Genes Related with Alzheimer's Disease: A Comparison of Evolutionary Search, Statistical and Integer Programming Approaches

Pablo Moscato¹, Regina Berretta¹, Mou'ath Hourani¹, Alexandre Mendes¹, and Carlos Cotta²

Newcastle Bioinformatics Initiative School of Electrical Engineering and Computer Science The University of Newcastle, Callaghan, NSW, 2308, Australia

² Dept. Lenguajes y Ciencias de la Computación, University of Málaga ETSI Informática, Campus de Teatinos, 29071, Málaga, Spain

Contact email: moscato@cs.newcastle.edu.au

Abstract. Three different methodologies have been applied to microarray data from brains of Alzheimer diagnosed patients and healthy patients taken as control. A clear pattern of differential gene expression results which can be regarded as a molecular signature of the disease. The results show the complementarity of the different methodologies, suggesting that a unified approach may help to uncover complex genetic risk factors not currently discovered with a single method. We also compare the set of genes in these differential patterns with those already reported in the literature.

1 Introduction

Alzheimer's disease affects ten percent of the population aged over 65 and that nearly half of all individuals aged above 85 will experience its effects. Since the first clear onset of the symptoms, a sufferer of Alzheimer's would an average life span of eight years and occasionally up to two decades. Since this disease is a progressive brain disorder that affects the patient ability to learn new things, to make judgements, and to accomplish simple daily activities, Alzheimer also has an emotional impact on families. This is aggravated by the fact that the sufferers may have an increasingly complex pattern of changes in personality and behavior as well as increasing anxiety, suspiciousness, agitation and forms of delusions or hallucinations. Memory loss is generally the first and worst symptom, but others include difficulty performing familiar tasks, communication, temporal/spatial disorientation, poor or decreased judgment, difficulty with task that require abstract thinking, misplacing objects, and passivity and loss of initiative ³.

After a certain point, people with Alzheimer's generally require 24-hour care. Estimations of total cost abound but differ very little, with the average lifetime

http://www.alz.org/Resources/FactSheets.asp

cost of care per patient around US\$200,000. Alzheimer's is the third most expensive disease in the United States, only second to heart disease and cancer, with the latter being a generic name for many different disease forms. According to figures provided by the Alzheimer's Association, the cost of 24-hours care plus diagnosis, treatment, and paid care costs, is estimated to be US\$100 billion annually. In the US, only a small percentage is covered, in almost equal parts by the federal government and the states (US\$8.5 billion total), the rest is paid by both patients and their families, putting a severe stress, that apart of the emotional cost, may also compromise their finances. With an increasingly aging population in the developed world, there is a need for more research on the causes of Alzheimer's disease. While a skilled physician would be able to diagnose it with 90 percent accuracy, an early genetic diagnosis of risks would help enormously. Quoting Zaven Khachaturian, "If we can push back the onset of Alzheimer's for just five years, we can reduce by 50 percent the number of people who get the disease, add years of independent functioning to people's lives, reduce the amount of care they need, and save this country billions of dollars in healthcare costs." ⁴ Accordingly, more research on the genetic basis of this disease is needed as demographically the picture is not good for developed countries, with a large number of their aging individuals getting the disease thus impacting on society as a whole.

We have conducted an extensive search in the scientific literature to try to identify which genes have already been reported as linked to Alzheimer's disease, resulting in a set of 95 genes. For 29 of them, we have found that there is a microarray data study in the public domain that contains their gene expression in control and Alzheimer's affected brains. A visual inspection of the relative gene expression of this dataset (containing approximately 2,000 genes), clustered with our memetic algorithm [1], clearly shows a pattern of differentiated gene expression in healthy and Alzheimer's affected brains (see Fig. 1).

From the set of 29 genes (out of those 95 identified as somewhat related), seven have already been reported in [2] (COX7B, IDI1, MAPK10, PRKCB1, RARS, SMS, WASF1 and YWHAH). The others are: ABCB1 [3], ADAM10 [4], ATOX1 [5], ATP [6], BCL2L2 [7], BRD2 [8], CRH [9], CTCF [10], GSK3B [11], HFETF [12], HTR2A [13], LAMC1 [14], NCSTN [15], NRG1 [16], NUMB [17], PRDX2 [18], PRDX5 [19], PRKR [20], PSEN1 [21], MAPK14 [22], and VSNL1 [23]. The two in boldface have been found by the methods we will present in this paper and were not reported in [2]. It is clear that a differential pattern of expression exists for these genes between Normal and AD brains, as shown in Fig. 1(b). In this paper we present a set of 70 genes which also show correlated patterns that may be useful to understand the genetic risk factors of the disease.

2 Modeling the Gene Subset Selection Problem

In order to model the problem of finding gene subsets of interest, we will proceed in two steps. Firstly, we will introduce the $(\alpha, \beta) - k$ -FEATURE SET Problem,

⁴ http://www.fda.gov/fdac/features/1998/398_alz.html

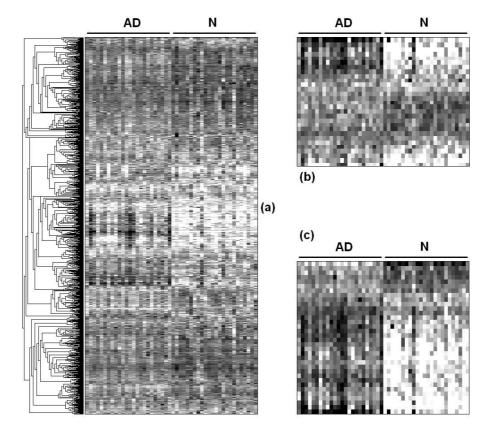


Fig. 1. (a) Gene expression of 2,100 genes in both Alzheimer (AD) and normal (N) brains (dataset from [2]). The columns correspond to different voxels (as described in [2]). It is clear that there exist a relatively large number of genes which are differentially expressed within different regions in AD and N brains. (b) Gene expression of the 29 genes found in our literature search (from a total of 95 we identified as possibly related) which are also present in the dataset. (c) The 34 genes highlighted in Table 3 of Ref [2]. These 34 genes are those of the union of the four subsets (I, II, III, and V) from Fig. 2 and have been obtained using a singular-value decomposition approach [2]. In all figures, we provide a high-quality clustering of the gene expression patterns using the memetic algorithm described in [1].

so as to provide a combinatorial setting for the target problem. Then we will address how to discretize microarray measurements to obtain a problem instance of the $(\alpha, \beta) - k$ -FEATURE SET Problem.

2.1 The $(\alpha, \beta) - k$ -Feature Set Problem

The $(\alpha, \beta) - k$ -Feature Set Problem is a generalization of k-Feature Set and it has been introduced with the aim of obtaining subsets of features of

robust discriminatory power [24]. Its use coupled with standard data mining algorithms has led us to successfully predict the outcome of the 2004 US Presidential election, two months in advance of the actual voting, only based on historical information from previous elections [25]. The problem can be formally defined as follows:

- **Instance**: A set of m examples $X = \{x^{(1)}, \ldots, x^{(m)}\}$, such that for all i, $x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}, t^{(i)}\} \in \{0, 1\}^{n+1}$, and three integers k > 0, and $\alpha, \beta \geqslant 0$.
- **Question:** Does there exist an (α, β) k-feature set $S, S \subseteq \{1, \dots, n\}$, with $|S| \leq k$ and such that:

 - for all pairs of examples $i \neq j$, if $t^{(i)} \neq t^{(j)}$ there exists $S' \subseteq S$ such that $|S'| \geqslant \alpha$ and for all $l \in S'$ $x_l^{(i)} \neq x_l^{(j)}$?
 for all pairs of examples $i \neq j$, if $t^{(i)} = t^{(j)}$ there exists $S' \subseteq S$ such that $|S'| \geqslant \beta$ and for all $l \in S'$ $x_l^{(i)} = x_l^{(j)}$?

We remark that the set S' is not fixed for all pairs of examples, but it is a function of the pair of examples chosen, so in the definition we mean S' = S'(i, j). Obviously, the problem is NP-hard as it contains the k-Feature Set Problem as special case [26]. Furthermore, the $(\alpha, \beta) - k$ -Feature Set problem is not likely to be fixed-parameter tractable for parameter k as Cotta and Moscato have recently proved that the k-Feature Set Problem is W[2]-complete [27].

We mentioned before that robustness is the goal. Indeed, robust feature identification methods are essential since microarray data measurements are notoriously prone to errors. This robustness comes at a price though. When this problem is used as a modelling tool for pattern recognition, robustness comes with redundancy in the number of features required for discrimination of a pair of examples. This may appear as counter-intuitive at first sight. A large number of approaches in data mining, and particularly in Bioinformatics, are concerned with finding "minimal" cardinality solutions. In the area of microarray data analysis, however, the true requirement is different. A small number of examples, as compared with the number of features, means that by just random chance a certain feature could dichotomize a set of examples. This said, the problem is how to preserve in our solutions a potentially useful set of features that could explain the examples, since they could be left aside due to the requirements of finding a minimal cardinality solution. Given a set of measurements obtained by means of a microarray experiment on m samples/conditions, the 0-1 values for each one of the features would correspond to under- or over-expressed genes respectively after a threshold value is determined.

2.2 Threshold Selection Issues

An instance of the (α, β) – k–Feature Set problem can be obtained once thresholds for discretizing gene-expression values have been set. The associated problem (finding appropriate thresholds given the particular values of α , β and k in the instance sought) can be formalized as follows:

(α, β) -Threshold Selection

- **Input**: A $m \times n$ \mathbb{R} -matrix \tilde{X} , class identifiers $t^{(i)} \in \mathbb{N}$ for every row i, $1 \leq i \leq m$, and two integers $\alpha, \beta \geq 0$.
- **Question**: Does there exist an array of m thresholds $\theta_1, \dots, \theta_{\tilde{m}}$ (i.e., one for each of the rows in \tilde{X}) such that each entry in the ith row of \tilde{X} greater than the θ_i is given the value 1, and 0 otherwise, and such that
 - 1. $\forall i, j, t^{(i)} \neq t^{(j)}$, the number of columns where $\tilde{X}_l^{(i)} \neq \tilde{X}_l^{(j)}$ (disagree) is at least α , and
 - 2. $\forall i, j, t^{(i)} = t^{(j)}$, the number of columns where $\tilde{X}_l^{(i)} = \tilde{X}_l^{(j)}$ (agree) is at least β ?

We note that this is a necessary but not a sufficient condition to create a yes-instance of the $(\alpha, \beta) - k$ -FEATURE SELECTION problem. Unfortunately, it is unlikely that an efficient algorithm for (α, β) -THRESHOLD SELECTION would be found as it is NP-complete [24]. However, Cotta, Sloper and Moscato have shown that evolutionary search strategies may help in practice to find thresholds allowing $(\alpha, \beta) - k$ -feature sets to be found in microarray data in lymphomas, opening the possibility of using the methodology in other domains as well.

3 Methodologies for Gene Subset Selection

We will now present three different, complementary methodologies for gene subset selection.

3.1 The Statistical Approach and the Microarray Dataset

The gene expression dataset is obtained from samples of normal and Alzheimer-affected diseased humans (for the complete description see [2]). Samples are obtained from spatially registered voxels (cubes) which produce multiple volumetric maps of gene expression. The technique is analogous to the reconstructed images obtained in biomedical imaging systems. A total of 24 voxel images of coronal hemisections at the level of the hippocampus of both the normal human brain and Alzheimer's disease affected brain were acquired for 2,100 genes. The statistical methods involve the use of a standard singular value decomposition (SVD) analysis. They show the most strongly differentially expressed genes between Alzheimer's affected and normal brains (having p-values ≤ 0.05). They show, notably, that the SVD results allow to produce images which correlate well with the neuroanatomy, including cortex, caudate, and hippocampus. This suggests that this technique will be a useful approach for understanding how the genome, and gene expression, constructs and regulates the brain.

3.2 The Evolutionary Search Approach

The evolutionary method used is similar to the one described in [24]. Therein, the authors present results of an evolution strategy that allowed to find, on

a microarray dataset of two different types of diffuse large B-cell lymphoma (each one containing 4 samples, and gene-expression profiles for 2,984 genes), an $\alpha = \beta = k = 100$ feature set. For $\alpha = \beta = 200$ and $\alpha = \beta = 300$, the ES found gene subsets of 227.3 and 360.5 genes on average respectively, 25% smaller gene subsets that those provided by a greedy heuristic.

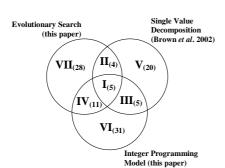
In this case, we have utilized a (1,10)-ES with binary tournament selection, gaussian mutation with independent self-adaptive stepsizes for each variable, and no recombination. For each candidate set of thresholds, the algorithm generates an $(\alpha, \beta) - k$ -Feature Selection problem instance, and uses a combination of kernelization techniques and greedy heuristics to solve it. The particular dataset we have considered seems to be difficult in practice for this algorithm, due to the fact that the size of this underlying $(\alpha, \beta) - k$ -Feature Selection problem that is being continuously generated and solved scales quadratically with the number of columns. Nevertheless, we have been able to identify several (10, 10) - k-feature sets with $17 \le k \le 19$ (see Fig. 3(c)).

3.3 The Integer Programming Approach

For the initial exploratory tests, we have used an standard integer programming (IP) formulation of the $(\alpha, \beta) - k$ -FEATURE SET problem as described in [28]. For some values of α and β we have been able to solve the instances to optimality using the CPLEX 9.0 mathematical software package. Treating all voxels' samples in the Alzheimer and control brains as 48 examples of two different classes (24 from each), allows us to find groups of genes differentially expressed in all regions of the brain (and expressed consistently within a class due to the large values of $\beta > 0$ obtained). However, we expect a degree of gene expression variation within different parts of a brain (both AD and control) to be present (due to normal functional differentiation). As a consequence, we adapted the IP formulation to look at the problem from a different perspective.

The aim of our new IP approach is now to find genes that are diffentially expressed in the same voxel in both the AD and control brains. This said, the number of pairs of examples corresponding to different classes drops from 24^2 to just 24 and the number of pairs of examples that belong to the same class drops to zero. This said, the IP model reflects our aim to find $(\alpha) - k$ -feature sets with large values of α and small values of k (the parameter β makes no sense here as we are treating any individual example as a member of a different class). This allows us as to find minimal sets of genes that are differentially expressed in both the AD and control brains in the same voxel.

In addition, in [28] good results were obtained by using an IP model in which we fix the number of features required to be in the solution with the objective of finding those features that maximize the "coverage". The coverage represents the sum of the cardinalities of all the sets S'(i,j) as defined in Sect. 2.1. Due to our good experience with this model, we fixed k=52 and we have been able to find a feature set with $\alpha=40$ and maximum coverage (998). The thresholds were fixed in this case, unlike the ES methodology, at the median value of the expression of each individual gene on the 48 samples.



I – DNCI1, KIAA0069, LOC51628, NR1I3, TAF2F

II - IDI1, MAPK10, WASF1, RAP2A

III - ICAP-1A, FOXJ3, KIAA0992, LOC51235, YWHAH

IV - HAX1, LOC54460, PRDX5 + 8 ESTs (Clone IDs 377827, 395436, 669471, 858450, 884653, 1032362, 1161775, 1500241)

V – BICD1, CCS, COX7B, DRAP1, DSCR1L1, IDH3A, LIMS1, NFATC3, PRKCB1, PSCA, PSCD2, PTPRN2, RAB2, RARS, SALL2, SEPW1, SMS, TIF1, XPO1. ZNF142

VI – ADD1, ATP6F, CANPX, CYBA, EIF2C1, EIF4B, FLJ11132, FLJ11200, GLGI, GNG10, INSL4, KIAA0154, KIAA0668, MAPK14, MCF2, NFIX, THBS1, TPD52, USP16 + 12 ESTs (Clone IDs 246116, 308788, 768324, 824479, 867751, 868188, 1034472, 1291971, 1292501, 1292893, 1499181, 1505783)

VII – APOC4, ATP5G3, FLJ11220, FLJ12895, FLJ20323, GAPDH, IL11RA, KIAA0308, LOC153561, NFKBIB, PPP2RIA, S100A11, SIAHI, SLC2A5, SLC9A6, SMAP, SRI, SRP46, Z39IG + 9 ESTs (Clone IDs 48906, 147192, 462944, 469379, 796548, 813813, 126858, 1493137, 1505240)

Fig. 2. A Venn diagram helps us to present the results of our comparison. We have uncovered a total of 70 genes not reported in [2] obtained from the solution of two different methodolgies, one is a variant of the evolutionary search methodology presented in [24] and the other is a truncated complete anytime algorithm based on the integer programming model discussed in Sect. 3.3 (with $\alpha=40$ and 52 genes) based on the method presented in [28] (union of the subsets IV, VI, and VII in the diagram). Our solutions also contained 14 genes already reported in [2] (subsets I, II and III). Ten genes from these sets, marked in boldface, have been also linked to Alzheimer's and neurodegenerative diseases (references are provided in Sect. 1)

4 Results

The main results are described with the aid of Fig. 2 and Fig. 3 and their accompanying captions. Another result worth mentioning is that the combined use of these three methodologies has uncovered that, in the union of all the genes (see Fig. 3) that provide a clear pattern of differential gene expression, there exists a peculiarity in the gene expression of area D2 of the normal brain (following the grid labeling used of [2]) ⁵. The gene expression pattern (for this subset of the genes) for region D2 in normals seems to be highly similar to the pattern of activities for other regions of the Alzheimer's brain. Puzzled by this fact, we conducted an experiment where we performed a hierarchical clustering of the columns as well as genes and indeed the pattern of activities for region D2 in normals was clustered together to those of Alzheimer's, though it appeared as an outgroup. This is intriguing, as this differential pattern of activity for this voxel is also clear in Fig. 1(b) and Fig. 1(c) where a distinctive dark column clearly stands out within a pattern of under expressed genes in the normal brain. All the methods revealed a similar characteristic and it could be visually appreciated even with our clustering of the entire set of 2,100 genes (Fig. 1(a)). While more research is needed, we believe that new analysis on this area, as differential to other areas in the normal and AD brains, may help to provide a bridge between genomics, functional differentiation and disease.

 $^{^{5}\} http://labs.pharmacology.ucla.edu/smithlab/genome_research_data/voxelgrid2.htm$

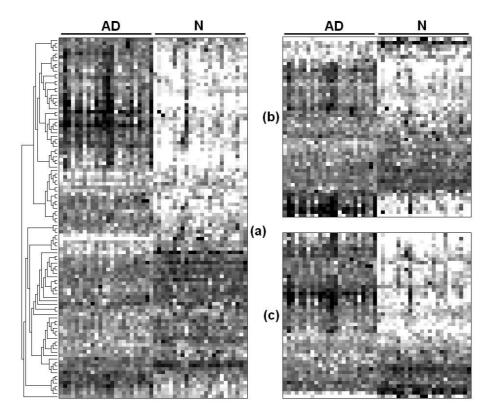


Fig. 3. (a) A clustering using the memetic algorithm of [1] of the union of all the genes reported in Fig. 2. (b) Clustering of the 52 genes found with the truncated exact search method based on the integer programming model (subsets I, III, IV, and VI of Fig. 2). (c) Clustering of the 48 genes found with the evolutionary search strategy proposed in this paper (subsets I, II, IV, and VII of Fig. 2). The evolutionary method searches for appropriate thresholds that will allow an $\alpha = \beta = 10$ -feature set to exist.

5 Conclusions

We have observed that the combination of results obtained in separately from evolutionary search, statistical and integer programming methods, allows the identification of a large number of genes differentially expressed in normal and Alzheimer's affected brain. The analysis has shown that there exist at least one hundred candidates for further exploration, mainly by correlating them with the activities of some of the genes in this study. This issue, as well as a combined methodology that would encompass the three methods so far applied to this dataset, plus a user-defined bias based on annotated information from biologists (and the biomedical literature) will be the subject of further studies.

References

- 1. Cotta, C., Mendes, A., Garcia, V., França, P., Moscato, P.: Applying memetic algorithms to the analysis of microarray data. In Raidl, G., et al., eds.: Applications of Evolutionary Computing. Volume 2611 of Lecture Notes in Computer Science. Springer-Verlag, Berlin (2003) 22–32
- 2. Brown, V., Ossadtchi, A., Khan, A., Cherry, S., Leahy, R., Smith, D.: High-throughput imaging of brain gene expression. Genome Research 12 (2002) 244–254
- 3. Vogelgesang, S., Cascorbi, I., Schroeder, E., Pahnke, J., Kroemer, H., Siegmund, W., Kunert-Keil, C., Walker, L., Warzok, R.: Deposition of Alzheimer's beta-amyloid is inversely correlated with P-glycoprotein expression in the brains of elderly non-demented humans. Pharmacogenetics 12 (2002) 535–541
- Colciaghi, F., Borroni, B., Pastorino, L., Marcello, E., Zimmermann, M., Cattabeni, F., Padovani, A., Di Luca, M.: Alpha-secretase ADAM10 as well as alpha-APPs is reduced in platelets and CSF of Alzheimer disease patients. Journal of Molecular Medicine 8 (2002) 67–74
- Bellingham, S., Lahiri, D., Maloney, B., La Fontaine, S., Multhaup, G., Camakaris, J.: Copper depletion down-regulates expression of the Alzheimer's disease amyloidbeta precursor protein gene. Journal of Biological Chemistry 279 (2004) 20378– 20386
- Sergeant, N., Wattez, A., Galvan-Valencia, M., Ghestem, A., David, J., Lemoine, J., Sautiere, P., Dachary, J., Mazat, J., Michalski, J., Velours, J., Mena-Lopez, R., Delacourte, A.: Association of ATP synthase alpha-chain with neurofibrillary degeneration in Alzheimer's disease. Molecular Brain Research 7 (2003) 293–303
- Zhu, X., Wang, Y., Ogawa, O., Lee, H., Raina, A., Siedlak, S., Harris, P., Fujioka, H., Shimohama, S., Tabaton, M., Atwood, C., Petersen, R., Perry, G., Smith, M.: Neuroprotective properties of BCL-W in Alzheimer disease. Journal of Neurochemistry 89 (2004) 1233–1240
- 8. Johnson, N., Bell, P., Jonovska, V., Budge, M., Sim, E.: NAT gene polymorphisms and susceptibility to Alzheimer's disease: identification of a novel NAT1 allelic variant. BMC Medical Genetics 5 (2004) 6
- 9. Rehman, H.: Role of CRH in the pathogenesis of dementia of Alzheimer's type and other dementias. Current Opinion in Investigational Drugs 3 (2002) 1637–1642
- Burton, T., Liang, B., Dibrov, A., Amara, F.: Transforming growth factor-betainduced transcription of the Alzheimer beta-amyloid precursor protein gene involves interaction between the CTCF-complex and Smads. Biochemical and Biophysical Research Communications 295 (2002) 713–723
- 11. Bhat, R., Budd, S.: GSK3beta signalling: casting a wide net in Alzheimer's disease. Neurosignals 11 (2002) 251–261
- Robson, K., Lehmann, D., Wimhurst, V., Livesey, K., Combrinck, M., Merryweather-Clarke, A., Warden, D., Smith, A.: Synergy between the C2 allele of transferrin and the C282Y allele of the haemochromatosis gene (HFE) as risk factors for developing Alzheimer's disease. Journal of Medical Genetics 41 (2004) 261–265
- Assal, F., Alarcon, M., Solomon, E., Masterman, D., Geschwind, D., Cummings, J.: Association of the serotonin transporter and receptor gene polymorphisms in neuropsychiatric symptoms in Alzheimer disease. Archives of Neurology 61 (2004) 1249–1253
- Palu, E., Liesi, P.: Differential distribution of laminins in Alzheimer disease and normal human brain tissue. Journal of Neuroscience Research 69 (2002) 243–256

- 15. Dermaut, B., Theuns, J., Sleegers, K., Hasegawa, H., Van den Broeck, M., Vennekens, K., Corsmit, E., George-Hyslop, P., Cruts, M., Van Duijn, C., Van Broeckhoven, C.: The gene encoding nicastrin, a major gamma-secretase component, modifies risk for familial early-onset Alzheimer disease in a Dutch population-based sample. The American Journal of Human Genetics 70 (2002) 1568–1574
- 16. Chaudhury, A., Gerecke, K., Wyss, J., Morgan, D., Gordon, M., Carroll, S.: Neuregulin-1 and erbB4 immunoreactivity is associated with neuritic plaques in Alzheimer disease brain and in a transgenic model of Alzheimer disease. Journal of Neuropathology and Experimental Neurology **62** (2003) 42–54
- Chan, S., Pedersen, W., Zhu, H., Mattson, M.: NUMB modifies neuronal vulnerability to amyloid beta-peptide in an isoform-specific manner by a mechanism involving altered calcium homeostasis: implications for neuronal death in Alzheimer's disease. Neuromolecular Medicine 1 (2002) 55–67
- Kim, S., Fountoulakis, M., Cairns, N., Lubec, G.: Protein levels of human peroxiredoxin subtypes in brains of patients with Alzheimer's disease and down syndrome. Journal of Neural Transmission. Supplementum 61 (2001) 223–235
- Krapfenbauer, K., Engidawork, E., Cairns, N., Fountoulakis, M., Lubec, G.: Aberrant expression of peroxiredoxin subtypes in neurodegenerative disorders. Brain Research 967 (2003) 152–160
- Peel, A., Bredesen, D.: Activation of the cell stress kinase PKR in Alzheimer's disease and human amyloid precursor protein transgenic mice. Neurobiology of Disease 14 (2003) 52–62
- Goldman, J., Reed, B., Gearhart, R., Kramer, J., Miller, B.: Very early-onset familial Alzheimer's disease: a novel presentilin 1 mutation. International Journal of Geriatric Psychiatry 17 (2002) 649–651
- Johnson, G., Bailey, C.: The p38 MAP kinase signaling pathway in Alzheimer's disease. Experimental Neurology 183 (2003) 263–268
- 23. Schnurra, I., Bernstein, H., Riederer, P., Braunewell, K.: The neuronal calcium sensor protein VILIP-1 is associated with amyloid plaques and extracellular tangles in Alzheimer's disease and promotes cell death and tau phosphorylation in vitro: a link between calcium sensors and Alzheimer's disease? Neurobiology of Disease 8 (2001) 900–909
- 24. Cotta, C., Sloper, C., Moscato, P.: Evolutionary search of thresholds for robust feature set selection: Application to the analysis of microarray data. In Raidl, G., et al., eds.: Applications of Evolutionary Computing. Volume 3005 of Lecture Notes in Computer Science., Berlin, Springer-Verlag (2004) 21–30
- 25. Moscato, P., Mathieson, L., Mendes, A., Berretta, R.: The Electronic Primaries: Predicting the U.S. presidency using feature selection with safe data reduction. In: (ACSC 2005), Newcastle, Australia, Australian Computer Society (2005)
- 26. Davies, S., Russell, S.: *NP*-completeness of searches for smallest possible feature sets. In Greiner, R., Subramanian, D., eds.: AAAI Symposium on Intelligent Relevance, New Orleans, AAAI Press (1994) 41–43
- 27. Cotta, C., Moscato, P.: The k-Feature Set problem is W[2]-complete. Journal of Computer and Systems Science **67** (2003) 686–690
- Berretta, R., Mendes, A., Moscato, P.: Integer programming models and algorithms for molecular classification of cancer from microarray data. In: Proceedings of the 28th Australasian Computer Science Conference (ACSC 2005), Newcastle, Australia. (2005)