REVIEW

Microarray, SAGE and their applications to cardiovascular diseases

Shui Qing Ye[1],*, Tera Lavoie[1], David C Usher[2], Li Q. Zhang[3]

[1] Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, School of Medicine, Baltimore, MD 21224, USA
[2] Department of Biological Science, University of Delaware, Newark, DE 19716, USA
[3] Department of Surgery, University of Maryland, Baltimore, MD 21201, USA

ABSTRACT

The wealth of DNA data generated by the human genome project coupling with recently invented high-throughput gene expression profiling techniques has dramatically sped up the process for biomedical researchers on elucidating the role of genes in human diseases. One powerful method to reveal insight into gene functions is the systematic analysis of gene expression. Two popular high-throughput gene expression technologies, microarray and Serial Analysis of Gene Expression (SAGE) are capable of producing large amounts of gene expression data with the potential of providing novel insights into fundamental disease processes, especially complex syndromes such as cardiovascular disease, whose etiologies are due to multiple genetic factors and their interplay with the environment. Microarray and SAGE have already been used to examine gene expression patterns of cell-culture, animal and human tissues models of cardiovascular diseases. In this review, we will first give a brief introduction of microarray and SAGE technologies and point out their limitations. We will then discuss the major discoveries and the new biological insights that have emerged from their applications to cardiovascular diseases. Finally we will touch upon potential challenges and future developments in this area.

Key words: *Microarray, SAGE, cardiovascular diseases, RNA, gene expression profiling.*

INTRODUCTION

A draft human genome DNA sequence was reported in Science[1] and Nature[2] early in 2001. As of Dec. 31, 2001, 63.0 % of the human genome DNA sequence has been finished[3]. However, the physical map of human genome only represents one level of genetic complexity. A second and more important level is the ordered and timely expression of these genes within an organism[4],[5]. Thus, the challenge of discovery has begun to shift from the identification of genes to the elucidation of their functions.

A powerful approach involves the analysis of gene expression profiles within an organism. The global snapshots of gene expression, both among different cell types and among different states of a particular cell type can help in identifying candidate genes that may be involved in a variety of normal or disease processes. Additionally, characterization of genes abnormally expressed in diseased tissues may lead to the discovery of genes that can serve as diagnostic markers, prognostic indicators or targets for therapeutic intervention[5]. Some available methods such

* Corresponding author: Shui Qing Ye, M.D., Ph.D. Division of Pulmonary and Critical Care Medicine Johns Hopkins University, School of Medicine, 5501 Hopkins Bayview Circle, JHAAC/4A. 30A Baltimore, MD 21224 USA. Tel: (410)-550-5984, Fax: (410)-550-8159, E-mail: syq@jhmi.edu

as cDNA subtraction[6], mRNA differential display [7], EST[8] can only analyze a limited number of genes or are nonquantitative. Two newly developed high-throughput technologies, microarrays of cDNA [9] or oligonucleotide[10] and SAGE[11], allow researchers to determine the expression pattern of thousands of genes simultaneously. This promises to provide insight into the pathophysiology of human syndromes such as cardiovascular diseases, whose etiologies are due to multiple genetic factors and their interaction with the environment. As of February 2002, 51 publications have cited Microarray and SAGE methods in examining gene expression profiles for cell-culture, animal and human tissues models of cardiovascular diseases. On Feb. 17, 2002, we carried out a PubMed search using the combination of words "microarray and cardiovascular diseases", we found 32 publications. Using the combination of words "SAGE and cardiovascular diseases", we found 21 publications. Evidently, the application of microarray and SAGE to the cardiovascular disease field is still in its infancy. In this review we will first give a brief introduction of microarray and SAGE technologies and point out their limitations. We will then discuss the major discoveries and the new biological insights that have emerged from their applications to cardiovascular diseases. Finally we will touch upon potential challenges and future developments in this area.

Microarray and SAGE technologies

Microarray technology

Microarray Technology evolved from the initial report in the mid 1970s that labeled nucleic acids could be used to monitor the expression of nucleic acid molecules attached to a solid support occurred [12]. Microarrays spot cDNA samples on a solid substrate, usually a glass microscope slide, in a high-density pattern. Spots created by special pins fitted to the head of a robot currently range between 80 and 150 mm in diameter and are placed at 100 mm intervals on the slide. Although most commercial arrays contain 10,000-20,000 spots, arrays containing 80,000 spots can be created. Nucleotide arrays can also be produced on a nylon membrane with spots of slightly larger size in a lower-density pattern.

Patrick Brown's laboratory at Stanford Univer-

sity created the first xyz robot microarrayer, and instructions on how to build an arrayer can be found on their website (http://www.cmgm.stanford.edu/pbrown). Many companies now produce microarrayers commercially which primarily vary on the way the spot is placed on the substrate. Advances in array printing technology, including the development of automated spotting devices with very precise movement controls on the x-, y- and z-axes, and the improvement of the pin technology to provide very accurate, reproducible spots on the arrays, have sparked a renewed interest in microarrays within the wider research community. Spotted arrays allow a greater degree of flexibility in the choice of arrayed elements, particularly for the preparation of smaller, customized arrays for specific investigations. As a result, cDNA gridded arrays have been the technique most frequently used in academic labs. In addition, arraying of unsequenced clones from cDNA libraries can be useful for gene discovery. With prices for oligonucleotide synthesis falling, spotted long-oligonucleotide arrays could be a viable alternative for the future. A typical cDNA microarray operational schema is illustrated in Fig 1A.

Oligonucleotide microarrays[13] are produced in situ, either by photolithography onto silicon wafers (high-density-oligonucleotide arrays from Affymetrix, http://www.affymetrix.com) or by ink-jet technology (developed by Rosetta Inpharmatics, http://www.rii.com, and licensed to Agilent Technologies). Alternatively, presynthesized oligonucleotides can be printed onto glass slides. Methods based on synthetic oligonucleotides offer an advantage that production and handling cDNA is not required. A typical oligonucleotide microarray operational schema is illustrated in Fig 1B.

Affymetrix (Santa Clara, CA, USA) has pioneered the use of this form of array with the development of the Genechip. The Genechip series is designed with 11-20 different oligonucleotides to represent each gene on the array. Each oligonucleotide has a partner with a single base mismatch to control for non- specific binding (http://www.affymetrix.com/products/index.affx). The degree of hybridization to the probe compared to the single base mismatched probe allows Affymetrix software to determine hybridization errors, which in turn improves the qual-
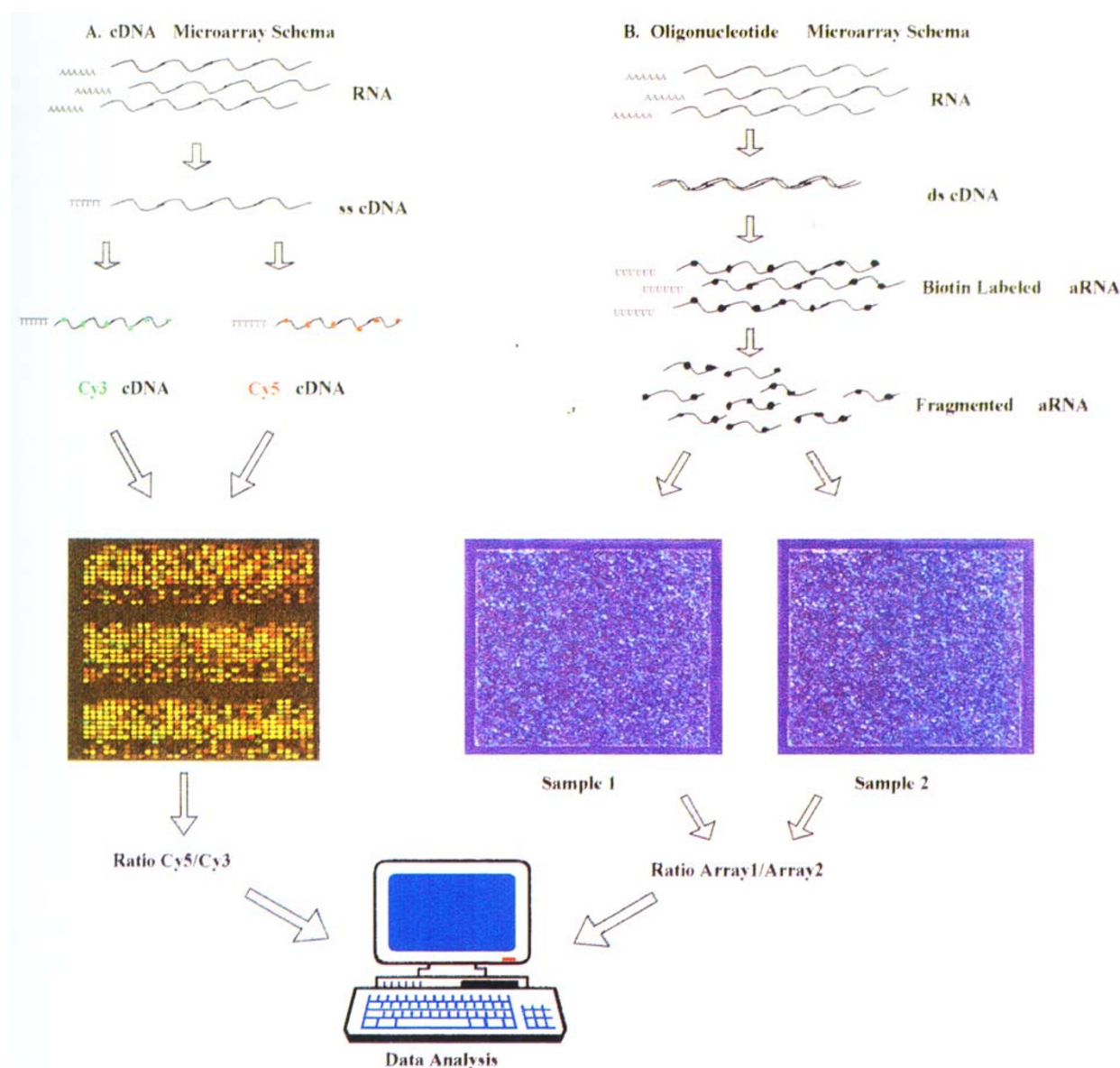
Fig 1. cDNA microarray schema (A) and oligonucleotide microarray schema (B). A. cDNA microarray schema. RNA from two different tissues or cell populations is used to synthesize single-stranded cDNA in the presence of nucleotides labeled with two different fluorescent dyes (for example, Cy3 labeled on sample A and Cy5 labeled on sample B). Both samples are mixed in a small volume of hybridization buffer and hybridized to the array surface, usually by stationary hybridization under a cover-slip, resulting in competitive binding of differentially labeled cDNAs to the corresponding array elements. High-resolution confocal fluorescence scanning of the array with two different wavelengths corresponding to the dyes used provides relative signal intensities and ratios of mRNA abundance for the genes represented on the array. Green(Cy3) spots indicate the genes upregulated in sample **A.** Red (Cy5) spots indicate the genes down-regulated in sample A. Yellow spots indicate the equal expressions of those genes in sample A and sample B. **B.** Oligonucleotide microarray schema. RNA from different tissues or cell populations is used to generate double-stranded cDNA carrying a transcriptional start site for T7 DNA polymerase. During in vitro transcription, biotin-labeled nucleotides are incorporated into the synthesized cRNA molecules. Each target sample is hybridized to a separate probe array. The arrays are stained with a streptavidin-phycoerythrin conjugate that binds to biotin tags and emits fluorescent light when exited with a laser. Arrays are scanned using the GeneChip laser confocal fluorescence scanner (Agilent GeneArray Scanner), the sensitivity of which is sufficient to detect as few as 400 phycoerythrin molecules in a 20×20-mm probe site. Automated image analysis software measures fluorescence by calculating signal intensity units at each discreet probe site or feature on the array. Signal intensities of probe array element sets on different arrays are used to calculate relative mRNA abundance for the genes represented on the array.

ity of the data.In this technology, cloning, spotting, or polymerase chain reaction (PCR) is not required. Errors inherent in cloning and PCR are thus avoided. A nucleotide of 25 bases (25mer) can efficiently be used to distinguish closely related genes. Individual genes are represented on a chip by a probe set comprised of 11 probe pairs (for the latest HG-U133 set). This allows collecting quantitative information across a range of transcript abundance. The GeneChip®Human Genome U133 Set (HG-U133A and HG-U133B, released in Jan., 2002) is comprised of two microarrays with over 1,000,000 unique oligonucleotide features covering more than 39,000 transcript variants which in turn represent greater than 33,000 of the best characterized human genes. Sequences used in the design of the array were selected from GenBank, dbEST, and RefSeq. Sequence clusters were created from Build 133 of UniGene (April 20, 2001) and refined by analysis and comparison with a number of other publicly available databases including the Washington University EST trace repository and the University of California, Santa Cruz golden-path human genome database (April 2001 release). In addition, ESTs were analyzed for low-quality sequence information, correct orientation, false priming, false clustering, alternative splicing and alternative polyadenylation.

SAGE Technology

SAGE Technology was originally developed by Velculescu et al[11]. Two basic principles underlie the SAGE methodology: 1) a short sequence tag (10 bp) within the defined position contains sufficient information to uniquely identify a transcript; 2) the concatenation of tags in a serial fashion allows for an increased efficiency in a sequence-based analysis. The SAGE procedure, depicted in Fig 2, has been extensively reviewed by Bertelsen and Velculescu [5], and Madden et al[14]. Briefly, poly(A)$^+$ RNA is isolated by oligo-dT column chromatography. cDNA is synthesized from poly(A)$^+$ RNA using a primer of biotin-5'-T18-3'. The cDNA is cleaved with an anchoring enzyme(e.g., NlaIII) and the 3'-terminal cDNA fragments are bound to streptavidin-coated beads. An oligonucleotide linker containing recognition sites for a tagging enzyme (e.g., BsmFI) is linked to the bound cDNA. The tagging enzyme is a

class II restriction endonuclease that cleaves the DNA at a constant number of bases 3' to the recognition site. This results in the release of a short tag plus the linker from the beads after digestion with BsmFI. The 3'-ends of the released tags plus linkers are then blunted and ligated to one another to form 102 bp linked ditags. After PCR amplification of the 102 bp ditags, the linkers and tags are released by digestion with the anchoring enzyme. The 26-28mer tags are then gel purified, concatenated and cloned into a sequence vector. Sequencing the concatemers enables individual tags to be identified and the abundance of the transcripts for a given cell line or tissue determined[11].

Although SAGE has become an extremely powerful technique for global analysis of gene expression, its requirement for a large amount of input mRNA (2.5-5.0 mg, which is equivalent to 250 to 500 mg total RNA) limits its utilization. Several labs attempted SAGE gene profiling using smaller amounts of RNA, but they all involved either PCR amplification of starting cDNA materials, such as SAGE-lite and PCR-SAGE[15],[16], or PCR reamplification of SAGE ditags generated by a first round of PCR amplification such as microSAGE and SADE (SAGE adaptation for downsized extracts)[17],[18]. These additional PCR amplifications potentially introduced bias and compromised the quantitative aspects of the SAGE method. A recently modified method, miniSAGE, was successfully applied to profile gene expressions of human fibroblasts from 1mg total RNA without extra PCR amplification[19]. Three key modifications contributed to the establishment of the miniSAGE: 1) The application of Phase Lock Gel (Eppendorf) to purify DNA after phenol extraction in seven steps of SAGE procedure to increase significantly the recovery of DNA during purification and get purer DNA; 2) the addition of 25 fold less linkers (10 ng/per reaction) in the ligation reaction, thus reducing its interference with the SAGE ditag amplification and increasing the SAGE ditag yield; 3) the employment of the mRNA Capture Kit (Boehringer Mannheim) to carry out the first four steps depicted in Fig 2. By isolating mRNA, synthesizing and digesting the cDNA, binding the cleaved biotin-cDNA to the streptavidin-magnetic beads, ligating linkers to the bound cDNA and releasing
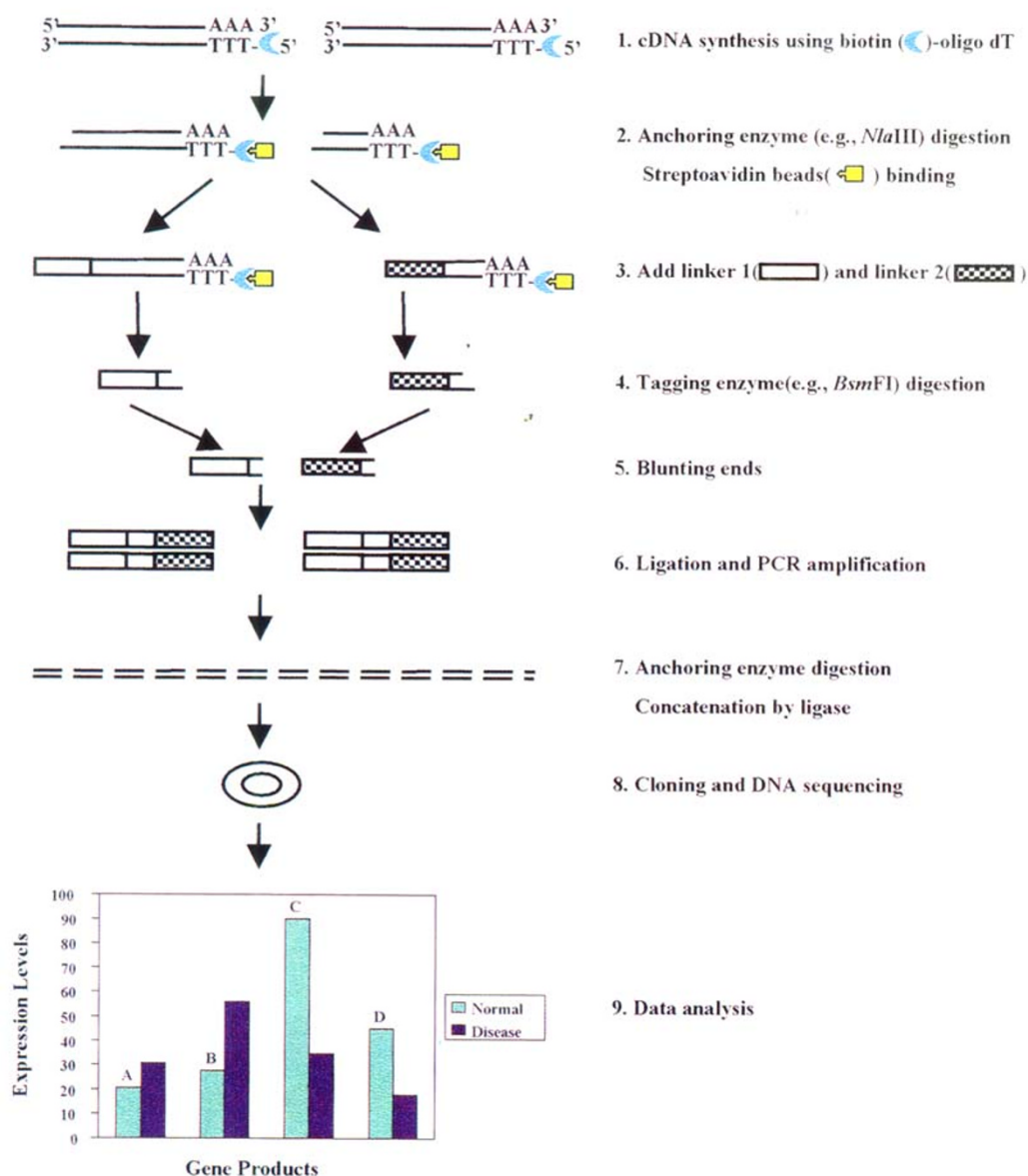
Fig 2. The schematic of the serial analysis of gene expression (SAGE) method

Step 1, cDNA is synthesized from poly (A)$^+$RNA using the biotin-oligo-dT. Step 2, the cDNA is cleaved by the anchoring enzyme (e.g., NlaIII), which is a frequent-cutting (four base-pair recognition sequence) enzyme and is expected to cleave most transcripts at least once, and 3'- portion of the cDNA is captured by the streptavidin-coated magnetic beads. Step 3, two different linkers containing a five base-pair recognition site for a type II restriction enzyme called tagging enzyme such as BsmFI are ligated to two aliquots of the captured cDNA, respectively. Step 4, the tagging enzyme BsmFI, which cleaves 14-15 bp 3' of its recognition sequence, is then used to release a short tags plus linkers are released from the cDNA by tagging enzyme digestion. Step 5, the ends of the released short tags are blunted by the treatment of the DNA polymerase I, large (Klenow) fragment. Step 6, the linker-tag molecules are ligated tail to tail to form ditags. Ditags are amplified by PCR. Step 7, ditags are released from linkers by the anchoring enzyme digestion and concatenated by ligase. Step 8, the concatemers are cloned into a sequencing vector and are subjected to DNA sequencing. Step 9, The sequencing of concatemer clones reveals the identity and abundances of each tag. Relative abundance can be calculated by dividing the observed abundance of any tag by the total number of tags analyzed to generate quantitative result. The digital data format of transcripts in both normal and diseased samples can be easily analyzed by SAGE software to identify the genes of interest in specific physiologic or disease states. This figure was modified from Ye et al.(19)

cDNA tags all in one tube, the loss of the DNA between the successive steps is significantly reduced, which greatly increases the efficiency of the procedure, which in turn lowers the amount of starting mRNA that is needed. Recently, Invitrogen[20] developed a convenient I-SAGE kit, which packages all necessary quality controlled reagents together enough for 5 SAGE library synthesis. These modifications have contributed to the broad application of SAGE.

Comparison between microarray and SAGE technology

One of the main differences between DNA microarray and SAGE is that microarray preparation requires prior knowledge of the sequence of the gene transcripts to be analyzed. This is a serious limitation, even for organisms with completely sequenced genomes such as humans, because genome annotation and gene prediction remain technical challenges. SAGE, however, can be used to analyze gene expression in organisms whose genomes are largely uncharacterized. Another advantage of SAGE over microarrays is the portability of the data generated. Although SAGE data are directly comparable, the differences in microarray formats and normalization methodologies make direct comparison of data sets between microarray platforms very difficult. SAGE not only can accurately determine the absolute abundance of mRNAs but also can detect even slight differences in expression levels between samples. Microarrays have advantages, too, of course, because they are relatively easy to use and more suitable for high-throughput applications. The choice of highthroughput gene expression technique is determined by the specific experiment to be performed. Expression profiling of hundreds of disease samples is certainly more efficient with the use of microarrays. SAGE, however, seems to be a better choice for the identification of new genes and alternatively processed transcripts that are unique to a specific cell type and for the analysis of previously uncharacterized organisms.

Microarray's application to cardiovascular diseases

So far, the utilization of microarray to examine gene expression patterns of cardiovascular diseases of either cell-culture models, animal models or human tissues can be technically classified into three categories: Parallel Approach, Temporal Approach, and Spatial Approach.

Parallel approach

Common microarray experiments on cardiovascular diseases are designed as two sample (parallel) comparisons. In this approach, microarray has been used to identify differences in gene expression between two samples representing a test condition and a control condition. The following two examples demonstrated this approach.

Tung et al[21] compared the gene expression profile of 1176 gene products in human abdominal aortic aneurysms (AAA) tissue with that in normal aorta using AtlasArray Human 1.2 I (Clontech, Palo Alto, Calif). They found that differences in expression for 20 of these gene products between AAAs and normal aorta, with the greatest increases seen for myeloid cell nuclear differentiation antigen (31-fold), cathepsin H (30-fold), platelet-derived growth factor-A (23-fold), apolipoprotein E (13-fold), gelatinase B/matrix metalloproteinase-9 (12-fold), and interleukin-8 (11-fold). The only gene products substantially decreased in AAAs were myosin light chain kinase (39-fold) and beta-1 integrin (twofold). AAA tissues thereby exhibited a distinct pattern of gene expression reflecting chronic inflammation, extracellular matrix degradation, atherosclerosis, and smooth muscle cell depletion. This study helped shedding light on the molecular mechanisms responsible for aneurysmal degeneration. Further studies will be needed to elucidate the functional and pathophysiologic significance of the individual genes that exhibit altered levels of expression in AAA tissue.

Geraci et al[22] reported that the transgenic mice overexpressing prostacyclin synthase were protected from the development of pulmonary hypertension after exposure to chronic hypobaric hypoxia. To investigate the mechanism underpinning prostacyclin synthase's protective effect, they examined the gene expression pattern of lung tissue from the transgenic mouse in comparison to that in transgene negative control. The mRNA from five transgenic mouse lungs was pooled and compared with five nontransgenic, sex-matched littermates. Using strict criteria (a twofold change in expression), they found

that of the 6500 genes surveyed, 32 genes showed an increase in expression and 26 showed a decrease in expression. Among them, peroxisome proliferator-activated receptor (PPAR)γ, RAS GTPase, focal adhesion kinase, keratinocyte growth factor, IL-7 and IL-17 receptor, cathepsins C, D, and E are significantly increased while PPAR δ, cyclooxygenase-2, α-catinin, TGF-β and its receptor are significantly decreased. This differentially expressed gene pattern has important implications for the pathogenesis and treatment of severe pulmonary hypertension.

Temporal approach

Another popular microarray experiments on cardiovascular diseases are called temporal approach. In this approach, microarray has been used to identify differences in gene expression among the different stages of diseases or different time points of treatments. Two examples employing this approach are presented below.

To seek out genes and pathways whose expression is altered at different stages in the process of cardiac hypertrophy, Friddle et al[23] screened more than 4000 genes for changes during the induction and regression of hypertrophy in a mouse model. In their study, administration of angiotensin II and isoproterenol by osmotic minipump produced increases in heart weight (15 and 45%, respectively) that returned to preinduction size after drug withdrawal. From multiple expression analyses of left ventricular RNA isolated at daily time-points during cardiac hypertrophy and regression, they identified sets of genes whose expression was altered at specific stages of this process. While confirming the participation of 25 genes or pathways previously shown to be altered by hypertrophy, a larger set of 30 genes was identified whose expression had not previously been associated with cardiac hypertrophy or regression. Of the 55 genes that showed reproducible changes during the time course of induction and regression, 32 genes were altered only during induction, and 8 were altered only during regression. This study identified both known and novel genes whose expression is affected at different stages of cardiac hypertrophy and regression and demonstrates that cardiac remodeling during regression utilizes a set of genes that are distinct from those used during induction of hypertrophy.

Shiffman et al[24] carried out a detailed time-course analysis of foam-cell formation by oxidized low-density lipoprotein (ox-LDL) treatment of macrophages. As ox-LDL-engorged macrophages constitute a large volume of early and late atherosclerotic lesions, they wanted to study the manifestation of mRNA expression patterns during foam-cell formation in a tissue-culture model. They probed a ~ 10,000-gene cDNA microarray with time-matched samples of ox-LDL-treated cells. In this study, 268 genes were differentially at least regulated at one time point. They used several approaches for gene expression analysis, including interpretation of the largest expression changes, thematic analysis of global gene expression alteration and agglomerative clustering of genes based on similarities of their expression patterns. Several biological themes emerged from the data. They observed a clear suppression of cell-cycle genes at late time points. An induction of integrin genes, which agreed with the increased adhesion phenotype characteristic of foam cells, was also observed. A complex response of pro- and anti- inflammatory genes indicated an altered inflammatory potential in these experiments. These observations elaborate on previously reported anti-inflammatory and pro-inflammatory responses of cells to ox-LDL treatment. The observed induction of the orphan receptor LXR-α and retinoid X receptor (RXR) and the previously reported induction of peroxisome proliferator- activated receptor (PPAR)- γ under these conditions points to the potential involvement of nuclear receptors in the macrophage response to ox-LDL loading.

Spatial approach

An emerging approach to apply microarray into cardiovascular diseases is called spatial approach. In this approach, microarray has been proposed to identify differences in gene expression among the different individual cells and the different locations. This approach has been heralded in the biomechanical force and atherosclerosis research. Improved precision in experimental models of spatially defined flow has recently been combined with regional and single-cell gene-expression profiling to investigate the relationships linking haemodynamics to vessel-wall pathobiology[25].

It has been recognized for many years that

haemodynamic characteristics determine the location of atherosclerotic lesions. The pulsatile nature of blood flow generates a complex interplay of three distinct types of fluid mechanical forces: wall shear stresses, cyclic strains, and hydrostatic pressures[26]. These hemodynamic factors not only influence the structure and function of the heart as a pump, but also act on the cells that comprise the walls of the distributive vascular network. Local vessel geometries (e.g. arterial branching and curvatures), and constraints on vessel motion by the surrounding tissues (e.g. coronary arteries) lead to flow instabilities and separations that correlate with sites of lesion development. Thus, the localization of atherosclerosis is determined primarily by the local flow conditions. Intervention to target a small number of 'pro-atherosclerotic' endothelial cells will become a reasonable strategy as the principal regulatory mechanisms of endothelial haemodynamic dysfunction that are associated with the disease process become better understood. Because the haemodynamics and location of lesions are defined spatially, transcriptional profiling of endothelial cells as a function of location is a promising approach to this goal. With some refinement of methods of cell isolation, it should be possible to profile endothelial cells at sites of lesion susceptibility in intact arteries. Davies et al [25] proposed that differences in haemodynamic signalling and gene expression that have been observed from region to region and from cell to cell in endothelium (both in culture and in tissues) are likely to arise from microscopic topographic differences at the interface of the fluid and the cell surface. Examples of this heterogeneity include the variable expression of endothelial vascular-cell-adhesion-molecule 1 and intercellular-adhesion-molecule 1 proteins from cell to cell in vivo and in vitro, the elevation of intracellular calcium measured in vitro and in vivo, the induction of synthesis and nuclear localization of c-Fos in vitro, and the expression of major-histocompatibility-complex antigens in vitro. In these studies, highly variable responses were observed in adjacent cells of the same endothelial monolayer exposed to a nominally identical flow field. An important implication of these findings is that a limited number of strongly expressed genes in only a few endothelial cells can play a dominant role in vascular regulation and pathogenesis. However, their identity may be masked because of message dilution in mRNAs isolated from a larger pool of cells. Quantitative profiles of gene expression, including the use of high-throughput hybridization to screen many genes simultaneously, now allow endothelial heterogeneity to be addressed in a detailed (single cell) yet comprehensive (multiple genes, high-throughput) approach that increases the probability of finding new therapeutic targets.

SAGE's application to cardiovascular diseases

SAGE was successfully applied to determine the yeast transcriptome[27], to obtain gene expression patterns in the different human tissues of both normal and diseased status[28], and to identify a new marker for colorectal cancer metastasis, PRL-3 gene [29]. Recently, based on over 2.45 million SAGE transcript tags from 12 tissue types, Caron et al[30] was able to derive a global perspective from the partial human transcriptome map: chromosome 4, 13, 18, and 21 show an overall low gene expression. The latter three chromosomes are responsible for most constitutional trisomies, suggesting that the low expression and low gene density could limit the lethality of an extra copy of them.

SAGE's application to the cardiovascular disease filed is just a beginning. The following three examples illustrated the usage of SAGE in this field.

To assess the endothelial cell response to an atherogenic stimulus

Activation of human arterial endothelial cells (ECs) is an early event in the pathogenesis of atherosclerosis. To identify the repertoire of genes that are differentially expressed after activation, de Waard et al[31] used serial analysis of gene expression (SAGE) to compare the mRNA spectrum of quiescent ECs with that of ECs activated for 6h with a strong atherogenic stimulus, oxidized LDL. About 5% of over 12000 tags analyzed were derived from genes that are differentially expressed (at least 5-fold up- or downregulated). These transcript tags are derived from only 56 genes, close to 1% of the total number of analyzed genes. Among these 56 differentially expressed genes are 42 known genes, including the hallmark endothelial cell activation markers interleukin 8 (IL-8), monocyte chemoattractant pro-

tein 1 (MCP-1), vascular cell adhesion molecule 1 (VCAM-1), plasminogen activator inhibitor 1 (PAI-1), Gro-α, Gro-β and E-selectin. Differential transcription of a selection of the upregulated genes was confirmed by Northern blot analysis. A novel observation is the upregulation of activin βA mRNA, a member of the transforming growth factor beta family. The authors demonstrated that for the application of SAGE, a moderate number of analyzed transcript tags suffice to reveal the significant alterations of EC transcription that results from a strong atherogenic stimulus.

To profile gene expressions of LPS-stimulated human monocytes

Monocytes play a pivotal role in various human infectious and inflammatory diseases. To reveal a whole picture of pathophysiologic function of activated human monocytes, Suzuki et al[32] used the serial analysis of gene expression (SAGE) procedure in lipopolysaccharide (LPS)-stimulated human monocytes. A total of 35,874 tags corresponding to more than 12,000 different transcripts were sequenced. Comparison of gene expression profile with that of resting monocytes revealed the LPS-inducible gene expression profile. Many cytokines and chemokines, including interleukin (IL)-6, IL-1α, IL-1β, tumor necrosis factor (TNF)-α macrophage inflammatory protein (MIP)-1β, MIP-2 β, MIP-2α, liver and activation- regulated chemokine (LARC), MIP-1α, thymus and activation-regulated chemokine (TARC), macrophage-derived chemokine (MDC), regulated on activation, normal T cell expressed and secreted (RANTES), growth-regulated oncogene (GRO) α, and IL-8, were observed in the highest inducible transcripts. Other genes encoding plasminogen activator inhibitor type 2 (PAI-2), Hc-gp39, apolipoproteins, malate dehydrogenase, matrix metalloproteinase-9 (MMP-9), and cyclooxygenase (COX2) were also highly elevated in LPS- stimulated monocytes. Moreover, up-regulation of Naf1 β, IL-7 receptor, adenosine receptor A2a, and many novel genes was newly identified. These results suggest that the LPS-inducible gene products may be involved in cell activation and migration, angiogenesis, tissue remodeling, and metabolism, and thus may orchestrate the inflammatory reactions. On the other hand, the expression of numerous sets of novel genes was discovered to be down-regulated

on LPS stimulation. This study represents the first comprehensive analysis of LPS-inducible gene expression in human monocytes and provides tremendous novel information for the function of LPS-activated monocytes and targets for diagnosing, monitoring, and treating sepsis and various human infectious and inflammatory diseases. Since atherosclerosis is an inflammatory disease and monocyte/macrophage plays an important role in atherogenesis[33], information derived from this study will be of great interest to scientists in the field of atherosclerosis research.

To dissect the molecular anatomy of an intracranial aneurysm

Approximately 6% of human beings harbor an unruptured intracranial aneurysm. Each year in the United States, > 30,000 people suffer a ruptured intracranial aneurysm, resulting in subarachnoid hemorrhage. Despite the high incidence and catastrophic consequences of a ruptured intracranial aneurysm and the fact that there is considerable evidence that predisposition to intracranial aneurysm has a strong genetic component, very little is understood with regard to the pathology and pathogenesis of this disease. To begin characterizing the molecular pathology of intracranial aneurysm, Peters et al [34] used a global gene expression analysis approach (SAGE-Lite) in combination with a novel data-mining approach to perform a high-resolution transcript analysis of a single intracranial aneurysm, obtained from a 3-year-old girl. SAGE-Lite provides a detailed molecular snapshot of a single intracranial aneurysm. The study indicates that the aneurysmal dilation results in a highly dynamic cellular environment in which extensive wound healing and tissue/extracellular matrix remodeling are taking place. Specifically, authors observed significant overexpression of genes encoding extracellular matrix components (eg, COL3A1, COL1A1, COL1A2, COL6A1, COL6A2, elastin) and genes involved in extracellular matrix turnover (TIMP-3, OSF-2), cell adhesion and antiadhesion (SPARC, hevin), cytokinesis (PNUTL2), and cell migration (tetraspanin-5). Although these were preliminary data, representing analysis of only one individual, the study presented a unique first insight into the molecular basis of aneurysmal disease and define numerous candidate

markers for future biochemical, physiological, and genetic studies of intracranial aneurysm. Products of these genes will be the focus of future studies in wider sample sets.

Future perspectives

Analyses of gene expression patterns may soon result in direct improvement in the diagnosis and treatment of cardiovascular diseases. Availability of early diagnostic markers could greatly reduce morbidity and mortality for many illnesses, including cardiovascular diseases. Genes that are highly expressed in disease tissues, especially if they encode secreted protein, could be measured in the blood or other body fluids. The hope is that early diagnosis, especially of diseases with insidious onsets, would lead to specific interventions and better outcomes. Improvements in staging of a variety of pathological states might be realized by the identification of new tissue-or state-specific markers. Analysis of gene expression differences in treatment responders vs. nonresponders could delineate difference between various patient populations and provide insight into the mechanism of action of different treatments such as statins on cardiovascular diseases. In addition, gene expression patterns could be useful in identifying new targets for therapeutic agents. Comprehensive gene expression approaches like microarray and SAGE will have a fundamental role in providing basic information integral to biologic and clinical investigation for years to come.

REFERENCES

[1] Venter JC, Myers EW, PW Li et al. The sequence of the human genome. Science 2001; **291**:1304-51.

[2] Lander ES, Linton LM, Birren B et al. Initial sequencing and analysis of the human genome. Nature 2001; **409**: 860-921.

[3] http://www.ncbi.nlm.nih.gov/genome/seq/, March 14, 2002.

[4] Velculescu VE. Essay: Amersham Pharmacia Biotech & Science prize. Tantalizing transcriptomes-SAGE and its use in global gene expression analysis. Science 1999; **286**:

1491-2.

[5] Bertelsen, AH, Velculescu VE. High-throughput gene expression analysis using SAGE. DDT 1998; **3**:152-9.

[6] Hedrick SM, Cohen DI, Nielsen EA and Davis MM. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature 1984; **308**:149-53.

[7] Liang, P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 1992; **257**:967-71.

[8] Adams, MD, Kerlavage AR. Fleischmann RD et al. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 1995; **377**:3-174.

[9] Schena, M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995; **270**:467-70.

[10] Lockhart, DJ, Dong H, Byrne, MC et al. Expression monitoring by hybridization to high- density oligonucleotide arrays. Nat. Biotechnol. 1996; **14**:1675-80.

[11] Velculescu, VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995; **270**:484-7.

[12] Burgess JK. Gene expression studies using microarrays. Clin Exp Pharmacol Physiol. 2001; **28**:321-8.

[13] Schulze A, Downward J. Navigating gene expression using microarrays - a technology review. Nat Cell Biol. 2001; **3**:E190-5.

[14] Madden SL, Wang CJ, Landes G. Serial analysis of gene expression: from gene discovery to target identification. DDT 2000; 5:415-25.

[15] Peters DG, Kassam AM, Yonas H, O'Hare EH, Ferrell RE and Brufsky AM. Comprehensive transcript analysis in small quantities of mRNA by SAGE-Lite. Nucleic Acids Res 1999; **27**:e39(i-vi).

[16] Neilson L, Andalibi A, Kang D et al. Molecular phenotype of the human oocyte by PCR- SAGE. Genomics 2000; **63**:13-24.

[17] Datson NA, Jong J, van den Berg MP et al. MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. Nucl Acids Res 1999; **27**:1300-7.

[18] Virlon B, Cheval L, Buhler J-M, Billon E, Doucet A. J-M Elalouf. Serial microanalysis of renal transcriptomes. Proc Natl Acad Sci USA 1999; **96**:15286-91.

[19] Ye SQ, Zhang LQ, Zheng F, Virgil D, Kwiterovich PO. miniSAGE: gene expression profiling using serial analysis of gene expression from 1 microgram total RNA. Anal Biochem 2000; **287**:144-52.

[20] Cat.No. T500001, Invitrogen Corporation, Carlsbad, CA, USA (www.invitrogen.com)

[21] Tung WS, Lee JK, Thompson RW. Simultaneous analysis of 1176 gene products in normal human aorta and abdominal aortic aneurysms using a membrane-based complementary DNA expression array. J Vasc Surg 2001; **34**:143-50.

[22] Geraci MW, Gao B, Hoshikawa Y, Yeager ME, Tuder RM, Voelkel NF. Genomic approaches to research in pulmonary hypertension. Respir Res 2001; **2**:210-5.

[23] Friddle CJ, Koga T, Rubin EM, Bristow J. Expression

profiling reveals distinct sets of genes altered during induction and regression of cardiac hypertrophy. Proc Natl Acad Sci USA 2000; **97**:6745-50.

[24] Shiffman D, Mikita T, Tai JT et al. Large scale gene expression analysis of cholesterol- loaded macrophages. J Biol Chem 2000; **275**:37324-32.

[25] Davies PF, Polacek DC, Handen JS, Helmke BP, DePaola N. A spatial approach to transcriptional profiling: mechanotransduction and the focal origin of atherosclerosis. Trends Biotechnol 1999; **17**:347-51.

[26] Garcia-Cardena G, Comander J, Anderson KR, Blackman BR and Gimbrone MA Jr. Biomechanical activation of vascular endothelium as a determinant of its functional phenotype. Proc Natl Acad Sci USA 2001; **98**:4478-85.

[27] Velculescu VE, Zhang L, Zhou W et al. Characterization of the yeast transcriptome. Cell 1997; **88**:243-51.

[28] Velculescu VE, Madden SL, Zhang L et al. Analysis of human transcriptomes. Nat Genet 1999; **23**:387-8.

[29] Saha S, Bardelli A, Buckhaults P et al. A phosphatase associated with metastasis of colorectal cancer. Science 2001; **294**:1343-6.

[30] Caron H, van Schaik B, van der Mee M, et al. The Human Transcriptome Map: Clustering of highly Expressed Genes in Chromosomal Domains. Science 2001; **291**: 1289-92.

[31] de Waard V, van den Berg BM, Veken J, Schultz-Heienbrok R, Pannekoek H, van Zonneveld AJ. Serial analysis of gene expression to assess the endothelial cell response to an atherogenic stimulus. Gene 1999; **226**:1-8.

[32] Suzuki T, Hashimoto S, Toyoda N et al. Comprehensive gene expression profile of LPS-stimulated human monocytes by SAGE. Blood 2000; **96**:2584-91.

[33] Ross R. Atherosclerosis-an inflammatory disease. N Engl J Med 1999; **340**:115-26.

[34] Peters DG, Kassam AB, Feingold E et al. Molecular anatomy of an intracranial aneurysm: coordinated expression of genes involved in wound healing and tissue remodeling. Stroke 2001; **32**:1036-42.