
Preface

This is a book about the procedures and methods that are used to describe the structure of the messenger RNAs and non-coding RNAs that are transcribed as the immediate gene products by RNA polymerase II in mammalian cells. It is intended for researchers working on a biological problem that involves characterization of the expression of “gene X.” The book is focused on the structure of the RNA products of gene X and mapping of the proteins associated with these RNAs. The book is mainly intended for the non-specialist in RNA biology.

Recent insight into the transcripts generated from the mammalian genome (i.e., the transcriptome) has revealed that transcription is a far more complex phenomenon than previously thought. In a sense, the present situation is comparable to the mid-1970s when the exon–intron organization of genes was discovered. Prior to that, it was generally believed that the mature mRNA was co-linear with the gene from which it was transcribed. This view was challenged by the extraordinary size of the genomes and the puzzling observation of very long nuclear RNA molecules that were capped and polyadenylated similar to the mature mRNAs. The introduction of recombinant DNA technology allowed for a direct comparison of the genes and their RNA products and led to the surprising conclusion that almost all of the genes in vertebrates are a mosaic of mRNA encoding exons interrupted by, on the average, six introns that are subsequently removed by RNA splicing. These intronic sequences comprise 30% of the mammalian genome. In hindsight, it is interesting to note that such a manifest phenomenon could be overlooked for many years.

It now appears that we are confronted with a similar dramatic change in the view of our genes and their products. New methods designed to give a complete and unbiased view on transcription have shown that the transcriptional landscape of the genome is far more complex than previously believed. Most of the genome is transcribed, and, within a given locus, the typical picture is that of multiple, overlapping transcripts generated from both strands of the DNA. Furthermore, characterization of the mature transcripts shows that half of the capped, spliced, and polyadenylated transcripts do not encode a protein. This class of non-coding RNAs are essentially in search of a function, but their characterization is included as part of the scope of this book because their biosynthesis is parallel to that of the mRNAs and because they may belong to a parallel regulatory universe to that of their protein encoding cousins, the mRNAs.

Organization of the Volume

The volume is organized into two parts. The first part deals with preparation and analysis of RNA. The second part is about the proteins and miRNAs that bind to RNA to regulate its function. The final chapter does not follow this outline. It deals with the problems of

outsourcing experimental work for high-throughput services. Each part has a conceptual chapter that introduces the new concepts in the field. Bioinformatics and experimental chapters are mixed to emphasize that bioinformatics should become an integral part of the experimental work, although this may be a bit optimistic at present. The volume contains both very basic and advanced chapters. The reason for the former is to have the basics at hand while embarking on the new and more advanced techniques.

Part I: RNA Methods

The first chapter of the volume introduces the new view on the transcriptional landscape characterized by multiple, overlapping transcripts from both strands of DNA. The complexity of transcripts derived from virtually any genomic region suggest that the operational unit in description of gene expression should be the transcript rather than the DNA from which it was transcribed.

Chapter 2 is about the basics of working with RNA. The chemical nature of RNA is briefly introduced followed by a description of how to create a working environment for RNA work in particular, with respect to maintaining the integrity of the RNA. This is followed by introductions to all of the basic procedures, including extraction, precipitation, quantitation, and storage. Recommendations for preparation of standard reagents and short protocols are also included. Another basic procedure, synthesis of RNA by *in vitro* transcription is described in **Chapter 3**. Beckert and Masquida provide the protocols for template preparation, synthesis of RNA, and purification of transcripts. They also discuss the synthesis of transcripts that are modified at the 5' end or internally for specialized purposes as well as the use of ribozymes to create populations of transcripts with homogeneous ends for NMR or X-ray crystallography. Continuing with a classic and very basic technique, Josefsen and Nielsen in **Chapter 4** present variations of northern blotting and hybridization analysis. Recent developments have made northern blotting analysis almost as sensitive as nuclease protection analysis and to many it remains the most convincing method for analysis of the size and quantity of an RNA transcript.

The present volume is focused on RNA polymerase II transcripts that with few exceptions are polyadenylated. These RNAs constitute 1–4% of cellular RNA and have to be purified from other RNAs in many protocols. In **Chapter 5**, Jacobsen and colleagues describe a variation of the classical oligo(dT) chromatography for purification of poly(A)⁺ RNA using Locked Nucleic Acid (LNA) oligo(T) capture of the poly(A)⁺. This is a very efficient method, and the chapter also serves to introduce LNA which has proven to be a particularly useful tool in many hybridization-based applications in RNA biology, including *in situ* hybridization and microarray analysis. The poly(A) tail of mRNA has several functions including stability and translational control which both depend on the length of the tail. Unfortunately, the tail length is quite difficult to assess. Meijer and de Moor provide a simple method for fractionation of mRNA according to tail length in **Chapter 6**. The method is based on differential elution from oligo(dT) and can be used for preparation of samples for microarray analysis. In **Chapter 7** by Yeku and Frohman, both ends of the RNA molecule are addressed. The chapter presents improvements to the Rapid Amplification of cDNA ends (RACE) technique. The method provides easy access to

full-length cDNA which is of particular significance because an important aspect of diversity in gene expression involves the use of alternative 5' and 3' ends.

The sequencing of the human genome was a milestone in biology, and the public access to genome data organized in genome browsers is a beautiful testimony to the openness of scientific endeavors. In **Chapter 8**, Torarinsson provides a primer to two such browsers (UCSC and Ensembl) with short exercises. The following chapter, **Chapter 9**, by George and Tenenbaum, is aimed at the much more experienced researcher. Here, a comprehensive list of web-based resources for the identification and study of RNA structural motifs is presented. The list comprises databases as well as analytical tools, each with a link, a brief description and a primary literature reference. These motifs are of particular importance for understanding protein binding and regulatory functions associated with the RNA molecules. RNA motifs are also amenable to experimental analysis of their structure, and two chapters in the electronic supplementary materials present such methods. First, in ESM1, Regulski and Breaker describe the use of in-line probing in the characterization of riboswitches in the bacterial world. Riboswitches are found in mammalian systems, but the technique is applicable to all RNA structures. This chapter was originally published as Chapter 4 in *Methods in Molecular Biology*, Vol. 419, *Post-Transcriptional Gene Regulation*, edited by Jeffrey Wilusz. Then, in ESM2, Wakeman and Winkler, in addition to providing a protocol on in-line probing, present structure probing of RNA by SHAPE (Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension). This is a very useful technique that has been used in structure probing of large molecules such as the HIV-1 genome. SHAPE can also be used to study the folding of RNA molecules provided that a fast-reacting acylation reagent is used. This chapter was originally published as Chapter 4 in *Methods in Molecular Biology*, Vol. 540, *Riboswitches: Methods and Protocols*, edited by Alexander Serganov.

The next two chapters deal with the most powerful of post-transcriptional modification processes: alternative splicing. This process is a major contributor to the diversity of gene products derived from the relatively few genes in the human genome. Furthermore, an increasing number of errors in gene expression leading to diseases are found to involve splicing errors. In **Chapter 10**, Zhang and Stamm provide an overview along with a description of bioinformatics tools to predict the influence of a mutation on alternative pre-mRNA splicing and the experimental testing of these predictions. Then, in **Chapter 11**, Lützelberger and Kjems show how the classical S1-nuclease protection method can be used to quantitate alternatively spliced mRNA isoforms. The method requires no specialized equipment and allows detection of as few as a couple of hundred femtograms of a specific RNA.

RNA interference (RNAi) is the method of choice for inactivation of cellular RNA molecules. In **Chapter 12**, Sioud provides a broad review of the use of RNAi as a research tool and in therapy. After an introduction to the RNAi pathway, the rules for design of siRNA are presented. This is followed by a thorough discussion of the detection of exogenous RNA by the immune system. Particular attention is given to separation of the effects of gene silencing from unwanted effects that have led to many erroneous conclusions in the literature. **Chapter 13** by Henriksen and Einvik describes one of the ways of introducing siRNA into cells. The procedure involves construction of vectors expressing short-hairpin RNA (shRNA) that are processed into siRNA by the cellular RNAi machinery. Detailed descriptions of target site selection, shRNA construction, shRNA transfection, and target knockdown validation are provided. The most obvious method for validation of target knockdown is quantitative RT-PCR, also known as real-time PCR. Josefsen and

Lee (**Chapter 14**) describe the application of a very general method for quantitation of RNA in a sample. The chapter includes other general protocols, e.g., on RNA isolation and cDNA synthesis.

Northern blotting, nuclease protection, and qRT-PCR are used to analyze the steady-state level of RNA. Chromatin immunoprecipitation (ChIP) using RNA polymerase II antibodies is a technique that in combination with measurements of mRNA levels can be used to measure transcription rates as an alternative to the cumbersome nuclear run-on method. Nelson and colleagues have developed a fast version of ChIP outlined in **Chapter 15**. ChIP is a general method that can be used with antibodies raised against other components of chromatin to provide a detailed description of the chromatin state of individual genes.

Part II: RNP Methods

The second part opens with an introduction to the post-transcriptional operon by Tenenbaum and colleagues. The mRNAs, and probably also the non-coding RNAs, are associated with protein factors throughout their lifetime. Some remain stably bound to the RNA while others are exchanged. The proteins are involved in coupling the various steps in the processing of genetic information. Transcription factors influence the pattern of splicing, and splicing factors influence translation. Ultimately, the associated proteins dictate the cytoplasmic fate of the mRNAs. Thus, a description of the structure of mRNAs and non-coding RNAs is very incomplete without a description of their protein partners. The post-transcriptional operon is a set of monocistronic mRNAs encoding functionally related proteins that are co-regulated by a group of RNA-binding proteins. The model is used to describe data from an assortment of methods (e.g., RIP-Chip, CLIP-Chip, miRNA profiling, ribosome profiling) that globally address the functionality of mRNA. Thus, the conceptual **Chapter 16** is followed by **Chapter 17**, by Jain and colleagues from the Tenenbaum lab, describing RIP-Chip analysis in which an antibody directed toward an RNA-binding protein is used to pull-down a collection of mRNAs that are subsequently identified by microarray analysis.

A different approach to the same problem is taken by Jönsson and colleagues in **Chapter 18**. Here, a tag (FLAG-tag) is attached to the RNA-binding protein that is expressed at endogenous levels under tetracycline control. The tag is used as a handle for immunoprecipitation of RNP granules that are visualized by atomic force microscopy. Like in Chapter 17, the RNA can be recovered from the granules, and the RNA content is subjected to microarray or deep sequencing analysis. Further characterization of RNPs as well as the detailed characterization of binding of individual proteins to RNA frequently involves analysis by electrophoretic mobility shift assay (EMSA), a technique that is also known for the characterization of DNA-binding proteins. Gagnon and Maxwell have refined this technique for protein-RNA complexes and demonstrate its usefulness in **Chapter 19**.

The steady-state levels of mRNA and protein are poorly correlated for a large fraction of genes. Polysome profile analysis is a method that can be used to study the translation status of cells and to isolate and characterize mRNAs actively engaged in translation. In **Chapter 20**, Masek and colleagues introduce translational control and present methods for sucrose-gradient-based analysis of polysomes followed by extraction of RNA suitable

for a wide-range of downstream applications, including microarray and qRT-PCR. miRNAs are mostly, but not exclusively, involved in translational repression. In the context of the post-transcriptional operon model, they can be considered formally equivalent to RNA-binding proteins. Many research projects involve miRNA profiling with the aim of identifying particular miRNAs that are up- or downregulated followed by a search for the targets of those identified miRNAs. Target-finding has proven to be one of the major challenges in bioinformatics. In **Chapter 21**, Lindow gives guidelines on how to use the existing tools for target-finding.

Many of the protocols described in this volume end with a sample for subsequent analysis by high-throughput technologies, such as deep sequencing or microarray analysis. In many research institutions, the options are to have this analysis done in a core facility or as a commercial service. **Chapter 22** provides some hints to the non-specialist with respect to choice of analytical tool and sample preparation for the outsourcing of experiments.

One of the surprises of the human genome project was the small number of genes (ca. 25,000) identified compared to that of, say, fruit flies (14,000) and nematodes (19,000). The new insights have challenged the concept of the gene and shown that a simple counting of the number of genes completely misses the point in understanding the complexity of an organism. The new view on the transcriptional landscape and the appreciation of the role that proteins play in the processing and interpretation of genetic information can account for many more products and much more sophisticated regulatory networks than the traditional DNA view. It is our hope that this volume will help researchers to reveal many new examples of this.

Finally, I would like to thank the authors for their contributions and for their patience during the preparation of this volume. Special thanks go to the editors at MiMB, who have been very supportive. One of the characteristics of the contributions to MiMB is the solidarity among scientists that is expressed in the willingness by the authors to share protocols and the very direct advice that is given in the extensive notes sections. It is my sincere hope that this volume lives up to the tradition.

Henrik Nielsen

RNA

Methods and Protocols

Nielsen, H. (Ed.)

2011, XV, 329 p. 53 illus. With online files/update.,

Hardcover

ISBN: 978-1-58829-913-0

A product of Humana Press