Bioinformatics Approaches in the Study of Cancer

David A. Hanauer*, Daniel R. Rhodes, Chandan Sinha-Kumar and Arul M. Chinnaiyan

Room 6303 CCGC, 1500 E Medical Center Dr, Ann Arbor, MI 48109-0942, USA

Abstract: A revolution is underway in the approach to studying the genetic basis of cancer. Massive amounts of data are now being generated ***via*** high-throughput techniques such as DNA microarray technology and new computational algorithms have been developed to aid in analysis. At the same time, standards-based repositories, including the Stanford Microarray Database and the Gene Expression Omnibus have been developed to store and disseminate the results of microarray experiments. Bioinformatics, the convergence of biology, information science, and computation, has played a key role in these developments. Recently developed techniques include Module Maps, SLAMS (Stepwise Linkage Analysis of Microarray Signatures), and COPA (Cancer Outlier Profile Analysis). What these techniques have in common is the application of novel algorithms to find high-level gene expression patterns across heterogeneous microarray experiments. Large-scale initiatives are underway as well. The Cancer Genome Atlas (TCGA) project is a logical extension of the Human Genome Project and is meant to produce a comprehensive atlas of genetic changes associated with cancer. The Cancer Biomedical Informatics Grid (caBIG™), led by the NCI, also represents a colossal initiative involving virtually all aspects of cancer research and may help to transform the way cancer research is conducted and data are shared.

INTRODUCTION

The approach to studying the genetic basis of cancer is undergoing a revolution. Rather than focusing on individual genes, scientists are now exploring substantial components of the expressed genome. The wealth of molecular information being generated from the laboratory as well as the volume of data being stored in the patient record is continuing to increase at an astounding rate. Finding new ways to integrate these data has been crucial to developing novel insights into the genetics of cancer. Thus bioinformatics, which is the convergence of biology, information science, and computation is continuing to emerge as a crucial component of cancer biology research.

Bioinformatics is playing an increasingly important role not only for the computational methods and tools that have been developed, but for the continuing work of creating standards and repositories as well. The following review will summarize some of the recent developments in the use of bioinformatics to study the genetics of cancer. This review is not meant to characterize all of the important progress in the field of bioinformatics, but will serve to highlight a few areas that are representative of current trends. Namely, the push to integrate more and more data in new and different ways to help speed cancer research and ultimately lead to better treatments for patients.

NEW BIOINFORMATICS TECHNIQUES

DNA microarray technology has revolutionized the process for discovering the relationships between gene expression and disease patterns. The use of microarrays has resulted in the generation of vast quantities of data with the need for high-throughput computational techniques for analysis. Traditional methods of analysis such as clustering have revealed highly informative information for hypothesis generation but have generally been limited to single experimental runs focused on a specific cancer type. Newer techniques have sought to move beyond the single chip analysis and look for higher-level patterns that are evident only across multiple, disparate microarray experiments.

Module Maps

The use of "module maps" [1, 2] has shown promise for delineating common patterns of gene expression across heterogeneous tissue types and disease processes for cancer. Modules are determined by comparing biologically relevant gene sets to expression data and extracting a subclass of genes that are co-expressed in a statistically significant manner. These modules are thought to better reflect the true biological processes involved since they are directly linked to the actual expression in the samples. The module patterns can then be compared across all tissue types to look for similar signatures suggesting common underlying processes.

A recent study published in Nature Genetics [3] demonstrates the interesting insights that can be achieved for cancer research using module maps. The authors collected results from almost 2,000 microarray experiments for a wide variety of cancers and annotated them with clinically relevant data including both diagnostic and prognostic information. Beginning with approximately 3,000 gene sets obtained from multiple sources, 456 modules were

*Address correspondence to this author at the Bioinformatics Core, Room 6303 CCGC, 1500 E Medical Center Dr, Ann Arbor, MI 48109-0942, USA; Tel: 734-615-0599; E-mail: hanauer@med.umich.edu

1566-5240/07 $50.00+.00

found. These modules were then compared across the different types of cancers. Their findings revealed gene expression patterns common to unrelated cancers, providing useful information about mechanisms of cancer that could improve our general understanding of how these disease processes occur. For example, a bone osteoblastic module, containing genes predominantly associated with the proliferation and differentiation of both chondrocytes and osteoblasts, was found to be significantly associated and upregulated in certain breast cancers and downregulated in lung cancer, hepatocellular carcinoma, and acute lymphoblastic leukemia. Such a common mechanism, once better understood, could lead to the development of therapeutic targets against a broad spectrum of cancer types.

While this technique was useful in elucidating global patterns of disease that would not have been evident by exploring single microarray experiments, issues still surround this approach. The lack of detailed, standardized clinical information with which to annotate experiments makes such broad cross-comparisons difficult. Furthermore, the variability in results obtained from different normalization techniques also can impact the utility of this type of study. The choice of an ideal normalization method [4,5] is still an area in need of further research.

SLAMS

Standard methods for classifying tumors based on histology, grade, and staging continue to represent the standard of care for almost all cancers. Even detailed histologic specimens offer only a crude view of what might be occurring at the molecular level since two tissue types may look alike but may not act alike, and complex interactions cannot be characterized. Such molecular differences likely underlie the heterogeneity in outcomes and response to therapy for similarly classified tissues and leads to great uncertainty when clinicians must tailor a treatment regimen for a particular patient.

Gene expression profiles, derived from microarray analyses, while still not widely used in clinical care, offer the promise of a much greater detailed approach to classifying tumors. When linked to clinical data, these profiles, or signatures, have proven to be effective for diagnosis [5] as well as for predicting clinical outcomes including both prognosis [6, 7] and therapeutic response to chemotherapeutic drugs [8, 9].

Even this approach has its limitations. The ultimate goal is not to simply find genetic markers that aid in prediction but to determine how those genes function in the disease process with identification of potential therapeutic targets to positively influence outcomes. Current techniques aimed at elucidating function allow for only one or a few genes in a signature to be studied in more detail at a time. Furthermore, specific oncogenetic regulators with large downstream effects may not even be included in the genetic signature itself.

A team at Stanford University has developed a method using linkage analysis of gene expression data linked to DNA copy number changes in order to identify chromosomal regions containing candidate oncogene regulators. In this case the phenotype in the linkage analysis is the gene expression profile associated with a known tumor type. Their method, SLAMS (Stepwise Linkage Analysis of Microarray Signatures) [10], involves a four-step process Fig. (**1**), the first of which is sorting tumors into two groups based on the presence or absence of a known genetic signature. Next, significant associations between changes in copy number and the expression profile are identified. Both amplifications and deletions can be detected by correlating them with either up- or down-regulation of the genetic signature. The third step involves identifying candidate regulators by comparing their level of expression with that of the expression signature. The final step in SLAMS is to use the expression levels of the candidate regulators to predict the expression profiles of other cancer samples not used in the first three steps, which serves to validate the association.

The SLAMS method was tested by applying it to breast cancer samples obtained from the Stanford Microarray Database [11]. A previously known "wound signature", comprising 512 genes, was used to separate the samples. This signature had previously been shown to predict clinical prognostic risk. Copy number changes at more than 6,500 loci were used in this analysis. More than half of the 57 DNA probes significantly associated with the signature were mapped to the proximal region of chromosome 8q. Interestingly, the MYC gene (known to be an oncogenic transcription factor) is located within the distal region of chromosome 8q, an area that was amplified in samples without the wound signature. The authors thus concluded that a gene (or genes) in the proximal 8q region might interact with MYC to express the wound signature. Further studies revealed a candidate co-regulator which, when activated, was able to induce expression of the previously known signature. This technique shows great promise, but it will have to be validated on other cancer types. Potential problems are that the DNA copy number might not always be associated with a change in the activity of a regulator gene and that correlations made with this method may simply be due to chromosomal proximity.

COPA

Traditional techniques for identifying oncogene expression profiles are limited in their ability to identify patterns across multiple samples, especially when those patterns may not be predominant enough to stand out. This is due to the heterogeneity of expression patterns in samples, making it difficult to separate actual patterns from

other background variations. To solve this problem, a new method called COPA (Cancer Outlier Profile Analysis) [12] was developed to find profiles that might only be expressed in a subset of tumor samples.

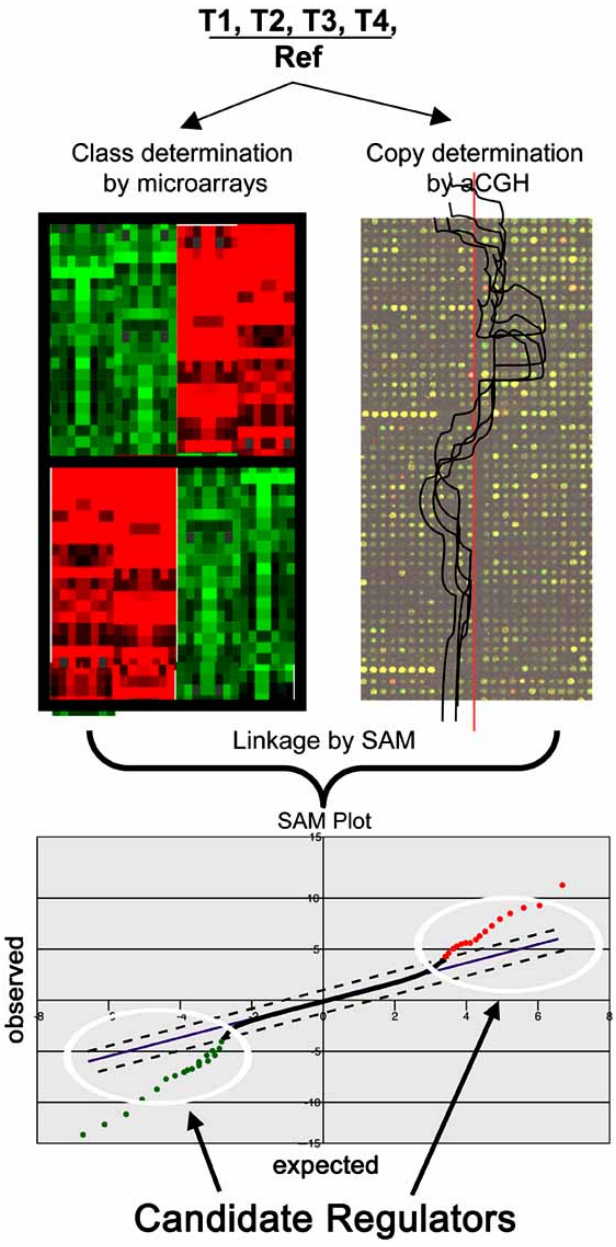

Fig. (1). Diagram representing the process for performing a SLAMS analysis. (1) Tumors are sorted into two groups based on the presence or absence of a known genetic signature; (2) Significant associations between changes in copy number and the expression profile are identified. Significance Analysis of Microarrays (SAM) is used to determine the association between the genetic signature and copy number changes; (3) Candidate regulators are identified by comparing their level of expression with that of the expression signature; (4) Expression levels of the candidate regulators are used to predict the expression profiles of other cancer samples not used in the first three steps, in order to validate the association.

The COPA technique is able to identify outlier expression profiles where the overall expression levels for a particular set of genes is low and only a small subset of tumor samples shows overexpression Fig. (**2**). The first step in this process is to set the median value of expression to zero and then determining the median absolute deviation (MAD) for each gene. The expression values are then divided by their corresponding MAD, resulting in a set of transformed expression values. This technique serves to flatten the average dynamic range of most expression profiles where most cancerous samples demonstrate overexpression and serves to highlight profiles that deviate from the average level of expression. The level of deviation from the average expression pattern is then ranked for each gene to provide a list of outlier genes worthy of further exploration.

When COPA was applied to the Oncomine database [13], comprising 132 gene expression data sets and over 10,000 microarray experiments, several outlier profiles were identified. What was most intriguing is that two of the outlier genes identified, ERG and ETV1 (located on 21q22.3 and 7p21.2, respectively), were shown to be among the top ten outliers to exist in prostate samples from six independent studies even though their prior association with cancer involved a gene fusion in Ewing's sarcoma and myeloid leukemia. Analysis of three independent prostate microarray studies found that more than half of the cancer samples revealed overexpression of either ERG or ETV1 whereas no such overexpression was identified in any of the benign controls. Furthermore, overexpression of ERG and ETV1 was mutually exclusive.

Additional laboratory work provided the reason for this: about 60-70% of prostate cancers were found to have a gene fusion of TMPRSS2 and either ERG or ETV1.

This was the first time a non-random, recurrent gene fusion had been found in a common epithelial solid tumor (such rearrangements were previously thought to occur primarily in sarcomas, leukemias, and lymphomas). Not only might this rearrangement be responsible for the majority of prostate cancers but, due to the high incidence of this common tumor, it may actually represent the most common rearrangement in human cancers overall. This discovery will surely encourage scientists to take another look at other tumors that may exhibit similar rearrangements and may lead to new diagnostic tests or therapeutic targets. Because the COPA method can be applied to any expression data, only time will tell how useful COPA will be in unmasking other novel associations worthy of exploration.

DATA REPOSITORIES, STANDARDS, AND ANALYSIS TOOLS

The analyses described above would not have been possible without the investigators being able to

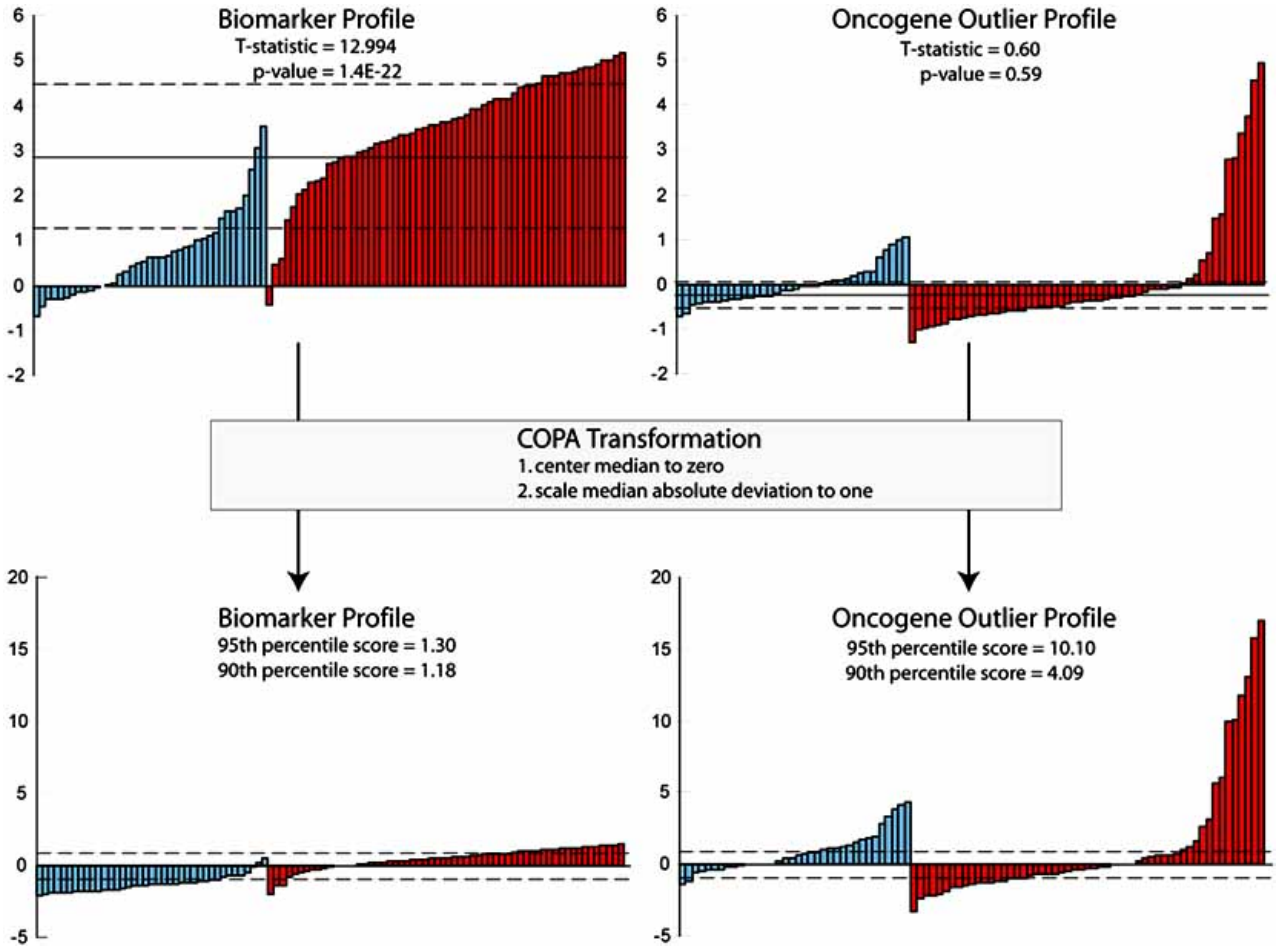

Fig. (2). Biomarker profiles (upper left) are usually characterized by extensive overexpression in cancer (red) compared to normal tissues (blue). Oncogene outlier profiles are usually characterized with overexpression in only a small subset of cancer samples (upper right). Application of the COPA transformation flattens the range of expression values in biomarker profiles (lower left) but can highlight the expression values in outlier profiles (lower right).

bring together many publicly accessible microarray gene expression datasets. Advantages of making such data readily available were outlined in an open letter submitted to various journals by the Microarray Gene Expression Data (MGED) Society [14] and included the ability to link microarray results to other types of data sources to enhance the ability to interpret gene expression patterns. Clearly this is a successful strategy.

Various repositories exist for storing microarray data and include the Stanford Microarray Database (SMD) [11], European Bioinformatics Institute's ArrayExpress [15], and the National Cancer Institute's Gene Expression Omnibus (GEO) [16] which is maintained by the National Center for Biotechnology Information (NCBI). While primarily for storage, these repositories do offer some limited data analysis options such as hierarchical clustering. Standard data formats have also been adopted in order to ensure the transportability and reuse of data sets. Examples of such standards include the Microarray Gene Expression Markup Language (MAGE-ML) [17] and the Minimum Information About a Microarray Experiment (MIAME) [18].

Microarray analysis tools specific to cancer also exist. Gene Logic's BioExpress® System Oncology Suite [19] is a commercially available application designed to allow investigators to explore microarray data for cancers involving over 30 tissue types. It includes negative controls, as well as primary, metastatic, and benign tumors. To aid in the discovery process the data sets are annotated with richly detailed clinical information including clinical staging, pathology reports, and complete blood counts. The tool also integrates data from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, BioCarta™, and other sources. The BioExpress® Oncology Suite offers visualization and statistical analysis tools as well. Gene Logic also offers their ASCENTA® System [20] which contains about 8,700 Affymetrix GeneChip® mouse, human, and rat arrays. While not specific to oncology, it does provide analysis tools to explore co-expression and differential gene expression between the samples.

Similar to Gene Logic's offerings is the Oncomine [13] database Fig. (**3**). True to its name, Oncomine is a data mining tool specific to oncology-related microarray analysis with free access to academic

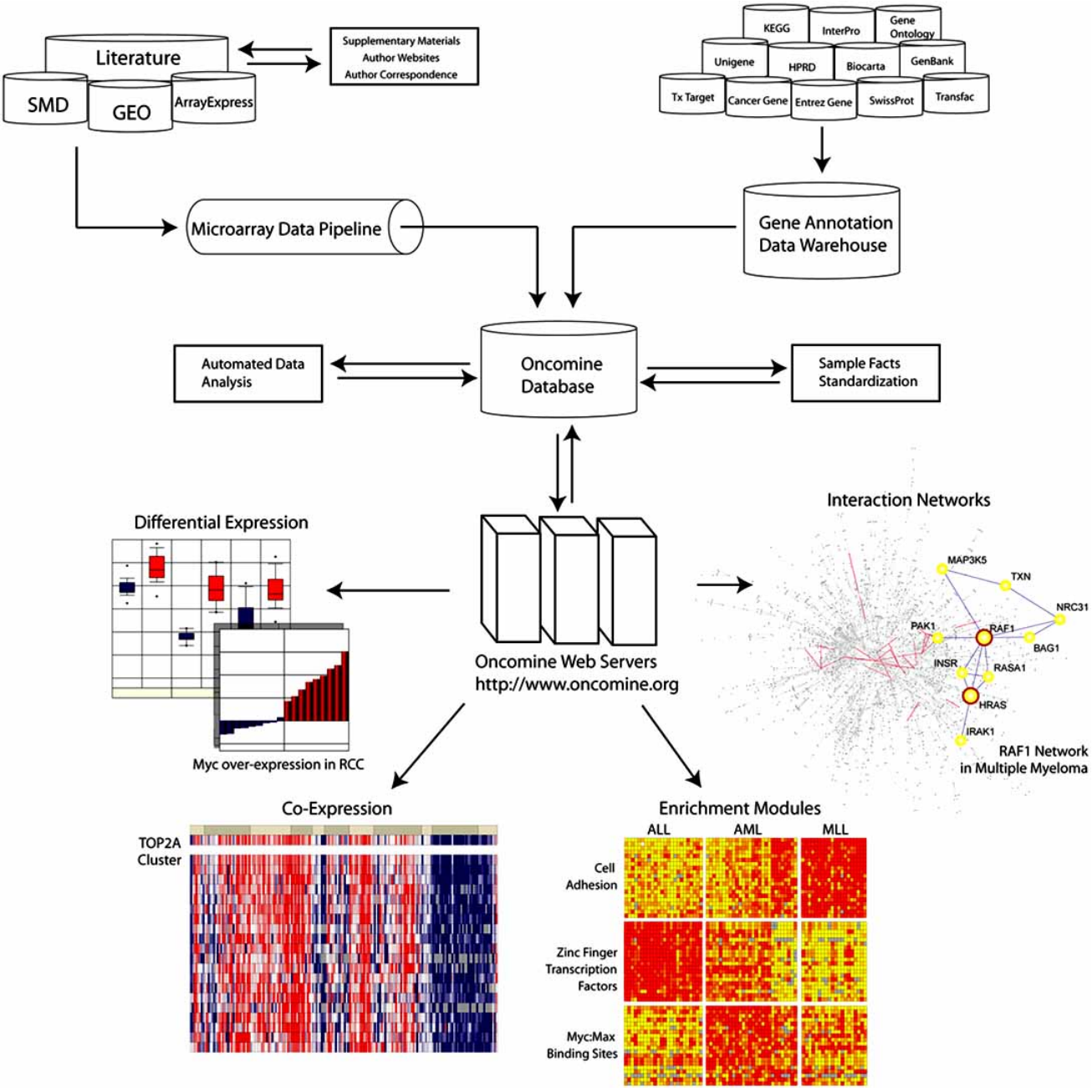

Fig. (3). Oncomine: A Cancer Microarry Compendium. This diagram shows a general overview of the key components of Oncomine. Data are imported to the Oncomine database via a data pipeline as well as a gene annotation data warehouse. Data are then analyzed and standardized for comparison across studies. End users interact with the Oncomine web servers. These servers provide many visual representations (e.g. Interaction networks, co-expression, differential expression, and enrichment modules) of the data in the compendium to aid in the exploration of interesting or novel relationships.

users. Oncomine focuses only on microarrays of human cancer tissues and normal controls, although it is not limited to Affymetrix chips. As of this writing it included nearly 10,000 microarrays representing over 30 tissue types. Data are normalized for easy comparisons across studies and analyses are pre-computed to make data exploration extremely quick. Microarrays are annotated with data from 14 external sources and analysis tools include visualization of differential expression, co-expression, enrichment analysis, and an interaction network analysis.

A relatively recent addition to the suite of cancer-specific analysis tools are caArray [21] and caWorkbench [22], developed through the caBIG™ initiative (described below). The caArray tool is an open source data repository for microarray data and

annotations that will support diverse file formats including Affymetrix, GenePix, and ImaGene, as well as the XML-based MAGE-ML format. While repositories such as the SMD have released their source code so that individuals or institutions could install a local version, what differentiates the caArray initiative from others is the potential for seamless sharing of public datasets across a federated grid of databases without any central storage or control of the information. The adoption of low-level data transfer standards across the caBIG™ grid will also allow for data to be imported into other caBIG™ compliant tools including caWorkbench. The caWorkbench application is a tool being developed for a myriad of uses including analysis of microarray data, pathways, and sequences. It is built using a plug-in architecture so that users can add desired components including normalization, filtering, and clustering. While its user base is currently limited, increased adoption will likely occur as the caBIG™ initiative becomes more mature.

LARGE-SCALE INITIATIVES

A growing trend with the tools and techniques currently in vogue is the integration of hundreds of samples, studies, and data types combined from multiple sources. Not all samples were originally created by the authors but were shared through public resources. The benefits of this data integration are numerous. Nevertheless, data integration from disparate data silos is very labor intensive and takes away from time spent conducting valuable research. Recent initiatives are seeking to eliminate the research bottleneck caused by a lack of standards and interoperability. Other initiatives are looking to characterize cancer on a global scale so that investigators will have more useful data to analyze and integrate in novel ways.

caBIG™

The cancer biomedical informatics grid (caBIG™) [23] Fig. (**4**) is an ambitious NCI-funded initiative to

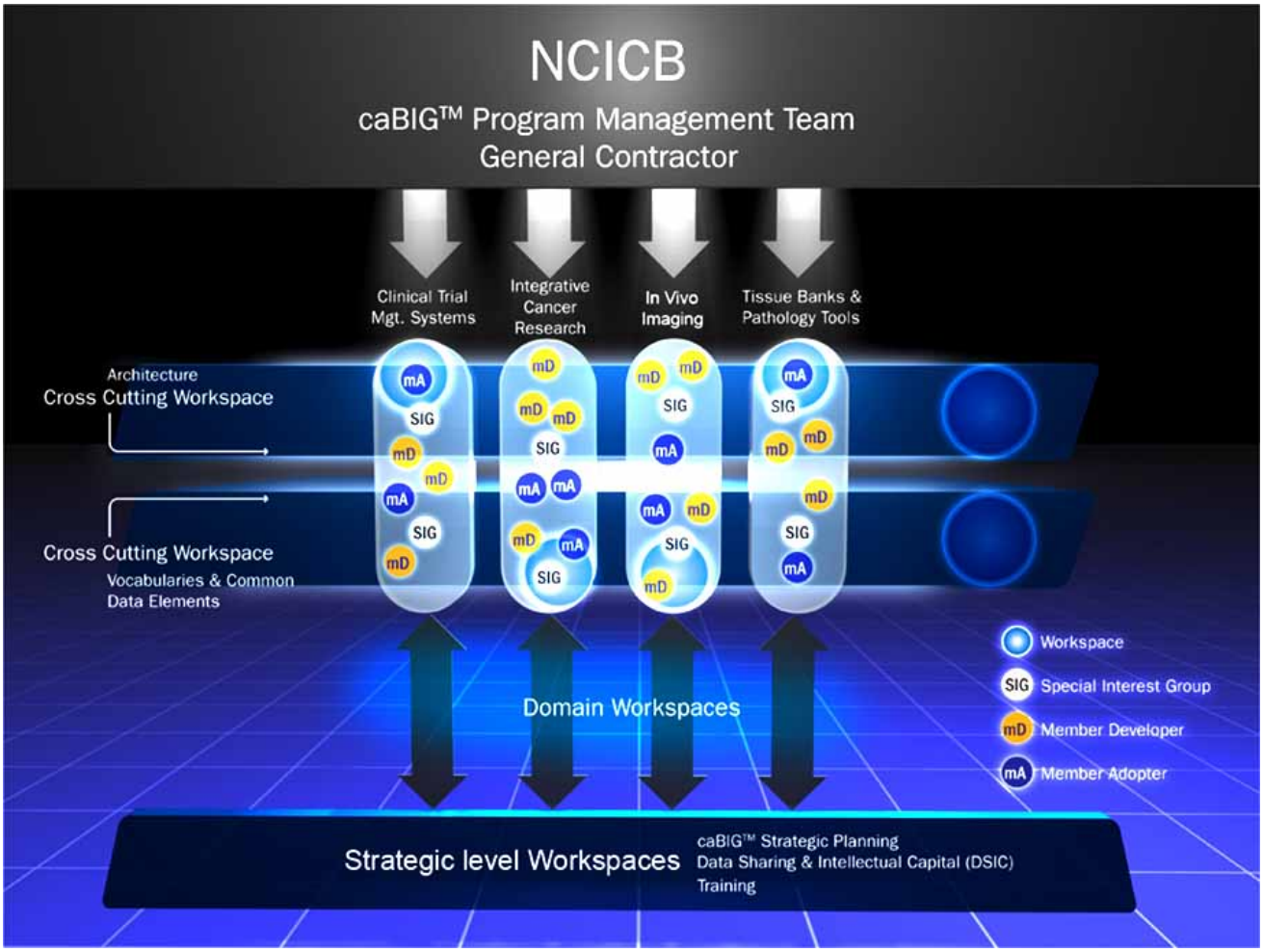

Fig. (4). Diagram showing the general organizational structure of the caBIG™ initiative. Domain Workspaces encompass projects with similar themes and include Clinical Trial Management Systems, Integrative Cancer Research, ***In Vivo*** Imaging, and Tissue Banking & Pathology Tools. Cross-Cutting Workspaces are those which impact the activities of all other workspaces and include Architecture and Vocabularies & Common Data Elements. Strategic level Workspaces include Strategic Planning, Training, as well as Data Sharing & Intellectual Capital.

connect Cancer Centers and other research institutions in a network, or grid, to seamlessly share and analyze biomedical data with the ultimate goal of developing novel and effective approaches for preventing, detecting, and treating cancer. The caBIG™ community is comprised of more than 800 people from over 80 organizations on more than 70 projects ranging from management of clinical trials to tools for analysis of gene expression data. Those involved include both funded participants and volunteers from academia, industry, and the community. Tools developed by caBIG™ will be made freely available to all researchers in order to foster their dissemination and adoption.

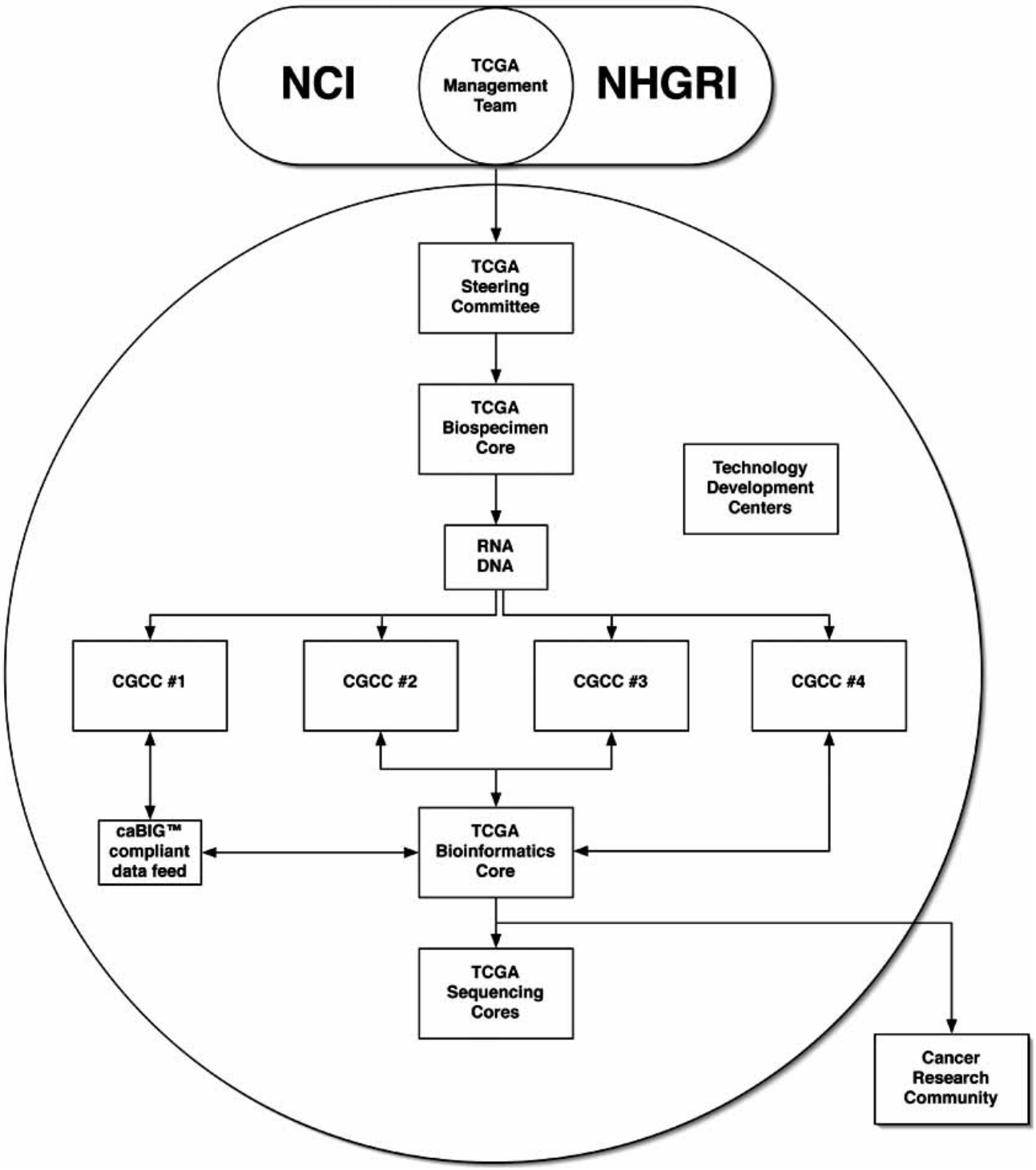

Fig. (5). A schematic diagram representing the structure of the proposed Cancer Genome Atlas (TCGA) program, a joint initiative between the NCI and the National Human Genome Research Institute (NHGRI). Collected specimens will be distributed to each of four cancer genome characterization centers (CCGC) for analysis. Data will then be processed *via* the TCGA bioinformatics core and shared with the greater cancer research community for further investigation.

By adopting standards across all aspects of cancer research, investigators using caBIG™ tools should be able to generate and test hypotheses faster than ever before. At the core of caBIG™ is caGrid, which will provide the underlying infrastructure for seamless and secure transmission of data to other grid nodes around the nation and perhaps eventually the world. Conforming to the standards developed by caBIG™ will ensure both syntactic interoperability with the grid itself as well as semantic interoperability with the controlled vocabularies and common data elements currently being adopted for use in caBIG™ applications. Compatibility with the grid architecture must be achieved in four general categories: Interfaces, Vocabularies/Terminologies and Ontologies, Data Elements, and Information Models.

The caBIG™ initiative is still relatively new. Many tools are not yet ready for adoption and other details, including those related to compatibility, are still being determined. Large scale adoption is likely many years away, although the adoption rate may increase as the NIH requires caBIG™ compatibility as a condition for funding.

The TCGA

The completion of the Human Genome Project (HGP) in 2003 marked a major milestone in genomics research and helped to establish the importance of bioinformatics for storing and analyzing the vast amounts of data that were generated. The success of the HGP helped build the foundation for further research into details of the human genome, with the recent establishment of The Cancer Genome Atlas (TCGA) [24]. Whereas the HGP was meant to provide the baseline genetic code of the human genome, the TCGA, a joint endeavor between the NCI and the National Human Genome Research Institute (NHGRI), is meant to provide a centrally maintained, comprehensive atlas of genetic changes related to cancer.

Tissue samples from patients with cancer will be sent to high-throughput genome sequencing centers to characterize the cancer-associated changes present Fig. (**5**). These data will be integrated with de-identified clinical information about the patient and subjected to analytical techniques to identify genetic fingerprints associated with each tumor type. This is not significantly different from the type of research currently being conducted except that the ultimate scope of this project is to characterize all cancers with the information stored in a central repository accessible by all investigators.

The role of bioinformatics in this project will focus on collection, storage, distribution, and analysis of the data and will likely rely on the tools and infrastructure developed through the caBIG™ initiative.

FUTURE CHALLENGES

While the application of bioinformatics to cancer research has allowed for staggering progress, significant challenges remain. New analysis techniques will continue to be needed in order to make sense of the myriad genetic changes associated with cancer, especially when those changes are subtle and poorly understood. Methodologies for data integration will be needed as well as normalization algorithms for samples brought together from different laboratories under different conditions. Also needed will be continued development and adoption of standard formats for data storage and sample annotation.

The ability to link microarray data with detailed clinical patient information in a de-identified and computable manner still does not occur as often as it should. It will be necessary to move beyond simply labeling a specimen in a microarray study as being positive or negative for cancer; rather, detailed clinical annotation will be necessary for elucidating the genetic role in cancer subtypes, reactions to therapies, as well as other difficult to quantitate parameters effecting the quality of life for patients. Not all medical records are in electronic format, and much clinical information that is electronic is still stored as free-text. Extracting that information reliably remains a significant challenge.

Because caBIG™ touches on almost all aspects of cancer, from clinical bedside to basic bench research, it may have the biggest potential to positively change the standard research paradigm. If successful, caBIG™ will likely transform the way research is conducted and data are shared. Perhaps once this occurs the cancer research community will be poised to meet the challenge set forth by the current director of the NCI, Andrew von Eschenbach, to “eliminate the suffering and death due to cancer by 2015”.

ACKNOWLEDGEMENTS

We would like to thank Scott Tomlins, Shanker Kalyana-Sundaram, and Terrence Barrette with helpful discussions and use of Figures. The caBIG™ Figure was provided courtesy of the National Cancer Institute Center for Bioinformatics.

This work was supported, in part, by the National Institutes of Health through the University of Michigan's Cancer Center Support Grant (5 P30 CA46592) as well as the Burroughs Welcome Foundation and the Department of Defense.

REFERENCES

[1] Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y. and Barkai, N. (2002) ***Nat. Genet.***, **31**, 370-7.

[2] Tanay, A., Sharan, R., Kupiec, M. and Shamir, R. (2004) ***Proc. Natl. Acad. Sci. USA***, **101**, 2981-6.

[3] Segal, E., Friedman, N., Koller, D. and Regev, A. (2004) ***Nat. Genet.***, **36**, 1090-8.

[4] Chua, S.W., Vijayakumar, P., Nissom, P.M., Yam, C.Y., Wong, V.V. and Yang, H. (2006) ***Nucleic Acids Res***., **34**, e38.
[5] Park, T., Yi, S.G., Kang, S.H., Lee, S., Lee, Y.S. and Simon, R. (2003) ***BMC Bioinformatics***, **4**, 33.
[6] Chen, C.N., Lin, J.J., Chen, J.J., Lee, P.H., Yang, C.Y., Kuo, M.L., Chang, K.J. and Hsieh, F.J. (2005) ***J. Clin. Oncol***., **23**, 7286-95.
[7] Dyrskjot, L., Zieger, K., Kruhoffer, M., Thykjaer, T., Jensen, J.L., Primdahl, H., Aziz, N., Marcussen, N., Moller, K. and Orntoft, T.F. (2005) ***Clin. Cancer Res.,*** **11**, 4029-36.
[8] Gyorffy, B., Surowiak, P., Kiesslich, O., Denkert, C., Schafer, R., Dietel, M. and Lage, H. (2006) ***Int. J. Cancer***, **118**, 1699-712.
[9] Gyorffy, B., Serra, V., Jurchott, K., Abdul-Ghani, R., Garber, M., Stein, U., Petersen, I., Lage, H., Dietel, M. and Schafer, R. (2005) ***Oncogene***, **24**, 7542-51.
[10] Adler, A.S., Lin, M., Horlings, H., Nuyten, D.S., van de Vijver, M.J. and Chang, H.Y. (2006) ***Nat. Genet***., **38**, 421-430.
[11] Ball, C.A., Awad, I.A., Demeter, J., Gollub, J., Hebert, J.M., Hernandez-Boussard,T., Jin, H., Matese, J.C., Nitzberg, M., Wymore, F., Zachariah, Z.K., Brown, P.O. and Sherlock, G. (2005) ***Nucleic Acids Res.,*** **33**, D580-2.
[12] Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran,S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J.E., Shah, R.B., Pienta, K.J., Rubin, M.A. and Chinnaiyan, A.M. (2005) ***Science***, **310**, 644-8.
[13] Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A. and Chinnaiyan, A.M. (2004) ***Neoplasia***, **6**, 1-6.
[14] Ball, C., Brazma, A., Causton, H., Chervitz, S., Edgar, R., Hingamp, P., Matese, J.C., Icahn, C., Parkinson, H., Quackenbush,J., Ringwald, M., Sansone, S.A., Sherlock, G., Spellman, P., Stoeckert, C., Tateno, Y., Taylor, R., White, J., Winegarden, N.; Microarray Gene Expression Data (MGED) Society**.** (2004) ***Microbiology***, **150**, 3522-4.
[15] Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., Holloway, E., Kapushesky, M., Kemmeren, P., Lara, G.G., Oezcimen, A., Rocca-Serra, P. and Sansone, S.A. (2003) ***Nucleic Acids Res***., **31**, 68-71.
[16] Barrett, T., Suzek, T.O., Troup, D.B., Wilhite, S.E., Ngau, W.C., Ledoux, P., Rudnev, D., Lash, A.E., Fujibuchi, W. and Edgar, R. (2005) ***Nuceic Acids Res***., **33**, D562-6.
[17] Spellman, P.T., Miller, M., Stewart, J., Troup, C., Sarkans, U., Chervitz, S., Bernhart, D., Sherlock, G., Ball, C., Lepage, M., Swiatek, M., Marks, W.L., Goncalves, J., Markel, S., Iordan, D., Shojatalab, M., Pizarro, A., White, J., Hubley, R., Deutsch, E., Senger, M., Aronow, B.J., Robinson, A., Bassett, D., Stoeckert, C.J. Jr. and Brazma, A. (2002) ***Genome Biol***., **3**, RESEARCH0046.
[18] Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J. and Vingron M. (2001) ***Nat. Genet***., **29**, 365-71.
[19] http://www.genelogic.com/docs/pdfs/OncologyProductSheet.pdf Accessed on April 22, 2006.
[20] http://www.genelogic.com/docs/pdfs/ascentaProductSheet.pdf Accessed on April 22, 2006.
[21] http://caarray.nci.nih.gov/ Accessed on April 23, 2006.
[22] http://gforge.nci.nih.gov/projects/caworkbench/Accessed on April 23, 2006.
[23] von Eschenbach, A.C. and Buetow, K. (2006) ***Cancer Inform.***, **2**, 22-4.
[24] http://cancergenome.nih.gov/index.aspAccessedon April 24, 2006.

Received: May 30, 2006 **Revised: June 06, 2006** **Accepted: June 07, 2006**