Transcript profiling analysis of Alzheimer's disease brains

Mariet Allen[1], Daniel Serie[4], Michael Walsh[4], Zhifu Sun[3], Saurabh Baheti[3], Fanggeng Zou[1], High Seng Chai[3], Curtis S. Younkin[1], Julia Crook[4], V. Shane Pankratz[3], Minerva M. Carrasquillo[1], Asha A. Nair[3], Sumit Middha[3], Sooraj Maharjan[3], Thuy Nguyen[1], Li Ma[1], Kimberly G. Malphrus1, Sarah Lincoln[1], Gina Bisceglio[1], Christopher P. Kolbert[6], Jin Jen[6], Ronald C. Petersen[7], Neill R. Graff-Radford[2], Dennis W. Dickson[1], Steven G. Younkin[1], Yan W. Asmann[4], Nilufer Ertekin-Taner[1,2]

1. Mayo Clinic in Florida, Department of Neuroscience, 2. Mayo Clinic in Florida, Department of Neurology, 3. Mayo Clinic Minnesota, Department of Health Sciences Research, 4. Mayo Clinic in Florida, Department of Health Sciences Research, 5. Mayo Clinic in Florida, Department of Psychiatry and Psychology. 6. Mayo Clinic in Minnesota, Microarray Core. 7. Mayo Clinic in Minnesota, Department of Neurology.

Abstract

Objective: To use microarray and methylome data to identify genes that are differentially expressed in AD vs. non-AD brains.

Background: Despite the success of recent LOAD GWAS studies, much of the heritability of LOAD remains unexplained. We hypothesize that evaluation of brain gene expression provides an additional avenue for identification of novel LOAD genes and pathways that may be potential drug targets. We previously performed a gene expression GWAS that assessed mRNA levels of ~24,000 transcripts in two brain regions (temporal cortex and cerebellum) for ~ 200 AD subjects and ~200 subjects with other pathologies We have now performed transcript profiling analysis comparing gene expression levels between AD and non-AD subjects in both brain regions and conducted pathway analysis on these data. In addition, we collected whole genome CpG methylation data (methylome) on a subset of these subjects for further evaluation.

Design/Methods: Gene expression levels were measured by the Illumina HT-12 V4 Expression BeadChip arrays using the Whole-Genome DASL HT assay and appropriate data quality control was implemented. Transcript profiling analysis was carried out in R using linear regression with appropriate covariates. Gene pathway analysis was executed using MetaCore for the two brain regions separately.

Results: Following QC ~17,000 gene expression measures were robustly detected in each brain region. Transcript profiling analysis identified 743 targets in the cerebellum and 2839 targets in the temporal cortex (un-corrected p-value <0.01) that were selected for pathway analysis. In the temporal cortex several significant pathways and GO Processes were identified including but not limited to oxidative phosphorylation and lipid metabolism. Some of the differentially expressed genes have correlative methylation levels.

Conclusions: Transcript profiling and pathway analysis of brain gene expression data has identified several interesting targets for further exploration of molecular pathways involved in AD.

Overview

Transcript Profiling: Analysis of whole transcriptome measures collected as part of the Mayo Clinic brain expression GWAS (Zou et al, *PLOS* Genetics, 2012). Compared brain gene expression measures, collected from two brain regions, for ~200 AD subjects with measures for ~200 subjects with other pathologies (~100 with Progressive Supranuclear Palsy (PSP) diagnosis).

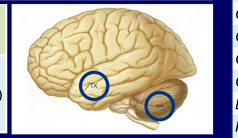
Pathway analysis: Using the functionalities in MetaCore, performed enrichment analysis to identify pathways and processes that are enriched for differentially expressed genes.

RRBS methylome: Analysis of DNA methylation measures collected for a subset of the samples with expression measures: specifically analyzed promoter region (1kb 5' of TSS) methylation for association with gene expression levels.

Samples

	Temporal Cortex (TX)			Cerebellum (CER)		
RNA: WG-DASL	**AD**	**Non-AD**	**PSP**	**AD**	**Non-AD**	**PSP**
N	202	197	107	197	177	98
Females (%)	108 (53%)	78 (40%)	45 (42%)	101 (51%)	63 (36%)	37 (38%)
Age: Mean (SD)	73.6 (5.5)	71.6 (5.6)	72.0 (5.3)	73.6 (5.6)	71.7 (5.5)	71.8 (5.1)
RIN: Mean (SD)	6.3 (0.9)	6.9 (1.0)	7.0 (1.0)	7.2 (1.0)	7.2 (1.0)	7.1 (1.0)

	Temporal Cortex (TX)		
DNA: RRBS	**AD**	**Non-AD**	**PSP**
N	55	na	55
Females (%)	24 (44%)	na	23 (42%)
Age: Mean (SD)	74 (5.4%)	na	73 (5.2)



RNA and DNA was isolated from frozen brain tissue of pathologically confirmed AD subjects and subjects with other pathologies (Non-AD); Mayo Clinic Brain Bank, Dr Dennis Dickson. Over half of the subjects with other pathologies had pathologically confirmed PSP.

Transcription Profiling

Diagnosis	**AD vs NonAD**		**AD vs PSP**	
Tissue	**TX**	**CER**	**TX**	**CER**
N	**202 vs 197**	**197 vs 177**	**202 vs 107**	**197 vs 98**
Probe Count	**17,473**	**15,150**	**17,523**	**15,147**
#Probes (p_{diff}<3.0E-06)	168	6	590	7
#Unique Genes (p_{diff}<3.0E-06)	163	5	569	5
Common Probes	14,900		14,892	
Common p_{diff} <3.0E06 (beta)	*HSD11B2* (++)		*HSD11B2* (++)	
p_{diff}<1.0E-04	478	44	1,255	15
Common p_{diff}<1.0E-04 (beta)	*VGF* (–)		*BDNF (–), MPPED2(–)*	

Gene expression levels were measured in two brain regions (TX and CER) using the Illumina Whole-Genome DASL microarray, with 24,526 probes, designed for partially degraded fresh frozen and FFPE tissues.
Genes, differentially expressed between AD and either Non-AD or PSP subjects were identified by multi-variable linear regression analysis, in R, using normalized gene expression measures, including age at death, gender, #*APOE*e4 alleles, plate, RIN, and cell specific gene levels as covariates.
Only those probes that were reliably detected > a background threshold in at least one group were considered for analysis.

References: Zou et al, Brain Expression Genome-Wide Association Study (eGWAS) Identifies Human Disease-Associated Variants , *PLoS Genetics*, 2012

Study Supported by: R01 AG032990, P50 AG016574, Mayo Clinic Center for Individualized Medicine Epigenomics Grant

LOAD genes

	AD-NonAD				AD-PSP			
	CER		TCX		CER		TCX	
Closest Gene	Beta	P-value	Beta	P-value	Beta	P-value	Beta	P-value
ABCA7	0.05	0.25	0.07	0.07	0.01	0.83	0.12	0.03
BIN1	0.04	0.09	0.05	0.01	0.03	0.35	**0.10**	**1.07E-04**
CASS4	na	na	-0.03	0.49	na	na	-0.04	0.57
CD2AP	0.01	0.61	0.01	0.70	-0.01	0.71	-0.02	0.65
CD33	na	na	-0.04	0.42	na	na	-0.11	0.10
CELF1	0.07	0.08	-0.02	0.43	0.06	0.20	-0.01	0.86
CLU	-0.12	0.03	-0.10	0.01	-0.10	0.12	-0.15	2.06E-03
CR1	na	na	0.16	0.03	na	na	0.24	0.01
EPHA1	0.05	0.31	na	na	0.08	0.20	na	na
FERMT2	-0.01	0.71	-0.05	1.30E-03	0.00	0.76	**-0.08**	**5.81E-05**
HLA-DRB1	0.27	0.12	-0.01	0.95	0.43	0.03	0.04	0.85
HLA-DRB5	na	na	na	na	na	na	na	na
INPP5D	0.02	0.56	0.01	0.86	0.01	0.83	0.00	0.99
MEF2C	0.02	0.52	-0.02	0.54	0.00	0.90	-0.02	0.65
MS4A4A	0.06	0.27	-0.11	0.08	-0.01	0.86	-0.19	0.01
MS4A6A	-0.04	0.49	-0.09	0.13	-0.06	0.39	-0.14	0.04
NME8	na	na	na	na	na	na	na	na
PICALM	0.03	0.46	-0.04	0.06	-0.02	0.42	-0.04	0.11
PTK2B	0.00	0.86	-0.07	0.10	0.01	0.71	-0.16	1.13E-03
RIN3	0.01	0.73	-0.04	0.29	na	na	na	na
SLC24A4	0.06	0.26	**0.16**	**9.31E-06**	0.07	0.29	**0.20**	**2.56E-07**
SORL1	0.05	0.04	-0.07	0.01	0.03	0.23	-0.10	1.27E-03
ZCWPW1	0.01	0.71	0.01	0.56	0.03	0.31	0.04	0.22

Recent LOAD GWAS have identified 20 novel LOAD risk loci. The most proximal genes to the LOAD GWAS index SNPs are candidate LOAD genes, listed here.
The results of transcription profiling analysis for these genes is shown, in the table above.
Of the 23 genes tested, 10 show nominal significant association (p_{diff}<0.05) with 3 genes achieving study wide significance in bold.

Pathway Analysis

Diagnosis	AD vs NonAD		AD vs PSP	
Tissue	TCX	CER	TCX	CER
N	202 vs 197	197 vs 177	202 vs 107	197 vs 98
Probe Count	17,473	15,150	17,523	15,147
# Probes P_{diff} <0.01	2,839	743	4,466	338

Pathway analysis enables identification of cellular pathways enriched in genes that are differentially expressed.
We implemented the tools in MetaCore [http://thomsonreuters.com/metacore] to conduct enrichment analysis.
The analysis was limited to probes >background threshold values and with a p_{diff} <0.01 for 4 groups of comparison.

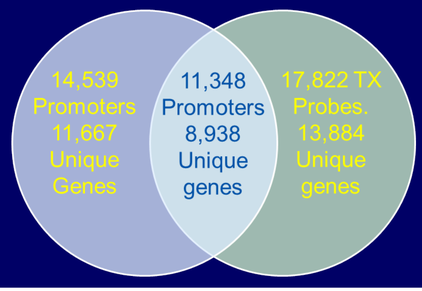
Top Pathways

AD vs NonAD (TX) **2,839 Probes P_{diff} <0.01**	**p-value**	**FDR**	**# Genes in active data**
Oxidative phosphorylation	1.124E-06	0.001	28
Ubiquinone metabolism	1.061E-04	0.036	17
Beta-alanine metabolism	2.785E-04	0.064	7
cAMP/ Ca(2+)-dependent Insulin secretion	4.187E-04	0.072	10
Beta-alanine metabolism/ Rodent version	1.117E-03	0.153	6
Histidine-glutamate-glutamine and proline metabolism/ Rodent version	3.186E-03	0.364	14

AD vs PSP (TX) **4,466 Probes P_{diff} <0.01**	**p-value**	**FDR**	**# Genes in active data**
Oxidative phosphorylation	1.81E-06	0.001	35
Regulation of lipid metabolism_Alpha-1 adrenergic receptors signaling via arachidonic acid	5.22E-05	0.018	21
Development_Activation of ERK by Alpha-1 adrenergic receptors	1.30E-03	0.302	16
Translation_Translation regulation by Alpha-1 adrenergic receptors	3.70E-03	0.525	18
ATP/ITP metabolism	3.75E-03	0.525	31
Neurophysiological process_NMDA-dependent postsynaptic long-term potentiation in CA1 hippocampal neurons	5.33E-03	0.566	20

The top MetaCore pathways identified for the temporal cortex analysis are show above. Oxidative phosphorylation is the most significant pathway for both the AD vs Non-AD and AD vs PSP analysis. 28 genes in this pathway were identified to have a p_{diff} <0.01 in the AD vs Non-AD profiling analysis.

DNA Methylation



CpG methylation measures collected using reduced representation bisulfite sequencing (RRBS) techniques implemented on the Illumina HiSeq2000.
Following QC, 46 AD samples, 46 PSP samples and calls for 2,083,781 CpGs were retained for additional analysis.
We tested for correlations between WG-DASL gene expression measures and promoter CpG methylation using functionalities in R, including important covariates in the model.
The majority of the promoters (66%) are hypomethylated (Average methylation <1%).

We observed 672 methylation-expression correlations (564 genes) with a p< 0.05; 10 of which remain significant after Bonferroni corrections. As expected 9/10 of the top promoters were negatively associated with gene expression levels

Conclusions

Transcript Profiling: Identified differentially expressed genes. Some of the top LOAD GWAS candidate genes are also differentially expressed. We postulate that regulatory genetic variants that influence LOAD risk account for this differential expression.

Pathway Analysis: Identified enrichment of differentially expressed genes between AD and non-AD subjects in pathways related to oxidative phosphorylation and mitochondrial function implicating regulatory genetic variants within energy metabolism genes in LOAD risk.

DNA Methylation: Detected correlations between promoter CpG methylation and gene expression levels for genes that display differential expression. We postulate that some genes may confer LOAD risk via expression changes due to methylation differences.