Blood genomic profiling: novel diagnostic and therapeutic strategies for stroke?

A.E. Baird1
Stoke Neuroscience Unit, NINDS (National Institute of Neurological Disorders and Stroke)/NIH (National Institutes of Health), 10 Center Drive, MSC 1294, Room 3N258, Bethesda, MD 20892-1294, U.S.A.

Abstract
Findings from gene expression profiling studies are leading to new diagnostic and therapeutic strategies that can be applied in medical practice, especially in the field of oncology. Promising results of gene expression profiling of the peripheral blood in patients with ischaemic stroke have been obtained in recent pilot studies, demonstrating a partially reproducible gene signature of acute cerebral ischaemia. However, questions remain. Given that blood is at least in part a surrogate tissue for ischaemic stroke, the specificity of these signatures needs to be evaluated. Furthermore, it needs to be determined whether standardization of this methodology is required and whether clinical signatures can be identified that are improvements over the tools currently used in clinical practice. Clinically useful signatures would include those of haemorrhagic as well as ischaemic stroke, reclassification of stroke type and prognosis, and vascular disease risk. If these conditions are met, then it should be possible to develop cost-effective and rapid assays.

Introduction
The sequencing of the human genome and the advent of microarray technology are opening up new possibilities for applying genomic information in medicine by permitting multiple genes to be studied simultaneously [1–3]. Information from multiple genes seems to offer considerable hope for providing new insights into the pathogenesis of complex but common disorders such as stroke. With the use of genome-wide scanning, combinations of SNPs (single-nucleotide polymorphisms) in DNA are now being studied in diverse disease states such as heart disease, Alzheimer’s disease and diabetes.

Another approach for studying the human genome is to measure gene expression. Gene expression is a dynamic process and reflects changes in mRNA levels resulting from normal physiology or disease. Gene expression profiling, also known as transcriptional profiling, permits changes in gene expression resulting from structural variation in DNA as well as environmentally induced changes, to be studied simultaneously in thousands of genes, and to be measured serially over time. Gene expression profiling has recently been described in the literature as a methodology that ‘is coming of age’ [4] and that has ‘important implications for clinical research and practice’ [5]. The cancer literature has led the way in studies of tumour specimens to the extent that gene expression signatures are now being used in the management of patients with breast cancer in the Netherlands [4,6,7]. Also the results from gene expression experiments are being applied to improve diagnostic accuracy in Burkitt’s lymphoma, permitting more accurate differentiation from diffuse large-B-cell lymphoma [8,9].

Current stroke management is based largely on the clinical examination and brain and cardiovascular imaging studies, but is imperfect at best. For example, the accuracy of early stroke diagnosis may only be as much as 60% [10], and in as many as 30–40% of ischaemic strokes the cause cannot be found [11]. There is also at present only one treatment available for ischaemic stroke, the thrombolytic agent rtPA (recombinant tissue plasminogen activator) that must be administered within 3 h of onset. Therefore, given the lack of treatment options for stroke and the need for rapid and accurate diagnosis, there is great potential to apply gene expression profiling in stroke. As brain tissue is rarely available, the peripheral blood is a most practical source of mRNA in the clinical setting, despite being in part a surrogate tissue for cerebral ischaemia. Blood is also the most practical source of mRNA for many other disease states and there has been a rapid rise in the number of blood-based studies over the past 2–3 years, for example in multiple sclerosis and rheumatoid arthritis, along with studies of therapeutic intervention [12,13].

Promising results of blood genomic profiling in stroke have been obtained in recent pilot studies. Moore et al. [14] reported that a gene expression signature of acute ischaemic stroke could be identified from profiling of PBMCs (peripheral blood mononuclear cells) using Affymetrix microarrays on which there are 22283 gene probes. The work was validated with real-time PCR and also in an independent cohort of subjects. Tang et al. [15] used the PreAnalytiX (PAX)geneTM (Hombrechtikon, Switzerland) whole blood RNA stabilization platform and, moreover, have also come up with a gene expression signature of acute ischaemic stroke. There was surprisingly an overlap of approx. 20% of the
genes in the two studies, despite the different time points after stroke, and the use of different arrays, different cell types, and different statistical analyses.

This pioneering work leads to many new questions. As blood is a surrogate tissue for stroke, how specific are these signatures? Furthermore, early gene expression studies have varied markedly in their quality, in particular, the quality of RNA that was used, and have been plagued by issues of reproducibility: should the methodology be standardized? Are the clinical signatures an improvement on current diagnostic and management approaches for stroke? And if clinically useful panels of genes can be identified, can they be made into cost-effective tools for use in the clinical setting?

**Specificity of gene expression signatures for ischaemic stroke**

The rationale for using the peripheral blood in gene expression profiling studies of stroke is that after stroke there is a well-characterized selective migration of blood-borne leukocytes into the ischaemic focus. There is an initial and short-lived neutrophil migration within the first few hours, followed by a more prolonged and sustained infiltration of monocytes and lymphocytes. This infiltration is believed to impact significantly on tissue and clinical outcome after stroke, by having diverse effects ranging from harmful (for example, exacerbation of reperfusion injury) to beneficial (promotion of tissue repair and remodelling). Recent studies have reported that there is a massive and rapid activation of the peripheral immune system that appears to be a dynamic reflection of inflammatory changes occurring in the brain, and have implied an increasingly important role for lymphocytes early on after cerebral ischaemia [16–18]. Furthermore, since the 1970s, it has been recognized that peripheral leukocytes become sensitized to brain antigens after stroke in a highly specific manner [19–21]. Promising and interesting results of gene expression profiling of the peripheral blood have been reported in experimental stroke models [22,23] as well as in human studies in patients with multiple sclerosis [24]. Both the studies of Moore et al. [14] and Tang et al. [15] also showed evidence of gene changes suggestive of an adaptive response to the altered cerebral microenvironment (up-regulation of genes induced by hypoxic stress, inhibition of neuronal apoptosis, and vascular repair). Moore et al. [14] compared their listing with the blood genomic profiles of two other disease states, (i) multiple sclerosis that is associated with chronic brain inflammation and (ii) sickle cell disease that is associated with ischaemic crises and chronic inflammation, and found minimal overlap of the gene signatures [14,25]. However, the gene expression signatures obtained may be reflective of a non-specific response to stress. Therefore the specificity of these stroke signatures needs to be determined in studies comparing the blood gene expression signature of acute ischaemic stroke with those of other disease states, for example, hospitalized patients with acute myocardial infarction. Approaches could include a consortium approach (involving multiple medical specialties), or meta-analyses, or otherwise testing in the field or the emergency department (for example, the diagnostic accuracy of the 22-gene panel for acute ischaemic stroke by Moore et al. [14]).

**Methodological standards**

Early gene expression studies have been plagued with variable reproducibility, but it has been shown that if strict laboratory methods are adopted, then good reproducibility can be shown, even across array platforms [12,26–28]. The most difficult aspect of an array experiment in the clinical setting is obtaining high-quality mRNA [12,13]. Once the blood is drawn, RNA degrades very quickly, and so must be isolated as quickly as possible. This is achievable in specialized laboratories, as used by Moore et al. [14] where mRNA was isolated from PBMCs, but is not usually practical in the hospital or clinical setting. The PAXgene approach used by Tang et al. [15] has merit (RNA is stabilized and so sample processing can be delayed for up to 24 h) but it has the limitation of a severe reduction in sensitivity relative to mRNA isolated from PBMCs.

The issue of which leucocyte type(s) should be used is also an interesting one, and it is not clear to what extent this impacts on the results obtained in the clinical setting. The PAXgene methodology uses all of the peripheral leucocytes, whereas PBMCs as used by Moore et al. [14] consist primarily of lymphocytes and monocytes. Du et al. [29] have recently been addressing the issue of the importance of cell type and have reported that the genes expressed in acute ischaemic stroke are most likely to be expressed in neutrophils under physiological conditions, although this does not exclude the possibility that cerebral ischaemia induces the expression of some genes by monocytes and lymphocytes that are not expressed under normal physiological conditions. This finding of Du et al. [29] also does not explain the 85% accuracy of the panel of 22 genes used by Moore et al. [14] when tested by Tang et al. [15], in which the blood draw was obtained within 3 h of stroke onset. It may be that at present only the most robust and prolonged gene expression changes are being detected reproducibly in stroke patients. While it is crucial that the most exact and strictest of laboratory methods should be employed, it is currently not clear as to what extent the methodology needs to be standardized in terms of cell populations and time of blood draw.

**Clinically useful gene expression signatures need to be better than current clinical tools**

The clinical purpose of the ‘-omics’ techniques (Table 1) is to find clinically useful genes that improve on the current diagnostic, management and prognostic tools in stroke [4]. As described in the Introduction section, this is much needed in the field of stroke. Current stroke diagnosis relies on the clinical examination and getting the patient to the hospital for a brain scan, either using computed tomography or magnetic
Table 1 | Glossary of ‘-omics’ terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomics</td>
<td>The study of genomes and the complete collection of genes that they contain. A short time ago, this collection might have been limited to protein-encoding genes, but genomics has shown that many other elements have important functions in the genome, such as transcription factor-binding domains, regions encoding microRNAs, and antisense transcripts, and large, evolutionarily conserved regions. The primary technique used in genomics is high-throughput genome sequencing.</td>
</tr>
<tr>
<td>Functional genomics or transcriptomics</td>
<td>Also known as gene expression profiling and attempts to analyse patterns of gene expression and to correlate the patterns with underlying disease biology. There is a wide range of techniques used, including DNA microarray analysis and serial analysis of gene expression.</td>
</tr>
<tr>
<td>Metabolomics or metabonomics</td>
<td>A large-scale approach for monitoring as many as possible of the compounds involved in cellular processes in a single assay to derive metabolic profiles. Although metabolomics first referred to the monitoring of individual cells and metabonomics referred to multicellular organisms, these terms are not often used interchangeably. Techniques applied to metabolic profiling include NMR and MS.</td>
</tr>
<tr>
<td>Proteomics</td>
<td>The examination of a collection of proteins to determine how, when and where they are expressed. Techniques include two-dimensional gel electrophoresis, MS and protein microarrays.</td>
</tr>
</tbody>
</table>

resonance imaging, to exclude intracerebral haemorrhage and other causes of the stroke-like symptoms. Blood tools offer the potential of being widely available, rapid, practical and cheap. There is only a 3 h window period for the administration of the thrombolytic agent rtPA that is effective for the treatment of acute ischaemic stroke, and the earlier the drug is administered, the more effective it is.

Panels of genes for acute ischaemic stroke show promise as a diagnostic tool. As mentioned above, the 22-gene panel of Moore et al. [14] may be as accurate as 85% for stroke diagnosis in the first 3 h when applied to the sample of Tang et al. [15]. This may be developed into a diagnostic tool and tested out for clinical use. Other panels of genes could be developed for haemorrhagic stroke to permit the two major types of stroke to be differentiated. A particularly promising area of clinical development is in disease reclassification and the identification of stroke types that allow improved prognosis, diagnosis and treatment options along with prediction of stroke recurrence and response to treatment (whether with rtPA or with potential new agents currently in clinical trial). As one example, the 22-gene panel could possibly identify patients at risk of ischaemic complications after Factor VIIa therapy for acute intracerebral haemorrhage; the agent Factor VIIa is in trial but may lead to thrombotic complications in older patients. A further application is for the prediction of stroke and cardiovascular disease risk. A number of examples already exist in the cancer literature, for tumour metastatic risk, response to chemotherapy and poor outcome. Gene expression profiling is increasingly being applied to the study of atherosclerotic vascular disease, for example in the aorta, and coronary artery disease, for example to study the factors associated with in-stent coronary artery restenosis [30–33].

It may be that the panels of genes identified are not those that give the best biological insights: it may be that the most reproducible results are coming from the most up-regulated/down-regulated genes or, as suggested above, from those with the most robust and prolonged expression. The panels in essence could be biomarkers, with all that that entails, and would need to show the properties of ideal biomarkers to be clinically useful (highly sensitive and specific for the disease under study, and easily measurable). This so far has not been a problem in other disease states, as some of the best cancer biomarkers in use include the plasma protein markers carcinoembryonic antigen used for the detection of bowel cancer and prostate-specific antigen used for the diagnosis of prostate cancer; for neither of these has the biological function been precisely elucidated. Therefore, perhaps it needs to be regarded that obtaining new biological ‘insights into mechanism (will be) a bonus’ [4] from gene expression profiling, although there is significant potential for novel information to be obtained from relevance network analyses, in which gene–gene interactions can be studied.

Development of assays for use in the clinic

Gene expression profiling represents a new approach to clinical and translational research, requiring teams with multiple areas of expertise, particularly in clinical medicine, molecular biology and bioinformatics as well as in bioengineering. This type of work is also extremely costly and raises questions as to how results that appear to have clinical utility can
ultimately be applied to clinical practice. As stated in a recent editorial by Harris and Horning [9]: ‘RNA extraction and microarray analysis are laborious and expensive and are therefore not ready for real time diagnosis in clinical practice’.

However, a number of options are beginning to emerge. One approach would be through the development of custom-made arrays of a small number of genes that could be analysed with current machine readers. The first Food and Drug Administration (U.S.A.)-approved array is the AmpliChip® CYP450 Test for the analysis of 31 SNPs and mutations in two genes in the cytochrome P450 system that can greatly influence the metabolism of many commonly used drugs. The test identifies a patient’s genotype and categorizes their predicted phenotype as a poor, intermediate, extensive, or ultrarapid metabolizer. The AmpliChip® uses PCR and Affymetrix technology and can take as few as 8 h to get a result. It should be noted that this technology is not based on gene expression profiling. Conceivably, custom-made chips could be made from the results of gene expression profiling in a similar vein, although cost-utility could be a concern along with the time taken for processing. Multiplex PCR is becoming increasingly available and it is now possible to measure up to 4–8 (or even more) mRNAs simultaneously. The protein products of up- or down-regulated genes could also provide valuable diagnostic tools through the development of serum- or plasma-based assays. In the case of Burkitt’s lymphoma, gene expression profiling led to the recognition of new immunophenotyping markers that may be useful in the setting of acute stroke and sepsis where rapid answers are needed, preferably within minutes.

Conclusions

Gene expression profiling is gaining acceptance in the field of oncology and is starting to impact on the management of some diseases. Promising results of gene expression profiling of the peripheral blood in stroke have recently been reported, even though blood is, in part, a surrogate tissue in brain ischaemia. These results have shown reproducibility, although it is vital that strict methodology is used, that issues of cell type be addressed and that the specificity of signatures for stroke be worked out. Also it needs to be determined whether clinically useful signatures such as those for the diagnosis of ischaemic and haemorrhagic stroke, prediction of vascular risk, and for stroke reclassification and outcome can be developed, and whether these offer improvements over currently used methods. Ultimately, panels of genes or gene/protein/lymphocyte products may be developed that provide new disease biomarkers with the potential to increase sensitivity while maintaining specificity for disease. Ideally rapid and cost-effective assays would subsequently be developed. This new era of clinical and translational research may lead to the development of tools that begin to make personalized medicine a reality.

References

30 Seo, D., Wang, T., Dressman, H., Herderick, E.E., Iversen, E.S., Dong, C.,
31 Chon, H., Gaillard, C.A., van der Meijden, B.B., Dijstelbloem, H.M.,
Kraaijenhagen, R.J., van Leenen, D., Holstege, F.C., Joles, J.A., Bluyssen,
201–205
33 Ganesh, S.K., Skelding, K.A., Mehta, L., O’Neill, K., Joo, J., Zheng, G.,
Pharmacogenomics 5, 952–1004
34 Biberthaler, P., Bogner, V., Baker, H.V., Lopez, M.C., Neth, P., Kanz, K.G.,

Received 20 July 2006