heather Marriott*^{1,2}, Guy P Hunt^{1,2,3,4}, Abigail L Pfaff^{3,4}, Ahmad Al Khleifat², Brett Adey^{1,9}, Ashley Jones², Claire Troakes¹⁰, John P Quinn⁵, Richard J B Dobson^{1,7,8}, Sulev Koks^{3,4,\$}, Ammar Al-Chalabi^{2,6}, Alfredo Iacoangeli^{1,2,8,\$}

*R.K. and H.M. have contributed equally to the manuscript

^{\$}correspondence should be addressed to alfredo.iacoangeli@kcl.ac.uk and Sulev.Koks@murdoch.edu.au

¹Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

²Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 9NU, UK

³Perron Institute for Neurological and Translational Science, Nedlands, WA 6009, Australia

⁴Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Murdoch, WA 6150, Australia

⁵Department of Pharmacology and Therapeutics, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 3BX, UK

⁶King's College Hospital, Bessemer Road, Denmark Hill, London, SE5 9RS, UK

⁷Institute of Health Informatics, University College London, London, UK

⁸National Institute for Health Research Biomedical Research Centre and Dementia Unit at South London and Maudsley NHS Foundation Trust and King's College London, London, UK

⁹Social Genetic and Developmental Psychiatry Centre, Memory Lane, London, SE5 8AF

¹⁰MRC London Neurodegenerative Diseases Brain Bank, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

Supplementary Methods

RNaseq protocol for samples from the MRC London Neurodegenerative Diseases Brain Bank (King's College London)

Purification, isolation and quality control

The 30mg tissue block for RNA was homogenised using a Qiagen PowerLyzer 24 Homogenizer. Total RNA was purified from the homogenate using the standard protocol of the RNeasy Lipid Tissue Mini Kit (Qiagen), with on-column DNase digestion. RNA integrity was estimated using Agilent Bioanalyzer 2100's RNA 6000 Nano assays. RNA quantification was performed using a NanoDrop.

Library preparation and RNA sequencing

Library preparation was performed using the standard Illumina TruSeq Stranded Total RNA Sample Preparation Guide with Ribo-Zero Human/Mouse/Rat (October 2013 Rev E.). Fragmentation steps were tailored to degrees of degraded RNA samples using Agilent Bioanalyzer 2100's Nano assay results from the previous section. Libraries were validated using an Agilent Bioanalyzer 2100 to assess fragment size distribution. Library concentrations were estimated using a Qubit RNA High Sensitivity Assay Kit. Nanomolar (nM) concentrations were estimated using $nM = ng/ul \times (1500/Average\ bp)$. Libraries were sequenced using Illumina HiSeq 4000 flow cells with 150bp paired-end reads with a target depth of 30 million clusters (60 million reads per sample).

TargetALS Sample Selection

In this study, we included 234 samples from 132 subjects from the TargetALS dataset. The distribution of the samples was 112 samples in the lateral motor cortex, 112 in the medial motor cortex and a further 10 samples in unspecified motor cortex. 30 subjects contributed a single sample, while 102 subjects contributed two samples from two different regions of the motor cortex. Figure S1 shows the PCA of the 234 samples, coloured by the motor cortex region.

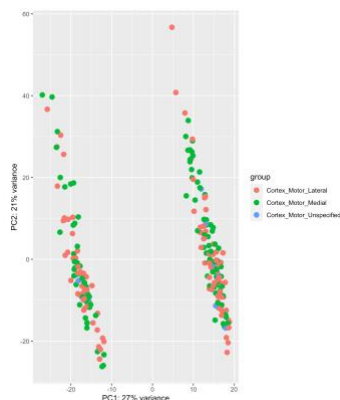


Figure S1. PCA of 241 selected samples from TargetALS. Coloured by motor cortex region.

To assess whether merging the three motor cortex regions to gain statistical power was a suitable approach, we performed separate case-control differential expression analyses of the lateral motor cortex and medial motor cortex and compared them to the results of the same analysis when all motor cortex regions are pooled together. As the unspecified motor cortex includes only ten samples, we did not perform a separate differential expression analysis. In the lateral motor cortex (96 ALS cases and 16 controls), 8 genes crossed genome wide significance (corrected p -value <0.05). All these genes are significant in the merged dataset. In the medial motor cortex (93 ALS cases and 19 controls), 42 genes crossed genome wide significance (corrected p -value <0.05), with 39 of these genes (93%) also being significant in the merged dataset. As the merged dataset gives us more statistical power and still detects the vast majority of differentially expressed genes in the two smaller subsets, we decided to use the merged dataset in our study.

Post-sequencing Bioinformatics Analyses

Clip adapter sequences and low-quality reads were removed using BBDuk (<http://jgi.doe.gov/data-and-tools/bb-tools/>). We used STAR v2.7.10a¹ and Homo_sapiens.GRCh38.104.gtf reference assemblies for alignment of RNAseq reads and HTSeq⁹ version 2.0.0 for transcript quantification. This pipeline is available at https://github.com/rkabiljo/RNASeq_Genes_ERVs. For the KCL BrainBank dataset, we included age, sex, RIN number, PMD and a surrogate variable obtained by SVA² covariates. In addition to these, we also included the sequencing centre as a covariate in the TargetALS dataset. For differential expression analysis, we used DESeq2, with the Wald test used for differential expression, and the Benjamini-Hochberg method used for *p*-value adjustment. The differential expression code is available at https://github.com/rkabiljo/DifferentialExpression_Genes. For MA Plots, apeglm was used for LFC shrinkage³. For enrichment analysis we used MetaCore. Due to a non-one-to-one mapping of the MetaCore algorithm of uploaded gene lists to their proprietary concept of network objects, the resulting enrichment lists consisted of 1,004 objects for KCL BrainBank and 366 for TargetALS. For background genes, we used a list of 41,709 transcripts which had at least 10 counts in at least 5 subjects, which were subsequently mapped in MetaCore to 26,484 network objects. To assess if gene enrichment results were a product of the contribution of different cell types in the bulk RNA sequencing data, we used the BRETIGEA cell type estimation R package (version 1.0.3) under default settings to derive cell proportions for the following cell types; astrocytes, endothelial cells, microglia, neurons, oligodendrocytes and oligodendrocyte progenitor cells (OPCs). Case-control differences in cell type proportions were assessed using one-way ANOVA corrected for age and sex, with a *p*-value of < 0.05 denoting significance. Pearson correlation between melanocortin receptors and their ligands, POMC and AGRP, was performed using the *cor* function of WGCNA⁴ (version 1.71).

ALS Gene Expression Explorer Webserver

The webserver code is publicly available on [GitHub](#) and requires R version 4.1.2 ≤.

References

1. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. Jan 1 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635
2. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet*. Sep 2007;3(9):1724-35. doi:10.1371/journal.pgen.0030161
3. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*. Jun 1 2019;35(12):2084-2092. doi:10.1093/bioinformatics/bty895
4. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. Dec 29 2008;9:559. doi:10.1186/1471-2105-9-559