

Chapter 1

What Is the Transcriptome and How it is Evaluated?

Amanda F. Assis, Erna H. Oliveira, Paula B. Donate, Silvana Giuliatti, Catherine Nguyen and Geraldo A. Passos

Abstract The concept of the transcriptome revolves around the complete set of transcripts present in a given cell type, tissue or organ and encompasses both coding and non-coding RNA molecules, although we often assume that it consists only of messenger RNAs (mRNAs) because of their importance in encoding proteins. Unlike the nuclear genome, whose composition and size are essentially static, the transcriptome often changes. The transcriptome is influenced by the phase of the cell cycle, the organ, exposure to drugs or physical agents, aging, diseases and a multitude of other variables, all of which must be considered at the time of its determination. However, it is precisely this property that makes the transcriptome useful for the discovery of gene function and as a molecular signature. In this chapter, we review the beginnings of transcriptome research, the main types of RNA molecules found in a mammalian cell, the methods of analysis, and the bioinformatics pipelines used to organize and interpret the large quantities of data generated by the two current gold-standard methods of analysis: microarrays and high-throughput RNA sequencing (RNA-Seq). Attention is also given to non-coding RNAs, using microRNAs (miRNAs) as an example because they physically interact with mRNAs and play a role in the fine control of gene expression.

G. A. Passos (✉) · A. F. Assis · E. H. Oliveira
Molecular Immunogenetics Group, Department of Genetics, Ribeirão Preto Medical School,
University of São Paulo, 14049-900, Ribeirão Preto, Brazil
e-mail: passos@usp.br

P. B. Donate
Inflammation and Pain Group, Department of Pharmacology, Ribeirão Preto Medical School,
University of São Paulo, 14049-900, Ribeirão Preto, Brazil

S. Giuliatti
Bioinformatics Group, Department of Genetics, Ribeirão Preto Medical School,
University of São Paulo, 14049-900, Ribeirão Preto, Brazil

C. Nguyen
Laboratory TAGC INSERM U1090, National Institute for Health and Medical Research,
Scientific Park of Luminy, 13100, Marseille, France

© Springer International Publishing Switzerland 2014
G. A. Passos (ed.), *Transcriptomics in Health and Disease*,
DOI 10.1007/978-3-319-11985-4_1

1.1 What is the Transcriptome, How it is Evaluated and What Types of RNA Molecules Exist?

Strictly speaking, the *transcriptome* can be conceptualized as the total set of RNA species, including coding and non-coding RNAs (ncRNAs), that are transcribed in a given cell type, tissue or organ at any given time under normal physiological or pathological conditions. This term was coined by Charles Auffray in 1996 to refer to the entire set of transcripts. Soon after, this concept was applied to the study of large-scale gene expression in the yeast *S. cerevisiae* (Velculescu et al. 1997; Dujon 1998; Pietu et al. 1999).

However, due to the importance of messenger RNAs (mRNAs), which represent protein-coding RNAs, the term transcriptome is often associated with this set of RNA and as an analogy species. Researchers later coined the analogous term *miRNome* to refer to the total set of miRNAs.

The *proteome* is conceptually similar to the transcriptome and refers the total set of proteins translated in a given cell type, tissue or organ at any given time during normal physiological or pathological conditions. Nevertheless, despite its importance, the proteome will not be discussed in this book, and we suggest the following reviews for further reading: Anderson 2014; Forler et al. 2014; Padron and Dormont 2014; Altelaar et al. 2013; and Ahrens et al. 2010.

Analyses of the transcriptome began well before its conceptualization. Large-scale analyses of gene expression in the murine thymus gland (Nguyen et al. 1995), the human brain and liver (Zhao et al. 1995) and human T cells (Schena et al. 1996) have been performed since the mid-1990s. These independent groups used cDNA clones arrayed on nylon membranes or glass slides to hybridize labeled tissue- or cell-derived samples. These arrayed cDNA clones represented the prototypes of the modern microarrays currently used in transcriptome research (Jordan 2012).

1.1.1 How the Transcriptome is Evaluated: The Birth of Transcriptome Methods

Although the first method used to analyze transcriptional gene expression emerged in 1980 with the development of Northern blot hybridization (Wreschner and Hersberg 1977), this method was not and still is not capable of being performed on a large scale, and thus cannot be considered a transcriptome approach. In 1990s, the human genome project, through partially automated DNA sequencing, had the ambition to identify, characterize and analyze all of the genes in the human genome (Watson 1990; Cantor 1990). This revolutionary approach led to thousands of entries that were constructed via the tag-sequencing of randomly selected cDNA clones (Adams et al. 1991, 1992, 1993a, b; Okubo et al. 1992; Takeda et al. 1993), thus opening an avenue for high-throughput approaches by making these data widely available in repositories such as the dbEST database (<http://www.ncbi.nlm.nih.gov/dbEST>). As more and more genes are identified, efforts are now being

redirected towards understanding the precise temporal and cellular control of gene expression. The advances provided by the current progress in high-throughput technologies have enabled the simultaneous analysis of the activity of many genes in cells and tissues, essentially depicting a molecular portrait of the tested sample. The transcriptome approach, based on the large-scale measurement of mRNA, became the method of choice among the emerging technologies of so-called “functional genomics”, primarily because this method was rapidly identified as one that can be performed at a reasonably large scale using highly parallel hybridization methods, and it has allowed a more holistic view of what is really happening in the cell (Sudo et al. 1994; Granjeaud et al. 1996, 1999; Botwell 1999; Jordan 1998).

As mentioned above, the first transcriptome analysis was performed on large nylon arrays using high-density filters containing colony cDNA (or PCR products) followed by quantitative measurements of the amount of hybridized probe at each spot. A common platform used spotted cDNA arrays, where cDNA clones representing genes were robotically spotted on the support surface either as bacterial colonies or as PCR products. These “macroarrays”, or high-density filters, were made on nylon membranes measuring approximately 10 cm². Although this is now considered a dated approach, it was nonetheless effective enough to test sets of hundreds or even a few thousand genes.

DNA arrays allow the quantitative and simultaneous measurement of the mRNA expression levels of thousands of genes in a tissue or cell sample. The technology is based on the hybridization of a complex and heterogeneous RNA population derived from tissues or cells. Initially, this was referred as a “complex probe”, i.e., a complex mix that contains varying amounts of many different cDNA sequences, corresponding to the number of copies of the original mRNA species extracted from the sample. This complex probe was produced via the simultaneous reverse transcription and ³³P labeling of mRNAs, which were then hybridized to large sets of DNA fragments, representing the target genes, arrayed on a solid support. Thus, each individual experiment provided a very large amount of information (Gress et al. 1992, Nguyen et al. 1995; Jordan 1998; Velculescu et al. 1995; Zhao et al. 1995; Bernard et al. 1996, Pietu et al. 1996, Rocha et al. 1997).

1.1.2 Miniaturization, an Obvious Technological Evolution Towards Microarrays

One of the major challenges that researchers faced was to obtain the highest possible sensitivity when working with a limited amount of sample (biopsies, sorted cells, etc.). In this regard, five parameters were taken into account: 1) the amount of DNA fixed on the array support; 2) the concentration of RNA that should be labeled with the ³³P isotope; 3) the specific activity of the labeling; 4) the duration of the hybridization; and 5) the duration of exposure of the array to the phosphor imager shields.

The miniaturization of this method lay in the intrinsic physical characteristics of nylon membranes, which allowed a significant increase in the amount of immobilized DNA. The feasibility of miniaturizing nylon was demonstrated in the Konan Peck (Academia Sinica, Taiwan) laboratory in 1998 using a colorimetric method as the detection system (Chen et al. 1998). A combination of nylon microarrays and ^{33}P -labeled radioactive probes was subsequently shown to provide similar levels of sensitivity compared with the other systems available at the time, making it possible to perform expression profiling experiments using submicrogram amounts of unamplified total RNA extracted from small biological samples (Bertucci et al. 1999).

These observations had important implications for basic and clinical research in that they provided a cheaper alternative approach that was particularly suitable for groups operating in academic environments and led to a large numbers of expression profiling analyses when only small amounts of biological material were available.

Microarrays based on solid supports, typically coated glass, were simultaneously developed in different academic and industrial laboratories. These arrays boasted the advantage of performing dual hybridization of a test sample and a reference sample, as they could be labeled with two different fluorescent compounds, namely the fluorochrome “Cy-dyes” cyanine-3 (Cy3) and cyanine-5 (Cy5) (Chee et al. 1996).

Around the same time, another well known DNA array platform was developed by Affymetrix (Santa Clara, CA, USA). Their array used oligonucleotide chips featuring hundreds of thousands of oligonucleotides that were directly synthesized *in situ* on silicon chips (each measuring a few cm^2) using photochemical reactions and a masking technology (Lockhart et al. 1996). This microarray platform promised a rapid evolution in miniaturization because it was based on the synthesis of short nucleic acid sequences, which could be updated on the basis of the current knowledge of the genome.

It quickly became clear in the academic community, as well as in industry, that the available microarray technologies represented the beginning of a revolution with considerable potential for applications in the various fields of biology and health because gene function is one of the key elements that researchers want to extract from a DNA sequence. Microarrays have become a very useful tool for this type of research (Gershon 2002). Therefore, the development of the microarray opened the door to various DNA chip technologies based on the same basic concept. For example, the maskless photolithography used to produce oligonucleotide arrays was originally developed in 1999 using the light-directed synthesis of high-resolution oligonucleotide microarrays with a digital micromirror array to form virtual masks (Singh-Gasson et al. 1999). However, this technology was barely accessible to academic laboratories at the time because of the high initial cost, the limited availability of equipment, non-reusability, and the need for a large amount of starting RNA (Bertucci et al. 1999).

This development formed the basis for the NimbleGen company, which in 2002 demonstrated the chemical synthesis quality of maskless arrays synthesis (MAS) and its utility in constructing arrays for gene expression analysis (Nuwaysir et al.

2002). Currently, NimbleGen is focused on products for sequencing (<http://www.nimblegen.com/>).

Similarly, in 2005, Edwin Southern's team developed a method for the *in situ* synthesis of oligonucleotide probes on polydimethylsiloxane (PDMS) microchannels through the use of conventional phosphoramidite chemistry (Moorcroft et al. 2005). This became the basis of the Oxford Gene Technology company (<http://www.ogt.co.uk/>), which today develops array products centered on cytogenetics, molecular disorders and cancer.

It is also widely known that Affymetrix (<http://www.affymetrix.com/estore/>) and Agilent (<http://www.home.agilent.com/agilent/home.jsp?lc=eng&cc=US>) developed the most popular microarray technology for expression profiling based on ink jet technology, which is still widely available in the transcriptome market.

1.1.3 Reliable Microarray Results Depend on a Series of Complex Steps

The reliability of transcriptome results has concerned scientists since the beginning of transcriptome research, resulting in a number of studies comparing the different platforms, which was a real challenge in the early 2000s. Transcriptomic results largely depend on the technology used, which itself is dependent on several complex steps, ranging from the fabrication of the microarray to the experimental conditions, in addition to the chosen detection system, which also determines the method of analysis.

The results obtained with one microarray platform cannot necessarily be reproduced on another, and differences in the presence of different target sequences representing the same gene on different arrays can make it extremely difficult to integrate, combine and analyze the data (Järvinen et al. 2004).

The fabrication of high-quality microarrays has been a challenging task, taking a decade to reach several stabilized solutions, and has become an industry of its own. There are a large number of parameters and factors that affect the fabrication of a microarray, as performance depends on the array geometry, chemistry, and spot density, as well as on characteristics such as morphology, probe and hybridized density, background and sensitivity (Dufva 2005). Among the different methods used to fabricate DNA microarrays, *in situ* synthesis is the most powerful because a very high spot density can be achieved and because the probe sequence can be chosen for each synthesis.

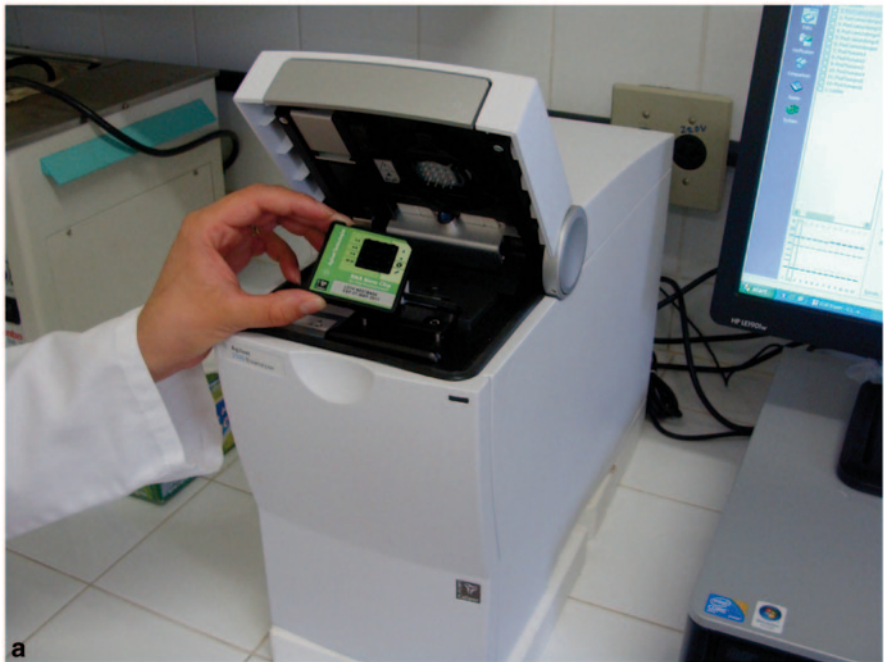
To achieve a 10^5 -fold dynamic range, which is an important parameter for gene expression analysis, the spots must contain at least 10^5 molecules, and the optimal spot size should be large enough to acquire the maximum hybridized density to obtain good sensitivity. Bead arrays that have different combinations of fluorescent dyes, which essentially constitute a barcode tag associated with the different immobilized probes, appeared to be the next evolution because they are in suspension and are therefore suitable for automation using standard equipment, leading

to extremely high-throughput approaches. Optical microarrays that are detected via flow cytometry can use a large number of different beads because each bead can be decoded using a series of hybridization reactions following the immobilization of the beads to the optical fibers (Ferguson et al. 2000; Epstein et al. 2003). This increases the multiplex capacity to several thousands of different beads (Gunder-son et al. 2004). Optical fiber microarrays have been commercialized by Illumina (<http://www.illumina.com/>), currently the leader in high-throughput sequencing technology, which allow the measurement of expression profiles by counting the amount of each RNA molecule expressed in a cell.

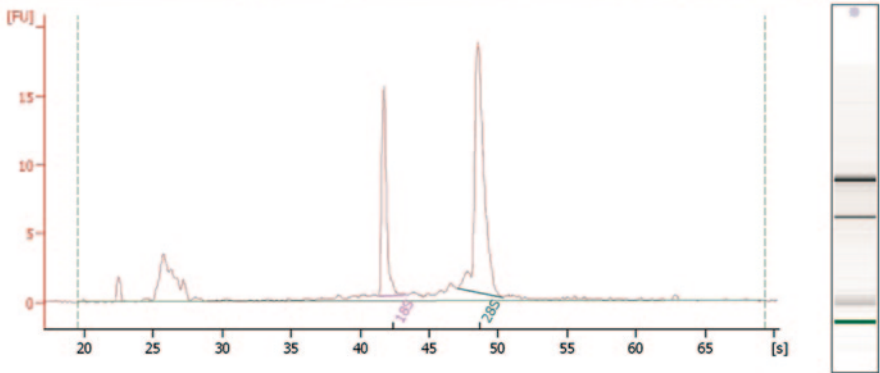
Experimental conditions also vary from lab to lab, as the preparation is dependent on the array platform. Variations in the quality of RNA preparations can be evaluated using the 2100 Bioanalyzer instrument developed by Agilent, which has become a standard, even if some slight variations have been observed from time to time. This system provides sizing, quantitation and quality control for RNA and DNA, as well as for proteins and cells, on a single platform, providing high-quality digital data (<http://www.genomics.agilent.com/en/Bioanalyzer-System/2100-Bioanalyzer-Instruments/?cid=AG-PT-106>) (Fig. 1.1).

The preparation of RNA prior to hybridization can affect microarray performance, particularly in terms of data accuracy, by distorting the quantitative measurement of transcript abundance. To obtain enough material from an initial nano- or picogram range of starting material, the RNA is transcribed *in vitro* and amplified using different protocols, which can introduce bias. In 2001, several publications discussed the different commercial protocols that were available. A publication from Charles Decreane's team examined the methods for amplifying picogram amounts of total RNA for whole genome profiling. The authors set up a specific experiment to compare three commercial RNA amplification protocols, Ambion messageAmp™, Arcturus RiboAmp™ and Epicentre Target Amp™, to the standard target labeling procedure proposed by Affymetrix, and all of the samples were tested on Affymetrix GeneChip microarrays (Clément-Ziza et al. 2009). The results obtained in this study indicated large variations between the different protocols, suggesting that the same amplification protocol should always be used to maximize the comparability of the results. Additionally, it was found that the RNA amplification affects the expression measurements as well, which was in agreement with earlier observations seen at the nanogram scale, as well as with other studies that were concerned with this question (Nygaard and Hovig 2006; Singh et al. 2005; Wang et al. 2003; Van Haaften et al. 2006; Degrelle et al. 2008).

In 2012, questions surrounding RNA amplification were still relevant. Indeed, even if the amplification of a small amount of RNA is reported to have a high reproducibility, there is still bias, and this can become time consuming. Even taking into account a correlation coefficient of 0.9 between microarray assays using non-amplified and qRT-PCR samples, the matter should still be reconsidered. In one study, the authors used the 3D-Gene™ microarray platform and compared samples prepared using either a conventional amplification method or a non-amplification protocol and a probe set selected from the MicroArray Quality Control (MAQC) project (<http://www.fda.gov/ScienceResearch/BioinformaticsTools/>



a



Overall Results for sample 1

RNA Area:	82.3	RNA Integrity Number (RIN):	9.7 (B.02.07)
RNA Concentration:	118 ng/μl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	2.0	Result Flagging Label:	RIN: 9.70

b

Fig. 1.1 Agilent Bioanalyzer model 2100 showing in **a** A RNA Nano Chip and in **b** A typical result of a microfluidic electrophoresis of a total human RNA sample extracted from leukocytes. On the *right* side of this figure appears a virtual gel with the respective bands of 28S and 18S rRNAs and 5S rRNA plus 4S tRNAs (from *top to bottom*). On the *left* side is shown the densitometry of this gel where appears the respective peaks of 28S rRNA, 18S rRNAs, 5S rRNA and 4S tRNAs. The rRNA ratio (28S/18S)=2.0 enabled a RNA integrity number (RIN=9.7), which indicated that this sample was intact (not degraded)

MicroarrayQualityControlProject/). They found that the samples from the non-amplification procedure had a higher quantitative accuracy than those from the amplification method but that the two methods exhibited comparable detection power and reproducibility (Sudo et al. 2012).

However in the above study, the researchers also used a few micrograms of RNA and a large volume of hybridization buffer. It is known that the ability to reduce the quantity of input RNA while maintaining the reaction concentration can be achieved in a device that decreases the hybridization reaction volume. Devices developed for use with beads have this characteristic; therefore, would hybridization using a bead device resolve this issue?

1.1.4 Bioinformatics and Standardization Approaches: A Possible Solution?

With regard to bioinformatics and standardization approaches, the MAQC project was initiated in 2006 to address these questions, as well as other performance and data analysis issues. The Microarray Quality Control (MAQC Consortium 2006) (<http://www.fda.gov/ScienceResearch/BioinformaticsTools/MicroarrayQualityControlProject/>) study tested a large number of laboratories, platforms and samples and found that there were notable differences in various dimensions of performance between microarray platforms. Each microarray platform has different trade-offs with respect to consistency, sensitivity, specificity and ratio compression. One interesting result was that platforms with divergent approaches for measuring expression often generated comparable results. The authors of this study concluded that the technical performance of microarrays supports their continued use for gene expression profiling in basic and applied research and may lead to the use of microarrays as a clinical diagnostic tool as well. This project has provided the microarray community with standards for data reporting, common analysis tools and useful controls that can help promote confidence in the consistency and reliability of these gene expression platforms (MAQC Consortium 2006). Similarly, in 2007, another meta-analysis of microarray results suggested several recommendations for standardization under the Standard Microarray Results Template (SMART) to facilitate the integration of microarray studies and proposed the implementation of the Minimum Information About a Microarray Experiment (MIAME) (<http://www.mged.org/Workgroups/MIAME/miame.html>) to facilitate the comparison of results (Cahan et al. 2007).

Given that measurement precision is critical in clinical applications, the question of the measurement precision in microarray experiments was addressed again in 2009 through an inter-laboratory protocol. In this study, the authors analyzed the results of three 2004 Expression Analysis Pilot Proficiency Test Collaborative studies using different methods. The study involved thirteen participants out of sixteen, each of whom provided triplicate microarray measurements for each of two reference RNA pools. To facilitate communication between the user and developer,

this study sought to set up standardized conceptual tools, but the result of this analysis was relatively disappointing and did not allow the creation of a gold standard, though it did put forth several recommendations (Duewer et al. 2009).

All of these studies focus on the same concept that has been defended since 2001 by the Microarray Gene Expression Data Society (<http://www.mged.org>) – the reanalysis and reproduction of results by the scientific community. The MGED society was the first to define the MIAME, which describes the minimum information required to ensure that microarray data can be easily interpreted and that the results derived from their analysis can be independently verified. This protocol became the standard for recording and reporting microarray-based gene expression data and for inserting it in databases and public repositories (Brazma et al. 2001, Ball et al. 2002). Currently, raw and/or normalized microarray data are deposited either in the ArrayExpress databank (<https://www.ebi.ac.uk/arrayexpress/>) or in the Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>), providing the scientific community with data for further analysis.

1.1.5 Analysis of the Expression Data

The past two decades have seen the development of methods that allow for a nearly complete analysis of the transcriptome, in the form of microarrays and, more recently, RNA-Seq, which are the most popular technologies used in genome-scale transcriptional studies. These high-throughput gene expression analysis systems generate large and complex datasets, and the development of computational methods to obtain biological information from the generated data has been the primary challenge in bioinformatics analysis.

Even a simple microarray experiment generates a large amount of data, which places certain demands on the analysis software. Fortunately, microarrays have benefited from the availability of many commercial and open-source software packages for data manipulation that have been developed over the years. RNA-Seq, however, demands more bioinformatics expertise. There are publicly available online tools such as the Galaxy platform (Goecks et al 2010, but a basic knowledge of UNIX shell programming and Perl/Python scripting is necessary for data modification. Furthermore, similar to microarray analysis, a familiarity with the R programming environment is useful, as the software programs for many of the downstream analyses are collected in the Bioconductor (<http://www.bioconductor.org/>) (Gentleman et al 2004) suite of the R package. Other important considerations regarding the choice for RNA-Seq include the need for data storage resources and computing systems with large memories and/or many cores to run parallel, sophisticated algorithms efficiently and faster.

In this section, we present the main steps for analyzing multi-dimensional genomic data derived from the application of microarray or RNA-Seq assays based on a common pipeline illustrated in Fig. 1.2.

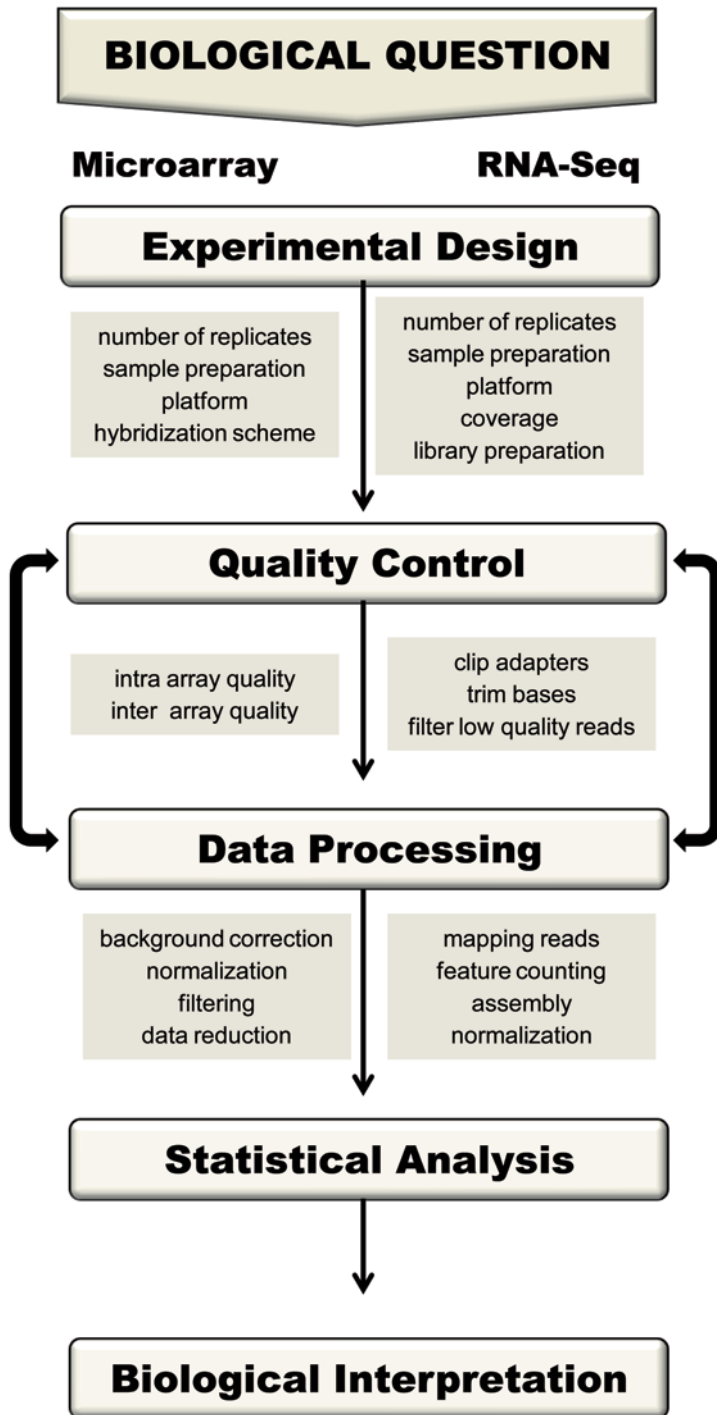


Fig. 1.2 An overview of the steps in a typical gene expression microarray or RNA-Seq experiment

1.1.5.1 Experimental Design

The aim of the experimental design is to make the experiment maximally informative given a certain amount of samples and resources and to ensure that the questions of interest can be answered. All of the decisions made at this initial step will affect the results of all the subsequent steps. The consequences of an incorrect or poor design range from a loss of statistical power and an increased number of false negatives to the inability to answer the primary scientific question (Stekel 2003).

The basic principles of experimental design rely on three fundamental aspects formalized by Fisher (1935), namely, replication, randomization and blocking.

Randomization dictates that the experimental subjects should be randomly assigned to the treatments or conditions to be studied to eliminate unknown factors that may potentially affect the results (Fang and Cui 2011).

Replication is essential for estimating and decreasing the experimental error and, thus, to detect the biological effect more precisely. A true replicate is an independent repetition of the same experimental process and an independent acquisition of the observations. There are different levels of replication in gene expression experiments: (1) a technical replicate provides measurement-level error estimates and (2) a biological replicate provides estimates of the population-level variability. If the goal is to evaluate the technology, technical replicates alone are sufficient. Otherwise, if the goal is to investigate the biological differences between tissues/conditions/treatments, biological replicates are essential (Alison et al 2006; Fang and Cui 2011). Replication is widely used in microarray experiments, though technical replicates are generally no longer performed, as analyses have shown that the results will be relatively consistent overall (Slonin and Yanai 2009). However, in RNA-Seq studies, replication is still neglected primarily due to the current high costs of these experiments. Studies conducted on the variability of this technology, both technical (Marioni et al. 2008) and biological (Bullard et al. 2010), underscore the importance of including replicates in the study design. The fundamental problem with generalizing the results gathered from unreplicated data is a complete lack of knowledge about the biological variation. Without an estimate of variability (i.e., within the treatment group), there is no basis for inference (i.e., between the treatment groups) (Auer and Doerge 2010).

As with microarray studies, RNA-Seq experiments can be affected by the variability coming from nuisance factors, often called technical effects, such as the processing date, technician, reagent batch and the hybridization/library preparation effect. In addition to these effects, in RNA-Seq experiments, there are also other technology-specific effects. For example, there is variation from one flow cell to another, resulting in a flow cell effect and variation between the individual lanes within a flow cell due to systematic variation in the sequencing cycling and/or base calling. A blocking design dictates comparisons within a block, which is a known uninteresting factor that causes variation, such as the hybridization scheme (microarray) or flow cell effect (RNA-Seq) (Fig. 1.3) (Alison et al. 2006, Slonin and Yanai 2009, Auer and Doerge 2010, Fang and Cui 2011, Luo et al 2010).

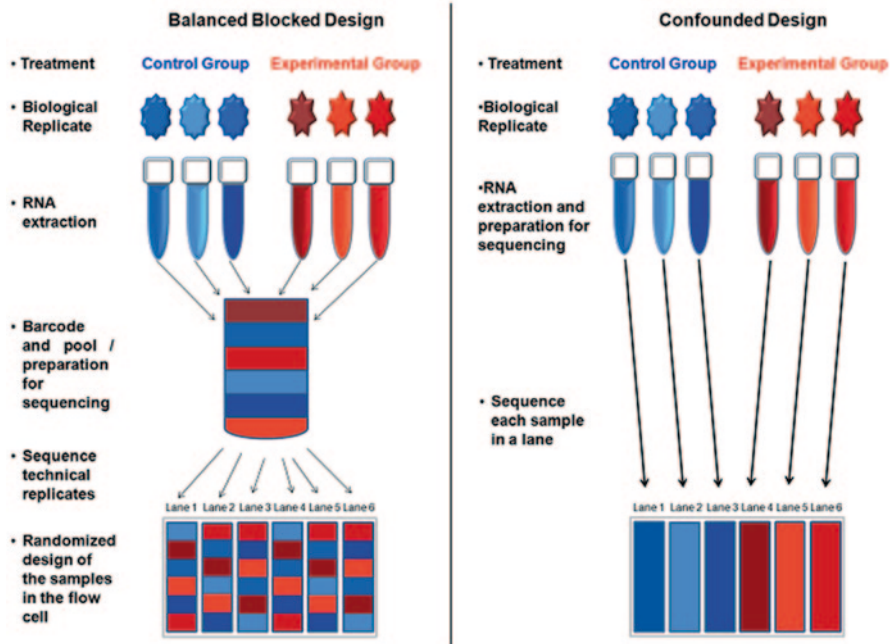


Fig. 1.3 Comparison of two methods for testing differential expression between treatments. **a** (red) and **b** (blue). In the ideal balanced block design (left), six samples are barcoded, pooled, and processed together. The pool is then divided into six equal portions that are input into six flow cell lanes. The confounded design (right) represents a typical RNA-Seq experiment and consists of the same six samples, with no barcoding, and does not permit batch and lane effects to be distinguished from the estimate of the intra-group biological variability (adapted from Auer and Doerge 2010)

In the case of microarray and RNA-Seq experiments, design issues are intrinsically dependent on hybridization and library construction, respectively. It is beyond the scope of this section to discuss and compare the different technologies available, but we recommend reading the following articles for microarray technologies: Paterson et al. (2006), Alison et al. (2006), Stekel (2003), Churchill (2002), Kerr and Churchill (2001), Jordan (2012). For RNA-Seq technologies, please see Auer and Doerge (2010) and Fang and Cui (2010), as well as chapter 2 of this book.

1.1.5.2 Quality Control

To assure the reproducibility, comparability and biological relevance of the gene expression data generated by high-throughput technologies, several research groups have provided guidelines regarding quality control (QC):

- **Minimum Information About a Microarray Experiment (MIAME):** describes the minimum information required to ensure that microarray data can be

easily interpreted and that the results derived from their analysis can be independently verified (Brazma et al. 2001).

- **External RNA Control Consortium (ERCC):** develops external RNA controls useful for evaluating the technical performance of gene expression assays performed by microarray and qRT-PCR (Baker et al. 2005).
- **MicroArray Quality Control (MAQC) Consortium:** a community-wide effort, spearheaded by the Food and Drug Administration (FDA), that seeks to experimentally address the key issues surrounding the reliability of DNA microarray data. Now in its third phase (MAQC-III), also known as Sequencing Quality Control (SEQC), the MAQC project aims to assess the technical performance of next-generation sequencing platforms by generating benchmark datasets using reference samples and evaluating the advantages and limitations of various bioinformatics strategies in RNA and DNA sequencing (Shi et al. 2006, Shi et al. 2010, (www.fda.gov/MicroArrayQC)).
- **Standards, Guidelines and Best Practices for RNA-Seq:** a guideline for conducting and reporting on functional genomics experiments performed with RNA-Seq. It focuses on the best practices for creating reference-quality transcriptome measurements (The ENCODE Consortium 2011) (<http://www.genome.gov/en-code>).

However, there are several sources of variability originating from biological and technical causes that can affect the quality of the resulting data, including biological heterogeneity in the population, sample collection, RNA quantity and quality, technical variation during sample processing, and batch effects, among others. Some of these issues can be avoided with an appropriate and carefully designed experiment that controls for the different sources of variation, but others require a quality assessment of the raw data through computational support tools. Therefore, regardless of the technology used to measure gene expression, ensuring quality control is a critical starting point for any subsequent analysis of the data (Churchill 2002, Geschwind and Gregg 2002, Cobb et al. 2005, Larkin et al. 2005, Irizarry et al. 2005, Heber and Sick 2006).

With regard to microarray technology, many tools applying diagnostic plots have been developed to visualize the spread of data and compare and contrast the probe intensity levels between the arrays of the dataset. These qualitative visualization plots include histograms, density plots, boxplots, scatter plots, MAplots, score plots of the PCA, hierarchical clustering dendrograms, and even chip pseudo plots and RNA degradation plots (Fig. 1.4). Comparing the probe intensity between samples allows us to observe if one or more of the arrays have intensity levels that are drastically different from the other arrays, which may indicate a problem with the arrays. For a better review of the use of diagnostic plots in quality control metrics, please see Gentleman et al. (2005) and Heber and Sick (2006).

In regard to RNA-Seq, several sequence artifacts are quite common, including read errors (base calling errors and small *indels*), poor quality reads and adaptor contamination. Such artifacts need to be removed before performing downstream analyses, otherwise they may lead to erroneous conclusions. Performing a quality

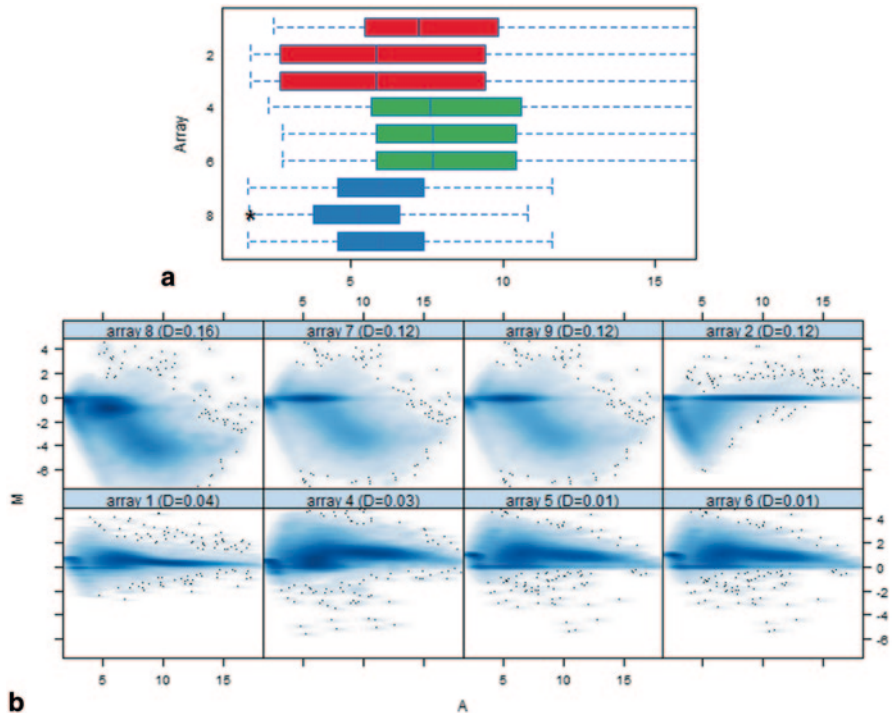


Fig. 1.4 Quality control plots of raw data sets. **a** Boxplots presenting various statistics for a given data set. The plots consist of boxes with a central line and two tails. The central line represents the median of the data, whereas the tails represent the upper (75th percentile) and lower (25th percentile) quartiles. These plots are often used to describe the range of log ratios that is associated with replicate spots. **b** *MA* plots are used to detect artifacts in the array that are intensity dependent

assessment of the reads allows us to determine the need for filtering (or cleaning) the data, removing low quality sequences, trimming bases, removing linkers, determining overrepresented sequences and identifying contamination or samples with a low sequence performance. The most important parameters used to verify the quality of the raw sequencing data are the base quality, the GC content distribution and the duplication rate (Guo et al. 2013, Patel and Jain 2012).

In addition to the QC pipelines provided commercially by the sequencing platform, there are online/standalone software packages and pipelines available as well (see: http://en.wikipedia.org/wiki/List_of_RNA-Seq_bioinformatics_tools). These packages present different features, and many are designed for a particular sequencing platform, such as NGS QC for the Illumina and Roche 454 platforms (Patel and Jain 2012) or Rolexa for Solexa sequencing data (Rougemont et al. 2009), or for a specific data storage format, such as FastQC toolkit and FastQScreen, which were both developed by the Brabraham Institute. The FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and FASTX-Tool kits (http://hannonlab.cshl.edu/fastx_toolkit/) include many of the tools used to remove indexes, barcodes and

adapters and filter out the reads based on the quality metrics of the FASTQ files. For a comparison of some of the available QC tools for RNA-Seq, please refer to Patel and Jain (2012).

1.1.5.3 Data Processing

Once the quality of the data has been assessed and the applicable changes have been made, it is still necessary to perform additional processing before analyzing the differentially expressed genes. The primary objective in processing raw data is to remove unwanted sources of variation, thereby ensuring the accuracy of the final results. There are several different methods to process the data being assayed, and the specific method used depends on how the data were generated.

According to Geleher et al. (2008), the data being assayed should be processed using several different methods, and the results should be compared to identify the most suitable method. The most appropriate method should then be used to process the raw data before the differential expression analysis.

Essentially, microarray processing involves three steps depending on the type of array: (1) background adjustment, which divides the measured hybridization intensities into a background and a signal component; (2) summarization, which combines the probe-level data into gene expression values, thereby reducing multiple probes representing a single transcript to a single measurement of expression; and (3) normalization, which aims to remove non-biological variations between arrays (Heber and Sick 2006). Other potential processing steps include transformation of the data from the raw intensities into log intensities and data filtering to remove flagged features, which are problematic features detected by the image-processing software (Stekel 2003, Allison et al. 2006).

Microarray data must also be background corrected to remove any signals arising from non-specific hybridization or spatial heterogeneity across the array. The background is a measure of the ambient signal obtained, generally, from the mean or median of the pixel intensity values surrounding each spot (Ritchie et al. 2007). The traditional correction is to subtract the local background measures from the foreground values, but the main problem with this procedure is that it can give negative corrected intensities, and there is high variability in the low-intensity log-ratios when the background is higher than the feature intensity (Stekel 2003). Instead, several different methods have been developed as alternatives. Some examples include the empirical Bayes model developed by Kooperberg et al. (2002), setting a small threshold value as suggested by Edwards (2003), the variance stabilization method (Vsn) of Huber et al. (2002), the *normexp* (normal-exponential convolution) method implemented by the RMA algorithm (Irizarry et al. 2003), and the MLE method (maximum likelihood estimation for *normexp*) (Silver et al. 2009). A detailed comparison of several of these methods can be found in the article by Ritchie et al. (2007).

The normalization of the microarray signal intensity has been widely used to adjust for experimental artifacts within the array and between all of the samples

such that meaningful biological comparisons can be made (Quackenbush 2001, Lou et al. 2010). According to Stekel (2003), the methods for normalization may be broadly classified into two categories:

1. Within-array normalization (normalizes the M-values for each array separately) – these methods are applicable for two-channel arrays, in which the aim is to adjust the Cy3 and Cy5 intensities to equal levels. Methods such as the linear regression of Cy5 against Cy3 and linear or non-linear (Loess) regression of the log ratio against the average intensity can correct for the different responses of the Cy3 and Cy5 channels. However, these methods rely on the assumption that the majority of the genes on the microarray are not differentially expressed. If this assumption is not true, a different normalization method, such as using a reference sample, would be more appropriate.
2. Between-array normalization (normalizes the intensities or log-ratios to be comparable across multiple arrays) – this method is used for one- and two-channel arrays. Various methods have been proposed for this approach, such as scaling to the mean or median, centering and quantiles. Bolstad et al. (2003) presented a review of several methods and found quantile normalization to be the most reliable method.

After processing, it is strongly recommended to verify the performance of the chosen method. This can be achieved by applying the aforementioned diagnostic plots during a Quality Control session. Several studies have been published on the performance of the various processing methods (Bolstad et al. 2003, Ploner et al. 2005), but most studies have found the Robust Multichip Average method (RMA) (Irizarry et al. 2003) to be the best method. This method applies a model-based background adjustment followed by quantile normalization and a robust summary method (median polish) on the log₂ intensities to obtain the probeset summary values.

The RNA-Seq data processing steps that were considered in our pipeline are as follows: (1) mapping reads; (2) transcriptome assembly; and (3) normalization of the read counts.

A common characteristic of all high-throughput sequencing technologies is the generation of relatively short reads, which should be mapped to a reference sequence, be it a reference genome or a transcriptome database. This is a critical task for most applications of the technology because the alignment algorithm must be able to efficiently find the right location for each read from among a potentially large quantity of reference data (Fonseca et al. 2012). The assembly of the transcriptome consists of the reconstruction of the full-length transcripts, except in the case of small classes of RNAs that are shorter than the sequencing length and require no assembly. The methods used to assemble reads fall into two main classes: (1) assembly based on a reference genome and (2) *de novo* assembly (Martin and Wang 2011). The strategies used to map the reads and assemble the transcriptome, along with the available tools, will be presented in more detail in chapter 2.

Normalization should always be applied to read counts due to two main sources of systematic variability: (1) RNA fragmentation during library construction causes the longer transcripts to generate more reads compared with the shorter transcripts

that are present at the same abundance in the sample, and (2) the variability in the number of reads produced for each run causes fluctuations in the number of fragments mapped across the samples. Proper normalization enables accurate comparison of the expression levels between and within samples (Garber et al. 2011, Dillies et al. 2013). The RPKM (reads per kilobase of transcript per million mapped reads) is the most widely used normalization metric. It normalizes a transcript read count by both its length and the total number of mapped reads in the sample (Mortazavi et al. 2008). This approach facilitates comparisons between genes within a sample and combines the inter- and intra-sample normalization. When data originate from paired-end sequencing, the FPKM (fragments per kilobase of transcript per million mapped reads) metric is used (Garber et al. 2011, Dillies et al. 2013).

In previous years, other methods for the normalization of RNA-Seq data have been proposed as well. These methods also applied inter-sample normalization using scaling factors and include the following: (1) Total count (TC), in which the gene counts are divided by the total number of mapped reads (or library size) associated with their lane and multiplied by the mean total count across all of the samples in the dataset; (2) Upper Quartile, which has a very similar principle to TC and in which the total counts are replaced by the upper quartile of counts different from 0 in the computation of the normalization factors; (3) Median, which is similar to TC, in which the total counts are replaced by the median counts different from 0 in the computation of the normalization factors; (4) DESeq, which is the normalization method included in the DESeq Bioconductor package (version 1.6.0) (<http://bioconductor.org/packages/release/bioc/html/DESeq.html>) and is based on the hypothesis that most genes are not differentially expressed; (5) Trimmed Mean of M -values (TMM), which is the normalization method implemented in the edgeR Bioconductor package (version 2.4.0) (<http://www.bioconductor.org/packages/release/bioc/html/edgeR.html>) and is also based on the hypothesis that most genes are not differentially expressed; and (6) Quantile, which was first proposed in the context of microarray data and consists of matching the distributions of the gene counts across lanes. These proposed normalization methods, in addition to the RPKM method, were comprehensively compared and evaluated by members of The French StatOmique Consortium. Based on this comparative study, the authors proposed practical recommendations for the appropriate normalization method to be used and its impact on the differential analysis of RNA-Seq data (Dillies et al. 2013).

1.1.5.4 Statistical Analysis and Interpretation

The primary goal of gene expression studies is to identify genes that are differentially expressed between RNA samples from two types of biological conditions. Differential gene expression can provide insights into biological mechanisms or pathways and form the basis for further experiments by determining the sample and gene similarity via clustering analyses or testing a gene set for enrichment.

Differential expression analysis searches for genes whose abundance has changed significantly across the experimental conditions. In general, this means taking the quantified and normalized expression values for each library and performing statistical testing between samples of interest. In theory, the transcript abundance of the mRNA would be directly proportional to the number of reads, thereby determining the expression level (Oshlack et al. 2010).

Many methods have been developed for the analysis of differential expression using microarray data. In the early days of microarrays, only the simple fold-change method was used (Chen et al. 1997). However, the evolution of the technology called for more accurate analytical methods, and many more sophisticated statistical methods have been proposed.

In addition to the traditional t-test and ANOVA approaches used to access differential gene expression in microarray assays, variations on these tests have been developed for the purpose of overcoming the problem of a small sample size when accessing such a large dataset: dealing with many genes but only a few replicates may lead to large fold-changes driven by outliers, as well as to small error variances (Lönstedt and Speed 2002). SAM (Significant Analysis of Microarrays) (Tusher et al. 2001) is a very popular differential expression method that uses a modified t-statistic to identify significant genes using non-parametric statistics.

Other statistical approaches for microarray data analysis have introduced linear models. The Bioconductor package Limma, developed by Smyth (2005), applies a gene-wise linear model that allows for the analysis of complex experiments (comparing many RNA samples), as well as more simple replicated experiments using only two RNA samples. Empirical Bayes and other shrinkage methods are used to borrow information across genes, making the analyses stable even for experiments with small numbers of arrays. Another powerful method to detect differentially expressed genes in microarray experiments is based on calculating the rank products (RP) from replicate experiments, while at the same time providing a straightforward and statistically stringent way to determine the significance level for each gene and allow flexible control of the false-detection rate and familywise error rate in the multiple testing situation of a microarray experiment (Breitling and Herzyk 2005).

Differential expression analysis methods that use probability distributions have also been proposed for use in modeling the count data from RNA-Seq studies, including Poisson and negative binomial (NB) distributions. The Poisson distribution forms the basis for modeling RNA-Seq counts. However, when there are biological replicates, the RNA-Seq data may exhibit more variability than expected by the Poisson distribution because it assumes that the variance is equal to the mean. If this occurs, the Poisson distribution will predict a smaller variation than that observed in the data, and the analysis will be prone to high false-positive rates that result from an underestimation of the sampling error (Anders and Huber 2010). Therefore, the NB model is the better method to address this so-called overdispersed problem because an NB distribution specifies that the variance is greater than the mean (Oshlack et al. 2010, Anders and Huber 2010, Garber et al. 2011).

Statistical analyses of RNA-Seq data will be discussed in more detail in chapter 2. There are also several reviews that discuss and compare the statistical methods

used to compute differential expression. For further information, please refer to Seyednasrollah et al. (2013) and Sonesson and Delorenzi (2013).

1.1.5.5 Classification and Enrichment Analysis

Classification can be performed either before or after the differential expression analysis. This process entails either placing the objects (in this case, the samples, genes or both) into pre-existing categories (known as a supervised classification) or developing a set of categories into which the objects can subsequently be placed (unsupervised classification) (Allison et al. 2006). Class discovery, or clustering analysis, is an unsupervised classification method that is widely used in the study of transcriptomic data because it allows us to identify co-regulated genes and/or samples with similar patterns of expression (biological classes). Various clustering techniques have been applied to identify patterns in gene-expression data. Most cluster analysis techniques are hierarchical: the resultant classification has an increasing number of nested classes, and the result resembles a phylogenetic classification. Non-hierarchical clustering techniques also exist, such as *k*-means clustering, which simply partition objects into different clusters without trying to specify the relationship between the individual elements (Quackenbush 2001). Eisen et al. (1998) is a classical reference for the use of hierarchical clustering with microarray data. In this study, the authors developed an integrated pair of open-source programs, Cluster and TreeView, for analyzing and visualizing clusters and heat maps (<http://rana.lbl.gov/EisenSoftware.htm>).

Biological insights into an experimental system can be gained by looking at the expression changes of sets of genes. Many tools focusing on gene set testing, network inference and knowledge databases have been designed for analyzing lists of differentially expressed genes from microarray datasets. Examples include Gene Set Enrichment Analysis (<http://www.broadinstitute.org/gsea/index.jsp>) (Subramanian et al. 2005) and DAVID (<http://david.abcc.ncifcrf.gov/tools.jsp>) (Dennis et al. 2003), which combine functional themes, such as those defined by the Gene Ontology consortium, (Ashburner et al. 2000), and metabolic and signaling pathways, such as KEGG pathways (<http://www.genome.jp/kegg/pathway.html>) (Kanehisa and Goto 2000) and Biocarta (<http://www.biocarta.com/>), with statistical enrichment analyses to determine whether specific pathways are overrepresented in a given list of differentially expressed genes. These approaches can also be applied to RNA-Seq, but the biases presented by this type of data should be taken into account (Oshlack et al. 2010). Therefore, specialized approaches (Bullard et al. 2010) and tools to perform enrichment analyses of RNA-Seq data are being developed, for example, GO-seq (<http://www.bioconductor.org/packages/release/bioc/html/go-seq.html>) (Young et al. 2010), SeqGSA (<http://www.bioconductor.org/packages/release/bioc/html/SeqGSEA.html>) (Wang and Cairns 2013) and generally applicable gene set enrichment for pathway analysis (GAGE) (Luo et al. 2009).

1.2 The Diversity of the Transcriptome

Unlike the genome, which is essentially static in terms of its composition and size (barring the rare occurrence of somatic and germline mutations or the rearrangement of immunoglobulin and T cell receptor genes), the transcriptome (and similarly, the miRNome) is extremely variable and depends on the phase of the cell cycle, the organ, exposure to drugs or physical agents, aging, diseases such as cancer and autoimmune diseases and a multitude of other variables, which must be considered at the time that the transcriptome is determined. This variability arises from the fact that RNAs are differentially transcribed (or transcribed at different rates) depending on the cell type and status, though this excludes ribosomal RNAs, as they are considered housekeeping molecules.

For many years, the central dogma of molecular biology stated that RNAs molecules were intermediates between DNA and protein. This idea presupposed that the function of RNA was primarily linked to the translation of the genetic material into polypeptide chains (proteins). The genetic material was interpreted as being involved in the synthesis of these RNAs, which were termed mRNAs (Brenner et al. 1961; Jacob and Monod 1961).

During the human genome sequencing era of the 1980s and 1990s, independently led by Francis Collins and Craig Venter, the latter individual and his coworkers conceived of expressed sequence tags (ESTs), which focus on mRNAs because they encode proteins. Libraries of mRNA-derived cDNA clones were generated based on first-strand synthesis using oligonucleotide primers for that are anchored at the 3' end of the transcript [the poly(A) tail of mRNA] (Starusberg and Riggins 2001) and then sequenced to create unique identifiers for each cDNA, with lengths ranging from 300 to 700 bp (Adams et al. 1992; Adams 2008).

ESTs were very useful for identifying new expressed genes in normal and diseased tissues (Strausberg and Riggins 2001), and transcriptome analysis at this time was largely, if not solely, based in this approach. The EST clones were distributed through the former IMAGE Consortium, whose sequences can now be retrieved via the National Center for Biotechnology Information (NCBI) dbEST Database (<http://www.ncbi.nlm.nih.gov/dbEST/>). The current number of public entries for all uni- or multicellular eukaryotic organisms that have been sequenced stands at more than 74 million ESTs, including more than eight million human and nearly five million mouse ESTs.

However, as was to be expected, imaginative new strategies were emerging around the same time as well. The Serial Analysis of Gene Expression (SAGE) method (Velculescu et al. 1995), which produces short sequence tags (usually 14 nucleotides in length) positioned contiguous to defined restriction sites near the 3' end of the cDNA strand (Strausberg and Riggins 2001), has also been widely used. At the time, the NCBI created the SAGEmap as a public repository for SAGE sequences. Currently, all of the SAGE libraries have been uploaded and accessioned through the Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) repository.

Another novel strategy, which had yet to be tested at that time, was the generation of open reading frame (ORF) ESTs (ORESTES). This approach was jointly developed by researchers funded by the São Paulo Research Foundation (FAPESP) and by the Ludwig Institute for Cancer Research (FAPESP/LICR)-Human Cancer Genome Project (Camargo et al. 2001). Unlike ESTs, ORESTES sequences are spaced throughout the mRNA transcript, providing a scaffold to complete the full-length transcript sequences. The authors generated a substantial volume of tags (700,000 ORESTES), which at the time represented nearly 20% of all human dbESTs (Strausberg and Riggins 2001).

The Transcript Finishing Initiative, another FAPESP/LICR project, was then undertaken for the purpose of identifying and characterizing novel human transcripts (Sogayar et al. 2004). This strategy was also novel and was based on selected EST clusters that were used for experimental validation. In this method, RT-PCR was used to fill in the gaps between paired EST clusters that were then mapped on the genome. The authors generated nearly 60,000 bp of transcribed sequences, organized into 432 exons, and ultimately defined the structure of 211 human mRNA transcripts.

However, the increasing use of modern transcriptome-wide profiling approaches, such as microarrays and whole-genome and transcriptome sequencing, allied to the precise isolation and characterization of different RNA species from eukaryotic (including mammalian) cells, led to an explosion of findings and revealed that although approximately 90% of the mammalian genome is actively transcribed into RNA molecules, only a tiny fraction (—2% of the total human genome) encodes mRNAs and, consequently, proteins (Maeda et al. 2006; Djebali et al. 2012).

In fact, the function of the genome can be seen from two different but complementary views. From a functional standpoint, only a fraction of the genome encodes RNA molecules (including coding and non-coding RNAs), and only a fraction of these are translated into proteins. In other words, when considering the genome in numerical terms, or rather the physical portion of DNA that is functional, we realize that only a small number of genes are transcribed specifically into mRNA molecules. However, a larger number of “variable” mRNA molecules are generated through alternative splicing, and these are translated into a greater number of proteins (including various isoforms). A large portion of the genome is then transcribed into non-coding RNAs, which play a role in the posttranscriptional control of mRNAs during their translation into proteins (Fig. 1.5).

Molecular mapping of the human genome has been largely resolved, revealing slightly more than three billion bp encompassing approximately 20–25,000 functional nuclear genes and mitochondrial DNA located in the cytoplasm. We suggest consulting the ENCODE Project (<http://www.genome.gov/encode/>) to follow ongoing progress in the identification of the functional elements in the human genome sequence. Nevertheless, the definition of the human transcriptome is still far from set, and it appears that most of the RNA molecules in eukaryotic cells are composed of ncRNAs that are involved in the fine control of gene expression.

Aside from knowing the exact number of mRNA molecules in a human cell, which is currently being investigated using new sequencing technologies (de Klerk

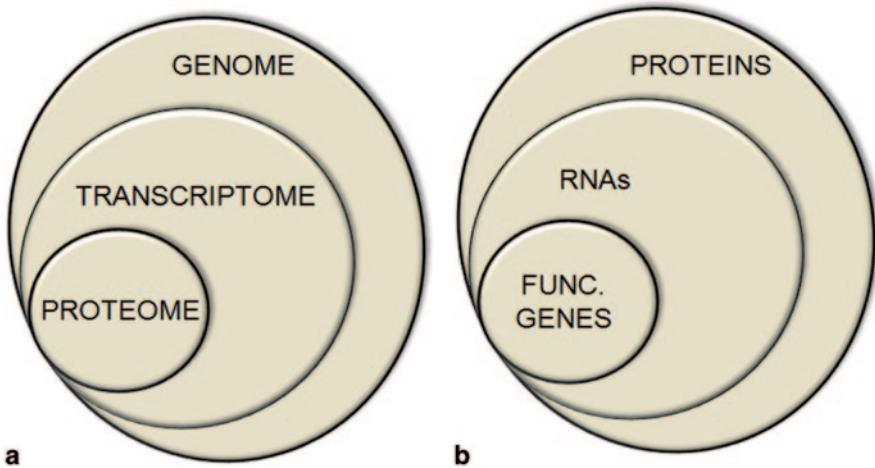


Fig. 1.5 Two ways to interpret the functioning the genome and the relative proportions of molecular entities. **a** In functional terms only a part of the genome encodes RNAs from which only a small fraction encodes proteins. **b** However, in numerical terms the set of functional genes transcribe a larger number of mRNAs from which a larger number of proteins is translated. The part A of this figure was conceived by Dr. Sven Diederichs (German Cancer Research Institute, DKFZ, Heidelberg, Germany) who allowed their use

et al. 2014; Kellis et al. 2014), one of the great challenges of the next decade will be to decipher the posttranscriptional interactions between coding and ncRNAs in the control of gene expression.

In fact, the human genome was revealed to be more than just a collection of protein-coding genes and their splice variants, rather, it displays extensive antisense, overlapping and ncRNA expression (Taft et al. 2010).

In mammals, the vast majority of the genome is transcribed into ncRNAs, which exceed the number of protein-coding genes (Liu and Taft 2013). These molecules are characterized by the absence of protein-coding capacity, but these RNAs have been described as key regulators of gene expression (Geisler and Collier 2013).

ncRNAs are grouped into two major classes based on their transcript size: small ncRNAs (19–30 nt) and long non-coding RNAs (200 nt to ~100 kilobases). These groups are distinct in their biological functions and mechanisms of gene regulation (Geisler and Collier 2013; Fatica and Bozzoni 2014; Neguembor et al. 2014).

Furthermore, ncRNAs can be grouped into a third class of housekeeping ncRNAs, which are normally constitutively expressed and include ribosomal (rRNAs), transfer (tRNAs), small nuclear (snRNAs), small nucleolar (snoRNAs) and regulatory noncoding RNAs (rnRNAs) (Ponting et al. 2009; Bratkovic and Rogelj 2014).

Small ncRNAs are primarily associated with the 5' or 3' regions of protein-coding genes, and based on their precursors and mechanism of action, they have been divided into three main classes: miRNAs, small interfering RNAs (siRNAs) and piwi-associated RNAs (piRNAs). These ncRNAs are involved in posttranscriptional gene regulation through translational repression or RNAi (Sana et al. 2012).

Interestingly, the aberrant expression of small ncRNAs has been associated with a wide variety of human diseases, including cancer, central nervous system disorders, and cardiovascular diseases (Taft et al. 2010; Sana et al. 2012) (Table 1.1).

For much of the last decade, special attention has been paid to research into long non-coding RNAs (lncRNAs), as these molecules tend to be shorter and have fewer introns than protein-coding transcripts (Ravasi et al. 2006). lncRNAs are considered to be the most numerous and functionally diverse class of RNAs (Derrien et al. 2011). Over 15,000 lncRNAs have already been identified, and this number is constantly increasing (Derrien et al. 2012; Fatica and Bozzoni 2014).

Amidst the great discoveries being made during this time of genome exploration, RNA is beginning to take center stage, and lncRNAs are a major part of this. These molecules are more abundant and functional than previously imagined, and they have been shown to be key players in gene regulation, genome stability, and chromatin modifications. Therefore, the identification and characterization of the function of lncRNAs has added a high degree of complexity to the comprehension of the structure, function and evolution of our genome.

lncRNAs can be grouped into one or more of five categories based on their position relative to protein-coding genes: (1) sense or (2) antisense, when they overlap with one or more exons of another transcript on the same or opposite strand, respectively; (3) bidirectional, when the expression of a lncRNA and a neighboring coding transcript on the opposite strand is initiated in close genomic proximity; (4) intronic, when the lncRNA is fully derived from the intron of a second transcript; or (5) intergenic, wherein a lncRNA is located within a gene (Potting et al. 2009). Most lncRNAs are transcribed by RNA Pol II and are often polyadenylated and have splice sites (Guttman et al. 2009; Mercer et al. 2013). However, they are devoid of obvious ORFs (Fatica and Bozzoni 2014).

The functional characterization of several mammalian regulatory lncRNAs has identified many biological roles, such as dosage compensation, genomic imprinting, cell cycle regulation, pluripotency, retrotransposon silencing, meiotic entry and telomerase length, and gene expression through chromatin modulation (Wery et al. 2011; Wilusz et al. 2009; Nagano and Fraser 2011).

The number of lncRNAs with described functions is steadily increasing, and many of these reports revolve around the regulatory capacity of lncRNAs. These molecules localize both to the nucleus and to the cytosol and can act at virtually every level during gene expression (Batista and Chang 2013; Van et al. 2014). Nuclear lncRNAs act as modulators of protein-coding gene expression and can be subdivided into *cis*-acting RNAs, which act in proximity to their site of transcription, or *trans*-acting lncRNAs, which work at distant loci. Both *cis*- and *trans*-acting lncRNAs can activate or repress transcription via chromatin modulation (Penny et al. 1996; Pandey et al. 2008; Nagano et al. 2008; Chu et al. 2011; Plath et al. 2003; Bertani et al. 2011).

Cytoplasmic lncRNAs can modulate translational control via sequences that are complementary to transcripts that originate from either the same chromosomal locus or independent loci. Target recognition occurs through base pairing (Batista and Chang 2013).

Table 1.1 Main RNA species found in eukaryotic cells including human

	Class	Symbol	Characteristic
Classical RNAs	Messenger RNAs	mRNAs	Variable in the size (average size about 2.2 kb) depending on the coded protein. Its linear structure includes a 5' G cap, the 5'UTR, AUG start codon, coding sequence (CDS), stop codon, 3'UTR and the poly A tail. Account 1–2% of the total cellular RNA.
	Transfer RNAs	tRNAs	This class of RNAs takes the form of “clover leaf” and has variable size ranging 70-100 nt. The residues 34, 35 and 36 are complementary to the mRNA codons located at CDS. For this reason they are considered as adaptors between mRNAs and elongation peptide chains.
	Ribosomal RNAs	rRNAs	Are components of ribosomes along with ribosomal proteins. The high molecular weight rRNAs are the 28S rRNA (~5000 nt), which is present in the large subunit of the ribosome and 18S rRNA (1870 nt) present in the small ribosomal subunit. The low molecular weight rRNAs are the 5.8 rRNA (156 nt) and the 5.0S rRNA (121 nt), both present in the large ribosomal subunit.
Small non-coding RNAs	MicroRNAs	miRNAs	Average size about 18–25 nt; account 1–2% of the human genome; control the 50% of protein-coding genes; guide suppression of translation; Drosha and Dicer dependent small ncRNAs.
	Small interfering RNAs	siRNAs	Average size about 19–23 nt; made by Dicer processing; guide sequence specific degradation of target mRNA.
	Piwi-interacting RNAs	piRNAs	Average size about 26–30 nt; bind Piwi proteins; Dicer independent; exist in genome clusters; principally restricted to the germline and somatic cells bordering the germline.
	Small nucleolar RNAs	snoRNAs	Average size about 60–300 nt; enriched in the nucleolus; in vertebrate are excised from premRNA introns; bind snoRNP proteins.

Table 1.1 (continued)

	Class	Symbol	Characteristic
	Promoter-associated small RNAs	PASRs	Average size about 20–200 nt; modified 5' (capped) ends; coincide with the transcriptional start sites of protein- and noncoding genes; made from transcription of short capped
	Transcription initiation RNAs	tiRNAs	Average size about 18 nt ; have the highest density just downstream of transcriptional start sites; show patterns of positional conservation; preferentially located in GC-rich promoters.
	Centromere repeat associated small interacting	crasiRNAs	Average size about 34–42 nt; processed from long dsRNAs.
	Telomere-specific small RNAs	tel-sRNAs	Average size about 24 nt; Dicer independent; 2'-O-methylated at the 3' terminus; evolutionarily conserved from protozoa to mammals; but have not been described in human up to now.
	Pyknons		Subset of patterns of variable length; form mosaics in untranslated and protein-coding regions; more frequently in 3' UTR.
Long non-coding RNAs	Long intergenic Non-coding RNAs	lincRNAs	Ranging from several hundreds to tens of thousands nts; lie within the genomic intervals between two genes; transcriptional cisregulation of neighbouring genes.
	Long intronic non-coding RNAs		Lie within the introns; evolutionary conserved; tissue and subcellular expression specified
	Telomere-associated ncRNAs	TERRAs	Average size about 100 bp to > 9 kb; conserved among eukaryotes; synthesized from C-rich strand; polyadenylated; form intermolecular G-quadruplex structure with single-stranded telomeric DNA.
	Long non-coding RNAs with dual functions		Both protein-coding and functionally regulatory RNA capacity.

Table 1.1 (continued)

	Class	Symbol	Characteristic
	Pseudogene RNAs		Gene copies that have lost the ability to code for a protein; potential to regulate their protein coding cousin; made through retrotransposition; tissue specific.
	Transcribed-ultraconserved regions	T-UCRs	Longer than 200 bp; absolutely conserved between orthologous regions of human, rat, and mouse; located in both intra- and intergenic regions.
	Circular RNAs	circRNAs	Noncoding RNAs generated during splicing through exon or intron circularization. They are transcription regulators or play their role as sponges for miRNAs.

Part of this table was adapted from Sana et al. (2012) *J Transl Med* 10: 103

RNA-Seq, the most powerful methodology for *de novo* sequence discovery, has been used to identify and analyze the expression of new lncRNAs in different cell types and tissues. Interestingly, sequencing experiments have shown that lncRNA expression is more cell-type specific than that of protein-coding genes (Riin and Chang 2012; Derrien et al. 2012; Guttman et al. 2012; Mercer et al. 2008; Cabili et al. 2011; Pauli et al. 2012).

The identification of lncRNAs relies on the detection of transcription from genomic regions that are not annotated as protein coding. However, other similarly robust methodologies have been used in the identification of lncRNAs, including the following: (1) Tiling arrays: this technology enables the analysis of global transcription from a specific genomic region and were initially used to both identify and analyze the expression of lncRNAs; (2) Serial analysis of gene expression (SAGE): this methodology allows both the quantification and the identification of new transcripts throughout the transcriptome; (3) Cap analysis gene expression (CAGE): this methodology is based on the isolation and sequencing of short cDNA sequence tags that originate from the 5' end of RNA transcripts; (4) Chromatin immunoprecipitation (ChIP): this method allows the isolation of DNA sequences that are associated with a chromatin component of interest, thereby allowing the indirect identification of many unknown lncRNAs; and (5) RNA-Seq: in a single sequencing run, this methodology produces billions of reads that are subsequently aligned to a reference genome (Fatica and Bozzoni 2014).

Transcriptome research began in parallel with the genome project because of Craig Venter's idea to sequence the "most important" genes, i.e., the functioning genome. This directive clearly fell upon mRNAs, as this type of RNA carries the protein code. Of course, this concept has not changed and mRNAs are still of central importance; however, what followed was the subsequent discovery of a large number of different ncRNAs whose functions are linked to the fine control of gene expression, often controlling the translation of mRNAs into proteins, i.e.,

posttranscriptional control as it is exerted by miRNAs. In its broadest sense, the transcriptome is undoubtedly more complex than anyone previously imagined.

1.3 The Transcriptome and miRNome are Closely Associated: The Role of MicroRNAs, a Class of Non-Coding Rnas Linked to the Fine Control of Gene Expression

Cellular gene expression is governed by a complex, multi-faceted network of regulatory interactions. In a very unique way, RNA molecules hybridize to each other. In the last decade, miRNAs have emerged as critical components of this cross-hybridization network. The miRNome was found to physically interact with the transcriptome, and this has important consequences for biological function.

The miRNA class of ncRNAs was first discovered in the worm *Caenorhabditis elegans* (Lee and Ambros 1993; Wightman and Ruykun 1993) and represents a family of small ncRNAs that posttranscriptionally regulate the stability of mRNA transcripts or their translation into proteins.

miRNAs participate in the regulation of a wide variety of biological processes, including cell differentiation and growth, development, metabolism chromosome architecture, apoptosis, and stress resistance. They are also involved in the pathogenesis of diseases as diverse as cancer and inflammation as well (Ambros 2004; Bushati and Cohen 2007; Stefani and Slack 2008). miRNAs are also promising candidates for new targeted therapeutic approaches and as biomarkers of disease. At approximately 22 nucleotides long, miRNAs are among the shortest known functional eukaryotic RNAs, and they repress most of the genes they regulate by just a small amount.

Many miRNAs are found in clusters and are transcribed from independent genes by either RNA Pol II or RNA Pol III (Chen et al. 2004; Borchert et al. 2006; Winter et al. 2009). They are normally found in three genomic locations: in the introns of protein-coding genes, in the introns of non-coding genes and in the exons of non-coding genes (Kim et al. 2006; Lin et al. 2008). Most miRNAs are derived from longer, double-stranded RNAs, which are termed primary miRNAs (pri-miRNAs).

Within these primary transcripts, miRNAs form stem-loop structures that contain the mature miRNA as part of an imperfectly paired double-stranded stem connected by a short terminal loop. pri-miRNAs are initially modified with a 5' 7-methylguanosine cap and a 3' poly-A tail (Cullen 2004) and contain hairpins that are further excised by the nuclear RNase III Droscha and its dsRNA-binding partner DGCR8 (DiGeorge syndrome critical region gene 8) (Gregory et al. 2004; Denli et al. 2004, Landthaler et al. 2004). The resulting pre-miRNA consists of an approximately 70-nucleotide double-stranded hairpin characterized by imperfect base-pairing in the stem-loop and a 2-nucleotide overhang at the 3' end (Lee et al. 2003).

The stem-loop of a pre-miRNA is recognized by the nuclear transport protein exportin-5, which exports the pre-miRNA to the cytoplasm, in combination with

the guanosine triphosphate (GTP) binding RAS-related nuclear protein (Ran-GTP) (Yi et al. 2003; Bohnack et al. 2004; Lund et al. 2004). In the cytoplasm, the pre-miRNAs are then cleaved by the RNase III enzyme Dicer and the double-stranded RNA-binding protein TRBP (TAR RNA-binding protein) into duplexes of miRNA and passenger strands of approximately 22 base pairs (Hutvagner et al. 2001; Zhang et al. 2002).

After the sequential processing of the miRNA precursors, one of the two strands of the miRNA duplex is incorporated into the RNA-induced silencing complex (RISC). This complex comprises the mature miRNA strand as well as several proteins from the Argonaute and Gw182 families (Chendrimada et al. 2005; Haase et al. 2005). RISC can then find and bind to complementary mRNA sequences and perform its silencing function (Kawamata and Tomari 2010, Czech and Hannon 2011). In addition, a few miRNAs are produced by alternative pathways, independent of Droscha and/or Dicer, by exploiting diverse RNases that normally catalyze the maturation of other types of transcripts (Yang and Lai 2011).

Although miRNAs typically function in the cytoplasm, there is increasing evidence that they can play important roles in the nucleus as well (McCarthy 2008; Politz et al. 2009). They can also be found in the mitochondria, where they may be involved in the regulation of apoptotic genes (Kren et al. 2009).

The regulatory roles of miRNAs have been the subject of intense research (Shimoni et al. 2007; Wang and Raghavachari 2011; Levine et al. 2007; Levine and Hwa 2008; Mehta et al. 2008; Osella et al. 2011; Mitarai et al. 2009; Bumgarner et al. 2009; Iliopoulos et al. 2009). In mammals, the majority of miRNAs are inferred to be functional on the basis of their evolutionary conservation.

The major determinant for recognition between an miRNA and a target mRNA is a region of high sequence complementarity that consists of an approximately 7-nucleotide domain at the 5' end of the miRNA known as the "seed" sequence (Bartel 2009). The remaining nucleotides are generally only partially complementary to the target sequence. Sequences that are complementary to the seed ("seed matches") trigger a modest but detectable decrease in the expression of an mRNA. Seed matches can occur in any region of an mRNA but are more likely to decrease mRNA expression when they are located in the 3' untranslated region (3' UTR) (Grimson et al. 2007; Forman et al. 2008, 2010; Gu et al. 2009) (Fig. 1.6). Because the region used to create the seed is so short, more than half of the protein-coding genes in mammals are regulated by miRNAs, and thousands of other mRNAs appear to have

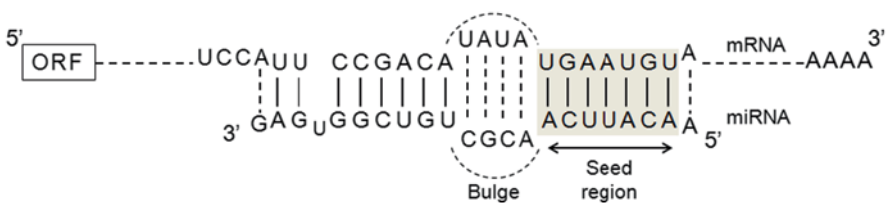


Fig. 1.6 Interaction of a miRNA with the 3'UTR of its mRNA target by base pairing. (Figure adapted from Filipowicz et al (2008) *Nat Rev Genetics* 9: 102–114)

undergone negative selection to avoid seed matches with miRNAs that are present in the same cell (Baek et al. 2008; Lewis et al. 2003, 2005; Farh et al. 2005, Stark 2005; Lewis 2005).

Despite the aforementioned basic features, a “seed” sequence is neither necessary nor sufficient for target silencing. It has been shown that miRNA target sites can often tolerate G:U wobble base pairs within the seed region (Miranda et al. 2006; Vella et al. 2004), and extensive base pairing at the 3' end of the miRNA may offset the absence of complementarity in the seed region (Brennecke et al. 2005; Reinhart et al. 2000). Moreover, centered sites showing 11–12 contiguous nucleotide base pairing with the central region of the miRNA without pairing to either end have also been reported (Shin et al. 2010). Adding to this repertoire, other studies have reported efficient silencing from sites that do not fit any of the above patterns and appear to be seemingly random (Lal et al. 2009; Tay et al. 2008), and even sites with extensive 5' complementarity can be inactive when tested in reporter constructs (Didiano et al. 2006).

How miRNAs repress or activate gene expression in animals is another important question, in addition to the high number of high-quality studies examining the biochemistry, biology and genomics of miRNA-directed mRNA regulation. The factors that determine which mRNAs will be targeted by miRNAs, or the mechanism by which they will be silenced, remain unclear. Extensive computational and experimental research over the last decade has substantially improved our understanding of the mechanisms underlying miRNA-mediated gene regulation (Ameres and Zamore 2013; Yue et al. 2009; Ripoli et al. 2010; Bartel 2009, Chekulaeva et al. 2009, Brodersen and Voinnet 2009).

miRNAs posttranscriptionally control gene expression by regulating mRNA translation or stability (Valencia-Sanchez et al. 2006, Standart et al. 2007; Jackson 2007, Nilsen 2007). What is known is that miRNAs can interfere with the initiation or elongation of translation; alternatively, the target mRNA may be affected by isolating it from the ribosomal machinery (Nottrott et al. 2006; Pillai et al. 2007). The binding of eIF4E to the cap region of an mRNA marks the initiation of initiation complex assembly. It has been demonstrated that miRNAs interfere with eIF4E and impair its function, and the function of the poly(A) tail can also be inhibited (Humphreys et al. 2005). There is additional evidence suggesting that miRNAs repress translation at the later stages of initiation as well. The miRNA *lin-4* targets the *lin-14* and *lin-28* mRNAs, but under inhibitory conditions, *lin-14* and *lin-28* are not altered, indicating that miRNAs inhibit translation after the initiation stage. Interestingly, in both cap-dependent and independent translation, the mRNAs are inhibited by synthetic miRNA, suggesting post-initiation inhibition. Another mechanism by which miRNAs inhibit translation is by ribosome drop off, in which the ribosomes engaged in translation are directed to prematurely terminate translation. There are also proposed mechanisms by which miRNAs can direct the degradation of nascent polypeptides by recruiting proteolytic enzymes (Olsen and Ambros 1999; Petersen et al. 2006).

Microarray studies of transcript levels in cells and tissues in which miRNA pathways were inhibited or in which miRNA levels were altered support the role

of miRNAs in mRNA destabilization (Behm-Ansmant et al. 2006; Giraldez et al. 2006; Rehwinkel et al. 2006; Schmitter et al. 2006; Eulalio et al. 2007). Reports have demonstrated the interaction of the P-body protein GW182 with Argonaute 1 is a key factor that marks mRNAs for degradation, as the depletion of these proteins leads to the upregulation of many mRNA targets. Moreover, knockdown experiments and analyses of the decay intermediates originating from repressed mRNAs in mammalian cells (Wu and Belasco 2006) support the role of decapping and 5'→3' exonucleolytic activities in these systems. Although many of the mRNAs that are targeted by miRNAs undergo substantial destabilization, it is not known what factors determine whether an mRNA follows the degradation or translational-repression pathway (Filipowicz et al. 2008).

In addition to their recognized roles in repressing gene expression, miRNAs have also surprisingly been linked to gene activation. The mechanism of activation is often indirect, with the repression of a repressor leading to the increased expression of specific transcripts. A relatively small number of studies have demonstrated that miRNAs can stimulate gene expression, indicating that these effects are mediated via gene promoters, extracellular receptors and the selective control of 3' or 5' UTRs. Below, we discuss three of the current examples of the role of miRNAs as stimulators of gene expression.

1) Promoter activation: Earlier studies have shown that the exogenous application of small duplex RNAs that are complementary to promoters activates gene expression in a manner similar to proteins and hormones, a phenomenon referred to as RNA activation (RNAa) (Li et al. 2006, Janowski et al. 2007). Soon afterwards, it was discovered that mir-373 targets sites in the promoters of e-cadherin and cold shock domain containing protein C2 (CSDC2), and its overexpression induced the transcription of both genes. Subsequently, mir-205 was discovered to bind to the promoter of the interleukin (IL) tumor suppressor genes IL-24 and IL-32 and, similar to mir-373, induce gene expression (Place et al. 2008; Majid et al. 2010).

2) Target activation: Several reports have shown that miRNAs can induce translation by binding to the 5' or 3' UTR of an mRNA. In the brain, a target sequence of mir-346 was found in the 5' UTR of a splice variant of receptor-interacting protein 140 (RIP140). Gain- and loss-of-function studies established that mir-346 elevated the RIP140 protein levels by facilitating the association of its mRNA with the poly-some fraction. This activity did not require Ago2, indicating that other proteins in complex with the miRNA or a different RIP140 mRNA conformation induced by the miRNA mediated the effect (Tsai et al. 2009). In another study, mir-145 was shown to regulate smooth muscle cell fate and plasticity by upregulating the myocardin gene (Cordes et al. 2009). Along with this, miR-4661, a miRNA discovered in mouse embryonic stem cells, upregulated IL-10 expression in TLR-triggered macrophages by antagonizing IL-10 mRNA degradation mediated by the RBP tristetraprolin (TTP) (Ma et al. 2010).

3) Receptor ligands: Mouse TLR7 and human TLR8, which are members of the Toll-like receptor (TLR) family that are expressed on dendritic cells and B lymphocytes, physiologically recognize and bind to and are activated by ~20-nucleotide

viral single-stranded RNAs (Heil et al. 2004; Lund et al. 2004). Because miRNAs can be secreted in exosomes and are of similar size, it was predicted that they may also serve as TLR7/8 ligands. It was also found that the tumor-secreted mir-21 and mir-29a were ligands for TLR7/8 and were capable of triggering a TLR-mediated prometastatic inflammatory response (Fabbri et al. 2012).

1.3.1 Control of miRNA Expression

Despite the substantial advances in our understanding of miRNA-mediated gene regulation, the mechanisms that control the expression of the miRNAs themselves are less well understood. Homeostatic and feedback mechanisms coordinate the levels of miRNAs with their effector proteins or harmonize the levels of the biogenesis factors that function within the complexes. Often we have the impression that these processes are constitutive and inflexible.

However, diverse mechanisms that regulate the biogenesis and function of small RNAs have been uncovered (Bronevetsky and Ansel 2013; Heo and Kim 2009). Notably, many of these mechanisms provide homeostatic control over the levels of biogenesis factors and/or the resultant miRNAs. Both transcriptional and post-transcriptional mechanisms regulate miRNA biogenesis (Carthew and Sontheimer 2009; Siomi 2010; Schanen and Li 2011).

The first and one of the most important mechanisms controlling miRNA abundance is the regulation of pri-miRNA transcription. pri-miRNAs can be positively or negatively regulated by different factors such as transcription factors, enhancers, silencers and epigenetic modification of the miRNA promoter (Ruegger et al. 2012; Macedo et al. 2013). Investigations in this area have been slowed by limitations in the methods used to define the promoters and measure the transcripts. pri-miRNAs are unstable, as they are processed by the nuclear microprocessor complex very soon after transcription. Therefore, they generally do not accumulate in great abundance in cells and are underrepresented in EST and RNA-Seq libraries.

Recently, these challenges have been overcome by epigenomic and transcriptomic experiments. One study took advantage of the fact that many pri-miRNAs accumulate in cells lacking Droscha to map pri-miRNAs using RNA-Seq (Kirigin et al. 2012).

It has long been known that the levels of mature miRNAs are not determined solely by their transcription. Measurements of pri-miRNAs and their corresponding mature miRNAs were poorly correlated, suggesting that specific miRNAs are subject to developmental regulation of their processing and/or stability (Thomson et al. 2006). Additionally, the expression of these miRNAs continues to be regulated after biogenesis is complete. Mature miRNA homeostasis can be influenced by signals that modulate the stability of the miRISC complex, by nucleases that degrade miRNAs, and/or by the abundance of their mRNA targets. It is estimated that 5–10% of mammalian miRNAs are epigenetically regulated (Breving and Esquela-Kerscher 2010, Brueckner et al. 2007, Han et al. 2007, Toyota et al. 2008).

Despite early reports indicating that miRNAs are often surprisingly stable in cells, displaying half-lives up to 12 days (van Rooij et al. 2007), cell differentiation and cell-fate decisions are frequently marked by dramatic changes in the expression of mature miRNAs.

The Argonaute proteins are limiting factors that determine the total abundance of cellular miRNAs. The deletion of these proteins, specifically Ago1 and Ago 2, was sufficient to drastically reduce miRNA expression (Bronevetsky et al. 2013; Diedrichs and Haber 2007; Lund et al. 2011). Conversely, overexpressing Ago2, but not the other proteins in the miRNA biogenesis pathway, increases miRNA expression in HEK293 cells. Thus, changes in the expression and stability of Ago proteins can have dramatic effects on the expression of mature miRNAs within cells.

The action of miRNA nucleases in the regulation of miRNAs is not well understood, especially in mammals. At least two ribonucleases have been shown to negatively regulate the expression of mature miRNAs. IRE1a, an endoplasmic reticulum (ER) transmembrane RNase activated in response to ER stress, cleaves precursors corresponding to miR-17, miR-34a, miR-96, and miR-125b and mediates the rapid decay of their expression in response to sustained cellular stress (Upton et al. 2012). Additionally, Eri1, a 3'-to-5' exoribonuclease with a double-stranded RNA-binding SAP domain, was discovered to limit miRNA abundance in CD4+T cells and natural killer (NK) cells (Thomas et al. 2012).

The sequence-specific degradation of miRNAs has also been observed with the addition of RNA targets. miRNA “antagomirs” and “miRNA sponges” are two technologies used to specifically knockdown miRNA expression, and both rely on miRNA degradation induced by high levels of miRNA-to-target complementarity (Krutzfeldt et al. 2005; Ebert et al. 2007; Plank et al. 2013). Further work is still needed to determine the extent to which miRNA expression is regulated by target mRNAs, as well as the molecular mechanisms that mediate this final step in the control of miRNA expression.

The posttranscriptional regulatory mechanisms that affect miRNA processing at different stages have recently been investigated (Siomi 2010). For example, p53 can form a complex with Drosha, which increases the processing of pri-miRNAs to pre-miRNAs (Suzuki et al. 2009). Histone deacetylase I can also enhance pri-miRNA processing by deacetylating the microprocessor complex protein DGCR8 (Wada et al. 2012). Additionally, cytokines such as interferons have been shown to inhibit Dicer expression and decrease the processing of pre-miRNAs (Wiesen and Tomasi 2009).

1.3.2 Extracellular miRNAs

RISC components and miRNAs have also been found in exosomes (Valadi et al. 2007). Exosomes isolated from the culture supernatant of many hematopoietic cells, including cytotoxic T lymphocytes, mast cells, and dendritic cells (DCs), as well as DC-derived exosomes, have been shown to stimulate CD4+T-cell activation and

induce tolerance (Zitvogel et al. 1998). Experimentally, vesicles containing both Ago2 and miRNAs, including miR-150, miR-21, and miR-26b, as well as the vesicle-derived miR-150, could be delivered to recipient HMEC-1 human endothelial cells and repress the target mRNAs in the recipient cells. These findings illustrate another mechanism by which immune cell stimulation/activation can lead to significant changes in mature miRNA levels. Interest in extracellular miRNAs in various body fluids has increased substantially as early findings indicated their utility as readily accessible biomarkers.

Circulating miRNAs have been studied in patient samples and animal models in the context of cardiovascular disease, liver injury, sepsis, cancer, and various other physiological and pathophysiological states (Cortez et al. 2011). The origin of extracellular miRNAs is still poorly understood, with blood cells appearing to be a major contributor to circulating miRNAs (Pritchard et al. 2012).

It has also become clear that extracellular miRNAs exist in several distinct forms in human plasma. In addition to miRNAs encapsulated in vesicles such as exosomes, there are stable non-vesicular miRNAs that can be copurified with Ago proteins, which are accessible for direct immunoprecipitation from plasma samples (Arroyo et al. 2011). Further research is needed to clarify the cellular sources of miRNAs, the forms in which they are released, and whether this process is regulated during biological processes.

1.3.3 An Example of the Biological Consequence of miRNAs: Their Role in the Immune System

The role of miRNAs in the immune system has been extensively investigated. Both innate and adaptive immune responses are highly regulated by miRNAs. By targeting the signal transduction proteins involved in the transmission of intracellular signals following initial pathogen recognition and by directly targeting mRNAs that encode specific inflammatory cytokines, miRNAs can have a significant impact on the innate immune response. In addition to their role in regulating the innate immune system, miRNAs have been implicated in adaptive immunity, wherein they control the development, activation and plasticity of T and B cells (Lu and Liston 2009; Xiao and Rajewsky 2009; O'Connell et al. 2010; O'Neill et al. 2011; Plank et al. 2013; Baumjohann and Ansel 2013; Donate et al. 2013).

Furthermore, the central role of miRNAs across many important aspects of innate and adaptive immunity strongly supports their potential in regulating inflammatory diseases. The identification of a broad range of miRNAs that play pathogenic roles is growing. To date, a relatively small number of miRNAs has been associated with specific inflammatory diseases, and most of the identified miRNAs are expressed across multiple tissues and cell types, and many have been shown to play roles in other disease settings, particularly in cancer. Despite the limited numbers of verified targets in inflammatory diseases, many of the targets that were verified in other experimental settings may also be relevant in inflammatory diseases (Plank et al. 2013).

1.4 Conclusion

Early on, transcriptome research was intertwined with the genome. Much of this was due to the mapping of ESTs, and sequencing dominated the scene. Through the use of EST clones and the application of technical concepts such as nucleic acid hybridization, researchers began to use arrayed filters to explore the transcriptional expression of a large number of genes in a single experiment.

The constant improvement of these DNA arrays led to the fabrication of high-density arrays and, finally, microarrays.

At the same time, sequencing also underwent significant changes involving automation and the endless quest to increase the number of reads, and this contributed substantially to a better understanding of the diversity of the transcriptome. Indeed, transcriptome research was rooted in these two major technological approaches (i.e., large-scale hybridization and sequencing).

What made microarrays robust and increased their popularity was the increase in the number of sequences deposited on the slides (currently, these slides contain the entire human or mouse functional genome), the sensitivity of the method (currently, experiments are being performed with nanogram amounts of total RNA to screen the entire functional genome), the simplicity of its use, its commercial availability and the availability of bioinformatics packages dedicated to analyzing the large amounts of data being generated.

Of key importance was the development of statistical procedures for the analysis of large amounts of data, which opened the door for biostatisticians and bioinformaticians.

All of these ongoing technological advances have contributed to the consolidation of the concept of the transcriptome. Unlike the genome, which is essentially static, the transcriptome is variable and is dependent on normal physiological, pathological or environmental conditions. Moreover, it is composed not only of mRNAs but also non-coding RNAs, including miRNAs.

This concept has provided the opportunity for all types of biomedical research to re-examine their results in light of transcriptomics.

References

- Adams J (2008) Sequencing human genome: the contributions of Francis Collins and Craig Venter. *Nat Educ* 1(1):133
- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merrill CR, Wu A, Olde B et al (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252:1651–1656
- Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, Fields C, Venter JC (1992) Sequence identification of 2,375 human brain genes. *Nature* 355:632–634
- Adams MD, Soares MB, Kerlavage AR, Fields C, Venter JC (1993a) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat Genet* 4:373–380

- Adams MD, Kerlavage AR, Fields C, Venter JC (1993b). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nat Genet* 4:256–267
- Ahrens CH, Brunner E, Qeli E, Basler K, Aebersold R (2010) Generating and navigating proteome maps using mass spectrometry. *Nat Rev Mol Cell Biol* 11:789–801
- Allison DB, Cui X, Page GP, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 7:55–65
- Altelaar AF, Munoz J, Heck AJ (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* 14:35–48
- Ambros V (2004) The functions of animal microRNAs. *Nature* 431:350–355
- Ameres SL, Zamore PD (2013) Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* 14(8):475–488
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Gen Biol* 11:R106
- Anderson L (2014) Six decades searching for meaning in the proteome. *J Proteomics*. doi:10.1016/j.jprot.2014
- Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, Mitchell PS, Bennett CF, Pogosova-Agadjanyan EL et al (2011) Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc Natl Acad Sci USA* 108:5003–5008
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS et al (2000) Gene ontology: tool for the unification of biology. *Nat Genet* 25:25–29
- Auer PL, Doerge RW (2010) Statistical design and analysis of RNA sequencing data. *Genetics* 185:405–416
- Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP (2008) The impact of microRNAs on protein output. *Nature* 455:64–71
- Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, Causton H, Conley MP, Elespuru R et al (2005) The external RNA controls consortium: a progress report. *Nat Methods* 2 731–734
- Ball CA, Sherlock G, Parkinson H, Rocca-Sera P, Brooksbank C, Causton HC, Cavalieri D, Gaasterland T, Hingamp P et al (2002) Microarray gene expression data (MGED) society. Standards for microarray data. *Science* 298:539
- Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233
- Batista PJ and Chang HY (2013) Long noncoding RNAs: cellular address codes in development and disease. *Cell* 152:1298–1307
- Baumjohann D, Ansel MK (2013) MicroRNA-mediated regulation of T helper cell differentiation and plasticity. *Nat Rev Immunol* 13:666–678
- Behm-Ansmant I, Rehwinkel J, Doerks T et al (2006) mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* 20:1885–1898
- Bernard K, Auphan N, Granjeaud S, Victorero G, Schmitt-Verhulst AM, Jordan BR, Nguyen C (1996) Multiplex messenger assay: simultaneous, quantitative measurement of expression for many genes in the context of T cell activation. *Nucleic Acids Res* 24:1435–1443
- Bertani S, Sauer S, Bolotin E et al (2011) The noncoding RNA *Mistral* activates *Hoxa6* and *Hoxa7* expression and stem cell differentiation by recruiting *MLL1* to chromatin. *Mol Cell* 43:1040–1046
- Bertucci F, Bernard K, Lorioid B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K, Jordan BR (1999) Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum Mol Genet* 9:1715–1722
- Bohnsack MT, Czaplinski K, Gorlich D (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* 10:185–191
- Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2):185–193

- Borchert GM, Lanier W, Davidson BL (2006) RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* 13:1097–101
- Botwell D (1999) Options available -from start to finish- for obtaining expression data by microarray. *Nat Genet* 21:2–32
- Bratkovic T, Rogelj B (2014) The many faces of small nucleolar RNAs. *Biochim Biophys Acta* 1839:438–443
- Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA et al (2001) Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat Genet* 29:365–371
- Breitling R, Herzyk P (2005) Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarray data. *J Bioinf Comp Biol* 3:1171–1189
- Brennecke J, Stark A, Russell RB et al (2005) Principles of microRNA-target recognition. *PLoS Biol* 3:e85
- Brenner S, Jacob F, Meselson M (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190:576–581
- Breving K, Esquela-Kerscher A (2010) The complexities of microRNA regulation: miRandering around the rules. *Int J BiochemCell Biol* 42:1316–1329
- Brodersen P, Voinnet O (2009) Revisiting the principles of microRNA target recognition and mode of action. *Nat Rev Mol Cell Biol* 10(2):141–1488
- Bronevetsky Y, Ansel MK (2013) Regulation of miRNA biogenesis and turnover in the immune system. *Immunol Rev* 253:304–316
- Bronevetsky Y, Villarino AV, Eisley CJ, Barbeau R, Barczak AJ, Heinz GA, Kremmer E, Heissmeyer V, McManus MT et al (2013) T cell activation induces proteasomal degradation of Argonaute and rapid remodeling of the microRNA repertoire. *J Exp Med* 210:417–432
- Bruceknor B, Stresemann C, Kuner R, Mund C, Musch T, Meister M, Sultmann H, Lyko F (2007) The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. *Cancer Res* 67:1419–1423
- Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinform* 11:94
- Bumgarner SL, Dowell RD, Grisafi P, Gifford DK, Fink GR (2009) Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proc Natl Acad Sci USA* 106:18321–18326
- Bushati N, Cohen SM (2007) MicroRNA functions. *Annu Rev Cell Dev Biol* 23:175–205
- Cahan P, Rovegno F, Mooney D, Newman JC, St. Laurent III G, McCaffrey TA (2007) Meta-analysis of microarray results: challenges, opportunities, and recommendations for standardization. *Gene* 401:12–18
- Camargo AA, Samaia HP, Dias-Neto E, Simão DF, Migotto IA, Briones MR, Costa FF, Nagai MA, Verjovski-Almeida S et al (2001) The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc Natl Acad Sci USA* 98:12103–12108
- Cantor CR (1990) Orchestrating the human genome project. *Science* 248:49–51
- Carthew RW, Sontheimer EJ (2009) Origins and mechanisms of miRNAs and siRNAs. *Cell* 136:642–655
- Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP (1996) Accessing genetic information with high-density DNA arrays. *Science* 274:610–614
- Chekulaeva M, Filipowicz W (2009) Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr Opin Cell Biol* 21:452–460
- Chen Y, Dougherty ER, Bittner ML (1997) Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J Biomed Opt* 2:364–374
- Chen JJ, Wu R, Yang PC, Huang JY, Sher YP, Han MH, Kao WC, Lee PJ, Chiu TF et al (1998) Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. *Genomics* 51:313–324
- Chen CZ, Li L, Lodish HF, Bartel DP (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science* 303:83–86

- Chendrimada TP, Gregory RI, Kumaraswamy E (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436:740–744
- Chu C, Qu K, Zhong FL et al (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA–chromatin interactions. *Mol Cell* 44:667–678
- Churchill GA (2002) Fundamentals of experimental design for cDNA microarrays. *Nat Genet* 32:490–495
- Clément-Ziza M, Gentien D, Lyonnet S, Thiery JP, Besmond C, Decraene C (2009) Evaluation of methods for amplification of picogram amounts of total RNA for whole genome expression profiling. *BMC Genomics* 26:10:246
- Cobb JP, Mindrinos MN, Miller-Graziano C, Calvano SE, Baker HV, Xiao W, Laudanski K, Brownstein BH, Elson CM et al (2005) Application of genome-wide expression analysis to human health and disease *PNAS* 102(13):4801–4806
- Cordes KR, Sheehy NT, White MP, Berry EC, Morton SU, Muth AN, Lee TH, Miano JM, Ivey KN et al (2009) miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature* 460:705–710
- Cortez MA, Bueso-Ramos C, Ferdin J (2011) MicroRNAs in body fluids—the mix of hormones and biomarkers. *Nat Rev Clin Oncol* 8:467–477
- Cullen BR (2004) Transcription and processing of human microRNA precursors. *Mol Cell* 16:861–865
- Czech B, Hannon GJ (2011) Small RNA sorting: matchmaking for argonautes. *Nat Rev Genet* 12:19–31
- De Klerk E den Dunnen JT t Hoen PA (2014) RNA sequencing : from tag-based profiling to resolving complete transcript structure. *Cell Mol Life Sci* (epub ahead of print) 71(18):3537–3551.
- Degrelle SA, Hennequet-Antier C, Chiappello H, Piot-Kaminski K, Piumi F, Robin S, Renard JP, Hue I (2008) Amplification biases: possible differences among deviating gene expressions. *BMC Genomics* 9:46
- Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature* 432:231–235
- Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 4:3
- Derrien T, Guigo R, Johnson R (2011) The long non-coding RNAs: a new (p)layer in the “dark matter”. *Front Genet* 2:107
- Didiano D, Hobert, O (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol* 13:849–851
- Diederichs S, Haber DA (2007) Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell* 131:1097–1108
- Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D et al (2013) A comprehensive evaluation of normalization methods for illumine high-throughput RNA sequencing data analysis. *Brief Bioinform* 14(6):671–683
- Djebali S, Davis CA, Merkel A et al (2012) Landscape of transcription in human cells. *Nature* 489:101–108
- Donate PB, Fornari TA, Macedo C, Cunha TM, Nascimento DC, Sakamoto-Hojo ET, Donadi EA, Cunha FQ, Passos GA (2013) T cell post-transcriptional miRNA-mRNA interaction networks identify targets associated with susceptibility/resistance to collagen-induced arthritis. *PLoS One* 8(1):e54803
- Duewer DL, Jones WD, Reid LH, Salit M (2009) Learning from microarray interlaboratory studies: measures of precision for gene expression. *BMC Genomics* 10:153
- Dufva M (2005) Fabrication of high quality microarrays. *Biomol Eng* 22:173–184
- Dujon, B (1998) European functional analysis network (EUROFAN) and the functional analysis of the *Saccharomyces cerevisiae* genome. *Electrophoresis* 19:617–624
- Ebert MS, Neilson JR, Sharp PA (2007) MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods* 4:721–726
- Edwards D (2003) Non-linear normalization and background correction in onechannel cDNA microarrays studies. *Bioinformatics* 19:825–833

- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *PNAS* 95(25):14863–14868
- Epstein JR, Leung AP, Lee KH, Walt DR (2003) High-density, microsphere based fiber optic DNA microarrays. *Biosens Bioelectron* 18:541–546
- Eulalio A, Rehwinkel J, Stricker M, Huntzinger E, Yang SF, Doerks T, Dorner S, Bork P, Boutros M et al (2007) Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes Dev* 21:2558–2570
- Fabbri M, Paone A, Calore F, Galli R, Gaudio E, Santhanam R, Lovat F, Fadda P, Mao C et al (2012) microRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. *Proc Natl Acad Sci USA* 109:E2110–E2116
- Fang Z, Cui X (2010) Design and validation issues in RNA-seq experiments. *Brief Bioinform* 12(3):280–287
- Fang Z, Cui X (2011) Design and validation issues in RNA-Seq experiments. *Brief Bioinformatics* 12:280–287
- Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP (2005) The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science* 310:1817–1821
- Fatica A, Bozzoni I (2014) Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* 15:7–21
- Ferguson JA, Steemers FJ, Walt DR (2000) High-density fiber optic DNA random microsphere array. *Anal Chem* 72:5618–5624
- Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of posttranscriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 9(2):102–114
- Fisher RA (1935) *The design of experiments*. Oxford, England. Oliver & Boyd, p 251
- Fonseca NA, Rung J, Brazma A, Marioni JC (2012) Tools for mapping high-throughput sequencing data. *Bioinformatics* 28(24):3169–3177
- Forler S, Klein O, Klose J (2014) Individualized proteomics *J Proteomics* 107C:56–61
- Forman JJ, Collier HA (2010) The code within the code: microRNAs target coding regions. *Cell Cycle* 9:1533–1541
- Forman JJ, Legesse-Miller A, Collier HA (2008) A search for conserved sequences in coding regions reveals that the let⁷ microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci USA* 105:14879–14884
- Garber M, Grabherr MG, Guttman M, Trapnell C (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods* 8:469–477
- Geeleher P, Morris D, Golden A, Hinde JP (2008) Handbook: bioconductor/Buntu users manual. <http://www3.it.nuigalway.ie/agolden/bioconductor/version1/handbook.pdf>
- Geisler S, Collier J (2013) RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 14:699–672
- Gentleman RC, Carey VJ, Bates DM (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5(10):R80
- Gentleman RC, Carey VJ, Huber W et al (2005) *Bioinformatics and computational biology solutions using R and bioconductor*. Springer, New York, p 473
- Gershon D (2002) Microarray technology, an array of opportunities; technology feature. *Nature* 416:885–891
- Geschwind DH, Gregg JP (2002) *Microarrays for the neurosciences: an essential guide*. The MIT Press
- Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312:75–79
- Goecks J, Nekrutenko A, Taylor J; Galaxy Team (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11:R86

- Granjeaud S, Nguyen C, Rocha D, Luton R, Jordan BR (1996) From hybridization image to numerical values: a practical, high throughput quantification system for high density filter hybridizations. *Genet Anal Biomol Eng* 12:151–162
- Granjeaud S, Bertucci F, Jordan BR (1999) Expression profiling: DNA arrays in many guises. *Bioessays* 21:781–790
- Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432:235–240
- Gress TM, Hoheisel JD, Lennon GG, Zehetner G, Lehrach H (1992) Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm Genome* 3:609–661
- Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27:91–105
- Gu S, Jin L, Zhang F, Sarnow P, Kay MA (2009) Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat Struct Mol Biol* 16:144–150
- Gunderson KL, Kruglyak S, Graige MS, Garcia F, Kermani BG, Zhao C, Che D, Dickinson T, Wickham E et al (2004) Decoding randomly ordered DNA arrays. *Genome Res* 14:870–877
- Guo Y, Ye F, Sheng Q, Clark T, Samuels DC (2013) Three-stage quality control strategies for DNA re-sequencing data. *Briefings in Bioinformatics* doi:10.1093/bib/bbt069
- Haase AD, Jaskiewicz L, Zhang H, Lainé S, Sack R, Gatignol A, Filipowicz W (2005) TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO* 6:961–967
- Han L, Witmer PD, Casey E, Valle D, Sukumar S (2007) DNA methylation regulates MicroRNA expression. *Cancer Biol Ther* 6:1284–1288
- Heber S, Sick B (2006) Quality assessment of Affymetrix GeneChip data. *OMICS* 10(3):358–368
- Heil F, Hemmi H, Hochrein H, Ampenberger F, Kirschning C, Akira S, Lipford G, Wagner H, Bauer S (2004) Species-specific recognition of single-stranded RNA via Toll-like receptor 7 and 8. *Science* 303:1526–1529
- Heo I, Kim VN (2009) Regulating the regulators: posttranslational modifications of RNA silencing factors. *Cell* 139:28–31
- Huber W, von Heydebreck A, Sülthmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 18(1):S96–S104
- Humphreys DT, Westman BJ, Martin DI, Preiss T (2005) MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci USA* 102:16961–16966
- Hutvagner G, McLachlan J, Pasquinelli AE, Bálint E, Tuschl T, Zamore PD (2001) A cellular function for the RNA interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293:834–838
- Iliopoulos D, Hirsch HA, Struhl K (2009) An epigenetic switch involving NF- κ B, Lin28, Let-7 microRNA, and IL6 links inflammation to cell transformation. *Cell* 139:693–706
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Bioinformatics* 4(2):249–264
- Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J et al (2005) Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2:345–350
- Jackson RJ, Standart N (2007) How do microRNAs regulate gene expression? *Sci STKE* 2007(367):re1
- Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3:318–356
- Janowski BA, Younger ST, Hardy DB, Ram R, Huffman KE, Corey DR (2007) Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol* 3:166–173

- Järvinen AK, Hautaniemi S, Edgren H, Auvinen P, Saarela J, Kallioniemi OP, Monni O (2004) Are data from different gene expression microarray platforms comparable? *Genomics* 83:1164–1168
- Jordan B (2012) The microarray paradigm and its various implementations. In Jordan B (ed) *Microarrays in diagnostics and biomarker development. Current and future applications*. Springer-Verlag, Berlin Heidelberg.
- Jordan BR (1998) Large scale expression measurement by hybridization methods: from high-density membranes to “DNA chips”. *J Biochem* 124:251–258
- Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30
- Kawamata T, Tomari Y (2010) Making RISC. *Trends Biochem. Sci* 35:368–376
- Kellis M, Wold B, Snyder MP et al (2014) Defining functional DNA elements in the human genome. *Proc Natl Acad Sci USA* 111:6131–6138
- Kerr MK, Churchill GA (2001) Experimental design for gene expression microarrays. *Biostatistics* 2:183–201
- Kim VN, Nam JW (2006) Genomics of microRNA. *Trends Genet* 22:165–173
- Kirigin FF, Lindstedt K, Sellars M, Ciofani M, Low SL, Jones L, Bell F, Pauli F, Bonneau R et al (2012) Dynamic microRNA gene transcription and processing during T cell development. *J Immunol* 188:3257–3267
- Kooperberg C, Fazzio TG, Delrow JJ, Tsukiyama T (2002) Improved background correction for spotted DNA microarrays. *J Comp Biol* 9:55–66
- Kren BT, Wong PY, Sarver A, Zhang X, Zeng Y, Steer CJ (2009) MicroRNAs identified in highly purified liver-derived mitochondria may play a role in apoptosis. *RNA Biol* 6:65–72
- Krützfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, Stoffel M (2005) Silencing of microRNAs in vivo with ‘antagomirs’. *Nature* 438:685–689
- Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, O’Day E, Chowdhury D, Dykxhoorn DM, Tsai P et al (2009) miR-24 inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to “seedless” 3’ UTR microRNA recognition elements. *Mol Cell* 35:610–625
- Landthaler M, Yalcin A, Tuschl T (2004) The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis. *Curr Biol* 14:2162–2167
- Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J (2005) Independence and reproducibility across microarray platforms. *Nat Methods* 2:337–344
- Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75:843–854
- Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O et al (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425:415–419
- Levine E, Hwa T (2008) Small RNAs establish gene expression thresholds. *Curr Opin Microbiol* 11:574–579
- Levine E, Zhang Z, Kuhlman T, Hwa T (2007) Quantitative characteristics of gene regulation by small RNA. *PLoS Biol* 5:e229
- Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. *Cell* 115:787–798
- Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120:15–20
- Li LC, Okino ST, Zhao H, Pookot D, Place RF, Urakami S, Enokida H, Dahiya R (2006) Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci USA* 103:17337–17342
- Lin SL, Kim H, Ying SY (2008) Intron-mediated RNA interference and microRNA (miRNA). *Front Biosci* 13:2216–2230
- Liu G, Mattick JS, Taft RJ (2013) A meta-analysis of the genomic and transcriptomics composition of complex life. *Cell Cycle* 12:2061–2072

- Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14:1675–1680
- Lönnstedt I, Speed T (2002) Replicated microarray data. *Stat Sinica* 12:31–46
- Lu LF, Liston A (2009) MicroRNA in the immune system, microRNA as an immune system. *Immunology* 127:291–298
- Lund JM, Alexopoulou L, Sato A, Karow M, Adams NC, Gale NW, Iwasaki A, Flavell RA (2004a) Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc Natl Acad Sci USA* 101:5598–5603
- Lund E, Güttinger S, Calado A, Dahlberg JE, Kutay U (2004b) Nuclear export of microRNA precursors. *Science* 303:95–98
- Lund E, Sheets MD, Imboden SB, Dahlberg JE (2011) Limiting Ago protein restricts RNAi and microRNA biogenesis during early development in *Xenopus laevis*. *Genes Dev* 25:1121–1131
- Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ (2009) GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinform* 10:161
- Luo J, Schumacher M, Scherer A, Sanoudou D, Megherbi D, Davison T, Shi T, Tong W, Shi L et al (2010) A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data. *Pharmacogenomics J* 10:278–291
- Ma F, Liu X, Li D, Wang P, Li N, Lu L, Cao X (2010) microRNA-466 l upregulates IL-10 expression in TLR-triggered macrophages by antagonizing RNA-binding protein tristetraprolin-mediated IL-10 mRNA degradation. *J Immunol* 184:6053–6059
- Macedo C, Evangelista AF, Marques MM, Octacilio-Silva S, Donadi EA, Sakamoto-Hojo ET, Passos GA (2013) Autoimmune regulator (Aire) controls the expression of microRNAs in medullary thymic epithelial cells. *Immunobiol* 218:554–560
- Maeda N, Kasukawa T, Oyama R et al (2006) Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. *PLoS Genet* 2: e62
- Majid S, Dar AA, Saini S, Yamamura S, Hirata H, Tanaka Y, Deng G, Dahiya R (2010) microRNA-205-directed transcriptional activation of tumor suppressor genes in prostate cancer. *Cancer* 116:5637–5649
- MAQC Consortium (2006) The microarray quality control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24:1151–1161
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18(9):1509–1517
- Martin JA, Wang Z (2011) Next-generation transcriptome assembly. *Nat Rev Genet* 12:671–682
- McCarthy JJ (2008) MicroRNA-206: the skeletal muscle-specific myomiR. *Biochim Biophys Acta* 1779:682–691
- Mehta P, Goyal S, Wingreen NS (2008) A quantitative comparison of sRNA-based and protein-based gene regulation. *Mol Syst Biol* 4:211
- Mercer TR, Mattick JS (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Biol* 20: 300–307
- Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126:1203–1217
- Mitarai N, Benjamin JA, Krishna S, Semsey S, Csiszovszki Z, Massé E, Snepken K (2009) Dynamic features of gene expression control by small regulatory RNAs. *Proc Natl Acad Sci USA* 106:10655–10659
- Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Brian KC et al (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci USA* 105:10513–10518
- Moorcroft MJ, Meuleman WR, Latham SG, Nicholls TJ, Egeland RD, Edwin M., Southern EM (2005) In situ oligonucleotide synthesis on poly(dimethylsiloxane): a flexible substrate for microarray fabrication. *Nucleic Acids Res* 33:e75

- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptome by RNA-Seq. *Nat Methods* 5(7):621–628
- Nagano T, Mitchell JA, Sanz LA et al (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322:1717–1720
- Neguembor MV, Jothi M, Gabellini D (2014) Long noncoding RNAs, emerging players in muscle differentiation and disease. *Skelet Muscle* 4:8
- Nguyen C, Rocha D, Granjeaud S, Baldit M, Bernard K, Naquet P, Jordan BR (1995) Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics* 29:207–216
- Nilsen TW (2007) Mechanisms of microRNA-mediated gene regulation in animal cells. *Trends Genet* 23:243–249
- Nottrott S, Simard MJ, Richter JD (2006) Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol* 13:1108–1114
- Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, Pitas A, Richmond T, Gorski T, Berg JP et al (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res* 12:1749–1755
- Nygaard VL, Hovig E (2006) Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. *Nucleic Acids Res* 34:996–1014
- O'Connell RM, Rao DS, Chaudhuri AA, Baltimore D (2010) Physiological and pathological roles for microRNAs in the immune system. *Nat Rev Immunol* 10(2):111–122
- O'Neill LA, Sheedy FJ, McCoy CE (2011) MicroRNAs: the fine-tuners of Toll-like receptor signalling. *Nat Rev Immunol* 11:163–175
- Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y, Matsubara K (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat Genet* 2:173–179
- Olsen PH, Ambros V (1999) The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental Biol* 216:671–680
- Osella M, Bosia C, Cora D et al (2011) The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLoS Comput Biol* 7:e1001101
- Oshlack A, Robinson MD, Young M (2010) From RNA-seq reads to differential expression results. *Genome Biol* 11:220–230
- Padron G, Domont GB (2014) Two decades of proteomics in Latin America: a personal view. *J Proteomics* 107C:83–92
- Patel RK, Jain M (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* 7:e30619.
- Pandey RR, Mondal T, Mohammad F et al (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 32:232–246.
- Penny GD, Kay GF, Sheardown, SA et al (1996) Requirement for Xist in X chromosome inactivation. *Nature* 379:131–137
- Petersen CP, Bordeleau ME, Pelletier J, Sharp PA (2006) Short RNAs repress translation after initiation in mammalian cells. *Mol Cell* 21:533–542
- Pietu G, Alibert O, Guichard V, Lamy B, Bois F, Leroy E, Mariage-Samson R, Houlgatte R, Soularue P, Auffray C (1996) Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res* 6:492–503
- Pietu G, Mariage-Samson R, Fayein NA, Matingou C, Eveno E, Houlgatte R, Decraene C, Vandembrouck Y, Tahi F et al (1999) The Genexpress IMAGE Knowledge Base of the Human Brain Transcriptome: a Prototype Integrated Resource for Functional and Computational Genomics. *Genome Res* 9:195–209
- Pillai RS, Bhattacharyya SN, Filipowicz W (2007) Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol* 17:118–126

- Place RF, Li LC, Pookot D, Noonan EJ, Dahiya R (2008) microRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci USA* 105:1608–1613
- Plank M, Maltby S, Mattes J, Foster PS (2013) Targeting translational control as a novel way to treat inflammatory disease: The emerging role of MicroRNAs. *Clin Exp Allergy* 43(9):981–999
- Plath K, Fang J, Mlynarczyk-Evans SK et al (2003) Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300:131–135
- Ploner A, Miller LD, Hall P, Bergh J, Pawitan Y (2005) Correlation test to assess low-level processing of high-density oligonucleotide microarray data. *BMC Bioinformatics* 6:80
- Politz JC, Hogan EM, Pederson T (2009) MicroRNAs with a nucleolar location. *RNA* 15:1705–1715
- Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* 136:629–641
- Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, Tait JF, Tewari M (2012) Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer Prev Res* 5:492–497
- Quackenbush J (2001) Computational analysis of microarray data. *Nat Rev Genet* 2:418–427
- Ravasi T, Suzuki H, Pang KC et al (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res* 16:11–19
- Rehwinkel J, Natalin P, Stark A, Brennecke J, Cohen SM, Izaurralde E (2006) Genome-wide analysis of mRNAs regulated by drosha and Argonaute proteins in *Drosophila melanogaster*. *Mol Cell Biol* 26:2965–2975
- Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403:901–906
- Ripoli A, Rainaldi G, Rizzo M, Mercatanti A, Pitto L (2010) The Fuzzy Logic of MicroRNA Regulation: a Key to Control Cell Complexity. *Curr Genomics* 11:350–353
- Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth GK (2007) A comparison of background corrections methods for two-color microarrays. *Bioinformatics* 23(20):2700–2707
- Rocha D, Carrier A, Naspetti M, Victorero G, Anderson E, Botcherby M, Nguyen C, Naquet P, Jordan BR (1997) Modulation of mRNA levels in the presence of thymocytes and genome mapping for a set of genes expressed in mouse thymic epithelial cells. *Immunogenetics* 46:142–151
- Rougemont J, Amzallag A, Iseli C, Farinelli L, Xenarios I, Naef F (2009) Rolex: statistical analysis of Solexa sequencing data. R package version 1.20.0 Available at Bioconductor (<http://bioconductor.org/packages/release/bioc/html/Rolex.html>)
- Rüegger S, Großhans H (2012) MicroRNA turnover: when, how, and why. *Trends Biochem Sci* 37:436–446
- Sana J, Faltejskova P, Svoboda M, Slaby O (2012) Novel classes of non-coding RNAs and cancer. *J Translat Med* 10:103–123
- Schanen BC, Li X (2011) Transcriptional regulation of mammalian miRNA genes. *Genomics* 97:1–6
- Schena M, Shanon D, Heller R et al (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* 93:10614–10619
- Schmitter D, Filkowski J, Sewer A et al (2006) Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Res* 34:4801–4815
- Seyednasrollah F, Laiho A, Elo LL (2013) Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in Bioinformatics* doi:10.1093/bib/bbt086 (in press)
- Shi L, Campbell G, Jones WD et al (2010) The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol* 28 (8):827–838
- Shi L, Reid LH, Jones WD et al (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24 (9):1151–1161

- Shimoni Y, Friedlander G, Hetzroni G et al (2007) Regulation of gene expression by small noncoding RNAs: a quantitative view. *Mol Syst Biol* 3:138
- Shin C, Nam JW, Farh KK et al (2010) Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell* 38:789–802
- Silver JD, Ritchie ME, Smyth GK (2009) Microarray background correction: maximum likelihood estimation for the normal-exponential convolution. *Biostatistics* 10(2):352–363
- Singh RL, Maganti RJ, Jabba SV, Wang M, Deng G, Heath JD, Kurn N, Wangemann P (2005) Microarray-based comparison of three amplification methods for nanogram amounts of total RNA. *Am J Physiol Cell Physiol* 288:C1179–C1189
- Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, Sussman MR, Cerrina F (1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat Biotechnol* 10:974–978
- Siomi H, Siomi MC (2010) Posttranscriptional regulation of microRNA biogenesis in animals. *Mol Cell* 38:323–332
- Slonin DK, Yanai I (2009) Getting Started in Gene Expression Microarray Analysis. *PLoS Comput Biol* 5(10):e1000543
- Smyth GK (2005) Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds) *Bioinformatics and computational biology solutions using R and bioconductor*. Springer, New York, 397–420
- Sogayar MC, Camargo AA, Bettoni F et al (2004) A transcript finishing initiative for closing gaps in the human transcriptome. *Genome Res* 14:1413–1423
- Soneson C, Delorenzi M (2013) A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinform* 14:91–108
- Standart N, Jackson RJ (2007) MicroRNAs repress translation of m7Gppp-capped target mRNAs in vitro by inhibiting initiation and promoting deadenylation. *Genes Dev* 21:1975–1982
- Stark A, Brennecke J, Bushati N et al (2005) Animal microRNAs confer robustness to gene expression and have a significant impact on 3' UTR evolution. *Cell* 123:1133–1146
- Stefani G, Slack FJ (2008) Small non-coding RNAs in animal development. *Nat Rev Mol Cell Biol* 9:219–230
- Stekel D (2003) *Microarray Bioinformatics*. Cambridge University Press, Cambridge. ISBN:9780521525879
- Strausberg RL, Riggins GL (2001) Navigating the human transcriptome. *Proc. Natl. Acad. Sci. USA* 98:11837–11838
- Subramanian A, Tamayo P, Mootha VK et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* 102:15545–15550
- Sudo K, Chinen K, Nakamura Y (1994) 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library. *Genomics* 24:276–279
- Sudo H, Mizoguchi A, Kawauchi J, Akiyama H, Takizawa S (2012) Use of non-amplified RNA samples for microarray analysis of gene expression. *PLoS ONE* 7:e31397
- Suzuki HI, Yamagata K, Sugimoto K, Iwamoto T, Kato S, Miyazono K (2009) Modulation of microRNA processing by p53. *Nature* 460:529–533
- Taft RJ, Pang KC, Mercer TR et al (2010) Non-coding RNAs: regulators of disease. *J. Pathol.* 220:126–139
- Takeda J, Yano H, Eng S, Zeng Y, Bell GI (1993) Construction of a normalized directionally cloned cDNA library from adult heart and analysis of 3040 clones by partial sequencing. *Hum Mol Genet* 2:1793–1798
- Tay Y, Zhang J, Thomson AM et al (2008) microRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 455:1124–1128
- The ENCODE Consortium (2011) Standards, guidelines and best practices for RNA-seq. http://encodeproject.org/ENCODE/protocols/dataStandards/ENCODE_RNAseq_Standards_V1.0.pdf
- Thomson JM, Newman M, Parker JS et al (2006) Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev* 20:2202–2207
- Thomas MF, Abdul-Wajid S, Panduro M et al. (2012) Eri1 regulates microRNA homeostasis and mouse lymphocyte development and antiviral function. *Blood* 120:130–142

- Toyota M, Suzuki H, Sasaki Y et al (2008) Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer. *Cancer Res* 68:4123–4132
- Tsai NP, Lin YL, Wei LN (2009) microRNA mir-346 targets the 5-untranslated region of receptor-interacting protein 140 (RIP140) mRNA and up-regulates its protein expression. *Biochem J* 424:411–418
- Tusher VG, Tibshirani R, Chu G (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98(9):5116–5121
- Upton JP, Wang L, Hand D et al (2012) IRE1a cleaves select microRNAs during ER stress to derepress translation of proapoptotic caspase-2. *Science* 338:818–822
- Valadi H, Ekstrom K, Bossios A et al (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol* 9:654–659
- Valencia-Sanchez MA, Liu J, Hannon GJ et al (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev* 20:515–524
- Van Haften RI, Schroen B, Janssen BJ, van Erk A, Debets JJ, Smeets HJ, Smits JF, van den Wijngaard A, Pinto YM, Evelo CT (2006) Biologically relevant effects of mRNA amplification on gene expression profiles. *BMC Bioinformatics* 7:200
- Van Heesch S, Van Iterson M, Jacobi J et al (2014) Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol* 15:R6
- Van Rooij E, Sutherland LB, Qi X et al (2007) Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* 316:575–579
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. *Science* 270:484–487
- Velculescu VE, Zhang L, Zhou W et al (1997) Characterization of the yeast transcriptome. *Cell* 88:243–251
- Vella MC, Choi EY, Lin SY et al (2004) The *C. elegans* microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3' UTR. *Genes Dev* 18:132–137
- Wada T, Kikuchi J, Furukawa Y (2012) Histone deacetylase 1 enhances microRNA processing via deacetylation of DGCR8. *EMBO Rep* 13:142–149
- Wang S, Raghavachari S (2011) Quantifying negative feedback regulation by microRNAs. *Phys Biol* 8:055002
- Wang X, Cairns MJ (2013) Gene set enrichment analysis of RNA-Seq data: integrating differential expression and splicing. *BMC Bioinform* 14(5):S16
- Wang J, Hu L, Hamilton SR, Coombes KR, Zhang W (2003) RNA amplification strategies for cDNA microarray experiments. *Biotechniques* 34:394–400
- Watson JD (1990) The human genome project: past, present, and future. *Science* 248:44–49
- Wery M, Kwapisz M, Morillon A (2011) Noncoding RNAs in gene regulation. *Wiley Interdiscip Rev Syst Biol Med* 3:728–738
- Wiesen JL, Tomasi TB (2009) Dicer is regulated by cellular stresses and interferons. *Mol Immunol* 46:1222–1228
- Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75:855–862
- Winter J, Jung S, Keller S et al (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* 11:228–234
- Wreschner DH, Herzberg M (1984) A new blotting medium for the simple Isolation and Identification of highly resolved messenger RNA. *Nucleic Acids Res* 12:1349–1359
- Wu L, Fan J, Belasco JG (2006) microRNAs direct rapid deadenylation of mRNA. *Proc. Natl Acad Sci USA* 103:4034–4039
- Xiao C, Rajewsky K (2009) MicroRNA control in the immune system: basic principles. *Cell* 136:26–36
- Yang JS, Lai EC (2011) Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. *Mol Cell* 43:892–903
- Yi R, Qin Y, Macara IG et al (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* 17:3011–3016

- Young MD, Wakefield MJ, Smyth GK et al (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11:R14
- Yue D, Liu H, Huang Y (2009) Survey of Computational Algorithms for MicroRNA Target Prediction. *Curr Genomics* 10:478–492
- Zamore PD, Haley B (2005) Ribo-gnome: the big world of small RNAs. *Science* 309:1519–1524
- Zhang H, Kolb FA, Brondani V et al (2002) Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J* 21:5875–5885
- Zhao N, Hashida H, Takahashi N, Misumi Y, Sakaki Y (1995) High-density cDNA filter analysis: a novel approach for large-scale, quantitative analysis of gene expression. *Gene* 156:207–213
- Zitvogel L, Regnault A, Lozier A et al (1998) Eradication of established murine tumors using a novel cell-free vaccine: dendritic cell-derived exosomes. *Nat Med* 4:594–600