Electronic Supplementary Material

Gene expression, methylation and neuropathology correlations at progressive supranuclear palsy risk loci.

Mariet Allen PhD[1], Jeremy D. Burgess BS[1], Travis Ballard BS[1], Daniel Serie BS[2], Xue Wang PhD[2], Curtis S. Younkin BS[2], Zhifu Sun MD[3], Naomi Kouri PhD[1], Saurabh Baheti MS[3], Chen Wang PhD[3], Minerva M. Carrasquillo PhD[1], Thuy Nguyen BS[1], Sarah Lincoln BS[1], Kimberly Malphrus BS[1], Melissa Murray PhD[1] Todd Golde[4], Nathan D. Price[5], Steven G. Younkin MD, PhD[1], Gerard D. Schellenberg PhD[6], Yan Asmann PhD[2], Tamas Ordog MD[7], Julia Crook PhD[2], Dennis Dickson MD[1], Nilüfer Ertekin-Taner MD, PhD[1,8*]

1. Mayo Clinic, Department of Neuroscience, Florida, USA, 32224
2. Mayo Clinic, Department of Health Sciences Research, Florida, USA, 32224
3. Mayo Clinic, Department of Health Sciences Research, Minnesota, USA, 55905
4. Center for Translational Research in Neurodegenerative Disease, McKnight Brain Institute, University of Florida, Department of Neuroscience, Gainesville, FL 32610, USA.
5. Institute for Systems Biology, 401 Terry Avenue N, Seattle, WA 98109, USA.
6. University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, 19104
7. Mayo Clinic, Department of Physiology and Biomedical Engineering and Center for Individualized Medicine, Rochester, Minnesota, USA, 55905
8. Mayo Clinic, Department of Neurology, Florida, USA, 32224

*Corresponding author: E-mail: taner.nilufer@mayo.edu, Phone: 904-953-7103, FAX:904-953-7370

Contents:

Online Resource: Methods

RNA sequencing (RNAseq)

Temporal cortex gene expression measures collected using RNA sequencing were available for a subset of the subjects in the previously described expression cohorts A and B (Cohort A, N= 78; Cohort B, N= 85). This available data was used for validation of the expression measures generated with the WG-DASL gene expression array. Total RNA for RNAseq was extracted from frozen brain tissue using either Trizol® reagent and cleaned using Qiagen RNeasy columns with DNase treatment (Cohort A) or the Ambion RNAqueous kit (Cohort B). RNA integrity number (RIN) was measured using an Agilent Technologies 2100 Bioanalyzer. Samples were randomized prior to transfer to the Mayo Clinic Medical Genome Facility Gene Expression and Sequencing Cores for library preparation and sequencing. The TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA) was used for library preparation. The library concentration and size distribution was determined on an Agilent Bioanalyzer DNA 1000 chip. Sequencing was performed on the Illumina HiSeq2000 using either 101 base-pair (bp) (Cohort A) or 50 bp (Cohort B) paired end sequencing, with triplicate multiplexing of barcoded samples (3 samples per flowcell lane). Base-calling was performed using Illumina's RTA 1.18.61 or RTA 1.17.21.3 (Cohort A) or v.1.12.4.2 (Cohort B). FASTQ sequence reads were aligned to the human reference genome using TopHat [5] and Bowtie [3] and either Subread 1.4.4 (Cohort A) or HTSEq (Cohort B) was used for gene counting [1]. FastQC was used for quality control (QC) of raw sequence reads, and RSeQC was used for QC of mapped reads. Raw read counts were normalized using Conditional Quantile Normalization (CQN) via the Bioconductor package; accounting for sequencing depth, gene length, and GC content. Normalized read counts were used as the gene expression phenotype for generating residuals.

Comparison with other brain gene expression data

We compared the significant *cis*-eSNPs from Tables 3 in our study to three other publically available brain eQTL datasets.

1. We have previously collected gene expression measures from temporal cortex tissue of 202 AD subjects, using the Illumina WG-DASL microarray [7]. Expression measures were available for *MOBP* (ILMN_2298464) and *LRRC37A4* (ILMN_2393693) and were tested for association with rs1768208 and rs8070723 respectively. Genotype data were available for 198 of these subjects. eQTL statistical methods used were as described for PSP expression cohort B in Methods for Main Text.

2. Ramasamy et al. [4] generated eQTL data using exon-specific RNA expression levels quantified by Affymetrix Human Exon 1.0 ST arrays and DNA genotypes obtained from Illumina Omni1-quad and Immunochip arrays followed by imputation using the 1000 Genomes Project (March 2012 release). Braineac website (http://www.braineac.org/) had TCX eQTL results only for rs1768208/*MOBP* (t2618407 = transcript ID). The beta coefficient for the *cis*-eSNP associations are not available at the Braineac website, therefore the direction of association with the SNP minor allele was based on the boxplots from this site.

3. Zhang et al. [6] utilized custom gene expression arrays manufactured by Agilent Technologies and two genotyping platforms, the IlluminaHumanHap650Y array and a custom Perlegen 300K array, for their eQTL analyses. The *cis*-eQTL results in their published supplementary tables for the dorsolateral prefrontal cortex (PFC), cerebellar (CER) and Visual Cortex (VC) were screened for the significant associations in our study from Table 3. Results were available for only

rs8070723 and only for VC. Zhang et al assessed subjects with a pathological diagnosis of AD, Huntington's disease and healthy controls. The group of subjects for which results were available is indicated in the Online Resource Table 4. The direction of the association (beta) is not available in this dataset and so only the p-value is reported.

Supplemental Tables

Online Resource Table 1. Genes and probes in *cis* with PSP index SNPs assessed in this study.

Index SNP	Chr	*Gene*	Probe(s)	# Tests
		PLEKHM1	ILMN_1709549	1
rs8070723	17q21	*LRRC37A4*	ILMN_2393693	1
		CRHR1	ILMN_1732426; ILMN_1753706	2
		MAPT	ILMN_1800049; ILMN_2310814	2
		STH	ILMN_1665311	1
		KANSL1	ILMN_2200636	1
		ARL17A	ILMN_1698680	1
		ARL17B	ILMN_3231502; ILMN_3231952	2
rs242557	17q21	*PLEKHM1*	ILMN_1709549	1
		LRRC37A4	ILMN_2393693	1
		CRHR1	ILMN_1732426; ILMN_1753706	2
		MAPT	ILMN_1800049; ILMN_2310814	2
		STH	ILMN_1665311	1
		KANSL1	ILMN_2200636	1
		ARL17A	ILMN_1698680	1
		ARL17B	ILMN_3231502; ILMN_3231952	2
rs1411478	1q25	*IER5*	ILMN_1721833	1
		MR1	ILMN_2167416	1
		STX6	ILMN_1777915; ILMN_2157951	2
rs7571971	2p11	*EIF2AK3*	ILMN_1724984	1
		RPIA	ILMN_1714809	1
rs1768208	3p22	*MOBP*	ILMN_1750271; ILMN_1768947; ILMN_2298464; ILMN_2414962	4
		RPSA	ILMN_1664910; ILMN_2411723	2
		SLC25A38	ILMN_1781231	1
		SNORA6	ILMN_3245365	1
rs2142991	10q11	*BMS1*	ILMN_1772713	1
rs11568563	12p12	*IAPP*	ILMN_1679527	1
		SLCO1A2	ILMN_1656097; ILMN_1720727; ILMN_1806979	3
rs6687758	1q41	*No Genes with probes in cis*		0
TOTAL				41

Online Resource Table 2. Sequence Alignment for WG-DASL Probes with Significant eQTL Results. BLAT results (UCSC genome browser, hg19) are shown where WG-DASL 50 bp probes aligned with a 100% match to RefSeq transcripts annotated to autosomal chromosomes.

		Probe Targets					
Probe	Sequence	Chr	Strand	Start	End	Gene	RefSeq Transcript (Exon)
ILMN_2298464	GATCTTGGCCAGGTGCCTTCTGCTCAAATATCGTCTCAGAGGTGCTTCCC	3	+	39557207	39557256	*MOBP*	NM_182935.3 (Exon 4)
ILMN_2393693	AGCAGCCCCGTGTGTATGCTGGTGCAGGTTCTAAGCAAAGTGAGCTGCCC	17	-	43584483	43584532	*LRRC37A4*	NR_002940.2 (Exon 6)
ILMN_3231952	CAAGAAATTATGCGGGGTCACTGGCACAAATGATGAGGCATCTCCTGGAA	17	-	44376951	44377000	*ARL17A*	NM_016632.2 (Exon 4)
ILMN_3231952	CAAGAAATTATGCGGGGTCACTGGCACAAATGATGAGGCATCTCCTGGAA	17	-	44376951	44377000	*ARL17B*	NM_001103154.1 (Exon 4)
ILMN_1698680	TCTGATTTGGCCCCTTCACACCTCACTCCTAGATTTTGCTAGACCTTTCT	17	-	44594196	44594245	*ARL17A*	NM_016632.2 (Exon 5)

Online Resource Table 3. Spearman rank correlations for gene expression measures collected using WG-DASL microarray and RNA sequencing methods. Cohort A underwent gene expression measurements with WG-DASL HT arrays which had probes for *MOBP, ARL17A, ARL17B* and *LRRC37A4*. Of these genes, only *MOBP* had both RNAseq measures and a probe on the older version of the reference genome and WG-DASL microarrays with which Cohort B was measured respectively. *MOBP* results for Cohorts A and B are shown individually as well as combined. RNAseq gene measurements correspond to the sum of all exons. RNAseq exon measurements correspond to the specific exon to which the WG-DASL probe binds. Online figures depicting each of the correlations are provided in the electronic supplemental materials as noted in the last column.

								Spearman rank		
Cohort	Gene	DASL probe	Start	End	RNAseq	Start	End	Corr.Coef	P-value	Figure
A	*MOBP*	ILMN_2298464	39557207	39557256	Gene	39508689	39570970	0.194	8.86E-02	ESM3.a
B	*MOBP*	ILMN_2298464	39557207	39557256	Gene	39509070	39567857	0.642	<2.20E-16	ESM3.b
A+B	*MOBP*	ILMN_2298464	39557207	39557256	Gene	NA	NA	0.433	1.15E-08	ESM3.c
A	*ARL17A*	ILMN_1698680	44594196	44594245	Gene	44594068	44657088	0.400	3.22E-04	ESM3.d
A	*ARL17B*	ILMN_3231952	44376951	44377000	Gene	44352150	44439130	0.505	3.32E-06	ESM3.e
A	*LRRC37A4*	ILMN_2393693	43584483	43584532	Gene	43578685	43627701	0.356	1.47E-03	ESM3.f
A	*MOBP*	ILMN_2298464	39557207	39557256	Exon	39554874	39557494	0.196	8.58E-02	ESM4.a
B	*MOBP*	ILMN_2298464	39557207	39557256	Exon	39554874	39557494	0.803	<2.20E-16	ESM4.b
A+B	*MOBP*	ILMN_2298464	39557207	39557256	Exon	NA	NA	0.497	8.64E-12	ESM4.c
A	*ARL17A*	ILMN_1698680	44594196	44594245	Exon	44594068	44594379	0.665	<2.20E-06	ESM4.d
A	*ARL17B*	ILMN_3231952	44376951	44377000	Exon	44376892	44377031	0.516	1.92E-06	ESM4.e
A	*LRRC37A4*	ILMN_2393693	43584483	43584532	Exon	43584107	43585905	0.140	2.21E-01	ESM4.f

Online Resource Table 4. Comparison of PSP temporal cortex eQTL data with results in other datasets. Comparative data was available for three of the results shown. a. eQTL results in PSP brains from Cohort A in this study. b. Our published Zou et al. study results from AD temporal cortex. c. Braineac eQTL beta values were not available from the website, therefore the direction of association with the SNP minor allele is shown: + indicates association of minor allele with higher gene levels based on boxplots from the website. Results are shown for brains of control subjects. d. Results from Zhang et al. study are for the combined subjects (All=376 LOAD, 194 Huntington's disease, 173 nondemented controls) or nondemented controls only as indicated. TCX=temporal cortex, VC=visual cortex. NA=Not available.

					Cohort A[a] (TCX)			Zou et al[b] (TCX)			Braineac Ramasamy et al[c] (TCX)			Zhang et al[d] (VC)	
SNP	Tested Allele	Probe	Chr	Gene Symbol	N	Beta	p-value	N	Beta	p-value	Transcript ID	Beta	p-value	Subect group	p-value
rs1768208	T	ILMN_2298464	3	*MOBP*	175	0.30	1.68E-02	198	0.31	1.38E-03	t2618407	(+)	0.02	NA	
rs8070723	G	ILMN_2393693	17	*LRRC37A4*	175	-0.45	1.60E-04	198	-0.53	1.48E-22		NA		Controls	1.25E-10
rs8070723	G	ILMN_3231952	17	*ARL17B*	175	-0.35	6.43E-03	NA				NA		All	8.55E-10

Online Resource Table 5. Chromosome 17 inversion region CpGs tested for association with rs8070723 and rs242557: The results for this table are deposited as a separate Excel file (ESM5_OnlineTable5.xlsx) due to their size. Chromosomal breakpoints were defined as described in Itsara et al[2] ranging from 43.4Mbp to 44.8Mbp, according to build hg19. A total of 1,373 Unique CpGs were captured across the locus on both strands. Excluding CpGs that were methylated in 100% or 0% of all subjects, 1,108 could be tested. The table has >1,373 rows, because some CpGs fall within the gene annotation of more than 1 gene/transcript and therefore are listed 2-3 times. Results are annotated with any known SNPs (dbSNP 135) at CpG positions, the gene in which the CpG resides, the specific transcript and distance to transcript start site (Tss) if relevant. Additionally whether the CpG is in a known exon or a CpG island is indicated (1= yes, 0 = no).

Supplemental Figure Legends:

Online Resource Fig 1. Histograms of RIN values for samples in a) cohort A and b) cohort B.

Online Resource Fig 2. Correlation plots of WG-DASL vs. RNAseq expression levels for gene counts. Linear regression plots of gene expression residuals obtained after adjusting for appropriate covariates as described in methods were plotted.

a) *MOBP* cohort A; b) *MOBP* cohort B; c) *MOBP* cohorts A+B; d) *LRRC37A4P* (a.k.a *LRRC37A4*) cohort A; e) *ARL17A* cohort A; f) *ARL17B* cohort A.

Online Resource Fig 3. Correlation plots of WG-DASL vs RNAseq expression levels for counts of exons corresponding to WG-DASL binding sites. Other information is as in Online Resource Fig. 2.

References

1 Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166-169 10.1093/bioinformatics/btu638

2 Itsara A, Vissers LE, Steinberg KM, Meyer KJ, Zody MC, Koolen DA et al (2012) Resolving the breakpoints of the 17q21.31 microdeletion syndrome with next-generation sequencing. American journal of human genetics 90: 599-613 10.1016/j.ajhg.2012.02.013

3 Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25 10.1186/gb-2009-10-3-r25

4 Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R et al (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 17: 1418-1428 10.1038/nn.3801

5 Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111 10.1093/bioinformatics/btp120

6 Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA et al (2013) Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 153: 707-720 10.1016/j.cell.2013.03.030

7 Zou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS et al (2012) Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS Genet 8: e1002707 10.1371/journal.pgen.1002707