Gene Expression

17

Julia Tree, Angela Essex-Lopresti, Stefan Wild, Ute Bissels, Barbara Schaffrath, Andreas Bosio, and Michael Elmore

17.1 Introduction

The phenotype of a cell is determined by the amount, the proportion and the condition of proteins present in this cell. Although every cell in an organism possesses the same genetic information, only certain genes are transcribed into MESSENGER RIBONUCLEIC ACID (mRNA) according to the function and demands of the cell. Based on the information provided by the mRNA, the information is translated into the corresponding protein, contributing to a distinctive set of proteins for every cell and every status of the cell, defining its phenotype. The mapping of the whole human genome was completed in 2004 [1]. Researchers are focusing now on the illumination of functions and interactions of genes and gene products by measuring, for example, the number of activated genes. An established method is DNA MICROARRAY technology, which, as well as other established DNA and RNA detection methods, utilises the characteristic of RNA strands to form helices due to complementary sequences. This process of combining two RNA strands to form a double helix is called HYBRIDISATION. Since Southern introduced the blotting technique [2] for DNA, the HYBRIDISATION process has been used in a wide range of techniques for the recognition and quantification of DNA or RNA. Such "classical" HYBRIDISATION techniques measure one DNA or RNA sequence per HYBRIDISATION using a specific probe. In contrast, DNA microarrays consist of several thousands of specific probes arrayed in a two-dimensional pattern allowing the parallel investigation of thousands of genes. A more recent development in measuring the expression levels of genes, using next-generation sequencing technology, is RNA-Seq. In this method, the entire transcriptome (mRNA content) of the sample is sequenced. The read depth, or number of sequence reads, corresponding to each gene is used as a proxy of the expression level of that particular gene. RNA-Seq analysis is still in its infancy but has distinct advantages over traditional microarray.

Final manuscript submitted on October 06, 2016.

J. Tree (✉) · M. Elmore
Public Health England (PHE), Salisbury, UK
e-mail: julia.tree@phe.gov.uk;
mike.elmore@phe.gov.uk

A. Essex-Lopresti
Defence Science and Technology Laboratory, Salisbury, UK
e-mail: aeelopresti@dstl.gov.uk

S. Wild · U. Bissels · A. Bosio
Miltenyi Biotec GmbH, Bergisch Gladbach, Germany
e-mail: macs@miltenyibiotec.de;
ute.bissels@miltenyibiotec.de;
andreasbo@miltenyibiotec.de

B. Schaffrath
Eppendorf Nordic A/S, Hørsholm, Denmark
e-mail: schaffrath.b@eppendorf.dk

© Springer International Publishing AG 2019

M. J. Parnham et al. (eds.), *Nijkamp and Parnham's Principles of Immunopharmacology*,
https://doi.org/10.1007/978-3-030-10811-3_17

17.2 Principle of Microarray Technology

Microarrays are tiny devices made for the analysis of targets of interest with a high degree of parallelism. Initially, the technology evolved around the analysis of mRNA levels in cells in different states, taking "classical" HYBRIDISATION-based technologies to a new level. For "classical" HYBRIDISATION-based analysis, genomic DNA (Southern) or RNA (Northern), extracted from the tissue of interest, is immobilised on a membrane. A single specific nucleotide sequence (the probe) that is complementary to the sequence of interest is labelled and applied to the membrane to subsequently detect the corresponding gene or gene transcript (Fig. 17.1). For array analysis, this principle is reversed and applied to thousands of sequences of interest by immobilising DNA fragments (probes) with distinct sequences on a SUBSTRATE (a membrane, glass, silicon or plastic slides) at defined positions (see Box 17.1). Nucleic acids from the cells of interest are labelled and applied to the SUBSTRATE for HYBRIDISATION, and the hybridised nucleic acids are identified by their position on the array.

Box 17.1: Production of Microarrays

A variety of different array substrates (membranes, plastics, glass), in combination with a range of different coatings, are used as the solid phase for microarray production. Coatings permit the functionalisation of substrates with reactive groups, like aldehyde, epoxy or isothiocyanate moieties, to bind DNA probes on the substrates.

The DNA probes can be directly synthesised on the microarray substrate (in situ synthesis) or the complete DNA probes are spotted on the substrate. The in situ synthesis, by photomediated synthesis or inkjet technology, allows a parallel production of

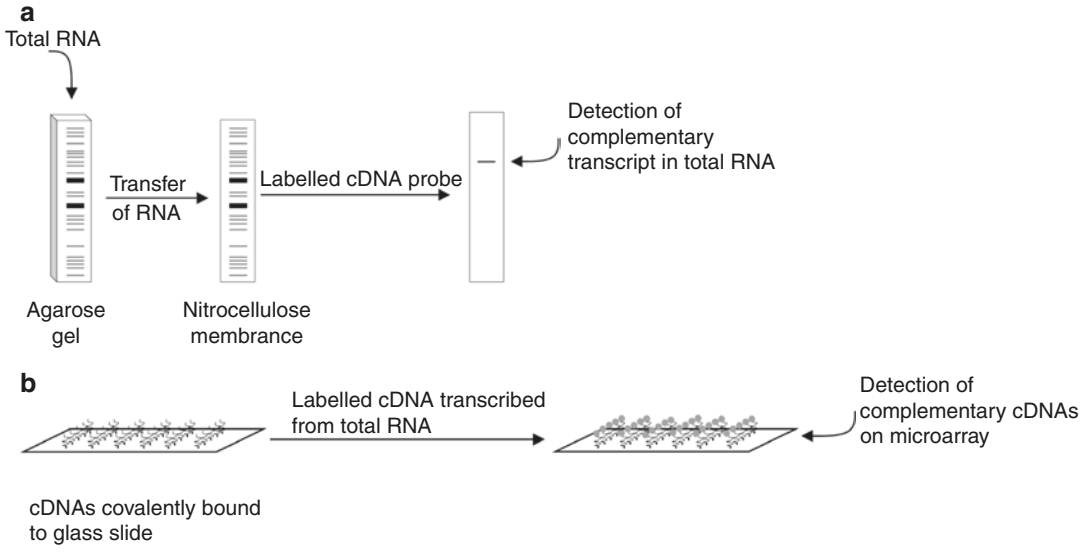

Fig. 17.1 Comparison of traditional Northern-blot and DNA microarray. (**a**) Total RNA of the tissue of interest is separated by gel electrophoresis and is blotted to a membrane. A labelled cDNA probe complementary to the transcript of interest is labelled and hybridised to the membrane. If the transcript is present in the total RNA, a signal can be detected due to hybridisation of probe and transcript. One experiment—one gene using a single labelled probe. (**b**) Several cDNAs (hundreds to thousands) complementary to mRNA transcripts of selected genes are covalently bound to a glass slide at defined positions (spots). Total RNA from the tissue of interest is transcribed into cDNA and labelled by reverse transcription. The labelled cDNA is hybridised to the bound cDNAs. Signals can be detected after hybridisation of two complementary cDNAs

OLIGONUCLEOTIDE ARRAYS, comprising oligonucleotides of 20–60 nucleotides in length [3, 4]. The use of short oligonucleotides (20–30 base pairs) is suitable to differentiate between perfectly matched duplexes and single-base or two-base mismatches [5–7]. When working with short oligonucleotide probes, the use of several different oligonucleotides corresponding to a single gene is typically required to enhance the reliability of the hybridisation signals [8].

Alternatively, cDNA fragments or presynthesised oligonucleotides with a length of up to 70 base pairs are spotted on the functionalised substrate in two manners: contact printing and non-contact printing.

CONTACT PRINTING typically involves rigid pins dipping into the spotting buffer containing the DNA probes. The drop at the tip of the pin is brought close to the surface at a given position, and a tiny drop remains on the surface. Non-contact printing methods are based on inkjet technology. The spotting buffer containing the DNA probes is dispensed as tiny droplets from the print head. Independent of the spotting mode, binding of the DNA probes occurs at the position of the drop. After the actual spotting process is completed, unbound DNA is removed, and the reactive substrate is blocked to avoid non-specific (independent of the provided sequence) binding of nucleic acids during hybridisation. The microarrays are now ready for processing.

The workflow of this process is illustrated by means of a DNA MICROARRAY experiment: In a typical scenario, GENE EXPRESSION of tumour cells, for instance, is compared to that in normal cells. RNA from tumour and normal cells is extracted from the respective tissue (Fig. 17.2). The RNA is transcribed into its reverse complimentary copy, the so-called cDNA. The cDNA derived from tumour cells and normal cells is labelled and applied to the DNA array. During the HYBRIDISATION step, the labelled nucleic acids bind to the complementary sequences of the respective probes. After washing away all unbound labelled nucleic acids, the signal intensities for each probe position are determined. After signal intensities have been generated for all probes on the array, signals derived from normal cells and tumour cells are compared, and differences in GENE EXPRESSION are identified. The altered expression of certain genes in the tumour, such as oncogenes, can help to typify the tumour. Combining the expression profile with clinical data may then be used to decide on the prognosis and the best therapy for the patient.

In addition to the described GENE EXPRESSION PROFILING, microarrays are also used to investigate other nucleic acids like genomic DNA [9] or non-coding RNAs [10] including MICRORNAS (miRNAs) [11–15]. In addition, the array principle has also been adapted to other ANALYTES such as proteins [16] or carbohydrates [17].

Due to the parallel measurement of up to thousands of ANALYTES, microarrays offer the opportunity to observe complex biological systems while using minimal amounts of sample material. Although in the following sections specifications and workflow procedures are mainly related to DNA microarrays for GENE EXPRESSION PROFILING, the general aspects hold true for other MICROARRAY-based technologies as well.

17.3 Application of Microarrays

17.3.1 Preparation and Quality of RNA

The first crucial step to achieve reliable GENE EXPRESSION results is RNA isolation. RNA is susceptible to chemical hydrolysis and to RNases, widespread enzymes that digest RNA molecules into small pieces. If the RNA is slightly degraded or contaminated by residual genomic DNA, for instance, the results may be biased and irreproducible (see also Box 17.2). Commonly, RNA is

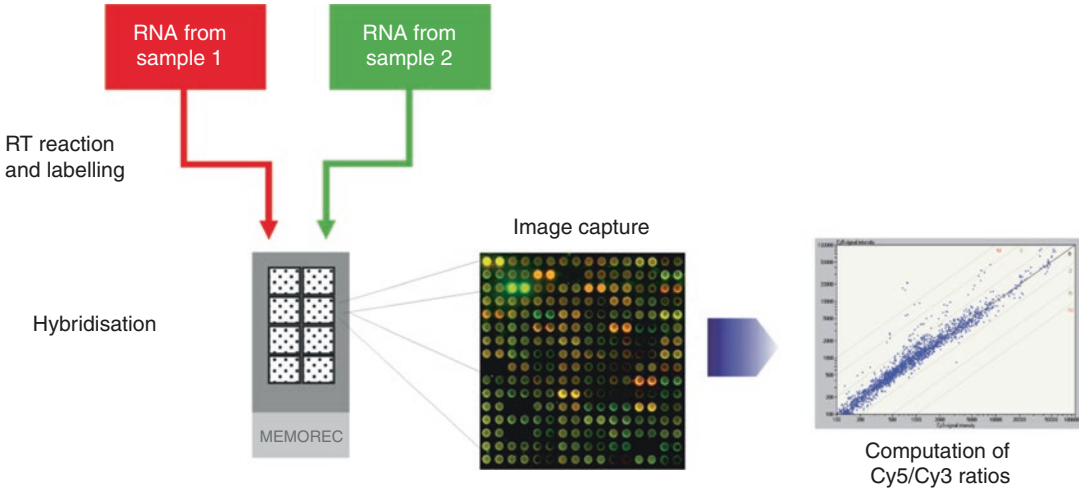

Fig. 17.2 Workflow diagram of microarray analysis

extracted from cells or tissues using organic solvents or silica filter-based methods. Since RNA extraction protocols may influence the outcome of the expression analysis, the same extraction procedure should be used for all samples analysed in one set of experiments.

Box 17.2: Quality of Total RNA

Integrity and purity are the most critical factors for the quality of RNA.

- Ratio of 28S rRNA and 18S rRNA should be 2, reflecting the higher molar mass of 28S rRNA compared to 18S rRNA. A more precise quality measure is given by the RNA integrity number (RIN) calculated by the Agilent Bioanalyzer.
- Ratio of the extinction 260 nm/280 nm should be between 1.8 and 2.0.
- The sample can be treated with RNase-free DNase to avoid contamination of genomic DNA.
- Protocols for RNA extraction have to be adapted according to the analysed tissue (e.g. high fat content or fibrous tissue).
- The choice of the preparation protocol may have an influence on the range of transcript lengths present in the extracted RNA (e.g. silica filters usually have a cut-off size of about 50–100 bases). Therefore, preparations derived in this way do not contain the whole range of fragment lengths. This might have an impact on the subsequent steps (labelling, amplification or hybridisation).

17.3.2 Amplification of RNA

The SENSITIVITY of MICROARRAY experiments strongly depends on the amount of material used for HYBRIDISATION. As the amount of RNA is usually limited, different AMPLIFICATION methods are available. The most common method utilises T7 DNA-dependent RNA polymerase to amplify RNA. The mRNA is first reverse transcribed to cDNA. The primer used for the reverse transcription additionally comprises the sequence of the T7 promotor. After the second strand synthesis, the T7 promotor is used by the T7 DNA-dependent RNA polymerase for in vitro transcription. The T7

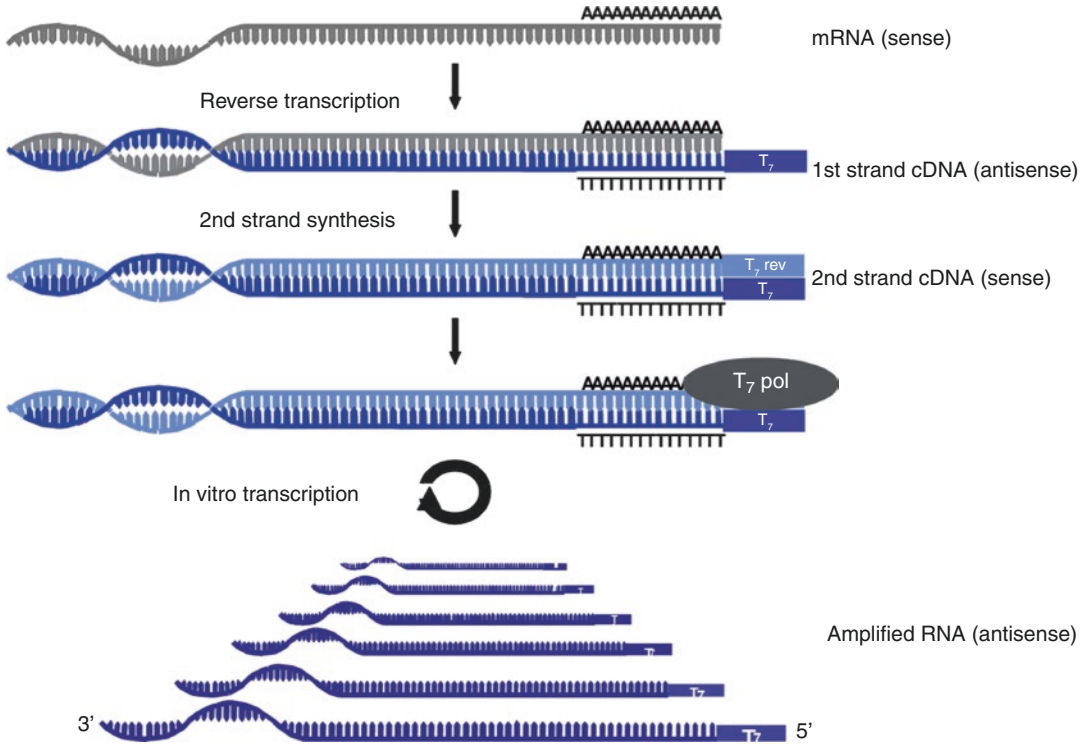

Fig. 17.3 Schematic diagram of T7 polymerase-based mRNA amplification

DNA-dependent RNA polymerase repeatedly transcribes the same cDNA thereby amplifying the original RNA (Fig. 17.3) [18]. In case even higher SENSITIVITY is needed, the amplified RNA can again be used as SUBSTRATE for cDNA synthesis and a second round of T7-based AMPLIFICATION. Alternatively, a variety of other AMPLIFICATION methods like PCR-based AMPLIFICATION methods have been developed (Fig. 17.4) [19]. Due to the slightly different properties of the different RNAs, such as length, sequence or GC content, the AMPLIFICATION efficiency can vary for different RNAs, again depending on the AMPLIFICATION method. Therefore, to allow comparison of different RNA samples, it is advisable to use the same AMPLIFICATION method for all samples. The most sensitive AMPLIFICATION methods allow MICROARRAY experiments from as little as a single cell (see also Sect. 5.1 and Fig. 17.4).

17.3.3 Dyes, Labelling and Hybridisation Methods

Most commonly, fluorescent dyes are used to detect the hybridised samples on microarrays, but alternative labelling methods using radioactivity or silver particles, for example, can also be applied.

In DIRECT LABELLING protocols, the labelled nucleotides are incorporated during the cDNA synthesis or the T7 DNA-dependent RNA polymerase-based AMPLIFICATION. Since the incorporation rate of labelled nucleotides is compromised by the partly bulky fluorescent dye, two-step labelling protocols (INDIRECT LABELLING) have also been established. During a two-step labelling procedure, nucleotides, labelled with a small molecule like biotin or an aliphatic amine, are incorporated by the polymerase. In a second step, the fluorescent dyes are linked to the modified nucleotides via

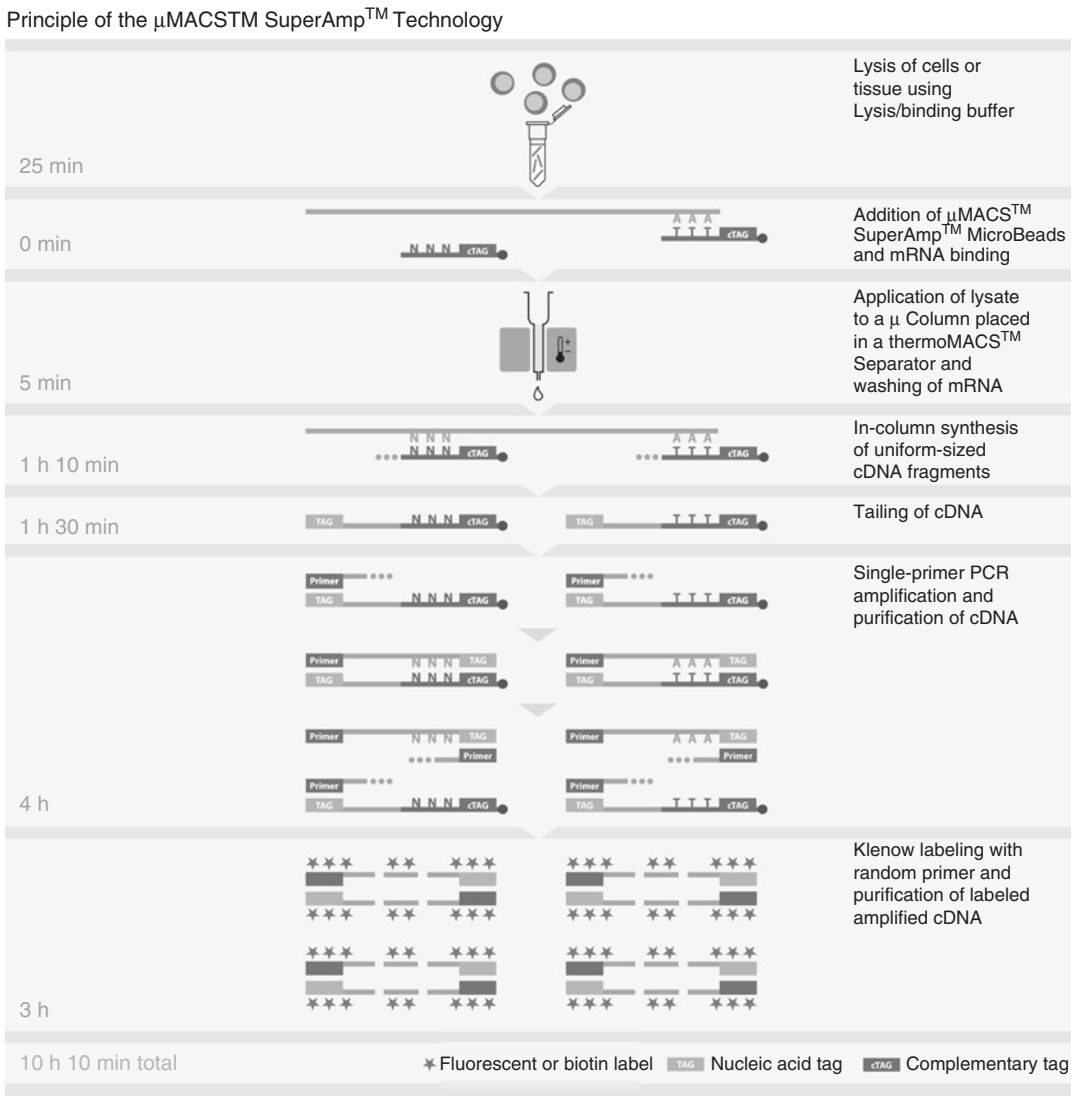

Fig. 17.4 Schematic diagram of a global PCR-based mRNA amplification

streptavidin or amine reactive groups like NHS esters. Depending on the system, the second step of the labelling protocol can also be performed after the HYBRIDISATION step (on-chip labelling).

After the labelling, the samples are hybridised on the MICROARRAY. The HYBRIDISATION can either be achieved by simple diffusion of the TARGET DNA molecules to the corresponding probes, or probe TARGET interaction can be assisted by moving the HYBRIDISATION mixture on top of the array. After the HYBRIDISATION step has been completed, unbound labelled TARGET molecules are removed by washing the array. Finally, the array is dried.

To minimise experimental variance caused by some of the processing steps, like the labelling or HYBRIDISATION, it is advisable to perform replicate MICROARRAY experiments using the same sample.

The HYBRIDISATION is usually performed as a one- or two-colour experiment. For one-colour experiments, each sample is hybridised on

one array, and the signal intensities derived from different arrays are compared. When using two colours, the two samples to be compared are labelled with different dyes and hybridised together on the same array. The direct comparison of the two samples on one array has the advantage that any experimental bias related to the array or the HYBRIDISATION step will affect both samples, therefore reducing detection of artefacts. When working with fluorescent dyes, however, the integrated dyes not only differ in their emission wavelength but also in the fluorescence intensity gained per dye, due to wavelength-dependent scanner properties, diverse quantum yields of the dyes or different stabilities of the dyes. Therefore, the raw data gained by two-colour MICROARRAY experiments has to be corrected for such dye effects. The methods used to centre or normalise the signal intensities for both wavelengths are based on the assumption that some of the genes, like housekeeping genes, are not regulated (see Box 17.3). The differences found for these genes can therefore be used to calculate a factor reflecting the different dye properties. As the reproducibility of array production and MICROARRAY HYBRIDISATIONS has dramatically improved, there is a trend in favour of single-colour HYBRIDISATIONS.

Box 17.3: Normalisation of Microarray Data

Integrity and purity are the most critical factors for the quality of RNA.

The main idea of NORMALISATION for dual-labelled samples is to adjust differences in the intensity of the two labels. Such differences result from the efficiency of dye integration, differences in amount of sample and label used and settings of laser power and photomultiplier. NORMALISATION of one channel arrays mainly corrects spatial heterogeneity. Although NORMALISATION alone cannot control all systematic variations, NORMALISATION plays an important role in the earlier stage of microarray data analysis because expression data can vary significantly due to different NORMALISATION procedures. A number of NORMALISATION methods have been proposed, but there is no general rule which method performs best. The NORMALISATION method strongly depends on several factors like the number of detectable genes, the number of regulated genes, signal intensities, quality of the hybridisation, etc.

For a rough classification, global NORMALISATION can be distinguished from local (signal intensity-dependent) NORMALISATION and NORMALISATION via transcripts known to be nonregulated or spike-in controls.

If global NORMALISATION is used, a single NORMALISATION factor is applied to all detectable genes, leading to a linear shift of all signal intensities. The underlying assumption is that constant systematic variations occur, including a lower integration rate of one dye in respect to the second dye. However, global NORMALISATION based on the median of all detected genes can only be used if a sufficient number of genes are nonregulated. If it is expected that most of the genes are regulated (which is of special interest regarding miRNA arrays), a set of "housekeeping genes" or spike-in controls should be included in the array configuration. Because housekeeping genes (by definition) are not regulated, the signal intensities of those genes should be the same on dual-labelled arrays. Using local NORMALISATION, a different NORMALISATION factor is calculated for every gene. Local NORMALISATION offers the opportunity of a signal intensity-dependent NORMALISATION. Some variations (e.g. laser settings) have different impacts on detected genes depending on their signal intensity. Thus, a non-linear shift of the signal intensities can be achieved based on the signal intensity of each single spot.

In the field of miRNA microarray research, NORMALISATION via spike-in controls is preferably used, as global NORMALISATION methods may fail due to (a) missing housekeeping miRNA, (b) limited number of expressed miRNAs and (c) a general up- or downregulation of many miRNAs. The used spike-ins represent a set of synthetic RNAs, which have no similarity to any known miRNA. The spike-ins are added to all experimental and control samples, and all signal intensities of the investigated samples are normalised using the median of the spike-ins.

17.3.4 Control Samples

The measurement of GENE EXPRESSION in a given sample is usually referred to the GENE EXPRESSION in other samples, here referred to as "control". Obviously, it is very important to choose the right control in order to gain valuable data. The best controls in most experiments are untreated cells or unaffected tissue of the same origin as treated cells or affected tissue, respectively. However, for practical or ethical reasons, it is not always possible to receive untreated cells or healthy tissue of the same origin, which is especially true for material derived from patients. If it is impossible to get matched control samples, a "related" control can be established, for instance, by pooling RNA from different individuals to reduce the effects of particular properties of single individuals in the control. In some cases, cell lines might also be a sufficient control. Alternatively, a pool of all samples used in an experimental series can work as the control (see Box 17.4). However, a sample pool carries the risk of missing genes that are consistently expressed differentially in all samples. In general, controls should either be case-matched to the samples of interest or consist of pooled material to compensate for individual differences.

Box 17.4: The Reference Strategy for Two-Colour Hybridisations

In microarray experiments, the direct comparison of absolute signal intensities of different microarrays can be critical due to different hybridisation efficiencies. To avoid this obstacle, two-colour microarray hybridisations can be performed. In two-colour microarray hybridisations, the sample, labelled with Cy5, for instance, and the control, labelled with Cy3, are hybridised on the same microarray. As the labelled molecules compete for the same probes on the microarray, the hybridisation efficiency is also the same and allows a direct comparison of sample versus control. Therefore, the ratio of the signal intensities of the two dyes represents the proportion of the analyte in the sample compared to the control. The principle of two-colour hybridisation can be extended to compare more than two samples by applying a reference scheme. For a microarray reference experiment, each of several samples and controls is hybridised versus the reference. The reference can then be used to compensate differences of the hybridisation efficiency for each microarray and allows standardisation and cross-referencing of microarray experiments. For the analysis of mRNA expression profiles, references consisting of total RNA mixtures are used [20]. For miRNA analysis, universal references consisting of known amounts of synthetic miRNAs are available [21]. Besides the cross-referencing of array experiments, such a reference allows the absolute quantification of miRNAs. The universal reference, consisting of an equimolar pool of about 1000 miRNAs, is labelled and hybridised versus each sample in a two-colour microarray approach. In this way, each single miRNA is quantified in comparison to an identical standard, compensating the bias related to sequence, labelling, hybridisation or signal detection.

17.4 Array Data: Acquisition, Analysis and Mining

17.4.1 Data Acquisition

Data acquisition of MICROARRAY experiments consists of two parts: the read-out of the MICROARRAY, meaning the detection of the signals, and the following image analysis. Whereas films have been used to detect radioactive signals, nowadays predominantly MICROARRAY scanners are used to excite commonly used dyes and measure the emitted fluorescence signals. The picture derived from the read-out of the MICROARRAY is saved as greyscale TIFF images for further analysis. During the next step, the signal intensity of each spot is determined and assigned to the gene represented by the given spot using appropriate image analysis software. In addition, the background signal, usually gained from the surrounding area of each spot, is subtracted from the signal to receive the net signal intensity. Spots of poor quality (empty or negative spots, irregular shape, spots showing background smears) can be excluded from further analyses. The set of data that results from the data acquisition step is referred to as primary data.

17.4.2 Data Analysis and Mining

For the analysis of the primary data, weak signals are excluded as non-reliable. The minimum reliable signal intensity of a spot can be determined by setting a minimum threshold for signal intensities, which is either dependent on the background or on negative controls. For some microarrays, *p*-values giving an estimate of the likelihood of the signal differing from background signals are used to indicate the reliability of the detected genes. To compare different samples, ratios of the signal intensities gained, such as for sample versus control, are computed for every detected gene. To correct for different labelling and HYBRIDISATION efficiencies, as well as for potential dye bias in two-colour MICROARRAY hybridisations, the signal intensities are centred or normalised prior to calculating the ratios (see Box 17.3).

Because of the multiparametric nature of MICROARRAY experiments, data mining and bioinformatics analysis are essential for interpretation of the numerical data produced by (series of) MICROARRAY experiments. Starting from relatively simple demands for appropriate visualisation of the data, bioinformatics tools are necessary to focus on candidate genes and reveal subtle changes in expression patterns.

A reliable identification of candidate genes by statistical methods is only possible if a sufficient number of replicate experiments have been performed. Technical replicates using the same starting material are usually performed to define the overall reproducibility of MICROARRAY experiments. Biological replicates are important to discriminate individual differences (e.g. patient specific) from general changes of GENE EXPRESSION (e.g. disease specific).

Additional bioinformatics methods can be used to identify groups of genes showing a comparable regulation. One method commonly used is the HIERARCHICAL CLUSTER ANALYSIS where genes and arrays are ordered by similarity in expression [22]. Due to the overwhelming amount of data, it is often difficult to understand MICROARRAY results in the light of certain biological questions. To assist researchers in interpreting the results, MICROARRAY data can be combined with knowledge stored in diverse databases like pathway information, genomic localisation or protein family classification.

Different data analysis tools can be applied to identify genes that may be related to a disease or treatment of interest. Linking the data to biological knowledge can also elucidate possible functions of the genes interest. Succeeding experiments using mostly molecular biology techniques like RT-PCR, in situ HYBRIDISATION, RNAi, knockout experiments, etc. are commonly performed to validate and corroborate the biological function concluded from the MICROARRAY data.

17.5 An Example of a Microarray Experiment

17.5.1 Global RNA Amplification and Microarray Analysis of T Cell Subpopulations

Naïve T cells differentiate in response to pathogens into multiple CD4⁺ and CD8⁺ subsets (see Chap. 3). To improve the understanding of this differentiation process as well as the nature of the different subsets, GENE EXPRESSION PROFILING has been used. As an example, MICROARRAY experiments were performed from ten different subpopulations covering the major stages of post-thymic CD4⁺ and CD8⁺ T cell differentiation (Fig. 17.5) [23]. The CD4⁺ and CD8⁺ subsets were isolated by immunomagnetic and flow cytometric cell sorting (see Chap. 16) based on the expression of CD4/CD8, CD27, CD28, CD45RA and CCR7. These markers characterise the major steps of T cell differentiation from naïve to highly differentiated cells in humans [24, 25]. The GENE EXPRESSION profiles were generated from multiple T cell subsets independently gained from two blood samples. As only limited cell numbers can be isolated from 20 mL of blood, a global PCR AMPLIFICATION method was applied allowing MICROARRAY experiments from 1000 cells per T cell population.

For the AMPLIFICATION of RNA from small cell numbers, loss of material is critical, and the pipetting of samples from one tube to another should be avoided as much as possible. For the global AMPLIFICATION, the cells were collected in a small volume of buffer and lysed (Fig. 17.4). Then, superparamagnetic oligo dT microbeads were directly added to the cell lysate binding the poly(A) residues of the mRNA. The labelled cell lysate was applied to a column that was placed in the magnetic field of a heatable permanent magnet. The magnetically labelled mRNA was retained in the strong magnetic field, while effective washing steps removed all other cell components. In-column cDNA synthesis and purification was performed in the same column used for mRNA isolation to avoid loss of material. Oligo dT and randomer oligonucleotides coupled to microbeads were used as primers for the cDNA synthesis. Thereby, cDNA fragments of uniform size were generated, and each transcript was represented by several cDNA

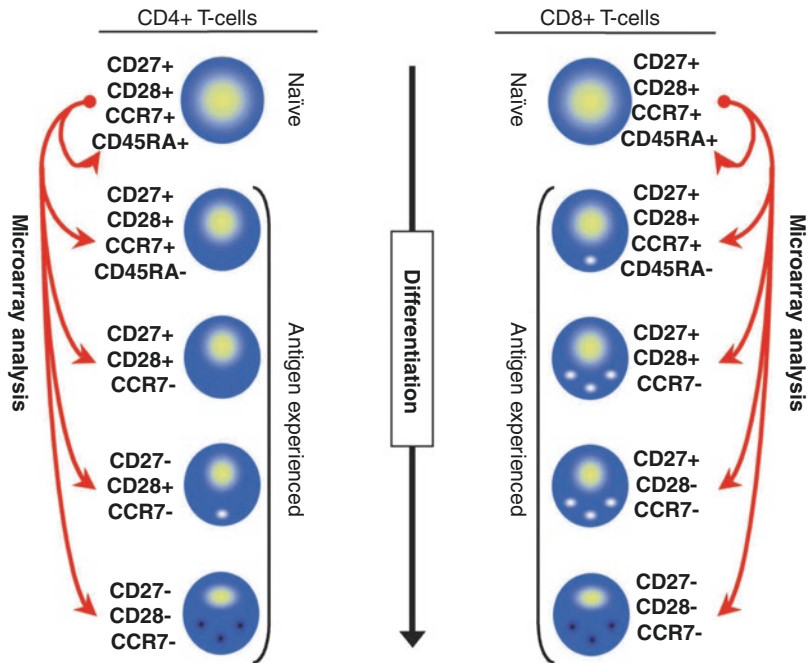

Fig. 17.5 Microarray analysis of CD8⁺ and CD4⁺ T cell subpopulations defining distinct stages of differentiation

fragments enabling uniform AMPLIFICATION during PCR. After eluting the cDNA fragments from the column, a tag was added to the 3′ end of each cDNA fragment by utilising a terminal deoxynucleotidyl transferase. A global PCR amplified the uniform-sized cDNA fragments 10^6-fold, resulting in sufficient TARGET material for MICROARRAY HYBRIDISATION. The PCR performed with a single primer enabled unbiased AMPLIFICATION due to the uniform annealing temperature. The primer binding site at the 3′ end was added during cDNA tailing. The complementary sequence of the tag was inserted at the 5′ end of the cDNA fragments during cDNA synthesis. After purification of the PCR products, a Klenow fragment labelling procedure with random primers in the presence of labelled nucleotides, in this case, Cy3-dCTP or Cy5-dCTP, yielded labelled DNA fragments that were used for MICROARRAY HYBRIDISATION.

All differentiated T cell subsets were hybridised against the corresponding naïve T cells as control in two-colour MICROARRAY experiments. Therefore, the genes found differentially expressed on the microarrays represented potential genes related to the differentiation from naïve to antigen-experienced T cells.

For the first differentiation stage ($CD27^+$/$CD28^+$/$CCR7^+$/$CD45RA^+$), about 15% of the detected genes were found to be differentially expressed, and this proportion increased for stages 2–5 to about 50%, which is consistent with the differentiation process.

A detailed analysis of the differentially detected genes revealed the acquisition of a cytolytic program by the highly differentiated T cells represented by the expression of genes encoding for the lytic granule membrane protein LAMP-3 and the CYTOTOXIC factors granzyme B and perforin. The up-regulation of these genes giving rise to lytic and CYTOTOXIC proteins supported the idea of CYTOTOXIC T cells as late differentiation state.

Another interesting set of genes was found downregulated in highly differentiated T cells. These genes encode for proteins involved in cell cycle entry and/or cell proliferation, as well as anti-apoptotic factors, suggesting a quiescence state and limited survival potential for the highly differentiated T cells under stress or upon activation.

Overall, during the differentiation process, the changes in GENE EXPRESSION for the differentiated T cells compared to the naïve T cells became increasingly similar between $CD4^+$ and $CD8^+$ T cells. So despite the clear differences between naïve $CD4^+$ and $CD8^+$ T cells, the differentiation process might be orchestrated by analogous changes in the GENE EXPRESSION profile.

In summary, the GENE EXPRESSION analysis using global RNA AMPLIFICATION for MICROARRAY experiments suggested functional changes especially during the late differentiation state pointing to CYTOTOXIC potential and limited lifespan. In addition, common changes in the GENE EXPRESSION pattern pointed to a similar differentiation process for $CD4^+$ and $CD8^+$ T cells.

17.6 RNA Sequencing

17.6.1 Introduction

Microarrays are currently the most popular choice for studying changes in the transcriptome, and significant advances in medical research have been made possible in the last 20 years by applying this technique [26]. Despite this, however, microarray technology does have limitations. There can be difficulties with probe design/performance, for instance, some probes cross hybridise with other genes, while some non-specifically hybridise [27]. The dynamic range of a probe can be restricting [28]; when an mRNA is abundantly expressed, a DNA microarray shows saturation, while at the low end of abundance, it suffers a loss of signal. Another disadvantage of microarray technology is that it is generally limited only to those genomes that have been previously sequenced [26].

In the last 5–10 years, a new technology for studying changes in the host transcriptome has emerged; this is known as RNA sequencing (RNA-Seq). Instead of using molecular hybridization to “capture” transcript molecules of interest, RNA-Seq samples transcripts in the

starting material by direct sequencing using next-generation sequencing (NGS) technologies (Box 17.5). Once detected, transcript sequences are then mapped back to a reference. Reads that map back to the reference are then counted to assess the level of gene expression, the number of mapped reads being the measure of expression level for that gene or genomic region [26].

A comparison of RNA-Seq versus microarray technology is presented in Table 17.1. RNA-Seq has many advantages compared to microarray analysis, for example, it can detect novel transcripts, allele-specific expression and splice junctions, and it can also be applied to any species even if the reference genome is unknown [27]. RNA-Seq has a larger dynamic range than microarrays and can detect more accurately transcripts in low abundance in the presence of highly abundant transcripts [30]. Another significant advantage offered by RNA-Seq is the need for a lower input of RNA starting material [29]. Currently, the challenges associated with RNA-Seq technology are the complexity of the data analyses and storage of large amounts of data, which should not be underestimated [26]. Despite this, RNA-Seq technology is revolutionising transcriptomic analysis and provides a powerful tool to decipher global gene expression patterns far beyond the limitations of microarrays.

Table 17.1 Comparison of microarray and RNA-Seq technologies

	Microarray	RNA-Seq
Amount of RNA required	High	Low
Resolution	Several to 100 bp	Single base
Distinguish splice forms?	Limited	Yes
Discover new genes?	No	Yes
Strandedness?	No	Yes
Dynamic range	Few hundredfold	>8000-fold
Reproducibility	Yes	Yes
Cost	Medium	High (due to computation)

Adapted from Bauer et al., BMC Bioinformatics 2014, 15(Suppl 11):S3 [29]

Box 17.5: Next-Generation Sequencing (NGS)

Next-generation sequencing is a term used to describe a number of different modern sequencing technologies which have in general replaced the traditional Sanger-based platform. NGS is also known as "high-throughput" or "deep" sequencing which reflects the vast increase in the number of sequenced bases per run (typically 10^3- to 10^6-fold greater), with a corresponding reduction in cost per sequenced base. There are a number of competing platforms each with different characteristics (e.g. read length, sequencing capacity, error rate, cost per base), with perhaps Illumina being the most widely used at present (a typical run generating 500 Gb of data), in preference to Ion Torrent, SOLiD and 454 platforms. These are characterised by relatively short reads (35–1000 bases) and also require significant sample preparation and amplification. More recently, single-molecule real-time (SMRT) methods have been introduced, which require less sample preparation time, and yield much longer read lengths ('000's of bases per run), although the error rate is relatively high and they are at present more expensive, currently limiting the widespread adoption of this technology. Examples of this include PacBio and Oxford Nanopore systems.

17.6.2 RNA-Seq Experimental Workflow

A typical RNA-Seq workflow, including important factors for consideration, is shown in Fig. 17.6. Experimental planning is one of the most important factors to consider, including determining if RNA-Seq is the most appropriate technique to use. RNA-Seq generates a huge, potentially bewildering, amount of data, and it is

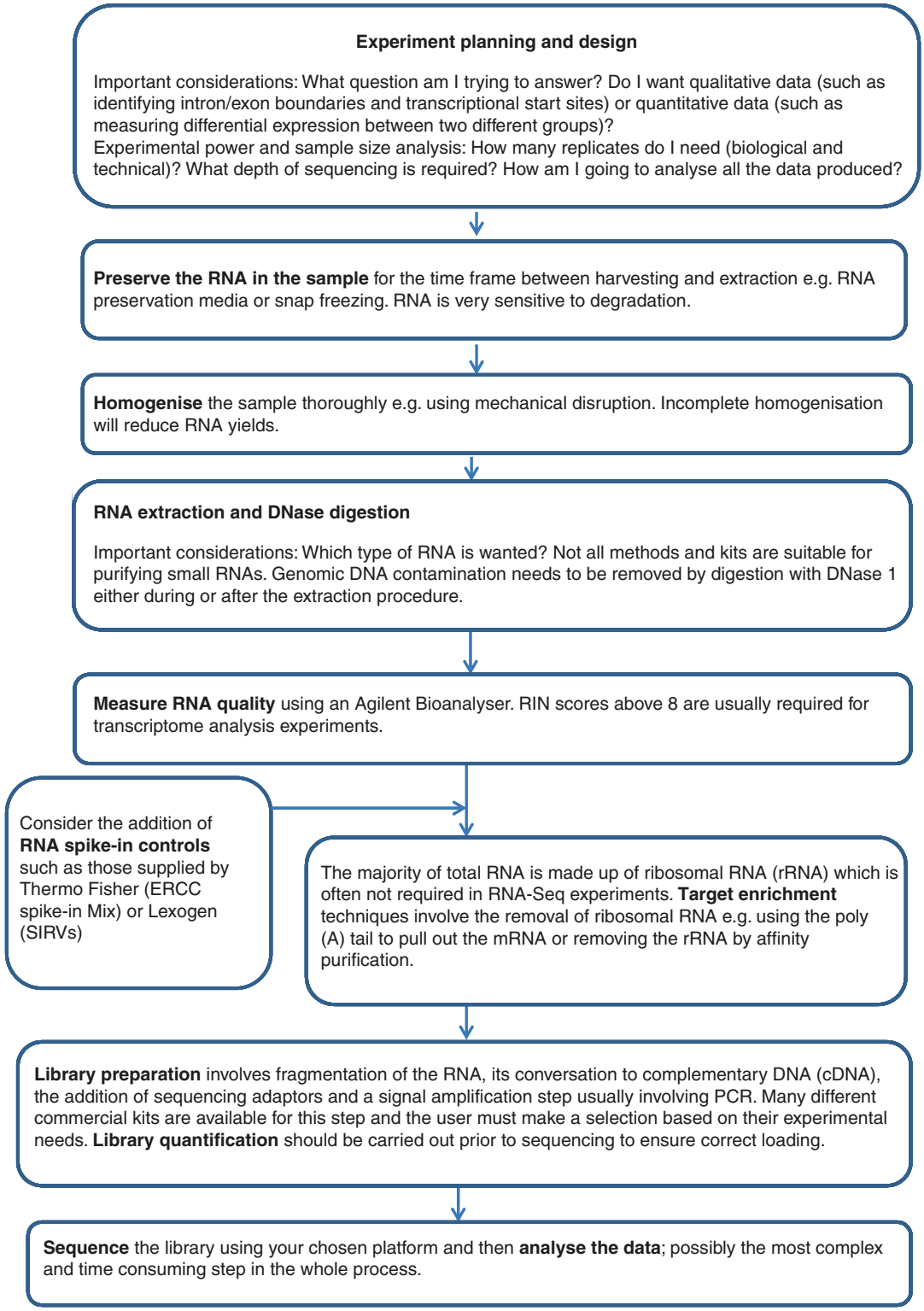

Fig. 17.6 Flow diagram for RNA-Seq experiment

vitally important to determine whether a more targeted approach such as reverse-transcription PCR would be preferable. The experimental design must ensure that it has the capability to answer the research question and is sufficiently powered including an appropriate number of biological replicates [31]. The success of RNA-Seq experiments is highly dependent upon recovering pure and intact RNA from samples which is free from DNA contamination. RNA is more labile than DNA, and RNases are very stable enzymes, so extra care should be taken when purifying and working with RNA as differential degradation of samples will adversely affect the experimental outcome. Several commercial kits exist for RNA extraction from various sample types ranging from bacteria and viruses to human and animal blood and tissue samples. Depending on the source of the RNA, it is vital to ensure the sample is thoroughly homogenised to ensure maximal isolation of RNA. Once the RNA has been isolated, the quality is usually assessed using an Agilent Bioanalyzer (see Box 17.2). In order to understand the biological variation in RNA-Seq results, it is important to understand and control for the technical variation that can be introduced at every step in the procedure from the operator carrying out the work, the method of homogenisation used to the methods of data analysis. To do this, spike-in external controls can be added to RNA-Seq experiments such as those designed by the External RNA Controls Consortium (ERCC; Thermo Fisher) or Spike-In RNA Variants (SIRVs, Lexogen).

The RNA isolated is usually the total RNA though most differential transcriptome analysis experimenters only want to look at the messenger RNA (mRNA). The majority (>95%) of the total RNA is made up of ribosomal RNA (rRNA). Before library preparation, many researchers choose to perform target enrichment to maximise the amount of their target RNA fraction in the final sample. Several commercial kits exist for either removing the rRNA from the sample or pulling out the mRNA using the poly-A tail.

The exact method of library preparation will depend on the sequencing platform being used and the experimental question being answered. Companies involved in sequencing sell a variety of library preparation kits. It is possible using RNA-Seq, if the appropriate library preparation method has been used, to obtain information from both the sense and antisense strands of the RNA template. This information is important for analysis of transcript orientation and the detection of overlapping transcripts. It is essential to accurately quantify the sequencing library before loading on the sequencing platform to ensure optimal performance and success of the sequencing run. Methods of sequencing library quantification include quantitative real-time PCR and spectrophotometry.

17.6.3 Choice of Sequencing Platform

A number of different NGS sequencing platforms (Box 17.5) are currently available for RNA-Seq experiments (see NGS review [32, 33]). Each platform has different characteristics (e.g. read length, number of reads per run, base-calling accuracy), and the choice will depend on the experimental question being asked. Typically, however, the Illumina platform (which yields relatively short reads (~150 bases), but has a low error rate and a high read depth) is the most commonly used. For more specialised applications where longer reads are required (e.g. novel transcriptome assembly), the PacBio RS instrument is more appropriate.

17.6.4 Data Analysis Pipeline

Once raw reads from the RNA-Seq experiment have been generated, they are processed and analysed through a series of software analysis steps. A schematic of a typical data analysis pipeline is shown in Fig. 17.7 and described as follows:

1. *Preprocessing and quality control of raw reads*: For samples that have been multiplexed (combined samples on one sequencing run), they must first be demultiplexed. Then adapter sequences are removed, generic quality

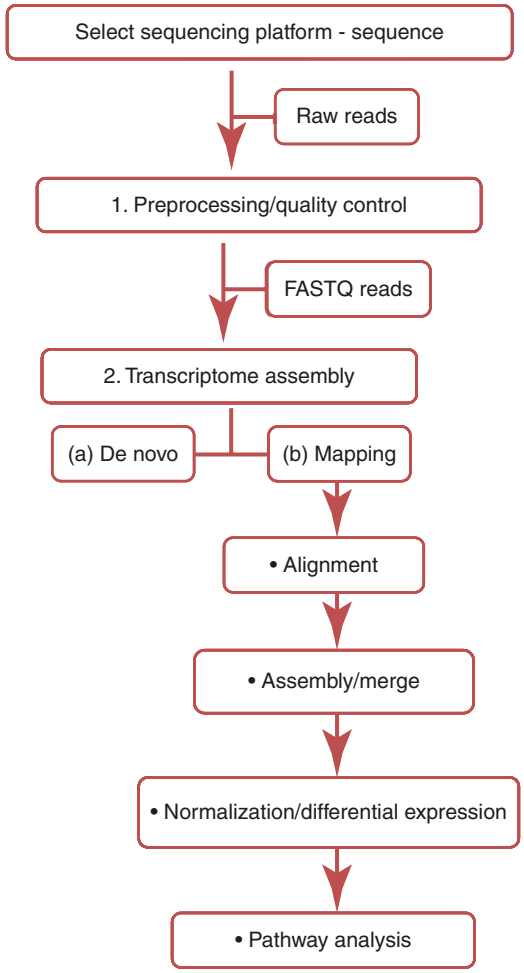

Fig. 17.7 Schematic data analysis pipeline

control steps may be performed (e.g. FastQC software), and reads are trimmed and filtered by base quality score to remove low-quality or contaminating reads (using "trimmomatic" for Illumina reads [34]). Sequence reads (base identity along with a quality score) are provided in a format known as FASTQ.

2. *Transcriptome assembly*: Assembly of the transcriptome falls into two methods depending on whether a reference genome or transcriptome of the organism under study exists. If a reference genome does not exist, then the de novo approach is taken (a); if a reference genome exists, a mapping approach is taken (b).

(a) *De novo approach*—This approach is more computer resource intensive than the mapping approach as each read has to be compared with every other read in order to generate a set of contigs, instead of just to one reference genome. Typically de novo transcriptome assembly software use a graph-based approach—examples are Velvet/Oases [35, 36], Trans-ABySS [37] and Trinity [38].

(b) *Mapping approach*—Assuming that a reference genome, transcriptome or annotation is available, many analyses and those leading to differential expression will use the mapping approach. As an example, typical programs from the "Tuxedo" protocol are used [39]: a representative pipeline could be:

- *Alignment*—Using TopHat software—Aligns an RNA-Seq read to the reference using the Bowtie short read aligner and then analyses the mapping results to identify splice junctions between exons.
- *Assembly/merge assemblies*—Using Cufflinks (assembles transcripts, estimates abundance and tests for differential expression and regulation in RNA-Seq samples) and Cuffmerge (merges together assemblies).
- *Normalization and differential expression*—The Cuffdiff software contains methods for normalization and calculation of statistically significantly changes in transcript expression. Numerous other methods exist, and those written for the free statistical software environment "R" are commonly used (e.g. edgeR [40], DESeq [41]).
- *Differentially expressed (DE) gene prediction*, using Cuffdiff, searches for significant changes in transcript expression. These genes may then be entered into a pathway analysis software package to identify key biological pathways.

An example of an RNA-Seq experiment which has used the described workflow is included in Box 17.6.

Box 17.6: Differential Transcriptome Analysis in an Animal Model of Bacterial Disease

The response a host makes to an infection is very complex. A full understanding of this response can lead to the development of new ideas for treatment [42]. RNA-Seq can be used to gain an understanding of this response by studying RNA isolated from an animal model over a time course of infection. BALB/c mice were exposed to an inhalational challenge of the Gram-negative pathogen *Burkholderia pseudomallei*, with a mean retained dose of 30 colony-forming units per mouse [43]. Animals were culled at predetermined time points, and tissues were harvested and stored immediately in RNA*later*. An identical number of control mice received a phosphate buffered saline exposure and handled in exactly the same manner as the infected groups. The mRNA was isolated from the lungs and sequenced using the workflow described above. The transcriptome obtained from the control group acted as the baseline comparator for the infected group. Once the transcriptome had been generated, Ingenuity Pathways Analysis (IPA; Qiagen, http://www.ingenuity.com) was used to interpret the data and apply biological meaning (Essex-Lopresti et al., personal communication).

When a host is exposed to a dose of bacteria, an important part of the response process is the body's ability to recognise the presence of the bacteria. The primary way of doing this is by the activation of pattern recognition receptors through the innate immune system [44]. Figure 17.8 shows the pattern recognition receptor pathway which has been overlaid with the transcriptome data from day 3, postexposure, of the *B. pseudomallei* aerosol infection. This transcriptome data shows that the mouse host has upregulated expression of several toll-like receptors (TLRs) which in turn leads ultimately to the release of cytokines via transcription factors such as NFκB. The cytokines released as a result of this pathway then circulate around the host influencing other response pathways and cell populations. Detailed knowledge of this and other response pathways helps researchers, trying to design novel drugs, to identify places where medical interventions might assist the host in its fight against the pathogen. © Crown copyright (2016), Dstl. This material is licensed under the terms of the Open Government Licence except where otherwise stated. To view this licence, visit http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

17.7 The Future of RNA-Seq

Developing technologies are likely to help shape the future of RNA-Seq for transcriptome profiling. For instance, advances in long-read sequencing technologies will impact upon RNA-Seq analysis. Currently, most RNA-Seq experiments are undertaken on platforms (e.g. Illumina) which yield relatively short-read lengths. Long-read sequencing technology (e.g. Pacific Biosciences (PacBio) single-molecule real-time (SMRT) sequencing approach) can produce reads matching longer transcripts. The advantage of longer reads is that a lot of mapping errors that occur after sequencing will be eliminated, and as a consequence the accuracy of sequencing will greatly improve.

Hand-held sequencing technologies are likely to impact on transcriptome profiling too.

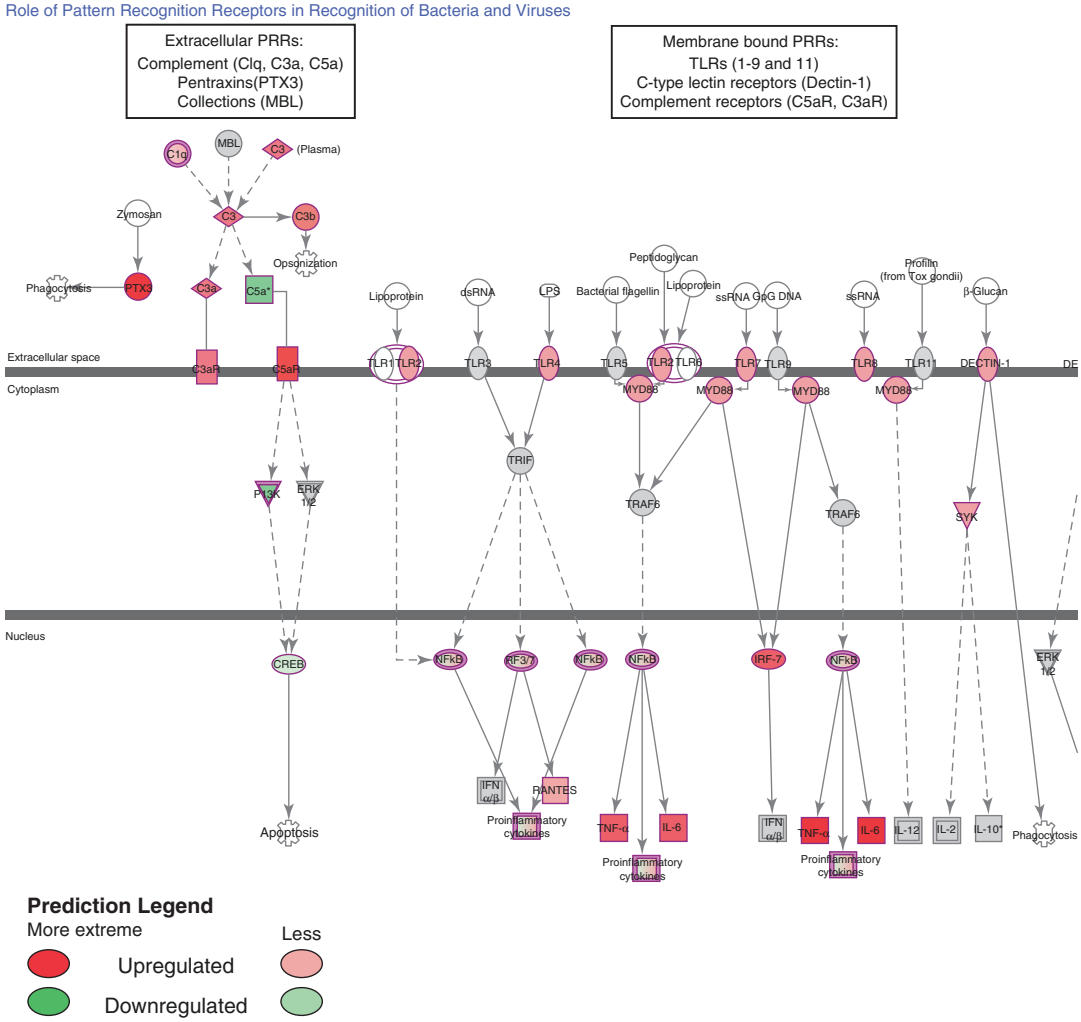

Fig. 17.8 Canonical pathway analysis. Canonical pathway, found within IPA, showing the role of pattern recognition receptors in recognition of bacteria and viruses overlaid with the transcriptome data from day 3 of the *B. pseudomallei* mouse aerosol infection model. Fold changes of up- and downregulated genes are indicated by red and green shading according to the prediction legend.

© Crown copyright (2016), Dstl. This material is licensed under the terms of the Open Government Licence except where otherwise stated. To view this licence, visit http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk

Currently MinIONs, palm-sized NGS sequencing devices from Oxford Nanopore (https://www.nanoporetech.com/), offer read lengths of tens of kilobases, limited only by the length of DNA molecules presented to it [45]. These devices were used to undertake the genomic surveillance of Ebola virus, in the field, in 2015 in West Africa [46]. Results were generated in less than 24 h after receiving an Ebola-positive sample where the sequencing process took as little as 15–60 min. Pathogen genomes are usually not very large and thus are suitable for use with this "miniaturised" technology. In the future however, RNA-Seq analysis might well become portable, and this might be aided by the use of direct RNA sequencing (DRS). DRS is another new pioneering technology, with which it is possible to carry out direct single RNA sequencing without prior conversion of RNA to cDNA. The advantage of DRS over RNA-Seq is that it removes the technical artefacts introduced

by having to create a cDNA library or by having amplification steps [47, 48].

RNA-Seq technology is currently evolving; some researchers have moved the science forward already and developed methods for analysing the transcriptome of individual cells, otherwise known as single-cell RNA sequencing (scRNA-Seq) [49]. This allows the complex analyses of heterogeneous samples and profiling of cell-to-cell variables on a genomic scale [50]. There are significant challenges with this technology (RNA losses, differences in strand-specificity and difficulty in distinguishing between noise and variability for low abundance transcripts), but it is hoped that advances in sequencing technology (such as nanopores, mentioned above) will overcome these barriers [51]. Given the high anticipated value of single-cell transcriptomics, explosive growth of scRNA-Seq data is expected in the next 5–10 years [49].

17.8 Overview

The ability to monitor changes in the mRNA expression of multiple genes by using microarray technology is firmly established. This technology is routinely used, reasonably cost-effective, reliable and highly reproducible. Many scientific advances in medical research and diagnosis have been made possible using this technique. RNA-Seq technology, however, is closely following in the footsteps of microarray technology, and once sequencing costs become lower, data analyses become more streamlined, and data storage issues are resolved, RNA-Seq is likely to significantly impact upon transcriptomic research in the future.

Selected Readings

Bosio A, Gerstmayer B, editors. Microarrays in inflammation (Progress in inflammation research). Basel: Birkhäuser Verlag; 2008.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51.

Müller UR, Nicolau DV, editors. Microarray technology and its applications. Berlin: Springer-Verlag; 2005.

Stekel D. Microarray bioinformatics. Cambridge, UK: Cambridge University Press; 2003.

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.

Recommended Websites

Functional Genomics Data Society – FGED Society: Minimum information about a microarray experiment – MIAME. http://fged.org/projects/miame/. Accessed Aug 2016.

Genetic Information Research Institute: Repbase. http://www.girinst.org. Accessed Aug 2016.

Illumina website – http://www.illumina.com/. Accessed Aug 2016.

miRBase: microRNA database. http://www.mirbase.org/. Accessed Aug 2016.

National Center for Biotechnology Information: Gene Expression Omnibus. http://www.ncbi.nlm.nih.gov/geo. Accessed Aug 2016.

RNA-seqlopedia – http://rnaseq.uoregon.edu/. Accessed Aug 2016.

Swiss-Prot: Curated protein sequence database. http://www.expasy.org/sprot/. Accessed Aug 2016.

Unigene: http://www.ncbi.nlm.nih.gov/unigene. Accessed Aug 2016.

References

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931–45.
2. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol. 1975;98(3):503–17.
3. McGall GH, Fidanza JA. Photolithographic synthesis of high-density oligonucleotide arrays. Methods Mol Biol. 2001;170:71–101.
4. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A. 1994;91(11):5022–6.
5. Hacia JG, Collins FS. Mutational analysis using oligonucleotide microarrays. J Med Genet. 1999;36(10):730–6.
6. Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, et al. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat Genet. 1999;22(2):164–7.
7. Okamoto T, Suzuki T, Yamamoto N. Microarray fabrication with covalent attachment of DNA using bubble jet technology. Nat Biotechnol. 2000;18(4):438–41.
8. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M,

Horton H, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol. 1996;14(13):1675–80.
9. Hester SD, Reid L, Nowak N, Jones WD, Parker JS, Knudtson K, Ward W, Tiesman J, Denslow ND. Comparison of comparative genomic hybridization technologies across microarray platforms. J Biomol Tech. 2009;20(2):135–51.
10. He H, Cai L, Skogerbo G, Deng W, Liu T, Zhu X, Wang Y, Jia D, Zhang Z, Tao Y, et al. Profiling *Caenorhabditis elegans* non-coding RNA expression with a combined microarray. Nucleic Acids Res. 2006;34(10):2976–83.
11. Barad O, Meiri E, Avniel A, Aharonov R, Barzilai A, Bentwich I, Einav U, Gilad S, Hurban P, Karov Y, et al. MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. Genome Res. 2004;14(12):2486–94.
12. Krichevsky AM, King KS, Donahue CP, Khrapko K, Kosik KS. A microRNA array reveals extensive regulation of microRNAs during brain development. RNA. 2003;9(10):1274–81.
13. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell. 2007;129(7):1401–14.
14. Liu CG, Calin GA, Meloon B, Gamliel N, Sevignani C, Ferracin M, Dumitru CD, Shimizu M, Zupo S, Dono M, et al. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. Proc Natl Acad Sci U S A. 2004;101(26):9740–4.
15. Thomson JM, Parker JS, Hammond SM. Microarray analysis of miRNA gene expression. Methods Enzymol. 2007;427:107–22.
16. Wulfkuhle J, Espina V, Liotta L, Petricoin E. Genomic and proteomic technologies for individualisation and improvement of cancer treatment. Eur J Cancer. 2004;40(17):2623–32.
17. Wang Z, Gao J. Microarray-based study of carbohydrate-protein binding. Methods Mol Biol. 2010;600:145–53.
18. Eberwine J. Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. Biotechniques. 1996;20(4):584–91.
19. Singh R, Maganti RJ, Jabba SV, Wang M, Deng G, Heath JD, Kurn N, Wangemann P. Microarray-based comparison of three amplification methods for nanogram amounts of total RNA. Am J Physiol Cell Physiol. 2005;288(5):C1179–89.
20. Novoradovskaya N, Whitfield ML, Basehore LS, Novoradovsky A, Pesich R, Usary J, Karaca M, Wong WK, Aprelikova O, Fero M, et al. Universal reference RNA as a standard for microarray experiments. BMC Genomics. 2004;5(1):20.
21. Bissels U, Wild S, Tomiuk S, Holste A, Hafner M, Tuschl T, Bosio A. Absolute quantification of microRNAs by using a universal reference. RNA. 2009;15(12):2375–84.
22. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25):14863–8.
23. Appay V, Bosio A, Lokan S, Wiencek Y, Biervert C, Kusters D, Devevre E, Speiser D, Romero P, Rufer N, et al. Sensitive gene expression profiling of human T cell subsets reveals parallel post-thymic differentiation for CD4+ and CD8+ lineages. J Immunol. 2007;179(11):7406–14.
24. van Lier RA, ten Berge IJ, Gamadia LE. Human CD8(+) T-cell differentiation in response to viruses. Nat Rev Immunol. 2003;3(12):931–9.
25. Appay V, Rowland-Jones SL. Lessons from the study of T-cell differentiation in persistent human virus infection. Semin Immunol. 2004;16(3):205–12.
26. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol. 2011;9:34.
27. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One. 2014;9(1):e78644.
28. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
29. Bauer MA, Chavan SS, Peterson EA, Heuck CJ, Johann DJ. Leveraging the new with the old: providing a framework for the integration of historic microarray studies with next generation sequencing. BMC Bioinformatics. 2014;15(Suppl 11):S3.
30. Weber AP. Discovering new biology through sequencing of RNA. Plant Physiol. 2015;169(3):1524–31.
31. Hart SN, Therneau TM, Zhang YJ, Poland GA, Kocher JP. Calculating sample size estimates for RNA sequencing data. J Comput Biol. 2013;20(12):970–8.
32. Mardis ER. Next-generation sequencing platforms. Annu Rev Anal Chem. 2013;6:287–303.
33. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51.
34. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
35. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18(5):821–9.
36. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics. 2012;28(8):1086–92.
37. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, et al. De novo assembly and analysis of RNA-seq data. Nat Methods. 2010;7(11):909–12.
38. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.

39. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562–78.
40. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
41. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.
42. Hancock RE, Nijnik A, Philpott DJ. Modulating immunity as a therapy for bacterial infections. Nat Rev Microbiol. 2012;10(4):243–54.
43. Lever MS, Nelson M, Stagg AJ, Beedham RJ, Simpson AJ. Experimental acute respiratory Burkholderia pseudomallei infection in BALB/c mice. Int J Exp Pathol. 2009;90(1):16–25.
44. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell. 2006;124(4):783–801.
45. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif. 2015;3:1–8.
46. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228–32.
47. Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM. Direct RNA sequencing. Nature. 2009;461(7265):814–8.
48. Ozsolak F. Attomole-level genomics with single-molecule direct DNA, cDNA and RNA sequencing technologies. Curr Issues Mol Biol. 2016;18:43–8.
49. Yu P, Lin W. Single-cell transcriptome study as big data. Genomics Proteomics Bioinformatics. 2016;14(1):21–30.
50. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol. 2015;33(3):285–9.
51. Saliba AE, Westermann AJ, Gorski SA, Vogel J. Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res. 2014;42(14):8845–60.