npg

EDITORIAL

Fine mapping the phenotype in autoimmune disease: the promise and pitfalls of DNA microarray technologies

The recent proliferation of microarray technologies for monitoring gene expression has presented enormous technical and analytical challenges for practising scientists. The ability to obtain massive amounts of information about the transcriptional state of cells and tissues is extremely seductive, with the promise of rapid progress. However, after an initial period of exhilaration, giving in to temptation and initiating these studies can lead to information overload, confusion and disappointment, unless one is prepared for a long-term commitment. Nevertheless, the field is beginning to move beyond some of these difficulties, and two papers in this issue provide evidence that gene expression phenotyping can provide insight into the underlying heterogeneity of autoimmune diseases.

The most convincing evidence that DNA microarrays can be useful for meaningful categorization of disease is in the field of cancer.[1] The seminal studies of Staudt *et al* have clearly shown that B-cell tumors exhibit distinctive gene expression profiles that correlate with the underlying pathophysiology and disease outcome.[2] More recently, expression profiling has been shown to outperform all other predictors of outcome in early breast cancer, including tumor size and local lymph node involvement, suggesting that microarray technology may have a dramatic influence on the approach to treatment of this disease in the future.[3] A major advantage of applying microarray techniques to tumors is the fact that tumors are highly clonal, thus largely eliminating the complication of interpreting results on mixtures of different cell types. For most autoimmune diseases, the exact cell types involved in the primary abnormality are uncertain at best. Some commentators have voiced skepticism concerning the utility of analyzing mixed cell populations for gene expression, and have suggested that a focus on specific cell types should be a criterion for evaluating the quality of rheumatic disease research that utilizes microarray technology.[4] While the analysis of pure cell populations is undoubtedly desirable and simplifies data interpretation, this restrictive approach may not be appropriate for more clinically oriented studies. The studies by Han *et al* and van der Pouw Kraan *et al* illustrate this point.

In an analysis of 10 Chinese lupus patients, Han *et al*[5] identified 61 genes that exhibited differential expression compared with controls. A mixed population of peripheral blood mononuclear cells was studied. The authors used a combination of criteria to focus on these genes, including a two-fold elevation of expression levels in at least half of the patients and evaluation of overall significance using a *t*-test statistic. Other criteria might have been applied, but one has to start somewhere in order to focus on genes that are clearly different in at least a subset of patients. Many of the common analytic approaches utilize some form of a *t*-test, along with various somewhat arbitrary criteria for fold changes, as detailed in a recent publication.[6] The basic goal is pattern discovery, and different criteria may be suitable depending on the experiment. Clearly, in order to detect heterogeneity in the disease group, one should not insist on uniform differences between cases and controls.

The lupus data show that a group of interferon-inducible genes are frequently upregulated in the lupus patients. This result is consistent with other gene expression studies implicating interferon pathways in lupus pathogenesis,[7] including our own studies in a set of 48 lupus patients and 42 matched controls using Affymetrix oligonucleotide array technology.[8] This larger sample size also allowed us to identify associations between this interferon 'signature' in peripheral blood and specific clinical manifestations, including CNS and renal disease. It will be of considerable interest to see whether Han *et al*[5] will observe similar associations in the Chinese lupus population as they expand their studies.

The work of van der Pouw Kraan *et al*[9] is also based on a relatively small sample size of 15 rheumatoid synovial samples. The analysis is again performed on a highly mixed cell population found in synovial tissue. Nevertheless, despite these complications, they make a provocative observation, namely that the synovial disease process can be stratified into subgroups based on gene expression patterns reflective of 'immune activation' *vs* 'tissue repair and remodeling'. Of course, these results are hardly definitive, since many questions remain. No correlation with histology was done, and it is not clear to what degree sampling bias contributed to these differences. For example, it would be of great interest to know whether a given gene expression pattern is consistent among many joints from the same patient. Therefore, these data should be strictly viewed as hypothesis generating, but as such they offer a benchmark for comparison with future studies. As pointed out by the authors, a major clinically relevant aspect of disease heterogeneity in rheumatoid arthritis relates to response to anti-TNF therapy; approximately 30% of patients have no clinical response to this treatment, whereas a substantial minority exhibit dramatic clinical improvement.[10] If microarray data can be utilized to identify responder and nonresponder subgroups, it would be a major step forward in the management of this disease. Ideally, these assays could be done on peripheral blood cells instead of synovial tissue.

Owing to its potential clinical utility, we have focused our efforts on the analysis of peripheral blood cells for gene expression patterns. As discussed above, our data in lupus is similar to the observations of Han *et al*.[5] We

hope to establish whether similar analyses in rheumatoid arthritis (RA) patients can be used to predict response to anti-TNF therapy. In the course of these studies, we have also found that sample handling, by overnight shipping or even in the first few hours after blood draw, can have a dramatic impact on gene expression levels in peripheral blood cells. After overnight shipment, approximately 2000 genes were found to be significantly dysregulated in peripheral blood mononuclear cells.[8] These include many genes in 'stress' pathways such as fos, jun, TGF beta-inducible early growth genes, as well as cytokines such as IL8 and the CD69 T-cell activation marker. Inasmuch as clinical studies are often carried out on samples shipped from various recruitment sites, these data are a caution for scientists considering microarray analysis in this context. We are currently exploring the use of blood collection tubes, which result in immediate stabilization of the RNA at the time of blood draw (PAXgene™).

The application of microarray technologies to the problem of human autoimmunity is still in its infancy. As more experience is gained and the costs of arrays become more manageable, this will allow for the study of larger sample sizes in multiple clinical contexts. The data so far suggest that such studies will lead to the discovery of clinically useful biomarkers that can be used to guide diagnosis and management of these difficult disorders.

Acknowledgements

PK Gregersen, *New York, USA*
TW Brehrens, *Minneapolis, USA*

References

1 Chung CH, Bernard PS, Perou CM. Molecular portraits and the family tree of cancer. *Nat Genet* 2002; **32** (Suppl 2): 533–540.

2 Staudt LM. Gene expression profiling of lymphoid malignancies. *Annu Rev Med* 2002; **53**: 303–318.

3 van de Vijver MJ, He YD, van't Veer LJ *et al*. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; **347**: 1999–2009.

4 Firestein GS, Pisetsky DS. DNA microarrays: boundless technology or bound by technology? Guidelines for studies using microarray technology. *Arthrit Rheum* 2002; **46**: 859–861.

5 Han G, Chen S, Shen N *et al*. Analysis of gene expression profiles in human systemic lupus erythematosus using oligonucleotide microarray. *Genes Immun* 2003; **4**(3): 177–186.

6 Slonim DK. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet* 2002; **32** (Suppl 2): 502–508.

7 Crow M, George S, Paget SA *et al*. Expression of an interferon-alpha gene program in SLE. *Arthrit Rheumat* 2002; **46**: S281.

8 Baechler E, Batliwalla F, Karypis G *et al*. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe SLE. *Proc Natl Acad Sci USA* 2003; **100**: 2610–2615.

9 van der Pouw Kraan T, van Gaalen FA, Huizinga TW *et al*. Discovery of distinctive gene expression profiles in rheumatoid synovium using cDNA microarray technology: evidence for the existence of multiple pathways of tissue destruction and repair. *Genes Immun* 2003; **4**(3): 187–196.

10 Weinblatt ME, Kremer JM, Bankhurst AD *et al*. A trial of etanercept, a recombinant tumor necrosis factor receptor: Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. *N Engl J Med* 1999; **340**: 253–259.