

## Network Analysis of Genes Regulated in Renal Diseases: Implications for a Molecular-Based Classification

**Suresh K. Bhavnani<sup>1,2\*§</sup> PhD, Felix Eichinger<sup>3\*</sup>, Sebastian Martini<sup>3\*</sup> MD, Paul  
Saxman<sup>2</sup>, H.V. Jagadish<sup>1,4</sup> PhD, Matthias Kretzler<sup>1,3</sup> MD**

<sup>1</sup>Center for Computational Medicine & Bioinformatics, 24 Frank Lloyd Wright Dr.,  
Domino's Farm, Ann Arbor, MI 48109-0738

<sup>2</sup>Michigan Institute for Clinical & Health Research, 24 Frank Lloyd Wright Dr.,  
Domino's Farm, Ann Arbor, MI 48106-0421

<sup>3</sup>Dept. of Internal Medicine, Division of Nephrology, University of Michigan Medical  
School, 1150 W. Medical Center Drive, MSRB2, SPC 5676 Ann Arbor, MI 48109-  
5676.

<sup>4</sup>Computer Science and Engineering, University of Michigan, 2260 Hayward, Ann  
Arbor, MI 48109-2121

\*These authors contributed equally to this work

§Corresponding author

Email addresses:

SKB: [bhavnani@umich.edu](mailto:bhavnani@umich.edu)

FE: [eichinge@med.umich.edu](mailto:eichinge@med.umich.edu)

SM: [sebmarti@umich.edu](mailto:sebmarti@umich.edu)

PS: [saxman@umich.edu](mailto:saxman@umich.edu)

HVJ: [jag@umich.edu](mailto:jag@umich.edu)

MK: [kretzler@umich.edu](mailto:kretzler@umich.edu)

# **Abstract**

## **Background**

Chronic renal diseases are currently classified based on morphological similarities such as whether they produce predominantly inflammatory or non-inflammatory responses. Such classifications do not reliably predict the course of the disease and its response to therapy. In contrast, recent studies in diseases such as breast cancer suggest that a classification which includes molecular information could lead to more accurate diagnoses and prediction of treatment response. This article describes how we extracted gene expression profiles from biopsies of patients with chronic renal diseases, and used network visualizations and associated quantitative measures to rapidly analyze similarities and differences between the diseases.

## **Results**

The analysis revealed three main regularities: (1) Many genes associated with a single disease, and fewer genes associated with many diseases. (2) Unexpected combinations of renal diseases that share relatively large numbers of genes. (3) Uniform concordance in the regulation of all genes in the network.

## **Conclusions**

The overall results suggest the need to define a molecular-based classification of renal diseases, in addition to hypotheses for the unexpected patterns of shared genes and the uniformity in gene concordance. Furthermore, the results demonstrate the utility of network analyses to rapidly understand complex relationships between diseases and regulated genes.

## Background

The rapid development of molecular biology and powerful analytical methods such as network analysis are enabling a shift in our understanding of diseases from a morphological (based on clinical and histological findings) to a molecular basis. This shift in focus has led to improvements in the classification of diseases [1, 2]. For example, gene expression analyses have been shown to improve prediction of treatment response in diseases such as breast cancer [3-5] and leukemia [6].

Unfortunately, relatively little is known about how renal diseases are similar and different at the molecular level. Currently, renal diseases are classified largely on morphological similarities. For example, systemic lupus erythematosus (SLE) is classified as a “predominant” *inflammatory* disease based on clinical findings, whereas focal and segmental glomerulosclerosis (FSGS) is classified as a “predominant” *non-inflammatory* disease based on histology. Several studies suggest that such morphology-based classifications could be significantly improved through the analysis of similarities and differences in gene expression, leading to more accurate diagnosis and targeted treatment options [4, 6, 7].

The analysis of gene expressions in chronic renal disease has either been studied at the level of a single renal disease [8], or by studying gene expressions across all known Mendelian disorders in the OMIM database [9]. The former obviously cannot reveal gene expressions that are common across renal diseases; the latter analyzed renal genes based on limited data, and at a high level (*glomerular* versus *tubular*), and therefore excluded important disease subcategories such as SLE. This article attempts to directly address the lack of understanding about gene expressions in chronic renal diseases. By using new data at the appropriate level of granularity, our goal was to

evaluate the current classification of renal diseases, and generate hypotheses about the molecular mechanisms underlying those diseases.

We begin by describing how we assembled a dataset of renal diseases and implicated genes, why and how we represented it using networks, and how we analyzed the networks using visualizations and quantitative verifications. We then discuss how the network analysis rapidly revealed unexpected overlaps of genes across the diseases. We conclude by discussing the utility of the network analysis approach to rapidly understand complex relationships, and the need to define a molecular-based classification of chronic renal diseases.

## **Methods**

Our research began with the question: *What are the molecular similarities and differences between chronic renal diseases?* If gene expressions occur in patterns that match the current classification of renal disease, then we can infer that the current classification is sufficient. However, if diseases have unexpected gene expression similarities, then we can infer that the current classification of renal diseases needs re-evaluation. To address our research question, we made critical decisions regarding *data selection, data representation and data analysis* as discussed below:

### **Data Selection**

Gene expression data were obtained from 106 patients with one of seven chronic renal diseases (classified in three categories as shown in Table 1) and compared to similar data obtained from biopsies of healthy kidney donors (control). Due to the rarity of three diseases (MCD, TMD, and DN, colored gray in Table 1) they currently have very small sample sizes (less than five) in the experimental and/or control conditions.

----- Table 1 about here -----

Microdissected renal tubuli in the biopsies underwent gene expression analysis of 12029 genes in each sample using Affymetrix HG-U133A microarrays. This analysis was done to identify the significantly regulated genes compared to pre-transplant living donor kidney tissues (controls) in each disease. A gene was considered to be significantly regulated in a disease if: (1) the difference of the normalized expression values between control and disease samples was significant at the 0.05 level (after correcting for multiple testing with the false discovery method), and (2) the regulation effect size (as defined by the log<sub>2</sub> fold change standardized by a pooled standard deviation) of that gene exceeded +0.3 (for up-regulated) or -0.3 (for down-regulated) when compared to controls. These statistical comparisons between experimental and controls were made within the same expression analysis batch to control for variations in equipment and context. The rigorous controls and tests resulted in a dataset of 747 genes significantly implicated in 7 renal diseases.

### **Data Representation**

Networks [10] have been used to represent a wide range of molecular phenomena related to human diseases [11]. These include networks to represent gene regulation [12], protein-protein interaction (PPI) [13], diseases-gene associations [9, 14], and disease-protein associations [15, 16]. Idekar and Sharan [15] identified four possible goals for the analysis of such disease networks: (a) identify network properties of disease genes such as the degree of differentially expressed proteins within a PPI network [17], (b) predict the role of disease-causing genes based on their relationship with existing known genes and proteins [18], (c) identify additional genes associated with particular diseases by analyzing the PPI network of known disease-causing

proteins [19], and (d) to identify highly predictive biomarkers that can be used to classify patients (e.g., those that have or do not have a disease) by identifying sets of biomarkers that are grouped within PPI networks [2]. However, none of these studies have used networks to analyze the similarities and differences between renal diseases based on significantly differentially regulated genes.

A network is a graph consisting of nodes and edges; nodes represent one or more types of entities (e.g., diseases or genes), and edges between the nodes represent a specific relationship between the entities (e.g., a disease is significantly correlated with a gene). Figure 1 shows a bipartite network (where edges exist only between two different types of entities) of diseases and their implicated genes. As shown, the bipartite network visually represents the explicit relationships between the 7 renal diseases and the 747 expressed genes. Furthermore, the size of a node is proportional to its *degree* (number of edges that connect to that node). Therefore, the larger a node, the more edges it shares with other nodes. Finally, the light edges (orange colored) and dark edges (blue colored) between diseases and genes are significantly up and down regulated respectively.

----- Figure 1 about here -----

Networks have three advantages for analyzing complex relationships. (1) They do not require *a priori* assumptions about the data, such as whether the data are hierarchically clustered or contain fuzzy clusters. Instead, by using a simple pair-wise representation of nodes and edges, networks enable rapid discovery of complex relationships using a single representation. (2) The specificity provided by the pair-wise representation between nodes can reveal details of relationships, for example

how specific diseases are connected to specific genes. (3) They can be rapidly visualized and analyzed using a set of network algorithms to reveal global regularities in the data. For example, Figure 1 shows how a *force-directed* layout algorithm [20] helps to visualize the relationship between diseases and genes. The algorithm pulls together nodes that are tightly connected, and pushes apart nodes that are not. As shown, the result is that diseases that share genes (e.g., FSGS and SLE in the center of Figure 1) are placed close to each other, and close to their implicated genes. Given these advantages, we used networks to explore the relationship between renal diseases and their implicated genes.

### **Data Analysis**

The insights from the network visualizations were verified using two standard quantitative network measures. (1) To provide a quantitative verification of the overall network topology observed in the visualization, we calculated the *degree* of each gene (number of diseases implicated with a gene) and plotted the *degree distribution* of the genes. (2) To understand more clearly how diseases had common gene expressions, we transformed the bipartite network to inspect only how the diseases shared genes in a method called a *one-mode projection* [10]. Here, all gene nodes were removed, and a weighted edge was placed between two diseases if they shared one or more genes as shown in Figure 3. The resulting network visually represented how pairs of diseases shared gene expressions, revealing how two diseases share more or less genes than expected in comparison to the current classification of renal diseases. The networks were created using *Pajek* (version 1.23) [21].

## Results

The bipartite network visualization revealed three critical patterns related to renal diseases and genes:

### 1. Many Specific Genes, Fewer Non-Specific Genes

There are a large number of *specific* genes in the periphery of the network that are connected to a single disease. These low degree nodes have been pushed out to the periphery by the force-directed layout algorithm due to their low connectivity with many diseases. In contrast, there are relatively fewer *non-specific* genes that are in the center of the network due to their high connectivity to many diseases. The degree distribution of genes (Figure 2) verifies this observation: more than half (54%) of the total 747 genes are implicated in just one disease, only two genes (0.3%) are implicated in 5 diseases, and no genes are implicated in 6 or all of the 7 diseases. This result provided the first glimpse into the pattern of overlap between diseases and implicated genes. The graph also shows the low number of overall genes for MCD, TMD, and DN which have been pushed to the periphery of the graph due to their low gene overlap. Because each of these diseases has a low sample size (as shown in Table 1) we removed them from the dataset to test its effect on the distribution. As shown in Figure 2, while the number of specific genes drops, the distribution still shows an overall pattern of many specific genes and fewer non-specific genes.

----- Figure 2 about here -----

### 2. Relationship between Diseases, Genes, and Regulation Type

The network visualization revealed a multidimensional relationship between diseases, genes, and regulation type. As shown in Figure 1, there exist three disease sets which share a disproportionately large number of genes: (a) the four dominant diseases

(SLE, FSGS, MGN, IgAN) on the right hand side of the network share 52 (88%) of the total 4-degree genes. These genes are mainly down-regulated. (b) A proper subset of the above disease set (SLE, FSGS, MGN) share 88 (79%) of the total 3-degree genes. These genes are mainly up-regulated. (c) A pair of diseases (SLE, FSGS) which overlap with the above sets share 130 (72%) of the total 2-degree genes. These genes are mainly up-regulated.

The gene expression overlap between diseases is verified by a one-mode projection on diseases. As shown in Figure 3, the one-mode projection removed the gene nodes, and placed edges between the diseases to correspond with how many genes each pair shares. The one-mode projection shows three dominant pairs of diseases (SLE-FGS, SLE-MGN, FGS-MGN) that share many genes. While the projection is not designed to reveal whether more than two diseases share the *same* genes, the network clearly shows the dominant relationship between SLE, FGS, MGN, and a less dominant relationship with IgAN. Future studies with larger samples of TMD, DN, and MCD should reveal how they relate to the other renal diseases.

----- Figure 3 about here -----

### **3. Uniform Concordance in Gene Regulation**

All the genes in the network, regardless of degree, are concordantly regulated. In other words, no gene was up-regulated in one disease, and down-regulated in another. This uniform concordance in gene regulation can be seen by the large areas of uniform color for edges connecting to high degree genes. Given the 100% uniformity of gene concordance, we re-examined the data to check for programmatic and bias errors, but found none. Furthermore, we examined another dataset containing biopsies from patients with acute renal failure. When the two data sets were merged, we found

two genes that were discordantly regulated. This suggests that the uniformity in gene regulation within chronic renal diseases (presented here) is most probably the result of similarity in biological mechanisms across chronic renal diseases, rather than a selection bias or error. Future research should further verify this conclusion.

## Discussion

Given that the data consisted only of the tubular compartment of renal biopsies, we expected to find a large number of non-specific (shared) genes. Instead, we found relatively few non-specific renal genes associated with a large number of diseases. This relationship held even when we removed diseases with a low sample size. Network analysis therefore helped to identify which genes might be involved in molecular mechanisms that are specific to a disease, and which genes are involved in common pathways activated in combinations of chronic renal diseases<sup>1</sup>. This approach could guide future research to identify specific and general drug therapies to treat kidney diseases.

Besides the distribution of specific and non-specific genes, the network analysis also revealed patterns of molecular similarity between diseases which do not match the current morphology-based classification of renal diseases. As shown in the first column of Table 1, SLE and IgAN belong to the class of *inflammatory* diseases. However, the network analysis revealed that SLE shares many more genes with FSGS

---

<sup>1</sup> It is important to note that the many specific genes in our dataset could be implicated in other renal diseases not included in our study, and therefore could be non-specific with respect to a wider scope of diseases.

and MGN (from the *non-inflammatory* class), compared to IgAN (from its own class). While molecular similarities between non-inflammatory and inflammatory renal diseases have been previously reported [22, 23] the unexpected finding was the strength of the association with members outside its class. Similarly, IgAN shares an equal number of genes with SLE (from its own class) as it does with FSGS, and MGN (from the *non-inflammatory* class). These results suggest that the current morphology-based classification of renal diseases does not match the pattern of shared tubulointerstitial gene expression, and therefore motivates future research to define a new molecular-based classification of renal diseases.

The above result also motivates future research to investigate how genes common to sets of disease can guide the identification of existing or new gene regulatory pathways [16]. For example, in a preliminary analysis we used *canonical pathways* (developed by Ingenuity<sup>®</sup> Systems) to search for existing regulatory pathways that best matched the genes shared by the three dominant disease sets. For the 59 genes shared by the disease set FSGS, MGN and IgAN, the search retrieved 45 canonical pathways that were significant at the 0.01 level. Many of these pathways (e.g., TGF- $\beta$ , JAK/STAT, NF- $\kappa$ B and VEGF) were experimentally-verified in these diseases. For example, involvement of the VEGF pathway is suggestive of vascular rarefaction as an underlying driving force for ischemic damage in renal failure [24], and is a potential biomarker for progressive renal disease. However, it is possible that shared genes in our network do not match known pathways, which would suggest the existence of new pathways yet to be discovered. These new pathways could be important in understanding the pathophysiology of the diseases.

Finally, the concordance in gene regulation and the fact that genes in the three dominant disease sets are either mostly up or down-regulated suggest that shared mechanisms have identical effects on gene regulation. This can be further investigated in the future by analyzing the properties of the shared genes using categories from the Gene Ontology database. Patterns in how the gene categories relate to different disease sets should lead to an understanding of this and other phenomena related to the type of gene regulation. The network analysis therefore led to several testable hypotheses about the underlying pathophysiological mechanisms involved in chronic renal disease.

## **Conclusions**

While networks have been used to analyze a wide range of molecular phenomena, they have not been used to analyze how genes are implicated in multiple renal diseases. Our analysis has enabled us to question the adequacy of the existing morphological based taxonomy of renal diseases. Furthermore, the analysis rapidly revealed useful biological insights, without requiring additional filtering to reveal complex but understandable relationships. This could be because the network was of medium size and density compared to many large networks, such as the Gnutella peer-to-peer file sharing network [25]. In addition, the resulting network quickly revealed overlapping, nested, and subset groups *from the same* representation, a result that would be difficult using traditional data mining techniques. However, it is important to note that like most data mining techniques, network analysis is essentially an exploratory tool, and most useful for generating hypotheses, which need rigorous testing using other techniques to arrive at definitive answers.

The limitation of the current analysis is the small sample sizes for three diseases, which new data will soon address. However, similar to the *Diseasome* project [9], studies that attempt to analyze gene expressions of many diseases simultaneously often have to deal with incomplete data. Networks are surprisingly useful for incomplete data because they enable us to visually inspect the data, and make appropriate choices for filtering and interpretation.

Our future research includes: (1) Using categories from the Gene Ontology database to annotate the genes in the bipartite network, with the goal of understanding why sets of shared genes have similar regulation type. (2) Analyzing the data using additional network analytical techniques [26], such as random network comparison [27, 28] and fuzzy cluster analysis [29, 30]. (3) Analyzing a network consisting of individual patients and expressed genes. The goal of analyzing individual patients is to construct a new classification of renal diseases using a bottom-up approach without the use of *a priori* disease classifications as was done in the current study. As previous research on biomarkers in renal diseases have stated, gene expression data should lead to a systematically constructed molecular-based classification, resulting in the identification of more targeted treatments for patients with chronic renal disease.

## **Authors' contributions**

SB conceived the initial idea to analyze renal diseases and regulated genes using bipartite and one-mode networks, and constructed them; FE extracted and formatted the data; SB, SM, FE, and PS analyzed the networks; MK supervised the project. SB, FE, SM, PS, HJ, MK wrote, discussed, revised, and approved the final manuscript.

## Acknowledgements

This study is funded in part by NIH grants # UL1RR024986, # U54 DA021519, P30 DK081943 and R01 DK079912, and was performed under the framework of the National Center for Integrative Bioinformatics (NCIBI) and the Applied Systems Biology Core of the O'Brien Kidney Research Center at the University of Michigan. We thank J. Kurlander, A. Ganesan and A. Feldman for their contributions, and the following members of the European renal cDNA Else Kroener Fresenius Bank (ERCB-KFB) for providing access to the samples used in this study: C. D. Cohen, M. Fischereeder, H. Schmid, M. Kretzler, and D. Schlöndorff (Munich); J. D. Sraer, and P. Ronco (Paris); M. P. Rastaldi, and G. D'Amico (Milano); F. Mampaso (Madrid); P. Doran, and H. R. Brady (Dublin); D. Mönks, and C. Wanner (Würzburg); A. J. Rees, P. Brown (Aberdeen); F. Strutz, and G. Müller (Göttingen); P. Mertens, and J. Floege (Aachen); N. Braun, and T. Risler (Tübingen); L. Gesualdo, and F. P. Schena (Bari); J. Gerth, and G. Wolf (Jena); R. Oberbauer, and D. Kerjaschki (Vienna); B. Banas, and B. K. Krämer (Regensburg); W. Samtleben (Munich); H. Peters, and H. H. Neumayer, (Berlin); K. Ivens, and B. Grabensee (Düsseldorf); R. P. Wüthrich (Zürich); V. Tesar, (Prague).

## References

1. Golub T, Slonim D, Tamayo P, Huard C, Gaasenbeek M, Mesirov J, Coller H, Loh M, Downing J, Caligiuri M, et al: **Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring.** *Science* 1999, **286**:531-537.
2. Chuang H, Lee E, Liu Y, Lee D, Ideker T: **Network-based classification of breast cancer metastasis.** *Molecular Systems Biology* 2007, **3**.

3. Wulfschlegel JD, Speer R, Pierobon M, Laird J, Espina V, Deng J, Mammano E, Yang SX, Swain SM, Nitti D, et al: **Multiplexed Cell Signaling Analysis of Human Breast Cancer Applications for Personalized Therapy.** *Journal of Proteome Research* 2008, **7**:1508-1517.
4. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415**:530-536.
5. Hall P, Ploner A, Björkholm J, Huang F, Lin C-Y, Liu E, Miller L, Nordgren H, Pawitan Y, Shaw P, et al: **Hormone-replacement therapy influences gene expression profiles and is associated with breast-cancer prognosis: a cohort study.** *BMC Medicine* 2006, **4**:16.
6. Cario G, Stanulla M, Fine B, Teuffel O, Neuhoff N, Schrauder A, Flohr T, Schafer B, Bartram C, Welte K, et al: **Distinct gene expression profiles determine molecular treatment response in childhood acute lymphoblastic leukemia.** *Blood* 2005, **105**:821-826.
7. Loscalzo J, Kohane I, Barabasi A-L: **Human disease classification in the postgenomic era: A complex systems approach to human pathobiology.** *Mol Syst Biol* 2007, **3**.
8. Martini S, Eichinger F, Nair V, Kretzler M: **Defining human diabetic nephropathy on the molecular level: Integration of transcriptomic profiles with biological knowledge.** *Reviews in Endocrine & Metabolic Disorders* 2008, **9**:267-274.

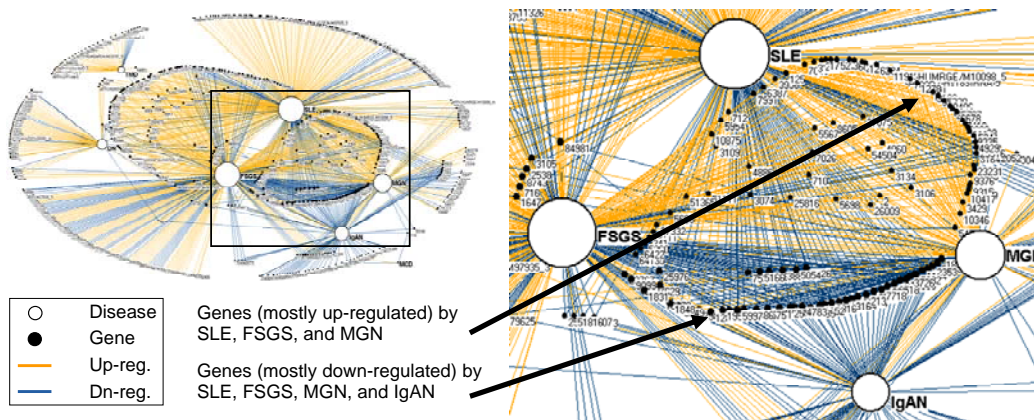
9. Goh K, Cusick M, Valle D, Childs B, Vidal M, Barabási A: **The human disease network.** *Proceedings of the National Academy of Sciences* 2007, **104**:8685.
10. Newman M: **The structure and function of complex networks.** *SIAM Review* 2003, **45**:167-256.
11. Junker BH, Schreiber F: *Analysis of Biological Networks (Wiley Series in Bioinformatics).* Wiley-Interscience; 2008.
12. Albert Rk: **Boolean Modeling of Genetic Regulatory Networks.** In *Complex Networks.* 2004: 459-481
13. Schwikowski B, Uetz P, Fields S: **A network of protein-protein interactions in yeast.** *Nat Biotechnol* 2000, **18**:1257-1261.
14. Oti M, Brunner H: **The modular nature of genetic diseases.** *Clinical genetics* 2007, **71**:1.
15. Ideker T, Sharan R: **Protein networks in disease.** *Genome Research* 2008, **18**:644.
16. Sam L, Liu Y, Li J, Friedman C, Lussier Y: **Discovery of protein interaction networks shared by diseases.** In *Pacific Symposium on Biocomputing.* 2007: 76-87.
17. Wachi S, Yoneda K, Wu R: **Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues.** vol. 21. pp. 4205-4208: Oxford Univ Press; 2005:4205-4208.

18. Oti M, Snel B, Huynen M, Brunner H: **Predicting disease genes using protein–protein interactions.** *Journal of medical genetics* 2006, **43**:691-698.
19. Pujana M, Han J, Starita L, Stevens K, Tewari M, Ahn J, Rennert G, Moreno V, Kirchhoff T, Gold B: **Network modeling links breast cancer susceptibility and centrosome dysfunction.** *Nature genetics* 2007, **39**:1338-1349.
20. Fruchterman T, Reingold E: **Graph drawing by force-directed placement.** *Software: Practice and Experience* 1991, **21**:1129-1164.
21. Batagelj V, Mrvar A: **Pajek - analysis and visualization of large networks.** In *Graph Drawing Software*. 2003: 77-103.
22. Ivanova L, Rudolph P, Shilov Y, Gieseler F, Alm P, Tareeva I, Proppe D: **Correlation between the expression of DNA topoisomerases I and IIalpha and clinical parameters in kidney disease.** *American journal of kidney diseases: the official journal of the National Kidney Foundation* 2001, **38**:1026.
23. Preston G, Waga I, Alcorta D, Sasai H, Munger W, Sullivan P, Phillips B, Jennette J, Falk R: **Gene expression profiles of circulating leukocytes correlate with renal disease activity in IgA nephropathy.** *Kidney international* 2004, **65**:420-430.
24. Lindenmeyer M, Kretzler M, Boucherot A, Berra S, Yasuda Y, Henger A, Eichinger F, Gaiser S, Schmid H, Rastaldi M: **Interstitial vascular**

- rarefaction and reduced VEGF-A expression in human diabetic nephropathy.** *Journal of the American Society of Nephrology* 2007, **18**:1765.
25. Ripeanu M, Iamnitchi A, Foster I: **Mapping the Gnutella Network.** *IEEE Internet Computing* 2002, **6**:50-57.
26. Costa, Rodrigues FA, Travieso G, Boas V: **Characterization of complex networks: A survey of measurements.** *Advances in Physics* 2007, **56**:167-242.
27. Wang X, Chen G: **Complex networks: small-world, scale-free and beyond.** *Circuits and Systems Magazine, IEEE* 2003, **3**:6-20.
28. Strogatz SH: **Exploring complex networks.** *Nature* 2001, **410**:268-276.
29. Zhang S, Wang R, Zhang X: **Identification of overlapping community structure in complex networks using fuzzy cc-means clustering.** *Physica A: Statistical Mechanics and its Applications* 2007, **374**:483-490.
30. Reichardt Jo, Bornholdt S: **Detecting Fuzzy Community Structures in Complex Networks with a Potts Model.** *Physical Review Letters* 2004, **93**.

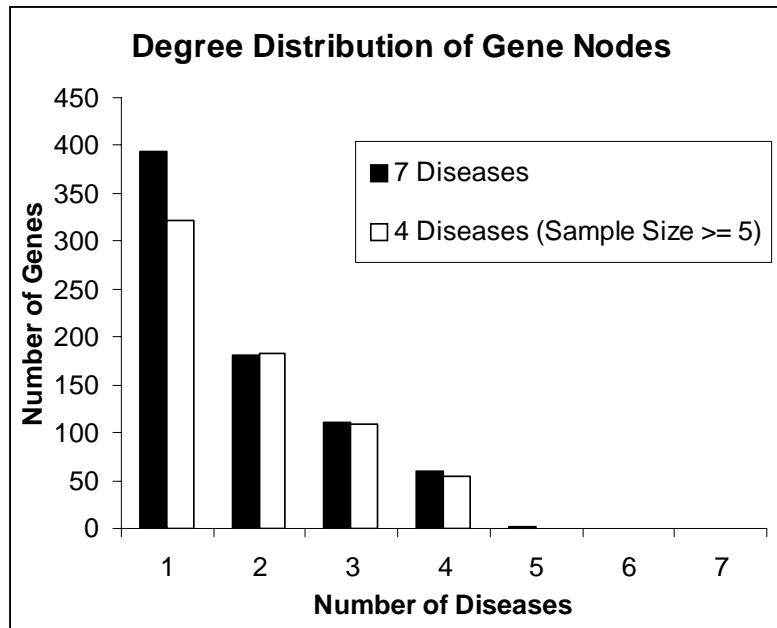
## Figures

Figure 1. Bipartite Network of Genes and Renal Diseases



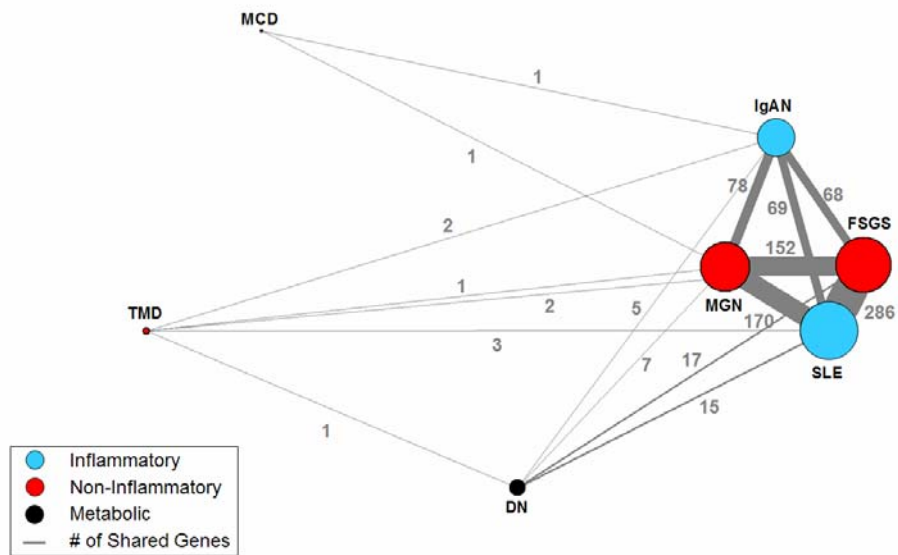
A bipartite network on the left (automatically generated by the *Fruchterman Rheingold* algorithm [20]) showing the relationship between 7 renal diseases (white nodes), and 747 genes (black nodes). The size of the nodes is proportional to the edges that connect to them. Therefore diseases with many genes have large nodes, whereas diseases with few genes have smaller nodes. The light edges (orange colored) represent up-regulated genes, the dark edges (blue colored) represent down-regulated genes, and gene node labels are the numerical identifiers for each gene from the *Entrez Gene* database. The inset shows that the genes regulated by all four of the high degree diseases (SLE, FSGS, MGN, and IgAN) are mostly down-regulated (shown as having mostly dark edges), whereas those regulated by a subset of the diseases shown (SLE, MGN, and FSGS) are mostly up-regulated (shown as having mostly light edges).

**Figure 2. Degree Distribution of Gene Nodes**



Degree distribution of the gene nodes for 7 diseases in the network showing the presence of more than half of the genes being regulated by one disease (specific genes), only two genes that are regulated by 5 diseases (non-specific genes), and no genes regulated by all diseases. The overall pattern of many specific genes and fewer non-specific genes holds even when 3 diseases with small sample sizes were excluded from the dataset.

**Figure 3. One-Mode Projection on Renal Diseases**



A one-mode projection on diseases of the network shown in Figure 1, showing how pairs of diseases share genes. The edge thickness is proportional to the number of shared genes (shown as numbers on the edges), and the size of the nodes is proportional to the sum of the edge weights incident to that node. Removal of diseases with small sample sizes (TMD, DN, and MCD) has no effect on the relationship between the other diseases.

## Tables

**Table 1. Current Classification for Chronic Renal Diseases, and the Number of Patients in the Experimental and Control Groups**

<b>Classification for Chronic Renal Diseases</b>	<b>Exp.</b>	<b>Ctrl.</b>
<i>"Predominant" Inflammatory</i>		
IgA Nephropathy ( <b>IgAN</b> )	25	6
Systemic lupus erythematosus ( <b>SLE</b> )	32	15
<i>"Predominant" Non-Inflammatory</i>		
Focal & segmental glomerulosclerosis ( <b>FSGS</b> )	10	15
Membranous glomerulonephritis ( <b>MGN</b> )	18	15
Minimal change disease ( <b>MCD</b> )	4	3
Thin membrane disease ( <b>TMD</b> )	6	3
<i>"Predominant" Metabolic</i>		
Diabetic nephropathy ( <b>DN</b> )	11	3
<b>Total</b>	<b>106</b>	<b>60</b>

The dataset consisted of biopsies from 106 patients diagnosed with one of 7 renal diseases, and compared with biopsies from 60 healthy patients within the same expression analysis batch. A test of significance between the two groups resulted in 747 genes implicated across the 7 chronic renal diseases.