A. Specific Aims

Diabetic nephropathy (DN) is an end organ complication responsible for significant morbidity and mortality in type 1 and type 2 diabetes mellitus. Association of genes and gene products with DN are currently defined in humans using genome wide association studies and renal tissue based gene expression studies (1). However, animal models are instrumental to define the functional impact of the associated molecules. The AMDCC has established an extensive data set defining the framework for DN in mice (2). A crucial next step will be to integrate the murine molecular information with corresponding comprehensive human data sets.

We have developed and employed a webbased analysis engine, Nephromine, for integrated systems biology analysis of renal gene expression for the renal research community. Nephromine accesses highly annotated human renal gene expression datasets in multiple search modes (see tutorial at nephromine.org), though Nephromine has been limited to human datasets to date. The goal of this application is to integrate murine DN gene expression data sets generated by the AMDCC and the renal research community at large with the human DN data sets to define shared and specific mechanism between mouse and man.

Specific aims to be addressed are:

Aim 1. Identify and annonate murine gene expression data sets of DN from AMDCC data repository, Gene Expression Omnibuts (GEO) and those accessed by direct interaction with investigative teams.

Aim 2. Compare murine DN gene expression data sets with human DN data sets and identify shared transcriptional changes on a network level.

Aim 3. Upload murine data sets into Nephromine for human-mouse comparative analysis by the renal research community.

Comparing the renal transcriptiome of human DN to the carefully characterized AMDCC murine DN models will allow the renal research community to define molecular mechanism shared between mouse and man. With this information investigators will be able to select the murine models that recapitulate the human disease mechanisms to be interrogated in their murine experimental studies. Defining mouse models *a priori* for their ability to recapitulate human disease will be a crucial step to increase comparability of murine studies to human disease mechanisms. This will aid in defining novel therapeutic targets for their efficacy in the specific murine models of DN and should decrease the attrition rate of compounds with activity in murine DN in their further validation as therapeutic targets of human DN.

B. Studies and Results

Specific Aim 1. *Identify and annonate murine gene expression data sets of DN from AMDCC data repository, GEO and those accessed by direct interaction with investigative teams.*

Hypothesis: Stage specific genome-wide transcriptional profiles in DN will allow identification of molecular pathways of disease initiation and progression.

Gene expression data sets were obtained from the AMDCC, public data repositories (GEO, Array Express, Stanford Microarray Database) and by direct interaction with investigators, if data were not deposited into the the public domain. For cross-species mapping, Affymetrix GeneChip Mouse Genome arrays image files were obtained and subsequently normalized using the RMA (Robust Multichip Average) approach and Mouse Entrez Gene custom CDF annotation version 10 (http://brainarray.mbni.med.umich.edu/Brainarray/default.asp) in GenePattern pipeline (http://www.GenePattern.org, Broad Institute,MIT)(7). Starting from the Affymetrix 430 2.0 plattform, 14501 of 16539 expressed mice genes were mapped to their respective putative human homologs using the NCBI homolog database (ftp://ftp.ncbi.nih.gov/pub/HomoloGene/current/, as of Feb 15, 2011).

If raw expression files of non-Affymetrix based expression data sets were available, files were annotated using custom CDF EntrezGene and then mapped to human using the same strategy as above. If only access to the expresssion matrix was available, GenBank Accession numbers were mapped to their Entrez GeneIDs using NCBI utilities and then mapped to the human orthologs. Table 1 provides the status of the currently identified and available murine gene expression data sets.

Table1: Murine DN data sets integrated into Nephromine

Model	Platform:	Reference Series:	Sample #	Diabetes	Background	Source
DN type II(DB/DB)	Affymetrix 430A	AMDCC website	21	8 wks	C57BLKs/J	AMDCC
DN type II (DB/DB)	Affymetrix U74 V.2	GSE642	12	16 wks	BKS	PMID: 15075187
DN type I (STZ)	Affymetrix 430 2.0	GSE33744	17	22 wks	DBA	our group
DN type II (DB/DB)	Affymetrix 430 2.0	GSE33744	10	24 wks	BKS	our group
DN type II (DB/DB eNOS-/-)	Affymetrix 430 2.0	unpublished	11	10 wks	BKS	our group
DN type II (DB/DB eNOS-/-)	Affymetrix 430 2.0	GSE33744	12	20 wks	BKS	our group
DN type I (STZ)	cDNA microarray	GSE710	10	18 wks	C57/B6	PMID: 14988265
DN type II (DB/DB)	cDNA microarray	GSE711	40	24 wks	C57BLKs/J	PMID: 14988265

These data sets have been annotated in Nephromine and are slated for release with the next data push of Nephromine..

Specific Aim 2. *Compare murine DN gene expression data sets with human DN data sets and identify shared transcriptional changes on a network level.*
Hypothesis: Pathway level analysis of networks conserved between mouse and man allows researchers the identification of shared disease mechanism not resolved by single gene comparison.

We performed an unbiased transcriptomic comparison of glomerular gene expression in diabetic humans and mouse models of DN. By generating glomerular gene expression profiles from diabetic humans and comparing these results with three well-characterized mouse models, we identified shared human-mouse cross-species glomerular transcriptional networks in DN. For the human-specific network, we compared glomerular gene expression in diabetic subjects who had high albuminuria (Halb) with those who had low albuminuria (Lalb) at the time of biopsy. For mouse-specific networks, we compared glomerular gene expression in diabetic mice with their appropriate non-diabetic controls. In the networks, genes were represented as gene nodes with connections that indicate functional dependencies.

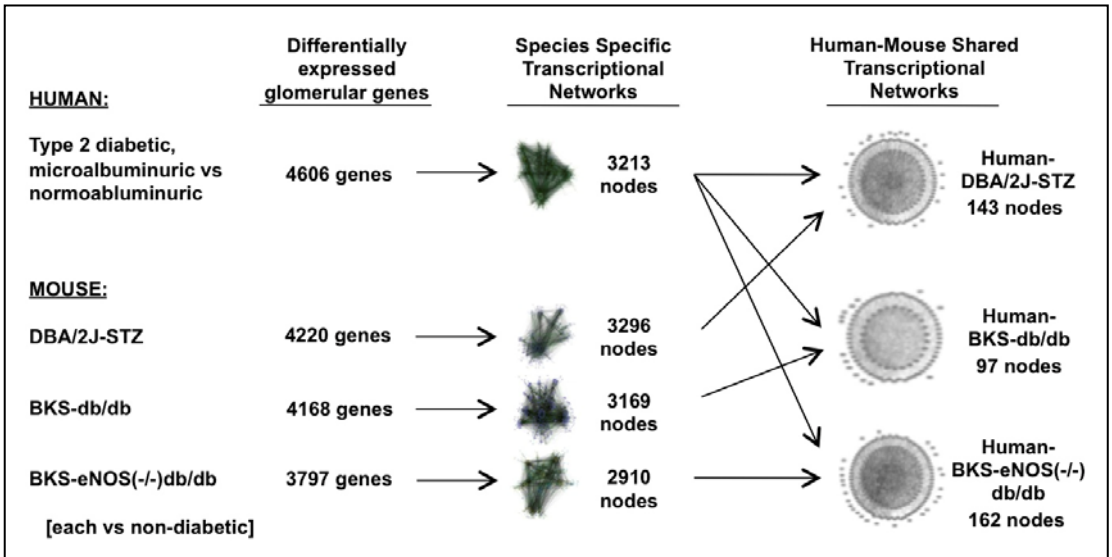

Figure 1. Cross-species transcriptional network generation using Bibliosphere networks and approximate graph-matching tool TALE. Strategy for integration of glomerular expression profiles of earl human DN with three AMDCC DN mouse models. (Hodgin et al. Diabetes, 2013)

The human and mouse networks were the overlaid to identifiy similar subnetworks operational in both species (Figure 1). Because all human subjects were diabetic and grouped according to progression (i.e. high and low albuminuria), we could identify shared human-mouse glomerular genes enriched for their specific contribution to nephropathy.

Nine genes were shared among the high connectivity gene nodes in all 3 networks (see Hodgin et al. Diabetes, 2013). These included STAT1 and STAT3, members of the JAK/STAT pathway, and genes expressed by endothelial cells and associated with endothelial cell dysfunction, including CD34, CD36, and FLT1. BCL2, FOS and FN1 were the three most connected gene nodes in the Human-DBA STZ and the Human-BKS eNOS-/- db/db networks. However these genes were not in the Human-BKS db/db network. PPARG (PPAR□) and MET (hepatocyte growth factor receptor) were found only in the Human-DBA STZ and Human-BKS db/db networks, whereas NR3C1, which encodes the glucocorticoid receptor, was shared only between Human-BKS db/db and Human-BKS eNOS-/- db/db networks.

Additionally, we used pathway enrichment analysis (GePS, Genomatix) of the gene node list from the human-mouse comparison. This approach assesses the biologic relevance of each cross-species network (Table 3). Canonical JAK-STAT signaling was the most enriched and shared by all three networks. VEGF receptor, FGF signaling, and HIF-1 gene regulation pathways, each with well characterized roles in DN, were also enriched among all three shared networks. The IL-7 signaling pathway was also enriched across all three. This pathway has not previously been implicated in the pathogenesis of DN. Finally, Table 3 lists top canonical pathways shared by any two networks and those that appear to be unique to network, suggesting important differences that will aid the research community in understanding the utility of separate models of DN.

Our human-mouse shared glomerular transcriptional networks will assist DN researchers in selecting mouse models most relevant to the human disease process of interest. Moreover, they will allow identification of new pathways shared between mouse models and humans.

Table 3: Top canonical pathways

Canonical pathway	# Total	Human-DBA STZ # Genes	Human-BKS db/db # Genes	Human-BKS eNOS-/- db/db # Genes
JAK/STAT pathway and regulation	272	**27**	**17**	**21**
AKT(PKB)-Bad signaling	178	**18**	**12**	**13**
IL-7 signaling pathway	177	**15**	**9**	**11**
Migration (VEGF)	180	**15**	**9**	**11**
VEGFR1 and VEGFR2 signaling	66	**11**	**8**	**8**
FGF signaling pathway	63	**9**	**4**	**6**
HIF-1-alpha transcription factor network	68	**8**	**4**	**7**
Angiopoietin receptor Tie2-mediated signaling	49	**6**	**5**	**---**
Hepatocyte Growth Factor Receptor signaling	55	**6**	**4**	**---**
Regulation of nuclear SMAD2/3 signaling	83	**---**	**5**	**9**
Regulation of androgen receptor activity	52	**---**	**6**	**5**
IL6-mediated signaling events	45	**8**	**---**	**8**
il-2 receptor beta chain in T cell activation	49	**6**	**---**	**7**
EGFR1 signaling	157	**10**	**---**	**---**
Signaling events mediated by PTP1B	52	**7**	**---**	**---**
Alpha6Beta4 Integrin	49	**---**	**5**	**---**
Endothelins	61	**---**	**5**	**---**
C-MYB transcription factor network	88	**---**	**---**	**9**
Androgen Receptor	87	**---**	**---**	**8**

Specific Aim 3. *Upload human-murine shared transcriptional networks into Nephromine for analysis by the renal research community.*

Hypothesis: A web based systems biology search engine integrating human and murine data sets will facilitate translational research in DN.

We have integrated the experimental data sets obtained in Aim 1 and 2 with the analysis capabilities currently available at the online search tool of renal genome-wide gene expression data sets at www.nehpromine.org. Datasets generated in Aim 1 have been annotated with the experimental parameters available from the murine study (DN status, gender, age, albuminuria,…). Datasets generated in Aim 2 are in the process of being uploaded as a shared human-murine transcriptome map for easy access and interpretation by clinician-scientists and molecular biologists.

Capabilities of Nephromine include automated searches of GO Molecular Function, Cellular Components and Biological Processes, KEGG pathways, Therapeutic Target Databases, prediction of microRNA targets, HPRD Interaction Sets, Literature-defined concepts and the Interactome database. A meta-analysis tool (metamap) allows for comparison of signatures from independent, but related studies. For further options and details of the currently expanding set of algorithms see www.nephromine.org. Use of Nephromine is free of charge and allows unrestricted use by the academic user.

We have delivered a data mining tool empowering the non-bioinformatic scientist to search large datasets of stage- specific, conserved human-mouse gene expression signatures in a specific and focused research context.

C. Significance

These data have established the first comprehensive maps of diabetic nephropathy in a stage specifici manner. The associated publication in Diabetes (Hodgin et al. 2013) presents these results to the scienfitic community and the respective data sets have been uploaded into GEO. Our work is received considerable feedback from the diabetes research community, summarized an edtiorial in by Dres. Bhalla, Velez and Chertow in Diabetes summarized our work as follows:

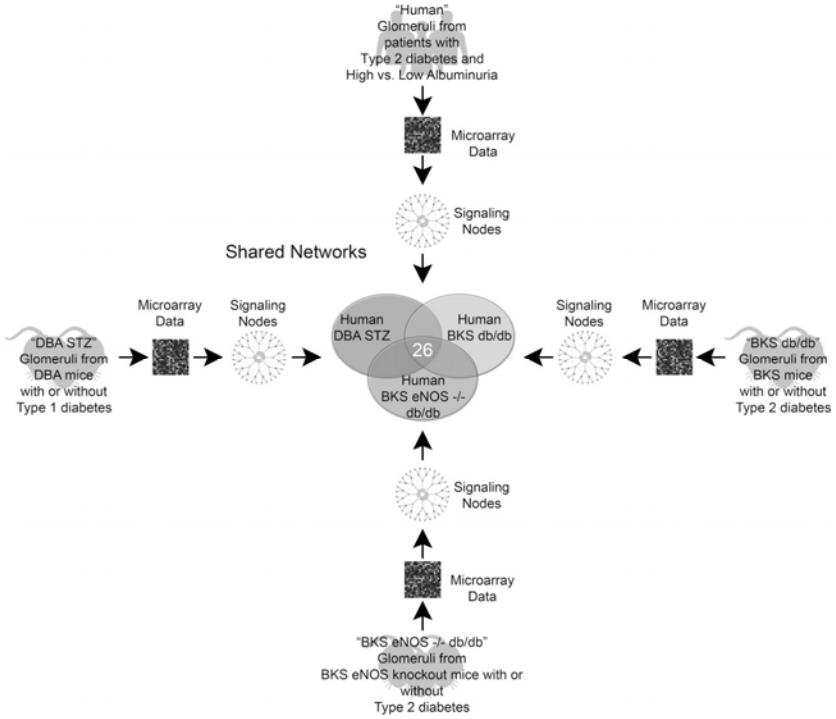

Fig.2: Summary of our study from Ballah et al. Diabetes, 2013.

" In summary, Hodgin et al. provide exciting, novel data that will allow investigators to compare gene expression profiles across different study cohorts and across species. By focusing on glomerular gene expression, the authors have enriched these profiles for specific gene products, which may be missed in transcriptome studies from the whole kidney as glomerular RNA comprises less than 5% of the total transcripts within the kidney (16), yet histological changes in glomeruli are the first visible signs of this disease (17). Investigators interested in studying a pathway relevant to human disease (e.g., epidermal growth factor signaling) can now choose from a menu of mouse models. Finally, patients with diabetes and kidney disease are rarely biopsied and thus, while cancers and other diseases have long been categorized by their molecular phenotype (18), kidney disease in patients with diabetes is crudely defined. This work may help to redefine the taxonomy of diabetic kidney disease based on glomerular gene expression rather than on nonspecific markers such as albuminuria and

serum creatinine. The authors have painstakingly derived, organized, and now shared a wealth of transcriptional data. Future experiments should build upon this impressive foundation and usher in an era of unprecedented progress in the treatment (and prevention) of diabetic kidney disease." (Bhalla et al. Diabetes, 2013). We are following up on our studies in multiple collaborations to continue to contribute to this critical progress in translational research in diabetic kidney disease.

D. Literature Cited

1. Hodgin JB, Nair V, Zhang H, Randolph A, Harris RC, Nelson RG, Weil EJ, Patel JM, Brosius FC, Kretzler M: Identification of cross-species shared transcriptional networks of diabetic nephropathy in human and mouse glomeruli. Diabetes, 62, 299-308, 2013.
2. Bhalla V, Velez MG, Chertow GM: A Transcriptional Blueprint for Human and Murine Diabetic Kidney Disease. Diabetes, 62, 231-3, 2013.
3. Keller B, Martini S, Sedor J, Kretzler M: Linking variants from genome-wide association analysis to function via transcriptional network analysis. *Semin Nephrol* 30:177-184, 2010
4. Brosius FC, 3rd, Alpers CE, Bottinger EP, Breyer MD, Coffman TM, Gurley SB, Harris RC, Kakoki M, Kretzler M, Leiter EH, Levi M, McIndoe RA, Sharma K, Smithies O, Susztak K, Takahashi N, Takahashi T: Mouse models of diabetic nephropathy. *J Am Soc Nephrol* 20:2503-2512, 2009

E. Vertebrate Animals

No de novo studies using vertebrate animals were performed for this application.

F. Sharing Plan

Nephromine platform is the main vehicle to share data sets generated in this application.