How Omics Data Can Be Used in Nephrology

Eugene P. Rhee

Advances in technology and computing now permit the high-throughput analysis of multiple domains of biological information, including the genome, transcriptome, proteome, and metabolome. These omics approaches, particularly comprehensive analysis of the genome, have catalyzed major discoveries in science and medicine, including in nephrology. However, they also generate large complex data sets that can be difficult to synthesize from a clinical perspective. This article seeks to provide an overview that makes omics technologies relevant to the practicing nephrologist, framing these tools as an extension of how we approach patient care in the clinic. More specifically, omics technologies reinforce the impact that genetic mutations can have on a range of kidney disorders, expand the catalogue of uremic molecules that accumulate in blood with kidney failure, enhance our ability to scrutinize urine beyond urinalysis for insight on renal pathology, and enable more extensive characterization of kidney tissue when a biopsy is performed. Although assay methodologies vary widely, all omics technologies share a common conceptual framework that embraces unbiased discovery at the molecular level. Ultimately, the application of these technologies seeks to elucidate a more mechanistic and individualized approach to the diagnosis and treatment of human disease.

Complete author and article information provided before references.

Am J Kidney Dis. 72(1): 129-135. Published online February 23, 2018.

doi: 10.1053/j.ajkd.2017.12.008

Introduction

Although it has undergone some revision, Francis Crick's "central dogma" of molecular biology remains the archetype for the flow of information in biological systems, namely that DNA makes RNA and RNA makes proteins. Some proteins, for example enzymes and transporters, then modulate metabolites. Omics approaches aim to provide comprehensive analysis of a given set of molecules and include genomics (DNA), transcriptomics (RNA), proteomics (proteins), and metabolomics (metabolites).[1,2] This Perspective provides a brief overview of these methodologies, with examples in nephrology research that outline their potential to advance clinical care.

Omics Approaches

Figure 1 depicts the conceptual inter-relationship of omics approaches, highlighting the approximate number of entities (eg, genes, transcripts, proteins, or metabolites) within each domain, and Table 1 outlines relevant analytical techniques. For comprehensive descriptions of underlying methods, the reader is referred to dedicated reviews.

Genomics

The human genome is composed of approximately 3 billion DNA base pairs, <2% of which are dedicated to coding approximately 20,000 genes. Strategies to analyze the genome in human studies vary.[3,4] For genome-wide association studies (GWAS) that can span many thousands of individuals, the most common approach has been to use DNA microarrays to genotype up to about 1 million nucleotides at prespecified positions in the genome known to vary across individuals (also known as single-nucleotide polymorphisms [SNPs]).[5] Using a case-control approach, GWAS aim to identify SNPs that are significantly enriched among individuals with a given disease compared with unaffected controls. These studies causally implicate specific genomic regions in disease pathogenesis, but do not necessarily identify the causal genetic alterations. Instead, the identified SNP is assumed to be in close proximity to the causal alteration in DNA sequence, which may not have been among the prespecified SNPs included on the DNA microarray. An alternative to microarrays is to sequence entire genomes (or key portions of the genome, such as all coding genes), an approach enabled by the development of high-throughput DNA sequencing techniques, often referred to as next-generation sequencing (NGS).[6] These methods provide more comprehensive information about each individual's genome, including rare mutations (DNA microarrays are usually restricted to relatively common SNPs). However, these methods are more costly and computationally demanding and thus are restricted to relatively smaller sample sizes.

Transcriptomics

Given the structural similarities between DNA and RNA, methods for analyzing the genome and transcriptome are similar. As with genomics, transcriptomic capabilities have evolved over time from microarray-based quantitation of prespecified transcripts to comprehensive analysis of all transcribed RNA using NGS, often termed RNA sequencing (RNA-seq). Traditional microarray approaches generally focus on protein-coding messenger RNA (mRNA).[7] Comprehensive sequencing with RNA-seq can provide more granular information, differentiating among mRNA isoforms or splice variants.[8,9] In addition, NGS-based methods have revealed significant expression of noncoding RNAs—RNAs that do not encode proteins—such as microRNAs and long noncoding RNAs that have diverse effects on regulating gene expression.

Proteomics

Proteins perform most of the actual work in cells and tissues and thus span an enormous diversity of form and function as structural proteins, transcription factors,

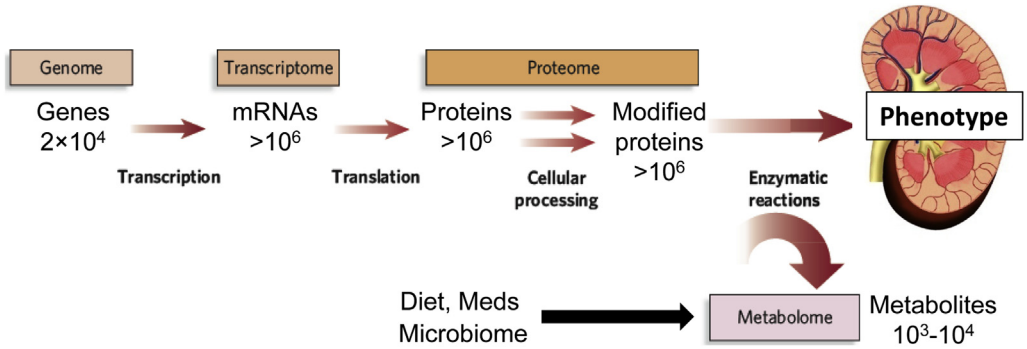

Figure 1. The flow of biological information from the genome to transcriptome, proteome, and metabolome. The estimated number of each type of molecule in a typical cell is indicated. The epigenome, not shown, resides conceptually between the genome and transcriptome; also not shown are various noncoding RNAs that are not transcribed into protein but are increasingly recognized to have important biological roles. Abbreviation: mRNA, messenger RNA. Modified from Gertzen et al[58] with permission of the copyright holder, Springer-Nature.

receptors, antibodies, hormones, ion and solute transporters, and enzymes. To add to this complexity, measuring the relative abundance of proteins provides incomplete information on the proteome. Alternative splice variants from a single gene can yield several distinct proteins, and posttranslational modifications such as phosphorylation and ubiquitination have a profound impact on protein localization, activity, and turnover. In addition, some proteins require specific protein-protein interactions or expression in specific cellular compartments to function properly. At present, different analytical techniques, along with dedicated bioinformatics methods, are required to assess these different facets of the proteome.[10-12]

Metabolomics

The metabolome is the global collection of small molecules, typically <1,500 Da, including carbohydrates, amino acids, organic acids, and lipids.[13-15] Downstream of transcription and translation, metabolite levels reflect gene, transcript, and protein abundance and function, thus providing information that is both complementary and in some cases summative of these other domains. Decades of research in metabolism have placed many metabolites

Table 1. Overview of Omics Methodologies

Methodology	Omics Approaches	Comment
DNA microarray	Genomics (GWAS); microarray-based transcriptomics	Relies on noncovalent bonding between DNA strands that are complementary. DNA sequences of interest (known as probes or oligos) are arranged on a surface. For genomics studies, probes are designed to be specific for SNPs across the genome. For transcriptomics, probes are designed to be specific for different gene transcripts. Experimental samples containing target DNA or cDNA (mRNA that has been reverse transcribed) are hybridized to the microarray and then a variety of techniques can be used to quantitate probe-target capture.
Next-generation sequencing	Genomics; RNA-seq–based transcriptomics	Methods vary, but a common theme is that many sequencing reactions of random DNA segments are performed in parallel, and enzymatic reactions are used to determine nucleotide sequence in real time (eg, using luminescence or fluorescence) rather than requiring gel electrophoresis as with traditional Sanger sequencing. Computational methods align and merge data from the multiple parallel reads to reconstruct the original full-length sequences.
Mass spectrometry	Proteomics; metabolomics	Resolves molecules based on their mass, or more precisely, their mass to charge ratio following ionization. Usually coupled to upfront capillary electrophoresis, liquid chromatography, or gas chromatography, which separate molecules on the basis of hydrophobicity and charge. Proteins require proteolytic cleavage prior to analysis.
Affinity-based assay	Proteomics	Traditionally refers to multiplexing antibodies of known specificity on an array to simultaneously measure multiple proteins. More recently, affinity reagents other than antibodies (such as oligonucleotides) have increased the breadth and throughput of these approaches.
NMR spectroscopy	Metabolomics	Uses magnetic properties of select atomic nuclei (eg, ^{1}H, ^{13}C, or ^{31}P) to determine the structure and abundance of metabolites.

Abbreviations: cDNA, complementary DNA; GWAS, genome-wide association studies; mRNA, messenger RNA; NMR, nuclear magnetic resonance; RNA-seq, RNA sequencing; SNP, single-nucleotide polymorphism.