REVIEW

Identification of disease genes by expression profiling

R. Bals*, B. Jany[#]

Identification of disease genes by expression profiling. R. Bals, B. Jany. ©ERS Journals Ltd 2001.

ABSTRACT: The human genome has been completely sequenced. The development of innovative methodologies and tools to understand the functions of human genes in health and disease will allow the data of the human genome project to be utilized.

This paper reviews methods that can be used to detect and isolate genes that are specifically expressed in certain diseases or that are specific to cell types. First, classical methods, such as differential screening of complementary deoxyribonucleic acid libraries and subtractive techniques, are described. Methods based on polymerase chain reaction (PCR), such as differential display PCR or serial analysis of gene expression, will then be discussed. Finally, recent developments in gene chip technology and basic principles of functional genomics will be illustrated.

Future developments will link the results of genomic approaches to data obtained by other systematic methods, such as proteomics (*i.e.* the systematic, large scale analysis of proteins), and will allow the production of a detailed molecular characterization of diseases, disease stages, tissues, or cell types.

Methods to detect disease or cell type-specific gene expression patterns will play an important role in the future of basic research, as well as the development of novel diagnostic procedures and identification of therapeutic targets.

Eur Respir J 2001; 18: 882–889.

*Medical Dept, Division of Pulmonology, Hospital of the University of Munich, Munich, and [#]Missionsärztliche Klinik, Würzburg, Germany.

Correspondence: R. Bals, Hospital of the University of Munich, Medizinische Klinik und Poliklinik I Großhadern, 81377 Munich, Germany.
Fax: 49 8970958877

Keywords: Differential display, gene expression, genomics, microarray chip analysis, proteomics

Received: January 5 2001

Accepted after revision June 12 2001

Work related to differential gene expression in the authors' laboratory was supported by grants of the Friedrich-Baur-Stiftung (Munich, Germany) and the Deutsche Forschungsgemeinschaft (Ba 1641/3-1).

The human genome is estimated to consist of 30,000–50,000 genes, up to 10–20% of which may be expressed in a cell at a certain time. The human genome has been completely sequenced by different approaches and huge amounts of data are being presented in public databases [1, 2]. Furthermore, a large number of human-expressed sequence tags, small complementary deoxyribonucleic acid (cDNA) sequences that represent parts of the transcript pool of a certain tissue and that are produced in a highly automated procedure, provide further information for sequence comparison and the detection of new genes [3]. The study of expressed genes has had a great impact on biological research [4]. However, these expression data do not provide information about the function of the genes or the significance of their expression in the specific state of development, cell differentiation, or disease. Genome sequencing projects and large-scale expression studies supply scientists with an enormous amount of potentially meaningful data. The challenge moves from identifying the parts of the human genome, to understanding their function in health and disease, a field that has been called "functional genomics" or "post-genomic area" [4]. When the Human Genome Project was initiated in the late 1980s, the scientific community debated the sense of a large-scale data acquisition approach in biomedicine. Subsequently, it has become clear that genomics

will have a great influence on the future of drug discovery and will change medicine profoundly. In the present review, the authors describe methodologies that can be used to detect genes that are differentially expressed in certain stages of development, in diseases, in response to therapy, or in specific cell types or tissues. First, classical methods, such as differential screening of cDNA libraries, subtractive hybridization and subtracted libraries, will be described. Polymerase chain reaction (PCR)-based approaches, such as differential display PCR (DD-PCR) and serial analysis of gene expression (SAGE), will then be explained. Finally, developments in gene chip approaches will be illustrated. Table 1 summarizes the methods explained in this article.

Classical methods to detect differentially-expressed genes

Differential screening of complementary deoxyribonucleic acid libraries

This method uses the transcript pools of two different cell types or tissues to screen a cDNA library, with the aim of finding sequences that are present in only one pool of transcripts. In detail, polyadenylated ribonucleic acid (poly (A)⁺ RNA) is

Table 1. – Overview of methods described in the present article

Methods	Short description
Classical methods to detect differentially-expressed genes	
Differential screening of cDNA libraries	Comparison of two transcript pools using a reference pool cloned into a cDNA library
Absorbed probes and subtracted libraries	Enrichment of rare transcripts that are used in subsequent procedures as an absorbed probe to screen cDNA libraries for specific clones or as a template to generate a subtracted cDNA library
PCR-based methods to detect differentially-expressed genes	
DD-PCR	PCR-based method to identify differentially-expressed genes
Serial analysis of gene expression	PCR-based method to monitor and compare different expression pools
Functional genomics and gene chips	
Gene chips	Prototypic genomic methodology to measure the abundance of several thousands of known DNA sequences
Proteomics	Systematic, large-scale analysis of proteins

DNA: deoxyribonucleic acid; cDNA: complementary DNA; PCR: polymerase chain reaction; DD-PCR: differential display PCR.

extracted from two samples of interest and used as a template to generate radiolabelled probes (fig. 1a). These are separately hybridized to duplicate copies of a cDNA library of interest [5]. In general, a cDNA library represents the pool of transcripts of a certain cell type or tissue that have been converted into cDNA and cloned into suitable vectors, such as lambda phages. To screen such libraries for a specific cDNA sequence, many vectors, in addition to the inserts, are poured onto agar plates together with host bacteria. The bacteria are then lysed by phage infection, resulting in visible plaques. Filters are lifted and hybridized to a labelled probe to detect plaques that correspond to the sequence of interest. To perform differential screening of cDNA libraries, duplicate filters are lifted from the same plate and used for separate hybridizations with the two radiolabelled cDNA probes of interest (fig. 1a). Clones that hybridize to both probes correspond to genes that are represented in both messenger ribonucleic acid (mRNA) pools, whereas clones that only hybridize to one probe correspond to mRNAs that are expressed differentially in only one pool of transcripts. These clones are isolated and further analysed. Differential screening is a classical method and has been used extensively to detect differentially-expressed genes (*e.g.* genes involved in cell differentiation in the respiratory system); however, it is quite laborious and only works effectively for those genes that are expressed in abundance in one of the two tissues or cell types [5]. Differential screening of cDNA libraries was used to identify genes that are upregulated in bronchial epithelial cells by treatment with retinol. Elongation factor-1 α is one of the genes that was induced in this experimental setup [6].

Subtractive methods: absorbed probes and subtracted libraries

When a gene is differentially expressed but its mRNA is not expressed in abundance, it is possible to increase the effective concentration of the sequence

by subtractive hybridization [7]. Approaches that use subtractive methods increase the percentage of relevant nucleotide sequences by elimination of non-relevant sequences. Poly (A)⁺ RNA isolated from one sample (the "tracer") is used to synthesize labelled cDNA, which is hybridized to an excess of poly (A)⁺ RNA (the "driver") from the other cell type or tissue (fig. 1b). cDNA and mRNA can be used either as the driver or the tracer. cDNAs corresponding to genes expressed in both tissues or cells will form DNA/RNA hybrids, and together with a nonhybridized driver, they can be separated from single-stranded cDNAs by different methods, *e.g.* by chromatography on hydroxyapatite columns or the use of a biotinylated or immobilized driver [8]. The pool of cDNA enriched in cell type- or tissue-specific transcript pools can then be used as a so-called "absorbed" probe to screen cDNA libraries for specific clones or as a template to generate a subtracted cDNA library. Methods using subtracted probes or templates are limited by the amount of mRNA that is initially available. After two or more rounds of hybridization and isolation, the yield of specific cDNA is limited in quantity and often not enough is available to perform subsequent steps. However, due to the >100–1,000-fold enrichment of specific clones, these methods can also detect genes expressed differentially at a low abundance [7]. To identify glucocorticoid-induced genes in foetal lung fibroblasts, WANG *et al.* [9] screened a cDNA library from cortisol-treated foetal lung fibroblasts with a subtracted cDNA probe, which was enriched for sequences specific to cortisol-treated foetal lung fibroblasts. Analysis of isolated clones identified rat transforming growth factor- β 3 as one gene induced by steroids [9].

Polymerase chain reaction-based methods to detect differentially-expressed genes

Differential display polymerase chain reaction

Methods using PCR to detect differentially-expressed genes are based on the fact that when

Fig. 1.—Classical methodologies of the differential screening of complementary deoxyribonucleic acid (cDNA) libraries and the use of subtractive methods. a) Differential screening of cDNA libraries is used to identify cell type-specific transcripts. Messenger ribonucleic acid (mRNA) is isolated from two different samples (A and B), reverse transcribed and labelled. The two samples of transcripts are compared to a reference cDNA library of the tissue or cell type of interest by hybridizing the labelled probes to duplicate lifts. Plaques that reveal differential patterns between the two lifts are isolated and further analysed. b) Subtractive methods aim to enrich the relative amount of a subset of transcripts that is associated with a certain cell type or cell differentiation. To generate absorbed probes, "tracer" nucleic acid (mRNA or cDNA) is isolated and hybridized to "driver" nucleic acid, which was previously obtained from nonspecific tissue or cells. After removal of tracer-driver hybrids and the unhybridized driver, the sample should contain sequences specific for the tracer. This sample is then used for screening or construction of libraries.

PCR is performed on a complex cDNA sample using one or two arbitrary primers, a reproducible fingerprint of the template is produced. Differences in these fingerprints are due to differences in the template material. DD-PCR provides a picture of the transcript pool of cells or tissues by displaying subsets of mRNAs. Subsets obtained from different cell types or tissues can be compared and used for isolation of the genes of interest. The general strategy for DD-PCR consists of the combination of: 1) reverse transcription using an anchor primer; 2) performing PCR with the anchor primer and an arbitrary primer; and 3) separation of the PCR product by electrophoresis and visualization (fig. 2a). The basic protocol for differential display described an initial reverse transcription of isolated mRNA using a deoxythymidine (dT) oligonucleotide (oligo(dT)_n) (n=8–10) primer with an "anchor" of two bases at its 5' end [11]. The use of 5'-oligo (dT)NN-3' for the reverse

transcription results in priming mainly at the 5' end of the poly (A)⁺ tail and partly in adenine-rich regions of the mRNA. Selectivity is introduced by the two anchor bases, theoretically approaching one out of 12 polyadenylated RNAs. The resulting cDNA is used as a template for a PCR using one arbitrary primer (*i.e.* a primer having a single random base at each position) and the anchor primer used for the reverse transcription (fig. 2a). Reverse transcription and PCR are performed for each transcript pool of interest. PCR products are then separated by gel electrophoresis, visualized by radioactive or nonradioactive methods and compared (fig. 2b). Bands found in only one sample are likely to represent differentially-expressed genes and are cut out, reamplified by PCR and cloned. In recent years, DD-PCR has been intensively used to isolate genes that are specific to lung differentiation [12], respiratory diseases, and cell types [10, 13]. Figure 2b illustrates the identification of transcripts

Fig. 2.—Illustration of differential display polymerase chain reaction (DD-PCR) to identify differentially-expressed genes. DD-PCR compares two transcript pools and allows the detection of differentially-expressed genes. a) Messenger ribonucleic acid (mRNA) is isolated from the samples of interest and reverse transcribed using a deoxythymidine (dT) oligonucleotide anchor primer (oligo(dT)_n) with an "anchor" of two bases at its 5' end. A polymerase chain reaction (PCR) is performed using the anchor and one random primer. The PCR products are separated by gel electrophoresis and a sample-specific band is cut out, cloned and analysed. b) Display of an example of nonradioactive DD-PCR [10]. mRNA of mucous and serous gland cells of the porcine airways were used to perform a differential display. PCRs with different primer combinations (1, 2, and 3) were performed in duplicate to control for reproducibility. Four serous cell-specific bands were identified (4), cut out and further analysed [10]. The results for three primer combinations are displayed, each in duplicates for serous (s) and mucous (m) cells. cDNA: complementary deoxyribonucleic acid.

that are expressed in serous gland cells of porcine airways. After enzymatic digestion of large airways, glandular cells were separated and analysed by differential display [10]. Detailed protocols for radioactive or nonradioactive DD-PCR are described elsewhere [14]. Differential screening is technically straightforward, although the percentage of false-positive results is high and thorough screening of the obtained sequences is necessary.

Serial analysis of gene expression

SAGE is a streamlined method of sequencing parts of cDNAs obtained from any type of transcript source to determine which genes are expressed and to quantify their level of expression [15]. Each mRNA population to be analysed by SAGE requires the preparation of a library containing SAGE tags that represent individual genes expressed in the cells

(fig. 3). For the production of SAGE libraries, mRNA is converted into cDNA, which is used to generate transcript-specific tags of 10 base pairs (bp) in length, obtained from a precise location relative to the 3' end of the individual mRNAs. Individual tags are linked together by subsequent digestions with certain endonucleases. They are then amplified by PCR using primers complementary to the adaptor sequences linked to the ends of the tags. Finally, individual tags are ligated and cloned into a conventional plasmid. Each SAGE library contains $\sim 2 \times 10^6$ tags with an individual cloned insert of 500 bp in size (40 SAGE tags). For most projects, up to 2,000 SAGE clones are sequenced to yield 50,000 SAGE tags. After the tags are sequenced, transcript profiling relies on computational data generation and analysis, which include: 1) detection of tag sequence and tabulation; 2) comparing tag abundances between individual libraries; and 3) database searches using the SAGE tags. SAGE has been used to monitor the expression

Fig. 3.—Serial analysis of gene expression (SAGE) allows the detection of formerly unknown, differentially-expressed transcripts and to monitor the expression levels of known genes to be monitored. As the first step, a library is generated that represents the transcript pools. To generate a SAGE library sample, messenger ribonucleic acid (mRNA) is reverse transcribed using biotinylated primers. After digestion with a so-called "anchoring" enzyme (here *Nla* III), a restriction endonuclease that would be expected to cleave most transcripts at least once, the 3' parts of the complementary deoxyribonucleic acids (cDNAs) are immobilized by biotin. The transcripts are divided and two different adaptor sequences are linked to the 5' end of the cDNAs containing a IIS restriction enzyme site. IIS restriction enzymes cleave at a defined distance from their recognition site and result, in this case, in the generation of liberated 5' parts of the cDNA, including the adaptor sequence and a single tag. After ligation of the tags, the ditags are amplified by polymerase chain reaction (PCR). Finally, the adaptor enzyme is used again to eliminate the adaptor sequences at the 3' and 5' ends of the ditags. The resulting sequences are concatenated and cloned into a suitable plasmid. In the analysis step, the inserts are sequenced and subjected to data analysis, including comparison to databases and libraries from other cellular sources.

of a large number of genes as well as the identification of novel genes. As an "open" system, SAGE can be applied to reveal genes, which are expressed without using a predetermined array of target genes [16]. SAGE has been used in several studies of lung biology, for example, for the detection of genes differentially regulated in lung cancer cells [17]. In this study, SAGE was applied to systematically analyse transcripts present in nonsmall cell lung cancer cells. SAGE tags ($n=226,000$) were sequence analysed from two independent primary lung cancers and two normal

human bronchial/tracheal epithelial cell cultures, and several transcripts associated with cancer phenotypes were identified.

Functional genomics and gene chips

Novel and powerful methods to analyse the expression of a large number of genes at the transcript and protein levels have been developed in the fields of genomics and proteomics. In the postgenomic area (the time after genomes have been completely sequenced), the focus shifts from the generation and acquisition of sequence information to a more functional view that involves various approaches to generating and understanding large sets of interconnected data [4]. Gene chips (also called DNA microarrays or oligonucleotide microarrays) represent the prototypical methodology that has been developed in parallel with the sequencing of the human genome [18–20]. Gene chips allow the expression of large numbers of genes to be determined, and will ultimately be used to perform genome-scale expression analysis. cDNA tags for thousands of potentially expressed genes are arrayed on glass or membrane support. The chips are then hybridized to a labelled cDNA probe of interest and analysed. Several different technical variations of this basic principle have been used [21] and are summarized in table 2. Typical steps in an array experiment include the following.

- 1) **Sample preparation and mRNA isolation.** Using conventional methods, mRNA is isolated from the tissues of interest. In most cases, cell cultures provide homogenous cellular material, whereas material obtained from patients may have to be microdissected in order to enrich the cell type of interest.
- 2) **cDNA generation and labelling.** cDNA is generated using reverse transcription. This step often involves the labelling the DNA using different reagents, such as radioactive or fluorescent substances. Labelled RNA can also be used in this step.
- 3) **Hybridization to the microarray.** The labelled nucleic acid is hybridized to the arrayed DNA using protocols optimized for the specific gene chip system. A single probe, or two probes labelled with different dyes, can be applied.
- 4) **Imaging.** After washes, the bound probes have to be detected using imaging tools that depend on the labelling method.
- 5) **Data analysis.** A critical step in the whole procedure is an appropriate analysis of the obtained data. This involves normalization to correct for differences in sampling, collection of raw data, assembly of an output format, and cluster analysis. Finally, data has to be integrated with the results of other systematic approaches. These steps are schematically displayed in figure 4. Critical to the success of the whole experiment is the analysis of the created data and the generation of an interactive virtual map, connecting different aspects of the underlying biological process (*e.g.* the phenotype, gene expression, protein abundance, and other cell biological categories), which are accessible by means of computerized approaches [20, 22]. The potential application outlined here describes an observational approach to investigating expression profiles of cells or diseases.

Table 2. – Technical approaches for monitoring gene expression using gene array technology

Support material	Arrayed DNA	Method used to array DNA	Label substance of the probe	Detection system
Nitrocellulose	cDNA produced by PCR (300–2000 bp)	Photolithography (on chip UV-light directed synthesis of polynucleotides)	Radioactive substance	Phosphoimaging for radioactive labelling
Glass slide	Oligonucleotides (20–50 bp)	Mechanical gridding (high-precision ink-jet or deposition-based methods)	Fluorescent dyes	Laser excitation of fluorescent molecules incorporated into the nucleic acid target

DNA: deoxyribonucleic acid; cDNA: complementary DNA; PCR: polymerase chain reaction; bp: base pairs; UV: ultraviolet. Different types of nucleic acids are used to produce an array on the support material. The probe used for the hybridization can be labelled with different substances that finally determine the kind of detection equipment to be used.

Microarray analysis of the genes involved in pulmonary fibrosis in a murine model detected clusters of genes associated with different cellular programmes,

Fig. 4. – Microarray chip analysis of gene expression. Microarray chips can be produced by various methods based on known genes or expressed sequence tag (EST) sequences. Different steps of a prototypical expression profiling experiment are displayed, which include sample preparation, messenger ribonucleic acid (mRNA) isolation and complementary deoxyribonucleic acid (cDNA) generation, labelling procedure, hybridization to the microarray, imaging, and data analysis. Finally, the results have to be validated using independent methods. The figure also shows an illustration of a commercially available oligonucleotide array from Affymetrix Inc. (Santa Clara, CA, USA) and a scanned micrograph of a labelled gene chip.

such as inflammation and fibrosis [23]. Expression screening can also be used for drug testing and toxicology purposes [24], for analysing metabolic pathways, or for the molecular classification of diseases [23]. In addition to the detection of differentially-expressed genes in a "see what happens" approach, gene chips have several other applications. For example, microarrays have been used to study normal and cystic fibrosis (CF) cells to evaluate the cellular response to drugs [25]. This approach may eventually be used to generate pharmacogenomic data for surrogate end-point testing of novel therapeutics in cell cultures prior to animal experiments or human studies. Gene chips have also been used to diagnose mutations of the human CF [26] and p53 gene [27]. Another application of gene chips is the large scale analysis of DNA markers, so-called single nucleotide polymorphisms (SNPs). The fast detection of SNPs by array technology facilitates mapping of disease genes [28] to further characterize mammalian genomes.

As a result of many genome projects, both ongoing and completed, a tremendous amount of data are currently available in a number of public and commercial databases. Although sequencing projects will continue, the challenge of discovery is shifting away from the generation of new sequence data to the identification and understanding of gene products and their functions. To meet these and other challenges, efficient integration and interpretation of generated data has become one of the critical tasks of bioinformatics. Bioinformatics is the application of computer science to the interpretation and management of biological data. It will link expression profiling on the transcript level (*i.e.* genomics) with other systematic approaches. The most prominent methodology besides genomics is proteomics, which deals with large-scale analysis of proteins [21, 29]. An approach in this area includes protein microcharacterization for large-scale identification of proteins and their post-translational modifications. "Differential display" proteomics compares protein levels in cell types or diseases. Studies on protein-to-protein interactions add more functional data. Methods applied in proteomics include two-dimensional gel electrophoresis [30], protein chips and isotope-coded affinity tags [31], yeast two-hybrid assays [32], phage display [33] and a variety of mass spectroscopic methods [34].

Future directions

The methods described in the present article are powerful tools for identifying genes associated with diseases and cell type differentiation. Together with the completion of the sequencing of the entire human genome, the development of systematic approaches, such as genomics or proteomics, is set to revolutionize medicine. As described in this paper, novel diagnostic (expression profiling, identification of disease genes, analysis of mutations) and therapeutic (pharmacogenetics, identification of drug targets) concepts are emerging. Whether or not the results obtained are useful will largely depend on data management, data collection in public databases, and the availability of interconnected approaches. Gene array techniques and methods of proteomics are large-scale approaches that cannot be established by a single conventional laboratory. Networks between well-funded core facilities and users (who are experts in their applications and also well funded) that facilitate free scientific exchange and access to novel technologies have to be established. The results of the Human Genome Project will predict increased risk, provide early detection of diseases, and promote more efficient treatment strategies. To be successful, developments in the areas of genomics and proteomics will have to cover all aspects of human disease, including basic and applied medical science and ethical considerations [35]. Although there is much excitement about the progress in human genome sciences, thorough ethical considerations are necessary to ensure the beneficial use of the techniques and knowledge of the postgenomic area.

References

1. Lander ES, Linton LM, Birren B, *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860–921.
2. Venter JC, Adams MD, Myers EW, *et al.* The sequence of the human genome. *Science* 2001; 291: 1304–1351.
3. Adams MD, Kelley JM, Gocayne JD, *et al.* Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 1991; 252: 1651–1656.
4. Vukmirovic OG, Tilghman SM. Exploring genome space. *Nature* 2000; 405: 820–822.
5. Sargent TD. Isolation of differentially expressed genes. *Methods Enzymol* 1987; 152: 423–432.
6. Ann DK, Wu MM, Huang T, Carlson DM, Wu R. Retinol-regulated gene expression in human tracheobronchial epithelial cells. Enhanced expression of elongation factor EF-1 alpha. *J Biol Chem* 1988; 263: 3546–3549.
7. Sagerstrom CG, Sun BI, Sive HL. Subtractive cloning: past, present, and future. *Annu Rev Biochem* 1997; 66: 751–783.
8. Love J, Deininger P. Hybridization selections, using immobilized driver DNA, to enrich for clones unique to a library. *Biotechniques* 1991; 11: 88–92.
9. Wang J, Kuliszewski M, Yee W, *et al.* Cloning and expression of glucocorticoid-induced genes in fetal rat lung fibroblasts. Transforming growth factor-beta 3. *J Biol Chem* 1995; 270: 2722–2728.
10. Bals R, Welsch U, Bosch TCG. Isolation and molecular characterization of serous and mucous gland cells of the porcine airways. *Am J Respir Cell Mol Biol* 1998; 18: 548–553.
11. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 1992; 257: 967–971.
12. Reddy PM, An G, Di YP, Zhao YH, Wu R. Isolation and characterization of nucleolin gene as one of the vitamin A-responsive genes in airway epithelium by a palindromic primer-based mRNA differential display method. *Am J Respir Cell Mol Biol* 1996; 15: 398–403.
13. Jin F, Nathan CF, Radzioch D, Ding A. Lipopolysaccharide-related stimuli induce expression of the secretory leukocyte protease inhibitor, a macrophage-derived lipopolysaccharide inhibitor. *Infect Immun* 1998; 66: 2447–2452.
14. Bosch TC, Lohmann JU. Identification of differentially expressed genes by nonradioactive differential display of messenger RNA. *Methods Mol Biol* 1998; 86: 153–160.
15. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 1995; 270: 484–487.
16. Madden SL, Wang CJ, Landes G. Serial analysis of gene expression: from gene discovery to target identification. *Drug Discov Today* 2000; 5: 415–425.
17. Hibi K, Liu Q, Beaudry GA, *et al.* Serial analysis of gene expression in non-small cell lung cancer. *Cancer Res* 1998; 58: 5690–5694.
18. Lockhart DJ, Winzler EA. Genomics, gene expression and DNA arrays. *Nature* 2000; 405: 827–836.
19. Harrington CA, Rosenow C, Retief J. Monitoring gene expression using DNA microarrays. *Curr Opin Microbiol* 2000; 3: 285–291.
20. Sherlock G. Analysis of large-scale gene expression data. *Curr Opin Immunol* 2000; 12: 201–205.
21. Celis JE, Kruhoffer M, Gromova I, *et al.* Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett* 2000; 480: 2–16.
22. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; 95: 14863–14868.
23. Kaminski N, Allard JD, Pittet JF, *et al.* Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. *Proc Natl Acad Sci USA* 2000; 97: 1778–1783.
24. Roses AD. Pharmacogenetics and the practice of medicine. *Nature* 2000; 405: 857–865.
25. Srivastava M, Eidelman O, Pollard HB. Pharmacogenomics of the cystic fibrosis transmembrane conductance regulator (CFTR) and the cystic fibrosis drug CPX using genome microarray analysis. *Mol Med* 1999; 5: 753–767.
26. Cronin MT, Fucini RV, Kim SM, Masino RS, Wespi RM, Miyada CG. Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Hum Mutat* 1996; 7: 244–255.
27. Ahrendt SA, Halachmi S, Chow JT, *et al.* Rapid p53 sequence analysis in primary lung cancer using an oligonucleotide probe array. *Proc Natl Acad Sci USA* 1999; 96: 7382–7387.
28. Lindblad-Toh K, Winchester E, Daly MJ, *et al.* Large-scale discovery and genotyping of single-nucleotide

- polymorphisms in the mouse. *Nat Genet* 2000; 24: 381–386.
29. Pandey A, Mann M. Proteomics to study genes and genomes. *Nature* 2000; 405: 837–846.
30. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci USA* 2000; 97: 9390–9395.
31. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999; 17: 994–999.
32. Legrain P, Selig L. Genome-wide protein interaction maps using two-hybrid systems. *FEBS Lett* 2000; 480: 32–36.
33. Pasqualini R, Ruosiahti E. Organ targeting *in vivo* using phage display peptide libraries. *Nature* 1996; 380: 364–366.
34. Heine G, Raida M, Forssmann W-G. Mapping of peptides and protein fragments in human urine using liquid chromatography-mass spectrometry. *J Chromatography* 1997; 776: 117–124.
35. Collins FS, Mansoura MK. The Human Genome Project. *Cancer* 2001; 91: 221–225.