Blood Reviews (2005) **19**, 223–234

www.elsevierhealth.com/journals/blre

REVIEW

Microarray-based gene expression profiling of hematologic malignancies: basic concepts and clinical applications

Ofer Margalit[1], Raz Somech[1], Ninette Amariglio, Gideon Rechavi*

Pediatric Hematology–Oncology, Safra Children's Hospital, The Chaim Sheba Medical Center, Tel-Hashomer and Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

KEYWORDS
Microarray;
Gene expression profiling;
Clustering;
Lymphoma;
Leukemia;
Multiple myeloma

Summary Each cell in our body contains a set of tens of thousands of genes, out of which a set of several thousands determines the cell's characteristics. The deciphering of the sequence of the human genome combined with the technical feasibility to simultaneously measure the gene expression levels of thousands of genes had revolutionized our understanding of cellular processes. This ability has great significance in our comprehension of the mechanisms that bring about diseases in general and hematologic malignancies in particular. Several new high-throughput technologies, commonly referred as microarrays, enable us to perform such measurements and concurrently, bioinformatic and statistical tools were developed to analyze the data obtained by using microarrays. In this review we present examples of analyses of hematologic malignancies using microarrays which contribute to refinement of diagnosis, identification of novel disease subtypes and of relationships between diseases that were previously considered to be unrelated, prediction of response to treatment and identification of genes and pathways linked to pathogenesis, thus defining targets to rational therapy.

Introduction

The complete elucidation of the human genome and the development of microarray technology have heralded a new era for biological sciences and medicine. Vast opportunities have opened to explore the transcriptional profiles of complex diseases in a fashion unknown prior to this era. Although all cells in our body possess the same inherited genomic DNA, each cell expresses different genes as mRNA according to the cell type, biological processes, normal/abnormal conditions, etc. This diversity in gene expression pattern led to an intensive research because of its biological

* Corresponding author. Present address: Institute of Hematology, Chaim Sheba Medical Center, 52621 Tel-Hashomer, Israel. Tel.: +972 353 02397; fax: +972 353 02377.
E-mail address: gidi.rechavi@sheba.health.gov.il (G. Rechavi).
[1] These authors contributed equally to this work.

0268-960X/$ - see front matter
doi:10.1016/j.blre.2004.11.003

and clinical relevancy. The microarray technology allows us to simultaneously profile the expression of tens of thousands of genes, thus painting a molecular portrait. This ability is extremely beneficial when trying to decipher complex diseases such as cancer. Evidently, since 1999, the year in which a mature microarray technology had become available, numerous studies in cancer research have used this tool. These studies provided new insights into the biological processes underlying cancer. The significant consequences of gene expression profiling include diagnosis, stratification and prediction of prognosis in many human cancers.

The first study utilizing microarray technology demonstrated the power of this tool to classify and predict human acute leukemias. These classification and prediction were based solely on gene expression monitoring and were independent of previous biological knowledge.[1] Although Histopathological evaluation, supplemented by cytogenetics and analysis of few molecular markers, is still the gold standard in evaluating diagnosis and prognosis, gene expression profiling had proved capable of replacing these evaluations. Diagnosis and prognosis of malignancies can occasionally require the combined expertise of several practitioners, such as oncologists, pathologists and cytogeneticists, and may also vary between experts, institutions and assays. Recent results obtained from studies performed using microarrays in several malignancies indicate that it may eventually supplant the intensive labor of these practitioners. Microarrays can even further magnify the precision of diagnosis and prognosis, as well as provide a single standardized platform. Furthermore, gene expression profiling revealed several molecularly distinct subtypes of diseases, which were formerly considered the same disease, based on morphological diagnosis. Indeed, this stratification was in correlation with the response to treatment and prognosis. Notably, prediction of outcome was based on gene expression profiling of samples taken at the time of diagnosis. Therefore, genomic large scale gene expression profiling should be included in designing clinical trials, in order to refine the diagnosis and matching treatment of each malignancy. In conclusion, the emergence of the microarray technology is destined to eventually enable physicians to tailor fit therapies for cancer patients.

We review henceforth the analyses of several hematological malignancies studied in recent years, that illustrate the approaches and tools that are in use in this dynamic field. The articles cited in this review are generally ordered chronologically, within each section.

Microarray platforms and data analysis

Several methods are used to quantify and simultaneously analyze a large number of RNA transcripts. The tools used to perform these measurements include complementary DNA (cDNA) microarrays,[2] oligonucleotide microarrays[3] and serial analysis of gene expression (SAGE).[4] Out of which, the most commonly used platforms are the two microarray technologies, in which each experiment can display the levels of expression of more than 20,000 genes. The cDNA microarrays are based on standard microscopic glass slides on which cDNA fragments have been spotted. The oligonucleotide microarrays are constructed with oligonucleotides, 25–60-mer in length, that are either synthesized in situ on a silicon wafer, or robotically spotted or injected on glass slides. The term commonly used to describe the DNA arrayed on a platform is ''probe'' and the cDNA or cRNA generated from a sample RNA, which represent the gene expression profile of the sample, are referred to as ''target''. The types of probes and targets used in each microarray platform differ, but these differences are becoming less significant. The main difference between the two platforms is the method by which the mRNA levels are determined. In cDNA microarrays the quantitation is made by comparing a selected sample to a 'control' sample, while in oligonucleotide microarrays a well-defined arbitrary unit is produced, without the need to compare with a different sample. The technical procedures of the two microarray platforms are depicted in Fig. 1.

Regardless of the microarray platform, each experiment produces a data set containing tens to hundreds of thousands of values of gene expression. This overwhelming abundance of data requires the use of powerful statistical and analytical tools. There are two basic approaches to analyze gene expression data set. The supervised approach is based on determining genes that fit a predetermined pattern, usually used to correlate between gene expression and clinical data. The two most common supervised techniques are: nearest neighbour analysis[1] and support vector machines.[5] The unsupervised approach is based on characterizing the components of the data set without the a priori input or knowledge of a training signal. This approach is usually used to identify a distinct subgroup of tumors that share similar gene expression profiles. The four most common unsupervised techniques are: principle-component analysis,[6] hierarchical clustering,[7] self organizing maps[8] and relevance network.[9]