Original Paper

Analysis of Correlation between Gene Expression and Aberrant Epigenetic Status in Alzheimer's Disease Brain

Kojiro Yano[1,a]

Abstract: Dysregulation of epigenetic mechanisms has been implicated in the pathogenesis of Alzheimer's disease (AD). It has been shown that epigenetic status in promoter regions can alter levels of gene expressions, but their influence on correlated expressions of genes and its dependency on the disease are unclear. Using publicly available microarray and DNA methylation data, this article infer how correlation in gene expression in non-demented (ND) and AD brain may be influenced by genomic promoter methylation. Pearson correlation coefficients of gene expression levels between each of 123 known hypomethylated genes and all other genes in the microarray dataset were calculated, and the mean absolute coefficients were obtained as an overall strength of gene expression correlation of the hypomethylated gene. The distribution of the mean absolute coefficients showed that the hypomethylated genes can be divided into two, by the mean coefficients above or below 0.15. The division of the hypomethylated genes by the mean coefficients was more evident in AD brain than in ND brain. On the other hand, hypermethylated genes had a single dominant group, and the majority of them had the mean coefficient below 0.15. These results suggest that the lower the DNA methylation, the higher the correlation of gene expression levels with the other genes in microarray data. The strength of gene expression correlation was also calculated between known AD risk genes and all other genes in microarray data. It was found that AD risk genes were more likely to have the mean absolute correlation coefficients above 0.15 in AD brain, when the evidence for their association with AD was strong, suggesting the link between DNA methylation and AD. In conclusion DNA methylation status is intimately associated with correlated gene expression, particularly in AD brain.

Keywords: microarray, gene expression, bioinformatics, Alzheimer's disease, DNA methylation

1.   Background

Alzheimer's disease (AD) is a neurodegenerative condition and the most common form of dementia. Recently, it has been suggested that AD pathogenesis may be associated with dysregulation of DNA methylation, which is an important feature of epigenetic mechanisms [1], [2]. In cancer cells, for example, promoter regions of tumor suppressor genes are often hypermethylated, resulting in decreased gene expression, whereas intergenic regions are commonly hypomethylated, which may lead to genomic instability [3]. In addition, DNA methylation status may affect neural differentiation in a number of ways [4]. For example, global hypomethylation by inhibition of DNMT1 caused premature astroglial differentiation of neural precursor cells [5]. Moreover, conditional DNMT1 mutant mice in which the DNA of cortical and hippocampal cells is severely hypomethylated showed significant neuronal cell death and impaired neuronal maturation [6]. In human AD brain and monkey AD-like regions, DNA methylation of amyloid precursor protein (APP) and presenilin 1 (PS1) genes are decreased in expression, compared to levels in non-demented (ND) brain [7], [8], and the brains of AD-affected twin siblings showed lower levels of DNA methylation [9]. These findings point to a close relationship between DNA hypomethylation and AD.

It has been shown that hypermethylation represses transcription [10], and that demethylation increases gene expression levels [11]. Therefore, if correlation in gene expressions is regulated at transcriptional levels, it is possible that altered transcription activities influence correlation in gene expression. In this report I examine whether correlation patterns of gene expressions can be associated with DNA methylation status and whether the association differs between in age-matched ND brain and in AD brain. This involved the uses of gene expression profiles and global DNA methylation profiling data from public data sources. Although the two sets of data came from different studies and therefore different collections of brains were used, I found that methylation status of gene promoters and the strength of gene correlations were closely linked. Namely, genes reported to be hypomethylated were more likely to have higher correlation with other genes in the microarray datasets while reportedly-hypermethylated genes showed the opposite. Moreover genes in the dataset from AD brain were likely to show the correlation patterns similar to those of hypomethylated genes and their patterns seemed to be correlated with their associated risk with AD.
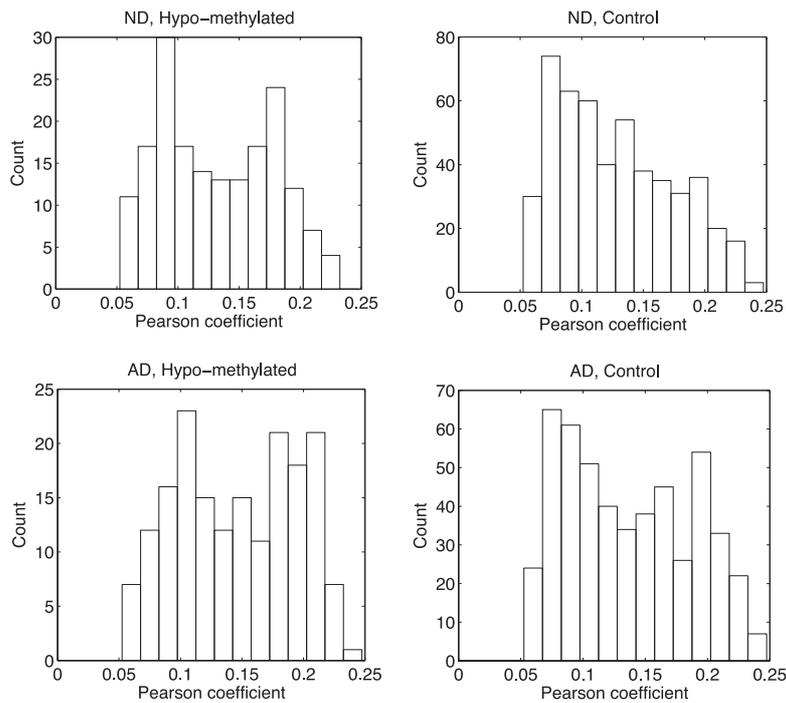
2.   Results and Discussion

2.1   Correlation Patterns of Hypomethylated and Hypermethylated Genes

First, I selected 123 genes, the promoters of which were specifically hypomethylated in brain compared to promoters in testis or

---
[1]   Osaka Institute of Technology, Hirakata, Osaka 573–0196, Japan
[a]   kyano@is.oit.ac.jp

Fig. 1   Distribution of mean absolute Pearson correlation coefficients of hypomethylated genes in non-demented (ND) and Alzheimer's disease (AD) brains. *Hypo-methylated* refers to the hypomethylated genes of the work of Schilling and Rehli [12], and *control* reflects data for 500 randomly selected genes.

Table 1   The peak counts in histograms for mean absolute Pearson coefficients of hypomethylated [12] and control genes in non-demented (ND) and Alzheimer's disease (AD) brains. Each cell in the table shows the peak counts below and above the coefficient = 0.15.

|	Hypomethylated coef.<0.15	Hypomethylated coef.>0.15	Control coef.<0.15	Control coef.>0.15
ND brain	30	23	74	36
AD brain	24	21	65	54

monocytes, according to Schilling and Rehli [12]. To characterize gene expression patterns, I obtained correlation coefficients between each of the hypomethylated genes and all other genes in the microarray dataset and calculated the mean absolute coefficient for each hypomethylated gene. In ND brain, the distribution of the mean correlation coefficients appeared to have two peaks which could be approximately separated when the threshold value of the correlation coefficient was set at 0.15 (**Fig. 1**, top left). The count of genes at the peak (termed the peak count) below the coefficient = 0.15 was 30, while the peak count above coefficient = 0.15 was 24. The two peaks were also observed in AD brain (Fig. 1, bottom left), and the peak counts were similar between above and below the coefficient = 0.15 (**Table 1**).

As a control, the mean correlation coefficients were calculated between 500 randomly selected genes and all other genes in the dataset. In ND brain, the peak above the coefficient = 0.15 was not evident (Fig. 1, top right) and its peak count was much smaller than that of the peak below the threshold (Table 1). In AD brain, two peaks (Fig. 1, bottom right), though less distinct compared to hypomethylated genes, could be seen in the histogram and similar peak counts below and above the coefficients = 0.15 (65 and 54, respectively) were observed.

The above results may indicate there were two groups of genes,

with different strength of correlations with other genes. The relative sizes of the two groups in ND brain were different between hypomethylated and control genes (Fig. 1 top left and right, respectively) suggesting the influence of methylation on gene expression correlations. However, the distribution of the correlation coefficients in AD brain looked similar between hypomethylated and control genes (Fig. 1 bottom left and right, respectively). This may be explained by a tendency of genes in AD brain as a whole toward hypomethylation, compared to in ND brain.

To examine this inference further, I carried out the same analysis for the list of hypo- and hyper-methylated genes of Christensen and colleagues [13]. **Figure 2** shows histograms for the mean absolute coefficients of the 200 least and most methylated genes against all other genes in microarray dataset. The histogram of the hypomethylated genes in ND brain did not have the distinct two peaks seen in the previous case (Fig. 2, top left), but the peak counts below and above the coefficient = 0.15 were comparable (**Table 2**). In AD brain, the mean absolute coefficient of the hypomethylated genes showed clearer two peak distribution (Fig. 2, bottom left) and the peak counts were now higher above the coefficient = 0.15 than below the threshold (Table 2). These results may imply a stronger hypomethylation tendency in AD brain compared to in ND brain. On the other hand, the histograms of hypermethylated genes in ND and AD brain had a dominant peak (Fig. 2, top right and bottom right, respectively) below the coefficient = 0.15 and an indistinct peak above it. Consequently the peak counts were much higher below the coefficient = 0.15 than those above the threshold (Table 2). These results seem to further support the hypothesis that the methylation status may influence the strength of gene expression correlations.
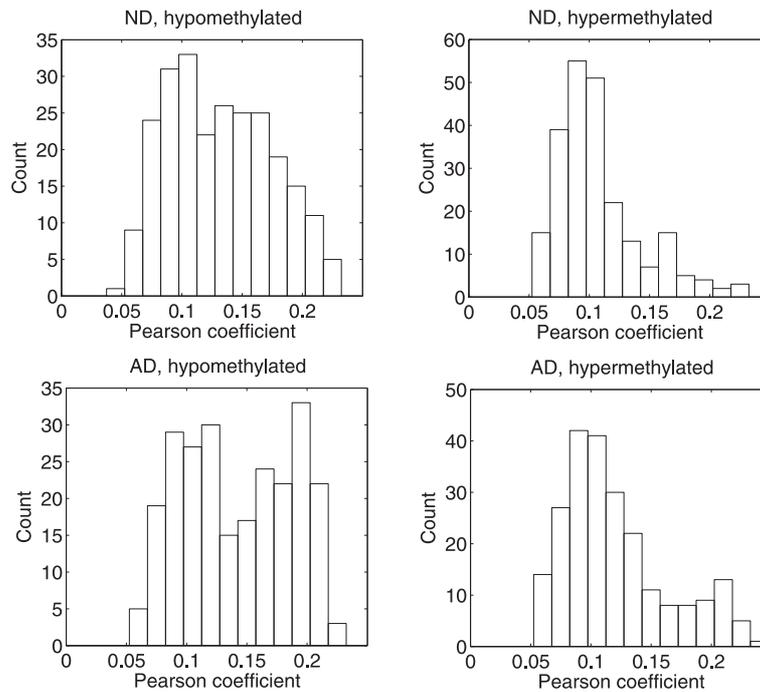
Fig. 2 Comparison of mean absolute Pearson correlation coefficients of hypo and hypermethylated genes in non-demented (ND) and Alzheimer's disease (AD) brains. The hypo- and hyper-methylated genes are those of Ref. [13].

Table 2 The peak counts in histograms for mean absolute Pearson coefficients of hypomethylated and hypermethylated genes [13] in non-demented (ND) and Alzheimer's disease (AD) brains. Each cell in the table shows the peak counts below and above the coefficient = 0.15.

	Hypomethylated coef.<0.15	Hypomethylated coef.>0.15	Hypermethylated coef.<0.15	Hypermethylated coef.>0.15
ND brain	33	25	55	15
AD brain	30	33	42	13

2.2 Correlation Patterns of AD Related Genes

The analysis of the absolute mean correlation coefficients of genes in ND and AD brains implied that DNA methylation may be linked to AD in one way or another. In order to gain more insights into this association, I selected the 30 genes most strongly associated with AD according to the Alzgene database [14] and calculated mean absolute correlation coefficients between such genes and all other genes in microarray datasets from ND and AD brain. The Alzgene database classifies the 30 genes into grades A, B, and C, based on the strength of epidemiological evidence. In **Table 3** these genes are shown, with the mean absolute correlation coefficients in ND and AD brains for each gene. Two of the eight genes of grade A had correlation coefficients above 0.15 in ND brain but the number increased to six in AD brain. The change in the mean coefficients was particularly large with CLU and TNK1, and these genes may be useful in the study of disease-dependent changes in DNA methylation and the influence thereof in correlated gene expression. On the other hand, only 1 of 4 genes of grade B had coefficients above 0.15 in both brains, while 4 and 5, respectively, of 18 genes of grade C, had correlation coefficients above 0.15 in ND and AD brains. This suggests that AD-related genes are more likely to show correlation patterns similar to those of hypomethylated genes when evidence for their involvement in AD is stronger, thus supporting a link between AD and DNA methylation.

3. Conclusions

In this paper the methylation status of genes was found to be associated with gene correlation patterns, particularly in AD brain. Namely, less methylated genes showed more correlations with other genes, probably because such genes exhibited higher transcriptional plasticity [15]. It appears that AD brain has a tendency toward hypomethylation, which may occur in a gene-specific manner, given that high-risk AD genes were prone to show high degrees of correlation, comparable to those of hypomethylated genes. This should be investigated further by global methylation profiling of ND and AD brains, combined with expression profiling, to test if modified methylation indeed affects gene expression levels and correlation patterns.

4. Methods

4.1 Microarray Data

The microarray data set GSE15222 from Gene Expression Omnibus [16] was used in analysis. This dataset was originally obtained by Webster and colleagues [17], and includes microarray data from 188 neuropathologically normal cortex samples and 176 cortex samples from patients with neuropathologically confirmed late-onset AD (LOAD) patients. Expression levels of 24,354 transcripts in the dataset were obtained using the Illumina Human Refseq-8 Expression BeadChip (Illumina, San Diego,

Table 3   Mean absolute coefficients of risk genes for Alzheimer's disease from Alzgene. Values above 0.15 are shown in **bold**.

Genes	Normal	AD	Grade
APOE	**0.170**	**0.166**	A
CLU	0.110	**0.161**	A
PICALM	**0.155**	**0.153**	A
TNK1	0.148	**0.223**	A
ACE	0.090	0.119	A
TFAM	0.147	**0.156**	A
CST3	0.131	**0.159**	A
IL1B	0.085	0.092	A
CR1	0.100	0.090	B
SORL1	**0.174**	0.103	B
CHRNB2	0.100	0.082	B
SORCS1	0.120	**0.192**	B
DAPK1	**0.207**	**0.204**	C
PRNP	**0.185**	**0.209**	C
MTHFR	0.075	0.098	C
GAB2	**0.165**	**0.226**	C
BDNF	0.117	**0.202**	C
NEDD9	0.128	0.132	C
CH25H	0.117	0.122	C
IL1A	0.096	0.114	C
TF	0.102	0.144	C
TNF	0.061	0.059	C
PGBD1	**0.174**	**0.213**	C
THRA	0.130	0.134	C
ENTPD7	0.103	0.113	C
IL33	0.135	0.130	C
GAPDHS	0.057	0.066	C
OTC	0.079	0.075	C
GALP	0.062	0.092	C
PSEN1	0.089	0.079	C

CA). I processed the data using the quantile normalization function (quantilenorm) of the Matlab Bioinformatics Toolbox (the MathWorks, Natick, MA).

4.2   DNA Methylation Data

DNA methylation data were obtained from the studies of Schilling and Rehli [12] and of Christensen and colleagues [13]. The first study examined CpG island methylations in about 20,000 promoters from testis, monocytes, and brain. Promoters were grouped as hypomethylated in brain if they were less methylated, by more than 2.4-fold, than those in testis and monocytes. The cited authors found 123 genes with hypomethylated promoters in the brain. The second study analyzed the methylation status of 214 normal human tissues, including brain, at 1,413 autosomal CpG loci. I selected the 200 least- and most-methylated loci and used genes associated with these loci in analysis.

4.3   List of AD Risk Genes

Thirty risk genes for AD showing the strongest epidemiological evidence were obtained from Alzgene website [14]. This site grades each gene by overall epidemiological credibility, based on the amount of reported evidence, consistency of data replication, and minimization of bias. The grades were used in this paper to classify AD risk genes and compare their patterns of gene expression correlations.

4.4   Statistical Analysis

Correlation of expression levels between any two genes was calculated as a Pearson correlation coefficient (PC) using the *corr* function in Matlab. The dendrogram of AD samples was created using the *pdist* (with the 'correlation' option), *linkage* (with the 'average' option), and *dendrogram* functions, from the Matlab Statistics toolbox.

Reference

[1]   Graff, J. and Mansuy, I.M.: Epigenetic dysregulation in cognitive disorders, *Eur. J. Neurosci.*, Vol.30, No.1, pp.1–8 (online), DOI: 10.1111/j.1460-9568.2009.06787.x (2009).

[2]   Sananbenesi, F. and Fischer, A.: The epigenetic bottleneck of neurodegenerative and psychiatric diseases, *Biol. Chem.*, Vol.390, No.11, pp.1145–1153 (online), DOI: 10.1515/BC.2009.131 (2009).

[3]   McCabe, M.T., Brandes, J.C. and Vertino, P.M.: Cancer DNA methylation: Molecular mechanisms and clinical implications, *Clin. Cancer. Res.*, Vol.15, No.12, pp.3927–3937 (online), DOI: 10.1158/1078-0432.CCR-08-2784 (2009).

[4]   Wen, S., Li, H. and Liu, J.: Epigenetic background of neuronal fate determination, *Prog. Neurobiol.*, Vol.87, No.2, pp.98–117 (online), DOI: 10.1016/j.pneurobio.2008.10.002 (2009).

[5]   Fan, G., Martinowich, K., Chin, M.H., He, F., Fouse, S.D., Hutnick, L., Hattori, D., Ge, W., Shen, Y., Wu, H., ten Hoeve, J., Shuai, K. and Sun, Y.E.: DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling, *Development.*, Vol.132, No.15, pp.3345–3356 (online), DOI: 10.1242/dev.01912 (2005).

[6]   Hutnick, L.K., Golshani, P., Namihira, M., Xue, Z., Matynia, A., Yang, X.W., Silva, A.J., Schweizer, F.E. and Fan, G.: DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation, *Hum. Mol. Genet.*, Vol.18, No.15, pp.2875–2888 (online), DOI: 10.1093/hmg/ddp222 (2009).

[7]   Wu, J., Basha, M.R., Brock, B., Cox, D.P., Cardozo-Pelaez, F., McPherson, C.A., Harry, J., Rice, D.C., Maloney, B., Chen, D., Lahiri, D.K. and Zawia, N.H.: Alzheimer's disease (AD)-like pathology in aged monkeys after infantile exposure to environmental metal lead (Pb): Evidence for a developmental origin and environmental link for AD, *J. Neurosci.*, Vol.28, No.1, pp.3–9 (online), DOI: 10.1523/JNEUROSCI.4405-07.2008 (2008).

[8]   Wang, S.-C., Oelze, B. and Schumacher, A.: Age-specific epigenetic drift in late-onset Alzheimer's disease, *PLoS One*, Vol.3, No.7, p.e2698 (online), DOI: 10.1371/journal.pone.0002698 (2008).

[9]   Mastroeni, D., McKee, A., Grover, A., Rogers, J. and Coleman, P.D.: Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer's disease, *PLoS One*, Vol.4, No.8, p.e6617 (online), DOI: 10.1371/journal.pone.0006617 (2009).

[10]   Rottach, A., Leonhardt, H. and Spada, F.: DNA methylation-mediated epigenetic control, *J. Cell Biochem.*, Vol.108, No.1, pp.43–51 (online), DOI: 10.1002/jcb.22253 (2009).

[11]   Yaqinuddin, A., Qureshi, S.A., Qazi, R., Farooq, S. and Abbas, F.: DNMT1 silencing affects locus specific DNA methylation and increases prostate cancer derived PC3 cell invasiveness, *J. Urol.*, Vol.182, No.2, pp.756–761 (online), DOI: 10.1016/j.juro.2009.03.082 (2009).

[12]   Schilling, E. and Rehli, M.: Global, comparative analysis of tissue-specific promoter CpG methylation, *Genomics*, Vol.90, No.3, pp.314–323 (online), DOI: 10.1016/j.ygeno.2007.04.011 (2007).

[13]   Christensen, B.C., Houseman, E.A., Marsit, C.J., Zheng, S., Wrensch, M.R., Wiemels, J.L., Nelson, H.H., Karagas, M.R., Padbury, J.F., Bueno, R., Sugarbaker, D.J., Yeh, R.-F., Wiencke, J.K. and Kelsey, K.T.: Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context, *PLoS Genet.*, Vol.5, No.8, p.e1000602 (online), DOI: 10.1371/journal.pgen.1000602 (2009).

[14]   Bertram, L., McQueen, M.B., Mullin, K., Blacker, D. and Tanzi, R.E.: Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database, *Nat. Genet.*, Vol.39, No.1, pp.17–23 (online), DOI: 10.1038/ng1934 (2007).

[15]   Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A., Haefliger, C., Horton, R., Howe, K., Jackson, D.K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., Pettett, R., Seemann, S., Thompson, C., West, T., Rogers, J., Olek, A., Berlin, K. and Beck, S.: DNA methylation profiling of human chromosomes 6, 20 and 22, *Nat. Genet.*, Vol.38, No.12, pp.1378–1385 (online), DOI: 10.1038/ng1909 (2006).

[16]   Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Muertter, R.N. and Edgar, R.:

NCBI GEO: Archive for high-throughput functional genomic data, *Nucleic Acids Res.*, Vol.37, Database issue, pp.D885–D890 (online), DOI: 10.1093/nar/gkn764 (2009).

[17] Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., McCorquodale, D.S., Cuello, C., Leung, D., Bryden, L., Nath, P., Zismann, V.L., Joshipura, K., Huentelman, M.J., Hu-Lince, D., Coon, K.D., Craig, D.W., Pearson, J.V., Group, N.A.C.C.-N., Heward, C.B., Reiman, E.M., Stephan, D., Hardy, J. and Myers, A.J.: Genetic control of human brain transcript expression in Alzheimer disease, *Am. J. Hum. Genet.*, Vol.84, No.4, pp.445–458 (online), DOI: 10.1016/j.ajhg.2009.03.011 (2009).

Kojiro Yano was born in 1973. He received his M.D. from Chiba University, M.Res. from University of Manchester, Ph.D. from University of Liverpool. Previously, he worked as a group leader at University of Cambridge and is currently an associate professor at Osaka Institute of Technology. He specializes in computational biology, bioinformatics and medical informatics.

(Communicated by *Masahiro Hattori*)