Genomics Data 2 (2014) 290–292

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Genome-wide profiling of alternative splicing in Alzheimer's disease

Mitchell K.P. Lai [a], Margaret M. Esiri [b], Michelle G.K. Tan [a,c,*]

[a] Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[b] Oxford Project to Investigate Memory and Ageing (OPTIMA), University of Oxford, John Radcliffe Hospital, Oxford, UK
[c] Department of Clinical Research, Singapore General Hospital, Singapore 169856, Singapore

ARTICLE INFO

Article history:
Received 27 August 2014
Accepted 1 September 2014
Available online 7 September 2014

Keywords:
Alzheimer's disease
Alternative splicing
Exon array
Transcript variant
Genome-wide profiling

ABSTRACT

Alternative splicing is a highly regulated process which generates transcriptome and proteome diversity through the skipping or inclusion of exons within gene loci. Identification of aberrant alternative splicing associated with human diseases has become feasible with the development of new genomic technologies and powerful bioinformatics. We have previously reported genome-wide gene alterations in the neocortex of a well-characterized cohort of Alzheimer's disease (AD) patients and matched elderly controls using a commercial exon microarray platform [1]. Here, we provide detailed description of analyses aimed at identifying differential alternative splicing events associated with AD.

Specifications	
Organism/cell line/tissue	Temporal cortex (BA22) from AD patients and matched elderly controls
Sex	5 male and 3 female in each group
Sequencer or array type	Affymetrix GeneChip® Exon 1.0 ST Array
Data format	Raw data: CEL files; normalized data: SOFT, MINiML and TXT
Experimental factors	Diagnosis (AD versus control); sex; age; post-mortem interval (PMI)
Experimental features	The samples processed in each batch were matched for AD versus control, sex and PMI
Consent	(1) Subjects who died with autopsy, consent was given by next-of-kin; (2) ethics review board approval was obtained from the respective institutions for the use of human postmortem tissues.
Sample source location	OPTIMA (Oxford Project to Investigate Memory and Ageing), a longitudinal, clinicopathologic study of aging and dementia (http://www.medsci.ox.ac.uk/optima)

Direct link to deposited data

Deposited data (at the exon-level): http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37264.

Deposited data (at the gene-level): http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37263.

Experimental design, materials and methods

Study population, clinical and neuropathological data

Frozen post-mortem tissues from the superior temporal gyrus (Brodmann area, BA22) of eight community-based AD patients recruited into OPTIMA (Oxford Project to Investigate Memory and Ageing), as well as eight age- and sex-matched controls, were thawed on ice and dissected free of meninges, blood vessel and white matter for further processing. Informed consent had been obtained from the subjects' next-of-kin before collection of brain tissues. Ethics review board approval was obtained in the respective institutions in Singapore and the UK for the use of human post-mortem tissues. Under OPTIMA, all subjects underwent clinical and neuropsychological assessments, including the Cambridge Cognitive Examination [2]. Subjects were clinically diagnosed with AD using the National Institute of Neurological and Communicable Diseases and Stroke (NINCDS) criteria [3], followed with pathological confirmation using the Consortium to Establish a Registry for AD (CERAD) criteria [4]. The age-matched controls did not have clinical evidence of dementia (CAMCOG scores >80 at the last assessment before death) and died from non-neurological causes. Braak staging of AD-related neuropathological changes [5] was also performed. Six out of eight AD subjects were staged at V/VI (with two at III/IV), while only one control was stage III/IV with the rest at 0–II.

* Corresponding author at: Department of Clinical Research, Singapore General Hospital, The Academia, Level 9, Discovery Tower, 20, College Road, Singapore 169856. Tel.: +65 63265388; fax: +65 62215142.
E-mail address: michelle.tan.g.k@sgh.com.sg (M.G.K. Tan).

http://dx.doi.org/10.1016/j.gdata.2014.09.002

Sample preparation and microarray procedures

Total RNA extraction of 25–50 mg aliquots of dissected tissue was performed using TRIzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Quality of extracted RNA was evaluated by agarose gel electrophoresis. 1 μg total RNA from each sample was then processed for ribosomal RNA (rRNA) depletion with a RiboMinus Human/Mouse Transcriptome Isolation Kit (Invitrogen). Removal of rRNA was confirmed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), and cDNA was generated with a GeneChip Whole Transcript (WT) cDNA Synthesis and Amplification Kit (Affymetrix, Santa Clara, CA). Subsequently, cDNA was fragmented and end-labeled using a Affymetrix WT Sense Target Labeling Kit before hybridization to Affymetrix GeneChip Human Exon 1.0 ST Arrays at 45 °C for 16 h, then washed and stained in a Fluidics Station 450 and scanned by a GCS3000 Scanner (both from Affymetrix) to generate cell intensity (CEL) files using the Affymetrix GCOS software. The sample process was limited by 4 available wash modules of the fluidic station. We thus matched 4 samples for AD versus control, sex and post-mortem interval (PMI) as far as possible in each run to minimize batch effects (Table 1).

Data analysis and normalization

Affymetrix CEL files were imported into Partek Genomics Suite 6.6 software (Partek Inc, Saint Louis, MO). Core meta-probeset file (HuEx-1_0-st-v2.r2.dt1.hg18.core.mps) was employed to include the most reliable probesets (exons) derived from NCBI's Reference Sequence Database (RefSeq). Probe summarization and probeset normalization were performed using Robust Multi-chip Average (RMA), which included RMA background correction, quantile normalization, $\log_2$ base transformation, and median Polish probe set summarization. Exploratory data analysis was done by plotting histogram for each sample with the intensity of the probeset on X-axis and the frequency of the probeset intensity on the Y-axis. Histogram plot revealed that all of the samples have a similar distribution pattern, indicating no obvious outliers of the data (Fig. 1). The dataset consisting of about 232,000 probesets has been deposited in the Gene Expression Omnibus database (GEO) under accession number GSE37264. Furthermore, we generated another dataset which consists of about 17,000 transcript clusters by averaging the expression of all probesets on the same gene locus for the corresponding transcript clusters. The dataset is available in GEO under accession number GSE37263. Genome wide profiling of altered gene expression in AD has been reported previously [1].

Alternative splicing data analysis

A Partek alternative splice analysis of variance (ANOVA) model was used to identify transcript clusters (genes) with one or more probesets (exons) which were significantly different from others in expression between AD and controls. To reduce the generation of false-positive results which can be caused by non-expressed exons, probesets with a $\log_2 < 3.0$ in all samples were excluded from further analysis except for those found significantly and differently expressed between the groups ($p < 0.05$). Using false discovery rate (FDR) adjusted values of $p < 0.01$, 487 genes were identified as candidates for differential alternative splicing events. An ideal differential alternative splicing event is presented as equal expression of most of the exons in a gene locus while one or a few exons are differentially expressed between groups. Thus, among these candidates, those which exhibited significant changes at the gene level (AD vs control, $p < 0.05$), or fold change >1.25 were excluded, resulting in 177 candidates for subsequent evaluation.

Interpreting the results of alternative splicing analysis can be the most challenging aspect of exon array analyses. For mutually exclusive exons, differential alternative splicing can readily be identified. However, in most cases, it is difficult to detect or interpret subtle changes in the relative proportions of different transcript variants between groups. Thus, the 177 candidates were manually inspected in Partek's Gene View which is incorporated with the University of California, Santa Cruz (UCSC) genome browser to localize differentially expressed probesets in the context of transcript variants. Among these candidates, we further classified the differential alternative splicing events as: (1) *Yes*, the differentially expressed probeset maps to a known alternative spliced exon, and the transcript variant can be easily be identified; (2) *Probably*, a single or more than one adjacent probesets behave similarly in mapping to an exon but do not correspond to any known transcript variants in RefSeq; and (3) *Unlikely*, differentially expressed probesets exhibited minor changes or high signal variance, or probeset mapped to regions of overlapping transcript clusters. Among the 177 candidates, 77 were removed due to the low likelihood of alternative splicing events. We identified 22 candidates (BEGAIN, BRWD1, CAMK2B, CYB5R3, EDF1, ELMO1, EPS15, FEZ1, FYN, GNAL, HK1, HPCAL1, KCNS3, KLC1, MAP4, NTRK2, PACSIN1, SCN4B, SIPA1L1, TKT, TRIM9 and ZNF365) with strong evidence of differential alternative splicing events, of which two representatives are depicted here in gene view plot (Fig. 2). The remaining candidates with possible

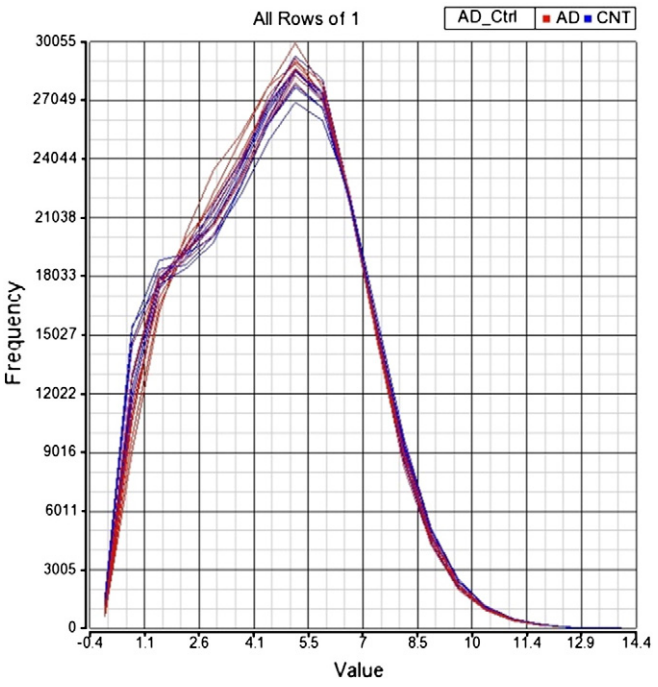

Fig. 1. Histogram plot revealed a similar distribution pattern of probeset intensities in all samples. Probeset intensity is on X-axis and the frequency of the probeset intensity is on Y-axis. Each line represents a sample.

Table 1
Matched samples for array processing.

Batch	Control sample	Sex	Age	PMI	AD samples	Sex	Age	PMI
1	op1044_CNT	M	85	16	op388_AD	M	83	29
	op1151_CNT	M	77	20	op417_AD	M	67	30
2	op1146_CNT	M	87	30	op91_AD	M	66	42
	op9_CNT	F	79	29	op80_AD	F	88	30
3	op223_CNT	M	66	47	op133_AD	M	69	44
	op1132_CNT	F	81	47	op340_AD	F	63	48
4	op1112_CNT	M	78	63	op450_AD	M	78	88
	op346_CNT	F	88	115	op107_AD	F	67	107

Batch, hybridization run; age, age of death (in year); PMI (in hours).

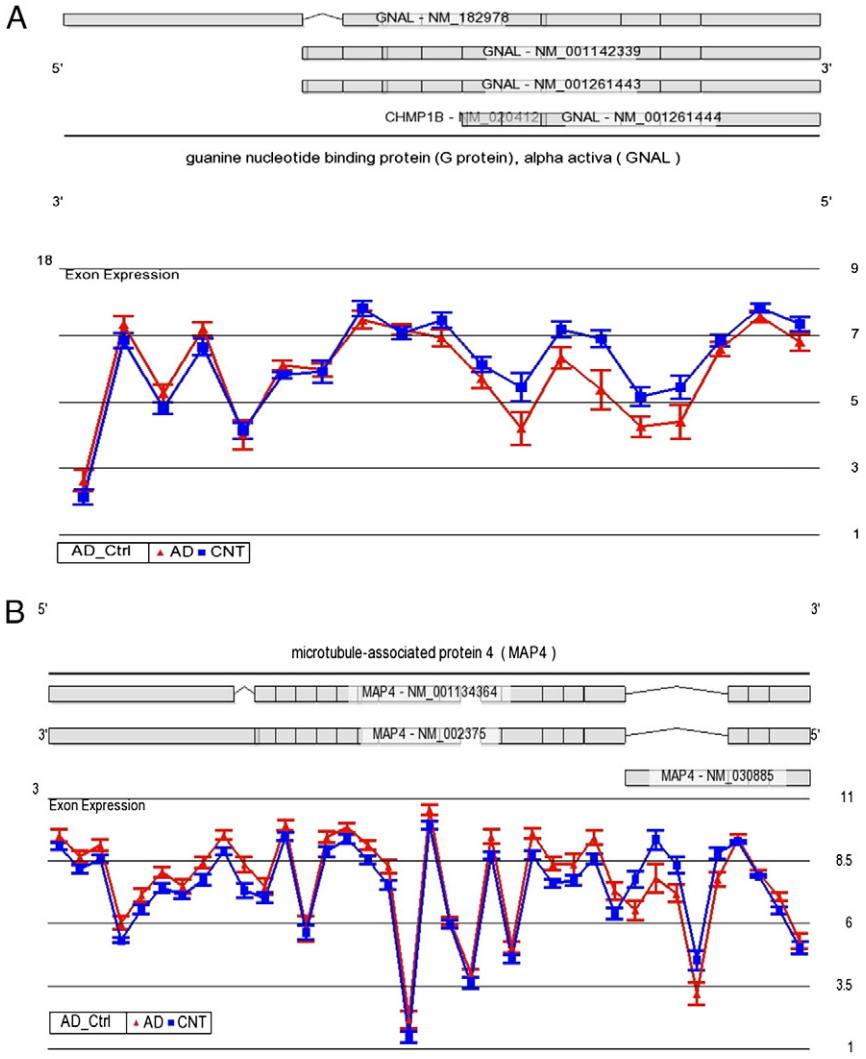

Fig. 2. Gene view plot revealed differential alternative splicing events associated with AD in GNAL and MAP4. The Y-axis is a $\log_2$ scale intensity value. (A) GNAL transcript variant 5 (NM_001261444) was significantly down-regulated in AD, whereas transcript variant 1 (NM_182978) exhibited no change or slight increase in AD. (B) MAP4 transcript variant 3 (NM_030885) was significantly down-regulated in AD, whereas transcript variant 1 (NM_002375) was significantly up-regulated.

differential alternative splicing events (78 genes) require further confirmation of the existence of corresponding transcript variants. The supplementary table of 100 candidates with differential alternative splicing associated with AD is available in GEO under accession number GSE37264.

Acknowledgments

This study was supported by Department of Clinical Research, Singapore General Hospital (Grant no: DCR/P09/2007).

References

[1] M.G. Tan, W.T. Chua, M.M. Esiri, A.D. Smith, H.V. Vinters, et al., Genome wide profiling of altered gene expression in the neocortex of Alzheimer's disease. J. Neurosci. Res. 88 (2010) 1157–1169.

[2] M. Roth, E. Tym, C.Q. Mountjoy, F.A. Huppert, H. Hendrie, et al., CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. Br. J. Psychiatry 149 (1986) 698–709.

[3] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, et al., Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34 (1984) 939–944.

[4] S.S. Mirra, A. Heyman, D. McKeel, S.M. Sumi, B.J. Crain, et al., The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41 (1991) 479–486.

[5] H. Braak, E. Braak, Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 82 (1991) 239–259.