Published in final edited form as:

Semin Nephrol. 2010 September ; 30(5): 439–442. doi:10.1016/j.semnephrol.2010.07.001.

Integrative biology of renal disease: towards a holistic understanding of the kidney's function and failure

Matthias Kretzler, M.D. [Professor] and

Nephrology/Internal Medicine, Center for Computational Medicine and Bioinformatics, University of Michigan, 1560 MSRB II, 1150 W. Medical Center Dr.-SPC5676, Ann Arbor, MI 48109-5676, Phone: 734-615-5757, Fax: 734-763-0982

Clemens D. Cohen, M.D. [Assistant Professor]

Division of Nephrology and Institute of Physiology, University and University Hospital of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland, Phone: +41-44-635 50 53, Fax: +41-44-635 68 14

Matthias Kretzler: kretzler@umich.edu; Clemens D. Cohen: clemens.cohen@access.uzh.ch

In biomedicine we are currently experiencing a fundamental transition from research focused on the functions of single entities (molecules or pathways) to an integrative biology analyzing biological systems as a unified whole. The need for systems biology approaches in medicine is rapidly expanding as genome wide technology platforms generate demands for multi-level, multi-dimensional analyses to model biological processes. Based on such data novel hypotheses of organ function and failure can be generated and subsequently tested.

The multilayered complexity of renal function mandates an integrative approach. Accordingly, nephrology offers unique opportunities for the transition to integrative biology. The kidney is the key regulator of the internal milieu of any vertebrate organism. To serve this function, the kidney represents a complex ultra-structure consisting of hundreds of thousands of functional units, nephrons, each of them featuring an internal functional heterogeneity serving a multitude of specific functional needs. Nephron function is closely linked with the integrity of the human organism in a bidirectional manner. Systemic metabolic, vascular or immune diseases impact renal function and vice versa; acute or chronic renal failure is one of the strongest predictors of progression of these systemic diseases towards system failure, i. e. death.

Focused scientific efforts have led to the identification of hundreds of essential processes in renal function and failure. This reductionistic approach mandates evaluation of each of these processes in isolation. With the advent of genome wide analytical capabilities it has become feasible to capture readouts of multiple regulatory biological systems at high precision and depth. This “pluralism of causes and effects in biological networks” (1) is at the core of what systems biology aims to observe. To gain this understanding requires a multidisciplinary approach for data generation, analysis and integration with available knowledge. In this issue of “Seminars in Nephrology” nephrologists, pathologists, biologists, bioinformaticians,

biochemists and computer-scientists joined forces to provide their perspective on large-scale data generation and their vision for an integrative understanding of the kidney.

The genome-phenome continuum

Systems biology considers disease development to progress along the genome-phenome continuum: from DNA, via RNA to protein and metabolite to intra-cellular organization, tissue homeostasis and organismal interaction. Along this continuum, a sequence of regulatory cascades integrates the genetic definition of an individual with environmental influences to manifest the phenotype of a human being. In this review series, the major areas of large-scale data generation in renal disease are presented along this continuum from genomic and transcriptomic techniques (2;3) to proteomic (4) and metabolomic data (5). In the March 2010 issue of *Seminars in Nephrology*, we presented strategies to integrate genetic with gene expression data sets (6). Each review will provide a unidimensional description of the current approaches to generate and analyze data sets from one domain followed by strategies to integrate parallel studies along the genome-phenome continuum. These approaches are currently mainly binary, integrating two complex data sets, but with a clear vision towards a multidimensional analysis.

Strategies to interrelate alterations on each level of data generation with the clinical phenotype are introduced and discussed. A plethora of public data resources regarding genes, proteins, metabolites, interactions, pathways, and their phenotypic and histological associations are presented and integrative strategies discussed. Dependencies detected *de novo* inside the data sets can be related with the published biological knowledge for further functional integration. Vice versa, prior biological knowledge might allow us to focus studies on certain sub-domains in the experimental data set (see (7;8)).

The complexities and opportunities of this integration challenge are illustrated in Figure 1, summarizing the multiple levels from which we will need to integrate data for multidimensional analyses. Only the initial few levels of the molecular biology paradigm have been targets by the ‘-omics’ approaches and have been mined with traditional bioinformatics and systems biology tools. A currently emerging challenge is the integration of these molecular data sets with the ‘meso-scale’ level of biological function, i.e. the level of cellular behavior, tissue alterations and regulation of organ function. Defining the impact and dependencies of the molecular studies with comprehensive measures of cell and tissue behavior will be crucial to understand the manifestation of complex clinical phenotypes. The strategies for compartment-specific proteomic studies presented by Velic et al. (4) and the concept of histogenomics introduced by Perco and Oberbauer (9) in this issue of the ‘Seminars’ are a first step in this direction. Work by the group of Eric Schadt has demonstrated the ability to detect interrelationships of regulatory networks across organ barriers (reviewed in (10)) and Ju and Brosius (11) in this issue present approaches to define shared networks or renal disease across species.

To make renal systems biology accessible to the research community

The information contained in large-scale data sets allows key driving forces of biological processes to be defined in an unbiased manner. The same data sets also enable the extraction of information of imminent relevance for ongoing reductionistic research studies and can guide these studies towards areas of particular relevance for human disease. A series of generic tools has been developed by the National Centers for Biological Computing (NCBC) and others (see Ju and Brosius (11) and Perco and Oberbauer (9) for summaries of available tools). These tools can be linked in analytical backbones, which allow addition of tools in a modular manner (i.e. ‘GenePattern’ from Broad (genepattern.org); ‘HIVE’ from i2B2 (i2b2.org/resrcs/hive.html)). An essential pre-requisite for the adoption of systems biology

approaches into biomedical research is the accessibility of large-scale data for the domain expert without systems biology background. Taking advantage of the available bioinformatic tool sets we have developed a renal specific systems biology search engine, Nephromine, specifically targeted towards the renal clinician scientist and biologist (Figure 2). Nephromine combines a rapidly growing compendium of publicly available human renal gene expression profiles, a sophisticated analysis engine, and a web application designed for data mining and visualization of gene expression data. It is freely available to the academic community at Nephromine.org and aims to fuel systems biology in kidney research.

We envision that trans-disciplinary, cross-species and multi-platform approaches, as presented in this issue, will allow the nephrology community to elucidate key networks of biological items relevant for the unified whole of the kidney. The challenge is ours to implement these opportunities in renal research and clinical nephrology.

References

1. Sauer U. Genetics. Getting closer to the whole picture. *Science* 2007;316:550–551. [PubMed: 17463274]
2. Hodgin JB, Cohen CD. Experimental approaches to the human renal transcriptome. *Sem Nephrol.* 2010
3. Mohtat D, Susztak K. Fine Tuning Gene Expression: The Epigenome. *Sem Nephrol.* 2010
4. Velic A, Macek B, Wagner CA. Towards quantitative proteomics of organ substructures: implications for renal physiology. *Sem Nephrol.* 2010
5. Wang JH, Byun J, Pennathur S. Analytical Approaches to Metabolomics and Applications to Systems Biology. *Sem Nephrol.* 2010
6. Keller B, Martini M, Sedor J, Kretzler M. Linking variants from genome wide association analysis to function via transcriptional network analysis. *Sem Nephrol* 30:177–184. in press.
7. Nelson PJ, Werner T. Pathways and promoter networks analysis provides systems topology for systems biology approaches. *Sem Nephrol.* 2010
8. Greene CS, Troyanskaya OG. Integrative Systems Biology for Data Driven Knowledge Discovery. *Sem Nephrol.* 2010
9. Perco P, Oberbauer R. Integrative analysis of ‘omics’ data and histological scoring in renal disease and transplantation: renal histogenomics. *Sem Nephrol.* 2010
10. Schaad EE. Molecular networks as sensors and drivers of common human diseases. *Nature* 2009;10:218–223.
11. Ju W, Brosius FC 3rd. Understanding Kidney Disease: Toward the Integration of Regulatory Networks Across Species. *Sem Nephrol.* 2010

Figure 1.
Multidimensional integration of large-scale data sets along the Genome – Phenome continuum integrating systems biology and patho- physiology

Fig. 2. Structure and capabilities of Nephromine

The Nephromine database consists of three layers: data input, data analysis, and data visualization. The data input layer has two components, the microarray data pipeline and the annotation data warehouse. The microarray pipeline is used internally to identify and prioritize microarray studies in the literature. The pipeline also draws data directly from the Stanford Microarray Database and the NCBI Gene Expression Omnibus. The data-analysis layer consists of sample facts standardization and automated statistical analysis. The sample facts standardization utilizes the NCI Thesaurus and manual annotation. The automated statistical analysis component is implemented in Perl and R. A series of scripts monitors the database for new data and sample parameters and automatically performs differential expression analysis, cluster analysis, and concept analysis, when needed. The Nephromine web servers query data from the Nephromine database and display tabular and graphical representations of the data and analysis results.