Manuscripts, Articles, Book Chapters and Other Papers

Techniques and Approaches to Genetic Analyses in Nephrological Disorders.

Laurel K. Willig
Children's Mercy Hospital

Recommended Citation

Techniques and Approaches to Genetic Analyses in Nephrological Disorders

Laurel K. Willig[1,2,3]

[1] Center for Pediatric Genomic Medicine, Kansas City, Missouri, United States
[2] Division of Nephrology, Children's Mercy Hospital, Kansas City, Missouri, United States
[3] Department of Pediatrics, University of Missouri-Kansas City, Kansas City, Missouri, United States

Address for correspondence Laurel K. Willig, MD, Division of Pediatric Nephrology, Children's Mercy Hospital, 2401 Gillham Rd, Kansas City, MO 64108, United States (e-mail: lkwillig@cmh.edu).

Abstract

Keywords
► next generation sequencing
► nephrology
► proteomics
► genomics
► metabolomics
► epigenomics
► transcriptomics

Inherited renal disease is a leading cause of morbidity and mortality in pediatric nephrology. High throughput advancements in genomics have led to greater understanding of the biologic underpinnings of these diseases. However, the underlying genetic changes explain only part of the molecular biology that contributes to disease manifestation and progression. Other omics technologies will provide a more complete picture of these cellular processes. This review discusses these omics technologies in the context of pediatric renal disease.

Introduction

In 1985, researchers identified the first renal disease gene, polycystic kidney disease 1 (*PKD1*), responsible for autosomal dominant polycystic kidney disease (ADPKD), but renal gene discovery progressed slowly prior to sequencing of the human genome and development of high throughput next generation sequencing (NGS).[1,2] Currently, there are more than 1,200 genes associated with renal disease and development, largely owing to advancements in genomic technology.[3] This collection of genes includes more than 160 genes associated with rare inherited kidney diseases that contribute greatly to morbidity and mortality in pediatric renal disease.[1,4] Investigators and clinicians are using this technology to diagnose Mendelian diseases such as Alport syndrome, search for novel pathogenic genes in diseases such as focal segmental glomerulosclerosis (FSGS), and identify risk genes in hypertension.[5–8] These genetic discoveries allow for further elucidation of the mechanistic underpinnings of inherited kidney disease.

However, genomics and NGS technology address only one aspect of the molecular biology. Other high throughput omics technologies attempt to address the missing pieces and complete the biological story. The high throughput methods of epigenomics and transcriptomics utilize basic NGS principals to describe regulation of genomic information, while proteomics and metabolomics use mass spectrometry (MS) to explore the downstream effects. This knowledge promises to improve prognostication and will translate into more effective treatments for children afflicted with these diseases.[1,9] However, the genetic complexity of many of these diseases coupled with immature high throughput research methodology may lead to improper application and inaccurate conclusions.[10]

This review discusses the available omics techniques, highlighting weaknesses and illustrating their use in pediatric nephrology research. It will not review microarray technology but will instead focus on NGS and MS high throughput methods. In addition, it will briefly discuss systems biology for high throughput data analysis and some of the model

systems used to test these integrative approaches in pediatric nephrology.

General Study Design Considerations for Omics Data

Biostatistical Considerations

Each omics field examines a specific aspect of a molecular pathway and as a result, multiple different techniques exist to study each element (►**Table 1**). These techniques require unique biostatistical methods and software to analyze the data. A detailed review of these methods and software is beyond the scope of this article, but this information may be found elsewhere.[10–22] However, researchers should consider some basic principles when using omics data. All experiments should apply appropriate power analyses prior to implementation. Power calculations must consider the genomic or biologic space interrogated to determine adjustments in the false discovery rates or p values due to multiple testing.[10,22–24] This adjustment relies on understanding the interdependency of the different tests which is often difficult to estimate. Thus, researchers must decide how conservative to be in their adjustment of type I error rates.[11,23] In genomic studies, researchers must consider the allele frequency, the effect size of a variant, and the genetic and phenotypic heterogeneity of the trait of interest when considering power estimates.[10] For omics studies of quantitative traits, power calculations should include consideration of the biological and measurement variability. Finally, the testing software selected must accommodate the massive amount and format of data generated from the experiments and be specific for the type of omics question being answered.[11,24,25]

In addition to these common considerations, there are specific considerations unique to each high throughput method. Researchers must consider genomic unit for comparison as this determines the number of tests performed and is important for the choice of downstream analysis.[23] In genomic association studies of rare variants for complex traits, rare variants are often collapsed into units based on gene parameters, linkage disequilibrium, or simply by proximity using sliding windows of a particular length.[11] Unlike genomic studies with low intra-sample variability due to homogeneity of cells and the relatively static nature of DNA, the other omics fields have much more intra-sample and inter-sample variability. Researchers must consider a statistical test's ability to capture and describe this variability when selecting their statistical approach.[12] Many of these studies are measuring quantitative traits. The addition of more samples will help improve some of the technical biases of the study and decrease variability for the measured trait. The addition of sample replicates will help control for within-sample biases.[12,16,18,19,22]

Utilizing Preexisting Datasets

Many omics studies use previously generated datasets for additional affected individuals, as validation cohorts, or as reference datasets due to the significant costs of generating data and the rarity of some conditions. There are many publicly accessible databases/repositories and ongoing initiatives to establish more (►**Table 2**). However, this pooling of data from multiple sites introduces the possibility of population stratification and systematic differences in high throughput methods and quality control.[10]

Phenotypic data plays an important role in identifying the underlying cause and the affected molecular pathways that cause inherited renal disease.[26] Often phenotypic data stored in databases and repositories are sparse or poorly categorized and thus may confound study results.[10] In attempts to improve omics data sharing and transparency, experts are establishing guidelines for both study publication and

Table 1 Overview of omics data and experimental approaches

Technology	Type of data	Common approaches	References
Genomics	Gene number, structure, function, and organization	Targeted gene panels, WES, WGS	32
Epigenomics	Regulatory elements: transcription factor binding sites, methylation patterns, nucleosome location, chromatin accessibility, histone modification	ChIP-seq, Methyl-seq, MethylC-seq, MeDip-seq, MBD-seq, ChIA-PET, ATAC-seq, FAIRE-seq, DNase-seq, MNase-seq	61,66
Transcriptomics	All RNA species including regulatory RNA, gene expression levels, posttranscriptional modifications	RNA-seq, NET-seq, CAPTURE-seq CLIP-seq, iCLIP, and PAR-CLIP	74,100
Proteomics	All proteins, protein–protein interaction	LC-MS	101
Metabolomics	Peptides, amino acids, nucleic acids, and organic acids	LC-MS	91

Abbreviations: ATAC-seq, assay for transposase-accessible chromatin using sequencing; CAPTURE-seq, RNA capture sequencing; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; ChIP-seq, chromatin immunoprecipitation sequencing; CLIP-seq, cross-linking immunoprecipitation sequencing; DNase-seq, DNase I sequencing; FAIRE-seq, formaldehyde-assisted isolation of regulatory elements; iCLIP, individual-nucleotide resolution cross-linking immunoprecipitation; LC-MS, liquid chromatography-mass spectrometry; MDB-seq, methylated DNA binding domain sequencing; MeDip-seq, methylated DNA immunoprecipitation sequencing; MNase-seq, micrococcal nuclease sequencing; NET-seq, native elongating transcript sequencing; PAR-CLIP, photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation; WES, whole exome sequencing; WGS, whole genome sequencing.

Table 2 Common and nephrologic databases with available omics datasets

Database	Data	Reference
General		
1000 Genomes	Low coverage human genome datasets	102
MGI Mouse Genome	Integrated genetic, genomic, and biologic information from laboratory mice	103
Human Epigenome Project	DNA methylation in human tissues	104
ENCODE	Functional genomic elements in multiple human tissues	105
Mouse Reference Transcriptome Project	Gene expression data on multiple mouse tissues	106
PRIDE	Proteomic datasets from large number of organisms	107
Human Metabolome Database	Detailed metabolites in the human body	108
Nephrology		
European Renal cDNA Bank	Gene expression data from renal biopsies	109
EURenOmics	Biorepository and phenotypic data for rare kidney diseases	110
GUDMAP	Gene expression datasets for genitourinary tract of mice	111,112
NEPTUNE	Biorepository and phenotypic data on patients with proteinuria	113

Abbreviations: ENCODE, Encyclopedia of DNA Elements; GUDMAP, Genitourinary Development Molecular Anatomy Project; NEPTUNE, Nephrotic Syndrome Study Network; PRIDE, PRoteomics IDEntifications database.

database deposition that will hopefully help clarify these issues.[10,27–30] Without this clarification, systematic confounding of data introduced by sequencing platforms and bioinformatics pipelines will lead to misinterpretation of results.

Basics of Next Generation Sequencing

NGS also known as massively parallel sequencing revolutionized the field of genomics. These methods create the foundation for a wide variety of experimental approaches in genomics, epigenomics, and transcriptomics. The basic approach to NGS starts with DNA fragmentation and platform-specific adapter ligation to create a DNA library, which can be amplified later using polymerases if desired. The sequencing reactions then take place in situ on a solid surface containing complimentary sequence to the adapters. This in situ sequencing generates a massive number of sequencing reads simultaneously (i.e., in parallel), vastly improving the efficiency compared with Sanger sequencing.[31] NGS platforms perform sequencing and nucleotide detection in one fluid process; nucleotides are added, a detection step is performed, a wash step occurs, and the process repeats itself.[31,32] These improvements decrease the per base sequencing costs and time compared with Sanger sequencing, allowing for larger adoption of NGS in research.[4]

Sequencing Platforms

All of the NGS high throughput methods require choice of a sequencing platform. Most platforms have built-in flexibility to perform sequencing for genomic, epigenomic, and transcriptomic experiments. While all NGS technologies have relatively high per base sequencing error rates and rely on

increased coverage to overcome these errors, recent comparisons of commonly used platforms demonstrate variable performance.[33,34] Common problems with this technology include short read lengths causing issues with alignment in repetitive regions of the genome, poor coverage in GC rich areas, and amplification bias for those technologies that use targeted enrichment (►**Table 3**).[33–37] Combining data from multiple platforms surmounts some of the weaknesses inherent to different sequencing methods.[33]

Modifications to the basic NGS workflow that are employed to study regulatory elements and RNA lead to additional biases largely in the generation of DNA fragments for sequencing and library preparation. For instance, fragmentation of chromatin for CHIP-seq occurs in a biased fashion with euchromatin being sheared more readily than heterochromatin. In RNA sequencing, commonly occurring biases include low complexity of transcripts, uneven transcript coverage, and antisense artifacts. There are methods that may be employed during library preparation and enrichment to help overcome some of the weaknesses.[38]

Bioinformatics Platforms

In addition to the variety of sequencing platforms, multiple bioinformatics pipelines exist to filter the raw data and create the data files used for analysis. In the most simplistic terms, NGS bioinformatics pipelines consist of an aligner and a variant caller. During alignment, each read is aligned to the best-fit reference sequence using a statistical algorithm and is assigned a mapping quality score.[39] These alignments are then fed to the variant caller for assignment of variant presence and zygosity. In-depth comparisons have revealed significant differences in the overall output as well as the

Table 3 Overview of next generation sequencing techniques

	Description	Strengths	Weaknesses	Reference
Platform				
Illumina HiSeq/Miseq	Reversible terminator sequencing	• Less GC biases in GC poor areas • More uniform coverage • Better coverage in simple repeat areas • High sensitivity • Low sample-to-sample variation	• GC biases in GC rich areas • Short read length	35,36,114
Complete genomics	DNA nanoball sequencing	• Less GC bias in GC rich areas • Highest total number of bases covered • High sensitivity	• GC biases in GC poor areas • Less uniform genome coverage • Large amount of DNA required	36,114
ABI/SOLiD 4	Ligation	• Most specific in SNP calling	• Significant GC bias • Less uniform coverage and smallest percentage of genome coverage	36,114
Ion Torrent	Proton detection	• Fast • Longer read length than many technologies	• Higher error rates • GC biases	35,36
Alignment				
Burrows Wheeler Aligner	Burrows–Wheeler algorithm based	• Fast • Memory efficient • Quality score to discard poorly aligned reads	• Sensitivity highly dependent on quality of sequence	40,115
Novoalign	Hash table-based approach	• Good for long indels if mapping quality is good • Performs well with variety of variant callers	• Sensitivity highly dependent on quality of sequence	115
GSNAP	Hash table-based approach—genome	• Genome indexing tool • Detects complex variants and splicing in individual reads	• Higher error percentage on mapped reads requiring filtering	115
Bowtie	Burrows–Wheeler algorithm based	• Highest quality reads • Fast	• Low SNP detection rate	40,115
Variant Caller				
GATK	Gold standard	• Gold standard for SNP calling • Most accurate for variants • High sensitivity and positive predictive value	• Time intensive	43,44
SAM tools	Manipulates alignments in SAM format	• Accurate at lower read depths	• Bias toward transition events • Performs poorly using multiple sample algorithm	43,44

Abbreviations: GATK, genome analysis toolkit; GSNAP, Genomic Short-read Nucleotide Alignment Program; SAM, sequence alignment/map; SNP, single nucleotide polymorphism; SOLiD, sequencing by oligonucleotide ligation and detection.

computational complexity and cost for different bioinformatics pipelines (►**Table 3**).[39–45] Epigenomic and transcriptomic analyses require additional data processing steps introducing additional levels of complexity to the bioinformatics pipeline.[46]

Genomics

Approach

The researcher generally selects one of three approaches for sequencing, namely, targeted panels, whole-exome

sequencing (WES), or whole-genome sequencing (WGS) (►**Table 4**). Targeted gene panels sequence a select number of genes and are often customized by the investigator to include candidate genes. These genes may have been previously implicated in disease, part of the same biologic pathways as disease genes, or essential to the function of the tissue or cell of interest.[47] WES is technically a targeted panel, but, due to its inclusion of all protein coding regions of the genome, it allows for a more all-inclusive approach than a targeted panel. It can examine known genes of interest as well as novel candidates. WGS identifies variants at almost all sites in the genome. It is the most inclusive testing for the identification of disease-causing variants, as it includes non-coding sequence important to gene regulation.[39] Often family data including linkage analysis or prior genome-wide association studies results are utilized in data analysis to improve results in genomic testing.[48,49]

In analyzing data from such experiments, researchers often consider previous evidence for genetic pathogenicity. Notably, gene pathogenicity and variant pathogenicity are two different things and must be considered independently by investigators. Recent studies have shown that the literature is rife with false-positive genetic associations and many poorly validated pathologic variants are propagated in the literature.[10,50,51] These misclassifications have led to recent attempts to establish more rigorous guidelines for variant classification including study design improvements, increased data sharing, and more comprehensive databases.

Applications
In nephrology, researchers have applied targeted NGS panels and WES to study both Mendelian disease and recent more complex diseases.[1] For example, FSGS, a leading cause of steroid-resistant nephrotic syndrome in children, is a genetically complex disease. Brown et al utilized linkage analysis in a large pedigree of individuals affected by familial FSGS which identified 17 areas on multiple chromosomes that had a logarithm of odds score of >1. They then used WES in five affected individuals and one unaffected individual to identify a mutation in LIM homeobox transcription factor 1 β (*LMX1B*). They discovered another family with a mutation in *LMX1B* in their validation cohort of 91 patients with familial FSGS.[52] Other studies used target gene panels to identify pathogenic or likely pathogenic mutations in patients with steroid-resistant nephrotic syndrome.[7] Other diseases undergoing extensive study with NGS include congenital anomalies of the kidney and urinary tract (CAKUT), nephronophthisis and other polycystic kidney disease, IgA nephropathy, and atypical hemolytic uremic syndrome.[6,53–56]

Epigenomics

Epigenomics is the study of the regulatory elements that control gene expression using high throughput technology. These elements are heritable and also reversible at the cellular level. Regulatory elements control many developmental processes and modification of these elements cause disease.[57,58] As these elements have diverse physical and chemical properties, many different methods have been developed to optimize the epigenome mapping resolution and overcome weaknesses inherent to the individual methods (►**Table 4**).[59–61] Commonly studied elements include histone modifications and DNA methylation. Histone modifications are important in the regulation of chromatin architecture and transcription. More than 100 different types of histone modification exist, including acetylation, methylation, and phosphorylation, all of which have been associated with transcription regulation.[62–65] Chromatin immunoprecipitation sequencing (CHIP-seq) is the most common technique used to study histone modifications and common issues with this technique include antibody affinity, GC biases, and sample bias due to biased fragmentation.[61,63,66] DNA methylation refers to the methylation of cytosine especially in CpG islands. DNA methylation plays a role in tumorigenesis and aging. More importantly, it is instrumental in mammalian cell differentiation and development.[63,67] Researchers use multiple different methods to describe the DNA methylome.

Approach
Unlike genomics which studies DNA specifically, epigenomic techniques must describe the architecture of chromatin, the molecules that alter that architecture and how these traits differ by cell type and change with environmental exposures and/or time. Thus, investigators have developed multiple different approaches to help address these questions. These techniques suffer from the same sequencing biases as those of genomic studies plus additional ones associated with the presequencing procedures (►**Table 4**). Researchers may utilize these methods with either observational or experimental designs. Traditional epidemiology study designs need modifications but may be used with epigenomic methods. Some specific observational study designs include birth cohorts, twin studies, and prenatal cohorts.[68,69] Experimental approaches often examine changes in regulatory elements before and after toxin or drug exposure in functional assays.

Application
Researchers have identified a role for both histone modification and DNA methylation in many renal diseases.[70] For instance, genome-wide studies of methylation in subjects with ADPKD reveal hypermethylation of *PKD1* and other genes involved in ion transport and cell adhesion. This hypermethylation led to decreased expression of these genes and suggests a role for epigenetics in cystogenesis that may explain the patchy nature of cyst formation.[71] Histone modification may also play a role in cyst formation. Knocking out histone deacetylase 5 in *pkd2* −/− mice led to a decrease in cyst formation, suggesting a role for histone deacetylases in cystogenesis.[72] These examples show that even in diseases with well-defined genetic pathogenicity, epigenetic mechanisms may improve understanding of disease progression and suggest avenues for new therapeutics. Other active areas of epigenetic study in nephrology include renal development, chronic kidney disease progression, and autoimmunity.[26,70,73]

Table 4 Next generation sequencing–based approaches and considerations for implementation

	Description	Genomic target	Considerations	Reference
Genomic				
Targeted gene panels	DNA library preparation, followed by targeted enrichment with probes designed against only genes of interest, MPS	DNA	• Less potential for novel gene discovery • Missing coverage in genes of interest due to gaps in probe design or biases during enrichment	116
WES	DNA library preparation, followed by targeted enrichment with probes designed for entire exome (many commercially available exome kits), MPS	DNA	• Missing coverage in genes of interest due to gaps in probe design or biases during enrichment • Potential for incidental findings	39,116
WGS	DNA library preparation and MPS	DNA	• Less coverage than whole exome or targeted panels • Potential for incidental findings	39,116
Epigenomic				
ChIP-seq	Cross-link proteins to DNA, followed by fragmentation, immunoprecipitation, amplification of DNA, and MPS	Protein–DNA interactions Histone modification	• Chromatin configuration introduces sample-specific biases • Amplification biases	61,66
MNase-seq	Use of micrococcal nuclease to digest DNA until it hits an obstruction like nucleosome, then size selection of fragments the size of nucleosomes, amplification of DNA and MPS	Nucleosome identification and quantification	• Cleavage biases for AT-rich sequences • Size selection biases • Variable TF binding characteristics affects enzymatic cleavage	66
DNase-seq	DNase I digests unprotected DNA, followed by size selection, amplification, and MPS	Maps open chromatin	• Efficiency for TF binding sites varies with fragment size • Cleaves DNA in a sequence-specific manner • Variable TF binding characteristics affects enzymatic cleavage	61,66
ATAC-seq	Transposase cuts DNA while simultaneously incorporating adapters into areas of open chromatin followed by PCR and MPS	Factor occupation, nucleosome position, and chromatin accessibility	• Cleavage biases based on sequence • Mitochondrial DNA contamination • Need for deep sequencing	66,117
FAIRE-seq	Cross-linking of protein and DNA with formaldehyde in vivo followed by sonication, phenol-chloroform extraction and MPS	Chromatin accessibility	• Low signal-to-noise ratio making computation difficult • Need for very deep sequencing • Fixation efficiency affects results	61,66,118
MethylC-seq	DNA fragmentation followed by ligation of methylated adapters, size selection using	DNA methylation	• Bias toward GC rich areas of genome but not necessarily CpG islands	119

(*Continued*)

Table 4 (*Continued*)

	Description	Genomic target	Considerations	Reference
	electrophoresis, bisulfate conversion, amplification and MPS		• Mappability of short reads	
ChIA-PET	Cross-linking protein and DNA, followed by sonication, immunoprecipitation, ligation of linkers with MME1 endonuclease, digestion with MME1, pull down of biotinylated linkers with streptavidin beads, amplification of DNA tags, and mapping to reference genome	Long-range chromatin interactions	• Limited by available antibodies • Inapplicable to repetitive regions of genome • Identifies transcription factor binding sites but cannot identify the specific proteins in the complex • Multiple different analysis programs providing different results	61
Transcriptomics				
RNA-seq	RNA isolated, fragmented, and copied into cDNA	Total RNA or subpopulations including mRNA, miRNA, siRNA, lncRNA	• Need replicates • GC biases • Less accurate at transcription start sites	120
CLIP-seq	Cross-links RNA and protein, protein pulled down by immunoprecipitation, protein is removed and RNA copied to cDNA, then MPS	RNA–protein interactions	• Need moderate amount of input material • Low efficiency of cross-linking • Potential bias due to cross-linking	78

Abbreviations: ATAC-seq, assay for transposase-accessible chromatin using sequencing; CAPTURE-seq, RNA capture sequencing; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; ChIP-seq, chromatin immunoprecipitation sequencing; CLIP-seq, cross-linking immunoprecipitation sequencing; DNase-seq, DNase I sequencing; FAIRE-seq, formaldehyde-assisted isolation of regulatory elements; iCLIP, individual-nucleotide resolution cross-linking immunoprecipitation; LC-MS, liquid chromatography-mass spectrometry; MDB-seq, methylated DNA binding domain sequencing; MeDip-seq, methylated DNA immunoprecipitation sequencing; MNase-seq, micrococcal nuclease sequencing; MPS, massive parallel sequencing; NET-seq, native elongating transcript sequencing; PAR-CLIP, photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation; TF, transcription factor; WES, whole exome sequencing; WGS, whole genome sequencing.

Transcriptomics

Transcriptomics encompasses all testing that identifies and quantifies RNA species. While microarray has traditionally been used for gene expression, RNA-seq's use has increased dramatically over the past few years due to its low background noise, ability to identify novel transcripts, a large dynamic range, and accuracy.[59] The development of new high throughput methods has also enabled researchers to examine RNA: protein interactions providing further information about RNA's role in gene expression.

Approach

Common experimental designs that utilize NGS technology include RNA-seq and cross-linking immunoprecipitation sequencing (CLIP-seq) (►**Table 4**). In RNA-seq, the sequencing process is similar to genomic sequencing but with some significant distinctions. RNA-seq's input is total RNA. Typically, during library preparation, RNA is separated by type using oligonucleotides attached to magnetic beads. Reverse transcription converts RNA species into cDNA which is then sequenced using NGS technology. RNA-seq utilizes NGS-

based techniques to identify the spectrum of RNA species, identify allelic imbalance, catalog alternative splicing patterns and posttranslational modifications, as well as quantify RNA. Study designs may be similar to genomic studies when identification of variants and transcripts are of interest. RNA-seq is also commonly implemented for gene expression studies allowing for similar epidemiological approaches and experimental designs to that of other regulatory elements.[74] RNA-seq has several weaknesses including transcript bias introduced by sample preparation (3′ bias from poly A tail selection), the necessity of deep sequencing to capture lower expressed genes, and possibility of base errors introduced during reverse transcription.[75,76]

RNA binding proteins (RBP) regulate posttranscriptional expression. There are now several different high throughput methods for studying RBPs including UV individual-nucleotide resolution cross-linking immunoprecipitation (iCLIP), RNA co-immunoprecipitation, and photoactivatable-ribonucleoside-enhanced CLIP (PAR-CLIP). CLIP-seq is most commonly employed to study RBPs.[77] In this approach, RBPs are cross-linked to RNA using ultraviolet light. These complexes are purified using immunoprecipitation. The unlinked

proteins and RNA are digested leaving only a peptide bound to the RNA. Then reverse transcription is performed and when the RNA transcriptase encounters the peptide, a mutation occurs or transcription prematurely stops. Bioinformatics pipelines then determine the sites of these mutations and where the truncated RNA aligns to identify the location of RBPs.[78]

Application

Both RNA-seq and CLIP-seq are relatively new technologies and remain underutilized in pediatric nephrology. Recently, investigators employed RNA-seq to describe the monocyte transcriptome in systemic lupus erythematosus (SLE) with associated glomerulonephritis. The monocyte was chosen for this study because of its important role in the pathogenicity of SLE. In this case–control design, investigators discovered that several developmental genes were underexpressed in SLE, suggesting that monocytes in SLE are more differentiated compared with controls. Total transcription of protein coding regions and antisense RNA were decreased in SLE subjects. While most of the noncoding RNAs were decreased in SLE, pri-miRNAs which prevent transcription were upregulated corresponding to decreased expression observed in protein coding genes.[79] Other transcriptomic studies have revealed a potential role for miRNA in lupus and have even begun to examine the urine transcriptome in hopes of identifying potential biomarkers for disease.[80] Preliminary studies of RNA-seq in ADPKD, acute kidney injury, and CAKUT have demonstrated promise for elucidation of disease pathogenicity.[81–83] One can easily imagine multiple uses for these methods in the study of inherited renal disease.

Mass Spectrometry–Based Approaches

For the high throughput study of proteomics and metabolomics, MS has become the leading experimental method. It has relatively high sensitivity and specificity. MS uses mass to charge ratio (m/z) to identify substances. Typically a compound is dissolved in a solvent and ionized leading to its fracture into charged fragments. These fragments are then separated using a combination of electric and magnetic fields and then detected as a spectrum based on their m/z. The spectra are compared with other spectra in databases to identify the compounds. Additionally, MS provides information about the relative abundance of the compounds. Different ionization methods and analyzers exist, each with their own strengths and weaknesses (►**Table 5**).[59,84] Notably, analyzers are often used in tandem to allow improved separation and identification of material.[84]

Proteomics

Proteomics relies heavily on high throughput methods such as MS to catalog all proteins in a cell or biological system. Proteomic experiments are difficult due to the issues mapping spectra and peptides to unique proteins, wide dynamic range in abundance of proteins within the cell, and posttranslational modifications.[85] It is estimated that only about two-thirds of the proteins present in humans have been detected using MS.[59]

Approach

The most common MS method utilized in proteomics is a bottom-up approach using trypsin digestion and liquid chromatography (LC-MS/MS).[84] Studies may aim to simply identify proteins in a sample or more commonly investigators are searching for quantitative differences between samples. In quantitative experiments, one must consider both the biological variation and technical variation given MS run variability. Two main approaches exist: one aims to characterize the relative or absolute concentration of proteins and the other aims to compare the changes in average abundance between conditions. Downstream analysis of this data including class discovery and class prediction often occurs. Class discovery attempts to identify functional protein classes or biological samples with similar protein profiles. Often systems biology approaches are applied to accomplish this. Class prediction is the use of protein abundance profile to predict disease status and is the first step in establishing biomarkers.[85]

Application

To date, most studies of proteomics in pediatric nephrology have focused on biomarker discovery. For instance, patients with posterior urethral valves (PUV) are now often diagnosed prenatally, but clinicians struggle with the extent of kidney damage that the neonate will manifest. Klein et al[86] examined the power of urinary proteins to predict renal disease progression in PUV, first taking an untargeted approach to identify urine proteins that were differentially expressed in patients with severe kidney disease/end-stage renal disease and those without end-stage renal disease at 2 years. Then they validated 26 peptides in a separate cohort, limited these tests further, and tested this algorithm in an independent cohort.[86] Follow-up studies have utilized this early work to start predicting outcomes for fetuses with PUV with the ultimate goal of using this information to drive fetal intervention.[87] These studies demonstrate some of the guiding principles put forward in clinical proteomic applications.[27] Other nephrologic areas of interest for proteomics include acute kidney injury, CAKUT, allograft rejection, and renal Fanconi syndrome.[88]

Metabolomics

Metabolomics is the high throughput study of small molecules such as peptides, amino, nucleic, and organic acids in a biological system. MS and nuclear magnetic resonance (NMR)-based techniques provide high throughput, sensitive, and specific information about the metabolome. These experiments often examine potential biomarkers using blood, urine, and serum. This review focuses on MS-based metabolomics, but thorough reviews of NMR may be found elsewhere.[89] The metabolome is a heterogeneous mixture of molecules with different physical and chemical properties preventing any one MS analytic method from becoming the gold standard. MS-based metabolomic experiments have many similar limitations and biases as are found in proteomics. These include

Table 5 Overview of mass spectrometry methods

	Description	Strengths	Weaknesses	Reference
Ionization methods				
ESI	High voltage applied to liquid, creating aerosol	• Leaves the molecule intact	• Prone to ion suppression	84,121
MALDI	MALDI matrix and sample co-crystalize and then matrix absorbs energy and sample desorbs into gas	• Rapid sample prep • More tolerant of salts and detergents	• Sensitivity and peak resolution suffer from suboptimal crystal formation • Not used for absolute quantification	84
Analyzers				
Triple quadrupole	Tandem quadrupole MS using time varied electric current using DC and RF voltage	• Good reproducibility • Low cost	• Limited resolution	84
Orbitrap	Uses electrostatic force balanced by centrifugal force to trap ions	• Low cost • High space charge capacity	• Slow scan speed • Less mass resolution than FT-IRC	122
Time of flight	M/Z determined by TOF through vacuum tube	• Detects largest molecules • Fast • Compatible with MS imaging	• -Less mass resolution than others	84
FT-IRC	Homogeneous application of magnetic field inducing cyclotron motion determined by m/z	• Highest mass resolving power • Highest mass accuracy	• Slow • Not used in quantification	84

Abbreviations: ATAC-seq, assay for transposase-accessible chromatin using sequencing; CAPTURE-seq, RNA capture sequencing; ChIA-PET, chromatin interaction analysis by paired-end tag sequencing; ChIP-seq, chromatin immunoprecipitation sequencing; CLIP-seq, cross-linking immunoprecipitation sequencing; DNase-seq, DNase I sequencing; FAIRE-seq, formaldehyde-assisted isolation of regulatory elements; iCLIP, individual-nucleotide resolution cross-linking immunoprecipitation; LC-MS, liquid chromatography-mass spectrometry; MDB-seq, methylated DNA binding domain sequencing; MeDip-seq, methylated DNA immunoprecipitation sequencing; MNase-seq, micrococcal nuclease sequencing; MPS, massive parallel sequencing; NET-seq, native elongating transcript sequencing; PAR-CLIP, photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation; TF, transcription factor; WES, whole exome sequencing; WGS, whole genome sequencing.

appropriate sample preparation, high analytic variability, and destruction of material during experimentation.[90]

Approach

Most studies attempt to identify metabolites and quantify their abundance. Often researchers use this information to detect difference between two conditional states using classic epidemiological approaches and may also employ experimental techniques especially in the area of pharmacology. Targeted approaches are commonly implemented in the study of drugs and other clinical biomarkers. This type of approach may utilize knowledge of metabolic pathways to study changes in biological networks. Targeted approaches allow for increased sensitivity and specificity of metabolite identification and improved quantification of products. Untargeted approaches are usually hypothesis generating and require further studies for validation.[91] Multiple in vitro study approaches attempt to identify protein–metabolite interactions.[59]

Application

Metabolomics especially on urine samples promises to provide biomarkers and metabolic profiles that help identify pathologic processes and potentially aid in prognosis in a noninvasive manner. For example, researchers recently examined the urine metabolome of pediatric renal transplant patients who underwent renal biopsy either as part of their surveillance or due to increased creatinine. The results revealed that a specific metabolic profile could be used to separate patients into three categories: acute T-cell rejection, borderline rejection, and no rejection. Baseline kidney function, inflammation, and pyuria did not affect the reliability of this metabolic profile.[92] This study demonstrates the promise of metabolic profiling in diagnosing renal disease states in a noninvasive manner. Other recent studies have examined the metabolome as a possible source of biomarkers in acute kidney injury and IgA nephropathy.[93]

Integration of Data

The development of high throughput omics methods has shifted the burden of genetic disease from data generation to data analysis. Each field has made important discoveries in isolation, but the richness of these data lies in its integration to more fully depict biological systems. While NGS technologies allow exploration of the genomic footprint, regulation, and transcription, the translation of this information into cellular function in health and disease requires exploration of proteins and metabolites. This incorporation of multiple different high throughput methods to describe biological systems currently underpins many principles of systems biology and has stimulated bioinformatics tools and statistical strategies to connect data from these large datasets.[94]

Investigators have developed multiple methods in attempts to integrate this omics information. There are two basic approaches to integration: unsupervised and supervised methods. Unsupervised methods ask one basic question. Is there a pattern in this data? It is a hypothesis generating technique. These processes tend to incorporate more of the raw data. These tests do not necessarily contribute knowledge to biological processes but instead provides fodder for a supervised project. Supervised integration starts from a hypothesis and often requires further experimentation. Often supervised approaches limit the dimensionality of the data. There are multiple ways to carryout either a supervised or unsupervised project. These include clustering, feature selection, prediction analysis, text mining, and pathway analysis.[59] While these methods attempt to address integration of omics data into a functional biologic picture, more methodological work in both systems biology and omics is required to build sophisticated hierarchal models of genetic disease.[94]

Functional Assays

To date, many of the systems biology approaches have utilized functional models to obtain the appropriate datasets for integration. Functional models are also used to validate disease hypotheses generated through these techniques. In nephrology, in vitro methods using cell line, ex vivo methods with murine kidney explants, and in vivo methods using model organisms exist for the study of genomic processes (►Table 6). Cell lines exist for almost all cells of the glomerulus and renal tubules, and cell-based methods such as cellular reporter assays including three-dimensional cell culture or tubulogenesis assays work well for renal diseases that primarily involve only one cell type (e.g., nephronophthisis or ADPKD).[4,95,96] Ex vivo manipulation of murine renal explants or in vivo methods such as conditional gene targeting holds promise for studying more complex genetic diseases such as CAKUT.[4] Animal models commonly used with gene targeting methods include mice and zebrafish.[97–99] Recent advances in the use of induced pluripotent stem cells directly from affected patients' fibroblast promise to provide a relatively easy genetic model system for studying the effect of specific alleles on renal differentiation and development but have yet to be available for use in inherited renal disease.[4] Researchers may also choose to apply omics methods to these model systems, especially if the relevant human tissue is inaccessible or when trying to study only one particular cell type within an organ system. Without more functional studies into genes and pathways identified using omics data, the ability to accurately predict pathogenic genetic mutations and describe the resultant molecular derangements identified in omics data will be incomplete.

Conclusion

Inherited renal disease is a leading cause of pediatric renal morbidity and mortality. Recently, NGS has identified a growing number of genes responsible for these diseases. Continued development of high throughput technologies provide further information on how genetic mutations manifest in disease. They offer new methods for diagnosis and prognostication and possible targets for new therapeutics. However, the omics fields are relatively young and further

Table 6 Functional assays for examining molecular pathways in renal disease

	Description	Reference
Short-interfering RNA assays	Knockdown of candidate genes in transfected cells using siRNAs that are homologous to the genes of interest	95
3D renal spheroid	Knockdown of candidate genes using siRNA in murine collecting duct cell lines looking for change in shape of cilia which should be spherical	96
Morpholinos	Antisense oligonucleotides complimentary to mRNA of interest bind mRNA preventing protein synthesis	97
Inducible Cre-Lox–based animal models	Site-specific recombinase technology that introduces deletions, insertions, translocations, and inversions into DNA. Targets specific cell types or can be triggered by specific external stimulus	99
CRISPR/Cas9 systems	May target and conditionally knockdown any gene in a mouse using guide RNA and Cas9 endonuclease to cut DNA at predetermined site of interest with the introduction of small indels or different base pairs during the repair	98

methodological development in the area of data processing and statistical analysis will likely improve interpretation of these experimental results. In order for translation of this research into clinical application, more coordinated efforts are needed in the field. Datasets that include not only genetic information but also protein and metabolite information will allow for more integrated data analysis further illuminating pathologic processes. Researchers need to focus on development of high throughput functional models to keep up with the demand that these omics experiments generate. Standardized methods of data sharing and more widely shared datasets will provide greater power to support genetic findings. Finally, greater education and awareness of these methodologies including their limitations among pediatric nephrologists will lead to well-designed research projects in a larger number of pediatric renal disease.

References

1 Devuyst O, Knoers NV, Remuzzi G, Schaefer F; Board of the Working Group for Inherited Kidney Diseases of the European Renal Association and European Dialysis and Transplant Association. Rare inherited kidney diseases: challenges, opportunities, and perspectives. Lancet 2014;383(9931):1844–1859

2 Reeders ST, Breuning MH, Davies KE, et al. A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. Nature 1985;317(6037):542–544

3 NCBI handbook. Bethesda, MD. National Library of Medicine, National Center for Biotechnology Information. 2002. Available at: http://www.ncbi.nlm.nih.gov/books/NBK21091/. Accessed October 1, 2014

4 Renkema KY, Stokman MF, Giles RH, Knoers NV. Next-generation sequencing for research and diagnostics in kidney disease. Nat Rev Nephrol 2014;10(8):433–444

5 Artuso R, Fallerini C, Dosa L, et al. Advances in Alport syndrome diagnosis using next-generation sequencing. Eur J Hum Genet 2012;20(1):50–57

6 Chatterjee R, Hoffman M, Cliften P, Seshan S, Liapis H, Jain S. Targeted exome sequencing integrated with clinicopathological information reveals novel and rare mutations in atypical, suspected and unknown cases of Alport syndrome or proteinuria. PLoS ONE 2013;8(10):e76360

7 McCarthy HJ, Bierzynska A, Wherlock M, et al; RADAR the UK SRNS Study Group. Simultaneous sequencing of 24 genes associated with steroid-resistant nephrotic syndrome. Clin J Am Soc Nephrol 2013;8(4):637–648

8 Munroe PB, Barnes MR, Caulfield MJ. Advances in blood pressure genomics. Circ Res 2013;112(10):1365–1379

9 Hildebrandt F. Genetic kidney diseases. Lancet 2010;375(9722):1287–1295

10 MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. Nature 2014;508(7497):469–476

11 Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet 2010;11(11):773–785

12 Robinson MD, Kahraman A, Law CW, et al. Statistical methods for detecting differentially methylated loci and regions. Front Genet 2014;5:324

13 Angelini C, Costa V. Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. Front Cell Dev Biol 2014;2:51

14 Marstrand TT, Storey JD. Identifying and mapping cell-type-specific chromatin programming of gene expression. Proc Natl Acad Sci U S A 2014;111(6):E645–E654

15 Lee S, Emond MJ, Bamshad MJ, et al; NHLBI GO Exome Sequencing Project—ESP Lung Project Team. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 2012;91(2):224–237

16 Soneson C, Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 2013;14:91

17 Matthiesen R, Carvalho AS. Methods and algorithms for quantitative proteomics by mass spectrometry. Methods Mol Biol 2013;1007:183–217

18 Matthiesen R, Azevedo L, Amorim A, Carvalho AS. Discussion on common data analysis strategies used in MS-based proteomics. Proteomics 2011;11(4):604–619

19 Pernemalm M, Lehtiö J. Mass spectrometry-based plasma proteomics: state of the art and future outlook. Expert Rev Proteomics 2014;11(4):431–448

20 Sugimoto M, Kawakami M, Robert M, Soga T, Tomita M. Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. Curr Bioinform 2012;7(1):96–108

21 Schäfer M, Lkhagvasuren O, Klein HU, et al. Integrative analyses for omics data: a Bayesian mixture model to assess the concordance of ChIP-chip and ChIP-seq measurements. J Toxicol Environ Health A 2012;75(8-10):461–470

22 Mensaert K, Denil S, Trooskens G, Van Criekinge W, Thas O, De Meyer T. Next-generation technologies and data analytical approaches for epigenomics. Environ Mol Mutagen 2014;55(3):155–170

23 Replogle JM, De Jager PL. Epigenomics in translational research. Transl Res 2015;165(1):7–11

24 Chi YY, Gribbin MJ, Johnson JL, Muller KE. Power calculation for overall hypothesis testing with high-dimensional commensurate outcomes. Stat Med 2014;33(5):812–827

25 Rizzo JM, Buck MJ. Key principles and clinical applications of "next-generation" DNA sequencing. Cancer Prev Res (Phila) 2012;5(7):887–900

26 Witasp A, Ekström TJ, Lindholm B, Stenvinkel P, Schalling M, Nordfors L. Novel insights from genetic and epigenetic studies in understanding the complex uraemic phenotype. Nephrol Dial Transplant 2014;29(5):964–971

27 Mischak H, Ioannidis JP, Argiles A, et al. Implementation of proteomic biomarkers: making it work. Eur J Clin Invest 2012;42(9):1027–1036

28 Weiss MM, Van der Zwaag B, Jongbloed JD, et al. Best practice guidelines for the use of next-generation sequencing applications in genome diagnostics: a national collaborative study of Dutch genome diagnostic laboratories. Hum Mutat 2013;34(10):1313–1321

29 Chervitz SA, Deutsch EW, Field D, et al. Data standards for Omics data: the basis of data sharing and reuse. Methods Mol Biol 2011;719:31–69

30 Field D, Garrity G, Gray T, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol 2008;26(5):541–547

31 Mardis ER. Next-generation sequencing platforms. Annu Rev Anal Chem (Palo Alto, Calif) 2013;6:287–303

32 van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet 2014;30(9):418–426

33 Riebler A, Menigatti M, Song JZ, et al. BayMeth: improved DNA methylation quantification for affinity capture sequencing data using a flexible Bayesian approach. Genome Biol 2014;15(2):R35

34 Ross MG, Russ C, Costello M, et al. Characterizing and measuring bias in sequence data. Genome Biol 2013;14(5):R51

35 Buermans HP, den Dunnen JT. Next generation sequencing technology: Advances and applications. Biochim Biophys Acta 2014; 1842(10):1932–1941

36 Hui P. Next generation sequencing: chemistry, technology and applications. Top Curr Chem 2014;336:1–18

37 Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012;2012:251364

38 van Dijk EL, Jaszczyszyn Y, Thermes C. Library preparation methods for next-generation sequencing: tone down the bias. Exp Cell Res 2014;322(1):12–20

39 Coonrod EM, Durtschi JD, Margraf RL, Voelkerding KV. Developing genome and exome sequencing for candidate gene identification in inherited disorders: an integrated technical and bioinformatics approach. Arch Pathol Lab Med 2013;137(3): 415–433

40 Pattnaik S, Vaidyanathan S, Pooja DG, Deepak S, Panda B. Customisation of the exome data analysis pipeline using a combinatorial approach. PLoS ONE 2012;7(1):e30080

41 Bao S, Jiang R, Kwan W, Wang B, Ma X, Song YQ. Evaluation of next-generation sequencing software in mapping and assembly. J Hum Genet 2011;56(6):406–414

42 Ruffalo M, LaFramboise T, Koyutürk M. Comparative analysis of algorithms for next-generation sequencing read alignment. Bioinformatics 2011;27(20):2790–2796

43 Pirooznia M, Kramer M, Parla J, et al. Validation and assessment of variant calling pipelines for next-generation sequencing. Hum Genomics 2014;8:14

44 Yi M, Zhao Y, Jia L, He M, Kebebew E, Stephens RM. Performance comparison of SNP detection tools with illumina exome sequencing data—an assessment using both family pedigree information and sample-matched SNP array data. Nucleic Acids Res 2014; 42(12):e101

45 Liu X, Han S, Wang Z, Gelernter J, Yang BZ. Variant callers for next-generation sequencing data: a comparison study. PLoS ONE 2013; 8(9):e75619

46 Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. Clin Chem 2009; 55(4):641–658

47 Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. Cell 2013;155(1):27–38

48 Goh G, Choi M. Application of whole exome sequencing to identify disease-causing variants in inherited human diseases. Genomics Inform 2012;10(4):214–219

49 Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA 2014; 312(18):1880–1887

50 Xue Y, Chen Y, Ayub Q, et al; 1000 Genomes Project Consortium. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. Am J Hum Genet 2012;91(6): 1022–1032

51 Hunt KA, Smyth DJ, Balschun T, et al; Type 1 Diabetes Genetics Consortium; UK Inflammatory Bowel Disease (IBD) Genetics Consortium; Wellcome Trust Case Control Consortium. Rare and functional SIAE variants are not associated with autoimmune disease risk in up to 66,924 individuals of European ancestry. Nat Genet 2012;44(1):3–5

52 Brown EJ, Pollak MR, Barua M. Genetic testing for nephrotic syndrome and FSGS in the era of next-generation sequencing. Kidney Int 2014;85(5):1030–1038

53 Hwang DY, Dworschak GC, Kohl S, et al. Mutations in 12 known dominant disease-causing genes clarify many congenital anomalies of the kidney and urinary tract. Kidney Int 2014;85(6): 1429–1433

54 Lemaire M, Frémeaux-Bacchi V, Schaefer F, et al. Recessive mutations in DGKE cause atypical hemolytic-uremic syndrome. Nat Genet 2013;45(5):531–536

55 Otto EA, Ramaswami G, Janssen S, et al; GPN Study Group. Mutation analysis of 18 nephronophthisis associated ciliopathy disease genes using a DNA pooling and next generation sequencing strategy. J Med Genet 2011;48(2):105–116

56 Douglas AP, Vance DR, Kenny EM, Morris DW, Maxwell AP, McKnight AJ. Next-generation sequencing of the mitochondrial genome and association with IgA nephropathy in a renal transplant population. Sci Rep 2014;4:7379

57 Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009;462(7271):315–322

58 Pieterse E, van der Vlag J. Breaking immunological tolerance in systemic lupus erythematosus. Front Immunol 2014;5:164

59 Ballereau S, Glaab E, Kolodkin A, et al. Functional genomics, proteomics, metabolomics and bioinformatics for systems biology. In: Prokop B, ed. Systems Biology. Dordrecht, The Netherlands: Springer Science and Business Media; 2013:3–41

60 Beckerman P, Ko YA, Susztak K. Epigenetics: a new way to look at kidney diseases. Nephrol Dial Transplant 2014;29(10): 1821–1827

61 Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet 2012;13(12):840–852

62 Zentner GE, Henikoff S. Regulation of nucleosome dynamics by histone modifications. Nat Struct Mol Biol 2013;20(3):259–266

63 Rivera CM, Ren B. Mapping human epigenomes. Cell 2013; 155(1):39–55

64 Banerjee T, Chakravarti D. A peek into the complex realm of histone phosphorylation. Mol Cell Biol 2011;31(24):4858–4873

65 Liu Y, Liu K, Qin S, Xu C, Min J. Epigenetic targets and drug discovery: part 1: histone methylation. Pharmacol Ther 2014; 143(3):275–294

66 Meyer CA, Liu XS. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat Rev Genet 2014;15(11):709–721

67 Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet 2013;14(3):204–220

68 Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet 2013;14(8):585–594

69 Heijmans BT, Mill J. Commentary: The seven plagues of epigenetic epidemiology. Int J Epidemiol 2012;41(1):74–78

70 Woroniecki R, Gaikwad AB, Susztak K. Fetal environment, epigenetics, and pediatric renal disease. Pediatr Nephrol 2011; 26(5):705–711

71 Woo YM, Bae JB, Oh YH, et al. Genome-wide methylation profiling of ADPKD identified epigenetically regulated genes associated with renal cyst development. Hum Genet 2014;133(3):281–297

72 Xia S, Li X, Johnson T, Seidel C, Wallace DP, Li R. Polycystin-dependent fluid flow sensing targets histone deacetylase 5 to prevent the development of renal cysts. Development 2010; 137(7):1075–1084

73 Deng Y, Tsao BP. Advances in lupus genetics and epigenetics. Curr Opin Rheumatol 2014;26(5):482–492

74 McHale CM, Zhang L, Thomas R, Smith MT. Analysis of the transcriptome in molecular epidemiology studies. Environ Mol Mutagen 2013;54(7):500–517

75 Lahens NF, Kavakli IH, Zhang R, et al. IVT-seq reveals extreme bias in RNA sequencing. Genome Biol 2014;15(6):R86

76 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10(1):57–63

77 Milek M, Wyler E, Landthaler M. Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing. Semin Cell Dev Biol 2012;23(2):206–212

78 Zhang C, Darnell RB. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol 2011;29(7):607–614

79 Shi L, Zhang Z, Yu AM, et al. The SLE transcriptome exhibits evidence of chronic endotoxin exposure and has widespread

dysregulation of non-coding and coding RNAs. PLoS ONE 2014; 9(5):e93846

80 Abulaban KM, Brunner HI. Biomarkers for childhood-onset systemic lupus erythematosus. Curr Rheumatol Rep 2015;17(1):471

81 Ben-Dov IZ, Tan YC, Morozov P, et al. Urine microRNA as potential biomarkers of autosomal dominant polycystic kidney disease progression: description of miRNA profiles at baseline. PLoS ONE 2014;9(1):e86856

82 Shapiro MD, Bagley J, Latz J, et al. MicroRNA expression data reveals a signature of kidney damage following ischemia reperfusion injury. PLoS ONE 2011;6(8):e23011

83 Marrone AK, Ho J. MicroRNAs: potential regulators of renal development genes that contribute to CAKUT. Pediatr Nephrol 2014;29(4):565–574

84 Lin SY, Hsu WH, Lin CC, Chen CJ. Mass spectrometry-based proteomics in Chest Medicine, Gerontology, and Nephrology: subgroups omics for personalized medicine. Biomedicine (Taipei) 2014;4:25

85 Käll L, Vitek O. Computational mass spectrometry-based proteomics. PLOS Comput Biol 2011;7(12):e1002277

86 Klein J, Lacroix C, Caubet C, et al. Fetal urinary peptides to predict postnatal outcome of renal disease in fetuses with posterior urethral valves (PUV). Sci Transl Med 2013;5(198):198ra106

87 Abdennadher W, Chalouhi G, Dreux S, et al. Fetal urine biochemistry at 13-23 weeks in low urinary tract obstruction: criteria for in utero treatment. Ultrasound Obstet Gynecol 2014; In press. Doi: 10.1002/uog.14734

88 Schanstra JP, Mischak H. Proteomic urinary biomarker approach in renal disease: from discovery to implementation. Pediatr Nephrol 2015;30(5):713–725

89 Smolinska A, Blanchet L, Buydens LM, Wijmenga SS. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. Anal Chim Acta 2012; 750:82–97

90 Dharuri H, Demirkan A, van Klinken JB, et al. Genetics of the human metabolome, what is next? Biochim Biophys Acta 2014; 1842(10):1923–1931

91 Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol 2012;13(4): 263–269

92 Blydt-Hansen TD, Sharma A, Gibson IW, Mandal R, Wishart DS. Urinary metabolomics for noninvasive detection of borderline and acute T cell-mediated rejection in children after kidney transplantation. Am J Transplant 2014;14(10):2339–2349

93 Hanna MH, Brophy PD. Metabolomics in pediatric nephrology: emerging concepts. Pediatr Nephrol 2015;30(6):881–887

94 Conesa A, Mortazavi A. The common ground of genomics and systems biology. BMC Syst Biol 2014;8(Suppl 2):S1

95 Doi N, Zenno S, Ueda R, Ohki-Hamazaki H, Ui-Tei K, Saigo K. Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors. Curr Biol 2003;13(1):41–46

96 Giles RH, Ajzenberg H, Jackson PK. 3D spheroid model of mIMCD3 cells for studying ciliopathies and renal epithelial disorders. Nat Protoc 2014;9(12):2725–2731

97 Ekker SC. Morphants: a new systematic vertebrate functional genomics approach. Yeast 2000;17(4):302–306

98 Hwang WY, Fu Y, Reyon D, et al. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. PLoS ONE 2013;8(7): e68708

99 Orban PC, Chui D, Marth JD. Tissue- and site-specific DNA recombination in transgenic mice. Proc Natl Acad Sci U S A 1992;89(15):6861–6865

100 Reyes-Herrera PH, Ficarra E. Computational Methods for CLIP-seq Data Processing. Bioinform Biol Insights 2014;8:199–207

101 Mischak H, Allmaier G, Apweiler R, et al. Recommendations for biomarker identification and qualification in clinical proteomics. Sci Transl Med 2010;2(46):46ps42

102 Abecasis GR, Auton A, Brooks LD, et al; 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491(7422):56–65

103 Blake JA, Bult CJ, Eppig JT, Kadin JA, Richardson JE; Mouse Genome Database Group. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. Nucleic Acids Res 2014;42(Database issue):D810–D817

104 Eckhardt F, Beck S, Gut IG, Berlin K. Future potential of the Human Epigenome Project. Expert Rev Mol Diagn 2004;4(5): 609–618

105 Consortium EP; ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol 2011;9(4): e1001046

106 Brandenberger R, Khrebtukova I, Thies RS, et al. MPSS profiling of human embryonic stem cells. BMC Dev Biol 2004;4:10

107 Vizcaíno JA, Côté RG, Csordas A, et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res 2013;41(Database issue):D1063–D1069

108 Wishart DS, Jewison T, Guo AC, et al. HMDB 3.0—The Human Metabolome Database in 2013. Nucleic Acids Res 2013;41(Database issue):D801–D807

109 Cohen CD, Kretzler M. Gene expression analyses of kidney biopsies: the European renal cDNA bank—Kröner-Fresenius biopsy bank [in German]. Pathologe 2009;30(2):101–104

110 EURenOmics. Available at: http://www.eurenomics.eu/home/ Accessed January 1, 2015

111 McMahon AP, Aronow BJ, Davidson DR, et al; GUDMAP project. GUDMAP: the genitourinary developmental molecular anatomy project. J Am Soc Nephrol 2008;19(4):667–671

112 Harding SD, Armit C, Armstrong J, et al. The GUDMAP database—an online resource for genitourinary research. Development 2011;138(13):2845–2853

113 Gadegbeku CA, Gipson DS, Holzman LB, et al. Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. Kidney Int 2013;83(4):749–756

114 Rieber N, Zapatka M, Lasitschka B, et al. Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies. PLoS ONE 2013;8(6):e66621

115 Hatem A, Bozdağ D, Toland AE, Çatalyürek ÜV. Benchmarking short sequence mapping tools. BMC Bioinformatics 2013; 14:184

116 Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. Genet Med 2015;17:444–451

117 Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10(12): 1213–1218

118 Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res 2007; 17(6):877–885

119 Harris RA, Wang T, Coarfa C, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol 2010; 28(10):1097–1105

120 Kawaji H, Lizio M, Itoh M, et al; FANTOM Consortium. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. Genome Res 2014;24(4):708–717

121 Annesley TM. Ion suppression in mass spectrometry. Clin Chem 2003;49(7):1041–1044

122 Perry RH, Cooks RG, Noll RJ. Orbitrap mass spectrometry: instrumentation, ion motion and applications. Mass Spectrom Rev 2008;27(6):661–699