Available online at www.sciencedirect.com
ScienceDirect
Neural Networks 21 (2008) 266–275
Neural Networks
www.elsevier.com/locate/neunet

2008 Special issue

Integrating evolving brain–gene ontology and connectionist-based system for modeling and knowledge discovery☆

Nikola Kasabov, Vishal Jain, Lubica Benuskova*

Knowledge Engineering & Discovery Research Institute (KEDRI), Auckland University of Technology (AUT), 581-585 Great South Road, Penrose, 1135 Auckland, New Zealand

Received 27 July 2007; received in revised form 27 November 2007; accepted 11 December 2007

1. Functional levels in the brain and their modeling

The brain is a dynamic information processing system that evolves its structure and functionality in time through information processing at different levels — Fig. 1: the quantum, molecular, single neuron, ensemble of neurons, cognitive, and evolutionary levels. At a molecular level, RNA and protein molecules evolve in a cell and interact in a continuous way, based on the stored information in the DNA and on external factors, and affect the functioning of a cell (neuron). At the level of a neuron, the internal information processes and the external stimuli in their interplay cause the neuron to produce a signal that carries the information to be transferred to other neurons (Arbib, 2003; Freeman, 2000), which is a continuous, evolving process. At the level of neuronal ensembles, all neurons operate in "concert", defining the function of the ensemble through continuous learning (Cooper, Intrator, Blais, & Shouval, 2004). At the level of the whole brain, cognitive processes take place in a life-long learning mode and global information processes are manifested, such as consciousness (Arbib, 2003; Chalmers, 1996; Grossberg, 1982; Taylor, 1999). At the evolutionary level, population of individuals and species evolve through generations, changing the genetic DNA code for a better adaptation (Darwin, 1859).

A project, called The Blue Brain Project, marks the beginning of a study on how the brain works by building very large scale models of neural networks (http://bluebrainproject.epfl.ch/index.html). This endeavor follows a century of experimental "wet" neuroscience and development of many theoretical insights of how neurons and neural networks function (Arbib, 2003). The Blue Brain Project was launched by the Brain Mind Institute, EPFL, Switzerland and IBM, USA in May 2005. Scientists from both organizations will work together using the huge computational capacity of IBM's Blue Gene supercomputer to create a detailed model of the circuitry in the neocortex — the largest and the most complex part of the human brain. The neocortex constitutes about 85% of the human brain's total mass and is considered to be responsible for the cognitive functions of language, learning, memory and complex thought. The Blue Brain Project will also build models of other cortical and subcortical parts of the brain and models of sensory and motor organs. By expanding the project to model other areas of the brain, scientists hope to eventually build an accurate, computer-based model of the entire brain. The project is a massive undertaking because of the hundreds of thousands of parameters that need to be taken into account. EPFL's Brain and Mind Institute's world most comprehensive set of empirical data on the micro-architecture of the neocortex will be turned into a working three-dimensional model recreating the high-speed electro-chemical interactions of the brain's interior. This step will be aimed at moving towards genetic level simulations of the neocortical column. With the most advanced phases of Blue Gene and Blue Brain Project development, scientists are planning to go to the molecular level in order to link gene activity with electrical and biochemical activity in neurons. Establishing this link will allow predictions of the cognitive consequences of genetic disorders and allow reverse engineering of cognitive deficits to determine the genetic and molecular causes.

The information processes at each level from Fig. 1 are very complex and difficult to understand as they evolve all the time, but much more difficult to understand is the interaction between the different levels. It may be that understanding the interaction

☆ An abbreviated version of some portions of this article appeared in Kasabov, Jain, Gottgtroy, Benuskova, Wysoski et al. (2007) as part of the IJCNN 2007 Conference Proceedings, published under IEE copyright.

* Corresponding author. Tel.: +64 9 921 9343; fax: +64 9 921 9501.
E-mail address: lbenusko@aut.ac.nz (L. Benuskova).

0893-6080/$ - see front matter
doi:10.1016/j.neunet.2007.12.041

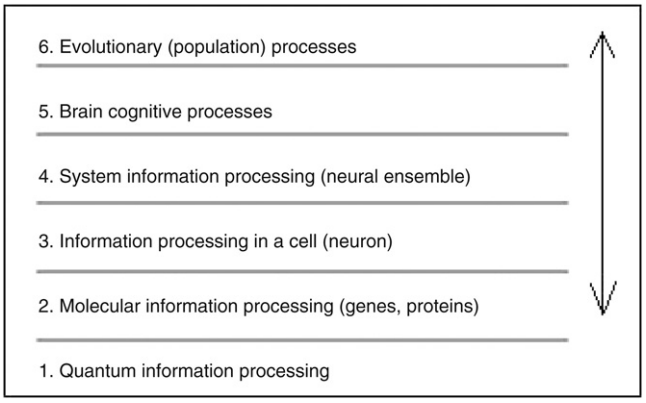

Fig. 1. Functional levels in the brain and the interaction between them.

through modeling would be a key to understanding each level of information processing in the brain and perhaps the brain as a whole. Using principles from different levels in one model and modeling their relationship can lead to a next generation of brain models as more powerful tools to understand the brain.

The Allen Brain Institute has completed a map of most of the genes expressed in different sections of the brain of a mouse and has published it free as the Allen Brain Atlas (http://www.alleninstitute.org). To integrate genetic, proteomic and brain activity data and to perform data analysis, modeling, prognosis and knowledge extraction that reveal the relationship between brain functions and genetic information, we need to build new global data and knowledge repositories and new mathematical and computational models. The explosion of biomedical data and the growing number of disparate data sources are exposing researchers to a new challenge — how to acquire, represent, maintain and share knowledge from large and distributed databases and how to use it for modeling and further discovery. This has led to the development of ontologies (Chandrasekaran, Josephson, & Benjamins, 1999). Here, however, we suggest an integration of evolving, globally shared brain–gene ontology and connectionist-based computational intelligence.

Ontology is a specification of the conceptualization of a knowledge domain. Ontologies range from taxonomies and classifications, database schemas, to fully axiomatized theories. Ontology captures the intrinsic conceptual structure of a domain. According to Gruber, the meaning of ontology in the context of computer science is the description of concepts and relationships that can exist for an agent or a community of agents (Gruber, 1993). By agent(s) we mean a database, software tool, or any computational system. Thus, ontology is a description (like a formal specification of a program) of the concepts and relationships between them to support the sharing and reuse of formally represented knowledge. In recent years, ontologies have been adopted in many business and scientific communities as a way to share, reuse and process domain knowledge (Fensel, 2004; Pisanelli, 2004).

For experimental purposes the disease ontology (http://diseaseontology.sourceforge.net/), biomedical ontology (http://www.bioontology.org/) and the Gene Ontology (GO) have been created (http://www.geneontology.org/). The GO project provides a controlled vocabulary to describe gene and gene product attributes in any organism addressing the need for consistent descriptions of gene products in different databases (Ashburner et al., 2000). The goal of a Biomedical Ontology is to allow scientists to create, disseminate, and manage biomedical information and knowledge in a machine-processable form for accessing and using this biomedical information in research. Disease Ontology is a controlled medical vocabulary designed to facilitate the mapping of diseases and associated conditions to particular medical codes such as ICD9, SNOMED and others.

The Semantic Web development of SenseLab of the Yale Center for Medical Informatics involves exporting data from NeuronDB, ModelDB, and BrainPharm to RDF and/or OWL format (http://neuroweb.med.yale.edu/senselab/). NeuronDB contains descriptions of anatomic locations, cell architecture and physiologic parameters of neurons linked to compartmental models. ModelDB is a large repository of computational neuroscience models and simulators. The mathematical models in ModelDB are annotated with references to NeuronDB. The BrainPharm database enhances descriptions in NeuronDB with descriptions of the actions of pathological and pharmacological agents and is intended to support research on drugs for the treatment of neurological disorders. The overall goal of the SenseLab project is to facilitate neuroscience data aggregation, integration, and reasoning using Semantic Web technologies.

The Brain–Gene Ontology (BGO) developed and presented here is focused on mammalian brain and has a broader scope than GO in a sense that we cover the gap in integration of knowledge that comes from different disciplinary domains such as neuroscience, bioinformatics, genetics, computer and information sciences (Kasabov, Jain, Gottgtroy, Benuskova, & Joseph, 2007) — Fig. 2. The BGO is an evolving system that is changing and developing with the addition of new facts and knowledge in it by multiple users. Linking selected structured bodies of physiological, genetic and computational information provides a pathway for different types of users. Designing an interface that enables users with different levels of expertise, specialization and motivation to access the BGO — either through a familiar or specialist approach or through a more general introduction is a critical issue.

In this paper we describe how the information is organized in the BGO system, the environment in which it is implemented, and how we can use the system to aid novel discoveries by means of computational intelligence and more specifically — evolving connectionist systems. We conclude with future directions for BGO development. Preliminary results were reported at IJCNN 2007 with the emphasis on computational neurogenetic modeling (CNGM) of genetic influence upon neural electrical activity and on teaching (Kasabov, Jain, Gottgtroy, Benuskova, Wysoski et al., 2007)

2. BGO structure, knowledge visualization and implementation

The overall system comprises three main parts: (a) brain organization and functions; (b) genes and gene regulatory

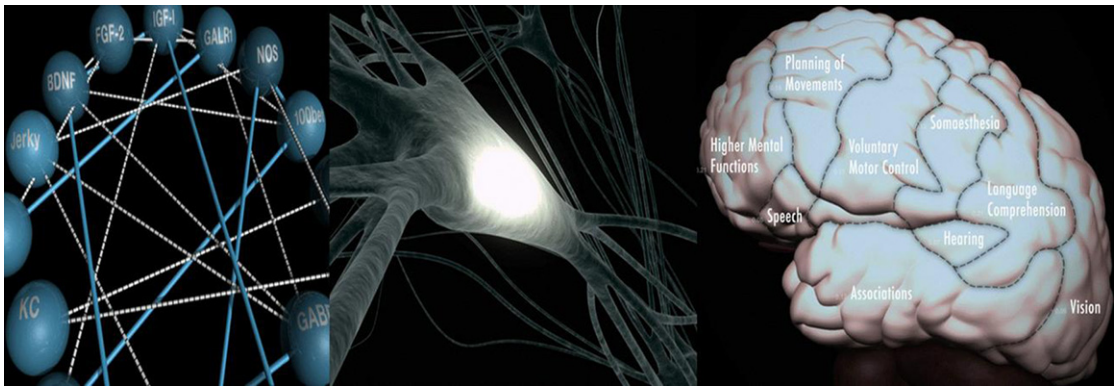

Fig. 2. The BGO is concerned with the accumulation and the use of data and knowledge for a better understanding and further discoveries of relationships between the brain, diseases and mammalian genes.

networks; and (c) a simulation module. The brain organization and function module contains information about neurons, their structure, process of spike generation and processes in synapses. It contains also the description and examples of high level brain data such as electroencephalogram (EEG) data for different brain states, e.g. for normal and epileptic states. The genes and gene regulatory networks (GRN) part is divided into sections on neurogenetic processing, gene expression regulation, protein synthesis and abstract GRNs. The third large part, the simulation module, has sections on computational neurogenetic modeling (CNGM), evolutionary computation, and evolving connectionist systems (ECOS). The CNGM methodology in general and particular case studies has been described in detail elsewhere, i.e. Benuskova, Jain, Wysoski, and Kasabov (2006), Benuskova and Kasabov (2007), Kasabov and Benuskova (2004), Kasabov, Jain, Gottgtroy, Benuskova, Wysoski et al. (2007).

The BGO represents existing concepts in the domain of brain and genes, their attributes and the relationships. It can be viewed as a declarative model. The first version of the BGO has been implemented in Protégé, which is open source ontology building environment developed by the Medical Informatics Department of the Stanford University (protege.stanford.edu). Protégé can be extended by way of a plug-in architecture and a Java-based Application Programming Interface (API) for building knowledge-based tools and applications.

We have developed a set of plug-ins that enable one to visualize, extract and import knowledge from/into different data sources and destinations. The information in the BGO is based on the two most used biological data sources, namely Gene Ontology, and Unified Medical Language System — UMLS, along with knowledge integrated from Entrez Gene, SwissProt, Interpro, Gene Ontology, Gene Expression Atlas, OPHID and others (Gottgtroy, Kasabov, & MacDonell, 2004). It also incorporates knowledge acquired from biology domain experts and from different literature databases such as PubMed.

Another feature of the BGO is the graphical presentation of relations by specific Protégé means (dynamic graphs, attached documents and pictures). There are many plug-ins which can be used to navigate, browse and visualize the information available within BGO. For example, as shown in Fig. 3, using OntoViz particular instances or appropriate classes can be selected and displayed in the form of a hierarchical graph. Few of the general domains are shown in this figure, out of many in the whole BGO.

As another example, using the plug-in called TGViz (touch graph visualization) we have explored the relationship of one gene, GABRA1, with several other molecules present in the BGO. Fig. 4 illustrates the detailed information available in the BGO about relations of GABRA1 with other genes, proteins, brain regions, molecular functions etc. The user can further navigate into each instance or class and their subsections down to the genetic level and use the relationship found for learning and research.

The BGO utilizes a novel evolving conceptual metadata structure which allows incorporating new discoveries and adapting its structure (Gottgtroy, Kasabov, & MacDonell, 2006). This evolving structure keeps track of change and provenance of source, date, among others. Thus, the BGO framework allows users to view in a complex evolving structure the hierarchical representation of relationships between genes, proteins, neurons and brain functions.

The data from the BGO can be used in simulation systems, such as computational neurogenetic simulation tool, CNGM (http://www.kedri.info), NeuCom http://www.theneucom.com), Siftware (http://www.peblnz.com), and Weka (http://www.cs.waikato.ac.nz/ml/weka/). NeuCom is a learning and reasoning computer environment based on connectionist models. It is designed to solve such problems as clustering, classification, prediction, adaptive control, data mining and pattern discovery from databases in a multidimensional, dynamic and changing data environment. Siftware is a software system for gene expression data analysis, modeling and profiling (http://www.aut.ac.nz/research/research_institutes/kedri/research_centres/centre_for_bioinformatics/siftware.htm). Weka is a collection of machine learning algorithms for data mining tasks. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. Results from these simulators can be added back to the BGO to update the BGO current knowledge base. Hence BGO evolves based on the knowledge input from outside and also based on creation of new knowledge by means of Computational Intelligence.

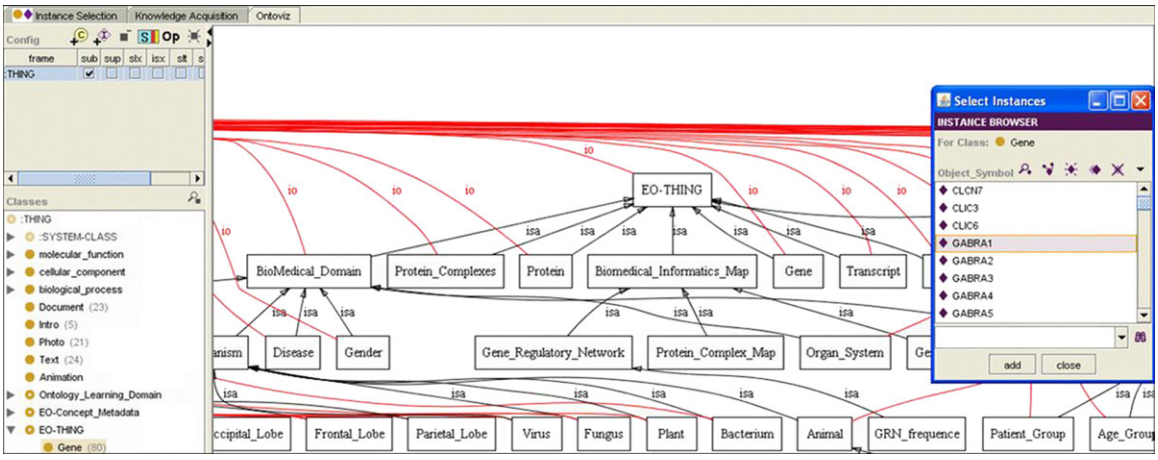

Fig. 3. BGO domain visualization using OntoViz.

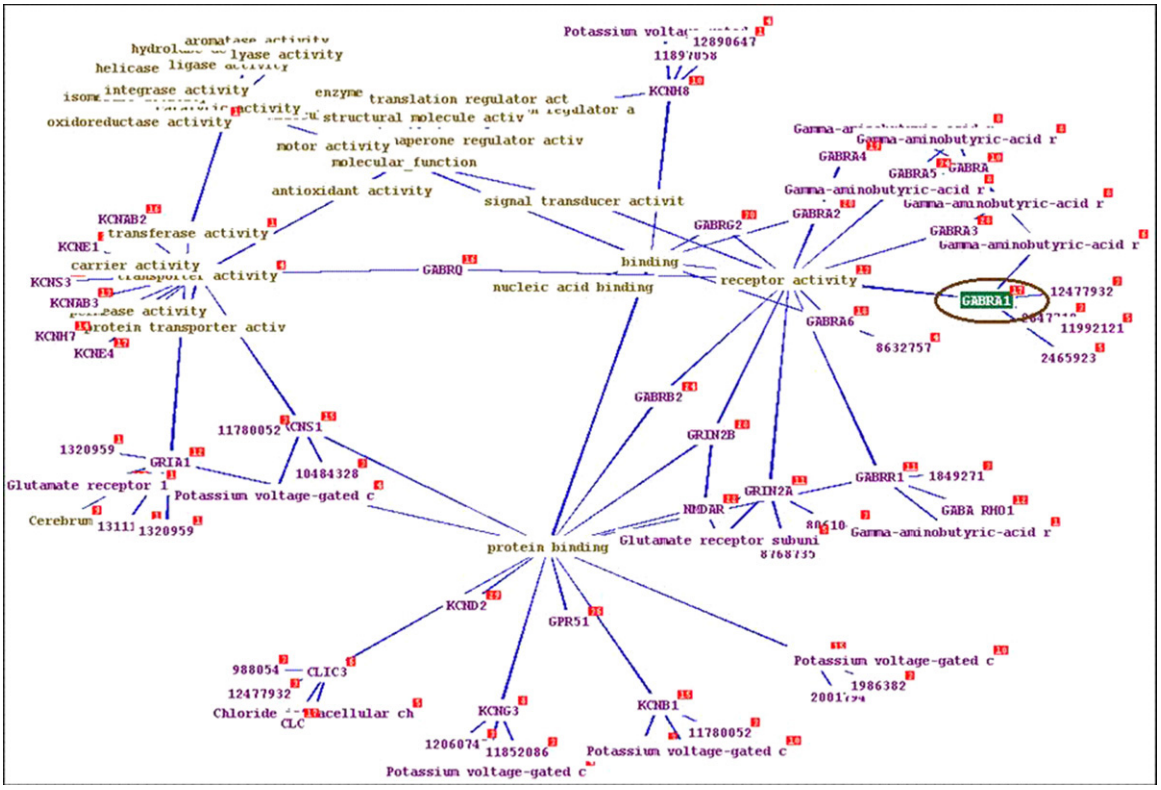

Fig. 4. GABRA1 relationship visualization in BGO using TGViz.

The current version of BGO (without animations) is downloadable at http://www.aut.ac.nz/research/research_institutes/kedri/research_centres/centre_for_neuroinformatics_and_brain_study/brain_gene.htm

In the next section we will use an evolving connectionist model to discover a set of relevant genes and the gene profiles of brain cancer survivals. The discovered genes will be entered back to the BGO to find their relationships with other genes and brain functions and diseases.

3. Brain–gene pattern discovery through evolving connectionist systems

The BGO system explained above provides conceptual links between data on brain functions and diseases, their genetic basis, experimental publications, graphical illustrations and the relationships between the concepts. Information items and their relationship are traceable through a query plug-in that allows, for example to answer questions such as – which genes are related to the occurrence of juvenile myoclonic epilepsy (JME)? – simply by typing the key word JME into the query window and selecting the class gene and the slot function comment — Fig. 5. The system returns the list of genes (here in this example 10 genes) potentially related to JME. By selecting any of them we can obtain detailed information about that particular gene, such as its GO function, chromosomal location, synonyms, brain expression profile, etc. Here we show the navigation for the GABRA1 gene. The window shows the detailed information available within BGO. Next we can

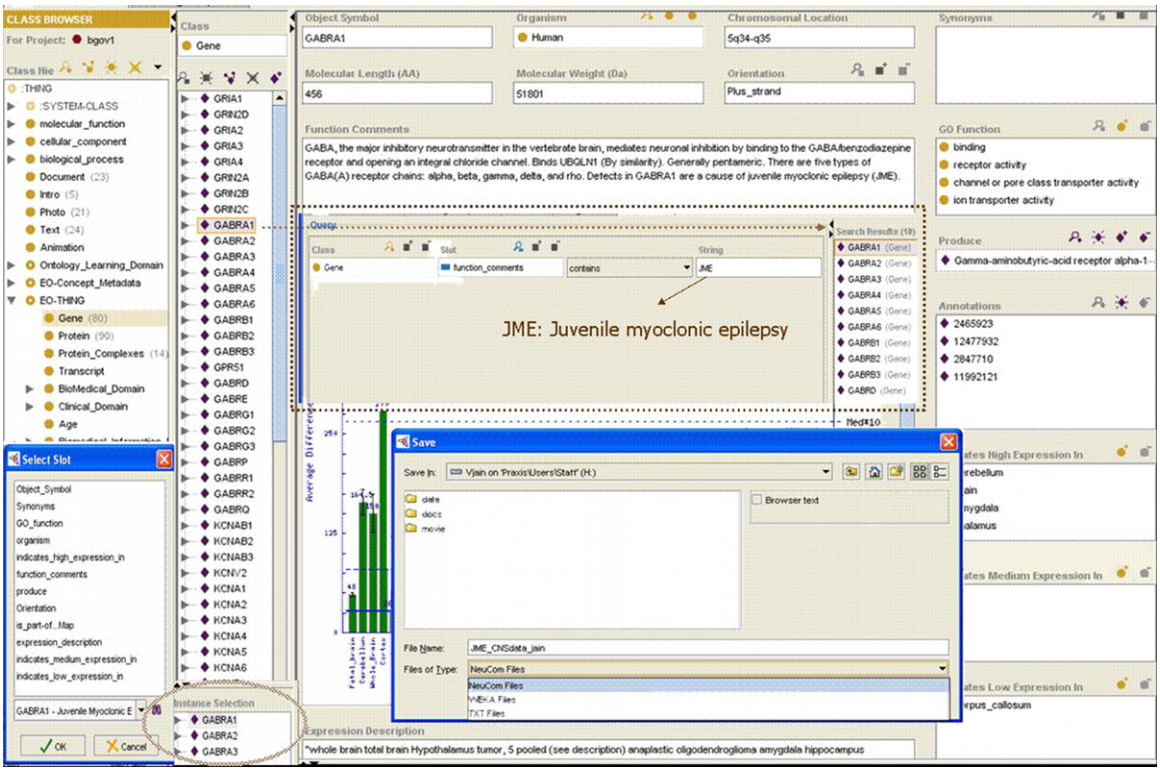

Fig. 5. Query search system looking for JME related genes; navigation of GABRA1 in BGO; gene instance selection window to export and save the data in the required format.

select gene(s) of interest to visualize their relationships to other concepts/instances in the BGO as illustrated in Fig. 4.

We have developed our own set of plug-ins that enable one to visualize, extract and import knowledge from/into different data sources and destinations (Gottgtroy et al., 2004, 2006). BGO thus allows users to select and export the specific data of their interest like chromosomal location or molecular sequence length, or expression patterns, which can then be analyzed in a software machine learning environment, such as WEKA and NeuCom to train prediction or classification models and to visualize relationship information. Such exported gene/protein data can also be analyzed in a different manner by standard bioinformatics software like BLAST and FASTA for revealing homology patterns for those genes/proteins of interest, etc.

One of the main applications of BGO is the integration between ontology and machine learning tools in relation to feature selection, classification and prognostic modeling with results incorporated back into the ontology. As an example, here we will take publicly available data, which is a gene expression data of 60 samples of CNS cancer (medulloblastoma) representing 39 children patients who survived the cancer after treatment, and 21 who did not respond to the treatment (Pomeroy, Tamayo, Gaasenbeek, & Sturla, 2002). Fig. 6 illustrates the selection of the top 10 genes out of 7129 genes, as numbered in the original publication, using a t-test method in a software environment Siftware.

Here is the list of the 10 selected genes (the first ID number is for reference of further analysis and the second ID number is the row number in the original data):

- G1 = G1352 = High mobility group protein (HMG-I(Y)) gene exons 1-8, L17131, high mobility group AT-hook 1, HMGA1
- G2 = G327 = D28124, NBL1- neuroblastoma, suppression of tumorigenicity 1
- G3 = G348 = Probable Ubiquitin Carboxyl-terminal Hydrolase, D29956 UBPY (ubiquitin specific peptidase 8, USP 8)
- G4 = G844 = Dynein, Heavy Chain, Cytoplasmic, HG2417-HT2513
- G5 = G2196 = Polyposis Locus Protein 1, M73547, adenomatosis polyposis coli, APC
- G6 = G2695 = TAR (HIV-1) RNA binding protein 2, U08998, TARBP2
- G7 = G3645 = Prostaglandin transporter hPGT mRNA, U70867
- G8 = G3320 = Leukotriene C4 synthase (LTC4S) gene, U50136
- G9 = G2496 = NTRK3 Neurotrophic tyrosine kinase, receptor, type 3 (TrkC), S76475 — (1 of 50 *markers of survival from* Pomeroy et al. (2002))
- G10 = G2996 = Gps2 (GPS2, G protein pathway suppressor 2) mRNA, U2896

Evolving Connectionist System (ECOS) can be used for building adaptive classification or prognostic systems and for extracting rules (profiles) that characterize data in local clusters. Evolving connectionist systems (ECOS) are modular connectionist-based systems that evolve their structure and functionality in a continuous, self-organized, on-line, adaptive,

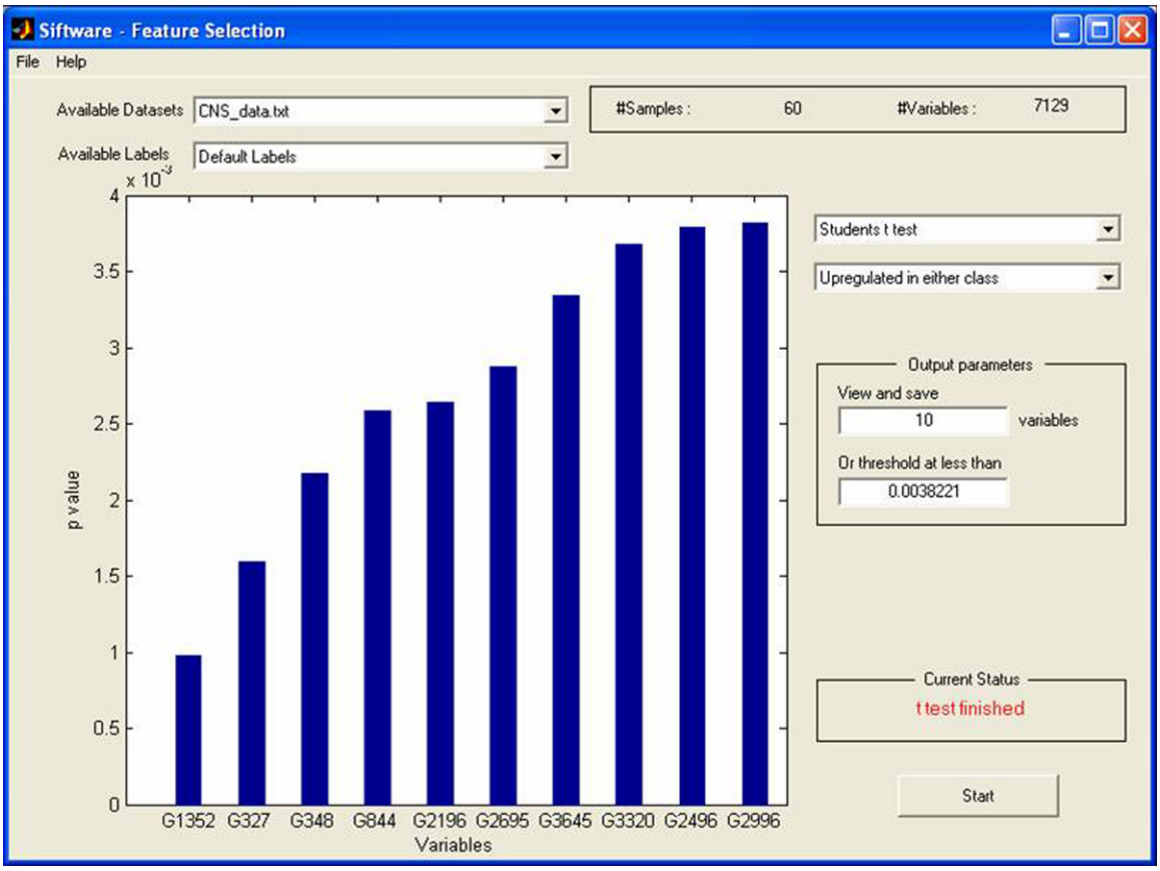

Fig. 6. Ten genes selected as top discriminating genes from the Central Nervous System (CNS) cancer data that discriminates two classes — survivals and not responding to treatment. The Siftware system is used for the analysis and the method is called t-test.

interactive way from incoming information; they can process both data and knowledge in a supervised and/or unsupervised way (Kasabov 2002b, Kasabov 2006). ECOS learn local models from data through clustering of the data and associating a local output function for each cluster. Clusters of data are created based on similarity between data samples either in the input space (this is the case in some of the ECOS models, e.g. the dynamic neuro-fuzzy inference system DENFIS (Kasabov and Song 2002), or in both the input space and the output space (this is the case in the Evolving Fuzzy Neural Network (EFuNN) models (Kasabov 2001). Samples that have a distance to an existing cluster center (rule node) N of less than a threshold $R_{\max}$ (for the EFuNN models it is also needed that the output vectors of these samples are different from the output value of this cluster center in not more than an error tolerance E) are allocated to the same cluster N_c. Samples that do not fit into existing clusters, form new clusters as they arrive in time. Cluster centers are continuously adjusted according to new data samples, and new clusters are created incrementally. The similarity between a sample and an existing rule node N can be measured in different ways, the most popular of them being the normalized Euclidean distance. ECOS learn from data and automatically create a local output function for each cluster, the function being represented in the connection weights, thus creating local models. Each model is represented as a local rule with an antecedent – the cluster area, and a consequent – the output function applied to data in this cluster. Implementations of the ECOS framework require connectionist models that support these principles. For a detailed theory of ECOS see Kasabov (2007, 2002).

Building adaptive classification and prognostic system for extracting rules (profiles) that characterize data in local clusters is illustrated in Figs. 7 and 8 on the 10 CNS genes from Fig. 6, where a classification system is evolved using the evolving classifier function method (ECF). Before the final classifier is evolved in Fig. 8, a leave-one-cross validation method is applied to validate the ECOS model on the 60 samples, where 60 models are created — each one on 59 samples, after one example is taken out, and then the model is validated to classify the taken out example. The average accuracy over all 60 examples is 85% as shown in Fig. 7. 51 samples are classified accurately, out of 60. Then an ECF classifier is evolved on the 10 CNS cancer genes from Fig. 6. Aggregated (across all clusters) general profiles for each of the two classes are shown in Fig. 8.

The selected smaller number of genes, out of thousands, can be further analyzed in terms of their relation to cellular processes or other types of cancer or other diseases. The results then can be imported back to BGO and conclusions can be made about the genetic differences between the two groups of patients. For instance after entering the information about the 10 selected genes from EntrezGene Database and combining it with the already present knowledge in BGO, we can discover that G1, the High mobility group protein (HMG-

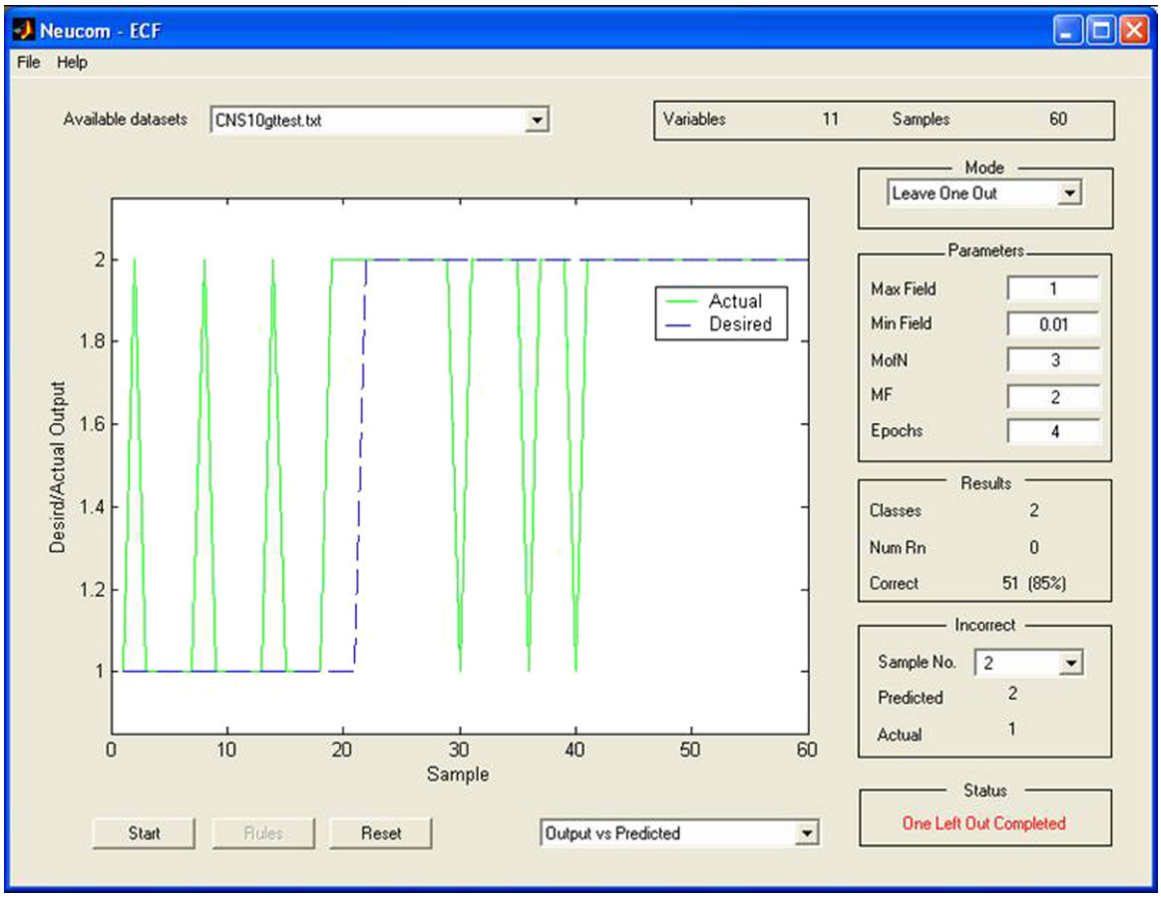

Fig. 7. A leave-one-cross validation method is applied to validate an ECF ECOS model on the 60 CNS cancer samples (Pomeroy et al., 2002), where 60 models are created — each one on 59 samples, after one example is taken out, and then the model is validated to classify the taken out example. The average accuracy over all 60 examples is 85%, where 51 samples are classified accurately and 9 incorrectly. Class 1 is the non-responding group (21 samples, 71.43% accuracy) and class 2 is the group of survivals (39 samples, 92.31%). The results are better than those achieved in Pomeroy et al. (2002), results of 78% (13 errors out of 60).

I(Y)), L17131, which is highly expressed in the treatment failures (see Fig. 8) encodes a non-histone protein involved in many cellular processes, including regulation of inducible gene transcription, integration of retroviruses into chromosomes, and the metastatic progression of cancer cells. Our analysis has revealed that over-expression of this gene is associated with a bad prognosis for medulloblastoma in connection with the over-expression of G6, TAR (HIV-1) RNA binding protein 2. The protein encoded by this gene activates HIV-1 gene expression in synergy with the viral Tat protein. Thus, maybe as our analysis points to, over-expression of this latter gene is related to a weaker immune response of an organism, which also makes sense from the point of view of failure to fight the disease. All other genes are under-expressed in the class of failures and relatively over-expressed in the class of survivors or at least not under-expressed. For instance, G2, NBL1-neuroblastoma, D28124, which is involved in suppression of tumorigenicity, is not under-expressed in the class of survivors, but is under-expressed in failures, which again makes sense in terms of an outcome prognosis. G3, the probable Ubiquitin Carboxyl-terminal Hydrolase, D29956 UBPY, labeling proteins for proteasomal degradation, is not under-expressed in the class of survivors, but is under-expressed in the class of failures. G4, Dynein, Heavy Chain, Cytoplasmic, HG2417-HT2513, which mediates the perinuclear aggregation of phagocytosed melanosomes, participates in the formation of the supranuclear melanin cap and serves as a mechanism to help protect the nucleus from ultraviolet-induced DNA damage, is over-expressed in the class of survivors meaning it might have a more general protective function not just against the UV light. G5, Polyposis Locus Protein 1, M73547, APC, adenomatosis polyposis coli, encodes a tumor suppressor protein that includes among its many intracellular functions one of nuclear export. Defects in this gene cause familial adenomatous polyposis (FAP), an autosomal dominant pre-malignant disease that usually progresses to malignancy. This gene is under-expressed in both classes reflecting the malignancy of medulloblastoma. G7, Prostaglandin transporter hPGT mRNA, U70867: so far only the role of PGT in the regulation of reproductive processes has been known. This study points to its role also in medulloblastoma, as one of the gene markers of survival, together with G4, Dynein, Heavy Chain, Cytoplasmic. The rest of the genes are under-expressed in failures and not under-expressed in survivals. G8, the Leukotriene C4 synthase (LTC4S) gene, U50136: This gene encodes an enzyme that catalyzes the first step in the biosynthesis of cysteinyl leukotrienes, potent biological compounds derived from arachidonic acid. Leukotrienes have

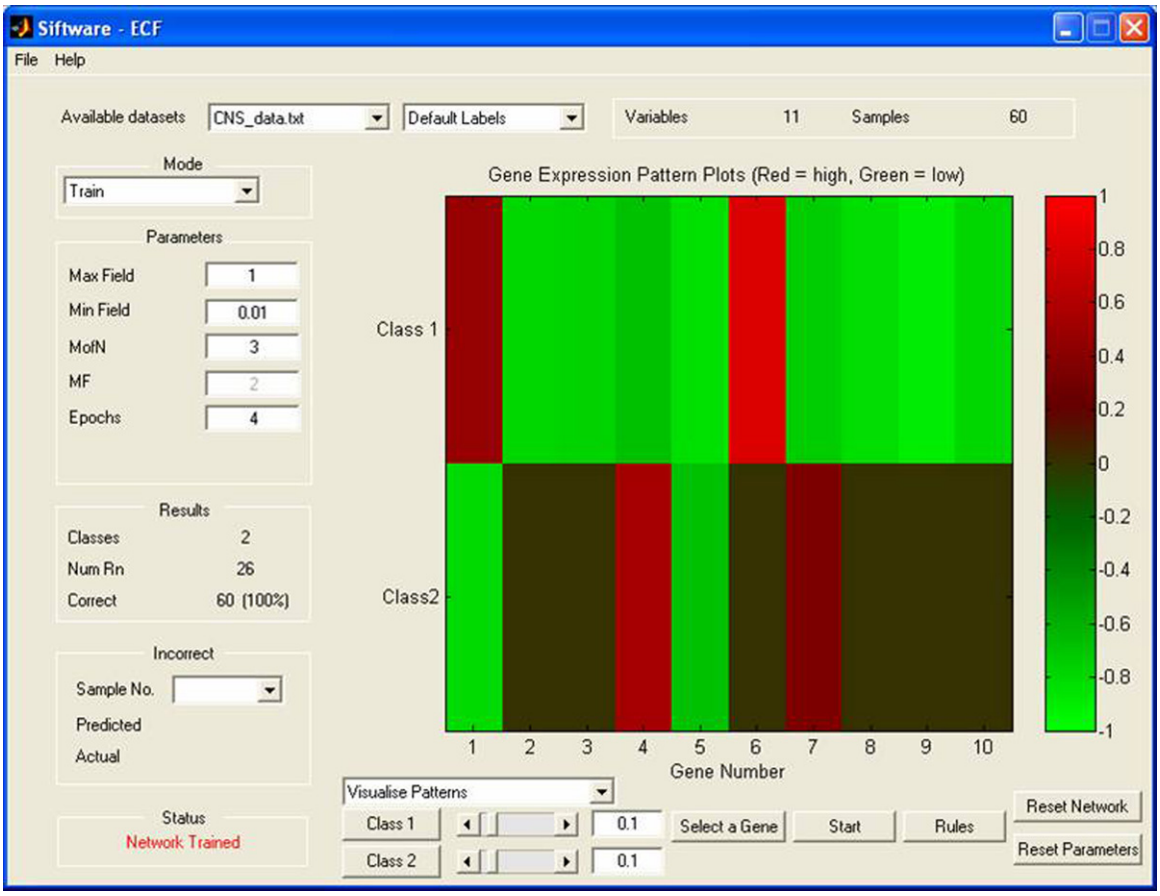

Fig. 8. An ECOS classifier is evolved on the 10 CNS cancer genes from Fig. 5. Aggregated (across all clusters) general profiles for each of the two classes are shown. Class 1 is the non-responding group (21 samples) and class 2 is the group of survivals (39 samples). The analysis is performed with the use of a proprietary software system SIFTWARE (http://www.peblnz.com).

been implicated as mediators of anaphylaxis and inflammatory conditions such as human bronchial asthma (under-expressed in failures). Mutations of G9 NTRK3/TrkC have been associated with secretory breast carcinomas and other cancers. Moreover, it plays a role in Long-Term Potentiation. It is under-expressed in failures. G10, Gps2 (GPS2, G protein pathway suppressor 2) mRNA, U2896, encodes a protein involved in G protein-mitogen-activated protein kinase (MAPK) signalling cascades. When over-expressed in mammalian cells, this gene could potently suppress a RAS- and MAPK-mediated signal and interfere with JNK (C-jun-amino-terminal kinase) activity, suggesting that the function of this gene may be signal repression. Ras proteins transmit extracellular signals that promote the growth, proliferation, differentiation and survival of cells. This G protein pathway suppressor 2 is under-expressed in the class of failures, which means signals leading to growth and proliferation are not suppressed efficiently, which can contribute to treatment failure. For instance anthrax toxin has been investigated as a therapeutic agent against cancer through inhibiting growth of RAS-transformed cells by regulation of MAPK (Ascenzi et al., 2002), thus indirectly confirming our analysis that genes involved in MAPK signalling pathways can influence the outcome of cancer.

Thus, by means of knowledge stored in BGO we can meaningfully interpret the results obtained by the CI analysis. In addition, for each of the genes, we can obtain the network of relations to other genes, gene functions, molecular processes and disease by means of Protégé TGVizTab as is illustrated in Fig. 9. We can dive into each node to obtain further information and relevant links.

4. Conclusion and future directions

This paper presents a brain–gene ontology that includes conceptual and factual information about brain and gene functions and their relationships. BGO can be viewed as a declarative model that defines and represents the concepts existing in the domain of brain and genes, their attributes and the relationships between them. It is represented as a knowledge base which is available to applications that need to use and/or share the knowledge of the domain. BGO is a tool for research and teaching across areas of bioinformatics, neuroinformatics, computer and information sciences at different levels of education and expertise.

BGO allows users to navigate through the rich information space, visualize relationship information and add new information as the BGO has an evolving structure. Various data can be used in a software machine learning environment, such as WEKA and NeuCom or Siftware to train prediction or classification models, in order to enter the results back to BGO for further knowledge discovery.

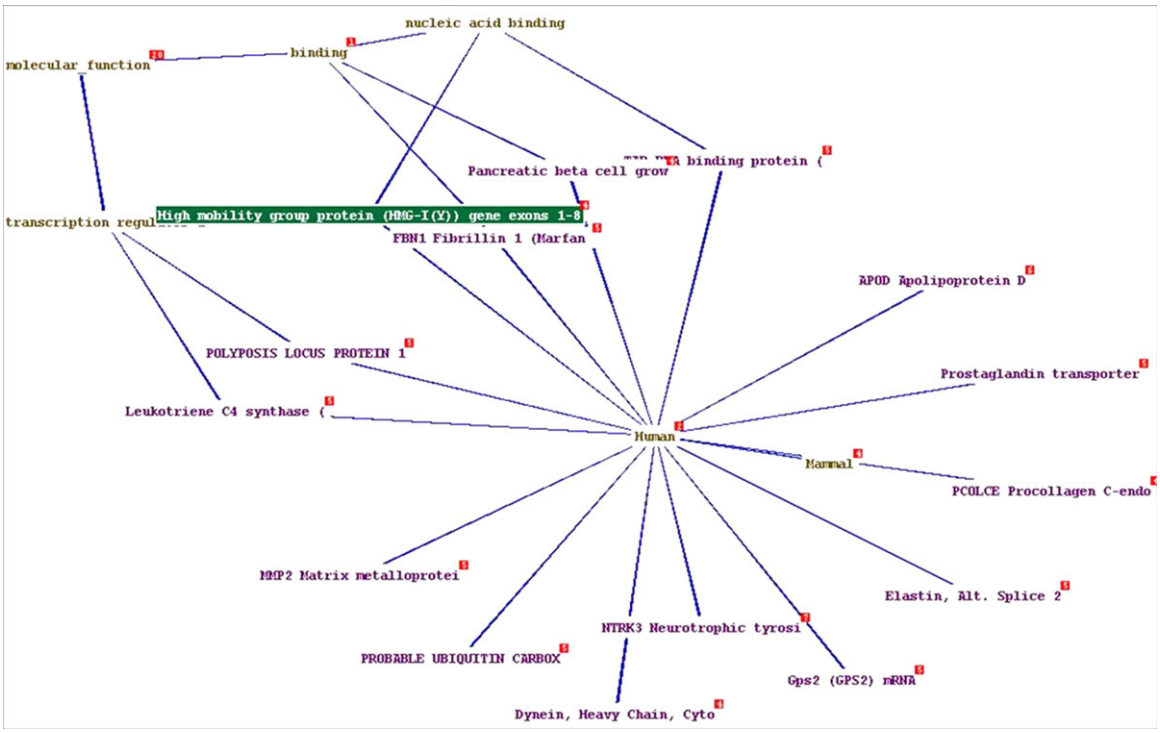

Fig. 9. Relationship visualization in BGO using TGViz for the High mobility group protein (HMG-I(Y)), L17131, one of the genes, which is differentially expressed between the group of survivors and treatment failures in medulloblastoma data.

Directions for further development are:

(1) To develop the BGO into a Web-based, multiple user, shared, open source environment;

(2) To develop new inference and knowledge discovery methods. Knowledge discovery (KD) has always been a critical aspect in ontology usage. Typical tasks for KD are the inference of new associations and relations between facts. Machine learning approaches for ontologies and suggested means like CLIPS, fuzzy CLIPS, Algernon and Jess, etc., can be applied in future.

(3) CNGM and hidden knowledge visualization. Results obtained by simulating the neurogenetic models will become new facts to be integrated within the BGO, which will close the loop of new knowledge discovery and representation in the ontology knowledge base. Visualization through clusters, trees, graphs, etc., using TGviz is under development.

The BGO will continue to be an evolving ontology that evolves its structure and content so that new information can be added in the form of molecular properties, disease related information and so on. All of this information can be re-utilized to create further models of brain functions and diseases that include models of gene interactions. We hope that by linking and integrating simulation results from the CNGM simulations with genetic information in the BGO, we can facilitate better understanding of metabolic pathways and modeling of gene regulatory networks, and ultimately a more complete understanding of the pathogenesis of brain diseases.

Acknowledgments

This work is supported by a NERF grant AUTX0201 and by KEDRI. We would like to thank Scott Heappey for creating all the animations, Simei Gomes Wysoski for programming the neurogenetic simulator, and Paulo C. M. Gottgtroy for initial work on BGO development. More information on the BGO is available at http://wwwwww.kedri.info.

References

Arbib, M. (Ed.) (2003). *The handbook of brain theory and neural networks* (2nd ed.). Cambridge, MA: MIT Press.

Ascenzi, P., Visca, P., Ippolito, G., Spallarossa, A., Bolognesi, M., & Montecucco, C. (2002). Anthrax toxin: A tripartite lethal combination. *FEBS Letters*, *531*(3), 384–388.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., & Cherry, J. M. (2000). Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nature Genetics*, *25*, 25–29.

Benuskova, L., Jain, V., Wysoski, S. G., & Kasabov, N. (2006). Computational neurogenetic modelling: A pathway to new discoveries in genetic neuroscience. *International Journal on Neural Systems*, *16*(3), 215–227.

Benuskova, L., & Kasabov, N. (2007). *Computational neurogenetic modeling*. New York: Springer.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them? *Intelligent Systems and Their Applications*, *14*(1), 20–26.

Cooper, L. N., Intrator, N., Blais, B., & Shouval, H. Z. (2004). *Theory of cortical plasticity*. Singapore: World Scientific Publishing.

Darwin, C. (1859). *The origin of species by means of natural selection*. London: John Murray.

Fensel, D. (2004). *Ontologies: A silver bullet for knowledge management and electronic commerce* (2 ed.). Heidelberg: Springer.

Freeman, W. J. (2000). *Neurodynamics. An exploration in mesoscopic brain dynamics*. London: Springer-Verlag.

Gottgtroy, P., Kasabov, N., & MacDonell, S. (2004). An ontology driven approach for knowledge discovery in Biomedicine. In C. Zhang, H. W. Guesgen, & W. K. Yeap (Eds.), *Lecture notes in artificial intelligence*: *Vol. 3157*. *PRICAI 2004: Trends in artificial intelligence. Proc. VIII pacific rim intl. conf. AI* (pp. 53–67). Berlin: Springer-verlag.

Gottgtroy, P., Kasabov, N., & MacDonell, S. (2006). Evolving ontologies for intelligent decision support. In E. Sanchez (Ed.), *Fuzzy logic and the semantic web (capturing intelligence)*. Amsterdam: Elsevier.

Grossberg, S. (1982). *Studies of mind and brain*. Boston: Reidel.

Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, *5*(2), 199–220.

Kasabov, N. (2007). *Evolving connectionist systems. the knowledge engineering approach* (2nd ed.). New York: Springer.

Kasabov, N., & Benuskova, L. (2004). Computational neurogenetics. *Journal of Computational and Theoretical Nanoscience*, *1*(1), 47–61.

Kasabov, N., Jain, V., Gottgtroy, P. C. M., Benuskova, L., & Joseph, F. (2007). Brain gene ontology and simulation system (BGOS) for a better understanding of the brain. *Cybernetics and Systems 38*, (5–6), 495–508.

Kasabov, N., Jain, V., Gottgtroy, P. C. M., Benuskova, L., Wysoski, S.G., & Joseph, F. (2007). Evolving brain-gene ontology system (EBGOS): Towards integrating bioinformatics and neuroinformatics data to facilitate discoveries. In *International joint conference on neural networks*.

Kasabov, N. K. (2002). *Evolving connectionist systems. methods and applications in bioinformatics, brain study and intelligent machines*. London: Springer-Verlag.

Pisanelli, D. M. (Ed.) (2004). *Ontologies in medicine*. Amsterdam: IOS Press.

Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., et al. (2002). Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, *415*(6870), 426.

Taylor, J. G. (1999). *The race for consciousness*. Cambridge, MA: MIT Press.