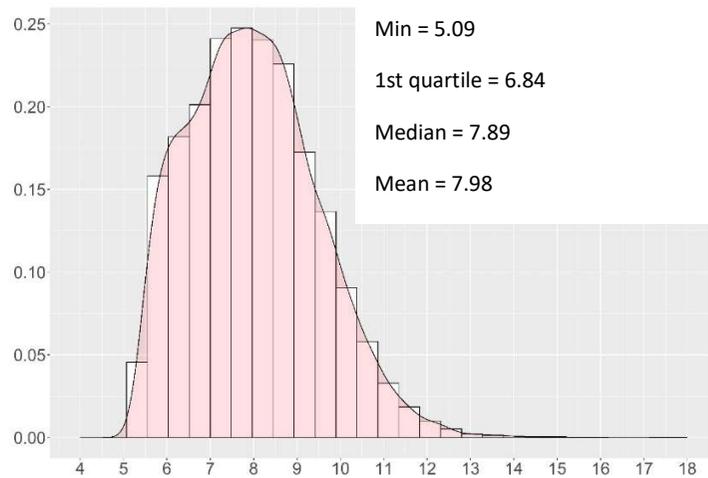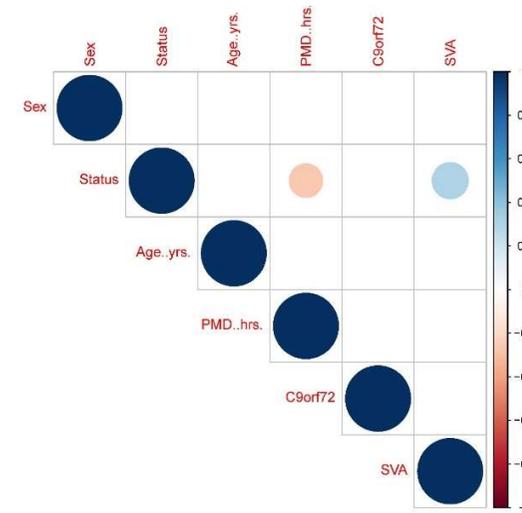# Supplementary Materials

Supplementary Figures

A

B

Supplementary Fig 1. **Distribution and correlational analyses of transcriptome-wide expression with covariates (A)** Histogram of average VST normalised expression by bin with kernel-density estimates by gene superimposed. **(B).** Pearson's correlation between covariates used in the post-mortem eQTL analyses. Size of circle within cell denotes statistical significance, whereas if the cell is empty no statistically significant correlation was found. PMD: Post-Mortem Delay. SVA: Surrogate variables
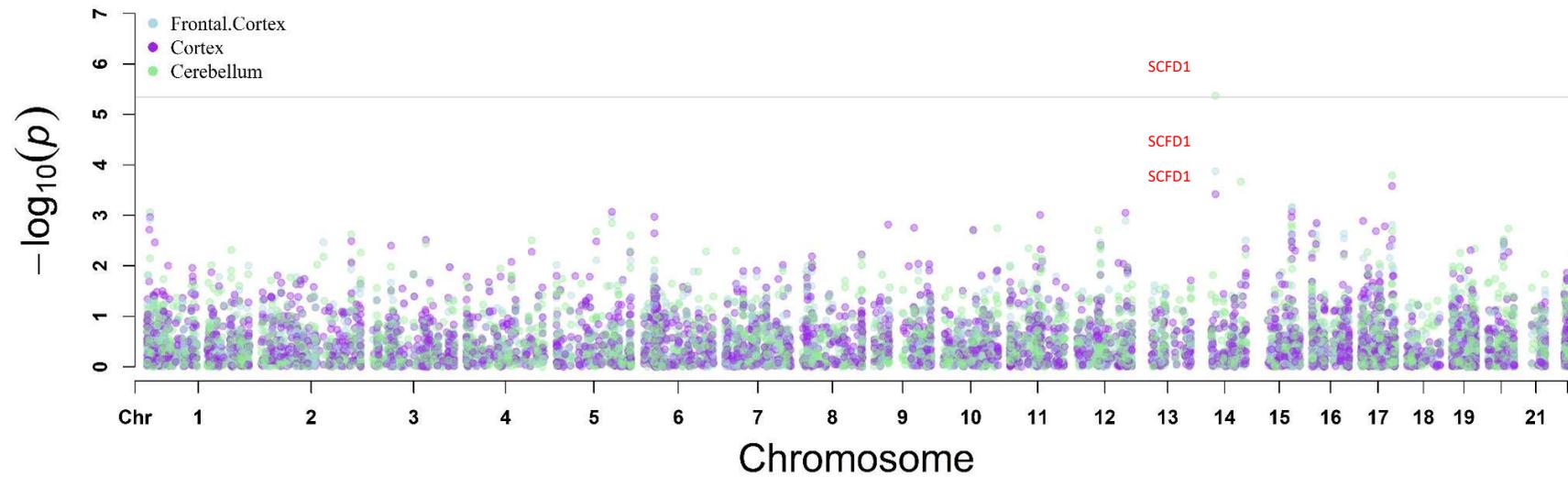
Supplementary Figure 2. Summary-data-based Mendelian Randomization with genes (pHEIDI < 0.05) **removed**. GTEx eQTL data was used from Frontal Cortex, Cortex and Cerebellum tissues (see colour-coded legend, top left) alongside ALS GWAS published by Nicolas et al. 2018. Y-axis: Gene-level -log10 of the p-value from the SMR analysis. X-axis: Chromosome location of gene. Dots: Genes.
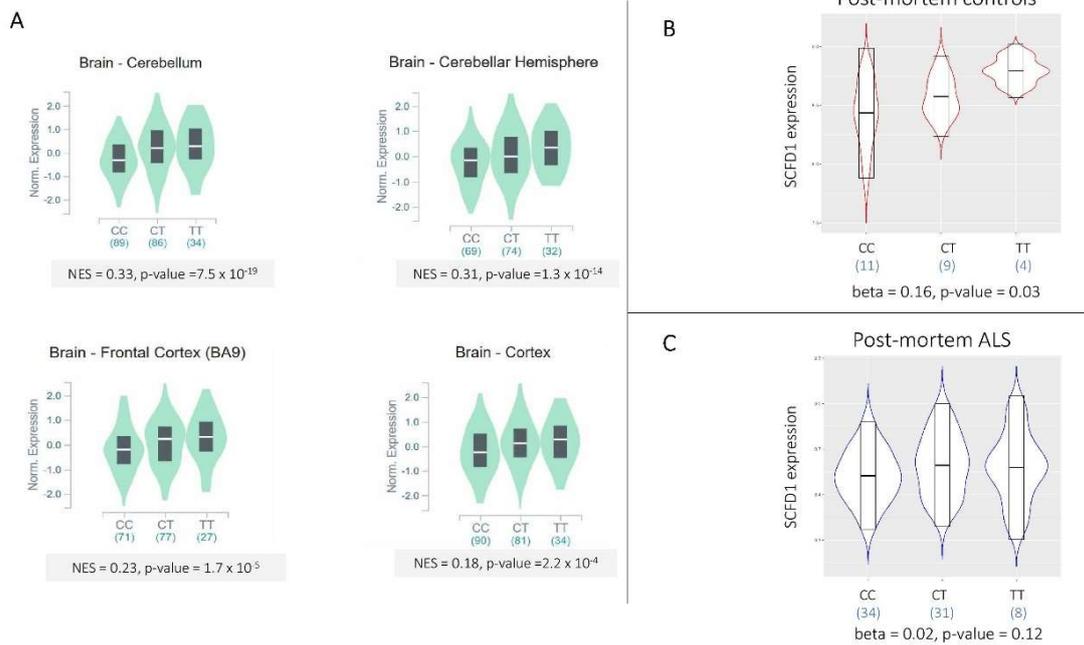
Supplementary Fig 3. Correlation of SCFD1 expression with the most significant SCFD1 ALS GWAS SNP. (A) Correlation between rs10139154 (GWAS SNP) and SCFD1 expression across different tissue types taken from GTEx (dbGaP Accession phs000424.v7.p2 accessed 07/01/2019). **(B)** Correlation between rs10139154 and SCFD1 expression in our post-mortem control dataset. **(C)** Correlation between rs10139154 and SCFD1 expression in our post-mortem ALS dataset. A and B show significant increase in SCFD1 expression with addition T alleles, whereas we do not see the same correlation in our post-mortem ALS samples.
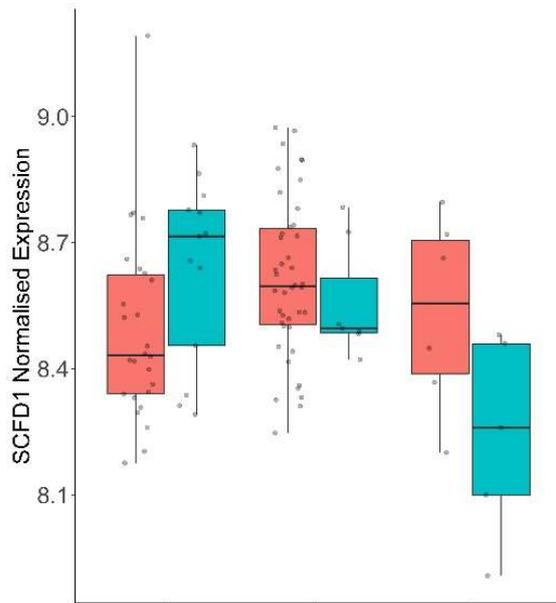
rs35330064 SCFD1 eQTL effect on survival

Reduction in ALS survival by 4.8 months (median) with rarer AA genotype

Hazard ratio = 1.11

P-value = 2.06×10^{-4}

	GG	GA	AA
ALS	27	42	6
Ctrls	13	7	5

Supplementary Fig 4. The relationship between *SCFD1 downstream eQTLS* and SNPs that modify ALS survival. (C) Jitter boxplot showing the effect of rs35330064 on the gene expression of *SCFD1* stratified by disease status

A   B

	TT	TC	CC
ALS	18	42	16
Ctrls	9	11	5
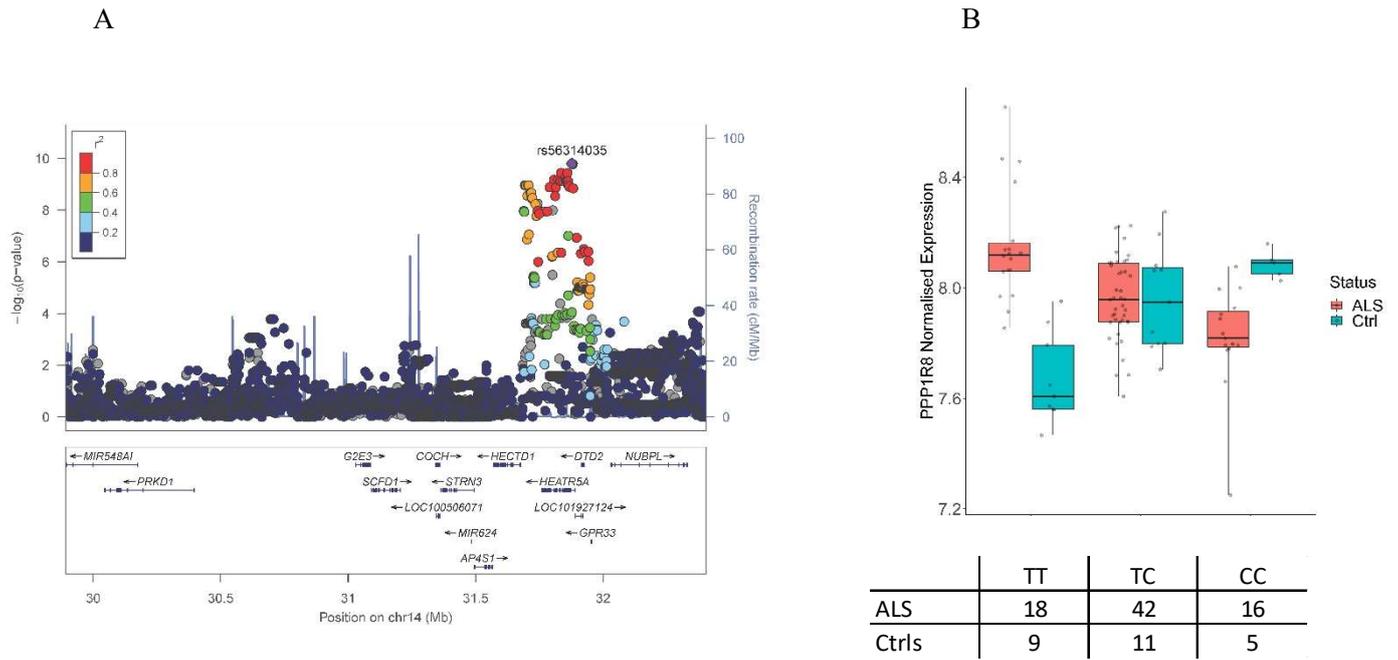
Supplementary Fig 5. The most significant genome-wide trans-acting eQTL to associate with post-mortem ALS is for gene *PPP1R8*, which are the same SCFD1 eQTLs differentially expressed in ALS. See Fig 2B for comparison. **(A).** Locus-zoom plot of trans-acting eQTLs for PPP1R8 that associate with ALS status. Y-axis: The p-value association of the trans-acting eQTLs with ALS. X-axis: genomic location of trans-acting eQTLs. Dots: eQTLs. Coloured dots: LD between SNPs and top trans-acting eQTL rs56314035. **(B).** Jitter-boxplot of PPP1R8 expression (y-axis) and rs56314035 genotypes (x-axis). Controls are coloured red, and ALS cases blue.
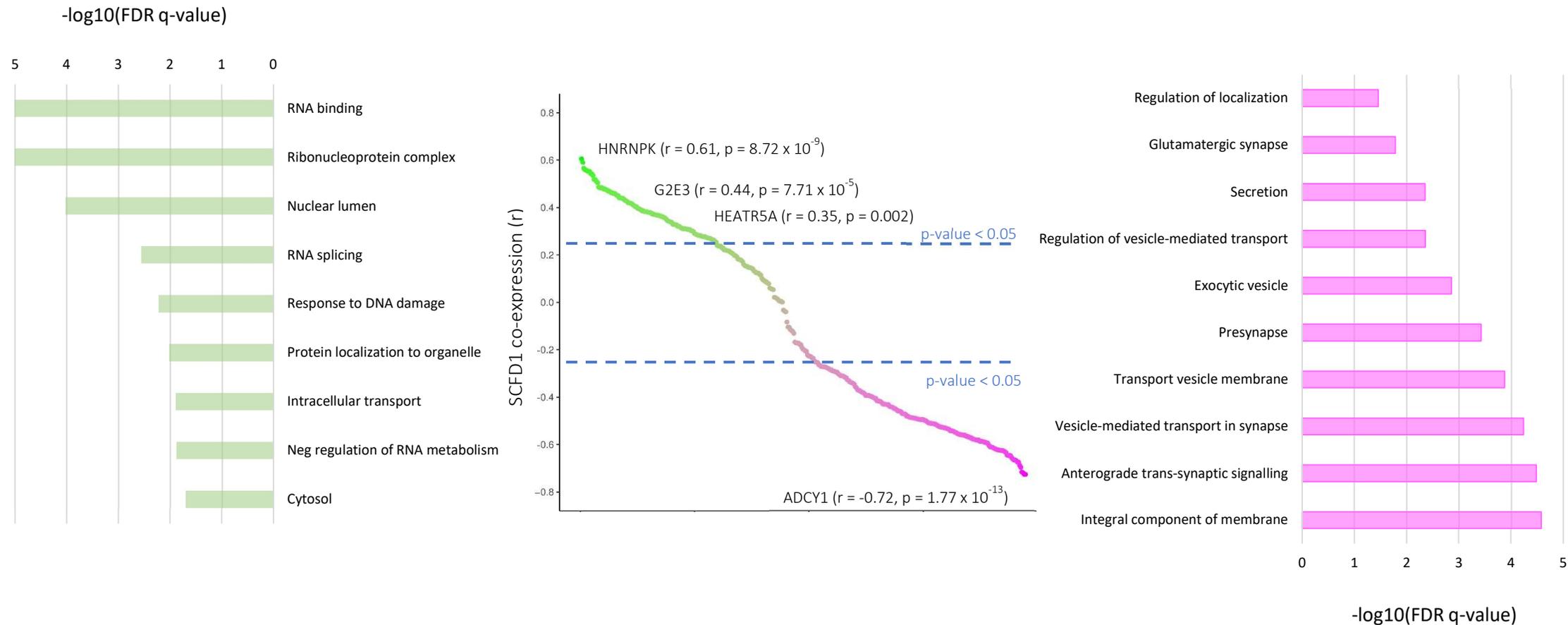
Supplementary Fig 6. Co-expression and GSEA analyses of genes that associate with the ALS *SCFD1* trans-eQTL hotspot. Left: GO categories that are enriched in genes that positively correlate with *SCFD1* expression and associate with the ALS via the *SCFD1* trans-eQTL hotspot. X-axis: -log10 of the FDR q-value derived from the GSEA analysis (capped at FDR q-value $< 1 \times 10^{-5}$). Y-axis: GO categories. Data

labels in bar: Normalised enrichment score (NES). **Centre**: Distribution of co-expression between genes that associate with *post-mortem* ALS via the *SCFD1* trans-eQTL hotspot and *SCFD1* itself. Dots: Genes that associate with ALS via the *SCFD1* trans-eQTL. X-axis: gene count (n = 389). Y-axis: Pearson's r co-expression statistic between gene (dots) and *SCFD1*. Dot-colour: Colour representation of Pearson's r statistic. **Right**: GO categories that are enriched in genes that negatively correlate with *SCFD1* expression and associate with the ALS via the *SCFD1* trans-eQTL hotspot. X-axis: -log10 of the FDR q-value derived from the GSEA analysis (capped at FDR q-value $< 1 \times 10^{-5}$). Y-axis: GO categories. Data labels in bar: Normalised enrichment score (NES).

In summary, Fig 4. shows that genes that associate with *post-mortem* ALS via the *SCFD1* trans-eQTL hotspot are significantly enriched for multiple ALS pathways, depending on their co-expression relationship with *SCFD1*. *SCFD1* trans-eQTL hotspot genes that positively correlate with *SCFD1* expression are enriched for RNA-binding and protein localisation. *SCFD1* trans-eQTL hotspot genes that negatively correlate with *SCFD1* are enriched for trans-synaptic signalling, vesicle-mediated transport and glutamate synaptic activity.

Supplementary Tables

	SAMPLES	SEX (M:F)	AGE (S.D)	PMD (S.D)	RIN (S.D)	C9ORF72 HREM
ALS	76	41:35	68 ($\pm$ 8)	26 ($\pm$ 13)	6.3 ($\pm$ 1.3)	5
CTRL	25	13:12	66 ($\pm$ 9)	36 ($\pm$ 22)	5.3 ($\pm$ 1.7)	0

Supplementary Table 1. Summary of donors with post-mortem motor cortex that were age-sex matched with RNA-sequence and genotype data with averaged attributes. Age: Age of Death in years. PMD: Post-mortem delay in hours. RIN: RNA Integrity Number. C9orf72 HREM: refers to the presence of a hexanucleotide repeat mutation found in C9orf72, which is a major cause of ALS. S.D: Standard Deviation

SYMBOL	CHROMOSOME	NO. OF SNPS	Z STATISTIC	P-VALUE
MOB3B	9	867	8.4106	2.04×10^{-17}
C9ORF72	9	122	7.5267	2.60×10^{-14}
IFNK	9	15	6.7243	8.82×10^{-12}
TNIP1	5	268	5.8297	2.78×10^{-8}
G2E3	14	154	5.2237	8.77×10^{-8}
SCFD1	14	327	4.8742	5.46×10^{-7}
ATXN3	14	294	4.8297	6.84×10^{-7}
GPX3	5	33	4.7056	1.27×10^{-6}
B4GALNT1	12	26	4.6085	2.03×10^{-6}

Supplementary Table 2. Genome-wide gene-level association with ALS using MAGMA and data from the 2018 GWAS (Nicolas et al. 2018).

eQTL rsID (minor allele)	Chromosome	Position	Target Gene Symbol	Beta	Standard Error	P-value	FDR Adjusted P-value	Count of eQTLs in locus (p < 5 x 10^{-5})
rs2782401_G	9	35020923	C9orf24	0.447	0.076	6.30×10^{-8}	0.022	8
rs8005942_A	14	31822149	SCFD1	0.344	0.063	4.45×10^{-7}	0.068	108
rs1167807_A	7	75183403	PMS2P3	0.362	0.069	1.05×10^{-6}	0.081	8
rs1204834_A	6	116513954	ZUP1	-0.300	0.058	1.67×10^{-6}	0.081	20
rs180181_A	17	68039434	ABCA5	0.293	0.058	2.63×10^{-6}	0.081	16
rs34969999_G	20	61737877	LIME1	0.405	0.081	3.48×10^{-6}	0.081	1
rs1451_G	19	58865857	ZNF324	0.197	0.040	4.63×10^{-6}	0.081	7
rs10406143_C	19	9548422	ZNF266	0.328	0.068	6.99×10^{-6}	0.081	155
rs12206581_C	6	28594342	ZKSCAN8	-0.326	0.068	7.83×10^{-6}	0.086	1

Supplementary Table 3. The most statistically significant cis-acting eQTLs found comparing post-mortem ALS donors with controls with p < 5 x 10^{-}

^{5}. The most significant eQTL by rsID is selected at each locus. The beta value denotes the direction and level of expression of the SNP allele attached to rsID (far-left column), when comparing ALS donors to controls. The p-value and FDR-adjusted p-value denote the statistical probability of the association with ALS status. The count of eQTLs in locus are the number of GTEx eQTLs showing an association in that genomic location.

Cell-type	Beta	Std. Error	P-value
Neurons	0.032	0.024	0.185
Astrocytes	0.014	0.024	0.573
Endothelial Cells	0.004	0.023	0.846
Microglia	-0.009	0.024	0.702
Oligodendrocytes	0.024	0.024	0.330
Oligodendrocytes PCs	0.027	0.024	0.271

Supplementary Table 4. Table showing differences in cell-type estimates between ALS and control donors using linear regression models.

BRETIGEA was used to estimate presence of cell-type for each sample. Differences between ALS and control donors were tested using a linear regression model with covariates sex, age of death, post-mortem delay, and surrogate variables. *Beta:* Differences in cell estimates per cell-type between ALS and control donors; *Std. Error*: Standard error; *P-value*: Non-adjusted p-values from the linear regression models.

CATEGORY	TERM NAME	TERM ID	ADJ. P-VALUE	TERM SIZE	INTER SECTIO N	SOURCE
GO:BP*	Cellular localization	GO:0051641	7.1E-07	2974	107	g:profiler
GO:BP*	Regulation of vesicle-mediated transport	GO:0060627	2.2E-05	558	34	g:profiler
GO:BP*	Regulation of transport	GO:0051049	1.1E-03	1915	70	g:profiler
GO:BP*	Regulation of localization	GO:0032879	1.8E-03	2893	94	g:profiler
GO:BP*	Regulation of cellular component organization	GO:0051128	1.8E-03	2488	84	g:profiler
GO:BP*	Cell morphogenesis	GO:0000902	3.0E-03	1050	45	g:profiler
GO:BP*	Regulation of cellular localization	GO:0060341	3.4E-03	918	41	g:profiler
GO:BP*	Cellular component morphogenesis	GO:0032989	7.3E-03	1156	47	g:profiler
GO:BP*	Vesicle-mediated transport	GO:0016192	9.0E-03	2146	73	g:profiler
GO:BP*	Macromolecule localization	GO:0033036	1.2E-02	3228	99	g:profiler
GO:BP*	Protein localization	GO:0008104	2.5E-02	2863	89	g:profiler
GO:BP*	Organelle localization	GO:0051640	3.1E-02	634	30	g:profiler
GO:CC*	Bounding membrane of organelle	GO:0098588	1.3E-03	2071	70	g:profiler
GO:CC*	Cytoplasmic vesicle	GO:0031410	4.1E-03	2346	75	g:profiler
GO:CC*	Intracellular vesicle	GO:0097708	4.3E-03	2349	75	g:profiler
GO:CC*	Organelle membrane	GO:0031090	1.2E-02	3500	100	g:profiler
GO:CC*	Endomembrane system	GO:0012505	1.2E-02	4490	122	g:profiler
GO:BP	Anterograde trans-synaptic signalling	GO:0098916	4.3E-09	706	46	g:profiler
GO:BP	Vesicle-mediated transport in synapse	GO:0099003	7.6E-06	218	21	g:profiler
GO:CC	Presynapse	GO:0098793	3.9E-09	516	37	g:profiler
GO:CC	Glutamatergic synapse	GO:0098978	4.2E-06	359	26	g:profiler
GO:CC	Axon	GO:0030424	1.1E-05	627	35	g:profiler
GO:CC	Exocytic vesicle	GO:0070382	3.1E-05	222	19	g:profiler
GO:CC	Axon terminus	GO:0043679	2.4E-04	120	13	g:profiler
GO:CC	Early endosome	GO:0005769	9.1E-03	355	20	g:profiler
GO:CC	Cytoplasmic vesicle membrane	GO:0030659	9.9E-03	777	33	g:profiler
GO:MF	Potassium ion transmembrane transporter activity	GO:0015079	1.7E-05	164	17	g:profiler

GO:MF	Passive transmembrane transporter activity	GO:0022803	1.2E-03	463	26	g:profiler
KEGG	Endocytosis	KEGG:04144	5.0E-04	243	17	g:profiler
KEGG	Cholinergic synapse	KEGG:04725	6.2E-03	112	10	g:profiler
REAC	Voltage gated Potassium channels	R-HSA-1296072	3.9E-06	43	10	g:profiler
REAC	Activation of NMDA receptors and postsynaptic events	R-HSA-442755	4.6E-03	89	10	g:profiler
REAC	Calmodulin induced events	R-HSA-111933	2.1E-02	34	6	g:profiler
GO:BP	Regulation of long-term neuronal synaptic plasticity	NA	5.8E-03	NA	NA	Enrichr
GO:BP	Cytoskeleton organization	NA	8.3E-02	NA	NA	Enrichr
GO:BP	Regulation of potassium ion transmembrane transport	NA	1.1E-01	NA	NA	Enrichr
GO:BP	Calcium ion transport into cytosol	NA	1.2E-01	NA	NA	Enrichr
GO:BP	Regulation of mRNA splicing via spliceosome	NA	2.3E-01	NA	NA	Enrichr
GO:CC	Ionotropic glutamate receptor complex	NA	2.2E-02	NA	NA	Enrichr
GO:CC	Dendrite	NA	3.8E-02	NA	NA	Enrichr
GO:CC	Perinuclear region of cytoplasm	NA	2.0E-01	NA	NA	Enrichr
GO:CC	Main axon	NA	2.1E-01	NA	NA	Enrichr
GO:CC	Microtubule	NA	2.1E-01	NA	NA	Enrichr
GO:CC	Nuclear body	NA	2.1E-01	NA	NA	Enrichr
GO:MF	Ligand-gated calcium channel activity	NA	7.6E-03	NA	NA	Enrichr
GO:MF	Syntaxin binding	NA	1.2E-01	NA	NA	Enrichr
GO:MF	Ionotropic glutamate receptor binding	NA	2.6E-01	NA	NA	Enrichr
KEGG	Glutamatergic synapse	KEGG:04724	2.0E-02	NA	NA	Enrichr
KEGG	Calcium signalling pathway	KEGG:04020	2.0E-02	NA	NA	Enrichr
KEGG	Nicotine addiction	NA	3.9E-02	NA	NA	Enrichr
KEGG	GABA-ergic synapse	NA	6.0E-02	NA	NA	Enrichr
ARCHS4 TISSUES	Brain	NA	3.6E-43	NA	NA	Enrichr
ARCHS4 TISSUES	Cerebral cortex	NA	1.4E-34	NA	NA	Enrichr
ARCHS4 TISSUES	Spinal cord	NA	7.9E-20	NA	NA	Enrichr

ARCHS4 TISSUES	Motor neuron	NA	2.8E-14	NA	NA	Enrichr
ARCHS4 TISSUES	Prefrontal cortex	NA	4.1E-09	NA	NA	Enrichr
DISGENET	Schizophrenia	NA	4.0E-04	NA	NA	Enrichr
MGI PHENO.	Abnormal CNS synaptic transmission	NA	4.1E-03	NA	NA	Enrichr
MGI PHENO.	Impaired coordination	NA	6.5E-02	NA	NA	Enrichr
MGI PHENO.	Abnormal inhibitory postsynaptic currents	NA	7.2E-02	NA	NA	Enrichr
MGI PHENO.	Abnormal learning/memory/conditioning	NA	8.9E-02	NA	NA	Enrichr

Supplementary Table 5. GO, KEGG, REAC, ARCHS4 Tissue, phenotype and DisGeNet ontological categories that show significant enrichment of genes that correlate with *SCFD1 post-mortem* eQTLs and show significant changes in post-mortem ALS.

* Indicates SCFD1 is a member of the ontological term.

GO ID	GO TERM	TERM SIZE	ES	NES	FDR Q-VALUE	FWER P-VALUE
GO:0099536	synaptic signalling	50	-0.553	-2.527	6.62E-05	2.00E-04
GO:0031224*	intrinsic component of membrane	134	-0.47	-2.495	6.57E-05	4.00E-04
GO:0099537	trans-synaptic signalling	48	-0.553	-2.491	4.38E-05	4.00E-04
GO:0098916	anterograde trans-synaptic signalling	46	-0.553	-2.488	3.28E-05	4.00E-04
GO:0016021*	integral component of membrane	133	-0.466	-2.486	2.63E-05	4.00E-04
GO:0097458	neuron part	91	-0.486	-2.479	3.31E-05	6.00E-04
GO:0007268	chemical synaptic transmission	46	-0.553	-2.47	4.73E-05	0.001
GO:0099003	vesicle-mediated transport in synapse	21	-0.661	-2.451	5.78E-05	0.001
GO:0044425*	membrane part	171	-0.449	-2.424	9.56E-05	0.003
GO:0044456	synapse part	63	-0.506	-2.412	1.06E-04	0.003
GO:0099504	synaptic vesicle cycle	19	-0.675	-2.4	1.15E-04	0.004
GO:0030658	transport vesicle membrane	16	-0.694	-2.392	1.33E-04	0.005
GO:0048167	regulation of synaptic plasticity	20	-0.648	-2.355	2.61E-04	0.010
GO:0006836	neurotransmitter transport	18	-0.658	-2.347	2.86E-04	0.012
GO:0045202	synapse	78	-0.47	-2.319	3.69E-04	0.017
GO:0001505	regulation of neurotransmitter levels	19	-0.642	-2.318	3.54E-04	0.017
GO:0007610	behaviour	23	-0.603	-2.307	3.80E-04	0.019
GO:0099177	regulation of trans-synaptic signalling	33	-0.547	-2.298	3.93E-04	0.021
GO:0007267	Cell-cell signalling	66	-0.474	-2.295	3.90E-04	0.022
GO:0098793	presynapse	37	-0.536	-2.294	3.73E-04	0.022
GO:0050804	modulation of chemical synaptic transmission	33	-0.547	-2.292	3.65E-04	0.023
GO:0005886*	plasma membrane	143	-0.428	-2.286	3.79E-04	0.025
GO:0098984	neuron to neuron synapse	29	-0.562	-2.277	3.94E-04	0.027
GO:0030133	transport vesicle	26	-0.577	-2.266	4.53E-04	0.032
GO:0071944*	cell periphery	144	-0.423	-2.246	5.76E-04	0.042
GO:0050877	nervous system process	30	-0.551	-2.245	5.54E-04	0.042
GO:0098794	postsynapse	40	-0.513	-2.242	5.51E-04	0.044

GO:0003723	RNA binding	38	0.549	3.139	0.00E+00	0.000
GO:1990904	ribonucleoprotein complex	16	0.706	2.936	0.00E+00	0.000
GO:0044428	nuclear part	107	0.37	2.771	4.07E-05	0.001
GO:0031981	nuclear lumen	99	0.373	2.694	9.62E-05	0.003
GO:0005654	nucleoplasm	84	0.379	2.646	1.56E-04	0.005
GO:0005634	nucleus	145	0.341	2.643	1.65E-04	0.005
GO:0003676	nucleic acid binding	76	0.388	2.637	1.66E-04	0.005
GO:0006396	RNA processing	21	0.556	2.563	2.74E-04	0.011
GO:0070013	intracellular organelle lumen	113	0.322	2.41	7.53E-04	0.040

Supplementary Table 6. Ontological categories that are significantly enriched (FDR q-value < 0.01) for genes that co-express with SCFD1 and associate with *SCFD1* trans-eQTL network. Red ontological IDs indicate negative correlation with *SCFD1* expression and blue ontological IDs indicate positive correlation with *SCFD1* expression. ES = Enrichment score; NES = Normalised enrichment score; FDR = False Discovery Rate; FWER = Family-wise error rate.

SOURCE	TERM_NAME	ADJUSTED P-VALUE
GO:MF	voltage-gated cation channel activity	1.18E-05
GO:MF	voltage-gated potassium channel activity	2.21E-05
GO:MF	potassium ion transmembrane transporter activity	3.46E-05
GO:MF	metal ion transmembrane transporter activity	3.72E-05
GO:MF	voltage-gated ion channel activity	3.04E-04
GO:MF	gated channel activity	3.18E-04
GO:MF	voltage-gated channel activity	3.20E-04
GO:MF	potassium channel activity	3.71E-04
GO:BP	synaptic signaling	6.95E-08
GO:BP	trans-synaptic signaling	2.33E-07
GO:BP	anterograde trans-synaptic signaling	1.02E-06
GO:BP	chemical synaptic transmission	1.02E-06
GO:BP	nervous system development	1.15E-06
GO:BP	neuron projection development	2.54E-06
GO:BP	neuron development	3.48E-05
GO:BP	modulation of chemical synaptic transmission	9.31E-05
GO:BP	localization	9.35E-05
GO:BP	regulation of trans-synaptic signaling	9.61E-05
GO:BP	regulation of neuron projection development	1.09E-04
GO:BP	plasma membrane bounded cell projection organization	1.25E-04
GO:BP	potassium ion transmembrane transport	1.93E-04
GO:BP	cell projection organization	2.22E-04
GO:BP	neurogenesis	2.78E-04
GO:BP	regulation of transport	2.95E-04
GO:BP	regulation of neuron differentiation	3.04E-04
GO:BP	regulation of synaptic plasticity	3.57E-04
GO:BP	transport	3.78E-04
GO:BP	regulation of localization	4.69E-04
GO:CC	synapse	1.44E-13
GO:CC	neuron projection	3.19E-07
GO:CC	postsynaptic density	1.63E-06
GO:CC	asymmetric synapse	1.92E-06
GO:CC	postsynapse	2.79E-06
GO:CC	axon	3.55E-06
GO:CC	postsynaptic specialization	4.56E-06
GO:CC	neuron to neuron synapse	4.92E-06
GO:CC	cell projection	1.04E-05
GO:CC	presynapse	2.28E-05
GO:CC	somatodendritic compartment	2.53E-05
GO:CC	dendrite	4.88E-05
GO:CC	dendritic tree	5.12E-05
GO:CC	voltage-gated potassium channel complex	9.39E-05
GO:CC	plasma membrane bounded cell projection	1.02E-04
GO:CC	transport vesicle	1.34E-04

GO:CC	synaptic vesicle	1.61E-04
GO:CC	potassium channel complex	2.31E-04
GO:CC	plasma membrane region	3.17E-04
GO:CC	exocytic vesicle	3.29E-04
GO:CC	synaptic membrane	3.61E-04
GO:CC	transmembrane transporter complex	0.002
GO:CC	neuromuscular junction	0.007
GO:CC	cell junction	0.011
GO:CC	whole membrane	0.012
GO:CC	secretory vesicle	0.016
GO:CC	cytoplasmic vesicle membrane	0.019
GO:CC	vesicle membrane	0.027
GO:CC	glutamatergic synapse	0.028
GO:CC	neuron projection terminus	0.031
GO:CC	intrinsic component of plasma membrane	0.033
GO:CC	postsynaptic density membrane	0.033
GO:CC	vesicle	0.034
GO:CC	integral component of plasma membrane	0.039
KEGG	Cholinergic synapse	0.000
KEGG	Long-term potentiation	0.001
KEGG	Aldosterone synthesis and secretion	0.009
REAC	Neuronal System	1.97E-08
REAC	Potassium Channels	3.97E-05
REAC	Long-term potentiation	2.29E-04
REAC	Voltage gated Potassium channels	4.63E-04
REAC	Ras activation upon Ca2+ influx through NMDA receptor	0.004
REAC	Negative regulation of NMDA receptor-mediated neuronal transmission	0.007
REAC	Unblocking of NMDA receptors, glutamate binding and activation	0.007
REAC	Post NMDA receptor activation events	0.013
REAC	Transmission across Chemical Synapses	0.019
REAC	CREB1 phosphorylation through NMDA receptor-mediated activation of RAS signaling	0.021
REAC	Activation of NMDA receptors and postsynaptic events	0.031
WP	Vitamin D-sensitive calcium signaling in depression	0.045

Supplementary Table 7. Enrichment analysis of 148 genes that associated with post-mortem ALS via the *SCFD1* trans-eQTL hotspot and showed significant association with increased Schizophrenia risk using MAGMA gene-set analysis.

Supplementary Methods

1. Gene-level genome-wide association analysis

To confirm association between SCFD1 and ALS, we used a gene-level analysis based on the 2018 ALS GWAS SNP-level summary statistics. This analysis assesses the association between all SNPs within each gene simultaneously and ALS. This approach has been shown to be better powered to detect genes containing many risk alleles that associate with a phenotype and it has been successfully used in other complex diseases including frontotemporal dementia. We performed this analysis using the MAGMA SNPtoGENE protocol on the FUMA webserver. The SNPs were mapped to 18,067 protein coding genes. Therefore, genome-wide significance was defined according to the conservative Bonferroni correction method at p-value = $0.05/18,067 = 2.8 \times 10^{-6}$.

For gene-level and gene-set analyses performed using the 2018 ALS GWAS and 2018 schizophrenia GWAS we used MAGMA for Linux. For the 2016 ALS GWAS there were 18,444 viable genes tested (Bonferroni p-value = $0.05/18,444 = 2.7 \times 10^{-6}$). For the 2014 schizophrenia GWAS there were 19,078 viable genes tested (Bonferroni p-value = $0.05/19,078 = 2.6 \times 10^{-6}$). For these analyses we implemented a flanking region 30kb upstream and 12.5kb downstream of gene-coding regions in accordance to previous analyses performed by the Psychiatric Genomics Consortium.

2. Additional Mendelian Randomisation and Colocalisation methods

Summary Based Mendelian Randomisation using Braineac dataset

To replicate our findings using an independent cis-acting eQTL dataset, we repeated the SMR analysis using the eQTL summary statistics from the Braineac study for cerebellar cortex, frontal cortex and temporal cortex tissues as exposure, and the ALS GWAS disease status as outcome. We followed the same procedure as described in the Methods using two p-value thresholds to select the eQTLs, p-value < 5×10^{-8} and p-value < 5×10^{-5}. Because of the very small number of SNPs selected with the SMR default p-value threshold (5×10^{-8}), the second less stringent threshold was used in our study to allow for a larger number of genes to be tested. This reflected the real number of SNPs per-transcript used in the Braineac analysis to identify cis-acting eQTLs (~1,000 independent tests per transcript). Only association signals with lead

SNP within 1 Mb of the transcription start site of the associated transcript were considered to be cis-acting in Braineac. We acknowledge that such a threshold might not fully account for the multiple testing burden of the Braineac cis-acting analysis.

3. Post-mortem DNA-sequencing and Genotyping

Purification, isolation and quality control

For each sample, a 25mg tissue block for DNA was homogenised using a Qiagen PowerLyzer 24 Homogenizer. DNA was purified from the homogenate using the standard protocol from Qiagen's DNeasy Blood and Tissue Mini Kit. DNA was quantified using PicoGreen (Quant-iT™ PicoGreen® dsDNA Reagent, ThermoFisher SCIENTIFIC) and measured using a Spectromax Gemini XPS (Molecular Devices).

Library preparation and DNA sequencing

Library preparation was performed using the Illumina DNA Sample Preparation HT Kit alongside the Illumina SeqLab DNA PCR-Free Library Prep Guide. Libraries were then quantified using qPCR and evaluated using gel electrophoresis. DNA libraries were clustered onto flow cells using the Illumina cBot System, as per cBot System Guide using Illumina HiSeq X HD Paired End Cluster Kit reagents. Sequencing was performed using an Illumina HiSeqX with 151bp paired-end runs using independent flow cell lanes and with a minimum of 30x average coverage per sample.

Genotyping, Quality Control and imputation

In parallel to whole genome sequencing, DNA was genotyped using the Infinium HumanOmni 2.5-8 v1.4 array. Resulting idat files underwent genotype calling, SNP clustering and quality control via Illumina GenomeStudio 2.0, generating a variant call format (vcf) file with variant quality filter data. Only biallelic variants with a pass filter status were selected.

The 176 post-mortem samples successfully genotyped and used in this study were merged with a larger UK dataset consisting of ALS cases and controls, which was lifted over from InfiniumOmni2-5-8v1-3_A1-b37 to InfiniumOmni2-5-8v1-4_A1-b37. The merged dataset was flipped and aligned to the positive

genomic strand and ambiguous SNPs (T/A, A/T, G/C, C/G) were removed. Samples were removed showing sex-mismatch, a genotype missingness rate greater than 2%, showed cryptic relatedness between samples (pi-hat $\geq$ 0.125), extensive autosomal homozygosity, or were outliers using PCA. SNPs were removed if it had a missingness rate of greater than 1%, minor allele frequency less than 0.01, or a Hardy-Weinberg p-value less than 1 x 10^{-7}. The remaining merged genotype dataset consisted of 1207 samples with 1,320,350 SNPs. Plink was used for multiple quality control steps.

Phasing and imputation were performed using SHAPEIT2 and IMPUTE2. The reference genome used was from 1000 Genomes (Phase3, V5b, date: 20130502). Once complete, imputed SNPs were with an info value (defined in Marchini et al. 2010) less than 0.4 and a certainty (average certainty of best-guess genotypes) value less than 0.9. The final imputation genotype file consisted for 9,172,093 SNPs with a minor allele frequency greater than 0.01. The 108 post-mortem samples were then extracted from the imputation dataset, with the remaining samples not used in the remainder of the analysis. Five samples were then removed as they were outliers in the genomic principal components' analyses and two samples removed showing significant shared identity by descent (pi-hat $\geq$ 0.125).

4. Post-mortem RNA Sequencing

Purification, isolation and quality control

The 100mg tissue blocks were divided; one for RNA purification and the other for DNA. For each sample, a 30mg tissue block for RNA was homogenised using a Qiagen PowerLyzer 24 Homogenizer. Total RNA was purified from the homogenate using the standard protocol from Qiagen's RNeasy Lipid Tissue Mini Kit. 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. Once library preparation was complete, 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 X (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).

Post-sequencing Quality Control, Alignment and Normalisation

Data was converted into Fastq format containing raw RNA sequence reads. MultiQC was used to assess read quality pre- and post- alignment across aligners and reference assemblies. SortMeRNA was used to filter and remove ribosomal RNA transcripts using rRNA reference databases. Clip adapter sequences and low-quality reads were removed using BBDuk (http://jgi.doe.gov/data-and-tools/bb-tools/). We used two RNA-sequence read aligners; STAR v2.7 and pseudo-aligner Kallisto v0.45.0. This was performed using both GRCh37.75 and GRCh38.89 reference assemblies. Using a mix of protocols was used to compare files, which contained transcript abundance per sample.

Tximport was used to import transcript abundances into R Statistics. Both gene-level and transcript-level abundances were estimated. Read counts with a total less than 10 were removed. The following sample-level phenotype and batch data was imported into R: disease status, gender, age (categorised into quintiles), post-mortem delay (categorised into quintiles), RIN (RNA Integrity Number), and flow-cell. Surrogate variable analyses were performed using SVA and SVAseq, while controlling for known covariates, where surrogate variables were appended to each sample. This is to estimate heterogeneity due to extraneous variables (such as cell heterogeneity).

Several Principal Components Analyses (PCA) were performed using DEseq2 and pcaExplorer to identify outlying samples and to assess potential confounding correlations between read counts, normalised expression profiles and covariates: sample-centric PCA, gene-centric PCA, hierarchal clustering of read counts and covariates, and hierarchal clustering of samples.

Data was normalised using variance stabilising transformations (VST) for use with analyses other than DESeq2 differential expression analyses.

5. Gene over-representation analysis

To test if genes significantly overlap between two gene-sets we implemented an over-representation adapted from: http://nemates.org/MA/progs/overlap_stats.html. In brief, the analysis tests if the overlap between two gene-sets statistically diverge from expectation, by calculating a representation factor (RF). The RF takes an expected number of genes to overlap between two random gene-sets ((n genes in set 1 * n genes in set 2) / n genes in the background genome). It then takes the overlap between sets, and divides it by the expected number of genes. This is the representation factor. Then a hypergeometric probability is used estimate the likelihood over overlap and over\under- representation.

Supplementary Results

1. Gene-level ALS GWAS analysis.

The gene *SCFD1* was first associated with ALS risk in a 2016 GWAS (van Rheenen *et al.*, 2016).

To confirm that *SCFD1* significantly associated with ALS risk in the 2018 ALS GWAS (Nicolas

et al., 2018), we performed a genome-wide gene-level association analysis using MAGMA.

SCFD1 associated with increased ALS risk after genome-wide correction for multiple testing,

where Z-Statistic = 4.874 and p-value = 5.46×10^{-7} (see Supplementary Table 2).

2. Mendelian Randomisation using Braineac eQTLs

Using Summary-Based Mendelian Randomisation and Braineac, we set the default p-value threshold (p =

5×10^{-8}) to select instrument eQTLs, where 461 eQTLs were tested. SCFD1 was not present among these.

451 eQTLs passed the heterogeneity test (pHEIDI > 0.05 and NSNP > 2). No genes were significant after

correcting for multiple-testing (p-value smr < 0.05 / 451 = 1.1×10^{-4}). Using the less stringent p-value

threshold (p = 5×10^{-5}), 4664 eQTLs were tested and 2006 passed the heterogeneity test (pHEIDI > 0.05

and NSNP > 2). One SCFD1 eQTL was selected (rs12433483, p-value$_{FrontalCortex}$ = 4.6×10^{-5}) but it did not

pass the heterogeneity test as it was the only selected SCFD1 SNP. No genes were significant after

correcting for multiple-testing (p-value smr < 0.05 / 2006 = 2.5×10^{-5})

3. Summary of post-mortem ALS and control samples with genomic and transcriptomic data

Post-mortem samples were age- and sex-matched, with no statistical difference in age between ALS cases

and controls using a Student's t-test (t = - 0.72, p-value = 0.47). The ALS cohort contained five donors

positive for the C9orf72 hexanucleotide repeat mutation (HREM) (Supplementary Tables 1).

Using the imputed genotype data from the post-mortem donors we filtered eQTLs that were statistically

significant cis-acting eQTLs in the GTEx database. We selected eQTLs present only in Cerebellum,

Cerebellar Hemisphere, Frontal Cortex or Cortex, resulting in 88,143 testable SNPs\eQTLs. If GTEx-

filtered eQTLs showed significant association with ALS, we then incorporated imputed SNPs to identify

novel eQTLs not identified in the GTEx database. For the transcriptomic data we removed genes with low read counts (less than 10) and performed surrogate variable analyses. We assessed potential correlation between SVAs and covariates. We performed VST to normalise read counts for use in the eQTL analyses. Mean normalised expression values by gene had a minimum = 5.09, 1st quartile value = 6.84, Median = 7.89, Mean = 7.98, 3rd Quartile = 8.96, and Maximum = 17.84 (Supplementary Fig. 1).

4. Genes proximal to SCFD1 overlap in function and show association with post-mortem ALS

There are three genes in the region (SCFD1, G2E3 and NUBPL) that have previously shown association with ALS risk. In addition, the genes in this locus share common functions namely vesicle-mediated transport and protein localisation (functions attributed to SCFD1), and protein ubiquitination. Therefore, we tested within a 1MB flanking region of SCFD1 for other eQTLs that may associate with ALS without filtering by low p-value. This analysis included imputed SNPs in addition to the GTEx-curated dataset used in the previous genome-wide cis-acting eQTL analysis.

In this region we identify other candidate eQTLs when comparing post-mortem ALS donors with controls, for G2E3 (beta = 0.397, p-value = 9.65×10^{-5}), NUBPL (beta = 0.208, p-value = 5.42×10^{-4}), and HEATR5A (beta = - 0.376, p-value = 9.92×10^{-3}).

To better understand the relationship between the expression of SCFD1 and the surrounding genes we performed co-expression analyses in the post-mortem ALS cohort. SCFD1 expression significantly correlated with the expression of neighbouring gene G2E3 (Pearson's r = 0.44, p = 7.71×10^{-5}), as well as HEATR5A (Pearson's r = 0.35, p = 0.001) and HECTD1 (Pearson's r = 0.48, p = 1.37×10^{-5}), in which the *SCFD1 downstream eQTLs* are located. In addition, G2E3 and HEATR5A (Pearson's r = 0.48, p = 1.58×10^{-5}), and G2E3 and HECTD2 (Pearson's r = 0.53, p = 6.76×10^{-7}) show very significant positive correlation. Very little correlation was found with other genes in the region (Supplementary Fig. 10, online resource).

We examined if rs8005942 and SNPs in LD with rs8005942 were significant eQTLs for SCFD1 in GTEx; they were not. rs8005942 is a significant eQTL in brain for genes more proximal to the locus, including HEATR5A (NES = 0.30, p-value = 2.8×10^{-8}), DTD2 (NES = -0.23, p-value = 8.8×10^{-17}), and NUBPL

(NES = -0.096, p-value = 1.7 x 10^{-5}). NUBPL has also shown association with ALS via TDP-43 conditional GWAS in 2018.

Given the genome-wide cis-eQTL analysis was confined to a distance of 850kb, from eQTL-to-gene coding regions, we tested whether any other eQTL or SNP on the genome (cis- or trans- acting) associated with ALS via SCFD1 expression. No other eQTL was more significantly associated with post-mortem ALS than the identified rs8005942.

5. SCFD1 and SCFD1 eQTLs co-express with major ALS pathways

The previous enrichment analyses treated *SCFD1* as a passive member to the ontological categories. As trans-acting eQTLs are known to regulate genes at a distance through regulation of nearby genes, we hypothesised that the direction of *SCFD1* expression could inform us on how it may have effect on larger pathways. To test this, we implemented co-expression statistics into the enrichment analyses using GSEA.

Firstly, we performed correlation analyses (Pearson's r) between *SCFD1* expression and the 388 genes that associate with *post-mortem* ALS via the *SCFD1* trans-eQTL hotspot. The most significant gene to positively correlate with *SCFD1* was *HNRNPK* (r = 0.605, p-value = 8.72 x 10^{-9}), a heterogeneous nuclear ribonucleoprotein, previously associated with ALS through its interaction with TDP-43 and FUS protein. The most significant gene to negatively correlate with *SCFD1* expression was *ADCY1* (r = -0.726, p-value = 1.77 x 10^{-13}), a calmodulin-sensitive adenylyl cyclase, also previously identified as an ALS candidate gene; see Fig. 4.

To put these co-expression results into context, we performed the same analysis between *SCFD1* expression and all remaining genes on the genome (n = 16,733 genes), applying a multiple-testing Bonferroni correction to each Pearson's p-value. 740 genes showed significant co-expression with *SCFD1*, with an adjusted p-value ≤ 0.05. The gene that most significantly co-expressed with *SCFD1* genome-wide was *ADCY1* (see results above). ALS genes *UNC13A* also negatively co-

expressed with *SCFD1* (r = -0.61, p-value = 5.71 x 10^{-9}). 97 genes overlapped between the 740 *SCFD1* co-expressed genes and the 388 genes that associated with *post-mortem* ALS via the trans-eQTL hotspot. This overlap was statistically over-representative, where RF = 5.5 (p-value = 1.74 x 10^{-45}). Of the 97 overlapping genes, 84 of them negatively correlated with *SCFD1* expression. In summary, most genes that associated with ALS via the *SCFD1* trans-eQTL hotspot were negatively co-expressed with *SCFD1*.

To explore why these genes negatively correlated with *SCFD1* in our *post-mortem* ALS cohort, we ran GSEA with the 388 genes ranked by their Pearson's r expression coefficient in relation to *SCFD1*; see Fig. 4. This analysis found that many of GO categories of which *SCFD1* is related, negatively correlated with *SCFD1* expression (see Supplementary Table 6). In the opposing direction, we found significantly enriched GO categories that positively correlated with *SCFD1* expression including RNA binding and metabolism, and the ribonucleoprotein complex (see Supplementary Table 6). In summary, these results indicate that *SCFD1* significantly co-expresses with at least two major ALS-disease related pathways: negatively with vesicle-mediated transport and transmembrane trafficking, and positively with RNA-binding and metabolism.

6. Overlapping pathways between ALS and schizophrenia identified via the *SCFD1* trans-eQTL hotspot

To find pathways that may be aberrantly regulated across the two diseases, we performed enrichment analysis on the 148 genes that showed association with schizophrenia risk (p-value $\leq$ 0.05) and association with *post-mortem* ALS via the trans-eQTL hotspot (see Table 2d and Supplementary Table 7). The results indicate that the genomic regulation of transmembrane potassium transport, synaptic signalling, and NMDA receptor activity, may be key to understanding why these diseases overlap.