Published in final edited form as:

Sci Signal. ; 1(51): eg11. doi:10.1126/scisignal.151eg11.

Microarrays Need Phylogenetics

Mones Abu-Asab

In the editorial on microarrays technology (1), Yaffe has voiced what many researchers have started to come to terms with, that better methods of analysis and modeling are needed for microarray data to be meaningfully interpreted. However, what is sorely missing from most mathematical and statistical analytical methods is their compatibility with core evolutionary concepts--the unifying framework of the biological sciences. Genetic modifications and their expansions within diseased tissue are evolutionary events at the cellular and tissue levels and should be analyzed as such. Furthermore, there are several characteristics that are fundamental to any efficient and meaningful analytical tool that, if applied to microarray data, will make the investment in microarrays justifiable, cost-effective, and rewarding. Here, I briefly list a few of these features that I feel are a must for better data modeling.

To construct evolution-compatible analytical tools, bioinformaticians should revisit the phenetic versus phylogenetic debate of the last century and learn from it why phylogenetics won among microbiologists, molecular biologists, and systematists. There are many issues that are better addressed by phylogenetics than by other paradigms. Take, for example, data heterogeneity, which is rampant throughout microarray and other high-throughput data. Genetic heterogeneities, both inter- and intra-populational, are most effectively analyzed and modeled with parsimony phylogenetics. Because the modeling is data-based rather than specimen-based, parsimony phylogenetics produces a highly predictive class discovery (classifying specimens in natural groupings that reflect their relatedness) based on the most parsimonious distribution of normal and altered gene expression. At the same time, it minimizes homoplasies (parallelisms and reversals), and produces a graphical tree (the cladogram), which makes it easy to sort out clonal and nonexpanding gene expression aberrations.

Clonal genetic changes within a population of diseased specimens that are responsible for disease initiation, progression, and maintenance need to be distinguished from nonexpanding changes in gene expression when deciphering the disease process (2). A phylogenetic cladogram maps the locations of both the clonal and nonexpanding expression alterations; the first are shared by a large number of specimens and are used in the circumscription of a group of specimens (the clade), while the second is randomly shared and/or restricted to a very small number of specimens. Because it reflects a hierarchical classification, a cladogram also reveals the direction of change accumulation among the study specimens and may also uncover multiple developmental pathways through gene-linkage of clonal changes shared by major clades.

In several parsimony analyses of microarray data performed by our lab, the cladograms showed transitional specimens that fell between the normal and diseased specimens; this makes the cladogram a tool for studying disease boundaries and definitions (another issue that needs deeper study). Thus far, there is no other analytical method that is capable of recognizing intermediate forms between healthy and diseased states.

Phylogenetic analysis generates unique insights, because it takes advantage of the qualitative aspect of microarray data. The qualitative nature of gene expression data can be explored through phylogenetic modeling; it incorporates the directionality of gene expression, as well

as the complex patterns of expression like those that violate normal distribution among a number of specimens. The latter is usually ignored in a quantitative statistical analysis, although it is indicative of pathway diversity (3).

Microarray data, when analyzed with an evolution-compatible tool, such as parsimony phylogenetics, can yield biologically meaningful results with a number of interesting implications. Parsimony phylogenetics provides disease modeling through an assessment of ontogenic pathways and phyletic relatedness of specimens on the basis of shared derived expression states. Researchers should give parsimony phylogenetics a try before giving up on microarrays.

References

1. Yaffe MB. Beyond microarrays. *Sci Signal* 2008;1:eg11. [PubMed: 19109235]
2. Bielas JH, Loeb KR, Rubin BP, True LD, Loeb LA. Human cancers express a mutator phenotype. *Proc Natl Acad Sci USA* 2006;103:18238–18242. [PubMed: 17108085]
3. Abu-Asab MS, Chaouchi M, Amri H. Phylogenetic modeling of heterogeneous gene-expression microarray data from cancerous specimens. *OMICS* 2008;12:183–199. [PubMed: 18699725]