iScience, Volume *23*

Supplemental Information

A Cell- and Tissue-Specific Weakness of the Protein Homeostasis System Underlies Brain Vulnerability to Protein Aggregation

Rishika Kundra, Christopher M. Dobson, and Michele Vendruscolo

Supporting Information

Transparent Methods

Dataset Acquisition

Healthy brain tissues. Microarray data for healthy brain tissues were acquired from the Allen Brain Atlas (Hawrylycz et al., 2012). Gene expression data for 6 healthy human brains were available, across 900 different tissues. Data were scaled and normalised using the 'scale' function of the Weighted Gene Correlation Network (WGCNA) package (Langfelder and Horvath, 2008). As the Allan Brain Atlas uses multiple probes for each genes, the 'collapseRows' function of WGCNA package was used to get a single expression value for each gene across all samples. The expression values for each gene was then averaged across all six brains to arrive at the final expression value associated with each gene (Kundra et al., 2017).

Tissue specific data. We obtained tissue specific data across various human tissues from a previously published dataset (Su et al., 2004). Certain cancerous tissues and cell lines were removed from the analysis as our aim was to study the expression levels in a healthy state, leaving 77 tissues for the study **(Table S1)**. From these, 27 were characterised as neural tissues and 50 body tissues **(Table S1)**. For the Braak and Non-Braak analysis, brain regions (from the 27 brain tissues) were assigned to either Braak or Non-Braak (Freer et al., 2016). Briefly, brain regions in the Allen Brain Atlas were matched with the closest regions mentioned in the original paper (Braak and Braak, 1991). 7 tissues were found to correspond to different Braak stages **(Table S1)**

Cell type specific data. Single-cell RNAseq data for 4 different cell types found in the brain were obtained from a published dataset (Darmanis et al., 2015). Data were scaled and normalised in R.

Coexpression network construction

We used the data obtained from the Allen Brain Atlas to construct a coexpression network for the genes corresponding to the metastable subproteome associated with Alzheimer's disease (MS) (Kundra et al., 2017) and the genes associated with the protein homeostasis system to get the protein

homeostasis complement (PHC) of the MS. WGCNA was used to construct the coexpression network (Kundra et al., 2017). Briefly, WGCNA is a clustering algorithm based on hierarchical clustering but which uses 'soft thresholding' and the concept of topological overlap or shared neighbours to identify clusters of coexpressed genes. The soft thresholding method assigns a weight to each pair of interacting genes and uses such weight along with the topological overlap to identify modules of coexpressed genes in the expression data.

Calculation of the protection factor *s*

Since our aim was to study the balance between the expression level of aggregation prone proteins and components associated with their regulation, we defined a protection factor, *s*, as the slope of the best-fit line for the relative expression of the MS and its PHC across different tissue or cell types. A line was fitted based on linear regression between the expression of MS and the PHC. The protection factor is a measure of the strength of the balance between the MS and its PHC.

Statistical testing

To evaluate the significance of our results, we used random sets of genes corresponding in number to the MS and PHC and calculated the s for them. We then calculated the difference in s between the body and brain tissues (*Δs*). We repeated these 1000 times to have a frequency distribution of Δs values for random sets of genes. The Δs for our genes of interest is 0.25, which is more than two SD away from the random sets of genes **(Figure S1)**. Thus, our s values are highly robust and significant. The statistical testing was performed using the Scipy module in Python. The comparison of different 's' values was done using a two-tailed t test with a value of α of 0.001.

Tissue	Classification (1)	Classification (2)
Adipocyte	Body	-
Adrenal cortex	Body	-
Adrenal gland	Body	-
Appendix	Body	-
Atrioventricular Node	Body	-
BDCA4+DentriticCells	Body	-
Bone marrow	Body	-
Bronchial Epithelial Cells	Body	-
CD105+_Endothelial	Body	-
CD14+_Monocytes	Body	-
CD19+_BCells(neg._sel.)	Body	-
CD33+_Myeloid	Body	-
CD34+	Body	-
CD4+_Tcells	Body	-
CD56+_NKCells	Body	-
CD71+_EarlyErythroid	Body	-
CD8+_Tcells	Body	-
Cardiac Myocytes	Body	-
Fetal Thyroid	Body	-
Fetal liver	Body	-
Fetal lung	Body	-
Heart	Body	-
Kidney	Body	-
Liver	Body	-
Lung	Body	-
Lymphnode	Body	-
Ovary	Body	-
Pancreas	Body	-
Pancreatic Islet	Body	-
Placenta	Body	-
Prostate	Body	-
Salivary gland	Body	-
Skeletal muscle	Body	-
Skin	Body	-
Smooth muscle	Body	-
Testis	Body	-
Testis Germ Cell	Body	-
Testis Intersitial	Body	-
Testis Leydig Cell	Body	-
Testis Seminiferous Tubule	Body	-
Thymus	Body	-
Thyroid	Body	-
Tongue	Body	-
Tonsil	Body	-

Trachea	Body	-
Uterus	Body	-
Uterus Corpus	Body	-
Whole Blood	Body	-
colon	Body	-
Small intestine	Body	-
Amygdala	Brain	Braak
Caudate nucleus	Brain	Non Braak
Cerebellum	Brain	Non Braak
Cerebellum peduncles	Brain	Non Braak
Ciliary ganglion	Brain	Non Braak
Cingulate cortex	Brain	Braak
Dorsal root ganglion	Brain	Non Braak
Fetal brain	Brain	Non Braak
Globus pallidus	Brain	Non Braak
Hypothalamus	Brain	Braak
Medulla oblongata	Brain	Non Braak
Occipital lobe	Brain	Braak
Olfactory bulb	Brain	Non Braak
Parietal lobe	Brain	Non Braak
Pons	Brain	Non Braak
Prefrontal cortex	Brain	Braak
Spinalcord	Brain	Non Braak
Subthalamic nucleus	Brain	Non Braak
Superior cervical ganglion	Brain	Non Braak
Temporal lobe	Brain	Braak
Thalamus	Brain	Braak
Trigeminal ganglion	Brain	Non Braak
Whole brain	Brain	Non Braak
Pineal day	Brain	Non Braak
Pineal night	Brain	Non Braak
Pituitary	Brain	Non Braak
retina	Brain	Non Braak

Table S1. List of the 77 different tissues used in the analysis. The Braak classification (Braak and Braak, 1991) is reported for brain tissues, related to Figure 1.

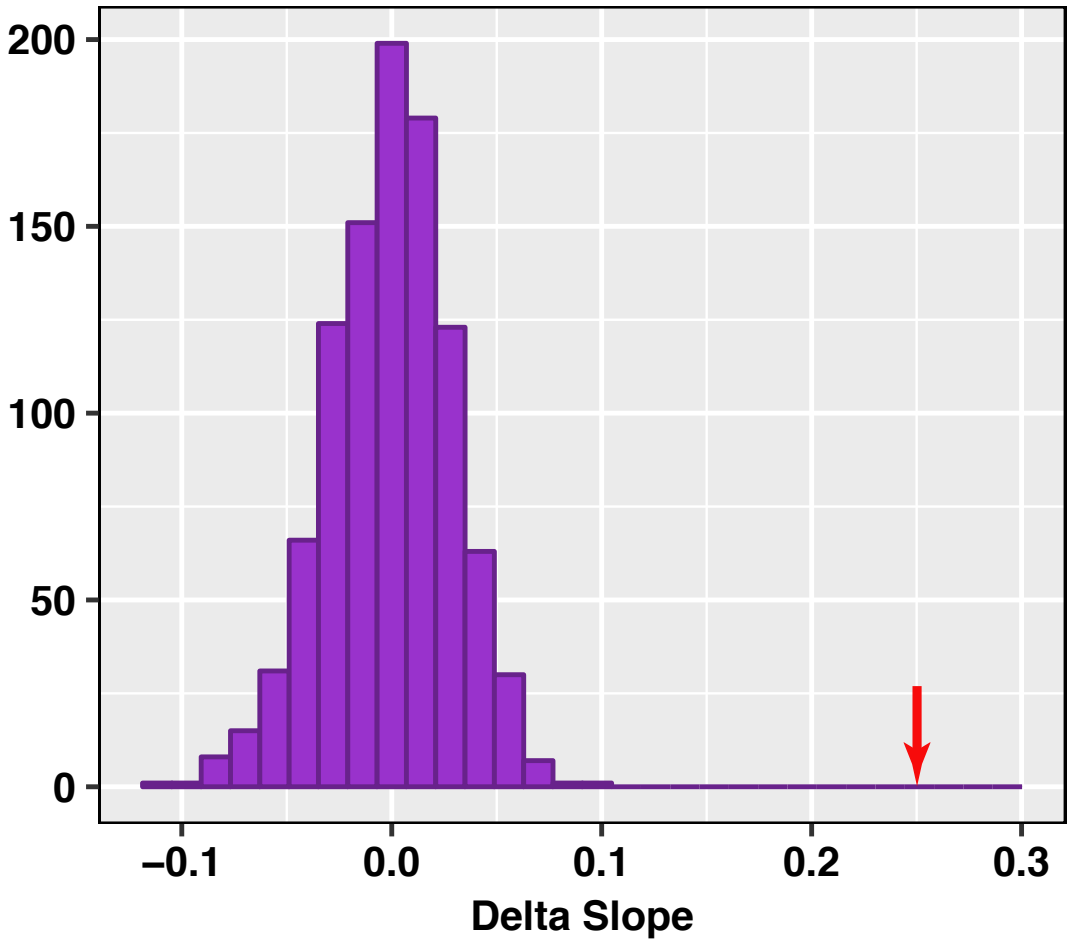

Figure S1. Histogram showing *Δs* (difference in *s* for body tissues and brain tissues) for 1000 random sets of genes, related to Figure 1. The *Δs* for our genes of interest (genes that are supersaturated and downregulated only in Alzheimer's disease, and the associated PHC) is 0.25 (shown by red arrow).

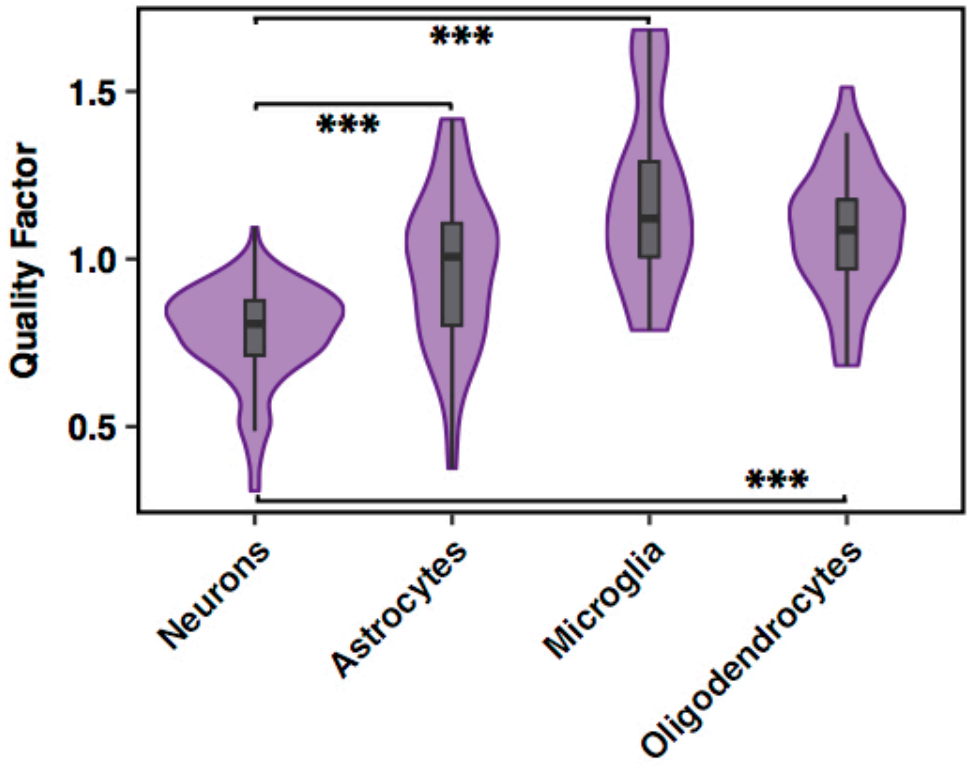

Figure S2. Neurons have a lower value of protection factor *s* compared to other non-neuronal cell types, related to Figure 3. *** $p<0.001$, see Methods.