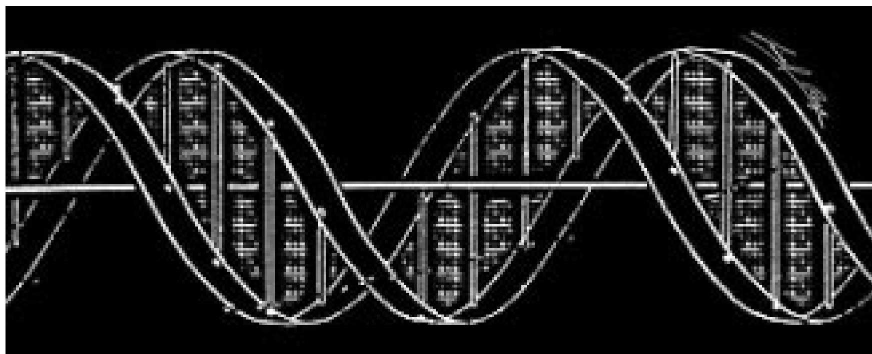
Toxicogenomics: The New Frontier In Toxicology and Environmental Health Science

Georgette M. Charles

Niched tightly between the Engineering and Biology departments, lies the Division of Bioengineering and Environmental Health (BEH). It is described as the merger of engineering and the biological sciences. Within the division, a particular group of scientists are now bubbling with excitement. These scientists are known as toxicologists, and their excitement is over toxicogenomics, an application of microarray technology that promises to bring toxicology to a new frontier.



Microarray analysis just became a widely used biological tool in the late 1990s. It uses silicon or glass chips, capable of being coated with dense, discrete spots of nucleic acids, which are termed "probes." A "target," fluorescent cDNA, RNA, or cRNA, can then be bound to the array of probes to determine the gene expression pattern–which genes' RNA products were indeed present in the "target" sample.[1] The target binds to the probe as a result of nucleic acid base pairing, also known as hybridization. This seemingly simple procedure allows scientists to screen tens of thousands of genes for an expression or particular sensitivity after cells have been subjected to a particular environmental stimulus. Toxicogenomics has taken advantage of this very principle.

Technology in the Hands of Toxicologists

In the field of toxicogenomics, microarrays are used to make research easier, systematic, more efficient, and faster. Toxicologists have the responsibility of establishing or disproving whether

exposure to molecules, including industrial and consumer by products, has a cause-effect relationship with human disease. Consider the example of the connection between cigarette smoke and lung cancer. Toxicologists have established both physiological and statistical data, piling a mile high, proving the relationship between cigarette smoke and the development of lung cancer. Presently, given the analytical power of microarrays, toxicologists can do the following: determine whether there are discrete gene changes linking cigarette smoke exposure with lung dysfunction and disease; assess whether an uncharacterized agent in cigarettes can be harmful by matching data with that of known toxicants; and assess drug sensitivities and effectiveness by identifying lung disease biomarkers, commonly occurring genetic traits in instances of bodily dysfunction, specifically tissue or organ dysfunction.[2] The primary objective of toxicogenomics is to identify and catalog the genomic expression fingerprint of a particular chemical or environmental agent. After cells are subjected via in vitro assay to a particular agent, the target cDNA, RNA, or cRNA is obtained, fluorescently tagged, and added to a microarray containing probes. The intensity of fluorescence detected under a microscope correlates with the amount of hybridization between the added target and the probe. All of the genes with detectable fluorescence can be evaluated as having sensitivity to the agent, by virtue of the gene expression pattern. Collectively, all of the genes with marked sensitivity would be included in this agents' Global Gene Expression Profile (GGEP).

So why might a GGEP be useful to other scientists, not only those in the field of toxicology? In order to assess this importance, let us first consider the fundamental impact that a GGEP has on toxicology. In the field of toxicology, scientists use many sensitivity assays and screens to profile the reaction of an organism or cell culture to a putative toxicant. They collect data for chronic, subchronic, subacute, and acute doses. By understanding the GGEP of a putative toxicant, it can be compared to and grouped with other known toxicants. For instance, if the putative toxicant had a similar GGEP to a known intoxicant with minimal to no chronic effects, some longer studies on the putative toxicant could be eliminated, reducing some of the project costs. This is but one illustration of the use of a GGEP and how, in short, it allows toxicologists to get a handle on the general classification of toxic agents.

Beyond the frontier of toxicogenomics and the GGEP output, scientists, particularly biologists and chemists, use this knowledge for new research opportunities within their own fields, whether it involves biochemistry, molecular and cellular biology, or organic chemistry. For example, suppose a toxicant's GGEP shows an overexpression of anti-apoptotic genes, those responsible for preventing inappropriate programmed cell death when it isn't necessary in cells. Additionally, suppose the toxicant was known to significantly increase the number of cases of leukemia in a population exposed to this toxicant via groundwater. This data might suggest important evidence linking leukemia to the over expression of anti-apoptotic genes. Clearly, biologists and biochemists studying leukemia would be interested in such data. With all of the new avenues that toxicogenomics is providing for other research fields, it remains to be mentioned that other fields have the ability to push this field forward.

Advances in Microarray Technology Evolving Toxicogenomics

One technique that has already proved important to toxicogenomics is the creation of tissue-specific gene arrays. A few to several thousand genes in a given tissue can be screened on a microarray. Such selectivity relative to the total number of genes present is important for analysis of diseases that affect or cause injury to a particular tissue or organ. To narrow this analysis even more, gene array filters can be created for articular class or classes of genes to be screened within the tissue-specific gene array. Investigators can pick and choose which genes are most applicable to their study and use only those selected for the microarray. Consider the example of air pollutant–induced lung injury that was investigated by Nadadar et al. They were interested in genes that were responsible for responses to injury. Specifically, out of all the genes present in the lung tissue, they selected an array of twenty-seven genes including inflammatory and anti-inflammatory cytokines, growth factors, adhesion molecules, stress proteins, transcription factors, and antioxidant enzymes.[3] So what experimental power would such a selective filter have for the whole field of toxicogenomics? Well, let's delve beyond the surface of cell culture and look at in vivo experiments. Toxicologists often have to deal with added complexity when it comes to modeling cell culture according to real-time physiological changes in organisms. Due to this difficulty, systems toxicologists use live animal models to investigate gene expression profiles. The animal is subjected to the chemical or

environmental stimulus in an acute or chronic manner. Tissue-specific arrays are then created as appropriate to the disease or toxicity under investigation. For this reason filters are needed to avoid having to screen every gene in each important tissue or organ relevant to the type of disease or localized toxicity. Despite the fact that toxicogenomics is a new field of research, its precursor, microarray technology, continues to evolve rapidly, bringing the possibilities for this research to new levels. Recently, scientists from MIT's Whitehead Institute, Junaid Ziauddin and David Sabatini, published a report in Nature, describing an experimental technique that uses whole cells with defined cDNAs as probes on microarrays. In this process, termed "reverse transfection," live cells are cultured on a glass slide with the cDNA probes in expression vectors, printed in defined locations. As the cells grow, they take up the DNA, making "spots of localized transfection within a lawn of nontransfected cells."[4] This procedure can be used both to identify drug targets and as a cloning system to discover gene products that change cellular physiology.[4] Looking at which genes are turned on by the stimulus provides investigators with preliminary information about putative targets of the stimulus. Also, by identifying very specific physiological changes in the treated cells on the array, the corresponding genes in accordance with the location of the cells can be extrapolated and assayed further. Toxicogenomics could benefit from this technique, especially in regards to the aforementioned application.

The key determinant in the application of "reverse transfection" to toxicology is that cells can be cultured first then added to the array of cDNAs. Toxicologists could culture the cells in media containing a chemical. The array would then determine which genes are expressed immediately, as in an acute response assay, and which are expressed over a long treatment time, a more chronic study. Essentially, toxicologists would have a handle on immediate reponses and could easily assay cells after determining the chemical's GGEP via this method.

Thus far, we know that whole-cell "reverse transfection" arrays and tissue-specific arrays carry great experimental power for toxicogenomics. But let's take it a step further. What if these applications were combined–would it be advantageous for toxicogenomics? A tissue-specific array using reverse transfection would change the scope of toxicogenomics and help toxicologists narrow their research in identifying tissue-specific responses to external stimuli. As mentioned before, reverse transfection can be extremely helpful in identifying drug targets and acting as a cloning system to discover gene products that cause physiological changes in cells. Combined, what we have is an array that helps pinpoint the gene targets of a particular chemical, drug, or stimulus within the context of a particular tissue. As technology develops, more experimental analysis will be feasible for toxicologists. But a question remains: Can data be consistently analyzed from one experiment to another? Here we must turn to bioinformatics.

Bioinformatics Bridges the Gap

Whereas in most research fields the analytical weight of the experiment lies with the design, this is not the case for toxicogenomics. Equal weight is given to the experimental design and the statistical analysis of the data. Bioinformatics is a research area that, in general, develops algorithms to process, manage, and present a vast amount of biological data.[5] For microarray technology and, more specifically, toxicogenomics, bioinformatics is responsible for creating necessary tools for statistical analysis.

The first critical task to be completed is processing the microarray data. How can a machine such as a scanner or computer read a microarray containing nucleic acids or live cells? It all has to do with fluorescence. In the case of nucleic acid arrays, the cDNA, mRNA, or cRNA targets are fluorescently labeled prior to complementary base-pairing to the probe. As for whole-cell arrays of reverse transfection, the expression vector with the target gene usually contains the green fluorescent protein (GFP) gene, fused downstream of the gene of interest and its promoter. For example, cells are cultured with a stimulus, then transfected by an expression vector containing a growth hormone. If the stimulus causes increased cellular production of this hormone, then this gene would be transcribed. GFP would be transcribed after this gene, and the same relative protein production would result according to the growth factor. The fluorescence intensity at each location of the microarray chip is then scanned by a microscope or camera.[1]

After scanning the microarrays, the toxicologist meets the critical step, matching each fluorescence intensity relative to every other one. Bioinformatics is essential for creating the computer software to handle this task. In order to analyze the qualitative data scanned from the microscope or camera, it must be translated into quantitative or numerical statistics. The software that toxicologists use therefore relies heavily on mathematical algorithms. The basic analysis that can be done without a computational strategy is

to note the observed signal. Since "the amount of probe on the array exceeds the amount of target, the observed signal at any given position is a good estimate of the abundance of cognate target in the sample."[1] To turn this relatively simple procedure into an algorithm requires more than a mathematics background. Rather, a complex understanding of statistics and regression analysis is needed for reliable recording of microarray data.

Now we are faced with two very critical questions. What constitutes a signal above background and how can different scans of the same microarray yield the same quantitative reading? Let us first look at signals and their background levels. One example of how bioinformatics specialists try to treat the problem is to scan a region without a bound nucleic acid target and subtract this reading from every data point.[1] Many other solutions have been developed and are still being worked out.

A concept that may be easier to understand is normalization, which is critical for scan-to-scan accuracy, yielding the same quantitative readings after each scan. The first method to normalize the data from one scan to another is to use appropriate controls whose signal is not expected to vary to normalize the scan intensity.[1] Essentially, normalization involves subtracting a baseline value from all of the sample values. The second method is to calculate the average overall intensity for all elements on the microarray and to use this average to normalize signal strength.[1] Beyond these complex computational tasks, the challenge remains to catalog all of the data for science to use.

Even though the above discussion brings light to normalization for one-color signals, it should be noted that it is most common to use two-color scanning (i.e., two fluorescence colors) to minimize experimental error. Currently, the most often used method is a Cy3/Cy5 ratio. This entails labeling the target with Cy3 and the probe with Cy5.[6] Based on the ratio of these signals, the relative abundance of the target is determined. Another solution to normalization is a standard probe set that is used for all experiments and minimizes variance from one experiment to another. Both of these techniques attack the problem of experimental variation, and additional techniques similar to these are currently under development.

Data management is very much a concern for the National Institute of Health (NIH). Specifically, the National Institute of Environmental Health Science (NIEHS) has established the National Center for Toxicogenomics (NCT), which seeks to create a database to catalog the plethora of toxicogenomics data and make it available to the public. With such a database, toxicogenomics will provide the global gene expression profiles to support other research. It is in this context that toxicogenomics must be recognized as a field moving forward, and, in exchange, its advancement will help catapult other sciences forward.

Despite all of these promising developments still it should be noted that this science is not perfect and still needs a lot of work. There are many intricate details that have to be figured out, such as elimination of erroneous data (i.e., "false positives"), undetected gene expressions (i.e., "false negatives"), conflicting data from different sites of identical experimentation, biological variation limitations, and, most importantly, human experimental error, which haunts all research fields. Given these problems, there are many avenues for scientific exploration and discovery. Hopefully more scientists will lend their expertise to this now-budding field.

Special thanks to Professor James Sherley of the Divison of Bioengineering and Environmental Health (BEH). It was his introduction to this subject in BEH.104J that encouraged me to research the topic further. Additionally, I would like to thank Professor Leona Samson of BEH, who directed me to useful information resources.

References:

1. Winzeler EA *et al.* "Fluorescence-based expression monitoring using microarrays." *Methods in Enzymology* 1999; 306:3-18.
2. NIEHS Division of Extramural Research and Training Web site: http://www.niehs.nih.gov/dert/programs/tgintro.htm
3. Nadadur SS *et al.* "A pulmonary rat gene array for screening altered expression profiles in air pollutant-induced lung injury." *Inhalation Toxicology* 2000; 12:1239-1254.
4. Ziauddin J and Sabatini D. "Microarrays of Cells Expressing Defined cDNAs." *Nature* 2001; 411:107-110.
5. NIEHS Microarray Center Web site: http://dir.niehs.nih.gov/microarray/bioinfo2.htm
6. Samartzidou H *et al.* "Lucidea Microarray ScoreCard: An integrated tool for validation of microarray gene expression experiments." Web site: http://www.apbiotech.com/product/publication/lsn/lsn8/lsn8-1.html

Recommended Readings:

1. Winzeler E A *et al.* "Fluorescence-based expression monitoring using microarrays." *Methods in Enzymology* 1999; 306:3-18.
2. Ziauddin J and Sabatini D. "Microarrays of Cells Expressing defined cDNAs." *Nature* 2001; 411:107-110.
3. NC160 Cancer Microarray Project Web site: http://genome-www.stanford.edu/nci60/explorer.shtml