Genomic Medicine

Kremena Star* and Barbara Birshtein†
Medical Scientist Training Program
*Department of Molecular Genetics
†Department of Cell Biology
Albert Einstein College of Medicine
Bronx, New York 10461

ABSTRACT

The human genome project created the field of genomics – understanding genetic material on a large scale. Scientists are deciphering the information held within the sequence of our genome. By building upon this knowledge, physicians and scientists will create fundamental new technologies to understand the contribution of genetics to diagnosis, prognosis, monitoring, and treatment of human disease. The science of genomic medicine has only begun to affect our understanding of health.

INTRODUCTION

The human genome is made of deoxyribonucleic acids (DNA). DNA carries the blueprints for all cellular proteins, and is a linear double-helical molecule. This double-helix is composed of two intertwined chains of nucleotides. Each nucleotide is built from a deoxyribose sugar, a nitrogen base (adenine (A), thymine (T), guanine (G), or cytosine (C)), and a phosphate. Two complementary chains of nucleotides are held together by weak hydrogen bonds between the bases A and T, or G and C. DNA molecules are coiled into higher order structures (e.g., chromosomes). In humans, DNA encodes about 25,000 protein-coding genes. These genes comprise only 2 percent of the human genome. The remaining 98 percent of the DNA largely consists of repetitive non-coding DNA sequences, some of which have regulatory functions. About 99.9 percent of the DNA is the same in all individuals. Small differences in DNA make each one of us unique and affect our responses to drugs as well as our susceptibility to disease. Except for germ cells, every cell of the human body contains two sets of 23 chromosomes, which are located in the cell nucleus. Each chromosome comes in two copies (alleles), maternal and paternal. At any given time, only a fraction of these genes are expressed. This expressed subset of genes confers unique properties to each cell in our bodies.

The information contained within the DNA is transcribed into messenger ribonucleic acids (mRNA) molecules. mRNA in turn is translated into the proteins that perform most the critical cellular functions. mRNA molecules are the working copies of genes. They are similar to DNA except that they are single stranded, have a ribose sugar instead of deoxyribose, and uracil (U) instead of thymine (T). Once a gene is transcribed and processed into mRNA, it leaves the nucleus and enters the cytoplasm where the protein synthesis machinery, ribosomes, is located. There the mRNA is read by the ribosomes in groups of three nucleotides called codons. Codons translate the nucleic acid code of DNA and mRNA into protein by specifying the particular amino acid to be incorporated into a growing polypeptide chain.

This polypeptide chain folds to make a protein. Proteins comprise the majority of cell structures. Since the genetic code (i.e., DNA) ultimately determines the proteins that are synthesized, point mutations, polymorphisms, and deletions or insertions in the DNA affect the sequence and the amount of the proteins synthesized. Missense (the substitution of a different amino acid in the polypeptide chain), nonsense (the premature termination of polypeptide synthesis), and frameshift (the disruption of the codon reading frame by one or two nucleotides) mutations can alter or abolish protein function. The inheritance of these types of mutations is the foundation of genetic disease.

Genetics contributes to the formation of disease (i.e., single gene and multiple gene diseases), disease susceptibility, differences in disease severity, and differences in therapeutic response. Recent scientific advances, like DNA microarrays, greatly facilitate the study of gene expression and the role of specific genes in the development of disease. The human genome project has provided copious amounts of information about the DNA sequence of the human genome. Consequently, scientists have identified a large number of novel genes. The main challenge today is to analyze, organize, and catalog this vast amount of information. Traditional methods for studying gene expression allow studies of a relatively small number of genes at a time. The development of new genomic tools enables researchers to address large scale problems and to uncover novel potential targets for diagnosis, prevention, and therapy.

DNA MICROARRAY TECHNOLOGY

DNA microarray technology allows researchers to analyze the expression of many genes in a single experiment. Microarrays are small, solid supports onto which the sequences from thousands of genes are attached at independent fixed locations. The supports can be glass

microscope slides, nylon membranes, or silicon chips. The DNA is spotted (e.g., cDNA microarray) or synthesized (e.g., Affymetrix Gene Chip) onto the solid support. The gene sequences are attached to their support in an orderly way, which allows researchers to use the location of each spot in the array to identify a particular gene sequence. The spots themselves can be DNA, cDNA (i.e., a DNA sequence that complements mRNA), or oligonucleotides (i.e., single-stranded DNA 5 to 50 nucleotides in length).

DNA microarrays take advantage of the properties of complementarity between DNA and RNA molecules (i.e., A must pair with U or T, and C must pair with G, one strand determines the sequence of the other and the sequence of the mRNA, hence complementary strands of DNA or RNA will anneal or bind to each other). The ability of cDNA to hybridize to the DNA template from which it was transcribed is paramount.

To run a DNA microarray, mRNA is isolated from the cells or tissue of interest as well as the control cells or tissue. The mRNA is reverse transcribed into a fluorescently labeled cDNA (either red or green, one color for the experimental sample and one for the control sample). These fluorescently labeled cDNA probes are hybridized to the microarray. After washing, the expression levels of the thousands of genes on the microarray are analyzed by a scanning fluorescence laser that can measure the intensity of red and green at each location on the array. By measuring the intensity of fluorescence at each spot on the array and comparing the intensity to the control sample, one can determine the activity of a particular gene or groups of genes. (For an internet animation of this process see http://www.bio.davidson.edu/courses/genomics/chip/chip.html.)

IMPACT OF DNA MICROARRAYS

Due to their small size and the fact that they contain a very large number of genes, microarrays are a significant advance over traditional methods for measuring gene expression (e.g., Southern blots). Microarrays can assay gene expression within a single sample or compare gene expression in two different cell types or tissues in large-scale genetic analysis. For example, microarrays can be used to study which genes in cells are actively making products under a particular set of conditions, as well as to detect differences in gene activity between healthy and diseased cells. One of the most important applications for arrays so far is the monitoring of gene expression (mRNA abundance). The collection of genes that are expressed or transcribed from genomic DNA, referred to as the expression profile (transcriptome), is a major determinant of cellular phenotype and function. Unlike the genome, the transcriptome changes rapidly and dramatically during normal cellular events such as DNA replication and cell division. In terms of understanding the function of genes, knowing when, where, and to what extent a gene is expressed is central to understanding the biological roles of its encoded protein. In addition, changes in the patterns of gene expressions can provide clues about regulatory mechanisms and biochemical pathways. In the context of human health, the knowledge gained from these types of measurements

TABLE 1 | DNA MICROARRAY APPLICATIONS

Microarray Type	Application
Gene Expression	Drug and Therapy Development Prognosis Prediction Genetic Testing
Comparative Genomic Hybridization	Tumor Classification Risk Assessment Prognosis Prediction Genetic Testing
Single Nucleotide Polymorphisms	Drug and Therapy Development Prognosis Prediction Genetic testing

may determine the source and outcome of a disease and identify gene products with therapeutic utility or targets for therapeutic intervention (Friend et al., 2002).

Microarrays are also used for studies of genomic gains and losses (i.e., Comparative Genomic Hybridization microarray), and DNA mutations (i.e., SNPs microarrays) (Friend and Stoughton, 2002) (Table 1). Changes in chromosomal DNA, such as mutations, genetic rearrangements, and amplifications or losses of particular chromosomal regions can often lead to abnormalities, such as Down's or Turner's syndrome. Comparisons of the copy number of genomic regions or the genotype of genetic markers may be used to detect chromosomal regions and genes that are amplified or lost in different cell types including cancerous and pre-cancerous cells. By applying arrays, scientists can identify when and where changes in copy number, chromosomal rearrangements, and DNA mutations have occurred. These changes can be used for diagnosis of developmental disorders (Dobrovolski et al., 1999), identification of tumor-suppressor genes, disease prognosis, and treatment selection.

Despite its impressive potential, microarray technology requires improvement, innovation, and increased accessibility. Nucleic acid array-based experiments are becoming routine as indicated by the increase in the number of publications that incorporate data obtained in this way (a PubMed search returned 1637 articles published in 2004, 363 more than the previous year). For example, measuring the expression level of essentially every gene (including variant splice forms) starting with RNA from a small number of cells, or even a single cell, is possible because of advances in single-cell handling, RNA amplification methods, and microarray technology. Microarrays, along with genomic sequence information and computational tools, facilitate the identification and classification of DNA sequence information and allow scientists to understand the function and regulation of genome. This knowledge aids physicians and scientists in discovering ways to prevent aberrant cellular processes and to improve human health.

CASE STUDIES

Metastatic Renal Cell Cancer

ND is a 53 year old male who presented to Jacobi Hospital with a 3 month history of left flank pain, gross hematuria, intermittent fever, and weight loss. On physical examination, he appeared pale. He was normotensive with a temperature of 38ºC. He was found to have enlarged supraclavicular lymph nodes and a palpable mass in his left upper quadrant. Chest X-ray showed densities in his left lung. Abdominal CT revealed a mass in his left kidney and enlarged retroperitoneal lymph nodes.

Laboratory findings included a white blood count (WBC) of $8.3x10^9$ cells/L with a normal differential, a hemoglobin (Hb) of 9.2 g/dL, and a hematocrit (Hct) of 26%. His serum calcium was 12.3 mg/dL. All other electrolytes were normal. Parathyroid hormone levels were elevated. Blood urea nitrogen (BUN) was 20 mg/dL and creatinine was 1.7 mg/dL. Urinalysis showed gross hematuria.

A diagnosis of stage IV metastatic renal cell cancer was made, and the patient underwent a left nephrectomy, with removal of adjacent metastases. Pathologic examination revealed clear cell histology. The patient was placed on a course of high dose interleukin-2, along with corticosteroids and calcitonin. Serum calcium levels returned to normal, and the patient felt better.

The patient was told of the poor prognosis for his condition and that it was characterized by a wide variability in the period of survival. His physician recommended a microarray study to answer this question. As a result, ND asked his physician to arrange for a gene-expression profile of the tumor. A study was conducted by researchers in the National Cancer Institute in 2003 (Vasselli et al., 2003), who found the presence of two distinct patterns of gene expression in the primary tumors of patients with metastatic renal cancer. These patterns correlated with difference in survival between the patient groups. Among the genes identified in the study, vascular cell adhesion molecule-1 (VCAM-1) was identified as the most predictive gene for survival. This gene was uniformly up-regulated in patients with longer survival (Vasselli et al., 2003). ND's gene expression profile was consistent with a poor prognosis. ND died three years after he was diagnosed.

Diffuse Large B-cell Lymphoma

AP is a 38 year old HIV positive male with a CD4+ T-cell count of 148 cells per cubic millimeter blood and a viral load of 477 copies/mL. He was admitted to Montefiore Medical Center with complaints of cough (productive of bloody sputum), low grade intermittent fever, weight loss, fatigue, and "swollen glands." The patient first noted swollen glands in his femoral and inguinal areas approximately six weeks prior to admission. A chest X-ray demonstrated a right-sided pleural effusion and possible infiltrate. He was started on ceftriaxone. There is no past history of exposure to tuberculosis and sputum smears were found to be negative.

On physical exam several days after admission the hematologists found maximum temperature (Tmax) of 101.3ºF, palpable small preauricular lymph nodes, bilateral high anterior cervical lymph nodes, bilateral supraclavicular lymph nodes along the clavicle, a small left axillary node, bilateral epitrochlear nodes, and bilateral large inguinal and femoral nodes. There was no tonsillar enlargement or splenomegaly.

Laboratory findings included a WBC of $5.9x10^9$ cells/L

with a normal differential, a RBC of 14,000 cells/L, Hb of 8.5 grams/dL, a Hct of 21%, platelet count of $111x10^9$ cells/L, mean cell volume (MCV) of 82 fL, and red cell distribution width (RDW) of 17.1. Patient's total protein was 7.1 g/L; albumin was 2.4 g/L; bilirubin, transaminases, alkaline phosphatase were normal, and lactate dehydrogenase (LDH) was 361 units/L. Serum iron was 35 µmol/L, total iron-binding capacity (TIBC) was 169 µmol/L, ferritin was 2640 µg/L, B12 was 264 pg/ml, and folate was 11.6 ng/ml. Pleural fluid pH was 7.157, LDH measured 1956 units/L. Renal function, thyroid function and artherial blood gasses were all normal. A complete blood count (CBC) performed 10 days prior to admission demonstrated a WBC of 4.5x109 cells/L, hemoglobin 13.4 grams/dL, platelets 283x109 cells/L.

Examination of his peripheral blood smear demonstrated mild aniso- and poikilocytosis with increased reactive-appearing monocytes. No platelet clumps were present. The patient underwent lymph node biopsy which demonstrated a diffuse large cell non-Hodgkin's lymphoma. He was further subtyped by immunohistochemistry and was found to have a B-cell lymphoma. Since only 40% of patients with large B-cell lymphoma respond well to therapy, his physician felt that AP would benefit from a recently developed microarray method for identifying this group of patients (Alizadeh et al., 2000). The authors of the study demonstrated a difference in gene expression among the tumors of diffuse large B-cell lymphoma patients, due to a variation in tumor proliferation rate and differentiation state, as well as host response. They identified two molecularly distinct forms of the disease whose gene expression patterns indicate different stages of B-cell differentiation. The first type expressed genes characteristic of germinal center B cells and the second type expressed genes induced during in vitro activation of peripheral blood B cells. Patients with germinal center expression pattern had a significantly better overall survival than the second group. Unfortunately, the results of the microarray test showed that AP had a tumor expressing genes induced during in vitro activation of peripheral blood B cells. Despite treatment, he died three years after being diagnosed.

Ductal Carcinoma of the Breast

MS is a 45 year old female. During routine clinical examination her physician discovered a nodule two centimeters in diameter in the upper quadrant of the right breast. Retraction of the skin and nipple was present. A mammogram revealed a radiopacity with irregular margins. The echographic image demonstrated an area with irregular margins, suggestive of a breast carcinoma. The patient underwent right total mastectomy. The post-surgical histological examination diagnosed infiltrating ductal carcinoma, without metastasis. An endocrine therapy with tamoxifen was prescribed for the period of five years. The patient felt very anxious about the possibility of developing distant metastasis. In order to alleviate the anxiety her physician referred her to a study for prognosis prediction based on DNA microarray technology. The authors applied DNA microarray analysis to primary breast tumors of 117 patients, and used supervised classification to identify a gene expression signature predictive of metastases in patients without tumor cells in local lymph nodes at the time of diagnosis. In addition, they identified a signature that separates tumors of BRCA1 carriers. The study concluded that women under 55 years of age diagnosed with lymph node-negative breast cancer have a poor prognosis if their tumors express a subset of genes regulating cell cycle, angiogenesis, invasion, and metastasis (van't Veer et al., 2002).

Acute Myelogenous Leukemia

A 66 year old man presented with one week of bruising and fatigue. No lymphadenopathy or splenomegaly were observed on physical examination. Laboratory studies showed WBC of 90,000 cells/L with 90% blasts, a Hb of 8.4 g/dL, a Hct of 25%, a platelet count of 8,500 cells/L, and reticulocytes 0.5%. Blood smear showed many blasts, Auer rods, and marked thrombocytopenia. An examination of the bone marrow lead to a diagnosis of Acute Myelogenous Leukemia (AML). Further analysis demonstrated that blasts were present in the spinal fluid. Treatment was begun with daunorubicin and cytarabine, combined with intrathecal methotrexate twice weekly. After two identical courses of chemotherapy, bone marrow examination showed remission of the AML. An alternative to enzyme-based immunohistochemistry of bone marrow for distinction between AML and ALL (Acute Lymphoblastic Leukemia) is cancer classification based on gene expression profiles of acute leukemia patients. This method uses a microarray to examine the expression levels of 50 genes and classify new patient samples as AML or ALL (Golub et al., 1999).

DISCUSSION

DNA microarrays are powerful tools for improving treatment selection, diagnostic classification, and prognosis prediction. Since they offer a new and potentially better way for differentiating between tumor "types," DNA microarrays have become tools for clinical oncologists. The main objective of these studies is to identify differentially expressed genes, which characterize phenotypically defined groups. The goal of this comparison between classes is to develop a method for class prediction which would correctly identify the diagnosis, treatment, and prognosis of a patient based on gene expression data. Microarrays also offer the possibility for disease screening and prevention. For example, Dobrowolski and colleagues used DNA from blood samples to screen simultaneously for Factor V Leiden, sickle cell disease, and alpha-1-antitrypsin deficiency (Dobrowolski et al., 1999). This illustrates how microarrays can increase the number of early detected treatable diseases by development of new assays and consolidating existing ones.

DNA microarrays are already a standard laboratory technique for studying biological mechanisms, discovering new disease classification, and developing diagnostic tests. They are an integral tool for identifying new drug targets and advances in the field of pharmacogenomics. Although promising, microarray technology does not come without drawbacks. Practical problems include obtaining sufficient amounts (10ng-20µg RNA depending on the type of microarray) and appropriate specimens. For example, heterogeneity of samples such as kidney which consists of numerous cell types (podocyte, endothelial, tubular, and mesangial cells) complicates the studies of cell-specific changes in gene expression. Another issue is the data analysis problem. Microarrays provide copious amounts of data points. While there is no doubt that microarrays are part of the future of clinical research and disease treatment, there is a clear need for careful study design and statistical analysis. Explication of the biologic basis of disease and development of effective therapies will increasingly require interdisciplinary teams of biologists, computer scientists, and statisticians working together in a large scale collaborative manner.

ACKNOWLEDGEMENTS

Special thanks to Dr. Richard Hays and Dr. Shirley Levine for providing the cases on renal cell cancer and diffuse large B-cell lymphoma, respectively.

REFERENCES

Alizadeh, A., Eisen, M, Davis, E., Ma, C., Lossos, I., Rosenwald, A., Boldrick, J., Sabet, H., Tran, T., Yu, X., Powell, J., Yang, L., Marti, G., Moore, T., Hudson, J., Lu, L., Lewis, D., Tibshirani, R., Sherlock, G., Chan, W., Greiner, T., Weisenburger, D., Armitage, J., Warnke, R., Levy, R., Wilson, W., Grever, M., Byrd, J., Botstein, D., Brown, P., Staudt, L. (2002) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403:**503-511.

Dobrovolski, S., Banas, A., Naylor, E., Powdrill, T., Thakkar, D. (1999) DNA microarray technology for neonatal screening. *Acta. Paediatr. Suppl.* **432:**61-64.

Friend, S. H. and Stoughton, R. B. (2002) The Magic Of Microarrays. *Scientific American* **286:**44-53.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. (1999) Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science* **286:**531-537.

Vasselli J.R., Shih J.H., Iyengar S.R., Maranchie J., Riss J., Worrell R., Torres-Cabala C., Tabios R., Mariotti A., Stearman R., Merino M., Walther M.M., Simon R., Klausner R.D., Linehan W.M. (2003) Predicting survival in patients with metastatic kidney cancer by gene-expression profiling in the primary tumor. *Proc Natl Acad Sci. U.S.A.* **100(12):**6958-63.

Van't Veer, L., Dai, H., Van de Vijver, M., He, Y., Hart, A., Mao, M., Peterse, H., Kooy, K., Marton, M.,Witteveen, A., Schreiber, G., Kerkhoven, R., Roberts, C., Linsley, P., Bernards, R. and Friend, S. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:**530-531.

DNA Microarray Animation. Retrieved August, 2005 http://www.bio.davidson.edu/courses/genomics/chip/chip.html.