

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/51667731>

Transcriptome Analysis

Article in *Advances in Biochemical Engineering/biotechnology* · September 2011

DOI: 10.1007/10_2011_102 · Source: PubMed

CITATIONS

14

READS

5,560

8 authors, including:



Frank Stahl

Leibniz Universität Hannover

209 PUBLICATIONS 4,771 CITATIONS

[SEE PROFILE](#)



Bernd Hitzmann

University of Hohenheim

309 PUBLICATIONS 3,528 CITATIONS

[SEE PROFILE](#)



Daniel Landgrebe

9 PUBLICATIONS 164 CITATIONS

[SEE PROFILE](#)



Cornelia Kasper

University of Natural Resources and Life Sciences Vienna

274 PUBLICATIONS 6,120 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Online monitoring of yeast cultivation using electronic nose [View project](#)



From 3D to 3D: isolation of mesenchymal stem cells into a 3D environment [View project](#)

Transcriptome Analysis

**Frank Stahl, Bernd Hitzmann, Kai Mutz, Daniel Landgrebe,
Miriam Lübbecke, Cornelia Kasper, Johanna Walter
and Thomas Scheper**

Abstract Transcriptome analysis technologies are important systems-biology methods for the investigation and optimization of mammalian cell cultures concerning with regard to growth rates and productivity. For the production of recombinant proteins, knowledge of the expression conditions of the influencing genes is a major issue in the improvement of cell lines by means of genome engineering. This chapter presents two main techniques for transcriptome analysis: microarray technology and next-generation sequencing. Protein-based methods are also briefly outlined. Furthermore, the impact of these technologies on mammalian cell culture improvement is discussed.

Keywords Cell culture · Transcriptome · Systems biology · Microarray · Next generation sequencing

Contents

1	Introduction.....	2
2	Transcriptome Analysis Using Microarray Technology.....	3
2.1	Fabrication of DNA Microarrays.....	4
2.2	Design of Microarray Experiments.....	5
2.3	Principle of Microarray Technology.....	7
3	Data Analysis.....	8
3.1	Microarray DataAnalysis.....	8
3.2	Exploratory Analysis.....	11
4	Transcriptome Analysis Using Next-Generation Sequencing.....	12
4.1	SOLiD System.....	14
4.2	454 Sequencing System.....	15
4.3	Illumina/Solexa Sequencing Technology.....	15

F. Stahl (✉) · K. Mutz · D. Landgrebe · M. Lübbecke · C. Kasper ·
J. Walter · T. Scheper
Institute for Technical Chemistry, Leibniz University,
Hannover Callinstr. 5, 30167 Hannover, Germany
e-mail: stahl@iftc.uni-hannover.de

B. Hitzmann
Fg Prozessanalytik und Getreidetechnologie, Universität Hohenheim,
Garbenstr. 23, 70599 Stuttgart, Germany

4.4	RNA-Seq	16
4.5	Applications	16
5	Protein Microarray Technologies.....	17
6	Impact of Transcriptome Analysis on Strain Improvement	19
6.1	High-Producing Cells	20
6.2	Cell Cycle Studies	20
7	Concluding Remarks	21
	References	22

1 Introduction

Several sophisticated technologies for genome-wide expression monitoring are available and widely used since the development of microarray technology in the 1990s and the complete sequencing of the human genome. The technical and experimental possibilities for studying gene expression offer a snapshot of the entire genome with a resolution that would have been inconceivable some years ago. The transcriptome displays a complete collection of messenger RNAs present under defined conditions. RNA synthesis is a central process in the flow of genetic information in eukaryotic and prokaryotic cells, and isolated RNA has therefore become the target for many analytical and diagnostic techniques. The availability of simple and efficient systems for the production of synthetic RNA has likewise led to the development of techniques for studying the interactions of biomolecules with defined RNA sequences. There is also an increasing interest in the clinical use of nucleic acids. The in vitro generation of randomized RNA transcripts has led to breakthroughs in the generation of high-throughput screens to identify sequences that may have enzymatic or therapeutic applications.

Traditional gene expression analysis involves techniques such as:

- RT-PCR
- Northern blotting
- Nuclease protection assay.

More progressive techniques are:

- Differential display
- Subtractive hybridization
- Representational difference analysis (RDA)
- Expressed sequence tags
- cDNA fragment fingerprinting
- Serial analysis of gene expression (SAGE)

These methods enable the discovery of unknown differentially expressed genes. The bottleneck of the traditional methods is the limitation in the number of genes which can be analyzed in parallel. While conventional methods focus on the examination of single genes, a DNA chip experiment delivers a complete gene

expressions pattern of the cell. This high degree of parallelization of biochips displays a great advantage over classic molecular biological methods. During the last decade, more and more microarrays containing probes for each annotated gene in the genome have become commercially available for completely sequenced organisms. For metagenomes and for unsequenced genomes, transcriptome analysis by new-generation sequencing (e.g. 454 or Solexa) is state of the art. Therefore, this chapter describes microarray technology and next-generation sequencing in detail. The impact of transcriptome analysis on cell culture techniques and improvement of productivity will be discussed from the user's point of view.

2 Transcriptome Analysis Using Microarray Technology

DNA microarrays, developed to determine gene expression levels in living cells, have revolutionized the way scientists study gene expression [1–3]. Since they enable the analysis of the mRNA levels of a large number of genes in a single assay, DNA microarrays have become standard tools for gene expression profiling. For the understanding of biological systems with up to 30,000 genes, the measurement of the complete set of transcripts of an organism is necessary. Thus, an ideal tool for such measurements is the DNA microarray technology, a large-scale and high-throughput application utilizing amino-modified oligonucleotides or PCR products arrayed on silylated microscope slides with high-speed robotics. These microscope slides containing many immobilized DNA samples—so-called targets—are typically hybridized with fluorescently labelled cDNA probes. This results in a highly parallel, addressable, and miniaturized format, in contrast to traditional molecular-biology methods.

Applications for this technology include, amongst others, the monitoring of gene expression [4, 5], mutation detection [6, 7], clone mapping, drug development [8], tailored therapeutics, single nucleotide polymorphism (SNP) research, detection of genetically modified organisms (GMO) [9], and high-throughput screening in general. DNA microarrays can considerably simplify and accelerate a number of expensive diagnostic methods and have a profound impact on biological research [4], industrial production [4], medicine [10], diagnostics [11], environmental research [12, 13], bioprocess optimization [14, 15], and pharmacology [8], and will likely be used as the biosensors of the future. Microarray technology represents a powerful tool that allows researchers to link hypothesis testing and data. The data generated by microarray experiments can provide a large amount of information about important cellular pathways and processes.

The complete DNA sequences of various microorganisms which have been determined in the past few years can be applied to optimize strains as well as recombinant protein production. Strain optimization involves measurement of genome-wide mRNA levels in wild-type and mutant strains using DNA microarrays. Furthermore, microarray analysis can help to identify unknown genes required for recombinant protein production.

Bioprocess optimization using microarrays facilitates metabolic control analysis, modeling, and molecular biology methods to create new mutants and strains, e.g. with an optimized protein production rate [16]. Microarrays allow the investigation on a genomic scale enabling the qualitative and quantitative characterization of the cell metabolism. A better understanding of the impact of recombinant protein production can thereby be achieved using the generated “snapshot” of the actual cellular composition and activity. Additionally, the knowledge of the interaction of host cell metabolism with recombinant protein production is improved and contributes to process optimization.

By applying high-throughput screening technologies, it is possible to screen large numbers of strains that are produced by random mutagenesis and to determine the valuable ones containing beneficial mutations. By going through several rounds of random mutagenesis on the one hand and screening for the desired phenotype on the other, this process will thus identify those strains with considerably improved properties for production.

Microarrays play a pivotal role in modern biological sciences. They enable the utilization and analysis of a great amount of genetic information, for example derived from the human genome project. As a result, microarrays facilitate the understanding of gene regulation and gene function. The microarray technology enables the simultaneous analysis of complex genetic changes (the so-called “differential gene regulation”) by its high degree of parallelization. This can be achieved by parallel measurements of thousands of interactions between mRNA-derived molecules and genome-derived target molecules, thereby producing large amounts of raw data.

2.1 Fabrication of DNA Microarrays

Microarrays use modified glass slides as substrate enabling high spot density (spot diameter < 200 microns). Various methods for the automated production of DNA microarrays are used at present. The oligonucleotides can either be generated directly in situ on the microarray surface in a so-called on-chip synthesis or can be synthesized separately followed by an immobilization to the surface using a so-called DNA arrayer. There are three primary technologies: photolithography, ink jetting, and contact printing, and variants thereof. Each of these manufacturing technologies has specific advantages and disadvantages.

The photolithographic approach relies on the in situ synthesis of 20–30mer oligonucleotides using photomasks (Fig. 1.1) [17–19]. By utilization of photolabile protection groups, each probe is individually synthesized on the surface of the microarray at a high density. Photolithography was developed by Fodor et al. [17] and commercialized by Affymetrix.

Affymetrix uses several 25mer oligonucleotides per gene in a perfect match and a mismatch manner, whereas Agilent uses one 60mer oligonucleotide per gene. In contrast, the ink jetting and contact printing methods attach pre-synthesized

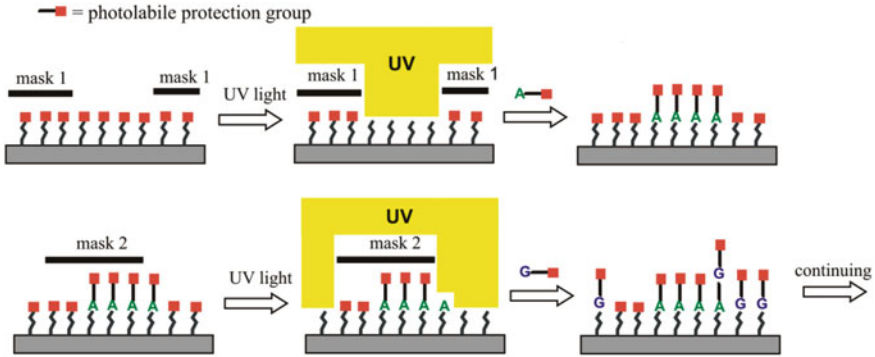


Fig. 1.1 Photolithographic process for the in-situ synthesis of oligonucleotides on microarrays

DNA probes to the chip surface. While the in situ probe synthesis requires sophisticated and expensive equipment, the contact and non-contact dispensing methods have made DNA microarrays affordable for academic research laboratories. In addition, the direct synthesis on the chip is less precise and the products cannot be sufficiently validated. In contrast, pre-synthesized oligonucleotides can be validated and thus produced at a high quality.

Since 1996, many DNA arrayers have become commercially available. Currently, the glass slide DNA microarrays represent the most popular format for gene expression profiling experiments.

2.2 Design of Microarray Experiments

The design of scientific experiments is an art of balancing several considerations including cost, equipment, and accuracy. For a given experiment, there would not be one ‘right’ design. Instead, different designs for the same scientific question may be chosen. Nevertheless, some commonsense principles are broadly accepted [20].

First of all, depending on the nature of the starting material, both, biological and technical replicates are to be selected. The biological variability of a given population needs to be calculated to enable conclusions from the investigation of a single measured effect on the entire target population of interest. Concretely, this means that, for example, investigating a nutritional supplement in a given animal model requires the measurement of several different animals (biological replicates), whereas the same effect in HepG2 cells can be tested on technical replicates.

Secondly, microarray experiments can be performed as single-component (colour) or two-component (colour) assays. Experimental standard designs of two colour assays are the so-called dye swap design (Fig. 1.2a), where the

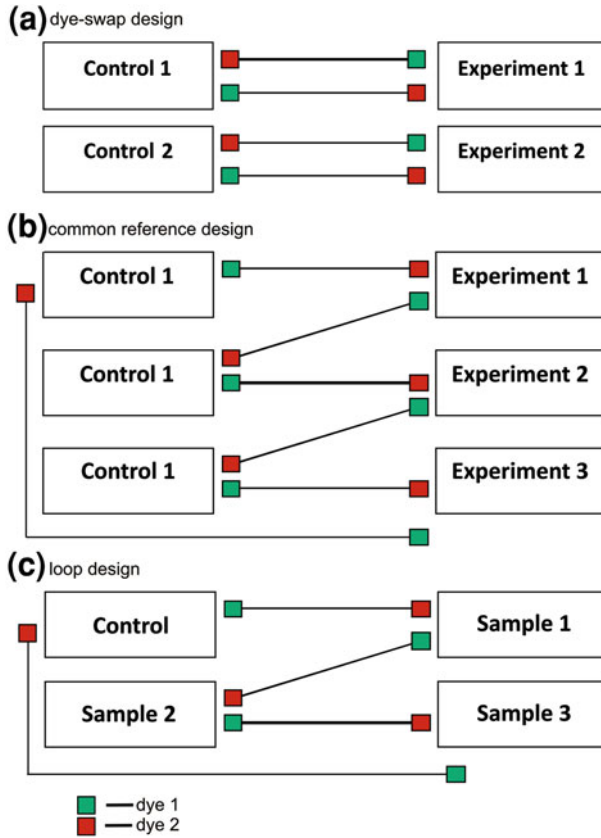


Fig. 1.2 Design of microarray experiments in two-component systems. Competitive hybridization of two labelled cDNA samples to the same microarray. The two mRNA targets are reverse transcribed into cDNA, labelled with different fluorescent dyes (usually *green* fluorescent dye, Cyanine 3 (Cy3, *dye 1*) and a *red* fluorescent dye, Cyanine 5 (Cy5, *dye 2*) mixed in equal proportions and hybridized to the spotted DNA probe molecules on the microarray surface. **a** Dye-swap design. **b** Common reference design. **c** Loop design

hybridization is repeated with a reverse labelling and the dye effect can therefore be minimized, and the so-called common reference design (Fig. 1.2b), where for every hybridization the same reference is always labelled with one dye and the samples of interest (e.g. different patients, cell lines or different time points) are labelled with the other dye. The most economic design, because a minimum number of chips is needed, is the so-called loop design (Fig. 1.2c). Here each sample is hybridized to each of two various different samples with different dye combinations [21–23].

The advantage of the two-component system is its independence of the absolute amount of the fixed DNA, as only the relative ratios of the Cy3 versus Cy5 signal intensities are analyzed for each spot separately. In contrast, one-component

systems (e.g. Affymetrix GeneChips) require larger numbers of hybridizations because of their higher array-to-array variance.

2.3 Principle of Microarray Technology

On a single microarray chip, a large set of genes is arrayed in a compact and regular manner. Due to the small size of the spots ($<200\text{ }\mu\text{m}$) several thousands of different oligonucleotides can be immobilized on one single slide. Each of these so-called probes binds to the complementary nucleic acid (“target”) isolated from the test and/or reference sample. The comparison of the binding efficiencies between two samples provides an easy and efficient survey of gene transcript level changes for numerous genes in a single experiment. Total or messenger RNA is used as starting material for target preparation. RNA isolation is one important step in the array experiment. Low quality RNA leads to poor hybridization results. In order to prepare the target for hybridization, first-strand cDNA is synthesized enzymatically from total RNA using oligo-d(T) primer or random primer. To exclude interference within the labelling reaction and during hybridization, DNase I digestion of the isolated RNA is strongly recommended after RNA extraction. During reverse transcription a fluorescently labelled nucleotide (Cy3-dC/UTP or Cy5-dC/UTP) is incorporated into the nascent first-strand cDNA. Subsequently, the template RNA is degraded by chemical treatment and the first-strand cDNA is separated from primers, unincorporated nucleotides, and RNA debris. Two sets of differently labelled cDNAs can be further combined and co-hybridized to the same array under stringent conditions. After hybridization, the unbound and non-specific bound cDNA is removed from the array by thorough washing. After subsequent scanning of the array with a confocal array scanner, the fluorescence intensity of each individual spot is determined and converted to grayscale values. Following normalization of individual grayscale values for each spot, the expression ratio of each gene on the array can be calculated semi-quantitatively. Data normalization is performed by using non-linear regression procedures.

A simple array experiment consists of five basic steps (Fig. 1.3):

1. The oligonucleotides are designed and spotted onto a substrate.
2. The sample RNA is isolated.
3. The cDNA is synthesized, a procedure that also involves fluorescent labelling for later detection.
4. The labelled cDNA target molecules are hybridized to the probe oligonucleotides on the substrate.
5. The hybridization results are imaged and analyzed.

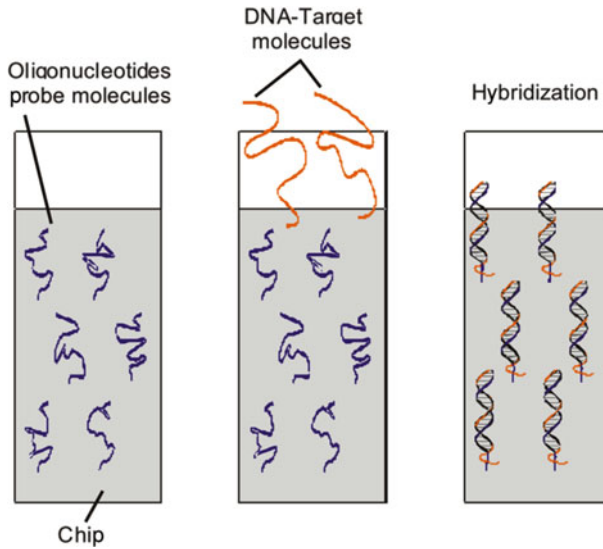


Fig. 1.3 Interaction of labelled cDNA target molecules with molecules immobilized on a glass array

3 Data Analysis

3.1 Microarray Data Analysis

Microarrays promise dynamic snapshots of cell activity, but microarray results are unfortunately not straightforward to interpret. The generation of complicated data sets and the difficulty of interpreting them requires a sound experimental design and particularly a coordinated and appropriate use of statistical methods [24, 25]. Tools for the efficient integration and interpretation of large datasets are needed. Despite the vast amount of literature available on microarray analysis, there is still a lack of standards for comparing and exchanging such data. Therefore, Minimum Information About a Microarray Experiment (MIAME) standards [26] have been established as a prerequisite for the worldwide comparability of gene expression data, and there are several URLs where these standards are available (e.g. <http://www.ncbi.nlm.nih.gov/geo/>).

For the planning and evaluation of microarray experiments, bioinformatics supply various procedures and algorithms [27]. One reason for the diversity is that the microarray experiments themselves are not performed in a consistent way but in different ways, depending on the objective of a project. In Fig. 1.4 the different contributions of bioinformatics are presented for the planning and evaluation of microarray experiments. They will be discussed below in detail. During all these steps, a fault detection and treatment is performed, which will not be discussed

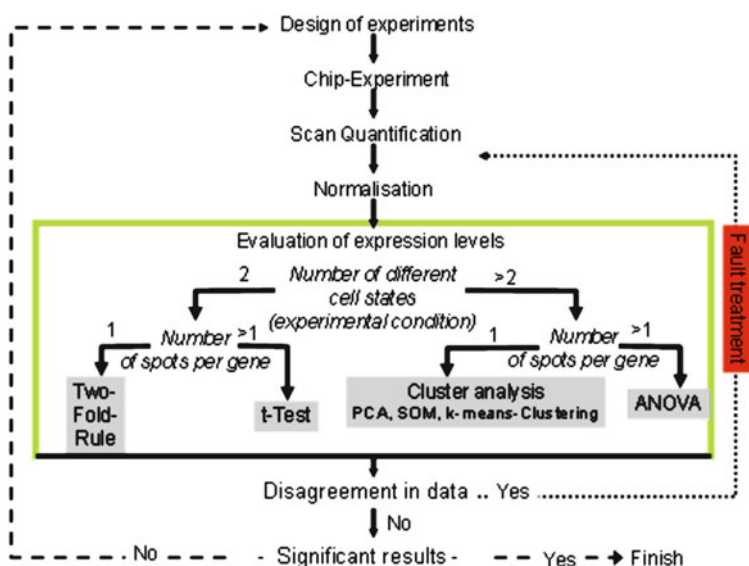


Fig. 1.4 Typical tasks and techniques of bioinformatics for the evaluation of microarray experiments

here. The first step is the design of the DNA microarray experiment. Here, for example, bioinformatics algorithms are used to select the sequence of oligonucleotides, to ensure that they exhibit high specificity for one single mRNA of the whole transcriptome. Software is used to specify the positioning of spots on the chips, in order to identify the corresponding gene in the evaluation procedure. Sometimes just the effect of an active pharmaceutical ingredient on a cell is investigated. In this case, a single chip experiment can be performed, where the transcriptomes of untreated and treated cells are compared to each other. However, to eliminate the influence of the dye, the dye-swap experiment is carried out, as discussed above. If replicates of spots are used, the positions of these replicates on the chip have to be considered carefully, to get the maximum information. All these issues are specified during the design of the microarray experiment.

After the microarray has been hybridized it must be scanned to acquire the raw data for further evaluation. The scanning is performed for the two dyes separately, measuring the fluorescence signal for each dye and obtaining two grayscale images for evaluation. The first step in analysis is the detection of the spots. Here, different segmentation procedures such as fixed circle, adaptive circle and adaptive shape segmentation can be applied. In the first and second procedures, a circle with a fixed or a variable radius is located optimally around a spot and the pixel intensity in the inner circle area is used to quantify it. The adaptive segmentation procedure does not postulate a circle shape but an arbitrary shape to locate the spot area for quantification [28]. To take into account changes of in the background over the

chip, this can be usually determined for each spot separately by using a kind of *o*-ring around a spot whose inner radius is clearly larger than the radius of the spot area. The outer radius is chosen, so that the ring will not interfere with other spots. The intensity in the ring will represent the background of the corresponding spot. For the quantification of the background and spot intensity different quantities can again be used, such as median, modal or mean values of the intensity. The difference between spot intensity and background intensity is a measure of the expression degree of the corresponding gene. In this way systematic variations in a DNA microarray experiment such as slide heterogeneity, spotting variation, changing background signals etc. can be compensated for. Image analysis software such as imaGeneTM and GenePixTM software can be used for this evaluation. However, due to the fact, that there are many variabilities from the harvesting of the mRNA to this quantified value, the degree of expression degree obtained cannot be regarded as an absolute measure. Therefore, at least two states of cells are investigated simultaneously (e.g. treated and untreated) whose mRNAs are labelled differently. The values obtained are evaluated relative to each other. Because the mRNAs obtained are treated almost in the same way (except for the dye used for labelling), the ratio of the values gives a relative measure to each other, i.e. the expression change.

However, the labelling as well as the fluorescence intensities of the two dyes are not exactly the same. A normalization procedure is therefore required, which is one of the most important steps during the evaluation. Applying normalization procedures, results from different experiments are made comparable and technical imperfections are compensated for. If housekeeping genes are available, they can be used for normalization. Housekeeping genes are genes whose products are necessary for fundamental cell maintenance and which are transcribed at an almost constant level. Therefore, it can be assumed that they will not change their expression grade under the situations investigated. For these genes the expression as well as the quantified expression grades must be the same. Thus, all expression values of one evaluated image can be transformed by using a multiplication factor in such a way that after normalization the expression values of the housekeeping genes of the two images (representing the transcriptomes of treated and untreated cells) are the same. Then the ratio of the corresponding grade should give the correct change in expression.

If a whole genome chip is under investigation, another normalization procedure can be carried out. Under the assumption that the overall expression of the mRNA is the same for the treated and untreated cells, the sum of the expression grade of all genes must be the same for both cases. Therefore, each expression grade is divided by the sum of the expression grades of all corresponding genes (as a consequence, the sum of the transformed grades will then be 1). After the transformation, the individual ratios are calculated as mentioned before and will give the change in expression.

If the expression change is calculated in this way, it has the disadvantage that a twofold upregulated (ratio = 2) and a twofold downregulated (ratio = 0.5) gene will be characterized by different numerical data. If the logarithm with respect to the

base 2 is calculated, then for a twofold upregulated gene the ratio is 1 and for a twofold downregulated gene the ratio is -1 . The absolute values are the same for both cases. A symmetric distribution for up- and downregulated genes is therefore obtained. The $\log_2(\text{ratio})$ obtained with the normalized expression grades are used as expression levels.

A further normalization procedure based on logarithms is frequently used. Here a special xy plot is considered, called an MA plot. In this plot the ordinate values represent the logarithm of the ratio of the corresponding expression grades and the abscissa values represent the logarithm of the multiplied corresponding expression grades. Then a linear regression or alternatively a locally linear regression (LOWESS regression) is carried out with the data. The theoretical values of the regression curve are subtracted from the ordinate values, so that afterwards the MA plot is more symmetric to the abscissa. Further normalization procedures are described in the literature [29].

After all normalization procedures are performed, the $\log_2(\text{ratio})$ is evaluated to give expression levels. Here, further analysis depends on whether replicates of spots have been considered or not. If the expression of a gene is represented by a single spot, then the twofold rule is applied, which should be considered just as an auxiliary release. Using the twofold rule an expression level greater than 1 is considered as upregulated, and an expression level less than -1 is considered as downregulated. However, this is difficult to interpret if the expression grades of both states (from treated and untreated cells) are small. Replicates should be therefore be performed. This offers the possibility of applying the t -test for the decision of a differently expressed gene.

If the response of cells to different conditions (for example whether they respond to different concentrations of an active pharmaceutical ingredient) is under consideration, then, as described above, replicates should be performed and multivariate evaluation techniques such as cluster analysis, principal component analysis (PCA) or self-organized maps (SOM) should be applied to elucidate the information gained from the chips. If replicates are available, then analysis of variance (ANOVA) is the best choice.

All these multivariate data evaluation techniques do not require explicit knowledge of transcriptome data analysis and will not be discussed here in detail, but can be found in general statistical textbooks.

3.2 Exploratory Analysis

The next step in the analytical pipeline is usually gene expression clustering: a preliminary examination of data to confirm that groups are homogeneous. Many studies aim to find unknown co-regulated genes. Such studies using multi-dimensional scaling and clustering can be summarized as exploratory analysis.

Various clustering techniques can be applied for the identification of patterns in gene expression data [30]. Several software tools for cluster analysis have been

developed, such as GeneSightTM from Biodiscovery or Eisen-cluster. Most cluster analysis techniques are hierarchical, where the classification results in nested classes resembling a phylogenetic tree. Non-hierarchical clustering techniques involve partitioning of objects into different groups, such as *k*-means clustering.

The concluding section of analysis deals with methods and problems in determining differentially expressed genes between groups of samples, covering for example *t*-tests or different multiple-testing corrections as well as analysis of variance. The purpose of finding differently expressed genes can be achieved by statistical tests rather than cluster analysis. Finally, the results as well as an additional indication of the statistical reliability are given in order to allow further, more precise studies of gene expression (e.g. qRT-PCR) or the publication of these microarray results as evidence for changes in gene abundance. In general, it is difficult to analyze data from low-density microarray experiments with commercially available programs. Low-density microarrays consist of a choice of a few relevant genes, thus offering the possibility of a very individual chip design in order to investigate specific experimental questions. One disadvantage of low-density microarrays is that the data analysis cannot be performed by commercial software.

Following the data analysis, *k*-means clustering could be conducted depending on the experimental design. Furthermore, the resulting files include not only information about the microarray analysis but also about the pathways of known genes on the microarray coming from the KEGG (Kyoto Encyclopedia of Genes and Genomes) database.

4 Transcriptome Analysis Using Next-Generation Sequencing

Microarray technology has begun to reach its limitations [31]. It shows a relatively small dynamic range due to background, and a limited spot density. In addition, mismatches and cross-hybridization significantly affect the results [32]. Furthermore, the comparison between different experiments usually needs complex bioinformatical normalization algorithms [33]. In contrast to microarrays, which provide only relative mRNA levels, sequencing methods have major advantages with regard to true quantification and they offer statistically complex data analysis [34, 35].

In the past, genetics was only an observational science. But the invention of the DNA sequencing technique was a revolutionary point for this biological science and represents the starting point for modern genomics which can enable mechanistic understandings. DNA sequencing means the assignment of the sequence of nucleotides in a DNA molecule. To date, the genomes of over 1,000 different organisms have been analyzed¹ in total and enhancements in DNA sequencing methods are increasingly used for transcriptome analysis to replace DNA

¹ <http://www.ncbi.nlm.nih.gov/guide/genomes/>.

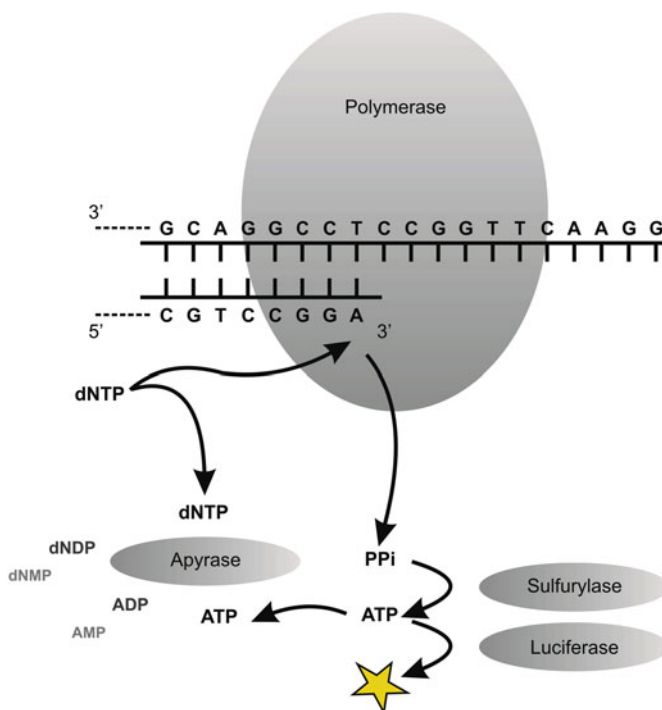


Fig. 1.5 Pyrosequencing is a genetic analysis based on sequencing by synthesis. It delivers explicit sequence data within a few minutes

microarray technology. Today, various methods for the achievement of sequence information are well established. Most of them are based on the dideoxy method drafted by Friedrich Sanger in the 1970s [36], which uses an enzymatic reaction. Starting from a short known primer, DNA polymerase elongates the complementary strand. The use of four differently labelled dideoxy-nucleotides leads to the identification of the unknown DNA sequence.

Recent developments in sequencing technology have led to a new key tool for transcriptomics: the pyrosequencing technology [37], which offers opportunities for accelerated sequencing by highly parallel approaches (Fig. 1.5). In analogy to Sanger et al., pyrosequencing-based next-generation sequencing methods use a DNA polymerase for synthesis of the complementary DNA strand. The incorporation of nucleotides during DNA sequencing is monitored by bioluminescence in real time. A luciferase-based multi-enzyme system generates visible light after nucleotide binding, which can be detected. The four different nucleotides are added successively, so that only the binding of the compatible nucleoside triphosphate generates a light signal. This shows whether a known nucleotide is incorporated or not.

The pyrosequencing technology was invented in 1986 based on the idea of following nucleotide incorporation by using released pyrophosphate (PPi) to generate a signal.

The release of an equimolar amount of PPi is a natural process during the binding of a nucleotide to the 3' end of the primer, which is used as the starting point of sequencing. The multi-enzyme system mentioned above consists of DNA polymerase, ATP sulfurylase, luciferase and apyrase. After incorporation and release of PPi, ATP sulfurylase catalyses the conversion of the PPi to ATP in a quantitative reaction.

The luciferase uses the ATP for catalyzing the conversion of luciferin to oxyluciferin. Within this reaction visible light is emitted, which can be detected by a CCD camera [37]. A special computer program displays the recorded data as peaks in a diagram. Free ATP and nucleotides are degraded by the apyrase. This disables light emission and regenerates the reaction solution. The complete enzymatic process can be performed in a single well, offering a fast reaction time of approximately 20 min per 96-well plate [37]. It is possible to determine the DNA sequence by computer-assisted analysis of the detected light signals, because of the consecutive addition of the four different dNTPs comparable to the well-known Sanger protocol [38]. Besides Sanger sequencing, pyrosequencing is the only method which is currently commercially available [39]. Pyrosequencing is applied in the analysis of single-nucleotide polymorphisms (SNPs), as well as tag sequencing and whole genome sequencing [37]. Beyond microarrays, next-generation sequencing is pushing transcriptomics further into the digital age [40].

Although those next-generation technologies generate many short DNA fragments with reduced time and costs [41], sequencing of a whole vertebrate genome is still an extensive task. Transcriptome sequencing of cDNA has the advantage that the templates are of relative small size. This enables high-throughput applications of gene expression profiling, genome annotation or discovery of non-coding RNA. Miscellaneous data offer a parallel analysis of gene expression, genomic loci structures and, e.g. SNPs [42]. At present, three popular next-generation platforms support gene expression profiling: SOLiD, Solexa and 454 Sequencing System [40]. In the following section these platforms are presented in detail.

4.1 SOLiD System

SOLiD is the acronym for “Sequencing by Oligonucleotide Ligation and Detection” and is available from Applied Biosystems, Inc. (Foster City, CA, USA). The underlying principle is sequencing by ligation. In this method, DNA fragments of the sample to be sequenced, modified with internal and external adapters, are coupled to magnetic microparticles. The adapters are cleaved and DNA rings are formed by ligation of the adapter ends. The rings are then split again at defined positions at the left and right domain of the adapter. The first few bases of the adherent end domains are sequenced afterwards and new corresponding adapters are ligated to the ends. The new adapters allow the subsequent attachment of the fragments to the microparticles and amplification by emulsion PCR. After accumulation of the particles, octamer degenerated oligonucleotides are hybridized to

the particles. These oligonucleotides are each labelled with a different fluorescent dye after the 5th base. Detection then takes place, followed by elimination of the last three bases after the analyzed nucleotide. By repeating this procedure in further cycles, the 10th and 15th bases will be identified. Other steps with shorter primers lead to the detection of the positions 4, 9, 14 and so on.²

4.2 454 Sequencing System

The 454 sequencing technology is the primary next-generation method in transcriptomics. The 454 Sequencing System by Roche Diagnostics Corporation (Branford, CT, USA) is an ultra-high-throughput system. It is the first next-generation sequencing technology released to the market [42] and can be described as pyrosequencing in high-density picoliter reactors. DNA fragments received by shearing are attached to streptavidin beads captured into separate droplets in an emulsion PCR. The droplets form small amplification reactors [43]. Then, any bead is transferred into a picoliter plate and analyzed by pyrosequencing. The instrument can sequence up to 120 million bases in a time of about 10 h. Limited by the pyrosequencing chemistry used, the single reading frames (250 nt) are considerably shorter than with Sanger technology (600 nt), but up to 400,000 reactions can be performed in parallel. With more than 100 research publications [42], it is the most widely published next-generation platform.³

4.3 Illumina/Solexa Sequencing Technology

This technique is based on a two-step mechanism, where amplification takes place first. Shear-stressed DNA fragments are tagged with different so-called “dense lawn” primers as adapters at both ends of their chain. Together with both complementary primers, all molecules are immobilized randomly to the surface of flow cell channels. DNA fragments hybridize with the complementary primers in a bridging way to start solid-phase bridge amplification immediately and the fragments become double-stranded. With further steps consisting of denaturation, renaturation and synthesis, a high density of equal DNA fragments is generated in an extremely small area. Several million of these dense clusters of double-stranded DNA are synthesized.

The second step is the real sequencing reaction. All four dNTPs labelled with different dyes and primers are added and successively incorporated by a DNA polymerase. After washing steps, a high-definition image is generated by laser

² www.appliedbiosystems.com.

³ www.454.com.

excitation from each cluster. The identity of the first base is recordable. The elimination of the 3' blocked terminus and the dyes follows. Within each new cycle, the DNA chain is elongated and more images are recorded for analysis. Here the reading frame is tenfold smaller (30 nt) than with common pyrosequencing. The whole system has is closely related to the method from Helicos BioScience.

Applications of this method are sold by both Illumina, Inc. (San Diego, CA, USA) as well as Solexa, Inc. (Hayward, CA, USA). Today the Genome Analyzer IIX is available on the market. It can be used for common DNA sequencing as well as for transcriptome analysis such as RNA-Seq, tag profiling or microRNA discovery.⁴

4.4 RNA-Seq

The transcriptomics alternative to pyrosequencing technology is called short-read high-throughput sequencing or RNA-Seq [41]. In recent years, RNA-Seq has rapidly emerged as the major quantitative transcriptome profiling system [13, 44]. RNA-Seq has been used, for example, for the global profiling of expression levels in human embryonic kidney and B cells [45], or for identification of differently expressed genes in mouse embryonic stem cells [46] and also for quantification of the whole mouse transcriptome [47]. Moreover, structural information or alternative splicing forms can be detected with this method [41].

Unlike microarrays, RNA-Seq can evaluate absolute transcript levels and detect novel transcripts and isoforms. As a consequence, it can be used to determine expression levels more precisely than microarrays [48]. Microarrays, on the other hand, have the power to measure the expression of thousands of genes in parallel, but they are not able to display the coding sequences of the transcripts. The results are calculated from indirect hybridization data, which gives rise to reproducibility and comparability problems [42]. One great advantage of RNA-Seq compared to microarrays is the possibility of capturing transcriptome dynamics across different cell culture conditions without normalization of data sets. Therefore, RNA-Seq is the method of choice in projects for transcript discovery, especially the analysis of metagenomes.

4.5 Applications

The major application area for next-generation sequencing technologies is biomedical research associated with key goals like “the USD1000 genome” [49]. Next-generation sequencing is used for the detection of sequence variations within

⁴ www.illumina.com.

an individual genome such as SNPs, deletions, insertions, or structural changes [50]. Here, RNA-Seq is typically used for the analysis of non-coding RNAs as crucial regulators [51].

Secondary next-generation sequencing was adopted for high-throughput research performed mostly by microarrays. To date, it has been possible to generate transcriptome sequencing libraries for important cell cultures for recombinant protein production, e.g. *Arabidopsis thaliana*, *Caenorhabditis elegans* and human cell line transcriptomes were successfully interrogated with 454 technology [52]. For HeLa cells Illumina technology was used [53]. The understanding of whole transcriptional networks was also enabled by RNA-Seq. Examination of small non-coding miRNAs allows a global view of the transcriptome [48].

A major focus of systems-biotechnology work is the quantitative understanding of molecular principles behind protein synthesis, modification and secretion derived from basic production strains as well as mutants and rationally engineered strains. Next-generation sequencing provides the tools for rationalize inverse metabolic engineering approaches so that they can be implemented in future into rational system-wide modeling and optimization strategies [54]. The functional complexity of a transcriptome cannot be fully elucidated with expressed sequence tags and microarrays. RNA-Seq can reveal more precisely the boundaries of untranslated regions at single nucleotide resolution and is useful for analyzing complex transcriptome and sequence variations, e.g. alternative splicing or gene fusion [55].

Additionally, metagenomics is an area of biological sciences concerned with acquisition of the whole genomic information of a biotope. Several microorganisms cannot be cultivated in a laboratory. To identify them without a cultivation process, metagenomic approaches can be performed. This can enhance knowledge of biodiversity or could lead to new biotechnological and pharmaceutical products [56]. Thus, next-generation sequencing can also raise metagenomics to a new level [56]. In this context, future analysis of mRNA levels under different conditions or in different cell types can be assessed by analysis of hybridization intensities and by application of methods using sequenced cDNA fragments, and both techniques will help to improve production output of cell cultures by optimizing cell growth and genetic activity.

5 Protein Microarray Technologies

Since the determination of the complete DNA sequences of a number of organisms, from bacteria to man, and the invention of new techniques like microarrays for monitoring biomolecular interactions, important milestones have been achieved in genomic and proteomic research. The results of such high-throughput screening approaches can change our fundamental understanding of life's cellular processes on a molecular level. However, gene expression analysis is not sufficient to predict the function of a protein. Monitoring protein interactions is an extremely

complex issue because the proteome is the quantitative representation of the complete protein expression pattern of a cell, tissue, organ or organism under exactly defined conditions.

Ideally, the analysis of the proteome delivers the complete available set of all proteins currently present in an organism. This data cannot be obtained on the transcriptome level, since no straightforward correlation exists between the amount of mRNA and the actual amount of protein. Parameters like mRNA stability, protein degradation, posttranslational modifications and others prevent a statement of the actual amount of protein based on transcriptome analysis. However, this information is of utmost importance, making a high-throughput analysis of the proteome necessary. One attractive method is the use of protein microarrays, which consist of a solid support, e.g. glass or synthetic material, with a modified or coated surface. Using special printers (preferably non-contact printing heads), capture probes—which may be proteins, peptides, receptors, enzymes or antibodies—are transferred to this surface in the form of micro spots ($<200\text{ }\mu\text{m}$) in a regular manner. Every micro spot contains only one kind of capture probe (in most cases, antibodies). These immobilized capture probes are able to bind their corresponding target molecule from a complex solution.

Different formats of protein microarrays are available. In the forward phase format, antibodies immobilized on the microarray surface are used as capture probes for their target. Since it is possible to immobilize many different antibodies on one single microarray, the forward phase format enables the parallel detection of many different targets within a complex sample. One disadvantage of the forward phase format is the necessity to label the target proteins. This labelling procedure may alter the composition of the sample, different proteins are labelled with varying labelling efficiencies, and the labels introduced may mask the proteins' epitopes essential for binding to the immobilized antibody.

In contrast, the reversed phase microarray offers the possibility of detecting unlabelled proteins. In this format, the protein of interest is directly immobilized onto the microarray surface and probed with fluorescently labelled detection antibodies. While this method allows the detection of the protein of interest in hundreds of different samples in parallel, its major limitation is the binding capacity of the microarray surface, resulting in a low dynamic range of the assay.

In the so-called sandwich format, capture antibodies are immobilized on the microarray surface and the binding of the corresponding protein is detected via labelled detection antibodies (Fig. 1.6). Therefore, for each target protein, two antibodies binding to different epitopes of the target are required. Sandwich-based microarrays avoid the difficulties associated with labelling reactions and exhibit high sensitivity [57]. Moreover sandwich assays are known to be highly specific, since the target must be recognized by two different antibodies.

The direct extrapolation of DNA microarray techniques to protein microarrays is limited due to the sensitive nature of the printed antibodies, which have to keep their native conformation in order to maintain activity. The fabrication of protein arrays is therefore particularly challenging and protein arrays lag behind in development because of the instability of the immobilized protein [58]. One

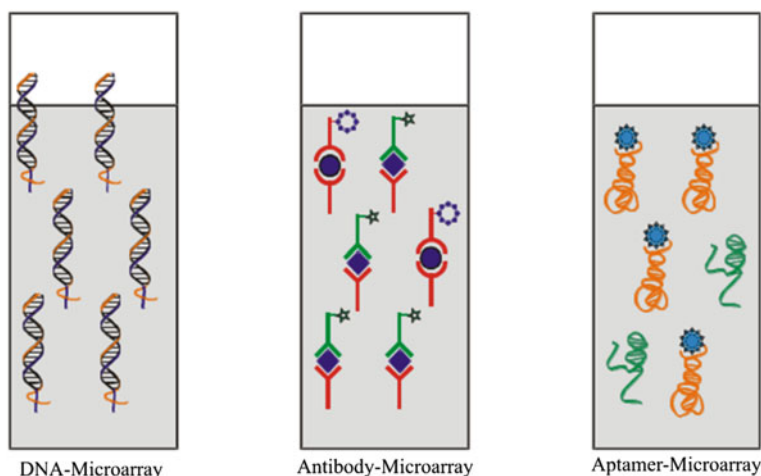


Fig. 1.6 Comparison of DNA, antibody and aptamer microarrays. An aptamer microarray consists of spotted DNA probes as capture molecules for protein detection

approach to overcoming this restriction is the utilization of a three-dimensional matrix for immobilization of the capture antibodies [59, 60]. In this structural environment, proteins are more likely to maintain their active configuration than on planar glass supports [58, 61]. Nitrocellulose membranes have shown their suitability for protein immobilization in Western blotting and the long-term stability of immobilized proteins on this support is known from immuno-diagnostic tests [62, 63]. Nitrocellulose membranes are therefore becoming the microarray substrate of choice in protein microarray applications [64, 65].

Another approach to overcoming the limitations caused by the low stability of immobilized antibodies is the utilization of more stable capture probes. In this context, aptamers have been investigated as an alternative to antibodies [66]. Aptamers are short single-stranded synthetic DNA or RNA oligonucleotides that can bind to a wide range of target molecules, including proteins.

As nucleic acids, aptamers can undergo denaturation, but the process is reversible. As a result of this stability and the possibility of automated selection of aptamers via systematic evolution of ligands by exponential amplification (SELEX), these oligonucleotides are highly promising capture molecules for protein microarrays.

6 Impact of Transcriptome Analysis on Strain Improvement

Mammalian cells in culture can be differentiated into two groups: primary and secondary cells. The latter ones are also known as immortal cells or cell lines. Primary cells are isolated directly from blood or tissue samples. These cells have a

restricted life span due to the fact that they undergo only a limited number of cell divisions. Primary cells better represent the tissue from which they are taken and are normally heterogeneous. These cells could be used for R&D applications, particularly for in vitro tests of new drugs and toxicity tests. Continuous cell lines rarely occur spontaneously from primary tissue cells; mostly they are developed by transformation with carcinogenic substances or viral genes. Besides their infinite growth, cell lines have further advantages including faster growth and the ability to be cultured in suspension. This makes them suitable for the production of recombinant proteins in large-scale cultivation [67]. Primary cells as well as cell lines are available as test systems for gene expression analysis, but there are a few drawbacks to both cell types. Primary cells display inter-individual differences, e.g. caused by age and gender of the donor, and the widely used cell lines are limited in their metabolic function because some pathways are different from those in normal tissues [68].

One important aim of strain improvement is to understand and characterize the functional heterogeneity in a given population at the cellular and molecular level. Another aim is to identify and isolate high-producing cells from the entire population and use them in production processes. Combining flow cytometric analysis and sorting of live cells with transcriptome analysis aids in relating molecular regulation processes within cellular subpopulations to the dynamics of the whole cell population [69]. Transcriptome analysis can thus be used to improve cell growth and to increase the productivity of mammalian cell cultures, e.g. by gene optimization.

6.1 High-Producing Cells

Numerous methods exist for developing high-producing populations via fluorescence activated cell sorting (FACS) and gene expression profiling of sorted subpopulations [70]. Co-expression of green fluorescent protein (GFP) is common. Cells which show high GFP fluorescence can be separated to obtain desired populations due to the fact that GFP expression is correlated with high productivity [71]. Simple surface staining is also accomplished [72]. In addition to this method, cultivation at lower temperature and treatment with chemicals such as butyrate can be used to increase the productivity of a cell line. Up to several hundreds of genes are upregulated in high producers; these genes may be involved in the secretory pathway including the Golgi apparatus and cytoskeleton, and may also include genes responsible for product formation [73].

6.2 Cell Cycle Studies

Cytometry is also applicable for cell cycle studies, since it is possible to stain DNA whose content can be correlated with the cell cycle. DNA replication occurs

exclusively during the S phase, such that G2-phase cells have twice the cellular DNA content of G1 cells [74]. For further analysis the cells can be synchronized into similar phases. This is achieved by various methods based on biological or physical effects. Due to the fact that the cell size changes during the mitosis, a separation by cell size is one of the practical physical methods. This separation can be performed by FACS or the centrifugal elutriation technique [75, 76]. Afterwards the cells grow in a synchronized cell cycle. This enables transcriptome analysis to search for genes involved in regulating the cell cycle or other cell-cycle-dependent gene activities like productivity rates [77]. The productivity of recombinant products depends on the cell cycle phase and the product. For recombinant proteins produced in CHO cells, it was ascertained that the productivity maximum occurs in the G₁ phase [78].

7 Concluding Remarks

The development of microarrays with DNA probes for gene expression analysis or antibody probes for proteomic applications based on hybridization processes (DNA probes) as well as on the immunological binding process (antibody probes) opens new horizons for biomolecular research. It can result in the production of new proteins, changes in membrane formation and various other alterations concerning cellular assembly [32].

Sequenced genomes are the basis for constructing DNA microarrays representing the common genes in a genome. Furthermore, they enable the synthesis of labelled cDNA from mRNA templates allowing high-throughput detection of transcript levels [33]. Various high-density oligonucleotide microarrays are now available commercially.

DNA microarrays have therefore already changed the way scientists study gene expression, but the real challenge starts with determining the function of all the genes discovered within the organisms. DNA microarrays are applied in industrial analytics and biomedical diagnostics, as well as in criminology. They can considerably simplify and accelerate a number of expensive diagnostic methods. Although conventional biosensors work well, their function can be validated and perhaps improved through a functional genomics study in which the induction of several thousand genes is detected simultaneously. Specifically, the use of microarrays can accomplish the following goals:

- Compare the time course of a sensor signal with the actual genomic response.
- Identify genes that respond earlier or more specifically to toxins.
- Identify gene induction patterns that can identify one toxin versus another (which can in turn be incorporated into multichannel sensors).

In the last few years, functional transcriptomics has been advanced by both microarray technology and genome sequencing. Certainly microarray technology has achieved its technical limits and is more and more complemented by

high-throughput next-generation sequencing technologies. Unlike microarrays, transcriptome sequencing (RNA-Seq) can evaluate absolute transcript levels, and detect novel transcripts and isoforms. Microarrays have the power to measure the expression of thousands of genes in parallel, but they are not able to reveal the coding sequences of the transcripts. The derived results are calculated from indirect hybridization data, which poses reproducibility and comparability problems. In fact, studies using both microarrays and RNA-Seq show a good correlation between the different data so that it is possible to compare results from one technology with the other [79], and both techniques help to improve production output of cell cultures by optimizing cell growth and genetic activity.

References

1. Khan J et al (1999) DNA microarray technology: the anticipated impact on the study of human disease. *Biochimica Biophysica Acta Rev Cancer* 1423(2):M17–M28
2. Duggan DJ et al (1999) Expression profiling using cDNA microarrays. *Nat Genet* 21:10–14
3. Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet* 21:33–37
4. DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278(5338):680–686
5. Harrington CA, Rosenow C, Retief J (2000) Monitoring gene expression using DNA microarrays. *Curr Opin Microbiol* 3(3):285–291
6. Park JH et al (2004) Oligonucleotide microarray-based mutation detection of the K-ras gene in colorectal cancers with use of competitive DNA hybridization. *Clin Chem* 50(9):1688–1691
7. Hegde MR et al (2008) Microarray-based mutation detection in the dystrophin gene. *Hum Mutat* 29(9):1091–1099
8. Walter G et al (2000) Protein arrays for gene expression and molecular interaction screening. *Curr Opin Microbiol* 3(3):98–302
9. Wang DG et al (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280(5366):1077–1082
10. Strauss KA et al (2008) Clinical application of DNA microarrays: Molecular diagnosis and HLA matching of an Amish child with severe combined immune deficiency. *Clin Immunol* 128(1):31–38
11. Gunn SR, Robetorye RS, Mohammed MS (2007) Comparative genomic hybridization arrays in clinical pathology—progress and challenges. *Mol Diagn Ther* 11(2):73–77
12. Sebat JL, Colwell FS, Crawford RL (2003) Metagenomic profiling: microarray analysis of an environmental genomic library. *Appl Environ Microbiol* 69(8):4927–4934
13. Wang RL et al (2008) DNA microarray application in ecotoxicology: experimental design, microarray scanning, and factors affecting transcriptional profiles in a small fish species. *Environ Toxicol Chem* 27(3):652–663
14. Richmond CS et al (1999) Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res* 27(19):3821–3835
15. Oh MK, Liao JC (2000) DNA microarray detection of metabolic responses to protein overproduction in *Escherichia coli*. *Metab Eng* 2(3):201–209
16. Wang M et al (2009) Microarray-based gene expression analysis as a process characterization tool to establish comparability of complex biological products: scale-up of a whole-cell immunotherapy product. *Biotechnol Bioeng* 104(4):796–808

17. Fodor SP et al (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science* 251(4995):767–773
18. McGall GH et al (1997) The efficiency of light-directed synthesis of DNA arrays on glass substrates. *J Am Chem Soc* 119(22):5081–5090
19. Gao X, Gulari E, Zhou X (2004) In situ synthesis of oligonucleotide microarrays. *Biopolymers* 73(5):579–596
20. Yang YH, Speed T (2002) Design issues for cDNA microarray experiments. *Nat Rev Genet* 3(8):579–588
21. Kerr MK, Churchill GA (2001) Statistical design and the analysis of gene expression microarray data. *Genet Res* 77(2):123–128
22. Dombkowski AA et al (2004) Gene-specific dye bias in microarray reference designs. *FEBS Lett* 560(1–3):120–124
23. Landgrebe J, Bretz F, Brunner E (2004) Efficient two-sample designs for microarray experiments with biological replications. *In Silico Biol* 4(4):61–70
24. Kim SY, Lee JW, Sohn IS (2006) Comparison of various statistical methods for identifying differential gene expression in replicated microarray data. *Stat Methods Med Res* 15(1):3–20
25. Klebanov L et al (2007) Statistical methods and microarray data. *Nat Biotechnol* 25(1):25–26 author reply 26–27
26. Brazma A et al (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 29(4):365–371
27. Stekel D (2003) *Microarray bioinformatics*. Cambridge University Press, Cambridge
28. Adams R, Bischof L (1994) Seeded Region Growing. *IEEE Trans Pattern Anal Machine Intell* 16(6):641–647
29. Quackenbush J (2002) Microarray data normalization and transformation. *Nat Genet* 32(Suppl):496–501
30. Eisen MB et al (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95(25):14863–14868
31. Bloom JS et al (2009) Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. *BMC Genomics* 10:221
32. van Vliet AH (2009) Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiol Lett* 302(1):1–7
33. Hinton JCD et al (2004) Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. *Curr Opin Microbiol* 7(3):277–282
34. Jiang H, Wong WH (2009) Statistical inferences for isoform expression in RNA-seq. *Bioinformatics* 25(8):1026–1032
35. Oshlack A, Wakefield MJ (2009) Transcript length bias in RNA-seq data confounds systems biology. *Biol Direct* 4:14
36. Sanger F et al (1977) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265(5596):687–695
37. Marsh S (2007) Pyrosequencing applications. *Methods Mol Biol* 373:15–24
38. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74(12):5463–5467
39. Ronaghi M, Uhlen M, Nyren P (1998) A sequencing method based on real-time pyrophosphate. *Science* 281(5375):5363
40. Blow N (2009) Transcriptomics: the digital generation. *Nature* 458(7235):239–242
41. Denoeud F et al (2008) Annotating genomes with massive-scale RNA sequencing. *Genome Biol* 9(12):R175
42. Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92(5):255–264
43. Margulies M et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376–380
44. Wang L et al (2009) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26(1):136–138

45. Sultan M et al (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321(5891):956–960
46. Cloonan N et al (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5(7):613–619
47. Mortazavi A et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621–628
48. de Magalhaes JP, Finch CE, Janssens G (2010) Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing Res Rev* 9(3):315–323
49. Zhou XG et al (2010) The next-generation sequencing technology: a technology review and future perspective. *Sci China Life Sci* 53(1):44–57
50. Nowrousian M (2010) Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryot Cell* 9(9):1300–1310
51. Simon SA et al (2009) Short-read sequencing technologies for transcriptional analyses. *Annu Rev Plant Biol* 60:305–333
52. Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* 10:135–151
53. Morin RD et al (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* 45(1):81
54. Graf A et al (2009) Yeast systems biotechnology for the production of heterologous proteins. *FEMS Yeast Res* 9(3):335–348
55. Wang B et al (2010) Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. *Nucleic Acids Res* 38(15):5075–5087
56. Simon C, Daniel R (2009) Achievements and new knowledge unraveled by metagenomic approaches. *Appl Microbiol Biotechnol* 85(2):265–276
57. Rubina AY et al (2005) Quantitative immunoassay of biotoxins on hydrogel-based protein microchips. *Anal Biochem* 340(2):317–329
58. Zhu H, Snyder M (2003) Protein chip technology. *Curr Opin Chem Biol* 7(1):55–63
59. Zong Y et al (2007) Forward-phase and reverse-phase protein microarray. *Methods Mol Biol* 381:363–374
60. Wu P, Castner DG, Grainger DW (2008) Diagnostic devices as biomaterials: a review of nucleic acid and protein microarray surface performance issues. *J Biomater Sci Polym Ed* 19(6):725–753
61. Kukar T et al (2002) Protein microarrays to detect protein–protein interactions using red and green fluorescent proteins. *Anal Biochem* 306(1):50–54
62. Tonkinson JL, Stillman BA (2002) Nitrocellulose: a tried and true polymer finds utility as a post-genomic substrate. *Front Biosci* 7:c1–c12
63. Grainger DW et al (2007) Current microarray surface chemistries. *Methods Mol Biol* 381:37–57
64. Reck M et al (2007) Optimization of a microarray sandwich-ELISA against hINF-gamma on a modified nitrocellulose membrane. *Biotechnol Prog* 23(6):1498–1505
65. Walter J-G, Reck M, Praulich I (2010) Protein microarrays: reduced autofluorescence and improved LOD. *Eng Life Sci* 10(2):103–108
66. Walter JG et al (2008) Systematic investigation of optimal aptamer immobilization for protein-microarray applications. *Anal Chem* 80(19):7372–7378
67. Doyle A, Griffiths J, Newel D (1994) *Cell & tissue culture: laboratory procedures*. Wiley, New York, pp 3:01–3:03
68. Wilkening S, Stahl F, Bader A (2003) Comparison of primary human hepatocytes and hepatoma cell line Hepg2 with regard to their biotransformation properties. *Drug Metab Dispos* 31(8):1035–1042
69. Castro-Melchor ML, Le H, Hu W-S (2011) Transcriptome data analysis for cell culture process. *Adv Biochem Eng Biotechnol*

70. Achilles J et al (2007) Isolation of intact RNA from cytometrically sorted *Saccharomyces cerevisiae* for the analysis of intrapopulation diversity of gene expression. *Nat Protoc* 2(9):2203–2211
71. Browne SM, Al-Rubeai M (2007) Selection methods for high-producing mammalian cell lines. *Trends Biotechnol* 25(9):425–432
72. Brezinsky SC et al (2003) A simple method for enriching populations of transfected CHO cells for cells of higher specific productivity. *J Immunol Methods* 277(1–2):141–155
73. Kantardjieff A et al (2009) Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment. *J Biotechnol* 145(2):143–159
74. Jayat C, Ratinaud MH (1993) Cell cycle analysis by flow cytometry: principles and applications. *Biol Cell* 78(1–2):15–25
75. Majore I et al (2009) Identification of subpopulations in mesenchymal stem cell-like cultures from human umbilical cord. *Cell Commun Signal* 7:6
76. Moretti P et al (2010) Characterization and improvement of cell line performance via flow cytometry and cell sorting. *Eng Life Sci* 10(2):130–138
77. Spellman PT, Sherlock G (2004) Reply: whole-culture synchronization—effective tools for cell cycle studies. *Trends Biotechnol* 22(6):270–273
78. Dutton RL, Scharer J, Moo-Young M (2006) Cell cycle phase dependent productivity of a recombinant Chinese hamster ovary cell line. *Cytotechnology* 52(2):55–69
79. Fu X et al (2009) Estimating accuracy of RNA-seq and microarrays with proteomics. *BMC Genomics* 10:161