Kidney International, Vol. 67 (2005), pp. 2114–2122

PERSPECTIVES IN BASIC SCIENCE

Transcriptome analysis and kidney research: Toward systems biology

MINGYU LIANG, ALLEN W. COWLEY, JR., MARTIN J. HESSNER, JOZEF LAZAR, DAVID P. BASILE, and JENNIFER L. PIETRUSZ

Department of Physiology; Department of Pediatrics; Department of Dermatology; and Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, Wisconsin

Transcriptome analysis and kidney research: Toward systems biology.

An enormous amount of data has been generated in kidney research using transcriptome analysis techniques. In this review article, we first describe briefly the principles and major characteristics of several of these techniques. We then summarize the progress in kidney research that has been made by using transcriptome analysis, emphasizing the experience gained and the lessons learned. Several technical issues regarding DNA microarray are highlighted because of the rapidly increased use of this technology. It appears clear from this brief survey that transcriptome analysis is an effective and important tool for question-driven exploratory science. To further enhance the power of this and other high throughput, as well as conventional approaches, in future studies of the kidney, we propose a multidimensional systems biology paradigm that integrates investigation at multiple levels of biologic regulation toward the goal of achieving a global understanding of physiology and pathophysiology.

High throughput analysis of gene expression at the mRNA level in a genome or near-genome scale (i.e., transcriptome analysis) has become one of the most widely used approaches in the genomic era of biomedical research. The development of these techniques, represented most prominently by DNA microarray, was embraced with great excitement and with promises to revolutionize biomedical research, including kidney research. The rapidly increasing popularity of these techniques is evidenced by more than 2000 *PubMed* entries that contained the term "microarray" in the year 2003 alone, less than 10 years after the introduction of DNA microarray techniques [1, 2].

Key words: gene expression, microarray, renal physiology, renal disease, genome.

Received for publication April 27, 2004
and in revised form July 9, 2004, and August 13, 2004
Accepted for publication February 9, 2005

While an enormous amount of data has been generated using transcriptome analysis, how much this data has meaningfully advanced our understanding of function and disease remains an open question. It is worthwhile at this time to reflect on the strengths and the weaknesses of this approach and how it may be better utilized in future studies. In this brief review article, we first describe the technical principles and characteristics of major techniques used for transcriptome analysis. We then summarize the progress in kidney research that has been made by using these techniques, emphasizing the experience gained and the lessons learned. Several technical issues regarding DNA microarray are highlighted. Finally, ways in which these techniques can be better applied in future studies are considered.

It appears clear from this brief survey that transcriptome analysis is an effective and important tool for question-driven exploratory science [3]. It is apparent, however, these approaches will be more powerful when they are appropriately integrated with proteomics, functional measurement, and other approaches into a multidimensional systems biology paradigm for the study of physiology and pathophysiology.

TRANSCRIPTOME ANALYSIS: WHY?

Biologic function at any given moment is determined by a myriad of factors such as ions, lipids, and catalytic activity of enzymes, nearly all of which involve, either directly or indirectly, the expression level of certain genes. Consequently, alterations of biologic function, especially sustained ones that occur in many diseases, are often the result of changes in gene expression levels.

In the simplified biologic axis of DNA-mRNA-protein function, protein is the final product of gene expression and the closest determinant of function. Intuitively, protein analysis would render measurement of mRNA (and DNA for that matter) unnecessary. Moreover, it is known that mRNA abundance does not always correspond to

protein abundance [4–6], which suggests one should be cautious when extrapolating mRNA to protein. It should be noted that quantitative estimates of the extent of discrepancy between mRNA and protein abundance based on high-throughput assays should be interpreted cautiously. This is because substantial noise exists in the measurement at each level, particularly when the abundance of mRNA and/or protein is low.

Even with some discrepancies between the abundance of mRNA and protein, however, measurements of mRNA are still very valuable for two major reasons. First, current techniques for quantifying protein, compared to mRNA, are more limited in throughput and possibly in accuracy and sensitivity. In many cases, mRNA levels are the best surrogates available for protein abundance. Second, mRNA analysis could provide mechanistic insights complementary to protein analysis, which can be significant not only scientifically, but also clinically. The clear value of mRNA analysis, however, by no means implies that mRNA measurement can replace protein measurement. The issue of multilevel analysis will be discussed later.

Techniques such as Northern blot, ribonuclease protection assay, and competitive reverse transcription-polymerase chain reaction (RT-PCR) are very useful for quantifying the mRNA expression levels of one or a few genes at a time. The more challenging goals, however, are to identify unsuspected changes in mRNA abundance and to obtain a global view of biologic regulation likely involving hundreds or more genes. Techniques allowing high-throughput analysis of mRNA levels would obviously be more appropriate for achieving these goals.

TRANSCRIPTOME ANALYSIS: HOW?

Techniques for high-throughput analysis of mRNA levels can be roughly divided into two categories based on the outcome of the analysis: those yielding estimates of absolute or relative levels of mRNA for all genes examined, and those primarily yielding the identity of a subset of genes whose mRNA levels differ substantially between the samples examined. DNA microarray, serial analysis of gene expression (SAGE), high-throughput real-time PCR, and sequencing of expressed sequence tags (EST) belong to the first category. The second category mainly includes mRNA differential display and subtractive hybridization.

The basic technical principles for each of these techniques are summarized in Figure 1. References containing extensive descriptions of these techniques, including their important variants, are provided in the figure legend.

The major technical advantages and disadvantages of each of these techniques are listed in Table 1. Some of these technical characteristics will be further discussed in the context of application in kidney research.

APPLICATION IN KIDNEY RESEARCH

Cataloging renal transcripts

Generating a comprehensive catalog of the identities and abundance of mRNA molecules is the most unique utility of transcriptome analysis. SAGE is most appropriate for generating such catalogs due to its quantitative nature, potential comprehensiveness, and independence of prior availability of sequence or cDNA clones. EST sequencing and DNA microarray could also be used for this purpose, although their ability to absolutely quantify may be more limited. DNA microarray is further limited by the prior availability of sequence or clones.

Several thousand distinct transcripts, including some previously unknown sequences, have been identified in mouse kidneys using SAGE [20, 21]. The most abundant renal transcripts were found to be those related to mitochondrial metabolism and tubular transport and primarily associated with the proximal tubule. The latter was not surprising since proximal tubules constitute the majority of the kidney mass. mRNA species of lower levels of abundance may not have been represented in these mRNA catalogs.

To obtain transcript profiles in specific nephron segments, a SAGE method using a much smaller amount of starting mRNA samples was developed and applied to mouse and human kidneys [21, 22]. Several hundred transcripts were reported to be differentially expressed among nephron segments. The grouping of nephron segments based on the number of shared transcripts appeared consistent with known functional similarities among nephron segments. A 3′-cDNA sequencing method has also been used to identify and quantify a few thousand transcripts in the proximal tubule and the inner medullary collecting duct [23, 24].

Baseline renal mRNA profiles in human and mouse have also been assessed using DNA microarray [25–28]. Some of these studies emphasized the observation that renal mRNA levels of many genes varied among "normal" individuals. This is consistent with the well-known biologic variability among individuals. However, the reported magnitude of such individual variability might highly depend on how technical variability was assessed and accounted for.

Identifying new regulatory mechanisms

The most common use of transcriptome analysis is to discover new regulatory mechanisms. This is most often achieved through comparing mRNA profiles between interventions, disease states, or developmental stages to identify differentially expressed genes.