

## Appendix A14.22

### The strategic profiling of gene expression changes to identify long term targets for mood stabilizing agents

As mentioned earlier, differential display technology was used to discover the upregulation of bcl-2 by lithium and valproate (via the identification of the upregulation of a transcription factor known to regulate bcl-2--PEBP2 $\beta$ ). The discovery of this unexpected target, which has had a major influence on our thinking about the long-term cellular effects of mood-stabilizers, illustrates the importance of hypothesis-generating (as opposed to hypothesis-dependant) techniques. Microarrays are another example of this type of research tool, and have already proven quite valuable in the study of psychiatric drugs. As the technology improves, current problems are eliminated, and new applications are developed, microarrays are likely to become an essential, indispensable tool for the neuroscience-psychiatric community. The methodological issues of microarray experiments have been extensively reviewed, but are discussed here briefly.

It should be noted that microarray technology should not necessarily be considered to be superior to, or a replacement of, differential display. Rather, each technique has unique advantages and disadvantages. Perhaps most obvious is the identification of transcripts: using a cDNA microarray, one is relatively certain what each differentially regulated mRNA is; with DD, on the other hand, one must clone and identify the transcript in question. While both methodologies require further validation to eliminate false positives, the immediate identification of the differentially expressed transcripts by the microarray methodology translates into a major savings in terms of time and labor. However, this very advantage of the microarray technology is also one of its weaknesses compared to the DD methodology; thus, the microarray methodology only allows for the identification of known transcripts (which the array in question comprises of); by contrast, DD allows the detection of entirely novel mRNA transcripts, and is thus truly unbiased. Oligonucleotide arrays are more like differential display, in that novel targets may be identified, and that the identity of the transcript binding to a given oligo must be determined.

One key difficulty in the use of microarrays lies in the signal-to-noise ratio. Particularly for transcripts of low abundance, the test-retest reliability of microarray results can be dismal. For this reason, repetition is of obvious importance; noise can be filtered, and variance can be used to assess the validity of results. Unfortunately, the high cost of microarrays often forces

researchers to forgo replication entirely, and thereby jeopardize the usefulness of their results. Likewise, the commonly-used approach of establishing a threshold of what magnitude of expression ratio constitutes a significant change fails to address the issue of deviation and error. As the reliability increases, and the cost decreases, microarray data will hopefully become subject to more statistically-rigorous analyses.

Even given the importance of replication, the usefulness of conventional statistical measures, such as the t-test, is another concern. Given a set of 10,000 transcripts, it is clear that one can expect ~500 false positives, if one uses a  $p < 0.05$  cutoff. On the other hand, if one uses Bonferroni correction for repeated analyses, one is all but guaranteed not to find any significantly regulated mRNA, and thus to acquire numerous false negatives.

In this light, the stringency with which microarray data is analyzed becomes a much fuzzier matter, and may be adjusted to address the specific question under investigation. While at first glance one may question the scientific rigor of such an approach, it needs to be emphasized that the microarray methodology is only a screening technique, and the results require much more independent validation. Thus, the cut-off threshold criteria should be determined not only by statistical considerations, but also the stringency and rigor of the experimental paradigm (e.g. identifying common targets of structurally highly dissimilar drugs), and the ability/willingness to subsequently validate positive results with independent methodologies. For example, if an experimenter hopes to discover single genes which are the target of the manipulation, high stringency is probably warranted (e.g.,  $p < 0.05$  and effect size of two-fold). This insures that less time will be wasted on false negatives at the PCR and/or protein validation stage. If, on the other hand, the goal is to find groups of related genes which are affected, or common targets of multiple agents, lower stringency may be more desirable. For example, if all 12 subtypes of a given receptor are upregulated 30% with p-values ~0.10, it may represent a very important finding, which would have been discarded with higher stringency.

Besides the t-test and fold-difference comparisons, more powerful means of analyzing microarray data are emerging. Clustering is a fairly common technique in which transcripts are associated based on their co-regulation across a number of samples. "Fingerprinting" refers to the association of a general pattern of microarray results with a particular variable. For example, this technique has already begun to show usefulness in oncology, where microarray data from biopsies can help predict treatment response and outcome. Likewise, ex vivo gene expression data psychiatric patients may one day predict optimal treatment strategies, risk of relapse, etc.

