Cell Reports, Volume *29*

Supplemental Information

Postmortem Cortex Samples Identify Distinct Molecular Subtypes of ALS: Retrotransposon Activation, Oxidative Stress, and Activated Glia

Oliver H. Tam, Nikolay V. Rozhkov, Regina Shaw, Duyang Kim, Isabel Hubbard, Samantha Fennessey, Nadia Propp, The NYGC ALS Consortium, Delphine Fagegaltier, Brent T. Harris, Lyle W. Ostrow, Hemali Phatnani, John Ravits, Josh Dubnau, and Molly Gale Hammell

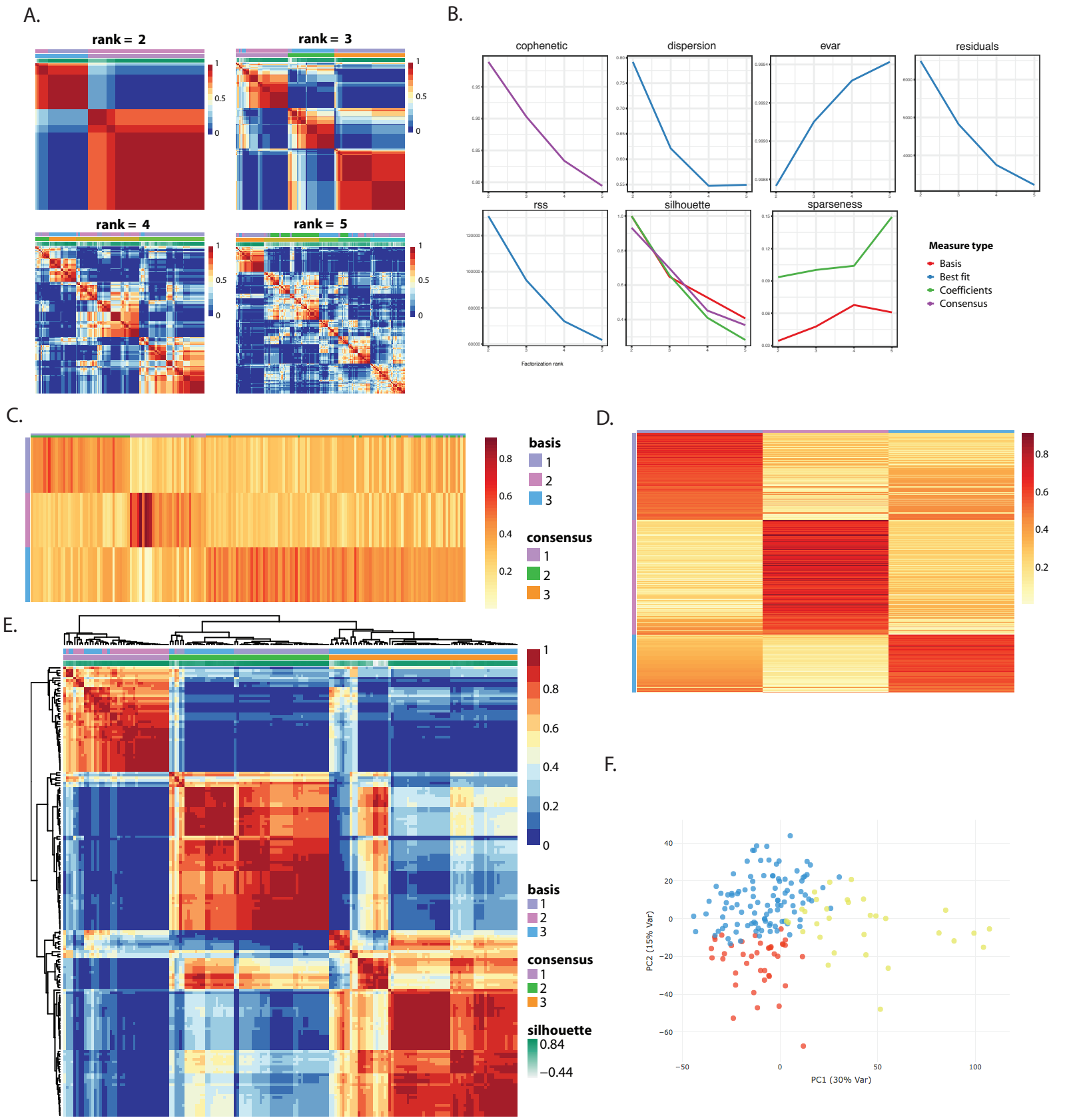


Figure S1. De Novo Clustering of the ALS Patient Transcriptomes. Related to Figure 1

(A) The SAKE algorithm (Ho et al. 2018) for Non-negative Matrix Factorization based de novo clustering was run for 1000 iterations each for a range of potential sub-clusters from k=2-5, with a cluster correlation heatmap shown for each potential value of k.

(B) Statistical measures of cluster consistency for each of the estimated cluster assignments in (A). These analyses indicated an optimum cluster number of k-3.

(C) A heatmap depicting the relative confidence of cluster assignments for each sample in the NYGC transcriptome for k=3 clusters.

(D) A heatmap depicting the relative confidence in each gene marker in determining cluster assignments for k=3 clusters.

(E) A final run of 1000 iterations X 100 SAKE runs was used to robustly determine sample assignments to the 3 clusters, and the updated sample correlation heatmap is displayed.

(F) Scatterplot of the first two principal components (PC1 and PC2) showing distribution of the SAKE identified clusters for ALS-TE (red), ALS-Ox (Blue), and ALS-Glia (yellow).

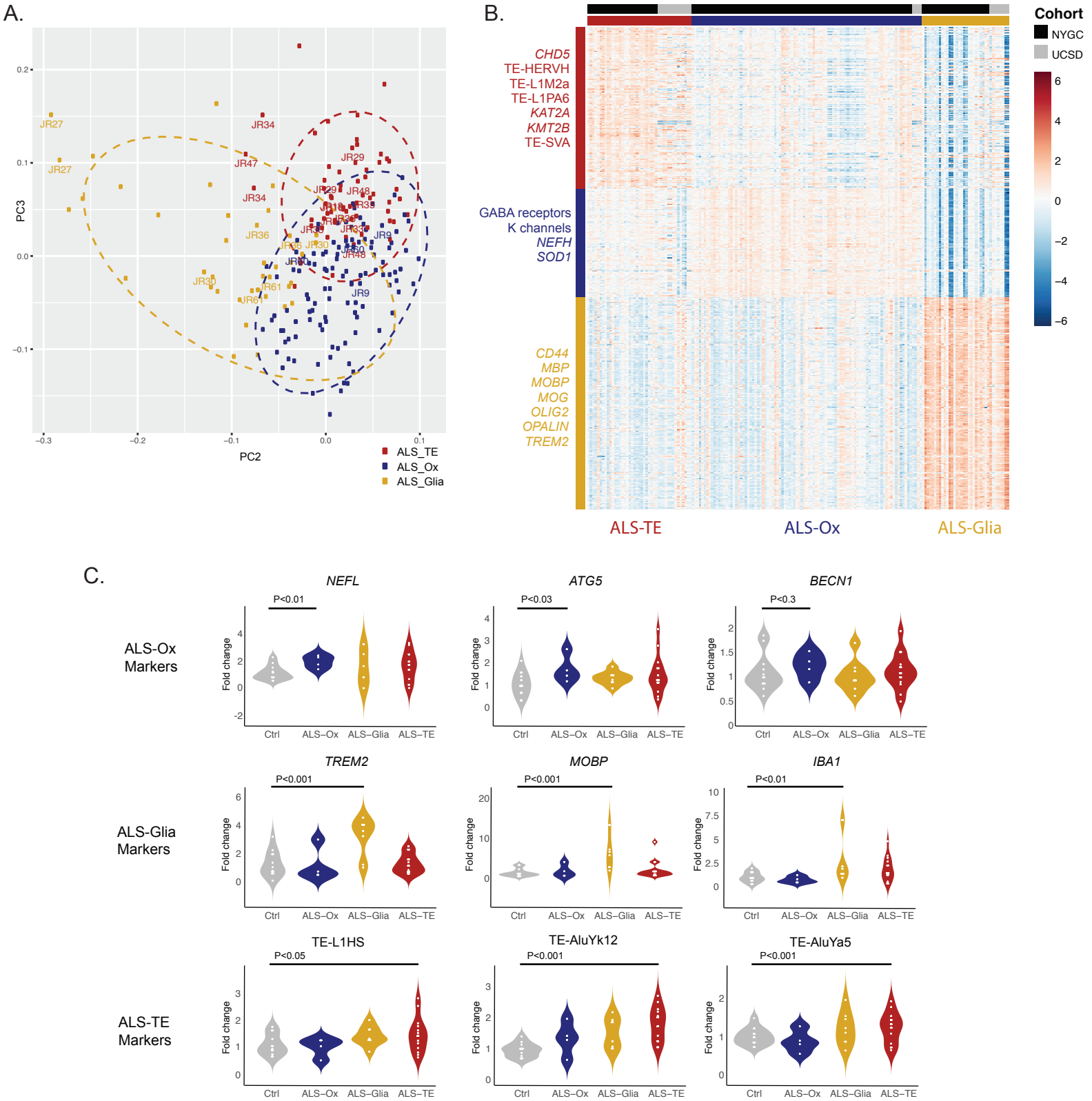


Figure S2. Transcriptome analysis of the UCSD Patient Cohort. Related to Figures 2-4.
The UCSD Patient cohort includes motor cortex tissue 13 sporadic ALS (sALS) patients and 6 non-neurological controls. RNA-seq libraries included biological replicates made from two separate pieces of tissue from each patient, such that the full transcriptome set represents 38 samples.
(A) PCA plot showing the position of UCSD samples (marked by JRxx patient ID numbers) relative to the subgroups within the NYGC ALS Consortium cohort. Samples assigned to the ALS-TE group (red), ALS-Ox group (blue), and ALS-Glia group (gold) fall largely within the 95% confidence interval ellipses defined by the ALS Consortium samples.
(B) Heatmap showing relative gene expression of ALS subtype marker genes for both the NYGC and UCSD cohort together. A barcode at the top of the heatmap shows NYGC samples in black, while UCSD samples are in grey.
(C) qPCR validation of the ALS subtype specific markers was performed for 3 representative markers for each group on all of the UCSD samples. All were statistically enriched in samples from the indicated group relative to control samples, with the exception of Beclin-1 (*BECN1*), a marker of lysosomal dysfunction which shows the sample general pattern of enrichment as the RNA-seq data in the NYGC cohort, but was not statistically significant for this smaller set of samples. For the ALS-TE group, in particular, three young and active retrotransposons were chosen (L1HS, ALuYk12, and AluYa5), which validated both the relative enrichment of expressed retrotransposons in these ALS-TE samples and also validated the informatics methods used to evaluate retrotransposon expression in the RNA-seq data.

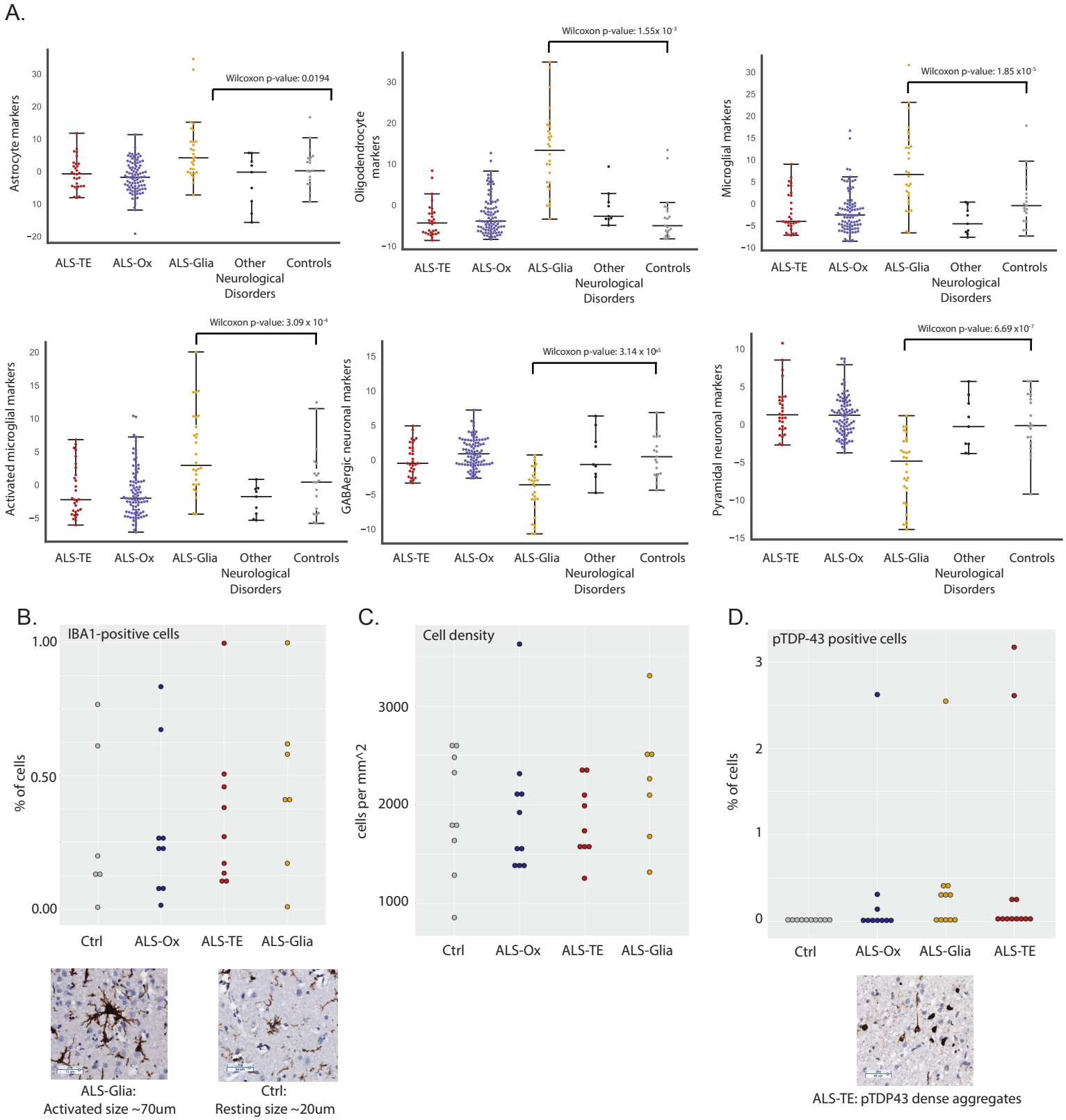


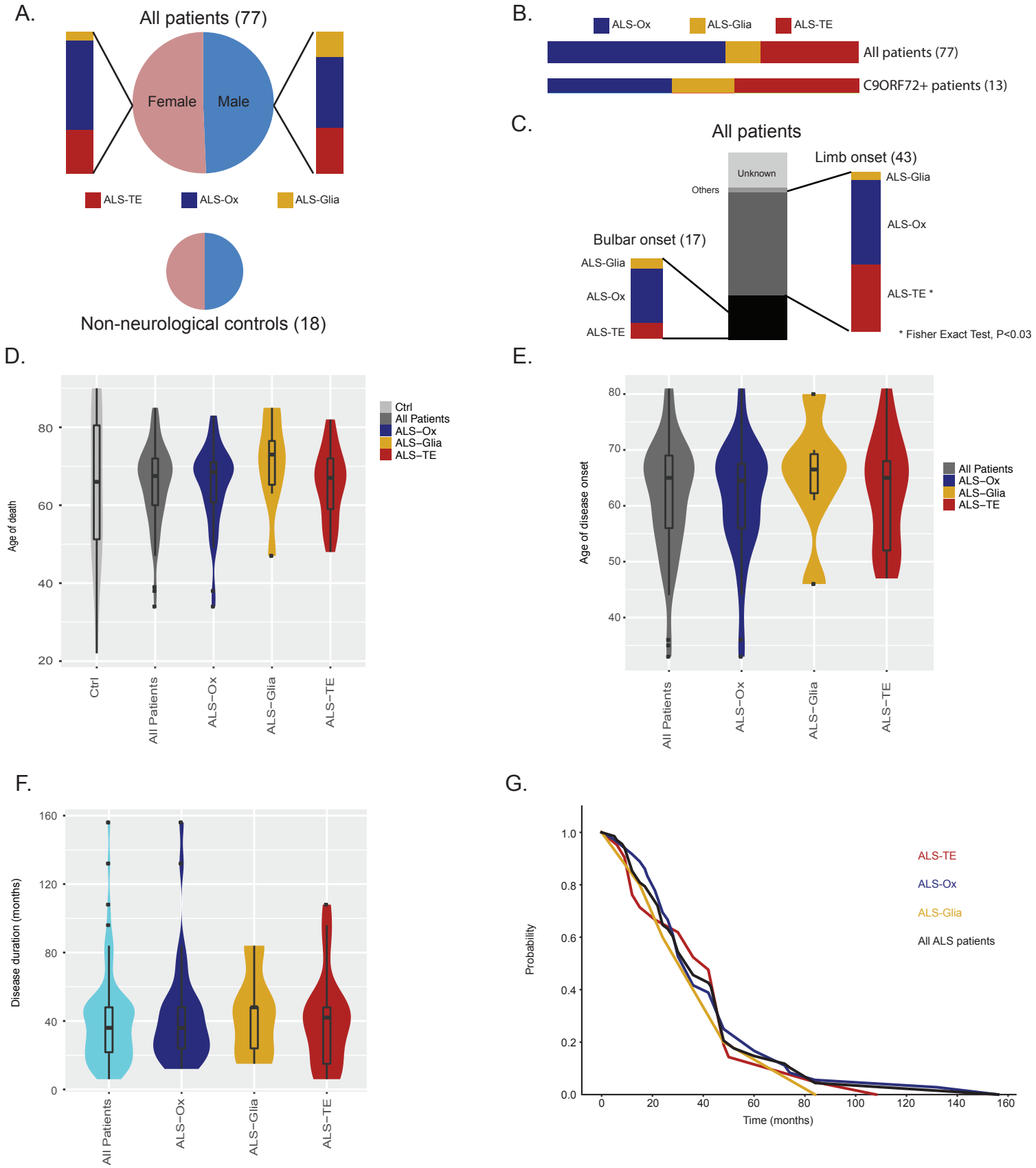
Figure S3. Cell Composition and Histological Analysis of NYGC cohort samples. Related to Figure 2-4.

(A) NeuroExpresso was used to computationally infer the cell types underlying the expression patterns in each of the samples from the three ALS subtypes and the control cortical samples. Significant differences in cell type expression markers were found only for the ALS-Glia samples, which showed enriched expression of markers that typically characterize astrocytes, oligodendrocytes, and microglia. These samples also showed relative depletion in genes that typically mark GABAergic and Pyramidal neurons

(B) IBA1 expression (by IHC) is typically seen in all microglia, both resting and activated, but activated microglia show both higher expression levels and larger cell bodies. Using a threshold cell diameter of 70 µm, relative proportions of activated microglia were counted for 35 ALS and control samples, showing the highest density in ALS-Glia patients. Example snapshots from the stained tissue slides for activated and resting microglia are shown for reference.

(C) Using the Hematoxylin channel, total cell counts in cortex were quantified from each of the subgroups, with no significant differences between the ALS subgroups.

(D) Adjacent slices from the same tissue were also stained for pTDP-43 positive aggregates, which are quantified as a relative percentage of all cells in the top cortical layers and where ALS-TE samples were the most likely to show dense regions of aggregated pathology. A snapshot of multiple adjacent pTDP-43 positive aggregates are shown for a representative slide form the ALS-TE group.



Supplementary Figure. S4. Association of Clinical Parameters with ALS Subtype. Related to Figure 1.
All available clinical parameters were evaluated to determine whether any of these was associated with one of the ALS subtypes.
(A) No statistically significant enrichments were seen between ALS Subtype and sex.
(B) No statistically significant enrichments were seen between ALS Subtype and C9ORF72 repeat expansions.
(C) ALS-TE samples showed a slight enrichment among Limb onset patients (Fisher's Exact P<0.03).
(D) No statistical associations were seen between ALS Subtype and age at death.
(E) No statistical associations were seen between ALS Subtype and age of disease onset.
(F) No statistical associations were seen between ALS Subtype and disease duration by Wilcoxon statistics.
(G) No statistical associations were seen between ALS Subtype and disease duration by Kaplan-Meier analysis.

Supplemental Table Legends.

Table S1. Clinical information for all ALS postmortem cortex samples. Related to Figure 1-4
See TableS1.xlsx
(A) Clinical information for the NYGC ALS Consortium Samples
(B) NYGC ALS Consortium Patient Sample Subtype Assignments
(C) Clinical information for the UCSD ALS Consortium Samples

Table S2. Full list of SAKE-identified features marking each of the ALS and associated enriched pathways. Related to Figures 1-4
See TableS2.xlsx
(A) Full list of genes identified by the SAKE algorithm as marking each ALS Subtype.
(B) GSEA gene sets with FDR < 0.05 in ALS-Ox subtype.
(C) GSEA gene sets with FDR < 0.05 in ALS-Glia subtype.
(D) GSEA gene sets with FDR < 0.05 in ALS-TE subtype.

Table S3. List of TDP-43 bound targets and TDP-43 dependent expression changes. Related to Figure 5
See TableS3.xlsx
(A) Full list of gene targets bound by TDP-43 in the eCLIP data.
(B) Full list of TE targets bound by TDP-43 in the eCLIP data.
(C) Enriched motifs under TDP-43 eCLIP peaks.
(D) Differentially expressed genes and TEs upon TDP-43 shRNA knockdown.

Table S4. Oligonucleotide sequences used in this study. Related to STAR methods

Primer name	Primer sequence	Source/ Reference	Identifier
AIF1/IBA1 forward primer	ATGAGCCAAACCAGGGATTTAC	Spandidos et al. 2010	PrimerBankID: 14574568a1
AIF1/IBA1 reverse primer	GGGATCGTCTAGGAATTGCTTGT	Spandidos et al. 2010	PrimerBankID: 14574568a1
AluYa5 forward primer	CAGCTACTTGGGAGGCTGA	Prudencio et al. 2017	NA
AluYa5 reverse primer	GATCTCGGCTCACTGCAAG	Prudencio et al. 2017	NA
AluYk12 forward primer	CGCCTGTAATCCCAGCTACT	Prudencio et al. 2017	NA
AluYk12 reverse primer	ACTGCAGTGGCGCTATCC	Prudencio et al. 2017	NA
ATG5 forward primer	AAAGATGTGCTTCGAGATGTGT	Spandidos et al. 2010	PrimerBankID: 92859692c1
ATG5 reverse primer	CACTTTGTCAGTTACCAACGTCA	Spandidos et al. 2010	PrimerBankID: 92859692c1
BECN1 forward primer	GGTGTCTCTCGCAGATTCATC	Spandidos et al. 2010	PrimerBankID: 187608304c2
BECN1 reverse primer	TCAGTCTTCGGCTGAGGTTCT	Spandidos et al. 2010	PrimerBankID: 187608304c2
C9 repeat PCR anchor primer	TACGCATCCCAGTTTGAGACG	Renton et al., 2011	NA
C9 repeat PCR forward primer	6-FAM-AGTCGCTAGAGGCGAAAGC	Renton et al., 2011	NA
C9 repeat PCR reverse primer	TACGCATCCCAGTTTGAGACGGGG GCCGGGGCCGGGGCCGGGG	Renton et al., 2011	NA
GAPDH forward primer	GGAGCGAGATCCCTCCAAAAT	Spandidos et al. 2010	PrimerBankID: 378404907c1

GAPDH reverse primer	GGCTGTTGTCATACTTCTCATGG	Spandidos et al. 2010	PrimerBankID: 378404907c1
L1-5'UTR forward primer	ACGGAATCTCGCTGATTGCTA	Macia et al. 2017	NA
L1-5'UTR reverse primer	AAGCAAGCCTGGGCAA	Macia et al. 2017	NA
MOBP forward primer	CCAGACTCTCCAAAAACCAGAAG	Spandidos et al. 2010	PrimerBankID: 109150417c1
MOBP reverse primer	CAGCCGCTCTTACAGATGCT	Spandidos et al. 2010	PrimerBankID: 109150417c1
NEFL forward primer	ATGAGTTCCTTCAGCTACGAGC	Spandidos et al. 2010	PrimerBankID: 197927150c1
NEFL reverse primer	CTGGGCATCAACGATCCAGA	Spandidos et al. 2010	PrimerBankID: 197927150c1
RPS24 forward primer	CCTGTATGAGAAGAAAAAGACC	Sigma Aldritch	H_RPS24_1
RSP24 reverse primer	AATCTCCAGCTCACTTTTTG	Sigma Aldritch	H_RPS24_1
TARDBP forward primer	GGGTAACCGAAGATGAGAACG	Spandidos et al. 2010	PrimerBankID: 42741653c1
TARDBP reverse primer	CTGGGCTGTAACCGTGGAG	Spandidos et al. 2010	PrimerBankID: 42741653c1
TREM2 forward primer	GGTCAGCACGCACAACTTG	Spandidos et al. 2010	PrimerBankID: 9507203a1
TREM2 reverse primer	CGCAGCGTAATGGTGAGAGT	Spandidos et al. 2010	PrimerBankID: 9507203a1