Revised Manuscript

Artificial Neural Networks link one-carbon metabolism to gene-promoter methylation in Alzheimer's disease

Enzo Grossi[1,2], Andrea Stoccoro[3,4], Pierpaola Tannorella[3], Lucia Migliore[3,5], Fabio Coppedè[3,5,*]

1) Villa Santa Maria Institute, Tavernerio, Italy.

2) Semeion Research Centre of Sciences of Communication, Rome, Italy.

3) Department of Translational Research and New Technologies in Medicine and Surgery, Section of Medical Genetics, University of Pisa, Pisa, Italy.

4) Doctoral School in Genetics, Oncology, and Clinical Medicine, University of Siena, Siena, Italy.

5) Interdepartmental Research Center Nutrafood "Nutraceuticals and Food for Health", University of Pisa, Pisa, Italy.

Running Title: Artificial intelligence and DNA methylation

Address for correspondence:

Prof. Fabio Coppedè, Ph.D.
Department of Translational Research
and New Technologies in Medicine and Surgery,
Medical Genetics Lab.
University of Pisa,
Medical School,
Via Roma 55,
56126 Pisa
Phone 050 2218544
E-mail: fabio.coppede@med.unipi.it

Abstract

BACKGROUND: There is increasing interest in DNA methylation studies in Alzheimer's disease (AD), but little is still known concerning the relationship between gene-promoter methylation and circulating biomarkers of one-carbon metabolism in the patients.

OBJECTIVE: to detect the connections among circulating folate, homocysteine (hcy) and vitamin B12 levels and promoter methylation levels of *PSEN1*, *BACE1, DNMT1, DNMT3A, DNMT3B* and *MTHFR* genes in blood DNA.

METHODS: We applied a data mining system called Auto Contractive Map to an existing database of 100 Alzheimer's disease (AD) and 100 control individuals.

RESULTS: Low vitamin B12 was linked to the AD condition, to low folates and to high hcy. Low *PSEN1* methylation was linked to low folate levels as well as to low promoter methylation of *BACE1* and *DNMTs* genes. Low hcy was linked to controls, to high folates and vitamin B12, as well as to high methylation levels of most of the studied genes.

CONCLUSIONS: The present pilot study suggests that promoter methylation levels of the studied genes are linked to circulating levels of folates, hcy, and vitamin B12.

Keywords: DNA methylation, folate, homocysteine, vitamin B12, Alzheimer's disease, PSEN1, BACE1, DNMT, MTHFR

Introduction

Alzheimer's disease (AD) individuals are characterized by impaired one-carbon metabolism, which is reflected in increased circulating homocysteine (hcy) and reduced levels of serum folates and related B-group vitamins [1]. Indeed, high hcy and low folates are regarded as AD risk factors, and could contribute to global and gene-specific DNA methylation changes in AD [2,3]. Results from either *in vitro* studies or animal models revealed that folic acid alone or formulations of B-group vitamins induce changes in the methylation and or expression of genes required for amyloid-β peptide (Aβ) production, such as *PSEN1* and *BACE1*, or involved in DNA methylation reactions, including *DNMTs* genes [4-7], but data in AD patients are missing.

In the present study we applied Artificial Neural Networks (ANNs) to unravel the connections between circulating folate, hcy and vitamin B12 and promoter methylation levels of genes required for Aβ production (*PSEN1* and *BACE1*) and DNA methylation reactions (*DNMT1, DNMT3A, DNMT3B,* and *MTHFR*) in blood DNA obtained from 100 late-onset AD (LOAD) patients and 100 matched controls. Particularly, we used the Auto Contractive Map algorithm (Auto-CM), a special kind of ANN able to define the strength of the associations of each variable with all the others and to visually show the map of their main connections and the basic semantic of their ensemble [8]. Auto-CM has been successfully applied to AD datasets, for example to link serum folate and hcy levels to brain atrophy [9], to unravel genetic polymorphisms linked to impaired one-carbon metabolism [10], to detect the connections among studied variables in the nun study [8], to differentiate among various forms of dementia [10], or to detect the predictors of response to cholinesterase inhibitors [12].

Materials and Methods

Study population

The analysis was based on 100 LOAD patients and 100 healthy matched controls (Table 1) selected from a previously described database [13] for whom all the following information was available: plasma hcy and serum folate and vitamin B12 levels; promoter methylation levels of *PSEN1, BACE1, DNMT1, DNMT3A, DNMT3B* and *MTHFR* genes; age at sampling; gender. All the methylation and biochemical analyses (folate, hcy, vitamin B12) were performed simultaneously after blood drawings. A trained neurologist from the Department of Neuroscience of the Pisa University Hospital visited cases and controls before inclusion in the study. LOAD patients met the diagnostic and statistical manual of mental disorders criteria (DSMIV) [14,15]. Healthy volunteer subjects were recruited simultaneously with LOAD patients and matched for age, gender, and ethnicity (all Italian Caucasians). All the control subjects underwent a rigorous neurological examination before inclusion in the study, in order to exclude the presence of cognitive impairment or any other kind of neurological disorder. Furthermore, control subjects were also investigated for their familial history of neurological disorders, and only individuals with no relatives who developed AD or related disorders were included in the study. In addition, individuals taking vitamins, drugs, or supplements known or suspected to interfere with one-carbon metabolism and DNA methylation reactions were not enrolled [13]. Each subject gave an informed and written consent for the inclusion in the study that received approval from the Ethics Committee of the Pisa University Hospital (Protocol number 3618/2012).

Epigenetic and biochemical data collection

Peripheral blood samples were collected from each subject for the evaluation of folate, hcy and vitamin B12 levels. Promoter methylation levels of *PSEN1, BACE1, DNMT1, DNMT3A, DNMT3B* and *MTHFR* genes were assessed with validated methylation sensitive high resolution melting (MS-

HRM) protocols [13]. All the data had been previously collected as detailed elsewhere [13] and were available in the database at the time of ANNs analysis. Table 1 shows the distribution of studied variables between LOAD and control subjects.

Semantic connectivity map

In order to graphically show the most important connections among variables we used an artificial adaptive system called Auto-CM [8], a special kind of ANN that develops weights that are proportional to the strength of the associations of all variables each other. The weights are then transformed in physical distances so that couples of variables whose connection weights are higher become nearer and vice versa. After the training phase, the weights matrix of the Auto-CM represents the warped landscape of the dataset. Subsequently, a simple filter to the weights matrix of the Auto-CM system was applied to obtain a map of the main connections between the variables of the dataset and the basic semantic of their similarities, defined connectivity map as detailed elsewhere [8]. Categorical variables were left as such and numerical variables (folates, hcy, vitamin B12, and promoter methylation levels of the six genes under investigation) have been transformed in input variables constructing for each of them, scaled from 0 to 1, its complement, as detailed elsewhere [9,10]. For example the variable *MTHFR* promoter methylation has natural values ranging from 5.9% (lowest observed value) to 74.6% (highest value) in our cohort. According to the transformation, 5.9 (the lowest value) becomes 0 and 74.6 (the highest value) becomes 1, and vice-versa. All the other observed *MTHFR* promoter methylation values are scaled in this new range. Doing this pre-processing for all the studied variables allows a graphical visualization of the projection of each variable in the map according to its low and high values.

Results

Figure 1 shows the semantic connectivity map obtained with Auto-CM. Variables showing the maximal amount of connections with other variables are called "hubs". A numerical value is applied to each edge of the graph and is proportional to the strength of the connection between two variables (s.a. = strength of association, ranges from 0 = not connected to 1 = highly connected). Results clearly indicate that LOAD patients and controls are well separated each other. Particularly AD patients are highly connected to low vitamin B12 levels (s.a. = 0.96), which in turn are connected to low folates (s.a. = 0.99) and high hcy (s.a. = 0.79). On the contrary, controls are highly connected to low hcy (s.a. = 0.96), which is in turn connected to high folates (s.a. = 0.75) and high vitamin B12 (s.a. = 0.82). Table 1 shows that vitamin B12 levels were significantly reduced in LOAD patients with respect to controls; furthermore, linear regression analysis pointed out a significant correlation between serum folate and vitamin B12 levels, and an inverse correlation between vitamin B12 and hcy levels, suggesting that individuals with lower vitamin B12 levels are likely to have relatively low folate and high hcy values (Supplementary Figure 1).

Very interestingly low *PSEN1* methylation works as hub of the system and is highly connected to low methylation levels of all the other investigated genes. Furthermore, low *PSEN1* methylation is highly connected to low folates (s.a. = 0.99) as well as to low hcy (s.a. = 0.99). Low hcy is another hub of the system, directly connected not only to the condition of being a control or to high folate and vitamin B12 levels, but also to high methylation levels of most of the studied genes (s.a. ranging from 0.53 to 0.96). Gender is not directly linked to gene promoter methylation, whilst increasing age (>80 years) is directly connected to low methylation of *DNMT1*, and younger ages are linked to low hcy (Figure 1).

Discussion

ANNs represent a valid tool to dissect complex non-linear interactions among high numbers of variables, and in the present pilot study we applied Auto-CM to reveal the connections between circulating biomarkers of one-carbon metabolism (folate, hcy, and vitamin B12 levels) and promoter methylation levels of selected genes in blood DNA of LOAD and control individuals. Very interestingly, Auto-CM revealed that low levels of vitamin B12 were directly connected to the LOAD condition; furthermore, low vitamin B12 levels were linked to low folates and high hcy. Conversely, low hcy connected to high folates and vitamin B12 were directly linked to the control condition. Similar findings have been already obtained by Auto-CM in two previous independent AD datasets [9,10], and confirm recent meta-analyses of a large amount of literature overall suggesting that reduced folates and increased hcy are both AD risk factors [1]. Conventional statistical approaches revealed that vitamin B12 levels were significantly reduced in the present LOAD cohort with respect to controls (Table 1). Furthermore, linear regression analysis revealed a positive correlation between serum vitamin B12 and serum folate levels, and an inverse correlation between serum vitamin B12 and plasma hcy levels (Supplementary Figure 1), suggesting that LOAD individuals with low vitamin B12 levels are also characterized by relatively low folates and high hcy than individuals with higher vitamin B12 levels, as revealed by Auto-CM analysis. ANNs pointed out also some connections between the study variables and gender. For example, male gender was linked to low vitamin B12 by Auto-CM, and a conventional statistical approach showed a trend for higher vitamin B12 in females than in males in our cohort (Supplementary Figure 1).

More interestingly, Auto-CM revealed connections between biomarkers of one-carbon metabolism and gene-promoter methylation levels of the selected genes. Particularly, present data confirm previous *in vitro* and animal model results suggesting a link between dietary folates and related B-group vitamins and the methylation levels of genes required for Aβ production, such as *PSEN1* and *BACE1* [4-7]. *PSEN1* codes for the presenilin 1 protein, involved in the γ-secretase

cleavage of the amyloid-β protein precursor (AβPP), and is mutated in familial early-onset AD cases [16]. *BACE1* codes for the β-secretase protein that, together with γ-secretase, is involved in the amyloidogenic cleavage of AβPP leading to the production Aβ peptides [17]. The reduction of folate and vitamin B12 in culture medium of neuroblastoma cell lines was linked to *PSEN1* demethylation, increased levels of presenilin 1 and BACE proteins, and increased production of Aβ peptides [4]. Similar results were observed in brain neurons of mice following a combined dietary deficiency of folate, vitamin B12 and vitamin B6 [18]. Studies in humans have revealed that both genes are hypo-methylated in blood DNA of AD and control individuals [13,19,20], but no correlation between gene-promoter methylation and folate or vitamin B12 levels is yet available in human AD tissues.

The present Auto-CM analysis also linked low hcy levels to high promoter methylation of genes required for DNA methylation reactions (*DNMT1, DNMT3A, DNMT3B,* and *MTHFR*), whilst low folates and advancing age were linked to their low methylation. So that, present results confirm previous cell culture studies revealing that folic acid can impact on DNA methylation levels through the modulation of proteins required for DNA methylation reactions [5-7]. Particularly, DNMT1 is required for the maintenance of DNA methylation patterns during development and cell division, whilst DNMT3a and DNMT3b are *de novo* methyltransferases that establish DNA methylation patterns during early development [21]. Studies in mouse neuronal cells have shown that folic acid stimulates DNMTs gene and protein expression, and DNMTs activities, but data in human LOAD patients are missing [5,6]. The *MTHFR* gene codes for methylenetetrahydrofolate reductase, a key protein for the inter-conversion of folate derivatives required for DNA methylation reactions, and *MTHFR* genetic polymorphisms have been associated to LOAD risk [22]. Previous studies in humans showed that the *MTHFR* promoter shows inter-individual variability in DNA methylation patterns that correlate with circulating folate, vitamin B12 and/or hcy levels [13,23].

We are well aware of the limits of the present study as most of the genes that we have investigated showed low methylation levels in blood DNA of AD patients in both present and

previous investigations [13,19,20], so that their clinical significance is still debated [19]. However, at best of our knowledge the present is one of the first evidence linking circulating folate, hcy, and vitamin B12 levels to DNA methylation in human LOAD subjects. Therefore, present data are indicative that conditions of reduced folate and increased hcy levels, such as those often observed in AD patients, might contribute to either global or gene-specific methylation changes that have been often observed in blood and brains of those subjects [24], suggesting that further studies are warranted to clarify the link between one-carbon metabolism and DNA methylation in AD. We also acknowledge that the present investigation is based on a cohort of only 100 LOAD cases and 100 matched controls, so that present data require confirmation in larger cohorts of individuals. In this regard, the present must be considered as a pilot study addressing how several variables are linked each other in a given dataset. As such, the present ANNs investigation does not infer about study power or statistically significant differences between groups, but helps addressing complex and often non-linear interactions among variables, suggesting which ones are closer or not to the others. In addition ANNs, at variance with the classical statistical tests, can manage complexity even with relatively small samples and to the subsequent unbalanced ratio between variables and records. In this connection, it is important to note that adaptive learning algorithms of inference, based on the principle of a functional estimation like ANNs, can overcome to a certain degree the problem of dimensionality [8]. Unfortunately, as recently reviewed by Bennett and coworkers, we are still at the beginning of our comprehension of the contribution of epigenetic mechanisms in LOAD pathogenesis, most of the available studies are limited to a few dozens or less than 100 cases, and correlations between epigenetic changes and environmental/dietary factors are often missing [25]. Within this context, we believe that the present observation of a connection between promoter methylation levels of a panel of genes and circulating levels of folate, vitamin B12 and hcy, coupled with similar evidence coming from previous *in vitro* or animal investigations [4-7], should stimulate novel questions in AD research to be addressed in future studies, including the following: 1) “Are the global methylation changes so far observed in DNA from AD brains and peripheral tissues

related to the nutritional status, and particularly to the bioavailability of methyl donor compounds?" 2) "Can we counteract age-related changes in global and/or gene-specific methylation by means of nutritional interventions?" These points could be rigorously addressed only by planning either prospective or retrospective studies aimed at linking both global and gene-specific methylation changes in AD tissues with the bioavailability of methyl donor compounds. Those studies could also clarify whether AD-related DNA methylation changes can be prevented by dietary interventions.

Acknowledgements

Research was funded by the Italian Ministry of Health (GR-2009-1606229, FC Principal Investigator).

References

[1] Shen L, Ji HF (2015) Associations between Homocysteine, Folic Acid, Vitamin B12 and Alzheimer's Disease: Insights from Meta-Analyses. *J Alzheimers Dis*. **46**,777-790.

[2] Coppedè F (2010) One-carbon metabolism and Alzheimer's disease: focus on epigenetics. *Curr Genomics* **11**, 246-260.

[3] Araújo JR, Martel F, Borges N, Araújo JM, Keating E (2015) Folates and aging: Role in mild cognitive impairment, dementia and depression. *Ageing Res Rev* **22**, 9-19.

[4] Fuso A, Seminara L, Cavallaro RA, D'Anselmi F, Scarpa S (2005) S-adenosylmethionine/homocysteine cycle alterations modify DNA methylation status with consequent deregulation of PS1 and BACE and beta-amyloid production. *Mol Cell Neurosci* **28**,195-204.

[5] Fuso A, Nicolia V, Cavallaro RA, Scarpa S (2011) DNA methylase and demethylase activities are modulated by one-carbon metabolism in Alzheimer's disease models. *J Nutr Biochem* **22**, 242-251.

[6] Li W, Jiang M, Zhao S, Liu H, Zhang X, Wilson JX, Huang G (2015) Folic Acid Inhibits Amyloid β-Peptide Production through Modulating DNA Methyltransferase Activity in N2a-APP Cells. *Int J Mol Sci* **16**, 25002-25013.

[7] Li W, Liu H, Yu M, Zhang X, Zhang M, Wilson JX, Huang G (2015) Folic acid administration inhibits amyloid β-peptide accumulation in APP/PS1 transgenic mice. *J Nutr Biochem* **26**, 883-891.

[8] Buscema M, Grossi E, Snowdon D, Antuono P (2008) Auto-Contractive Maps: an artificial adaptive system for data mining. An application to Alzheimer disease. *Curr Alzheimer Res* **5**, 481-98.

[9] Gallucci M, Zanardo A, Bendini M, Di Paola F, Boldrini P, Grossi E (2014) Serum folate, homocysteine, brain atrophy, and auto-CM system: The Treviso Dementia (TREDEM) study. *J Alzheimers Dis* **38**,581-587.

[10] Coppedè F, Grossi E, Buscema M, Migliore L (2013) Application of artificial neural networks to investigate one-carbon metabolism in Alzheimer's disease and healthy matched individuals. *PLoS One* **8**, e74012.

[11] Caffarra P, Gardini S, Dieci F, Copelli S, Maset L, Concari L, Farina E, Grossi E (2013) The qualitative scoring MMSE pentagon test (QSPT): a new method for differentiating dementia with Lewy Body from Alzheimer's disease. *Behav Neurol* **27**, 213-220.

[12] Gallucci M, Spagnolo P, Aricò M, Grossi E (2016) Predictors of Response to Cholinesterase Inhibitors Treatment of Alzheimer's Disease: Date Mining from the TREDEM Registry. *J Alzheimers Dis*, in press.

[13] Tannorella P, Stoccoro A, Tognoni G, Petrozzi L, Salluzzo MG, Ragalmuto A, Siciliano G, Haslberger A, Bosco P, Bonuccelli U, Migliore L, Coppedè F (2015) Methylation analysis of multiple genes in blood DNA of Alzheimer's disease and healthy individuals. *Neurosci Lett* 600, 143-147.

[14] The Dementia Study Group of the Italian Neurological Society (2000) Guidelines for the diagnosis of dementia and Alzheimer's disease. *Neurol Sci* **21**,187-194.

[15] American Psychiatric Association (1994) *Diagnostic and statistical manual of mental disorders*, *fourth ed.,* American Psychiatric Association, Washington DC.

[16] Steiner H, Capell A, Leimer U, Haass C (1999) Genes and mechanisms involved in beta-amyloid generation and Alzheimer's disease. *Eur Arch Psychiatry Clin Neurosci* **249,** 266-270.

[17] Vassar R, Bennett BD, Babu-Khan S, Kahn S, Mendiaz EA, Denis P, Teplow DB, Ross S, Amarante P, Loeloff R, Luo Y, Fisher S, Fuller J, Edenson S, Lile J, Jarosinski MA, Biere AL, Curran E, Burgess T, Louis JC, Collins F, Treanor J, Rogers G, Citron M (1999) Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* **286**, 735-741.

[18] Fuso A, Nicolia V, Cavallaro RA, Ricceri L, D'Anselmi F, Coluccia P, Calamandrei G, Scarpa S (2008) B-vitamin deprivation induces hyperhomocysteinemia and brain S-adenosylhomocysteine,

depletes brain S-adenosylmethionine, and enhances PS1 and BACE expression and amyloid-beta deposition in mice. *Mol Cell Neurosci* **37**, 731-746.

[19] Piaceri I, Raspanti B, Tedde A, Bagnoli S, Sorbi S, Nacmias B (2015) Epigenetic modifications in Alzheimer's disease: cause or effect? *J Alzheimers Dis* **43**,1169-7113.

[20] Wang SC, Oelze B, Schumacher A (2008) Age-specific epigenetic drift in late-onset Alzheimer's disease. *PLoS One* **3**, e2698.

[21] Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* **10**, 805-811.

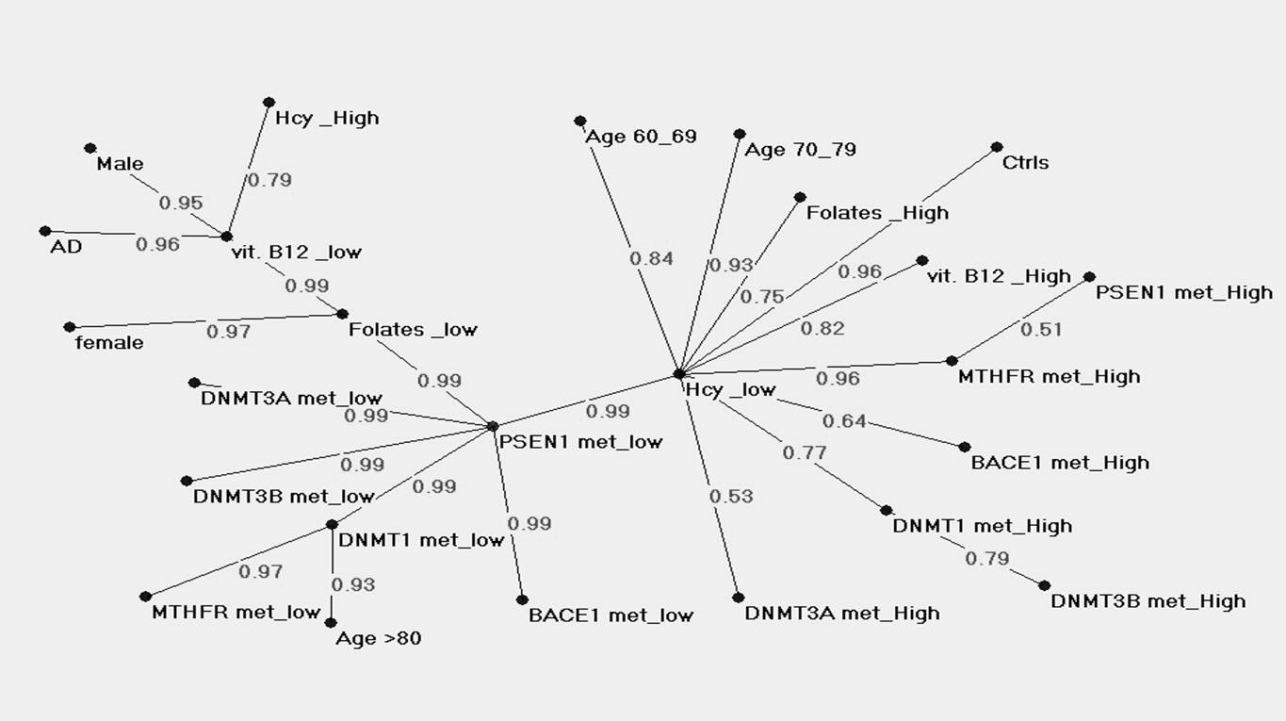
[21] Román GC (2015) MTHFR Gene Mutations: A Potential Marker of Late-Onset Alzheimer's Disease? *J Alzheimers Dis* **47**, 323-327.

[23] Wei LK, Sutherland H, Au A, Camilleri E, Haupt LM, Gan SH, Griffiths LR (2015) A potential epigenetic marker mediating serum folate and vitamin B12 levels contributes to the risk of ischemic stroke. *Biomed Res Int* **2015**, 167976.

[24] Millan MJ (2014) The epigenetic dimension of Alzheimer's disease: causal, consequence, or curiosity? *Dialogues Clin Neurosci* **16**, 373-393.

[25] Bennett DA, Yu L, Yang J, Srivastava GP, Aubin C, De Jager PL (2015) Epigenomics of Alzheimer's disease. *Transl Res* **165**, 200-220.

Figure 1.



Legend to Figure 1:

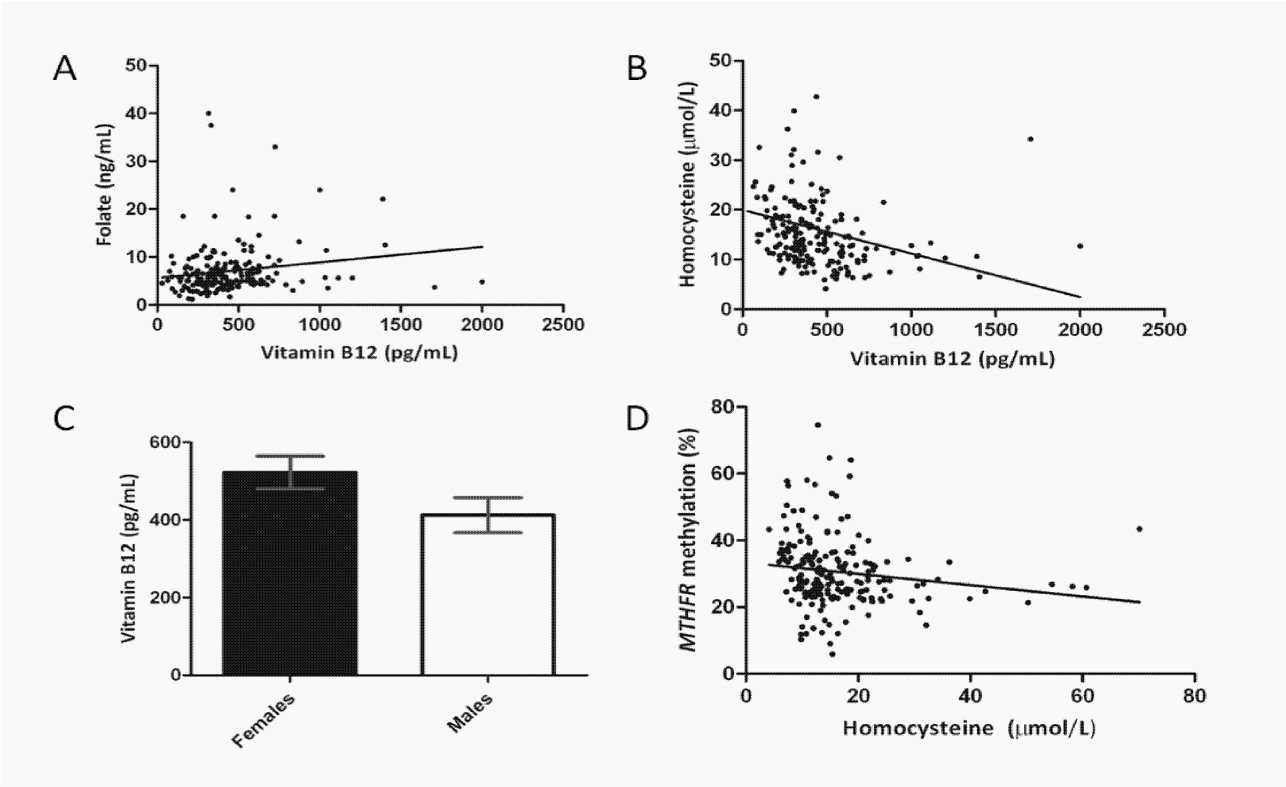
Semantic connectivity map obtained with Auto-Cm system. The numbers on the arches of the graph refer to the strength of the association between two adjacent nodes. The value ranges from 0 (not linked) to 1 (highly linked).

Table 1. Distribution of study variables between cases and controls

Variable	LOAD (n = 100)	Controls (n = 100)	*p*-value[a]
Gender (M/F)	46/54	46/54	1.00
Age at sampling (years): mean ± SD	77.5 ± 7.5	76.8 ± 8.4	0.53
Homocysteine (μmol/l): mean ± SD	17.4 ± 11.8	14.8 ± 6.3	0.17
Folate (ng/ml): mean ± SD	6.8 ± 4.9	6.9 ± 4.7	0.77
Vitamin B12 (pg/ml): mean ± SD	403 ± 218	472 ± 310	0.04
PSEN1 methylation (%): mean ± SD	0.6 ± 1.2	0.9 ± 1.9	0.18
BACE1 methylation (%): mean ± SD	0.5 ± 0.9	0.8 ± 1.2	0.06
DNMT1 methylation (%): mean ± SD	2.3 ± 2.9	1.9 ± 2.5	0.29
DNMT3A methylation (%): mean ± SD	0.9 ± 2.2	1.2 ± 2.2	0.33
DNMT3B methylation (%): mean ± SD	2.4 ± 4.1	2.7 ± 3.7	0.59
MTHFR methylation (%): mean ± SD	30.8 ± 11.2	30.3 ± 9.9	0.74

[a] *p*-value was obtained by chi-square analysis or Student's t test

Supplementary Material



Supplementary Figure 1: Linear regression analysis pointed out a positive correlation between vitamin B12 and folate values (r = 0.16; p = 0.019) (A), and an inverse correlation between hcy and vitamin B12 values (r = -0.25; p = 0,0004) (B). Vitamin B12 values tended to be higher in females than in males (p = 0.076; analysis of variance) (C). Similarly, linear regression analysis revealed an inverse correlation between hcy levels and *MTHFR* promoter methylation (r = -0.15; p = 0.03) (D). These data can help the reader to understand some of the connections between variables revealed by Auto-CM analysis.