Animal toxins: what features differentiate pore blockers from gate modifiers.

Jay Kola[1]
Kenneth Revett[2]

1 School of Biosciences
[2] Harrow School of Computer Science

Computational Modelling of the Gene Expression Profile From Acute Ischaemic Brain Injury

Jay Kola and Kenneth Revett

Abstract — **The ensuing events subsequent to cerebral ischaemia are complex and multi-faceted, making it difficult to extract causal relationships between the various pathways that are altered during ischaemia. In this study, we analyse a comprehensive DNA microarray dataset of acute experimental ischaemic stroke, in an effort to elucidate key regulatory elements that participate in the triggering of the pathways that lead to tissue damage. The data suggest that genes responsible for immediate early genes, apoptosis, neurotransmitter receptors (principally glutamate), and inflammation are differentially expressed at various time points subsequent to experimental ischaemia. Using unsupervised clustering (self-organising maps) and gene regulatory networks, we were able to establish a framework within which we could place the resultant gene expression changes into. Although not yet complete, the results from this study indicate that even a complicated pathology such as ischaemia can be analysed in a biologically meaningful way using DNA microarray technology.**

Keywords — **apoptosis, DNA microarrays, gene regulatory networks, ischaemia, & self-organising maps**

I. Introduction

THE World Health Organisation (1998) defined severe cerebral ischaemia (stroke) as: "a syndrome of rapidly developing clinical signs of focal or global disturbance of cerebral function, with symptoms lasting 24 hours or longer or leading to death, with no apparent cause other than of vascular origin" [1]. The underlying mechanism for stroke is ischemia characterised by transient or permanent reduction in cerebral blood flow in an area supplied by a blood vessel. The economic impact of stroke is staggering in both developed and developing countries. Wolfe estimated 4.5 million deaths a year from stroke world-wide with over nine million survivors and an overall incidence rate of 0.2 – 0.25 % [1,2]. Stroke per se is the leading cause of disability in adults, the second most important cause of dementia and third most important cause of death in developed countries [3]. The burden of stroke on healthcare systems around the world are staggering both in terms of the required resources with an estimated costs of 2–30 billion US dollars per year [4]. For these reasons, pathophysiology of stroke has been an area of active research.

The brain is the organ with the highest oxygen demand in the body, consuming 20% of the cardiac output. An ischemic event in the brain results in hypoxia (reduced tissue oxygen levels) and hypoglycaemia (reduced tissue glucose levels) leading to irreversible cell damage (stroke) in a few minutes if perfusion is not restored [1]. Cerebral ischaemia triggers a multitude of events of which three primary pathways have been reported to result in tissue damage: excitotoxicity, inflammation and apoptosis. These cascades form a complex series of pathophysiological events that evolve in time and space (see Figure 1 below). Together, these damage generating pathways, termed the 'Post Ischemic Cascade,' have been extensively studied using both in vitro and in vivo models to identify possible therapeutic targets in the treatment of stroke [5]. The three major pathways in the Post Ischemic Cascade are brought about by the interactions of numerous cellular components (e.g. transcription factors, caspases) that, until recently, have been identified and grouped by the functional roles these cellular components play in normal physiology [3]. That there is an underlying genetic basis for the altered activity of these cellular components involved in these pathways has been known for decades. What has hindered progress in elucidating the mechanism of action is a genetic basis that relates various elements within these pathways and their abnormal regulation in pathology. This is largely due to the magnitude of the changes associated with ischaemia, implicating possibly hundreds of cellular components. With the advent of DNA microarray technologies, new opportunities for studying large collections of genes in a single experiment provide an exciting opportunity to investigate the underlying

Jay Kola is from the Cavendish School of Biosciences, University of Westminster, London, England, UK. (e-mail: g0uk1@hotmail.com).

Kenneth Revett is from the harrow School of Computer Science, University of Westminster, London, England, UK HA1 3TPT. (e-mail: Biomodelling@aol.com).

genetic basis within and between various elements within the post-ischaemic cascade. [4,5,6,7].

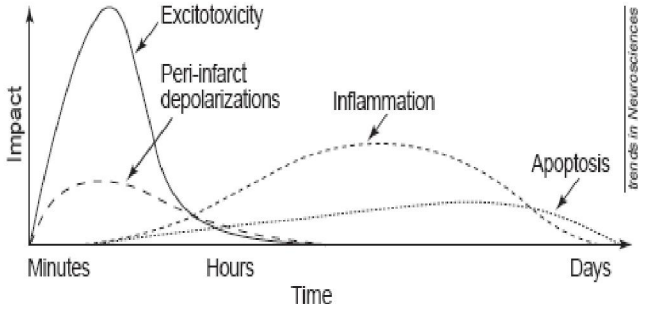

Fig. 1. *Putative cascade of damaging events in focal cerebral ischaemia. Very early after the onset of the focal perfusion deficit, excitotoxic mechanisms can damage neurones and glia lethally. In addition, excitotoxicity triggers a number of events that can further contribute to the demise of the tissue. Such events include peri-infarct depolarizations and the more-delayed mechanisms of inflammation and programmed cell death. The x-axis reflects the evolution of the cascade over time, while the y-axis aims to illustrate the impact of each element of the cascade on final outcome.*

If the post ischaemic cascade is under genetic control, then having a tool that can record changes in gene expression with sufficient temporal resolution (minutes preferably) may help elucidate the genetic basis for these pathways. Armed with this information, one can begin to build models of each pathway individually, and then connect these pathways together into a coherent framework. If checkpoints for these pathways exists and can be determined, than we may be able to regulate them in such a way as to curtail those pathways that generate tissue damage, while stimulating those pathways that are involved in tissue repair. In order to pursue this goal, we must have data regarding gene expression profiles of ischaemia and be able to analyse those data in a biologically realistic fashion.

II. METHODS

We present a preliminary study that investigates a comprehensive DNA microarray using an Affymetrix U34 Rat neurobiology array containing 1,322 functional genes across 6 time points (30 min, 4 hr, 8 hr, 24 hr, 3 days, and 7 days) obtained from experimentally induced cerebral ischaemia [8]. The dataset has been processed such that only genes displaying a 2-fold change at one or more time point(s) were considered to be differentially expressed. Those that met the call (267 of them) were subsequently classified into functional groups using available information from the literature. Table 1 lists the differentially expressed genes according to their particular functional grouping (from Lu et al.[8]). As can be seen in table 1, the three major tissue damage pathways in PIC are clearly identified (apoptosis, excitotoxicity -via neurotransmitter receptors, and inflammation). The first step in this work was to examine the expression profile of the genes classified on a purely functional basis matched the expression profiles of the differentially expressed genes – that is the whether the functional classification actually followed the expression patterns observed at these particular time points. We approached this problem using a self-organised

Table 1. Listing of the twelve functional categories that the genes in the DNA microarray study of Lu et al. [8] during a study on focal cerebral ischaemia. In addition, the authors found a set of 86 genes that were not previously identified using DNA microarray technologies in experimental models of cerebral ischaemia.

Functional Class	General Function of class
1. Immediate Early Gene	First to be elevated in response to ischemia
2. Transcription	Involved in transcription regulation
3. Heat shock Protein	Molecular chaperones
4. Inflammation	Involved in mediating inflammatory responses
5. Apoptosis	Involved in apoptosis
6. Cytoskeletal Structure	Maintain cytoskeletal structure
7. Metabolism	Involved in cellular metabolism
8. Growth Factor	Various growth factors for neural tissue
9. Signal Transduction	Involved in signal transduction between cells
10. Ion Channel	Ion channel genes
11. Neurotransmitter Receptor	Neurotransmitter receptors
12. Synaptic Protein Gene	Code for synaptic proteins

map – a standard unsupervised clustering algorithm using the package freely available software package GeneCluster 2.0 [http://www.broad.mit.edu/cancer/software/genecluster2/gc2.html]. The distance metric employed was the time series values across all time points for genes selected at random (85% were randomly selected for training). The results from the classification process are presented in Table 2. The data indicate that in this particular study, there is at best a 50% correspondence between genes classified functionally versus

classification via their temporal expression. Whether this is universally true will of course depend on the particular phenomenon under study. This is an important observation – as it indicates that a functional classification yields results

Table 2. Distribution of genes according to functional classes derived from the literature and the classification produced by using a SOM. As a parameter, we selected 12 classes in the SOM to see how well the functional classification matched the temporal expression patterns of all genes differentially expressed. The numbers indicate the percentage of the genes from that cluster that fall within the specific functional class. Highlighted values indicate maximal occurrence within a group

Gene Class \ Cluster Number	1	2	3	4	5	6	7	8	9	10	11	12
Apoptosis	-	18.7	-	-	18.7	25	31.2	-	-	-	6.25	-
Growth Factor	7.1	7.1	-	-	50	-	14.2		-	-	14.2	7.1
Heat Shock Protein	-	-	16.6	50	16.6	-	16.6	-	-	-	-	-
Immediate Early Gene	-	5.5	5.5	-	-	-	5.5	5.5	-	-	55.5	22.2
Inflammatory	4.5	27.2	-	-	13.6	31.8	-	4.5	4.5	4.5	4.5	-
Ion Channel	-	-	-	-	6.3	6.3	15.6	-	43.8	18.9	6.3	-
Metabolism	6.7	-	6.7	-	13.3	13.3	13.3	-	13.3	13.3	13.3	6.7
Neurotransmitter Receptor	-	-	-	-	8	4	-	-	56	20	12	-
Signal Transduction	-	2.8	-	-	5.7	-	17.1	-	22.8	25.7	27.1	11.4
Cytoskeletal Structure	7.1	14.2	-	-	7.1	-	35.7	-	14.2	14.2	-	7.1
Synaptic Protein	-	-	-	-	4.5	4.5	22.7	-	18.2	45.4	4.5	-
Transcription	-	-	-	-	5.9	5.9	23.5	5.9	17.6	5.9	11.7	17.6

that may be considerably different from a temporal expression profile. If one wishes to elucidate causal chains of gene induction, then functional classification alone may be insufficient for this purpose. One must also consider the temporal expression data as well.

In Figure 2, we present the clusters (12 in total) that were generated using GeneCluster 2.0. Each has a unique temporal profile, and each panel represents the prototype (centre plot) for the class and the min/maximal range of values for the genes within the cluster. The next stage was to develop a gene regulatory network from this data. Considering the large number of genes in each cluster (on average 22), we decided to determine if each cluster had a transcription factor associated with it. If a transcription factor was associated with each cluster, then one could make an assumption that the genes within that cluster were differentially expressed as a result of an associated transcription factor. A basic assumption of this hypothesis is that the transcription factor would be expressed *before* the other elements in the cluster. The temporal resolution of this study, and microarray studies in general is rather coarse – usually in the range of several hours. In this study, we required that the putative transcription factor must be expressed at the same time as any other gene under it regulatory control. We found that 7 clusters contained transcription factors (clusters 5,6,7,9-12). We therefore focused on those particular clusters, as we could in theory reduce the activity of the cluster to that of the specific transcription factor contained within it. (please note that all clusters have more than 1 transcription factor). The assumption here is that the activation of this group of proteins is under the regulation of 1 or more specific transcription

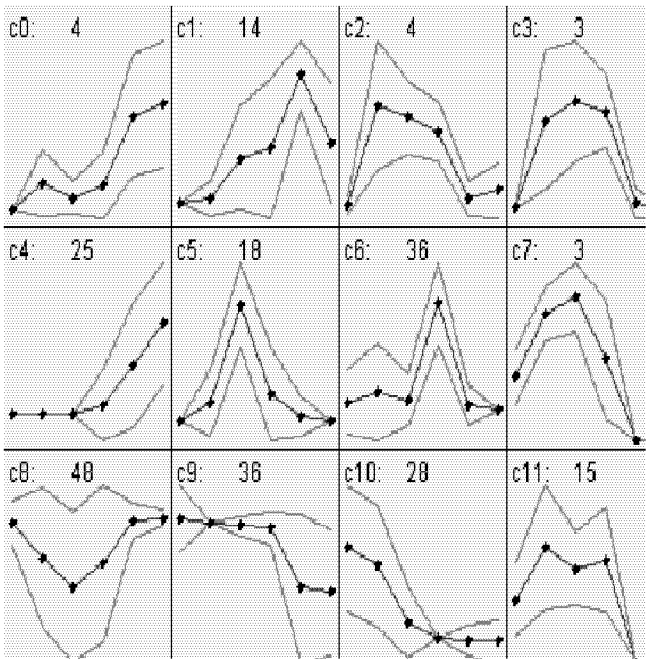

Figure 2. Average expression profile of the genes in the twelve clusters. The x-axis represents the various time points (30 min, 4 hr, 8 hr, 24 hr, 3 days, and 7 days) and the y-axis represents the magnitude of the expression levels

factors. In order to reduce the complexity of the cluster, we focus solely on the transcription factor. The gene regulatory network then becomes simplified to determining the relationship between a series of transcription factors. This reduces the complexity of the task, making the problem more tractable than trying to determine the relationship between 267 genes. Using literature results and the software package GeneNetwork we produced a series of GRNs for the clusters that contained transcription factors. We have produced a series of gene regulatory networks for 7 of the 12 clusters. Each GRN was generated based on a Bayesian network model, which, due to their probabilistic nature, handle biological variation inherent to microarray datasets [9].

III. RESULTS

We present a partial GRN for the apoptosis cluster in Figure 3. The genes listed under this cluster are Bcl-x, TGF-α, ANIA-6, c-jun, pJun B, JAK1 and JAK2, MAP kinase phosphatase and the transcription factors associated are Jun dimerisation protein, rNFIL-6 and Transcriptional repressor. From Fig 3 of

apoptotic pathway below, it can be seen that Bcl-x, c-jun, MAP kinase are all involved in the apoptotic pathway. The transcriptional factor Jun dimerisation protein is involved in transcriptional regulation of c-jun and c-fos family of proteins (c-jun, jun-B and jun-D) [9]. These proteins form homo and heterodimers with themselves and the DNA and are involved in the pro apoptotic pathways. Jun Dimerisation Protein (JDP) is considered to be a general repressor of transcription related to the c-jun family. It is supposed to form dimers with proteins of c-jun family and interfere in their pro apoptotic effects. Janus protein tyrosine kinases (JAK) are a family of proteins that are upstream regulators of several signalling pathways.

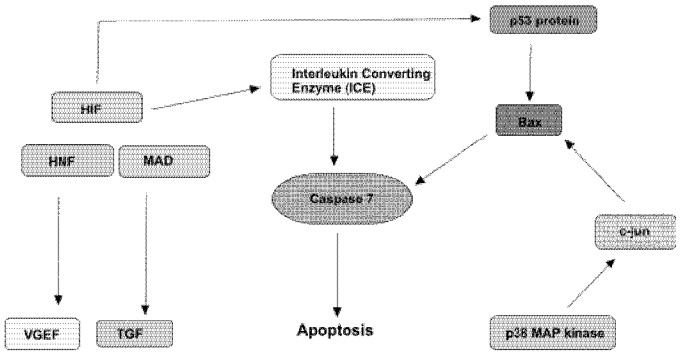

Figure 3. Representation of possible gene regulatory network for cluster 7 the primary apoptosis cluster. The pro-apoptotic actions of this cluster are followed by activation of downstream growth factors. If this were the case, then we would expect to find a cluster of growth factors with a temporal expression pattern falling after this cluster

They are involved in the JAK/STAT pathway that has an anti-apoptotic mechanism of action that is not clearly understood. The induction of JAK following ischemia has been shown to be protective in focal ischemia and could be mediated by the binding of Stat3 protein of the JAK/STAT pathway to the members of the c jun / c-fos family [128]. From figure 3 of

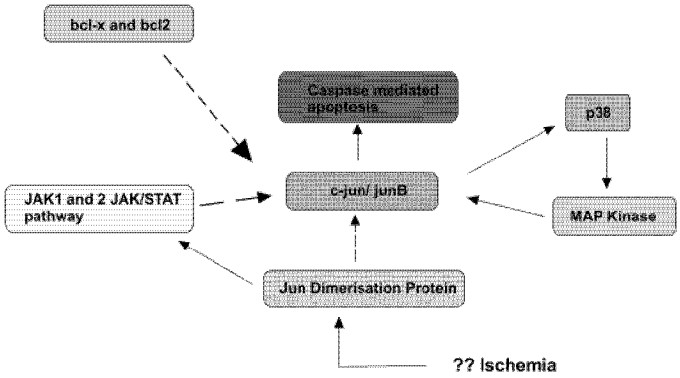

Figure 4. Representation of possible gene regulatory network for cluster 11 – the immediate early gene cluster. Inhibitory influencess are depicted by dotted arrows and excitatory influences depicted by the solid lines. In this figure, c-jun is the primary transcription factor activated by ischaemia that may lead to the process of apoptosis

apoptosis it can be seen that MAP kinase has a positive feedback on c-jun which is mediated through p38 protein and HSP27 [10]. This explains the possible relationships between all these genes except for bcl-x. Bcl-x is part of the Bcl2 family of anti apoptotic proteins and antagonises the pro apoptotic effects of c-jun. From this summary, we have 5 anti-apoptotic genes and 3 pro-apoptotic genes in this cluster.

It is possible to generate a regulatory pathway for this cluster with c-jun at the centre and JDP acting as the transcription factor (see figure 4 for details). The important genes listed under this cluster are Bax, Caspase 7, IL-1β Converting Enzyme (ICE), Inhibitor of Apoptosis, TGF receptor, p38 MAP kinase, SAP kinase, etc. The main transcription factors associated with this cluster are Hypoxia Inducible Factor (HIF), Hepatocyte Nuclear Factor 3 (HNF-3) and Mothers against dpp 1 protein (MAD-1). Investigation of the genes in this cluster revealed several interesting finding and associations with each other.

Hypoxia Inducible Factor (HIF) as its name indicates is induced in all states of tissue hypoxia and has been extensively studied in relation to stroke. It is believed that HIF mRNA is rapidly transcribed in response to hypoxia through a hypothetical oxygen sensing system which involves a flavoprotein [9]. It is believed that HIF is responsible for neuroprotection induced by hypoxic preconditioning through the induction of various downstream target genes. HIF target genes include those related to vasomotor control (NOS2), angiogenesis (VEGF, FLT-1), blood and iron metabolism (EPO, transferrin, transferrin receptor, ceruloplasmin), cell proliferation [IGF-1, IGFBP-1, -3, transforming growth factor (TGF)β], and energy metabolism (glucose transporter-1, -2, and -3, phosphoenolpyruvate carboxylase, lactate dehydrogenase A, aldolase, phosphoglucokinase-1, -L and -C, pyruvate kinase, enolase, and many others, including prolyl 4-hydroxylase and adenomedullin) [8]. Of these genes, the growth factors like VEGF induce new blood vessel formation (neoangiogenesis) and help reperfusion of the ischemic tissue in the later stages [52]. HIF also induces expression of Inhibitor of Apoptosis Protein (IAP) family. The Inhibitor of Apoptosis proteins (IAP) are a family of novel genes that function in the cell death pathway to block apoptosis induced by a variety of triggers [9]. It was shown recently that the mechanism by which the IAPs inhibit apoptosis is direct inhibition of key apoptotic proteases, caspase 3 and 7 [8]. However, on the other hand, HIF-1 can also initiate apoptosis by inducing high concentrations of proapoptotic proteins, such as BNIP3, and can cause stabilisation of p53 protein [9]. p53 is a well known mediator of apoptosis via Bax group of proteins as indicated in figure 3 [7]. P38 MAP kinase is a

signal kinase which mediates caspase mediated apoptosis along with JNK and ERP kinase [7]. It is believed that p38 MAP kinase is the major mediator of apoptosis acting via caspase 3 and caspase 7 and its upstream kinases like MAP kinase kinase kinase . In fact many experimental models of neuroprotection have been developed using inhibitors of p38 MAP kinase like SB203580, SB239063, etc.[148, 149] From the figure 5, it can also be seen that p38 MAP kinase also activates through Activating Transcription Factor 2 (AT2) [7]. The regulatory network for the genes in cluster 11 are displayed in figure 4.

Hepatic Nuclear Factors are a group of transcription factors that are involved in cell differentiation and embryogenesis [11]. Increased expression of HNF has been demonstrated in hepatic ischemia and it is believed to act in concert with HIF and Heat Shock Proteins in protecting the cell from damage [11]. It is very likely that HNF serves the same function in the brain. MAD 1 and 3 proteins are transcriptional factors and have been proved to be associated with the target genes of

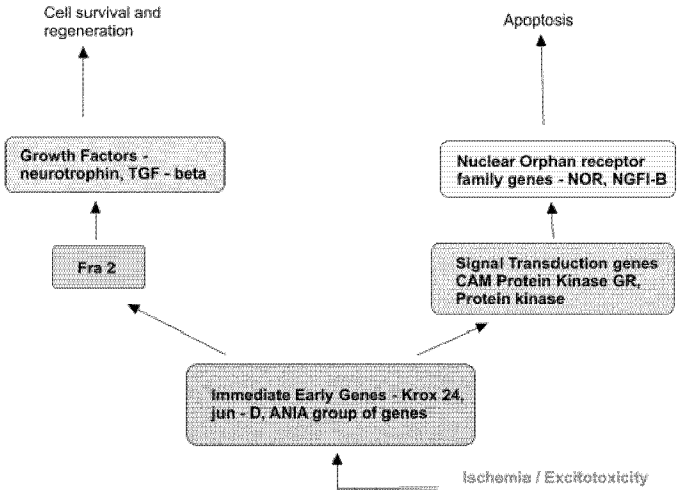

Figure 5. Representation of possible gene regulatory network for cluster 12 – a mixture of immediate early genes and transcription factors. Inhibitory influences are depicted by dotted arrows and excitatory influences depicted by the solid lines. In this figure, c-jun is the primary transcription factor activated by ischaemia that may lead to the process of alterations in systemic metabolism

The cluster presented in figure 5 presents some difficulties in generating the gene regulatory network (GRN). Since the cluster is composed predominantly of immediate early genes *and* transcription factors, it becomes confusing as to which gene is the one to be expressed earliest. In fact, some recent authors have classified the ANIA (activity and neurotransmitter-induced early gene) group of genes under transcription factors [11,12]. On the same lines, it is believed that the Immediate Early Genes (IEG) are in fact the first to respond to external stimuli and these then set off the transcription factors which then act as third messengers in the signalling pathway [12]. Apart from this confusion, it has been shown that Fra2 (Fos related antigen 2) and jun D proteins are part of the AT-1 family of inducible transcription factors [11]. The function of Fra 2 in vitro seems to be obscure and has been linked to delayed neuronal damage and also neuronal regeneration through induction of growth factors and cytokines . Krox 24 is also part of the AT-1 family and is similar in function to Fra and jun D proteins. They are involved in regulation of expression of late-response genes and may influence neuronal plasticity [13], but it has also been suggested that they play a role in neuronal damage. Specifically, it has been suggested that prolonged NGFI-A expression after an ischemic insult is associated with delayed neuronal degeneration [14].

Transforming Growth Factor β (TGF-β) and to potentiate its action [12]. IL-1β Converting Enzyme (ICE) is a member of the caspase family of cysteine proteases, which are associated with both proinflammatory and apoptotic actions [11, 12]. ICE is critical for processing pro-IL-1 to the biologically active form of the cytokine, IL-1 [12]. From this, it can be seen that the genes in this cluster are mainly pro apoptotic (caspase 7, Bax, p38 MAP kinase, ICE) but this is also accompanied by activation of neuroprotective factors like VGEF, TGF-β downstream by HIF, HNF and MAD-1. NGFI-B (also known as Nurr 77) is a transcription factor that belongs to the nuclear receptor family along with Nuclear Orphan Receptor 1 and are both constitutively expressed in the brain and other peripheral tissues [14]. These two proteins have been documented to play a role in T cell mediated apoptosis however their downstream targets remain elusive [15]. However more recently, these proteins have been shown to be activated through the Ca++/ CAM kinase pathway [13]. It is likely that the signal transduction genes CAM protein kinase GR, protein kinase, etc are involved in activation of these proteins. Following activation, these genes promote apoptosis [14]. From this, a GRN for this cluster can be attempted with the Early Intermediate genes activating the Nuclear Orphan Receptor (NOR) proteins through the signal transduction proteins. These NOR proteins then promote apoptosis via yet unknown mechanisms.

IV. CONCLUSION

In this work, we examined a comprehensive dataset that was used to investigate genes that are differentially expressed in and animal model of ischaemic stroke. Such studies by their nature tend to produce a bewildering amount of data that requires exacting interpretation to be useful. In general, studies of this sort reduce the complexity by making an initial classification based on literature reviews of the genes involved, attempting to place them into functional categories.

This approach has become feasible with the advent of international projects such as the Human Genome Project, which has made available a wide range of genetic information. At the same time, the rather arbitrary clustering of genes based on functional homology may not be sufficiently accurate to allow detection of subtle changes in gene expression profiles that appears to be required to elucidate the mechanism of action of disease. In this study, we also perform a temporal sequence analysis and have found that more than 50% of differentially expressed genes differ from strict functional classification. We also found that not all genes are created equal – for instance transcription factors, which are responsible for inducing the expression of other genes may need to be treated differently from other genes. In this study, 7 out of the 12 clusters that were generated using self-organised maps contained transcription factors. We were able to reduce the complexity of this series of genes (267 in total) by focusing on those clusters that contained differential expression of transcription factors. Although this process has not yet been completed, it is possible that this approach can reduce the complexity of the functional roles of differential expression of genes, as detected by DNA microarray studies by allowing collections of genes to be represented by their regulatory elements – in this case transcription factors. As microarray technology becomes more sophisticated and widely used, we may find that each experiment will contain hundreds to thousands of differentially expressed genes, making biologically plausible interpretation of the results extremely difficult.

REFERENCES

[1] Braunwald Eugene; Athony Fauci; Dennis Kasper; Stephen Hauser; Dan Longo; J. Jameson (2001); Harrison's Principles of Internal Medicine , Fifteenth

[2] Wolfe C. D A ; The impact of stroke ; British Medical Bulletin, 1 October 2000, vol. 56, no. 2, pp. 275-286(12)

[3] Neumar RW. 2000. Molecular mechanisms of ischemic neuronal injury. Ann Emerg Med 36: 483-50

[4] Bakhai A ; The burden of coronary, cerebrovascular and peripheral arterial disease ; Pharmacoeconomics. 2004;22 Suppl 4:11-8

[5] Matchar DB, Duncan PW. Cost of stroke. Stroke Clin Updates. 1994;5:9–12Sandercock P, Willems H. Medical treatment of acute ischaemic stroke. Lancet. 1992;339:537–539.

[6] Dirnagl, Costantino Iadecola and Michael A. Moskowitz (1999) ; Pathobiology of ischaemic stroke: an integrated view ; Trends Neurosci. (1999) 22, 391–397.

[7] Kathy Keyvani, Otto W Witte and Werner Paulus (2002) ; Gene Expression Profiling in Perilesional and Contralateral Areas After Ischemia in Rat Brain; Journal of Cerebral Blood Flow & Metabolism (2002) 22, 153 –160

[8] Lu XC, Williams AJ, Yao C, Berti R, Hartings JA, Whipple R, Vahey MT, Polavarapu RG, Woller KL, Tortella FC, Dave JR. (2004) ; J Neurosci Res. 2004 Sep 15;77(6):843-57

[9] Jung-Bin Kim, Chun-Shu Piao, Kang Woo Lee, Pyung-Lim Han, Joon Ik Ahn, Yong Sung Lee and Ja-Kyeong Lee (2004) ; Delayed genomic responses to transient middle cerebral artery occlusion in the rat; Journal of Neurochemistry Volume 89 Issue 5 Page 1271 - June 2004

[10] Vemuganti R, Dhodda V, Song G, Bowen KK, Dempsey RJ (2002) ; J Neurosci Research 71:208-219

[11] Derek E. Clevidence, D. G. Overdier, WUFAN TAOt, X. Qian, L. Pani, E. Lalt., & R.. H. Costa ; (1993); Proc. Natl. Acad. Sci. USA Vol. 90, pp. 3948-3952, May 1993 Biochemistry

[12] Yan Chen, Jean-Jacques Lebrun, and Wylie Vale (1996) ; Proc Natl Acad Sci U S A. 1996 November 12; 93(23): 12992–12997

[13] Abraham, W.C., Dragunow, M. & Tate, W.P. 1992 The role of immediate early genes in the stabilization of long-term potentiation. Mol. Neurobiol., 5, 297-314

[14] Akins, P.T., Liu, P.K. & Hsu, C.Y. 1996 Immediate early gene expression in response to cerebral ischemia. Friend or foe? Stroke, 27, 1682-1687

[15] Mario Maira, Christine Martens, Alexandre Philips, and Jacques Drouin (1999) ; Mol Cell Biol. 1999 November; 19(11): 7549–7557.