Assessing the Functional Bias of Commercial Microarrays Using the Onto-Compare Database

Sorin Draghici[1], Purvesh Khatri[1], Abhik Shah[1], and Michael A. Tainsky[2]

[1]Department of Computer Science, Wayne State University, and [2]Molecular Biology and Genetics, Karmanos Cancer Institute, Detroit, MI, USA

ABSTRACT

Microarrays are at the center of a revolution in biotechnology, allowing researchers to screen tens of thousands of genes simultaneously. Typically, they have been used in exploratory research to help formulate hypotheses. In most cases, this phase is followed by a more focused, hypothesis-driven stage in which certain specific biological processes and pathways are thought to be involved. Since a single biological process can still involve hundreds of genes, microarrays are still the preferred approach as proven by the availability of focused arrays from several manufacturers. Because focused arrays from different manufacturers use different sets of genes, each array will represent any given regulatory pathway to a different extent. We argue that a functional analysis of the arrays available should be the most important criterion used in the array selection. We developed Onto-Compare as a database that can provide this functionality, based on the Gene Ontology Consortium nomenclature. We used this tool to compare several arrays focused on apoptosis, oncogenes, and tumor suppressors. We considered arrays from BD Biosciences Clontech, PerkinElmer, Sigma-Genosys, and SuperArray. We showed that among the oncogene arrays, the PerkinElmer MICROMAX™ Oncogene Microarray has a better representation of oncogenesis, protein phosphorylation, and negative control of cell proliferation. The comparison of the apoptosis arrays showed that most apoptosis-related biological processes are equally well represented on the arrays considered. However, functional categories such as immune response, cell-cell signaling, cell-surface receptor linked signal transduction, and interleukins are better represented on the Sigma-Genosys Panorama™ human apoptosis array. At the same time, processes such as cell cycle control, oncogenesis, and negative control of cell proliferation are better represented on the BD Biosciences Clontech Atlas Select™ human apoptosis array.

INTRODUCTION

Microarrays have been introduced as powerful tools that can efficiently screen a large number of genes. The typical result of a microarray experiment is a number of gene expression profiles, which in turn are used to generate hypotheses and locate effects on many, perhaps unrelated, pathways. This is a typical hypothesis-generating experiment. For this purpose, it is best to use comprehensive microarrays that represent as many genes of an organism as possible. Currently, such arrays include tens of thousands of genes. For example, the HG-U133 Set from Affymetrix (Santa Clara, CA, USA) contains 44 928 probes that represent 42 676 unique sequences from GenBank® database corresponding to 28 036 UniGene clusters.

Typically, after conducting a microarray experiment, one would select a small number of genes (e.g., 10–50) that are found to be differentially expressed. These genes are analyzed from a functional point of view, either going through online databases manually or by using an automated data mining tool such as Onto-Express (1,2). This step identifies the biological processes, molecular functions, biochemical functions, and pathways impacted in the condition under study and generates specific hypotheses involving them. In many cases, only a small number of pathways are identified, and the next logical step is to focus on these specific pathways.

In many cases, it is desirable to construct a molecular classifier that can diagnose or classify samples into different categories based on their gene expression profiles (3–9). This involves a training process that suffers from a "curse of dimensionality" (10), which refers to the fact that the difficulty of building such a classifier increases exponentially with the dimensionality of the problem (i.e., the number of genes involved). Furthermore, constructing a classifier requires many more training examples (samples or patients) than variables (genes). Both issues strongly suggest that the number of genes used to build the classifier has to be reduced to a minimum. In other words, it is best if the set of genes is restricted to strictly relevant genes.

Therefore, focusing on a smaller number of genes is both the logical step that follows the initial screening experiment that generated the hypotheses as well as a required step if molecular classifiers are to be constructed. Unlike the first step of an exploratory search in which the hypotheses are generated, the second steps should be a "hypothesis-driven experiment" in which directed experiments are performed to test a small number of very specific hypotheses. However, specific hypotheses and a small number of pathways may still involve hundreds of genes. This is still too many for reverse transcription PCRs, Western

blotting, or other gene-specific techniques, so the microarray technology is still the preferred approach.

Many commercial microarray manufacturers have realized the need for and have started to offer focused arrays. For instance, BD Biosciences Clontech (San Jose, CA, USA) currently sells focused human microarrays for the investigation of the cardiovascular system, cell cycle, cell interaction, cytokines/receptors, hematology, neurobiology, oncogenes, stress, toxicology, and tumors. Many other companies have followed the same trend and offered focused arrays (e.g., PerkinElmer [Boston, MA, USA], Takara Bio [Shiga, Japan], SuperArray [Frederick, MD, USA], and Sigma-Genosys [The Woodlands, TX, USA]). Literally tens of focused arrays are available on the market with several companies offering customized arrays for the same pathways. Typically, a focused array includes a few hundred genes covering the biological mechanism(s) of choice. However, two microarrays produced by different companies are extremely unlikely to use the same set of genes. Consequently, various pathways will be represented to various degrees on different arrays even if the arrays are all designed to investigate the same biological mechanisms. This is an unavoidable functional bias. Such a bias will be associated with each and all arrays that include less than the full genome of a given organism.

The general criteria used to select an array include several categories. One category is the availability of a particular array. For instance, one laboratory or core facility may have certain arrays readily available because they have been purchased in a larger lot, or because they remained from previous experiments. Another important criterion is cost. Even if the cost of the array itself is a relatively small component of the overall cost of the experiment, smaller arrays do tend to cost less, and spending more money for a larger array is usually expected to be associated with a benefit. Another set of factors influencing the choice of the array is related to technological preferences. For instance, a researcher might prefer cDNA to oligonucleotide arrays, filters to glass, and so on. Finally, the choice might be influenced by data analysis issues. For instance, certain normalization techniques, such as dividing by the mean of all genes on a given array, assume that most genes do not change. This is probably true for an array containing thousands of genes but may not be true for a smaller array. Furthermore, the data analysis may be easier and more reliable if fewer genes are present on the array. Considerably fewer genes means fewer difficulties related to correction for multiple experiments in statistical hypothesis testing, much less computation in model fitting (e.g., expectation maximization approaches), less of a curse of dimensionality in the construction of a classifier for diagnosis purposes and in general a better ratio between number of dimensions (number of genes) and number of data points (mRNA samples).

The interplay of these factors eventually decides the choice of the array when, in fact, the choice of the array should be made primarily based on the scientific question at hand. If array A contains 10 000 genes but only 80 are related to a given pathway and array B contains only 400 genes but 200 of them are related to the pathway of interest, the experiment may provide more information if performed with array B instead of A. Furthermore, using a smaller array can also translate into significant cost savings. Since an array can contain thousands of probes and a typical user would compare several arrays at once, annotating the probes and comparing the arrays becomes a daunting computational task. We built a tool, Onto-Compare, that allows a researcher to perform this task in a fast and convenient way using terms from the Gene Ontology (GO) Consortium (11,12).

MATERIALS AND METHODS

To assess the biological bias of various arrays, we designed and implemented a custom database. This Onto-Compare database was populated with data collected from several online databases, as well as lists of genes (GenBank accession numbers) for each microarray as provided by their manufacturers. From the list of accession numbers, a list of unique UniGene cluster identifiers is prepared for each microarray, and then a list of LocusLink identifiers is created for each microarray in our database. UniGene is a system for automatically partitioning GenBank mRNA sequences and expressed sequence tags (ESTs), into a nonredundant set of gene-oriented clusters with identical 3' untranslated regions (3' UTRs) (13). LocusLink
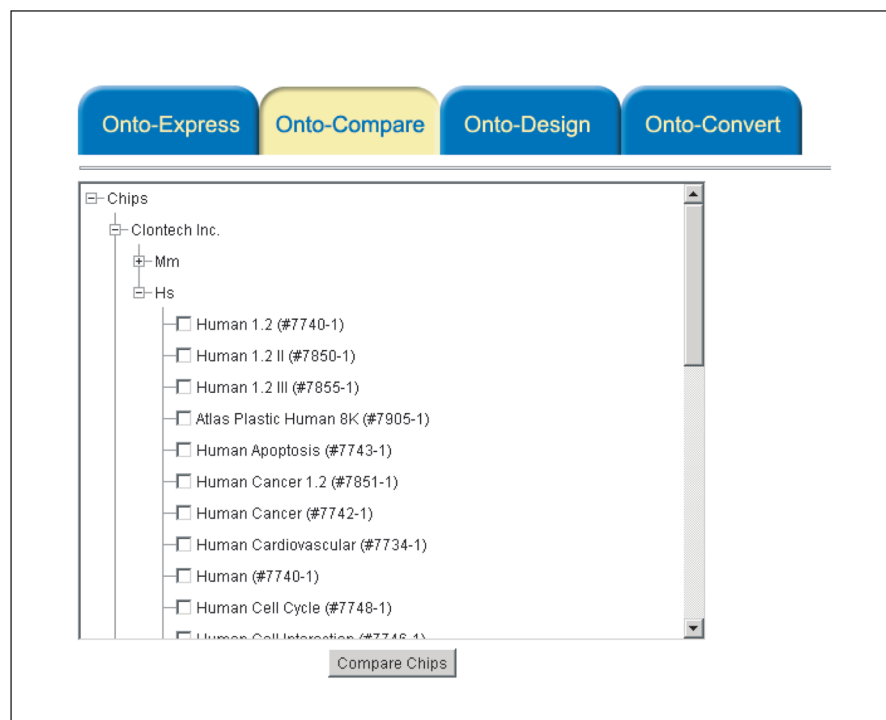

Figure 1. The input screen of Onto-Compare. The arrays are organized by manufacturer and organism. The user can select any number of arrays to be compared.

RESEARCH REPORT

(14) is a database of official gene names and other gene identifiers. Each locus in the LocusLink database is annotated using ontologies from the GO Consortium (15) and ontologies from other researchers and companies. The GO Consortium provides ontologies for biological processes, molecular function, and cellular components. The data from these databases and gene lists are parsed and entered into our Onto-Compare relational database. The Onto-Compare database is implemented in Oracle® using a schema designed to allow for efficient querying. A group of Java programs and Perl routines are used to download, parse, and enter the data into the Onto-Compare database as well as to update the database on a regular basis (currently done every two months) with minimal human intervention.

After creating a list of locus identifiers for each array, the list is used to generate the following profiles: biochemical functions, biological process, cellular role, cellular component, and molecular function. The profiles for each microarray are stored in the database. The list of genes deposited on a microarray is static, but the annotations for those genes keep changing and are updated automatically, as more information becomes available. A Java program is used to facilitate the update of the database, which recreates these profiles as more information becomes available.

An array can contain thousands of probes and a typical user would compare several arrays at once, so annotating the probes becomes a daunting computational task. For this reason, we precalculate the functional annotations for each array and store the results in the database. Thus, this very time-consuming computation is only done after database updates. During user interaction, the data is merely queried from the database. However, a researcher might choose to use a set of arrays (e.g., the Affymetrix HG-U133 is actually a set of two arrays). To accommodate this, we allow the user to merge arrays and calculate functional profiles of the sets of arrays. Since a user can merge an arbitrary number of arrays, we cannot realistically precalculate the functional annotation for every possible union of arrays. Thus, this computation is done every time a user merges arrays.

Onto-Compare runs as a Java® applet in a Web browser on a user computer. The input screen is shown in Figure 1. The results are presented as a table, in which the first column displays an ontology term and the rest of the columns corresponds to one of the arrays from the set of selected arrays (see Figure 2). The first row of the table represents the total number of unique GenBank sequences represented on each of the selected arrays. For each ontology term, Onto-Compare displays the total number of unique GenBank sequences found on each of the selected array. Onto-Compare also displays the total number of unique UniGene clusters found on each of the selected arrays for each ontology term in square brackets. Clicking an array name sorts the entire table by the total number of sequences for that array. Clicking on a value for total number of sequences, displays accession numbers of the sequences, their corresponding cluster ID and locus ID along with the cluster's official gene symbol.
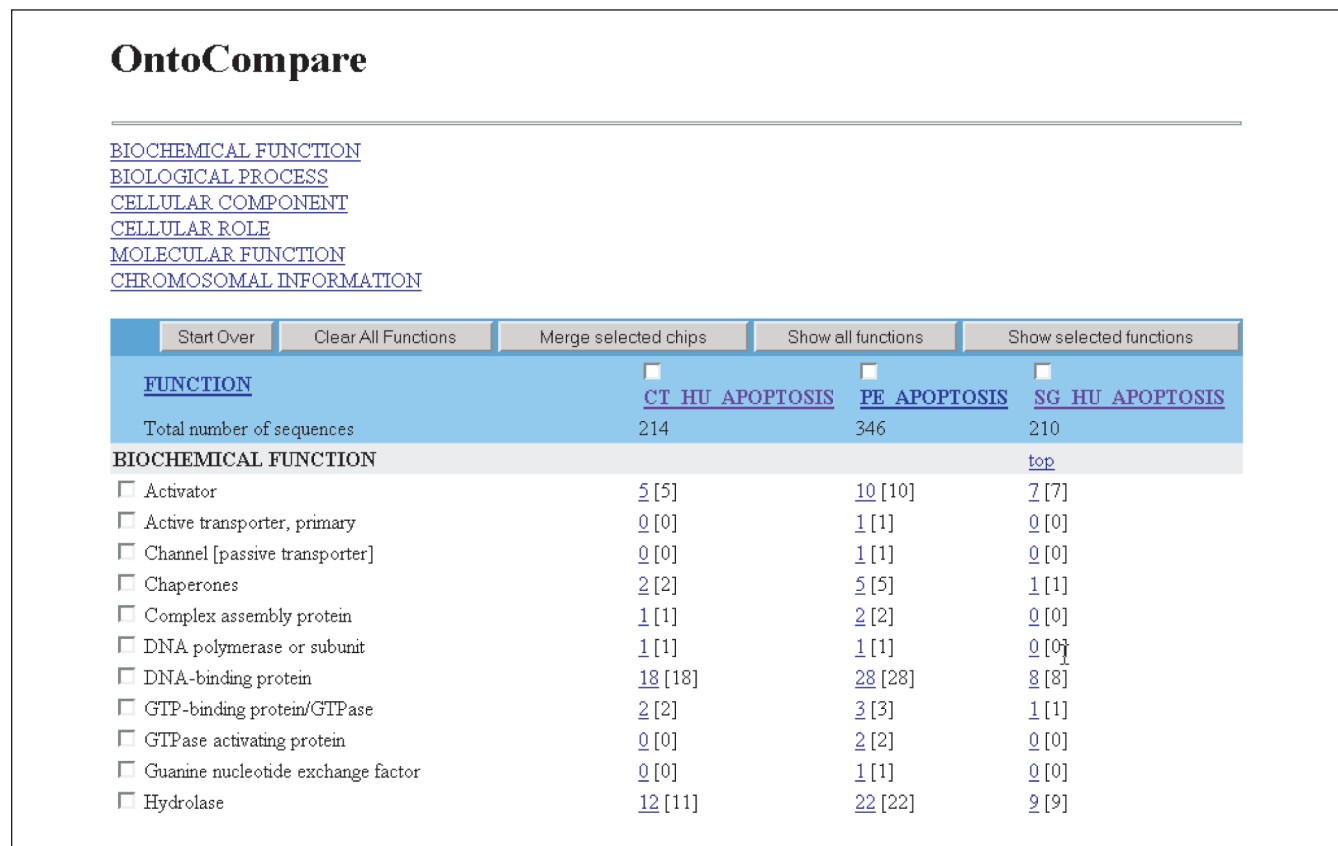

OntoCompare

BIOCHEMICAL FUNCTION
BIOLOGICAL PROCESS
CELLULAR COMPONENT
CELLULAR ROLE
MOLECULAR FUNCTION
CHROMOSOMAL INFORMATION

Start Over	Clear All Functions	Merge selected chips	Show all functions	Show selected functions

FUNCTION	CT_HU_APOPTOSIS	PE_APOPTOSIS	SG_HU_APOPTOSIS
Total number of sequences	214	346	210
BIOCHEMICAL FUNCTION			top
Activator	5 [5]	10 [10]	7 [7]
Active transporter, primary	0 [0]	1 [1]	0 [0]
Channel [passive transporter]	0 [0]	1 [1]	0 [0]
Chaperones	2 [2]	5 [5]	1 [1]
Complex assembly protein	1 [1]	2 [2]	0 [0]
DNA polymerase or subunit	1 [1]	1 [1]	0 [0]
DNA-binding protein	18 [18]	28 [28]	8 [8]
GTP-binding protein/GTPase	2 [2]	3 [3]	1 [1]
GTPase activating protein	0 [0]	2 [2]	0 [0]
Guanine nucleotide exchange factor	0 [0]	1 [1]	0 [0]
Hydrolase	12 [11]	22 [22]	9 [9]

Figure 2. A sample output screen of Onto-Compare. The functional categories analyzed are biochemical function, biological process, cellular component, cellular role, molecular function, and chromosomal location. The user can further merge arrays (e.g., for HG-133A and B) or focus the analysis on a subset of functional categories.

Table 1. A Comparison of Three Apoptosis-Specific Microarrays

Ontology Term	BD Biosciences Clontech	PerkinElmer	Sigma-Genosys
Total genes on array	214	346	210
Induction of apoptosis	16 [16]	27 [26]	23 [23]
Anti-apoptosis	15 [15]	20 [20]	21 [21]
Immune response	0 [0]	1 [1]	19 [19]
Cell-cell signaling	9 [9]	9 [9]	18 [18]
Cell surface receptor linked signal transduction	4 [4]	9 [9]	17 [17]
Oncogenesis	22 [22]	29 [29]	16 [16]
Regulation of cell cycle	30 [30]	30 [30]	12 [12]
Positive regulation of cell proliferation	5 [5]	5 [5]	12 [12]
Negative regulation of cell proliferation	16 [16]	20 [20]	10 [10]
Induction of apoptosis by DNA damage	3 [3]	4 [4]	3 [3]
Induction of apoptosis by extracellular signals	8 [8]	12 [12]	7 [7]
Induction of apoptosis by hormones	1 [1]	1 [1]	1 [1]
Induction of apoptosis by intracellular signals	2 [2]	2 [2]	2 [2]
Induction of apoptosis by oxidative stress	0 [0]	0 [0]	1 [1]
Induction of apoptosis via death domain receptors	4 [4]	5 [5]	7 [7]
Caspase	1 [1]	3 [2]	2 [2]
Caspase activator	1 [1]	3 [3]	2 [2]
Caspase-1	1 [1]	1 [1]	1 [1]
Caspase-10	1 [1]	2 [1]	1 [1]
Caspase-2	1 [1]	1 [1]	1 [1]
Caspase-3	1 [1]	1 [1]	1 [1]
Caspase-4	1 [1]	1 [1]	1 [1]
Caspase-8	2 [1]	1 [1]	1 [1]
Caspase-9	1 [1]	1 [1]	1 [1]
Caspase-activated deoxyribonuclease	1 [1]	2 [2]	2 [2]
Tumor necrosis factor receptor	2 [2]	2 [2]	2 [2]
Tumor necrosis factor receptor ligand	1 [1]	1 [1]	1 [1]
Tumor necrosis factor receptor, type I	1 [1]	1 [1]	1 [1]
Interleukin receptor	0 [0]	0 [0]	2 [2]
Interleukin-1 receptor	0 [0]	0 [0]	2 [2]
Interleukin-1, Type I, activating receptor	0 [0]	0 [0]	1 [1]
Interleukin-10 receptor	0 [0]	0 [0]	1 [1]
Interleukin-12 receptor ligand	0 [0]	0 [0]	2 [2]
Interleukin-2 receptor	0 [0]	0 [0]	3 [3]
Interleukin-2 receptor ligand	0 [0]	0 [0]	1 [1]
Interleukin-4 receptor	0 [0]	0 [0]	2 [2]
Interleukin-4 receptor ligand	0 [0]	0 [0]	1 [1]
Interleukin-7 receptor	0 [0]	0 [0]	1 [1]
Unique sequences [clusters]	105 [104]	142 [141]	130 [130]

Biological processes such as induction of apoptosis and molecular functions such as caspases and tumor necrosis are almost equally represented on each of the chips, but none of the interleukins are represented on BD Biosciences Clontech or PerkinElmer microarrays. Processes such as immune response, cell-cell signaling, cell surface receptor linked signal transduction are better represented on the Sigma-Genosys array. Processes such as cell cycle control, oncogenesis, and negative control of cell proliferation are better represented on the BD Biosciences Clontech array. The numbers represent sequences present on the arrays; the numbers in brackets represent distinct UniGene clusters.

Table 2. A Comparison of Commercially Available Arrays

Biological Process	BD Biosciences Clontech Oncogene	BD Biosciences Clontech Tumor Suppressor	PerkinElmer MICROMAX	SuperArray Oncogene
Total genes on array	514	199	335	26
Oncogenesis	23 [22]	38 [38]	65 [63]	8 [8]
Signal transduction	47 [47]	39 [39]	49 [49]	6 [6]
Cell proliferation	23 [23]	19 [19]	23 [23]	7 [7]
Protein phosphorylation	6 [6]	17 [17]	19 [19]	5 [5]
Regulation of cell cycle	20 [20]	30 [30]	16 [16]	1 [1]
Negative regulation of cell proliferation	5 [5]	15 [15]	16 [15]	0 [0]
Cell-cell signaling	10 [9]	8 [8]	11 [11]	0 [0]
Unique sequences [clusters]	88 [87]	101 [101]	125 [123]	13 [13]

A comparison of the Atlas Select Human Oncogene 7831-1 and Human Oncogene/Tumor Suppressor 7745-1, MICROMAX, and SuperArray Human Cancer/Oncogene arrays. MICROMAX is a better choice for conditions potentially involving oncogenesis, protein phosphorylation, and negative control of cell proliferation. The numbers represent sequences present on the arrays; the numbers in brackets represent distinct UniGene clusters.

Table 3. A Comparison of Sets of Arrays

Biological Process	BD Biosciences Clontech Oncogene and Tumor Suppressor	PerkinElmer MICROMAX	SuperArray Oncogene
Total number of sequences	676	335	26
Signal transduction	82 [82]	49 [49]	6 [6]
Oncogenesis	55 [54]	65 [63]	8 [8]
Regulation of cell cycle	42 [42]	16 [16]	1 [1]
Cell proliferation	38 [38]	23 [23]	7 [7]
Developmental processes	21 [21]	13 [13]	3 [3]
Protein phosphorylation	21 [21]	19 [19]	5 [5]
Negative control of cell proliferation	19 [19]	16 [15]	0 [0]
Cell-cell signaling	18 [17]	11 [11]	0 [0]
Signal transduction	10 [10]	6 [6]	0 [0]
Unique sequences [clusters]	175 [174]	125 [123]	13 [13]

A comparison of the BD Biosciences Clontech set of arrays (Atlas Select Human Oncogene 7831-1 and Human Oncogene/Tumor Suppressor 7745-1) with the PerkinElmer, and SuperArray (Human Cancer/Oncogene) arrays. When the two BD Biosciences Clontech arrays are used together, there is a good representation of general signal transduction, protein phosphorylation, and negative control of cell proliferation. However, oncogenesis is still better represented on the PerkinElmer array. The numbers represent sequences present on the arrays; the numbers in brackets represent distinct UniGene clusters.

Selecting check boxes and then clicking "show selected functions" only displays the selected terms. The user can merge two or more arrays, by selecting the check boxes next to the array names and clicking the "merge selected arrays" button.

RESULTS AND DISCUSSION

To illustrate the utility of Onto-Compare, we can consider the example of an anticancer drug candidate that inhibits *bcl2*, which is an anti-apoptotic factor. The drug candidate was obtained from a large-scale screening so the exact mechanism through which this drug promotes apoptosis is not yet known. We would like to use a microarray approach to study the interaction between the various genes and their respective proteins on the apoptotic pathway as this pathway is affected by the drug. This is typical example of a hypothesis-driven research. We have strong reasons to believe that a certain pathway or a cellular mechanism is involved and we would like to focus our experiments on it.

There are several companies that provide arrays designed for

research involving apoptosis. The Atlas™ human apoptosis array contains probes for 206 UniGene clusters; the MICRO-MAX™ human apoptosis microarray contains probes for 324 UniGene clusters, and the Panorama™ human apoptosis microarray contains probes for 198 UniGene clusters. There are 74 clusters common among all arrays. There are 167 clusters common between the Atlas and MICROMAX arrays, 92 clusters common between MICROMAX and Panorama arrays, and 82 clusters common between Atlas and Panorama arrays. Comparison of these three apoptosis-specific arrays using Onto-Compare is shown in Table 1. A biological process such as induction of apoptosis and a molecular function such as caspases are clearly relevant to the apoptosis (16–18). Indeed, the table shows that all three arrays have similar numbers of genes representing induction of apoptosis, caspases, and tumor necrosis factor receptors. Various interleukins are also reported to be mechanistically associated with apoptosis at both protein and gene levels (19–21). However, neither Atlas nor MICROMAX microarrays contains any interleukin-related genes. On the other hand, the Panorama human apoptosis microarray contains 14 genes related to various interleukins. Clearly, among the 3 arrays considered, the Panorama would be a better choice for testing any hypothesis involving the role of interleukins in apoptosis. Other processes such as immune response, cell-cell signaling, and cell surface receptor linked signal transduction are also better represented on the Panorama array. However, it is important to emphasize that the Panorama array is not necessarily better than the other two arrays. In fact, since the Panorama and the Atlas arrays have almost the same number of genes, there must exist some functional categories that are better represented on the Atlas array. Examples include processes such as cell cycle control, oncogenesis, and negative control of cell proliferation.

As another example, we compared commercially available oncogene arrays. Such arrays are available from BD Biosciences Clontech (Atlas Select™ Human Oncogene 7831-1 and Human Oncogene/Tumor Suppressor 7745-1), MICROMAX, and SuperArray. The numbers of distinct sequences available on these arrays are 514, 199, 335, and 26, respectively. A comparison between these arrays is shown in Table 2. It is interesting to note that the MICROMAX Oncogene array contains 73 sequences (71 UniGene clusters) related to oncogenesis. At the same time, the Atlas Oncogene array contains only 22 sequences (21 UniGene clusters) representing the same biological process. The results are remarkable in the light of the fact that that MICRO-MAX Oncogene array contains only 335 genes compared to the Atlas Oncogene array, which contains 514 genes. The same holds true for protein phosphorylation represented by 19 genes on the MICROMAX Oncogene array and 6 genes on the Atlas Oncogene array and negative control of cell proliferation represented by 15 genes on the MICROMAX Oncogene array and only 6 genes on the Atlas Oncogene array.

Finally, we performed another analysis in which we contemplated the use of both Atlas Oncogene and tumor suppressor as a two-array solution and compared this selection of genes with the genes used on the MICROMAX Oncogene array. Note that when sets of arrays are used, many genes may be present on more than one array. Thus, the coverage of a given pathway or biological process cannot be inferred by simply summing the number of genes covering the given pathway on each array.

Circle Reader Service No.

Table 3 shows the comparison between the set of two Atlas Oncogene arrays and the MICROMAX array. When the two Atlas arrays are used together, there is a good representation of general signal transduction, protein phosphorylation, and negative control of cell proliferation. Remarkably, oncogenesis is still better represented on the MICROMAX array in spite of the fact that the set of two Atlas Oncogene arrays deploy 676 genes while the unique MICROMAX Oncogene array uses only 335 genes.

The examples above show that each array or set of arrays has a certain biological bias. These examples should not be interpreted as proving that one particular array is better or worse than other similar arrays. However, these examples do show that in those situations in which a certain hypothesis exist, the choice of the array must be made based on a comprehensive functional analysis of the biological processes, biochemical functions, cellular components, and chromosomal locations of the genes represented on each of the arrays considered.

CONCLUSION

The first step of an experiment involving microarrays will probably be exploratory, aimed at generating hypotheses. Comprehensive arrays including as many genes as possible are useful at this stage. In most cases, the hypothesis-generating phase will be followed by successive steps of focused research. Such hypothesis-driven research often concentrates on a few biological mechanisms and pathways. However, even a single biological mechanism may still involve hundreds of genes that may make the microarray approach the preferred tactic. If microarrays are to be involved in any follow-up, focused, hypothesis-driven research, we argue that one should use the array(s) that best represent the corresponding pathways. Functional analysis can suggest the best array or set of arrays to be used to test a given hypothesis. This can be accomplished by analyzing the list of genes on all existing arrays and providing information about the pathways, biological mechanisms, and molecular functions represented by the genes on each array. Onto-Compare is a tool that allows such comparisons. This tool is available at http://vortex.cs.wayne.edu/Projects.html.

REFERENCES

1. **Draghici, S., P. Khatri, R.P. Martins, G.C. Ostermeier, and S.A. Krawetz.** 2003. Global functional profiling of gene expression. Genomics (In press.)
2. **P. Khatri, Draghici, S., G.C. Ostermeier, and S.A. Krawetz.** 2002. Profiling gene expression with onto-express. Genomics 79:266-270.
3. **Armstrong, S.A., J.E. Staunton, L.B. Silverman, R. Pieters, M.L. den Boer, M.D. Minden, S.E. Sallan, E.S. Lander, T.R. Golub, and S.J. Korsmeyer.** 2002. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat. Genet. 30:41-47.
4. **Bhattacharjee, A., W.G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, et al.** 2001. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc. Natl. Acad. Sci. USA 98:13790-13795.
5. **Golub, T.R., D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, et al.** 1999. Science 286:531-537.
6. **Shipp, M.A., K.N. Ross, P. Tamayo, A.P. Weng, J.L. Kutok, R.C. Aguiar, M. Gaasenbeek, M. Angelo, et al.** 2002. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat. Med. 8:68-74.
7. **Ramaswamy, S. P. Tamayo, R. Rifkin, S. Mukherjee, C.H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, et al.** 2001. Proc. Natl. Acad. Sci. USA 98:15149-15154.
8. **Staunton, J.E., D.K. Slonim, H.A. Coller, P. Tamayo, M.J. Angelo, J. Park, U. Scherf, J.K. Lee, et al.** 2001. Chemosensitivity prediction by transcriptional profiling. Proc. Natl. Acad. Sci. USA 98:10787-10792.
9. **Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub.** 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA 96:2907-2912.
10. **Bellman, R.** 1961. Adaptive Control Processes: A Guided Tour. Princeton University Press, Princeton, NJ.
11. **Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, et al.** 2000. Gene ontology: tool for the unification of biology. Nat. Genet. 25:25-29.
12. **Ashburner, M., C.A. Ball, J.A. Blake, H. Butler, J.M. Cherry, J. Corradi, K. Dolinski, J.T. Eppig, et al.** 2001. Creating the gene ontology resource: design and implementation. Genome Res. 11:1425-1433.
13. **Schuler, G.D.** 1997. Pieces of puzzle: Expressed sequence tags and the catalog of human genes. J. Mol. Med. 75:694-698.
14. **Pruitt, K.D., K.S. Katz, H. Sicotte, and D.R. Maglott.** 2000. Introducing Refseq and LocusLink: curated human genome resources at the NCBI. Trends Genet. 16:44-47.
15. **http://www.geneontology.org.** Gene Ontology Consortium.
16. **Hetz, C.A., M. Hunn, P. Rojas, V. Torres, L. Leyton, and A.F. Quest.** Caspase-dependent initiation of apoptosis and necrosis by the fas receptor in lymphoid cells: onset of necrosis is associated with delayed ceramide increase. J. Cell Sci. 115:4671-4683.
17. **Yoshida, T., N. Koide, T. Sugiyama, I. Mori, and T. Yokochi.** 2002. A novel caspase dependent pathway is involved in apoptosis of human endothelial cells by shiga toxins. Microbiol. Immunol. 46:697-700.
18. **Chen, F., R. Chang, M. Trivedi, Y. Capetanaki, and V.L. Cryns.** 2002. Caspase proteolysis of desmin produces a dominant negative inhibitor of intermediate filaments and promotes apoptosis. J. Biol. Chem. (In press.)
19. **Aqeilan, R., R. Kedar, A. Ben-Yehudah, and H. Lorberboum-Galski.** 2002. Mechanism of action of il2-bax; an apoptosis-inducing chimeric protein targeted against interleukin-2 receptor expressing cells. Biochem. J. (In press.)
20. **Hodge, S., G. Hodge, R. Flower, P.N. Reynolds, R. Scicchitano, and M. Holmes.** 2002. Up-regulation of production of TGF-beta and IL-4 and down-regulation of IL-6 by apoptotic human bronchial epithelial cells. Immunol. Cell. Bio. 80:537-543.
21. **Schuhknecht, S., S. Duensing, I. Dallman, J. Grosse, M. Reitz, and J. Atzpodien.** 2002. Interleukin-12 inhibits apoptosis in chronic lymphatic leukemia (CLL) B cells. Cancer Biother. Radiopharm. 17:495-499.