AMDCC Pilot Grant

Title: Murine-Human Transcriptomic Comparisons in Diabetic Nephropathy and Neuropathy

Authors: Brosius, Frank C.; Feldman, Eva; Kretzler, Matthias.

Progress Report

Background and Significance

Studying molecular networks in multiple relevant tissues and associating them with clinically relevant phenotype data to identify the networks driving disease is at the center of what has been loosely defined as 'systems biology' (1). By taking a holistic approach, systems biology aims to better understand the complex interplay among tissues, molecular networks, and environment that leads to endorgan damage. The challenge for systems biology is to decipher the biological function of individual genes, pathways, and networks that drive complex phenotypes, and human diabetic complications are prime examples of this challenge. Such efforts require integration of the diverse and manifold data pertaining to the complex system under study (for review see (2)).

In this pilot proposal we are employing analysis of co-expressed transcriptional networks leveraging pairwise interaction data among genes. This allows us to represent general relationships among genes in a comprehensive fashion. Central to this type of network is the identification of modules comprised of highly interconnected sets of genes, called nodes. These data sets form the functional units of the networks that are associated with complex phenotypes such as diabetic complications.

To obtain causal relationships among genes, and between genes and clinical phenotypes, simple networks comprised of only a handful of genes and clinical phenotypes will be generated by cross-species comparisons. By integrating gene expression with known and predicted transcriptional regulation and clinical phenotype data, the resulting networks, which we will refer to as "transcriptional networks" in this application, are capable of representing direction (e.g., genes regulated by transcription factors that show regulated expression) along the edges (connections) of the network, unlike co-expression networks, which do not show directionality. It is this directionality among the edges of the network that provides the causal information both among genes and between genes and clinical phenotypes (3).

A. Specific Aims

The overarching hypothesis of this application states that the identification of mechanisms of diabetic end-organ damage conserved between mice and humans will greatly facilitate the translation of findings from animal models to human disease and the development of improved murine models of diabetic complications.

Specific Aim 1: Define transcriptional networks activated in humans and in various AMDCC models of murine diabetic complications using transcriptional pathway mapping and promoter modeling tools.

We will generate annotated transcriptional microarray data for DN and DPN models. PIs who developed the models could then validate changes by mRNA and protein detection methods.

Specific Aim 2: Compare transcriptional networks between humans and mice with diabetic complications and define key pathways to be altered in murine models.

Using the regulatory networks defined in Aim 1 we will generate comparative subnetworks between human and murine DN and DPN. Conserved transcriptional networks could be screened for their suitability as therapeutic targets in murine models and ultimately human disease whereas divergent networks could be targets for transgenic or knockout approaches to further "humanize" the murine models.

We would provide initial network analysis that would be made available to all PIs for comparison to human analyses in Aim 3.

Specific Aim 3: Utilize a web-based search tool for effective mining of the data sets generated in Aims 1 and 2 by the diabetes complication research community.

A freely accessible internet-based search engine at www.nephromine.org, will empower the AMDCC PIs to effectively extract information from the DN and DPN datasets with direct relevance for design and execution of ongoing research studies in diabetic complications.

We would provide training to AMDCC investigators in an individualized, problem orientated manner.

B. Studies and Results

Specific Aim 1: Define transcriptional networks activated in human and murine diabetic complications.

In the past funding year, significant progress has been achieved in Specific Aim 1. In addition to the human (PIMA Indian protocol biopsies) differentially regulated gene sets derived from microdissected renal biopsies, we have generated transcriptional networks from gene sets using three well-described murine models of diabetic nephropathy (4,6) that we have successfully used (6-8), the db/db mouse, the streptozoticin-treated DBA (DBA-STZ) mouse, and the $eNOS^{-/-}$ db/db mouse. ChipInspector (Genomatix suite) (9) was used to identify significantly regulated genes (FDR < 1%) in glomeruli from experimental animals versus control. Analysis of glomerular RNA from 5 db/db mice versus 5 controls yielded 4708 differentially regulated genes (2186 up, 2522 down). Analysis of glomerular RNA from 4 DBA/STZ mice versus 3 controls yielded 1883 differentially regulated genes (1098,up, 785 down). Data analysis of the $eNOS^{-/-}$ db/db mouse is currently ongoing. Transcriptional pathway maps are being generated using natural language processing combined with promoter analysis and canonical pathway mapping. Canonical pathways overrepresented in the db/db gene set include extracellular matrix-receptor interaction, pyruvate metabolism, and valine-leucine-isoleucine degradation. In the DBA-STZ gene set, overrepresented canonical pathways include ECM-receptor interaction, leukocyte transendothelial migration, and cell adhesion molecules.

In a parallel study, significantly regulated genes have been identified in human sural nerve biopsies collected from a series of multicenter double-blind, placebo-controlled pharmaceutical trial of human DPN (10,11). Of the 4680 differentially expressed genes in the group of patients with progressive DPN versus non-progressive DPN, 1881 are up-regulated and 2799 genes are down-regulated. In db/db mice, the sciatic nerve excised after 24 weeks of diabetes yielded 7263 differentially expressed genes compared to control (3898 up, 3365 down).

Specific Aim 2: Compare transcriptional networks between humans and mice with diabetic complications and define key pathways to be altered in murine models.

We have constructed conserved transcriptional pathways using the DN gene sets described above and a gene set derived human DN using glomeruli from albuminuric PIMA Indians versus non-albuminuric PIMA Indians (i.e. progressors vs non-progressors) by overlaying human and murine transcriptional. Our goal is to identify shared transcriptional networks between humans and mice and between the mouse models to define conserved regulatory pathways in DN. This gene set includes 4630 significantly regulated genes (231 up, 4399 down). A multi-dimensional data integration approach was used to generate transcriptional networks for human and murine genesets. Network nodes were assigned to differential regulated transcripts in DN and edges assigned via knowledge of transcriptional relationships and/or automated sequence analysis of promoter regions of corresponding transcriptional factors using the Bibliosphere tool co-developed and available from our collaborator Genomatix at www.genomatix.com. The human glomerular DN transcriptional network was overlaid with each mouse network (db/db and DBA-STZ) to construct two shared networks based on conserved regulatory networks in DN. This was generated by using TALE (Tool for approximate large graph matching), which

was developed by our collaborator Dr. Patel, using human-mouse ortholog assignments extracted from NCBI database to define subnodes shared between models and species (12).

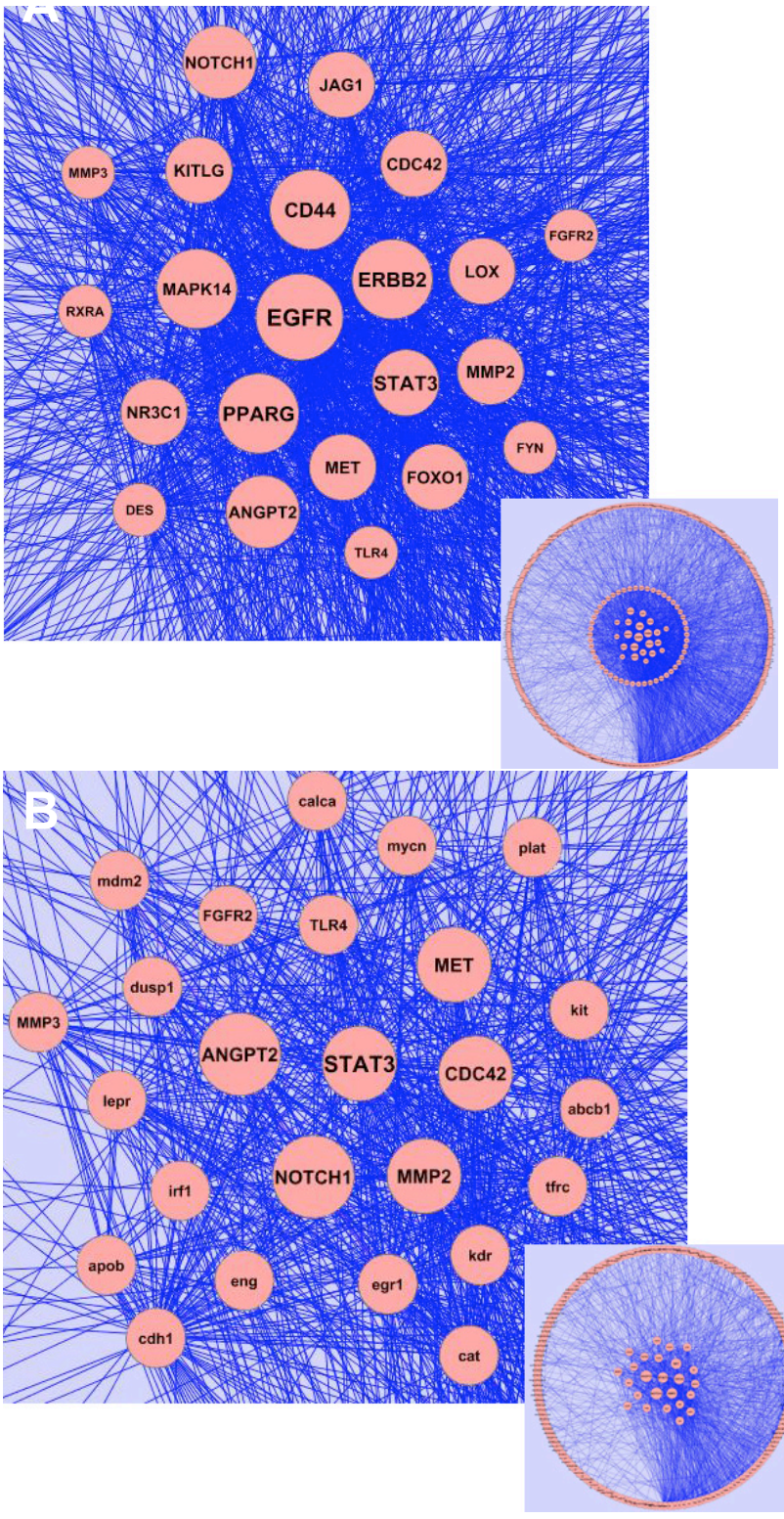

Figure 1: Cross-species conserved glomerular transcriptional network between human (PIMA) and db/db mouse DN (A) and human and DBA-STZ mouse (B). Transcriptional networks of genes regulated in human and db/db mouse DN from isolated glomeruli were compared using the graphical matching tool TALE. Nodes indicate aligned human-mouse gene pairs and the edges (blue lines) represent the conserved interactions. A subset of nodes with the most connections is shown in a grid format while the inset figures display the Genes entire conserved network.

The Human-db/db mouse network revealed 447 nodes and the Human-DBA-STZ mouse network has a187 nodes. Figure 1 highlights the key nodes (conserved genes) that have the most connections with other nodes. These nodes represent key hubs of cross-species conserved regulatory elements and will

serve as starting points for further analysis. The data set containing the gene-gene relationships is provided in the attached table. The Human-db/db mouse network contains several nodes that reflect known pathway activations such as JAK/STAT (jak3, stat3), notch (jag1, notch1), and PPAR signaling (pparg). The Human-DBA-STZ network shares 1/3 of its nodes with the Human-db/db network. However, the most overrepresented canonical pathways are shared and include adherens junction; valine, leucine, and isoleucine degradation; and PPAR signaling.

Specific Aim 3: Utilize a web-based search tool for effective mining of the data sets generated in Aims 1 and 2 by the diabetes complication research community.

Data obtained from the experiments described above and from gene expression data sets on diabetic end-organ damage generated by all AMDCC investigators as well as the diabetes research community at large are currently being processed by us for upload into the systems biology search tool nephromine.

The NIDDK team performing the Pima Indian interventional trial providing the biopsy material for our human gene expression studies informed us that due to constraint of their IRB, only aggregate data on the cohort but no data points from individual patients can be made publicly available. As Nephromine relies on correlation and description of individual data points for correlative analysis, we currently are exploring several strategies to be able to make the data sets available to the scientific community in a matter that is consistent with the IRB.

Our preferred strategy will be to report from the human-mouse conserved pathways the murine expression data points and their correlation with murine renal functional parameters. This would allow to us take advantage of all functionalities of Nephromine, but not reveal any individual human subject specific gene expression information. We are currently working on the implementation of the murine – human cross annotation and will upload the respective data sets into Nephromine.

In summary, during the pilot program we were able to generate genome wide expression profiles of several murine models of diabetic microangiopathy, generate transcriptional networks from these models, integrate the murine with the human regulatory pathways on a network level and provide a list of main network nodes to the AMDCC and shortly to the scientific community at large.

References:

1. Schadt EE, and Lum PY. 2006. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Reverse engineering gene networks to identify key drivers of complex disease phenotypes. *J Lipid Res* 47:2601-2613.
2. Quackenbush J. 2007. Extracting biology from high-dimensional biological data. *J Exp Biol* 210:1507-1517.
3. Yasuda Y, Cohen CD, Henger A, and Kretzler, M. 2006. Gene expression profiling analysis in nephrology: towards molecular definition of renal disease. *Clin Exp Nephrol* 10:91-98.
4. Breyer MD, Bottinger E, Brosius FC, 3rd, Coffman TM, Harris RC, Heilig CW, Sharma K. 2005. Mouse models of diabetic nephropathy. *J Am Soc Nephrol* 16:27-45.
5. Zhao HJ, Wang S, Cheng H, Zhang MZ, Takahashi T, Fogo AB, Breyer MD, Harris RC. Endothelial nitric oxide synthase deficiency produces accelerated nephropathy in diabetic mice. 2006. *J Am Soc Nephrol* 17(10):2664-9.
6. Berthier CC, Zhang H, Schin ML, Henger A, Nelson RG, Yee B, Boucherot A, Carter-Su C, Argetsinger LS, Rastaldi MP, Brosius FC, Kretzler M. 2009. Enhanced Expression of JAK-STAT Pathway Members in Human Diabetic Nephropathy. *Diabetes* 58:469-77.
7. Zhang H, Saha J, Byun J, Schin ML, Lorenz M, Kennedy RT, Kretzler M, Feldman EL, Pennathur S, Brosius III FC. 2008. Rosiglitazone reduces renal and plasma markers of oxidative injury and reverses urinary metabolite abnormalities in the amelioration of diabetic nephropathy. *American Journal of Physiology – Renal Physiology* 295:F1071-1081.

8. Wiggin TD, Kretzler M, Pennathur S, Sullivan KA, Brosius FC, Feldman EL. 2008. Rosiglitazone treatment reduces diabetic neuropathy in STZ treated DBA/2J mice. *Endocrinology* 149:4928-4937, 2008.
9. Cohen CD, Lindenmeyer MT, Eichinger F, Hahn A, Seifert M, Moll AG, Schmid H, Kiss E, Gröne E, Gröne HJ, Kretzler M, Werner T, Nelson PJ. 2008. Improved elucidation of biological processes linked to diabetic nephropathy by single probe-based microarray data analysis. *PLoS One* 13:e2937.
10. Greene, D.A., Arezzo, J.C., and Brown, M.B. 1998. Effect of aldose reductase inhibition on nerve conduction and morphometry in diabetic neuropathy. *Annals of Neurology*:1-52.
11. Greene, D.A., Arezzo, J.C., and Brown, M.B. 1999. Effect of aldose reductase inhibition on nerve conduction and morphometry in diabetic neuropathy. Zenarestat Study Group. *Neurology* 53:580-591.
12. Tian Y, Patel JM, Nair V, Martini S, Kretzler M: Periscope/GQ: A Graph Querying Toolkit. Proceedings of the VLDB Endowment, 1, 1404-1407, 2008.