ELSEVIER

Neurobiology of Aging 31 (2010) 257–269

NEUROBIOLOGY OF AGING

www.elsevier.com/locate/neuaging

Multivariable network associated with cognitive decline and dementia

Federico Licastro [a,*], Elisa Porcellini [a], Martina Chiappelli [a], Paola Forti [b], Massimo Buscema [c], Giovanni Ravaglia [b], Enzo Grossi [d]

[a] *Department of Experimental Pathology, Via San Giacomo 14, 40126 Bologna, Italy*
[b] *Department of Internal Medicine, Cardiology and Hepatology, School of Medicine, Via Albertoni 15, 40138 Bologna, Italy*
[c] *Semeion Research Center of Communication Science, Via Sersale 117, 00128 Roma, Italy*
[d] *Bracco SpA, via Folli 50, 20134 Milan, Italy*

Received 25 September 2007; received in revised form 25 March 2008; accepted 26 March 2008
Available online 15 May 2008

Abstract

Data mining of a large data base from the population longitudinal study named "The Conselice Study" has been the focus of the present investigation. Initially, 65 years old or older participants were interviewed, underwent medical and cognitive examination, and were followed up for 5 years: 937 subjects completed the follow-up. Relationships of 35 genetic and/or phenotypic factors with incident cognitive decline and dementia were investigated. The new mathematical approach, called the Auto Contractive Map (AutoCM), was able to show the differential importance of each variables. This new variable processing created a semantic connectivity map that: (a) preserved non-linear associations; (b) showed connection schemes; (c) captured the complex dynamics of adaptive interactions. This method, based on an artificial adaptive system, was able to define the association strength of each variable with all the others. Few variables resulted to be aggregation points and were considered as major biological hubs. Three hubs were identified in the hydroxyl-methyl-gutaryl-CoA reductase (HMGCR) enzyme, plasma cholesterol levels and age. Gene variants and cognate phenotypic variables showed differential degrees of relevance to brain aging and dementia.

This data analysis method was compared with another mathematical model called mutual information relevance network and results are presented and discussed.

Keywords: Brain aging; Dementia; Data base analysis; Connectivity map; Predictive factors

1. Introduction

Alzheimer' disease (AD) is a chronic progressive disease and the most frequent cause of mental disability and loss of independence among the elderly (Aronson et al., 1991). The disease is characterized by neuro-pathological hallmarks such as, synapsis loss, extracellular amyloid deposition, intracellular fibrillary tangle deposits and neuronal degeneration (Terry, 1994; Trojanowski et al., 1997). A prominent neuropathological feature of the AD brain is also represented by astrogliosis and microglia activation (Griffin et al., 1989; McGeer et al., 1993; Rogers et al., 1988). Abnormal activation of glia cells is now considered an early phenomenon associated with the development of the disease (Griffin et al., 1998) and has been suggested to be implicated in the pathogenesis of AD (Mrak et al., 1995). Genetic studies on inflammatory gene polymorphism associated with the disease have reinforced the notion that abnormal immune responses in the brain play a pivotal role in the disease (Licastro, 2002; Licastro and Chiappelli, 2003).

Some inflammatory genetic markers and the levels of their cognate proteins in the blood have been related to the conversion of pre-dementia states, such as subjects with mild cognitive impairment (MCI) or cognitive impairment and no dementia (CIND) to AD (Chiappelli et al., 2006a,b). A gene

* Corresponding author at: Dipartimento di Patologia Sperimentale, Via San Giacomo 14, 40126 Bologna, Italy. Tel.: +39 0512094730; fax: +39 0512094746.
E-mail address: licastro@alma.unibo.it (F. Licastro).

0197-4580/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.neurobiolaging.2008.03.019

polymorphism in the promoter region of an acute phase protein called alpha-1 antichymotrypsin (ACT) or SERPINA 3, has been found to be associated with an increased risk of early onset AD and levels of the ACT protein were elevated in sera from AD and CIND patients (Licastro et al., in press). These findings have raised the question whether genetic or phenotypic markers might be used for the screening of persons at high risk of developing cognitive decline and dementia before clinical manifestation of the diseases. The answer to this question might open the possibility of starting preventive protocols for high-risk healthy subjects with the goal of decreasing AD incidence.

AD is a complex multi-factorial disease and it is unlikely that a single biomarker may carry enough information for screening the potential risk of cognitive decline and dementia. Therefore, the use of several biomarkers, either genetic or phenotypic, may be necessary for a comprehensive screening protocol.

To approach this complex situation, informative biomarkers should be generated during longitudinal studies that can validate the clinical endpoints, e.g. cognitive decline, dementia or healthy cognitive performance.

The statistical evaluation of multiple variables in a sufficiently large population is another complex issue and new statistical models able to connect several factors with the disease, to evaluate the degree of linkage among variables and their association with the disease or its absence are needed.

The Conselice Study of brain aging is a population-based prospective study focused on an homogeneous elderly population from Northern Italy (Forti et al., 2001; Ravaglia et al., 2001). The principal aim of this investigation was to explore environmental, epidemiological and intrinsic risk factors for dementia in the elderly (Ravaglia et al., 2001).

From this study a biological and clinical data base during the 5-year follow-up has been generated and biological markers have been found individually associated or not with AD risk, incident cognitive decline and incident AD (Ravaglia et al., 2005; Ravaglia et al., 2006a, Sep. 28; Ravaglia et al., 2006b; Ravaglia et al., 2007). However, results were not conclusive or completely satisfactory, because of the limited power of classical statistical analysis and the difficulty in solving multiple concomitant variable analysis.

Here, we applied a novel data mining process to concomitantly explore the possible association of 35 different variables with CIND and AD and the possible presence of patterns or systematic relationship among variables, as recently described in other topics of medicine (Buscema and Grossi, in press).

This method of data mining is an analytical process designed to search a data base for consistent patterns and/or systematic relationships between variables. The method has the aim to detect patterns from new subsets of data. The ultimate goal of data mining is to discover hidden trends and associations among variables.

The more common algorithms of linear projections of variables are the principal component analysis (PCA) and the independent component analysis (ICA); the former requires a Gaussian distribution of data, while the latter does not require any specific distribution. These classical statistical techniques have limited power when the relationships between variables are non-linear. Moreover, PCA and ICA are not able to preserve the geometrical structure of the original space. Application of these methods may lose important information and establishing precise association among variables having only the contiguity as a known element is difficult. Another limitation of currently used statistical methods is that mapping is generally based on a specific kind of "distance" among variables (e.g. Euclidean, City block, correlation, etc.) and gives origin to a "static" projection of possible associations. In other words, the intrinsic dynamics due to active interactions of variables in living systems of the real world (which could be captured by means of artificial adaptive systems) is completely lost.

A connection scheme able to hypothesize links among variables, i.e. minimum spanning tree (MST) algorithm, as described by Kruskal (1956), could increase the information obtained by the map. The Kruskal MST algorithm of graph theory finds a minimum spanning tree for a connected weighted graph. MST method finds a subset of the edges that form a tree that includes every vertex, where the total weight of all the edges in the tree is minimized. This function has been recently applied in the medical field, especially in biology and medical imaging. However, the MST algorithm is still rare in medical clinics (Frimmel et al., 2004; Lee et al., 2006).

Here, we describe a new paradigm of variables mapping able to create a semantic connectivity map in which: (a) non-linear associations are preserved; (b) there are explicit connections schemes; (c) the complex dynamics of adaptive interactions is captured.

Data recorded during the 5-year follow-up from The Conselice Study participants were elaborated in relation to three different clinical endpoints: no cognitive decline, CIND and dementia. Three major biological hubs connecting variables with the three different cognitive conditions were identified in hydroxyl-methyl-gutaryl-CoA reductase enzyme (HMGCR), plasma cholesterol levels and age.

Biological hubs of variables are detected by the analysis. Related dependent variables converge to these hubs, that in turn may be considered as relevant biological variables in the connectivity map.

Several gene variants of different inflammatory genes and their cognate phenotypic factors showed a variable degree of relevance to brain aging and development of dementia. This is the first attempt to describe an integrated approach illustrating 35 variables in association with the risk of developing cognitive impairment and dementia in the elderly. The identification of biological hubs suggests possible patterns of pharmacological and non-pharmacological intervention with preventive potential against cognitive impairment.