Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

TRANSCRIPTIONAL PROFILING OF THE KIDNEY GLOMERULUS

Ying Sun

Karolinska Institutet

Stockholm 2010

Published by Karolinska Institutet. Printed by Larserics Digital Print AB.

ISBN 978-91-7409-959-1

To my families-
Hui, Donghe, Liqun and Siwei

We choose to go to the moon in this decade and do the other things, not because they are easy, but because they are hard…

John F. Kennedy

Abstract

The kidney glomerulus plays a crucial role in ultrafiltration of the blood through the glomerular capillary wall. Glomerulus injury may occur in many systemic and primary kidney diseases or as secondary effects to drugs and infections. It has been shown that many known glomerular specific/enriched genes play crucial roles in the maintenance of glomerular filtration function and in the pathogenesis of the known glomerular diseases of known etiology. With functional genomics and bioinformatics approaches, the identification of additional glomerular specific/enriched gene transcripts will provide a better understanding of the glomerular physiology and function, as well as the mechanisms of glomerular diseases where causes and pathogenic processes are still poorly understood.

To achieve the goal of identifying glomerular genes, we, firstly, performed comparative analysis of mouse glomerular EST libraries and whole kidney EST libraries. In this study, we found 497 glomerular enriched genes. These glomerular enriched genes were validated by different approaches. In situ hybridization confirmed the cellular location of six new glomerular markers. Apart from the mouse kidney EST library comparisons, we integrated the results of glomerular enriched genes from four additional expression-profiling platforms, and made a comprehensive catalogue of glomerular genes. Based on this, a protein-protein interaction network in glomerulus (GlomNet) was constructed. To further understanding the gene expression profile of human glomerulus, we performed transcriptional profiling of human kidney glomeruli which were obtained from renal needle biopsies taken from live renal transplant donors. By comparison with our mouse transcriptional profiling data, we observed a surprisingly large difference between the most glomerulus-specfic genes in humans and mice, respectively. Furthermore, the common glomerular enriched genes between human and mouse were classified into six different clusters according to their expression specificity in human and mouse glomeruli.

To gain insight into glomerular transcriptional changes that accompany pathological processes leading to proteinuria, we analyzed gene expression profiles following glomerular injury in the mouse lipopolysaccharide (LPS)-induced proteinuria model. LPS triggered the specific gene expression pattern changes in the glomerulus and regulated several different pathways. We observed that podocyte number decreased significantly in the model. Moreover, the genes encoded collagen $\alpha 1$, $\alpha 2$ and laminin $\beta 2$ chains which are expressed in the embryonic kidney glomerulus increased their expression in LPS-treated mice, suggesting remodeling of the glomerular basement membrane, caused by activation or dedifferentiation of glomerular cells.

In summary, the new glomerular gene transcripts identified in this thesis provide the foundation for targeted analysis of individual genes and pathways, and the GlomNet provides a preliminary systematic view of the relationships and interactions between different components in the glomerulus. The methodology of human glomerular transcription profiling can be applied for the identification of diagnostic and prognostic markers for renal diseases. The transcriptional profile changes accompanying LPS-induced proteinuria may provide clues about mechanisms involved in proteinuria. The information from this study contributes new insights for the glomerular function, physiology and the disease mechanisms.

LIST OF PUBLICATIONS

I. He L, **Sun Y**, Patrakka J, Mostad P, Norlin J, Xiao Z, Andrae J, Tryggvason K, Samuelsson T, Betsholtz C, Takemoto M. Glomerulus-specific mRNA transcripts and proteins identified through kidney expressed sequence tag database analysis. *Kidney Int.* 2007 May;71(9):889-900.

II. He L, **Sun Y**, Takemoto M, Norlin J, Tryggvason K, Samuelsson T, Betsholtz C. The glomerular transcriptome and a predicted protein-protein interaction network. *J Am Soc Nephrol.* 2008 Feb;19(2):260-8.

III. **Sun Y**, Nyström J, Ebefors K, Katayama M, Lal M, Patrakka J, Pikkarainen T, He L, Tryggvason K, Haraldsson B, Betsholtz C. Comparison of human and mouse glomerular transcription profiling data. *Manuscript*.

IV. **Sun Y**, He L, Takemoto M, Patrakka J, Pikkarainen T, Genové G, Norlin J, Truvé K, Tryggvason K, Betsholtz C. Glomerular transcriptome changes associated with lipopolysaccharide-induced proteinuria. Am J Nephrol. 2009; 29(6):558-70.

LIST OF ABBREVIATIONS

GBM	Glomerular Basement Membrane
CNF	Congenital Nephrotic Syndrome of the Finnish Type
PM	Perfect match probe
MM	Mismatch probe
SAGE	Serial Analysis of Gene Expression
QC	Quality Control
SAM	Significance Analysis of Microarray
PAM	Partitioning Around Medoids
SVM	Support Vector Machine
GO	Gene Ontology
PPI	Protein-Protein Interaction
LPS	Lipopolysaccharide
EST	Expressed Sequence Tag
ISH	*in situ* hybridization
IHC	Immunohistochemistry
HPA	Human Protein Atalas database
FDR	False Discovery Rate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
qPCR	Quantitative Polymerase Chain Reaction
EM	Electron Microscopy
RMA	Robust Multi-Array Average (RMA)
ANOVA	Analysis of Variance

Contents

List of Figures

Chapter 1

Introduction

1.1 Kidney glomerulus

The main function of the kidney is to remove various metabolic waste products by blood filtration. The kidney also plays an important role in maintenance of body homeostasis, including regulation of electrolytes, acid-base balance, and blood pressure [1]. The nephron, the functional unit of kidney, comprises the glomerulus, Bowman's capsule and renal tubules. Each human kidney contains around one million nephrons. The glomerulus, the most proximal component of the nephron, is composed of endothelial cells, mesangial cells and glomerular epithelial cells (podocytes), and it carries out ultrafiltration of the blood through the glomerular filtration barrier that resides in the capillary wall. In the healthy kidney, water, various solutes, and small proteins pass through the glomerular filtration barrier, whereas macromolecules, e.g. proteins larger than 60 kDa in size remain in the plasma. The glomerular filtration barrier is disrupted in many acquired and inherited nephropathies, which result in macromolecules, such as albumin, passing into the primary urine. In contrast to electrolytes and small proteins, the macromolcules do not get reabsorbed efficiently by the tubular system and therefore remain in the concentrated urine (proteinuria). Moreover, the macromolecular leakage into the primary urine causes secondary damage to the tubular cells of the nephron, further impairing the function of the nephron.

1.2 Glomerular filtration barrier

The glomerular capillary wall consists of three interactive lays: glomerular fenestrated endothelium, the glomerular basement membrane (GBM) and the specialized visceral epithelium-podocyte (Figure 1.1). Over decades, data generated suggested that all these three layers contribute to ultrafiltration [2-4] and they are usually considered together as the glomerular filtration barrier.

1.2.1 Glomerular endothelium

The endothelium of the glomerular capillary contains a large fenestrated area constituting 20-50% of the entire capillary surface [5]. The fenestrations contribute to efficient passage of large volumes of water and small solutes through the endothelial wall. This is important since the blood volume is filtered several times per hour [6]. Glomerular

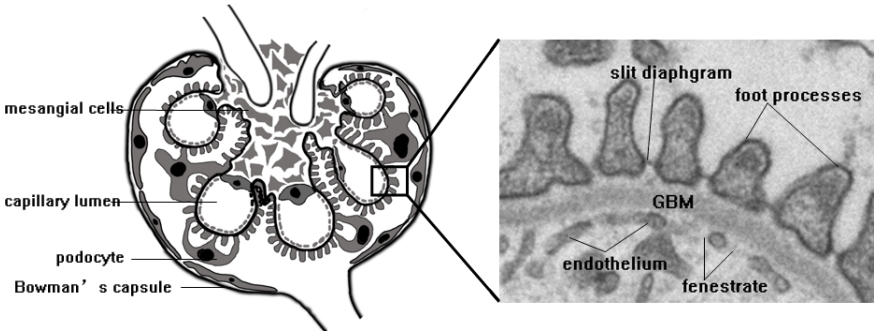

Figure 1.1: Schematic drawing of the kidney glomerulus in cross-section and EM picture of glomerular capillary wall

fenestrations, which in contrast to fenestrations in other vascular beds lack diaphragms, are too large to effectively exclude macromolecules, but the endothelial cells are covered with negatively charged surface layer-glycocalyx [7]. Jeansson et al. showed that intravenous injections of glycocalyx-degrading enzymes resulted in increased urinary albumin excretion, providing evidence that the glycocalyx is a very important component of the glomerular barrier, which could be central for charge selectivity [3, 6-8].

1.2.2 Glomerular basement membrane (GBM)

The GBM is a meshwork of type IV collagen, nidogen/entactin, heparan sulfate proteoglycans, laminin and fibronectin [1]. The GBM plays an important role to maintain the structural integrity of the filtration barrier. Since GBM is enriched in anionic charged heparan sulfate proteoglycans such as agrin [9], it has been hypothesized that GBM repels negative charged molecules such as plasma proteins and prevents them from crossing the filtration barrier [1, 10-13]. This hypothesis, referred to as the charge-selective barrier theory, has been challenged by recently published studies performed with genetically modified mice. For example, transgenic mice that lack agrin, the major GBM heparin sulfate-proteoglycan, do not develop proteinuria [14-16].

GBM Collagen IV and Laminin

Type IV collagen, the most abundant component of all basement membranes [17], is a triple-helical molecule composed of three α chains [18]. The collagen $\alpha1\alpha2\alpha1$ (IV) is found in the early developmental stages of glomerulogenesis (the "comma shaped", "S-shaped" and early "capillary loop" stages). Collagen $\alpha3-5$ chains (IV) begin to accumulate at the capillary loop stage and replace $\alpha1$ and $\alpha2$ chain (IV) in GBM in the mature glomerulus [17, 19, 20]. The possible mechanism for the collagen IV isoform substitution that occurs in the developing GBM is that embryonic $\alpha1\alpha2\alpha1$ (IV) network is more susceptible to endoproteolysis than the more heavily cross-linked $\alpha3\alpha4\alpha5$ (IV) chains [21, 22]. Endothelial cells, mesangial cells, and podocytes of immature glomeruli synthesize collagen $\alpha1$ and $\alpha2$, but collagen $\alpha3-5$ originates solely from podocytes [23].

The classical laminin molecule is cross-shaped and contains one α chain, one β chain,

and one γ chain [24]. During the glomerulogenesis, the $\gamma 1$ chain is found throughout all development stages of the GBM [1]. Laminin $\beta 1$ is enriched in the comma and S-shaped stage of glomerulogenesis. At the capillary loop stage, the $\beta 2$ chain begins to accumulate in the developing GBM, and gradually replaces the fetal isoform of laminin $\beta 1$, and eventually becomes the sole laminin β chain in the full mature GBM [19, 25]. Laminin $\alpha 5$ begins to accumulate and strongly expressed in the GBM during the capillary loop stage. Laminin $\alpha 5$ chains replace the $\alpha 1$ chains that are expressed in the early stage GBM development. Consequently, in the mature glomerulus GBM, the major laminin is laminin11 ($\alpha 5\beta 2\gamma 1$) [17, 26]

1.2.3 Podocyte and slit diaphragm

The visceral epithelial cells, known as podocytes, are lined outside of the GBM and wrap around the glomerular capillaries. The cell body of the podocyte extend into the cell's major processes, which branch to intermediate processes, and further into interdigitated foot processes. The foot processes are bridged by slit diaphragms that are around 30-40 nm wide (Figure 1.1). The podocyte has been intensively studied in recent years, and the foot process and its slit diaphragm is now believed to play a crucial role in the physiology of glomerulus filtration [4, 27-30], as well as in the pathology of renal disease. The foot processes contain an actin-based cytoskeleton that is linked to the GBM through focal contacts mediated by membrane anchoring proteins, such as the $\alpha 3\beta 1$-integrin complex [31] . Several different renal disease-causing mutations affecting podocyte proteins have been identified that lead to rearrangement of the foot process actin cytoskeleton [32-36]. The slit diaphragm forms a highly specialized continuous zipper-like structure that bridge the foot processes. The slit diaphragm contains pores permeable to water and solutes but relatively impermeable to plasma proteins [37, 38]. The slit diaphragm is formed by a complex of the plasma-membrane proteins nephrin, podocin, NEPH1-3, ZO-1, FAT1, TRPC6, VE-cadherin and P-cadherin [39-48]. Proteins that reside in slit diaphragm have been shown to be crucial for a functional slit diaphragm structure and function [37].

1.3 Genes involved in nephrotic syndrome

Nephrotic syndrome (NS) is characterized by proteinuria, hypoalbuminemia and edema. Recent genetic studies have demonstrated that several gene products of the glomerular filtration barrier have crucial roles in the function of the glomerular filtration barrier, and congenital and infantile NS often have a monogenic genetic cause [49, 50]. For example, it has been showed that mutations in the genes of NPHS1 and NPHS2 encoding nephrin and podocin proteins, which are both major component of the slit diaphragm, cause the the congenital nephrotic syndrome of the Finnish type (CNF) [51], and the familial form of steroid-resistant nephrotic syndrome [52], respectively. The TRPC6 (transient receptor potential cation channel, subfamily C, member 6) protein, also located in the slit diaphragm in close interaction with nephrin and podcin, is mutated in focal segmental glomerulosclerosis [53]. Mutations in the transcription factor gene WT1 arrests kidney development, and is associated with diffuse mesangial sclerosis. Moreover, PLCE1 (phospholipase C-epsilon-1) mutations in human result in nephrotic syndrome [54, 55]. A point mutation in the ACTN4 gene, which encodes alpha-actinin-4

protein, causes familial focal segmental glomerulosclerosis [34]. As mentioned, collagen $\alpha 3\alpha 4\alpha 5$ (IV) is present in the GBM of mature glomerulus. Mutations of either human COL4A3, COL4A4 or COL4A5 collagen genes cause Alport syndrome. This disorder is characterized by hematuria, proteinuria, progressive renal failure and sensorineural deafness [21, 22]. A interesting consequence of absence of collagen $\alpha 3\alpha 4\alpha 5$ (IV) in the GBM is that it leads to persistence of the developmental components collagen $\alpha 1\alpha 2\alpha 1$ [21, 22]. A mutation of LAMB2 (laminin $\beta 1$ chain) gene has been linked to Pierson syndrome, a disease characterized by congenital nephrotic syndrome, renal insufficiency and ocular abnormalities [56, 57].

1.4 Transcriptional profiling

Global patterns of expressed protein coding genes/transcripts represent one type of molecular signature of a given tissue or cell type, its composition, state of differentiation and function. These signatures can on the one hand be used as fingerprints for classification or correlation with e.g. cell type composition, developmental stage, disease type, stage, prognosis, therapeutic response, etc. This use does not require any prior knowledge of gene function. On the other hand, a map of the global transcriptome does also provide extensive, albeit yet incomplete, information about molecular functions in the tissue/cell, which can be used as a foundation for research into cellular mechanisms of e.g. physiological and medicine. In the current post-genomics era, huge amounts of information about gene structure, expression and function is rapidly being collected and published, and it is a challenge to take advantage of this information in the best way, i.e. for the purpose of specific scientific questions. Systematic collection and analysis of data for a given process is emerging as one of the most powerful ways of generating new information about disease processes, and, as a consequence provide new opportunities for drug target discovery and drug development. Transcriptional profiling is at the heart of these new systems biology/systems medicine approaches, and it is entering the clinic as a validated approach to disease phenotyping, such as in certain types of cancer. In this thesis I am exploring the possibility of applying transcriptional profiling not only to basic discovery of molecular mechanisms involved in the healthy and diseased kidney glomerulus, but also to develop new approaches to find biomarkers that may improve diagnosis and prognostication in human renal disease.

Microarray

Microarray technology is based on hybridization properties of nucleic acid and uses complementary molecules attached to a solid surface, referred to as *probes*, to measure the quantity of specific nucleic acid transcripts of interest that are present in a sample, referred to as the *target*. According to the different type of manufacture of gene probes (cDNA or oligonucleotides probes), DNA arrays can be divided into cDNA arrays and oligonucleotide arrays [58, 59].

cDNA microarray In cDNA microarray platforms, amplified cDNA fragments (500bp and up) are printed onto a glass slide as spots. Each spot represents one gene. The test and control RNA samples are labeled with different fluorescent dyes, and usually two differently labeled samples are mixed and hybridized to one array. With this two-

color hybridization scheme, the result of a fluorescence ratio from the spot indicates the relative mRNA abundance between the two samples (e.g. two different tissues, cell types or diseases states). Compared to oligonucleotide arrays, cross-hybridization is an especially severe problem for cDNA microarrays due to the length of the probes and targets and the lack of control for sequence homologies [60].

Oligonucleotide microarray- Affymetrix, GeneChip® Oligonucleotide arrays are produced by in situ synthesis of oligonucleotides on a solid surface using photolithographic technology [61]. In the Affymetrix 3' expression array design, each gene is represented by 11-20 pairs of 25-mer oligonucleotide probes. Each pair of probes is composed of one perfect match (PM) probe, which is designed to hybridize only with transcripts from the intended gene (specific hybridization) and one mismatch (MM) probe, which is used to measure the level of non-specific hybridization by altering the middle (13th) nucleotide of the PM probe [62]. High-density oligonecleotide arrays provide direct information about the expression levels in mRNA, because these arrays generally use a one-color biotin-labeling protocol [63] which employs one sample to be hybridized per array. The different abundance of a specific transcript between two samples (test and control) can be obtained by comparing the fluorescent signal from one sample with another sample. High-density oligonucleotide array can contain much more features than spotted cDNA arrays, for example, GeneChip® Human Genome U122 Plus 2.0 Array (Affymetrix) used in paper III comprises of more than 1,300,000 distinct oligonucleotide features which corresponding to more than 47,000 transcripts.

SAGE

SAGE is a sequence-based high-throughput method which measures the level of gene expression based on the frequency of occurrence of the tags. It is based on two principles. First, a short nucleotide sequence tag (13 or 14bp, including 4bp anchoring enzyme restriction site), is chosen that is sufficiently unique to identify a specific transcript. Second, concatamerization of SAGE tags and their subcloning into vectors allow the efficient analysis of transcripts in a serial manner by sequencing [64]. The "long SAGE" approach, further improve the specificity of tag-to-gene mapping in comparison to the original SAGE method by sequencing 21bp tags (including 4bp anchoring enzyme restriction site) [65], which allows a high percentage of long SAGE tags to be mapped directly to genomic sequence data, allowing the identification of novel genes and transcripts. However, from a economical point of view, the extra nucleotides sequenced in long SAGE can significantly increase the cost for large-scale gene expression projects. This is a balance; for the same costs, the original short SAGE yields a much greater depth of sequencing.

Chapter 2

Aims

To identify kidney glomerulus enriched/specific genes in mouse and human during normal healthy conditions and to identify causal mechanisms and putative biomarkers for the pathogenesis of proteinuria development during glomerular injury.

Chapter 3

Methods

3.1 Microarray analysis method

Figure 3.1 shows the scheme of the microarray data analysis.

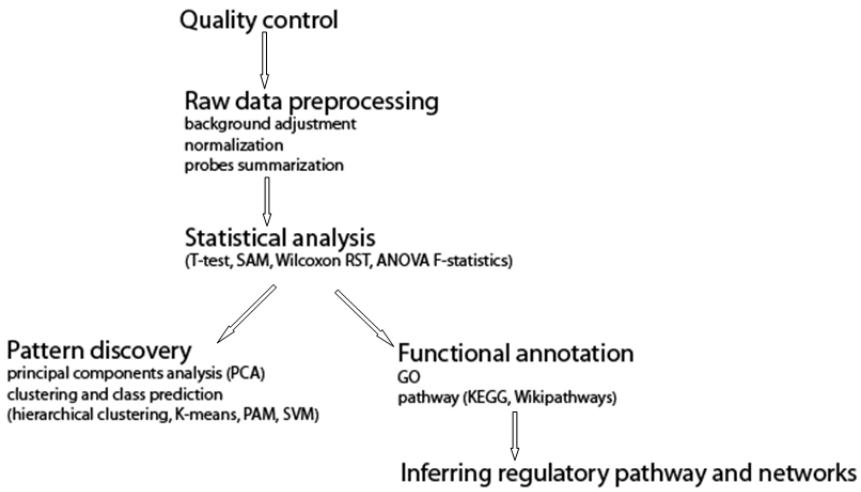

Figure 3.1: A schematic plot for general microarray data analysis workflow

3.1.1 Quality Control (QC)

QC assessment is a crucial first step in successful microarray data analysis. It is done in order to make sure that there are no problems with sample processing and that the arrays and the hybridization results are of sufficient quality to be included in the study. Affymetrix has developed several spike-in controls in order to help researchers monitor assay data quality. For examples, GAPDH and β-actin are used to assess RNA sample and assay quality; hybridization spike-in controls BioB, BioC, BioD and CreX aim to verify the efficiency of hybridization step [66].

3.1.2 Raw data preprocessing

The procedure extending from acquirement of signal intensities from the chip image to acquisition of genomic-level measurement data is commonly referred to as preprocessing. Preprocessing the high-density oligonucleotide array (Affymetrix GeneChip®, this array platform was used in paper II, III, IV) expression data usually includes background adjustment, normalization and summarization. Background adjustment is the essential step for preprocessing array data, which accounts for the background noise and non-specific binding. Normalization is the attempt to compensate for systematic technical differences and variation between chips. A series of multiple probes, eventually, need to be summarized in one quantity to represent the intensity from one transcript.

Affymetrix default algorithm MAS 5.0 for background noise adjustment is based on the difference PM-MM. However, MMs have been observed to detect some specific signal, so PM-MM transformation is likely to over-adjust [67, 68]. To overcome this drawback, RMA was developed with a background adjustment step that ignores the MM probes intensities [68]. This approach sacrifice accuracy for large gains in precision, but non-specific binding is often underestimated in this method. So, in order to balance between accuracy and precision, GCRMA algorithm was proposed, which adjusts the background based on probe sequence information combined with quantile normalization and a linear additive model for probe set expression measure. It has been shown that GCRMA outperforms other methods [69-71], and consequently we chose the GCRMA method for our analyses in papers II, III, IV.

3.1.3 Statistical analysis to identify differentially expressed genes

The typical question asked by the biologists from the transcriptional profiling experiment is which genes show differential expression with statistical significance between samples under the different conditions. Based on the assumption that microarray data is normally distributed, parametric approaches like the t-test is commonly used for comparing two groups. Another widely used statistical test, Significance Analysis of Microarray (SAM), is a t-test modified by adding a small positive constant s_0 to the denominator [72].

$$d(i) = \frac{\overline{x}_I(i) - \overline{x}_u(i)}{s(i) + s_0}$$

where $d(i)$ represents relative difference of gene(i) in expression level, $\overline{x}_I(i)$ and $\overline{x}_U(i)$ are defined as average levels of expression of gene (i) in states I and U, respectively. The $s(i)$ is the standard deviation of repeated expression measurements. The s_0 is a small positive constant number.

At low expression levels, the variance of $d(i)$ can be high because of small values of $s(i)$ leading to potential false postives. Therefore, to ensure that the variance of $d(i)$ is independent of the level of gene expression, SAM algorithm adds a correction factor s_0 to the denominator. The SAM statistical test was used in paper III and IV.

When the experimental design has multiple factors, ANOVA F statistical analysis can be employed [73]. If the microarray data is not assumed to have normal distribution, non-parametric rank-based statistic approaches can be used, such as Wilcoxon rank

sum test.

To decide the statistical significance, we usually use a p-value. The p-value is the probability that the observed result are by chance alone, when that null hypothesis is true. Using a p-value$\leq$0.05 for for significance is standard and generally acceptable when small numbers of tests are performed. However, when thousands of tests are run simultaneously (such as in microarray experiments), problems arise. If we have 10,000 genes and set threshold of the p-vaue=0.05, on average 500 (10,000 $\times$ 0.05=500) genes will have "significantly different expression levels" by chance. So, if from such an experiment we have 500 genes selected using the threshold p-value=0.05, we are unable to tell whether we have any true positives among them. Therefore, the p-values need to be adjusted according to how many tests are run together. This procedure is called multiple tests correction. Multiple tests correction is crucial step in the statistical analysis of microarray data. False Discovery Rate (FDR) method, a multiple tests correction approach that estimates the expected proportion of false positives among all rejected null hypotheses [74] is commonly used nowadays. If FDR=0.05 is used as the threshold and 500 genes are selected, we expect 500$\times$0.05=25 genes among the 500 genes are false positives. In SAM analysis, FDR is estimated by permutation test. Permutation test is based on repeatedly re-sampling in a way that scrambles the assignment of observations' class labels and computing the test statistics (i.e. t statistic) for all genes under rearrangement of the labels [75].

3.1.4 Cluster and classification analysis

Clustering is one of the most useful tasks for discovering groups and identifying interesting patterns for multivariate microarray data based on the similarity or dissimilarity. The goal of clustering is to group together objects (i.e. genes or experiments) with similar patterns of expression [76], which facilitates the identification of co-regulation. In a complex tissue, such co-regulation may represent e.g. the composition of different cell types. For example the invasion of inflammatory cells may be seen as a cluster of markers for that cell type. In a permutated cell type (e.g. stimulated by a growth factor), a cluster may represent e.g. a set of genes co-regulated by a common signaling pathway or transcriptional machinery.

Unsupervised clustering

Unsupervised clustering, a method that has no *a priori* knowledge about the data, plays an important role for the identification of new patterns in microarray data. The hierarchical clustering algorithm is the traditional and the most commonly used unsupervised clustering method, which proceeds iteratively by either merging smaller clusters into larger ones or by splitting larger clusters, yielding a dendrogram that assembles all elements into a single tree. The algorithm can be implemented to either clustering genes or to cluster experiment samples.

Another unsupervised clustering method is the partitional clustering algorithm. In this method, we usually need specify the number, k, of clusters and designate the k data points which are chosen either randomly or deliberately as 'centroids' - center points of clusters - of an initial set of clusters. The algorithm then partitions the

samples into the k clusters, optimizing some objective function (such as within-cluster similarity) by iteratively assigning objects to the nearest centroid's cluster and adjusting the centroids to represent the new clusters' center points [76]. The k-means is a well-known partitional clustering method. But the drawback of k-means clustering is that it is very sensitive to outliers. Partitioning Around Medoids (PAM) is a more robust version of the k-means clustering method [77], which partitions the data points to k clusters around k representatives, so called medoids, which minimize the sum of the dissimilarities of the observations within the cluster. Medoids are robust representations of the cluster centers that are less sensitive to outliers than centroids (because the medoid is an actual data point whereas the centroid in k-means clustering is the mean of a group of data points). The PAM clustering algorithm was used in the paper III.

Supervised classification

In contrast to unsupervised clustering methods, supervised classification usually assigns the unknown labels to the specific cluster using training samples from which a particular classification is already observed and recorded. Support Vector Machine (SVM) is becoming a very popular such technique for classification [78-80]. SVM classifies the genes by mapping the gene vectors into a high dimensional feature space, in which distance is measured using a mathematical function known as a kernel function, separating data by an optimal hyperplane [81]. SVM classification method was used in paper III.

3.1.5 Biological interpretation for transcriptional data

Functional annotation resource

In most cases microarray experiments generate lists of differentially expressed genes. The methods mentioned above using either unsupervised or supervised cluster methods are useful for describing changes in gene expression, but they are of limited use to describe cellular responses in the context of available knowledge. Therefore, such lists of differentially regulated genes usually need to be mapped to functional annotation resources. There are some knowledge databases of biological pathway and process available for such analysis, including Gene Ontology (GO), KEGG, Reactome and WikiPathways [82-84].

GO The goal of the GO Consortium is to produce a structured, precisely defined, common and controlled vocabulary for describing the roles of genes and gene products in any organism [85]. GO (`http://www.geneontology.org/`) organizes genes into hierarchical categories based on biological process, molecular function and cellular component. As of this date (March 3, 2010), there are 18581 terms related to the biological process, 2689 terms related to the cellular component and 8688 terms related to the molecular function categories.

KEGG Kyoto Encyclopedia of Genes and Genomes (KEGG, `http://www.genome.jp/kegg/`) provides a data resource that is primarily concentrated on pathways. The

pathways that represent the knowledge on the molecular interaction and reaction networks are: *metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems,human diseases* and *drug development.*

Enrichment Analysis in high-throughput genomics experiments

Enrichment analysis seeks to infer if any categories or pathways are 'enriched' by statistical methods. In paper III and IV, hypergeometric test was performed for the enrichment analysis on GO classes and KEGG pathways.

One useful computational method, called Gene Set Enrichment Analysis (GSEA, `http://www.broadinstitute.org/gsea/`), integrates a number of gene sets based on prior biological knowledge of published biochemical pathways, physical position of genes on chromosome and coexpression in previous experiments, and tests which gene sets show significant enrichment in the gene list [86].

Integration transcriptional data with biological networks

By combining expression profiles with "interactome networks information", including protein-protein interactions (PPI) and protein-DNA interactions, we may decipher the regulatory mechanisms underlying the observed changes in an activity of a biological process from the systematic level. For this type of analysis, there are a number of public repositories available, such as HPRD, MINT, IntAct [87-90]. With respect to human protein-protein interaction data, HPRD seems to be the most comprehensive [91]. In paper II, protein-protein interaction network was constructed based on HPRD annotation.

3.2 Collection of biological samples

3.2.1 Lipopolysaccharide (LPS)-induced proteinuria mouse model

LPS is a component of the cell wall of the Gram-negative bacteria. LPS activates toll-like receptor 4 (TLR4) in mammals. TLR4 is expressed on many cell types, including normal glomerular podocytes in the mouse [92]. More recently it was demonstrated that activation of podocyte TLR4 results in strong production and release of chemokines, which may promote the local recruitment of leukocytes. It has previously been suggested that LPS-induced activation of TLR4 in the glomerulus triggers podocyte foot process effacement and proteinuria through upregulated expression of the B7-1 protein and urokinase receptor uPAR [93, 94]. Moreover, LPS has been used in different studies. Haas et al. employed the LPS treatment to mouse to mimic systemic lupus glomerulonephritis [95]. They showed increase in ICAM-1 expression, increase in mesangial matrix, deposition of immune complexes and of complement. However, proteinuria was not observed, which indicated that LPS-induced glomerulonephritis mimics only some aspects of systemic lupus nephritis. Fu et al. reported the transcriptional reprogramming in LPS-treated mesangial cell lines in a time-course study [96].

In paper IV, we devised a protocol in which mice displayed a consistent proteinuric response to LPS stimulation. Nine-week-old female C57BL/6 mice were randomly

allocated to the LPS treatment group (n=6) or the PBS control group (n=4), and subsequently treated either by a single intraperitoneal injection of 13μg/g body weight of LPS or 500μl PBS control buffer. LPS-treated mice displayed consistent transient albuminuria, which peaked between 12 and 30 hour after LPS administration and returned to baseline levels after 72 hours.

3.2.2 Renal needle biopsies from pre-transplant kidneys

Healthy living-donor kidneys were selected in order to obtain an even distribution between male (M, n=8) and female sex (F, n=7) with a mean age of 50.1 $\pm$ 11.3. The biopsy specimens from fifteen kidneys were taken before transplantation and placed in RNAlater® (Qiagen, Roche) at the bedside. The biopsies were refrigerated for 24 hours and subsequently stored at -20° until preparation. Microdissection of biopsies in RNAlater® was performed manually under a stereomicroscope to separate glomeruli from the tubulointerstitial component of the biopsy. The glomeruli were washed in two steps to reduce contamination of tubular cells and placed in RLT buffer (RNeasy kits, Qiagen, Hilden, Germany). RNA was extracted from glomeruli with RNeasy Micro kit and from the tubulointerstitial component part with the RNeasy Mini kit.

Chapter 4

Results and discussion

4.1 Identification of mouse glomerular enriched genes by EST library comparison (Paper I)

The expression of eight known glomerular markers (Nphs2, Cdkn1c (P57), Nphs2, Tcf21, Podxl, Foxc2, Synpo and Ptpro) was analyzed using public oligonucleotide array data (GNF SymAtlas, Novartis) and public whole kidney EST libraries (UniGene, NCBI), but, both approaches failed to reveal higher expression of these markers in kidney as compared to other reference organs, except for Nphs2. However, we found that the above mentioned eight glomerular markers displayed high enrichment in the glomerular EST libraries [97] as compared with the whole kidney EST libraries. This encouraged us to use EST mining of glomerular librarie as an approach to identify new glomerular markers.

Our own newborn mouse glomerulus standard EST library was compared with a pool of five public newborn kidney EST libraries. From this comparison, we categorized 329 UniGene clusters (transcripts from a single gene) as glomerulus enriched by their significantly increased relative abundance in the glomerulus library. Likewise, the adult mouse glomerulus standard EST library was compared with the pool of two public adult kidneys leading to the identification of 217 UniGene clusters as glomerulus enriched at the adult stage. Comparison of the two sets of UniGene clusters revealed a total of 497 glomerulus enriched gene transcripts, 280 of which were found only at the newborn stage, 168 only at the adult stage, and 49 at both stages.

Validation by literatation mining We compared the 49 glomerular enriched transcripts present in both newborn and adult stages with published data of glomerular gene and protein expression. Fifteen (31%) had previously been demonstrated to be expressed in the kidney glomerulus by ISH, IHC, or reverse transcription-PCR. Among the transcripts enriched either in the adult or in the newborn glomerulus, 17% (28/168) and 7% (20/280), respectively, were previously demonstrated to be expressed by glomerular cells.

Validation by HPA database The glomerulus-enriched genes were examined for their corresponding protein expression pattern in the Human Protein Atlas database

(`http://www.proteinatlas.org/index.php`), which revealed nine examples of higher expression in glomerulus than in the kidney tubulus. These included CD34 (Hematopoietic progenitor cell antigen CD34), VIM (vimentin), COL4A1 (collagen alpha 1(IV)), ITGB1 (integrin beta-1), ENG (endoglin), CLDN5 (claudin-5), ACTB (actin, beta), MCAM (melanoma adhesion molecule) and MCL1 (induced myeloid leukemia cell differentiation protein Mcl1). For all of these, the glomerular expression of the human proteins was higher than in the tubular tissue.

Validation by ISH A selection of candidate glomerulus-enriched transcripts was also validated by ISH experiment. These were Clic5 (chloride intracellular channel 5), Dlk1 (delta-like 1 homologue), Gpr116 (G protein-coupled receptor 116), Plat/tPA (plasminogen activator, tissue), Tm4sf1 (transmembrane 4 superfamily member 1), and Igfbp7 (insulin-like growth factor binding protein 7). Clic5, Gpr116, Plat and Tm4sf1 showed restricted staining in kidney glomerulus across the different glomerular development stages observed in the new born kidney. Apart from showing expression in glomerulus, Dlk1 and Igfbp7 also showed expression in the kidney capsule and in certain tubular compartments, respectively.

Validation by Q-PCR We also employed the quantitative-real time PCR (Q-PCR) to quantify the transcript expression level for six new glomerular markers validated by ISH and two glomerular markers (MCAM and MCL1) validated by HPA. Using this method, they were all demonstrated to be significantly higher in their transcript abundance in the glomerulus than in whole kidney.

The most important advantage of quantitative EST profiling over hybridization-based transcription profiling methods is that it measures the abundance of EST sequence directly. The limitation of the method is the number of EST clones available for analysis. Thus, with the methodology employed for Paper I, reliable expression information is obtained only for a proportion of the genome, likely representing the most abundantly expressed genes in the glomerulus. Since our publication of Paper I the methodology of EST library synthesis and deep sequencing has developed rapidly and is likely to replace hybridization based methods in the coming years.

4.2 Integration of glomerulus-enriched transcripts and their protein-protein interaction network (paper II)

We summarized and combined the results from five different expression-profiling platforms. These data sets include a mouse kidney EST library comparison, a mouse cDNA microarray profiling (GlomChip), mouse Affymetrix Genome 430 2.0 Array profiling, a human SAGE profiling, a human Stanford cDNA microarray profiling [97-100]. In total, 1407 genes were identified as glomerulus-enriched in comparison with other parts of the kidney by at least one approach, but only seven genes (CDKN1C, ENG, EMCN, PTPRO, PLAT, IGFBP5 and PPAP2B) were identified in all studies. The five different approaches to identify glomerulus-enriched genes thus showed surprisingly limited overlap (figure 4.1). The possible reasons of limited overlapped genes are, firstly, that experiment samples were from different species (mouse and human) and

developmental stages (adult and newborn); secondly, the kidney glomerulus is likely also a target of immune-, stress-, and other environment-induced responses that may have varied between the studies; thirdly, most of the transcriptional differences reflect the usage of different technical platforms and protocols; moreover, raw data were also processed using different methods, and different criteria were used to categorize genes as glomerulus-enriched. The platform difference is probably a major reason limiting the gene expression overlap. However, paper III also identifies a surprisingly large discrepancy between human and mouse glomerular enriched transcripts (see below).

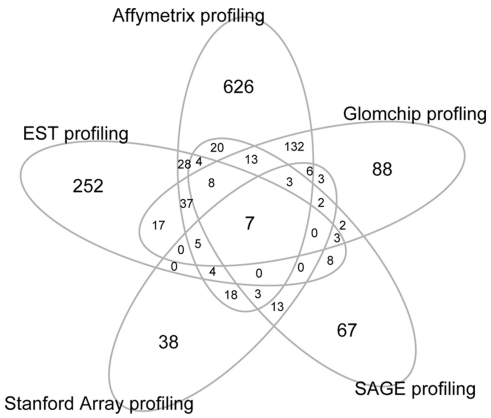

Figure 4.1: Venn diagrams showing the overlap among the five glomerulus-enriched gene lists identified by five different methods

Comparison and Evaluation of Different Approaches For assessment of the efficiency of each method in identifying glomerulus-enriched genes, the data were compared with published data on gene or protein expression in glomeruli by ISH or by IHC using literature mining and HPA database. The Affymetrix analysis, which assesses the expression of almost all genes, identified the largest number of glomerulus-enriched genes and also more literature- and HPA-confirmed genes than any of the other methods; however, the proportion of the Affymetrix glomerulus-enriched genes that were confirmed by literature or HPA was lower than for the other methods. A comparison between the results from the two cDNA microarray approaches, GlomChip and Stanford, showed that GlomChip identified more glomerulus-enriched genes and both a higher number and proportion of genes confirmed by literature and HPA than the Stanford array. This suggests that for the application to a specific tissue, a corresponding tissue-specific cDNA microarray might be preferable to a nonspecific cDNA microarray. Sequencing based EST sequencing and SAGE methods directly acquire the abundance of transcripts. As such, high abundance transcripts would also be predicted to be relatively easier to detect by other experiments methods, such as ISH or IHC. This prediction partially agreed with the result from the literature and HPA, in which a higher proportion of SAGE and EST results were confirmed as compared to other methods. But one source of false negatives from SAGE is that it will miss all those transcripts that do not have any anchoring enzyme site, which is one of the possible reasons why this method identifies relatively less glomerular-enriched genes than the other methods (except the Stanford cDNA array).

Construction of GlomNet By using the protein-protein interaction data from Human Protein Reference Database (HPRD), we predicted 772 direct interactions between 543 of the 1407 glomerulus-enriched gene products. On the basis of them an integrated glomerular protein-protein interaction network, GlomNet, was constructed (figure 4.2), Thus, GlomNet illustrates the protein interaction networks and, to some extent, signal transduction pathways in the glomerulus. By mapping the known genes whose mutation causes glomerular abnormality in humans or mice into GlomNet (highlighted by blue color in figure), we predict that known glomerular disease proteins are located mainly in the extracellular, plasma membrane, and nuclear compartments, indicating that a majority of the glomerular disease proteins involved into the cell-cell communication and signaling transduction pathway.

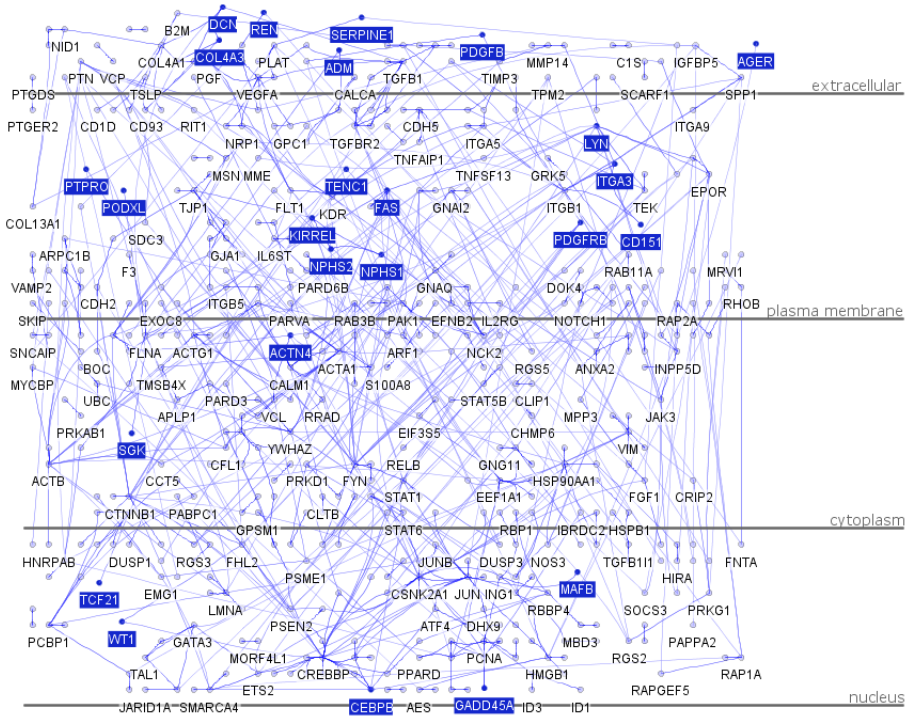

Figure 4.2: Protein-protein interaction network for glomerulus-enriched genes
Proteins are represented with blue nodes, and interactions are represented with edges. The known glomerulus disease genes are highlighted with deep blue color.

We also used the glomerulus-enriched gene catalogue and GlomNet to explore a published gene expression profiling data from patients with diabetic nephropathy (DN) in comparing with normal control subjects [101]. We found that a high proportion of the DN-regulated genes were represented in the glomerular-enriched catalogue and GlomNet. It is interesting to note that many of upregulated genes in DN are expressed in endothelial cells (NOTCH4, TIE1, TEK[TIE2], ENG, PTPRB, ICAM2, and others), whereas many of the downregulated genes are expressed specifically in podocytes (NPHS1, PTPRO, PODXL, CDKN1C, TCF21, WT1, VEGFA, COL4A3, COL4A4,

and others). Moreover, several of the DN-affected genes link directly to each other in GlomNet, suggesting specific pathways that may be targeted by or are part of the pathogenic disease process.

4.3 Comparison of human and mouse glomerular transcription profiling data (paper III)

In this study, we did transcriptional profiling for human kidney glomeruli and tubulointerstitial samples from renal needle biopsies, and compared their gene expression profiles with mouse glomerular and tubulointerstitial gene expression profiles obtained in a previous study. By analyzing overlapping genes by setting all possible combinations of fold changes for glomerulus overexpressed genes between human and mouse, it showed that surprisingly large differences exist between the most glomerulus-enriched genes in humans and mice, respectively. Nevertheless, if one-fold change was used as cut-off for both the human and mouse data, 3,119 genes were overlapping, which represented approximately 60% of the genes more than one fold upregulated in either human or mouse. The remaining 40% were selectively enriched in glomeruli in human or mouse only. The observed differences in gene expression pattern between human and mouse might be caused by true species-specific differences in gene expression, or by technical factors, such as probe design and hybridization affinity, isolation methods for glomeruli (microdissection procedure applied to the human samples may produce a more variable degree of contamination than the Dynabead perfusion [102] method used to isolate the mouse glomeruli).

Further, from 3119 overlapping glomerular genes upregulated more than one-fold in human and mouse, we firstly selected 58 known glomerular genes and grouped them into six clusters using the PAM clustering method according to their specificity in human and mouse glomerular samples. Then, by using these 58 known glomerular genes as a training data set, the remaining 3061 (3119-58) glomerulus enriched genes were classified into these six clusters using the SVM algorithm. Cluster NO.6 (Table 4.1) represented the core cluster consisting of the most glomerular specific genes compared with the other five clusters. The genes in cluster NO.5 and cluster NO.2 showed lower fold changes of enrichment compared with cluster NO.6. Moreover, cluster NO.1 and cluster NO.4 were characterized by a relative higher gene expression preference in human and mouse glomeruli, respectively. All other genes were gathered into cluster NO. 3 and most of genes in this cluster showed less than 2^2 fold enrichment both in human and mouse.

We did GO analysis for cluster NO.1 NO.2, NO.4, NO.5 and NO.6. Compared to the other clusters, cluster NO. 6 gathered the most glomerular specific genes, exemplified by NPHS1, PLCE1, TCF21 and ITGA8. Therefore, in the biological process category, *kidney development* and *glomerulus development* categories were overrepresented in cluster NO.6. Interestingly, we found that cluster NO. 5, NO.2 and NO.4 were enriched for the GO categories *angiogenesis, blood vessel morphogenesis, blood vessel development and vasculature development categories.* In cluster NO.1, the categories *immune system process, antigen processing and presentation of endogenous antigen,*and *negative regulation of cellular metabolic process* were significantly enriched.

Table 4.1: genes in Cluster No. 6

GeneSymbol	GeneName	Hs.fold(log2)	Mm.fold(log2)
PDPN	podoplanin	6,97	4,07
FZD2	frizzled homolog 2 (Drosophila)	6,35	5,83
SOST	sclerosteosis	6,28	4,08
HECW2	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 2	6,09	6,87
NPHS1	nephrosis 1, congenital, Finnish type (nephrin)	6,09	7,50
GJA5	gap junction protein, alpha 5, 40kDa	6,08	5,85
MRGPRF	MAS-related GPR, member F	5,96	5,80
TBXA2R	thromboxane A2 receptor	5,95	5,14
TCF21	transcription factor 21	5,84	7,95
CORO2B	coronin, actin binding protein, 2B	5,83	6,22
GJA3	gap junction protein, alpha 3, 46kDa	5,74	4,87
HS3ST3A1	heparan sulfate (glucosamine) 3-O-sulfotransferase 3A1	5,65	5,88
SGEF	Src homology 3 domain-containing guanine nucleotide exchange factor	5,50	6,43
PLCE1	phospholipase C, epsilon 1	5,48	6,66
EHD3	EH-domain containing 3	5,44	7,04
TYRO3	TYRO3 protein tyrosine kinase	5,38	5,12
C1orf21	chromosome 1 open reading frame 21	5,17	5,52
TBX3	T-box 3	5,07	7,77
CDKN1C	cyclin-dependent kinase inhibitor 1C (p57, Kip2)	5,05	7,95
PTPRO	protein tyrosine phosphatase, receptor type, O	4,98	6,94
TSPAN2	tetraspanin 2	4,97	7,53
KCND3	potassium voltage-gated channel, Shal-related subfamily, member 3	4,96	6,06
MSRB3	methionine sulfoxide reductase B3	4,86	5,89
SULF1	sulfatase 1	4,68	6,60
ARHGAP28	Rho GTPase activating protein 28	4,59	8,12
WT1	Wilms tumor 1	4,45	8,32
DAAM2	dishevelled associated activator of morphogenesis 2	4,41	6,43
DDN	dendrin	4,33	8,14
ITGA8	integrin, alpha 8	4,33	7,86
PARD3B	par-3 partitioning defective 3 homolog B (C. elegans)	4,28	6,43
NES	nestin	4,21	8,63
SYNPO	synaptopodin	4,04	7,48
CYP26B1	cytochrome P450, family 26, subfamily B, polypeptide 1	3,46	7,39

We also mapped the above mentioned 3119 genes to the HPA database and found an overlap of 1410 gene products represented by specific antibodies and IHC experiments in the HPA database. We found that 149 out of the 1410 (approx.10%) gene products showed stronger expression in glomeruli in comparison with renal tubules in the HPA database. Figure 4.3 shows three examples: MAP6 (microtubule-associated protein 6), NPNT (nephronectin) and TOM1L2 (target of myb1-like 2).

The human kidney glomeruli used for transcription profiling in this study were obtained from renal needle biopsies taken from live renal transplant donors as part of the transplantation routine. Thus, by focusing on biopsies from extensively screened (with regard to possible kidney problems) and confirmed healthy individuals, our method differs from previous published studies on the normal human glomerular transcriptome, which used archival material or kidney samples obtained from the uninvolved part of nephrectomy specimens from patients with renal neoplasia [98, 100, 103, 104]. The

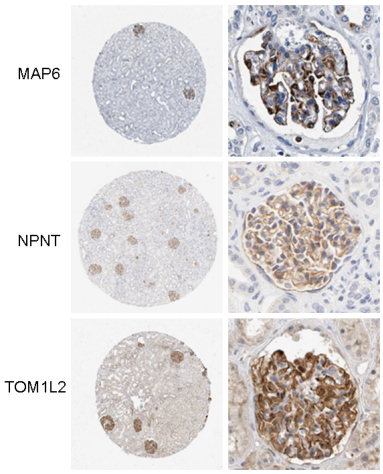

Figure 4.3: IHC shows three glomerular protein expression
IHC images were obtained from `www.proteinatlas.org`

methodology for human glomerular transcription profiling reported in this paper could be applied to the identification of new diagnostic and prognostic markers for renal diseases.

4.4 Study of glomerular transcriptome changes on LPS-induced proteinuria murine model (paper IV)

We devised a protocol in which mice displayed consistent and transient albuminuria response to LPS stimulation, which peaked between 12 and 30h after injection and returned to baseline levels after 72h. No gross morphological alterations were revealed by hematoxylin-eosin staining. However, by transmission electron microscopy (TEM) podocyte foot process effacement was observed in LPS-treated animals, which indicates that the expected rearrangement of the foot process actin cytoskeleton had taken place. Comparison of glomerular gene expression between PBS control and LPS-treated animal revealed that 4,147 different genes were differentially expressed, which corresponds to about 20% of the total number of known genes on the array (Figure 4.4 shows top 90 up- and downregulated genes in glomerulus under the LPS treatment). Moreover, in order to identify which of the LPS-induced changes that affect transcripts that are normally restricted to or enriched in the glomerulus, we compared the 4,147 genes mentioned above with 3,614 genes that were selected as more than two-fold upregulated in normal mouse glomerulus as compared to normal tubulointerstitium. We found an overlap of 1,624 genes, suggesting that roughly 45% of the glomerulus enriched gene transcripts were significantly affected in their abundance by LPS treatment. This indicated that dramatic transcriptional reprogramming of glomerular cells takes place in response to LPS.

GO and KEGG pathway analysis for the 1,624 dysregualted genes revealed that 'adherens junction' was one of the GO terms associated with significantly downregulated genes, which included α-actinin-4, catenin, vinculin and Fyn. 'focal adhesion' and

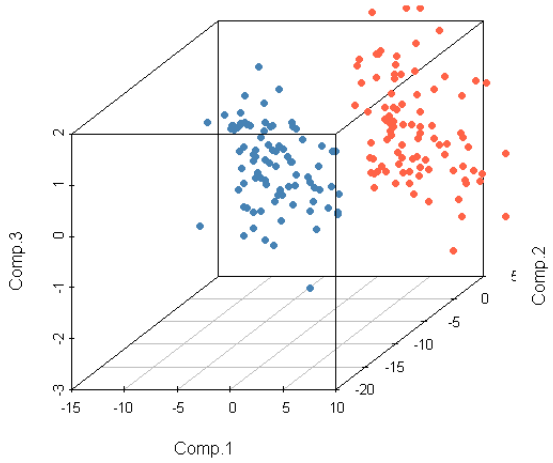

Figure 4.4: PCA plot for top 90 up- and downregulated genes following LPS treatment. Upregulated genes and downregulated genes are represented by red and blue color, respectively. X, Y, Z axis are first three principle components.

'regulation of actin cytoskeleton' were overrepresented pathways associated with both up- and downregulated genes, indicating that genes involved in these two processes are dysregulated in different ways in response to LPS. The figure 4.5 shows the 'regulation of actin cytoskeleton' pathway from the KEGG database and the plot is drawn by GenMapp tool (version 2.1) [105]. GO and KEGG analysis were in agreement with our ultrastructural analysis by TEM, which showed segmental podocyte foot process effacement. The data suggested that podocyte actin cytoskeleton undergoes reorganization in response to LPS. Furthermore, among the most strongly upregulated genes, we found numerous examples of genes/proteins implicated in Toll-like receptor signaling, immune and defense responses, apoptosis, cytokine production and inflammation. It has been reported that consecutive LPS administration to mice causes lupus-like nephritis and induced expression of Icam1 (intercellular adhesion molecule-1) and Vcam1 (vascular cell adhesion molecule) in the glomerulus [95, 106]. From our transcriptional profiling data, we observed that mRNA levels were upregulated twenty-four and five fold for Icam1 and Vcam1, respectively, after 24h by LPS administration, which might reflect the early stage of inflammatory processes leading to glomerulonephritis.

Nuclei counts of Wt-1, which is a nuclear protein specific for glomerular podocyte cells suggested a small (approx. 15%) but significant podocyte cell loss in LPS-treated mice. From our array data, we noticed that LPS dramatically changed podocyte gene expression. Thirty-four known podocyte-specific or podocyte-enriched transcripts present on the chip were 20-90% downregulated [97, 99, 107]. Therefore, the observed podocyte loss (approx. 15%) could not fully explain the decreased glomerular podocyte transcript levels (20-90%). Rather, we would propose that a combination of cell death and dedifferentiation explains the reduced relative levels of podocyte-specific

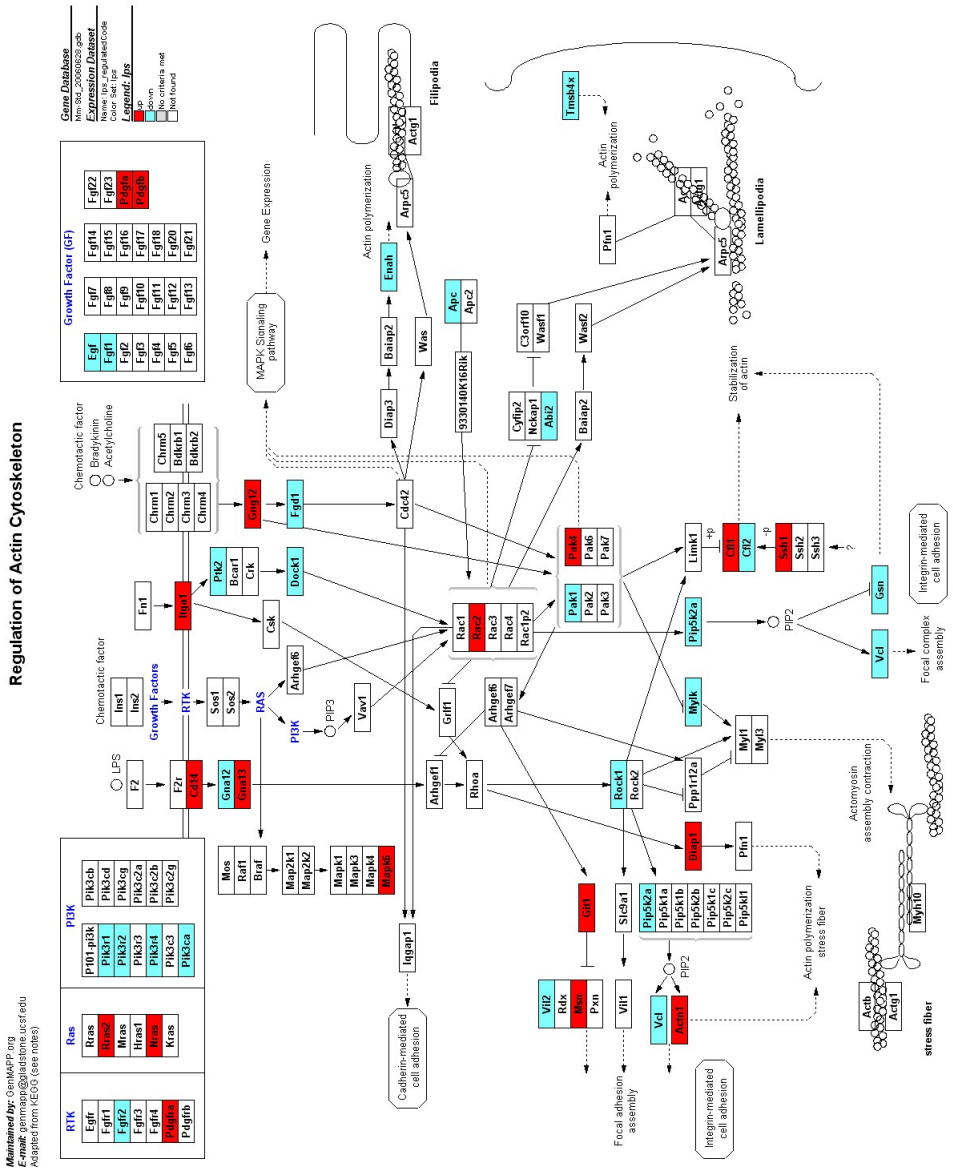

Figure 4.5: Dysregulated genes under LPS treatment in "regulation of actin cytoskeleton" pathway
The up- and downregulated genes are highlight by red and lightblue color, respectively.

transcripts. Nevertheless, part of the podocyte-specific transcriptional response observed may be involved in causing podocyte apoptosis since several genes involved in survival mechanisms were downregulated. These include Ccni (Cyclin I), Cd2ap and Nphs1. Cyclin I prevents apoptosis in postmitotic podocytes in vitro and in vivo [108]. Nphs1 and Cd2ap participate in PI13K/AKT signaling, which has been proposed to promote podocyte survival [109]. Nphs1 is phosphorylated on tyrosine residues by the Src family kinase Fyn [110, 111]. Fyn was also downregulated by LPS, possibly indicating reduced phosphorylation of nephrin, and, consequently, defects in foot process formation/maintenance.

Moreover, upregulated expression of collagen $\alpha 1$, $\alpha 2$ (IV) and laminin10 (laminin $\alpha 5 \beta 1 \beta 1$) suggested a glomerular (and podocyte) remodeling response to LPS. We confirmed the microarray data on collagen $\alpha 1$, $\alpha 2$ (IV) expression at the protein level by immunofluorescence staining, which showed increased staining intensity in the mesangial matrix and the GBM following LPS-treatment. It is known that Collagen $\alpha 1$, $\alpha 2$ (IV) and laminin $\beta 1$ are present in the GBM of early developmental stage glomeruli. At the capillary loop stage, collagen $\alpha 1$, $\alpha 2$ (IV) are replaced by collagen $\alpha 3$ $\alpha 4$ $\alpha 5$(IV) and laminin $\beta 1$ is replaced by laminin $\beta 2$ [19, 20, 22, 25]. These observations might indicate that LPS induces dedifferentiation of podocytes, leading to the expression of components of the embryonic kidney GBM. Also, increased expression of collagen $\alpha 1$, $\alpha 2$ (IV) in mesangial matrix in LPS-treated mice indicates the activation of mesangial cells.

LPS triggers dramatic transcriptional changes in glomeruli, affecting almost half of the genes that are normally expressed in a restricted or enriched fashion in the glomerulus, and also leading to the expression of a large number of immune, inflammation and stress response genes. This makes it difficult to single out specific critical signaling events. The significant downregulation of podocyte specific genes suggests that podocytes and the podocyte actin cytoskeleton are culprits in the proteinuric response to LPS. Moreover, the transcriptional changes also include re-expression of embryonic glomerular transcripts collagen $\alpha 1$, $\alpha 2$ (IV) and laminin $\beta 2$, suggesting remodeling of podocytes and GBM as part of the glomerular response to LPS.

Chapter 5

Conclutions and Future perspectives

Genetic studies have revealed that mutation of several glomerular specific or enriched genes such as NPHS1, NPHS2, PLCE1 and TRPC6 caused glomerular disease [51-53, 55]. Hence, from a reverse genetic point of view, it is reasonable to assume that the identification of additional glomerular specific or enriched genes will provide a better understanding of the glomerular physiology and functions, as well as of mechanism of the pathogenesis of unknown glomerular disease.

We performed comparative analysis of mouse glomerular EST library with whole kidney EST libraries in paper I; we integrated the different platforms of glomerular transcriptome data and put them into system biology context in paper II; and we did comparison analysis of human and mouse glomerular transcriptome profiling data in paper III. By these studies, we provide a useful source of glomerular specific/enriched genes and pinpoint a number of potential differences in the glomerular gene expression patterns between human and mouse species. GlomNet pictures the interactions network of the glomerular enriched gene products, which improve our systemic understanding of the glomerulus. In addition to exploring normal glomerular transciptome, we also move on to analyze a pathogenic condition of the glomerulus in paper IV. In this paper, the detected glomerular gene expression changes by LPS stimulation provide putative markers for early and transient glomerular injury and proteinuria.

The ultimate goals of the research that has lead to this thesis are to decipher the molecular mechanism underlying human glomerular disease, such as membranous nephropathy, IgA nephropathy and diabetic kidney disease, and try to identify molecular targets for novel therapeutic intervention strategies. The discovery of the glomerular enriched/specific genes by profiling the glomerulus in normal condition (paper I, II and III) has provided a valuable reference for this work. The next step is to gain insight to the glomerular transcriptome in pathogenic conditions in animal models and in humans. For example, using the micro-dissection method described in paper III, the possibility arises to map human glomerular transcriptome changes that accompany specific glomerular diseases, disease stages, response to therapy etc. We realize that high-throughput technologies develop rapidly, from microarray to today's deep sequencing of transcriptome technique-'RNA-Seq' [112-115] and 'next-next' generation sequencing

using single molecules [116-118], which provide us with more and more accurate high-throughput biologic data and pave the path for clinical diagnostics. From paper II, we learned that the sequencing-based method, SAGE, was more accurate than the spotted cDNA array method. Surely, new high-throughput sequencing techniques will replace microarray analysis in the long run [119]. Compared to array-based transcriptome profiling, RNA-Seq allows characterization of transcriptomes without prior knowledge of the genomic sequences; it is capable of single-base resolution, distinguishes different RNA isoforms and reveals sequence variants. Also the dynamic range of RNA-Seq for determining expression levels is greater than microarray and it performs well regarding quantitative detection; the background level is much lower and it is much more sensitive and capable to detect the most rarely expressed genes and differentially expressed genes [120].

Many of the glomerular diseases are not monogenetic. In fact, the most important renal diseases and causes of end-stage renal disease are all polygenic and likely influences also by environmental and hormonal factors. Therefore, we not only need to identify or learn about the functions of single genes or protein, we also need to discover pathologic signaling pathways and aberrant protein-protein interactions (interactome) that occur during disease states. With high-throughput functional genomics techniques, we need to integrate mass dispersed biological information and data into numerous systemic networks (regulation network, protein-protein interaction network and others) and infer the biological pathway or pathway organization underlying the pathology conditions.

Acknowledgements

I would like to express my sincere gratitude to those who provided me generous help during the years of my Ph.D study. I would like to especially thank the following persons:
Prof. Christer Betsholtz, my supervisor, for your great support, supervision, and constant encouragement, for your inspiration and shining ideas in science, for providing the great working environment, for sharing the enthusiasm, confidence, bravery and fearless both in and beyond science.
Prof. Karl Tryggvason, my co-supervisor, for your valuable guidance, advice and great patience for my questions, for creating nice scientific environment in Matrix Biology division.
Prof. Börje Haraldsson, for your kindness and wonderful collaboration, for your enthusiasm and effort in the project.
Prof. Ulf Eriksson, for your encouragement, great help and advice.
Jenny Nyström and **Kerstin Ebefors**, for providing good samples, wonderful collaboration and all the informative project discussions.
The current and former members in **Betsholtz** group:
Minoru Takemoto, for teaching me lab experiments;
Guillem Genove, for your enthusiasm, constant encouragement and helping me read through this whole book and giving me valuable comments;
Elisabet Wallgard, my sweet friend and intelligent former office mate, for your sincere encouragement and warm smiles, for all the happy and nice chats;
Miyuki Katayama, for the nice collaborations and discussions in the project;
Simin Rymo, for all the shining smiles lighting me up, for your kind words and big warm hugs; **Bárbara Laviña Siemsen**, for being a great office mate , for your enthusiasm and constant encouragement; **Konstantin Gaengel**, for your scientific attitude and help me with the microscopy; **Lwaki Ebarasi**, for the scientific discussions, for always being optimistic, encouraging and diligent; **Johanna Andrae**, for always being nice and give me help in the lab; **Kazuhiro Hagikura**, for your kindness, and telling me the tips for being alive during earthquake; **Maya Nisancioglu**, for being a great organizer of social activities in Betsholtz team, and for being inspiring; **Jenny Norlin**, for all the nice discussions and for sharing the happiness being a mother; **Annika Armulik**, for organizing the wonderful group retreat in Estonia; **Colin Niaudet**, for your enthusiasm and nice talks; **Maarja Mäe** , **Radiosa Gallini**, **Kamph Sara**, **Lars Jakobsson, Peter Lonnerberg, Johan Dixelius, Mattias Kalen, Mats Hellström , Irma Rymo** and **Karin Strittmater**, for all the enjoyable talks and generous help; **Longlong**, for teaching me basic Swedish and sharing your funny story of fishing.
The current and former members in **Tryggvason** group:
Jaakko Patrakka, for your great help in my project, for your patience, response my questions very quickly and all the informative scientific discussion, for your active spirit in the sports

which I always admire;
Mark Lal, for your enthusiasm and encouragement, for all project discussion and advice;
Timo Pikkarainen, for always being friendly and peaceful, and for your enthusiasm and inspiring commitment in science and research, for your great patience and all informative discussions;
Zhijie Xiao, for great help in the lab experiments and the good hands for making delicious cakes; **Anne-May Osterholm**, for always being nice and refilling the coffee machine; **Laleh Sistani**, for all the lovely chats and nice tomato soup; **Asmundur Oddsson, Juha Ojala, Masatoshi Nukui, Ljubica Perisic, Bing He, Berit Rydlander, Ann-Sofie Nilsson, Ann-Charlotte Andersson, Anna Domogatskaya, Sergey Rodin, Kan Katayama, Elisabeth Raschperger, Stefania Cotta-Done, Susan Warner, Yi Sun, Yunying Chen, Dadi Niu, Xiaojun Xu, Jianhua Mao, Eyrùn Hjörleifsdóttir, Marko Sankala, Ari Tuuttil**, for your generous help and being great colleagues during these years.
All the members in **Vascular Biology** division, for being friendly around and creating good scientific atmosphere.
Elisabeth Åhman, Eva Lindberg and **Alessandra Nanni**, for your kindness and all the great help.
Chad Tunell and **Anders Lundsjo**, for your excellent IT support.
My great office mates: **Hong Li**, for your humorous and interesting talks, for your fearless and being yourself. **Monika Ehnman**, for all the scientific discussions and nice talks, for your accompany during late evenings through this tough winter. **Charlotte Anderberg**, for your great kindness and all the lovely talks. Especially, **Monika** and **Charlotte**, I want to thank you for 'saving my life' when I suffered the emergent stomach pain.
Also, I thank the Bioinformatics Master's program in **Chalmers University of Techology**, for providing the great education.
Friends in MBB department: **Cheng Qing**, the great party organizer, for always being frank, trustworthy and helpful; **Wang Xun**, for always being helpful and your great knowledge in IT field; **Wan Min**, for your kind help and sharing your experience of being a good mother; **Zhang Xu**, for all interesting talks and your great sense of humor; **Lu Jun**, for always being kind and peaceful; **Xu Jianqiang**, for always being friendly, for your great enthusiasm in traditional Chinese literature and excellent Chinese calligraphy; **Sun Jia, Du Yatao and Huihui**, for your kindness and enthusiasm; **Chew Eng-Hui**, for all the lovely talks and sharing the happiness in the parties we had together; also, those who are in the badminton team of MBB, for sharing your happiness and positive attitude in the life.
Other Chinese friends met in Sweden: **Wu Hao** and **Mo Hao**, for your enthusiasm, great encouragement and help, for sharing your life experience in Sweden and nice trip to Kiruna; **Li Yu**, for your kindness, great help and the nice Viking Line boat trip; **Xian Xiaojie**, for picking me up from the airport when I came to Sweden for the first time and all the great help; **Xue Yuan**, a great photographer, for all the nice talks and your ambition; **Sun Di**, for your enthusiasm and persistence in science, for your fortitude throughout tough time; **Zhang Zhe**, for your optimism and sharing your adventurous trip stories, for your skilful techniques for taking professional pictures for us; **Wei Tianling**, for always being cheerful, for all the dinner we had together in Kungshamra; **Mi Jia**, for your great sense of humor and optimism, for your great help in Uppsala days; **He Xiaobing** and **Jiang Su**, for your enthusiasm and sharing the valuable experience for child's education; **Zhang Xiaoqun** and **Qi Hongshi**, for the good accompany of a nice Easter tip; **Zhang Bing and Li Yan**, for all the great help and valuable advice; **Zheng Huiyuan, Liu Tong, Fu Jie, Zhu Ju, Lu Ming, He Zhiqing, Qiu Ping, Jiang Tao**, for all the generous help and being great neighbors in Kungshamra; **Cao Lin, Dai**

Qi, Wang shengqiu, Zheng Lin,Wang Xiaoda, Gong Lili, Zhao Yuheng, Bai Xue, Zhao Yuwei, Shen JingFeng, Zheng Kang, Jing Xiao Na, Li Jing and Zhangfan, for your generous help.

Liu Shuang and **Ma Xiaochen**, my darling friends in China, for your great friendship and being like sisters through all of these years.

My dear parents: my mother **Zhao Hui** and my father **Sun Donghe**, for giving me life, bringing me up and encouraging me to leave the home to explore the world with a dream. Without your everlasting support, understanding and love, I wouldn't have been where I am today. Especially, thanks for giving me the great help for taking care my daughter. My husband, **He Liqun**, for your love, patience, friendship and encouragement every step of the way, for your smiling with a peaceful heart. Finally, my gorgeous, healthy and lovely daughter, **He Siwei**, for letting me know the love people would like to die for and for being a daily reminder that world is full of happiness!

Reference

1. Peter D. Vize ASW, Jonathan B.b. Bard: The Kidney From Normal Development to Congenital Disease: Elsevier Science (US); 2003.
2. Haraldsson B, Nystrom J, Deen WM: Properties of the glomerular barrier and mechanisms of proteinuria. Physiol Rev 2008, 88(2):451-487.
3. Jarad G, Miner JH: Update on the glomerular filtration barrier. Curr Opin Nephrol Hypertens 2009, 18(3):226-232.
4. Deen WM: What determines glomerular capillary permeability? J Clin Invest 2004, 114(10):1412-1414.
5. Bulger RE, Eknoyan G, Purcell DJ, 2nd, Dobyan DC: Endothelial characteristics of glomerular capillaries in normal, mercuric chloride-induced, and gentamicin-induced acute renal failure in the rat. J Clin Invest 1983, 72(1):128-141.
6. Satchell SC, Braet F: Glomerular endothelial cell fenestrations: an integral component of the glomerular filtration barrier. Am J Physiol Renal Physiol 2009, 296(5):F947-956.
7. Jeansson M, Haraldsson B: Morphological and functional evidence for an important role of the endothelial cell glycocalyx in the glomerular barrier. Am J Physiol Renal Physiol 2006, 290(1):F111-116.
8. Jeansson M, Bjorck K, Tenstad O, Haraldsson B: Adriamycin alters glomerular endothelium to induce proteinuria. J Am Soc Nephrol 2009, 20(1):114-122.
9. Harvey SJ, Miner JH: Revisiting the glomerular charge barrier in the molecular era. Curr Opin Nephrol Hypertens 2008, 17(4):393-398.
10. Tamsma JT, van den Born J, Bruijn JA, Assmann KJ, Weening JJ, Berden JH, Wieslander J, Schrama E, Hermans J, Veerkamp JH et al: Expression of glomerular extracellular matrix components in human diabetic nephropathy: decrease of heparan sulphate in the glomerular basement membrane. Diabetologia 1994, 37(3):313-320.
11. Raats CJ, Van Den Born J, Berden JH: Glomerular heparan sulfate alterations: mechanisms and relevance for proteinuria. Kidney Int 2000, 57(2):385-400.
12. Kanwar YS, Linker A, Farquhar MG: Increased permeability of the glomerular basement membrane to ferritin after removal of glycosaminoglycans (heparan sulfate) by enzyme digestion. J Cell Biol 1980, 86(2):688-693.
13. van den Born J, van den Heuvel LP, Bakker MA, Veerkamp JH, Assmann KJ, Berden JH: A monoclonal antibody against GBM heparan sulfate induces an acute selective proteinuria in rats. Kidney Int 1992, 41(1):115-123.
14. Rossi M, Morita H, Sormunen R, Airenne S, Kreivi M, Wang L, Fukai N, Olsen BR, Tryggvason K, Soininen R: Heparan sulfate chains of perlecan are indispensable in the lens capsule but not in the kidney. Embo J 2003, 22(2):236-245.
15. Harvey SJ, Jarad G, Cunningham J, Rops AL, van der Vlag J, Berden JH, Moeller MJ, Holzman LB, Burgess RW, Miner JH: Disruption of glomerular basement membrane charge through podocyte-specific mutation of agrin does not alter glomerular permselectivity. Am J Pathol 2007, 171(1):139-152.
16. Goldberg S, Harvey SJ, Cunningham J, Tryggvason K, Miner JH: Glomerular filtration is normal in the absence of both agrin and perlecan-heparan sulfate from the glomerular basement membrane. Nephrol Dial Transplant 2009, 24(7):2044-2051.
17. Miner JH: Developmental biology of glomerular basement membrane components. Curr Opin Nephrol Hypertens 1998, 7(1):13-19.
18. Timpl R, Dziadek M: Structure, development, and molecular pathology of basement membranes. Int Rev Exp Pathol 1986, 29:1-112.
19. Miner JH, Sanes JR: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequence, distribution, association with laminins, and developmental switches. J Cell Biol 1994,

127(3):879-891.
20. Harvey SJ, Zheng K, Sado Y, Naito I, Ninomiya Y, Jacobs RM, Hudson BG, Thorner PS: Role of distinct type IV collagen networks in glomerular development and function. Kidney Int 1998, 54(6):1857-1866.
21. Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG: Alport's syndrome, Goodpasture's syndrome, and type IV collagen. N Engl J Med 2003, 348(25):2543-2556.
22. Kalluri R, Shield CF, Todd P, Hudson BG, Neilson EG: Isoform switching of type IV collagen is developmentally arrested in X-linked Alport syndrome leading to increased susceptibility of renal basement membranes to endoproteolysis. J Clin Invest 1997, 99(10):2470-2478.
23. Abrahamson DR, Hudson BG, Stroganova L, Borza DB, St John PL: Cellular origins of type IV collagen networks in developing glomeruli. J Am Soc Nephrol 2009, 20(7):1471-1479.
24. Aumailley M, Bruckner-Tuderman L, Carter WG, Deutzmann R, Edgar D, Ekblom P, Engel J, Engvall E, Hohenester E, Jones JC et al: A simplified laminin nomenclature. Matrix Biol 2005, 24(5):326-332.
25. Noakes PG, Miner JH, Gautam M, Cunningham JM, Sanes JR, Merlie JP: The renal glomerulus of mice lacking s-laminin/laminin beta 2: nephrosis despite molecular compensation by laminin beta 1. Nat Genet 1995, 10(4):400-406.
26. Goldberg S, Adair-Kirk TL, Senior RM, Miner JH: Maintenance of Glomerular Filtration Barrier Integrity Requires Laminin alpha5. J Am Soc Nephrol.
27. Kriz W, Gretz N, Lemley KV: Progression of glomerular diseases: is the podocyte the culprit? Kidney Int 1998, 54(3):687-697.
28. Smoyer WE, Mundel P: Regulation of podocyte structure during the development of nephrotic syndrome. J Mol Med 1998, 76(3-4):172-183.
29. Huber TB, Benzing T: The slit diaphragm: a signaling platform to regulate podocyte function. Curr Opin Nephrol Hypertens 2005, 14(3):211-216.
30. Somlo S, Mundel P: Getting a foothold in nephrotic syndrome. Nat Genet 2000, 24(4):333-335.
31. Faul C, Asanuma K, Yanagida-Asanuma E, Kim K, Mundel P: Actin up: regulation of podocyte structure and function by components of the actin cytoskeleton. Trends Cell Biol 2007, 17(9):428-437.
32. Wharram BL, Goyal M, Gillespie PJ, Wiggins JE, Kershaw DB, Holzman LB, Dysko RC, Saunders TL, Samuelson LC, Wiggins RC: Altered podocyte structure in GLEPP1 (Ptpro)-deficient mice associated with hypertension and low glomerular filtration rate. J Clin Invest 2000, 106(10):1281-1290.
33. Kim YH, Goyal M, Wharram B, Wiggins J, Kershaw D, Wiggins R: GLEPP1 receptor tyrosine phosphatase (Ptpro) in rat PAN nephrosis. A marker of acute podocyte injury. Nephron 2002, 90(4):471-476.
34. Kaplan JM, Kim SH, North KN, Rennke H, Correia LA, Tong HQ, Mathis BJ, Rodriguez-Perez JC, Allen PG, Beggs AH et al: Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal segmental glomerulosclerosis. Nat Genet 2000, 24(3):251-256.
35. Asanuma K, Yanagida-Asanuma E, Faul C, Tomino Y, Kim K, Mundel P: Synaptopodin orchestrates actin organization and cell motility via regulation of RhoA signalling. Nat Cell Biol 2006, 8(5):485-491.
36. Asanuma K, Kim K, Oh J, Giardino L, Chabanis S, Faul C, Reiser J, Mundel P: Synaptopodin regulates the actin-bundling activity of alpha-actinin in an isoform-specific manner. J Clin Invest 2005, 115(5):1188-1198.
37. Tryggvason K, Wartiovaara J: How does the kidney filter plasma? Physiology (Bethesda) 2005, 20:96-101.
38. Rodewald R, Karnovsky MJ: Porous substructure of the glomerular slit diaphragm in the rat and mouse. J Cell Biol 1974, 60(2):423-433.
39. Liu G, Kaw B, Kurfis J, Rahmanuddin S, Kanwar YS, Chugh SS: Neph1 and nephrin interaction in the slit diaphragm is an important determinant of glomerular permeability. J Clin Invest 2003, 112(2):209-221.
40. Tryggvason K: Unraveling the mechanisms of glomerular ultrafiltration: nephrin, a key component of the slit diaphragm. J Am Soc Nephrol 1999, 10(11):2440-2445.
41. Sellin L, Huber TB, Gerke P, Quack I, Pavenstadt H, Walz G: NEPH1 defines a novel family of podocin interacting proteins. Faseb J 2003, 17(1):115-117.
42. Schnabel E, Anderson JM, Farquhar MG: The tight junction protein ZO-1 is concentrated along slit diaphragms of the glomerular epithelium. J Cell Biol 1990, 111(3):1255-1263.
43. Ruotsalainen V, Ljungberg P, Wartiovaara J, Lenkkeri U, Kestila M, Jalanko H, Holmberg C, Tryggvason K: Nephrin is specifically located at the slit diaphragm of glomerular podocytes. Proc Natl

Acad Sci U S A 1999, 96(14):7962-7967.
44. Roselli S, Gribouval O, Boute N, Sich M, Benessy F, Attie T, Gubler MC, Antignac C: Podocin localizes in the kidney to the slit diaphragm area. Am J Pathol 2002, 160(1):131-139.
45. Reiser J, Kriz W, Kretzler M, Mundel P: The glomerular slit diaphragm is a modified adherens junction. J Am Soc Nephrol 2000, 11(1):1-8.
46. Inoue T, Yaoita E, Kurihara H, Shimizu F, Sakai T, Kobayashi T, Ohshiro K, Kawachi H, Okada H, Suzuki H et al: FAT is a component of glomerular slit diaphragms. Kidney Int 2001, 59(3):1003-1012.
47. Pavenstadt H, Kriz W, Kretzler M: Cell biology of the glomerular podocyte. Physiol Rev 2003, 83(1):253-307.
48. Reiser J, Polu KR, Moller CC, Kenlan P, Altintas MM, Wei C, Faul C, Herbert S, Villegas I, Avila-Casado C et al: TRPC6 is a glomerular slit diaphragm-associated channel required for normal renal function. Nat Genet 2005, 37(7):739-744.
49. Tryggvason K, Patrakka J, Wartiovaara J: Hereditary proteinuria syndromes and mechanisms of proteinuria. N Engl J Med 2006, 354(13):1387-1401.
50. Zenker M, Machuca E, Antignac C: Genetics of nephrotic syndrome: new insights into molecules acting at the glomerular filtration barrier. J Mol Med 2009, 87(9):849-857.
51. Kestila M, Lenkkeri U, Mannikko M, Lamerdin J, McCready P, Putaala H, Ruotsalainen V, Morita T, Nissinen M, Herva R et al: Positionally cloned gene for a novel glomerular protein–nephrin–is mutated in congenital nephrotic syndrome. Mol Cell 1998, 1(4):575-582.
52. Boute N, Gribouval O, Roselli S, Benessy F, Lee H, Fuchshuber A, Dahan K, Gubler MC, Niaudet P, Antignac C: NPHS2, encoding the glomerular protein podocin, is mutated in autosomal recessive steroid-resistant nephrotic syndrome. Nat Genet 2000, 24(4):349-354.
53. Winn MP, Conlon PJ, Lynn KL, Farrington MK, Creazzo T, Hawkins AF, Daskalakis N, Kwan SY, Ebersviller S, Burchette JL et al: A mutation in the TRPC6 cation channel causes familial focal segmental glomerulosclerosis. Science 2005, 308(5729):1801-1804.
54. Mucha B, Ozaltin F, Hinkes BG, Hasselbacher K, Ruf RG, Schultheiss M, Hangan D, Hoskins BE, Everding AS, Bogdanovic R et al: Mutations in the Wilms' tumor 1 gene cause isolated steroid resistant nephrotic syndrome and occur in exons 8 and 9. Pediatr Res 2006, 59(2):325-331.
55. Hinkes B, Wiggins RC, Gbadegesin R, Vlangos CN, Seelow D, Nurnberg G, Garg P, Verma R, Chaib H, Hoskins BE et al: Positional cloning uncovers mutations in PLCE1 responsible for a nephrotic syndrome variant that may be reversible. Nat Genet 2006, 38(12):1397-1405.
56. Zenker M, Aigner T, Wendler O, Tralau T, Muntefering H, Fenski R, Pitz S, Schumacher V, Royer-Pokora B, Wuhl E et al: Human laminin beta2 deficiency causes congenital nephrosis with mesangial sclerosis and distinct eye abnormalities. Hum Mol Genet 2004, 13(21):2625-2632.
57. Hasselbacher K, Wiggins RC, Matejas V, Hinkes BG, Mucha B, Hoskins BE, Ozaltin F, Nurnberg G, Becker C, Hangan D et al: Recessive missense mutations in LAMB2 expand the clinical spectrum of LAMB2-associated disorders. Kidney Int 2006, 70(6):1008-1012.
58. Schena M, Shalon D, Davis RW, Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995, 270(5235):467-470.
59. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP: Accessing genetic information with high-density DNA arrays. Science 1996, 274(5287):610-614.
60. Chen YA, Chou CC, Lu X, Slate EH, Peck K, Xu W, Voit EO, Almeida JS: A multivariate prediction model for microarray cross-hybridization. BMC Bioinformatics 2006, 7:101.
61. Lipshutz RJ, Morris D, Chee M, Hubbell E, Kozal MJ, Shah N, Shen N, Yang R, Fodor SP: Using oligonucleotide probe arrays to access genetic diversity. Biotechniques 1995, 19(3):442-447.
62. Affymetrix I: Statistical Algorithms Description Document. 2002.
63. de Borst MH, Benigni A, Remuzzi G: Primer: strategies for identifying genes involved in renal disease. Nat Clin Pract Nephrol 2008, 4(5):265-276.
64. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: Serial analysis of gene expression. Science 1995, 270(5235):484-487.
65. Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW, Velculescu VE: Using the transcriptome to annotate the genome. Nat Biotechnol 2002, 20(5):508-512.
66. Affymetrix I: GeneChip Expression Analysis-Data Analysis Fundamentals. 2002.
67. Robert Gentleman VJC, Wolfgang Huber, Rafael A. Irizarry, Sandrine Dudoit: Bioinformatics and Computational Biology Solutions Using R and Bioconductor: Springer; 2005.
68. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatis-

tics 2003, 4(2):249-264.
69. Zhijin Wu RAI, Robert Gentleman, Francisco Martinez Murillo, Forrest Spencer: A Model Based Background Adjustment for Oligonucleotide Expression Arrays. Johns Hopkins University, Dept of Biostatistics Working Papers 2004.
70. Irizarry RA, Wu Z, Jaffee HA: Comparison of Affymetrix GeneChip expression measures. Bioinformatics 2006, 22(7):789-794.
71. Zhang L, Wu C, Carta R, Baggerly K, Coombes KR: Response to Preprocessing of oligonucleotide array data. Nat Biotechnol 2004, 22(6):658.
72. Tusher VG, Tibshirani R, Chu G: Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001, 98(9):5116-5121.
73. Pavlidis P: Using ANOVA for gene selection from microarray studies of the nervous system. Methods 2003, 31(4):282-289.
74. Storey JD, Tibshirani R: Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003, 100(16):9440-9445.
75. Slonim DK: From patterns to pathways: gene expression data analysis comes of age. Nat Genet 2002, 32 Suppl:502-508.
76. Frades I, Matthiesen R: Overview on techniques in cluster analysis. Methods Mol Biol, 593:81-107.
77. Kaufman L, Rousseeuw, Peter J.: Finding groups in data: An introduction to cluster analysis. New York: Wiley 1990
78. Segal NH, Pavlidis P, Noble WS, Antonescu CR, Viale A, Wesley UV, Busam K, Gallardo H, DeSantis D, Brennan MF et al: Classification of clear-cell sarcoma as a subtype of melanoma by genomic profiling. J Clin Oncol 2003, 21(9):1775-1781.
79. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D: Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 2000, 16(10):906-914.
80. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M, Jr., Haussler D: Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci U S A 2000, 97(1):262-267.
81. Noble WS: What is a support vector machine? Nat Biotechnol 2006, 24(12):1565-1567.
82. Kelder T, Pico AR, Hanspers K, van Iersel MP, Evelo C, Conklin BR: Mining biological pathways using WikiPathways web services. PLoS One 2009, 4(7):e6447.
83. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C: WikiPathways: pathway editing for the people. PLoS Biol 2008, 6(7):e184.
84. Vastrik I, D'Eustachio P, Schmidt E, Gopinath G, Croft D, de Bono B, Gillespie M, Jassal B, Lewis S, Matthews L et al: Reactome: a knowledge base of biologic pathways and processes. Genome Biol 2007, 8(3):R39.
85. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000, 25(1):25-29.
86. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005, 102(43):15545-15550.
87. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Gandhi TK, Chandrika KN, Deshpande N, Suresh S et al: Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res 2004, 32(Database issue):D497-501.
88. Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R et al: IntAct–open source resource for molecular interaction data. Nucleic Acids Res 2007, 35(Database issue):D561-565.
89. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G: MINT: the Molecular INTeraction database. Nucleic Acids Res 2007, 35(Database issue):D572-574.
90. Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G: MINT, the molecular interaction database: 2009 update. Nucleic Acids Res, 38(Database issue):D532-539.
91. Lehne B, Schlitt T: Protein-protein interaction databases: keeping up with growing interactomes. Hum Genomics 2009, 3(3):291-297.
92. Banas MC, Banas B, Hudkins KL, Wietecha TA, Iyoda M, Bock E, Hauser P, Pippin JW, Shankland SJ, Smith KD et al: TLR4 links podocytes with the innate immune system to mediate glomerular injury. J Am Soc Nephrol 2008, 19(4):704-713.

93. Wei C, Moller CC, Altintas MM, Li J, Schwarz K, Zacchigna S, Xie L, Henger A, Schmid H, Rastaldi MP et al: Modification of kidney barrier function by the urokinase receptor. Nat Med 2008, 14(1):55-63.
94. Reiser J, von Gersdorff G, Loos M, Oh J, Asanuma K, Giardino L, Rastaldi MP, Calvaresi N, Watanabe H, Schwarz K et al: Induction of B7-1 in podocytes is associated with nephrotic syndrome. J Clin Invest 2004, 113(10):1390-1397.
95. Haas C, Car B, Ryffel B, Le Hir M: Lipopolysaccharide-induced glomerulonephritis develops in the absence of interferon-gamma signaling. Exp Nephrol 1996, 4(4):222-230.
96. Fu Y, Xie C, Yan M, Li Q, Joh JW, Lu C, Mohan C: The lipopolysaccharide-triggered mesangial transcriptome: Evaluating the role of interferon regulatory factor-1. Kidney Int 2005, 67(4):1350-1361.
97. Takemoto M, He L, Norlin J, Patrakka J, Xiao Z, Petrova T, Bondjers C, Asp J, Wallgard E, Sun Y et al: Large-scale identification of genes implicated in kidney glomerulus development and function. Embo J 2006, 25(5):1160-1174.
98. Chabardes-Garonne D, Mejean A, Aude JC, Cheval L, Di Stefano A, Gaillard MC, Imbert-Teboul M, Wittner M, Balian C, Anthouard V et al: A panoramic view of gene expression in the human kidney. Proc Natl Acad Sci U S A 2003, 100(23):13710-13715.
99. He L, Sun Y, Patrakka J, Mostad P, Norlin J, Xiao Z, Andrae J, Tryggvason K, Samuelsson T, Betsholtz C et al: Glomerulus-specific mRNA transcripts and proteins identified through kidney expressed sequence tag database analysis. Kidney Int 2007, 71(9):889-900.
100. Higgins JP, Wang L, Kambham N, Montgomery K, Mason V, Vogelmann SU, Lemley KV, Brown PO, Brooks JD, van de Rijn M: Gene expression in the normal adult human kidney assessed by complementary DNA microarray. Mol Biol Cell 2004, 15(2):649-656.
101. Baelde HJ, Eikmans M, Doran PP, Lappin DW, de Heer E, Bruijn JA: Gene expression profiling in glomeruli from human kidneys with diabetic nephropathy. Am J Kidney Dis 2004, 43(4):636-650.
102. Takemoto M, Asker N, Gerhardt H, Lundkvist A, Johansson BR, Saito Y, Betsholtz C: A new method for large scale isolation of kidney glomeruli from mice. Am J Pathol 2002, 161(3):799-805.
103. Cuellar LM, Fujinaka H, Yamamoto K, Miyamoto M, Tasaki M, Zhao L, Tamer I, Yaoita E, Yoshida Y, Yamamoto T: Identification and localization of novel genes preferentially expressed in human kidney glomerulus. Nephrology (Carlton) 2009, 14(1):94-104.
104. Nystrom J, Fierlbeck W, Granqvist A, Kulak SC, Ballermann BJ: A human glomerular SAGE transcriptome database. BMC Nephrol 2009, 10:13.
105. Salomonis N, Hanspers K, Zambon AC, Vranizan K, Lawlor SC, Dahlquist KD, Doniger SW, Stuart J, Conklin BR, Pico AR: GenMAPP 2: new features and resources for pathway analysis. BMC Bioinformatics 2007, 8:217.
106. Baran D, Vendeville B, Ogborn M, Katz N: Cell adhesion molecule expression in murine lupus-like nephritis induced by lipopolysaccharide. Nephron 2000, 84(2):167-176.
107. He L, Sun Y, Takemoto M, Norlin J, Tryggvason K, Samuelsson T, Betsholtz C: The glomerular transcriptome and a predicted protein-protein interaction network. J Am Soc Nephrol 2008, 19(2):260-268.
108. Griffin SV, Olivier JP, Pippin JW, Roberts JM, Shankland SJ: Cyclin I protects podocytes from apoptosis. J Biol Chem 2006, 281(38):28048-28057.
109. Huber TB, Hartleben B, Kim J, Schmidts M, Schermer B, Keil A, Egger L, Lecha RL, Borner C, Pavenstadt H et al: Nephrin and CD2AP associate with phosphoinositide 3-OH kinase and stimulate AKT-dependent signaling. Mol Cell Biol 2003, 23(14):4917-4928.
110. Verma R, Wharram B, Kovari I, Kunkel R, Nihalani D, Wary KK, Wiggins RC, Killen P, Holzman LB: Fyn binds to and phosphorylates the kidney slit diaphragm component Nephrin. J Biol Chem 2003, 278(23):20716-20723.
111. Lahdenpera J, Kilpelainen P, Liu XL, Pikkarainen T, Reponen P, Ruotsalainen V, Tryggvason K: Clustering-induced tyrosine phosphorylation of nephrin by Src family kinases. Kidney Int 2003, 64(2):404-413.
112. Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009, 10(1):57-63.
113. Ramskold D, Wang ET, Burge CB, Sandberg R: An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol 2009, 5(12):e1000598.
114. Licatalosi DD, Darnell RB: RNA processing and its regulation: global insights into biological networks. Nat Rev Genet, 11(1):75-87.
115. Blow N: Transcriptomics: The digital generation. Nature 2009, 458(7235):239-242.
116. Treffer R, Deckert V: Recent advances in single-molecule sequencing. Curr Opin Biotechnol.

117. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A et al: mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 2009, 6(5):377-382.
118. McCaughan F, Dear PH: Single-molecule genomics. J Pathol, 220(2):297-306.
119. Ledford H: The death of microarrays? Nature 2008, 455(7215):847.
120. Voelkerding KV, Dames SA, Durtschi JD: Next-generation sequencing: from basic research to diagnostics. Clin Chem 2009, 55(4):641-658.