Diabetic Complications Consortium

Application Title: Comparative transcriptomics to identify key gene networks of diabetic neuropathy

Principal Investigator: Junguk Hur, Ph.D.

1. Project Accomplishments:

The overall objectives of this proposal were (1) to examine the gene expression profiles in degenerative and regenerative sural nerve biopsies collected from subjects with DN to identify crucial genes and pathways that may regulate the progression of human DN, and (2) to identify commonly dysregulated critical pathways in both human DN and animal models of DN. The rationale for the proposed research is that the identification of shared critical pathways and gene expression signatures between human and mouse DN will elucidate the core mechanisms of DN progression and expedite the development of mechanism-based therapeutic agents.

With the support of the DiaComp Pilot & Feasibility grant, we processed human sural nerve samples and performed a RNA-sequencing profiling on 78 samples with good quality and sufficient quantity. These 78 samples included 22 Degenerators, 29 Intermediators, and 27 Regenerators, whose grouping was based on the percent changes in myelinated fiber density over 52 weeks (Hur et al. Diabetologia 2013). Our initial Aim was to use only Degenerator and Regenerator samples, but we decided to include Intermediator samples as control as suggested by the reviewers.

Although our analyses on these RNA-Seq data are still to be completed, our preliminary analysis result suggests that there are not much consistent significant differences among these groups, possibly due to high variation within three groups. However, our data-driven analysis result suggests that the gene expression profiles in these sural nerves were largely affected by two clinical factors: HbA1c level and O'Brien score. We are currently focusing on these data-driven analyses.

While we were working on the RNA-Seq data generation and analysis to get the differentially expressed genes (DEGs) to be used in Aim 2, we also improved our analysis pipeline to identify commonly dysregulated genes and pathways across species, specifically between human and murine diabetes, using existing microarray data. Using previously published and unpublished human and various models of murine diabetes microarray datasets, we have identified a large transcriptional network, consisting of 688 genes that are shared between human DN and murine DNs. Our functional enrichment analyses on these networks revealed that genes involved in LXR/RXR activation, different rheumatoid arthritis pathways, and the adipogenesis pathway were highly affected by diabetes and conserved across species.

2. Specific Aims:

Specific Aim 1. ***Identify dysregulated genes and pathways in human DN using RNA-Seq.***

RNA-Seq: We performed an unbiased transcriptomic profiling assays of sural nerve samples obtained from subjects with DN using RNA-Seq. From our sural nerve repository, maintained at Dr. Eva Feldman's laboratory, University of Michigan, we performed a RNA-Sequencing on 78 samples with good quality and sufficient quantity (22 Degenerators, 29 Intermediators, and 27 Regenerators). The sequencing library was prepared using SMARTer Stranded RNA-Seq Kit (Clontech Laboratory, Inc.). The sequencing was performed using Illumina's HiSeq-2500 and approximately 90M paired-end 50bp sequencing reads were obtained per sample. About 96% of these reads were in good quality. FastQC was used to further examine the quality of the reads, and the first 10bps of these reads were trimmed due to relatively low-quality and any adapter sequences (contamination) were removed.

Sequence mapping and identification of differentially expressed genes (DEGs): The processed reads were mapped against the human hg19 genome RefSeq reference sequences to identify genes. The average unique mapping rate was about 77%, except one sample with 33% that was removed from further analyses. We employed DESeq2, a count-based tool for differential analysis, to identify DEGs at an adjusted p-value < 0.05 as the cutoff between three groups (Degenerator, Intermediator, and Regenerator). However, due to the high overlap between the three groups within the transcriptomics space, no DEG was identified using this approach. We tried alternative identification tools such as EdgeR, Cufflink/Cuffdiff and Limma, which also found no significant DEGs with the current classification. We also tried iterative sample grouping: two extreme groups from the highest and lowest %change in myelinated fiber density, and adding one sample to each group at each iteration. This iterative approach identified 41 DEGs with no enriched biological functions or pathways.

Data-driven grouping and DEGs: Due to the high overlap between the group, we decided to use a data-driven approach, coming up new grouping solely based on their global gene expression profiles. Figure 1A illustrates the hierarchical clustering dendrogram of these samples, where we found three groups (Groups1-3), while Figure 1B shows a principal component analysis (PCA) plot.

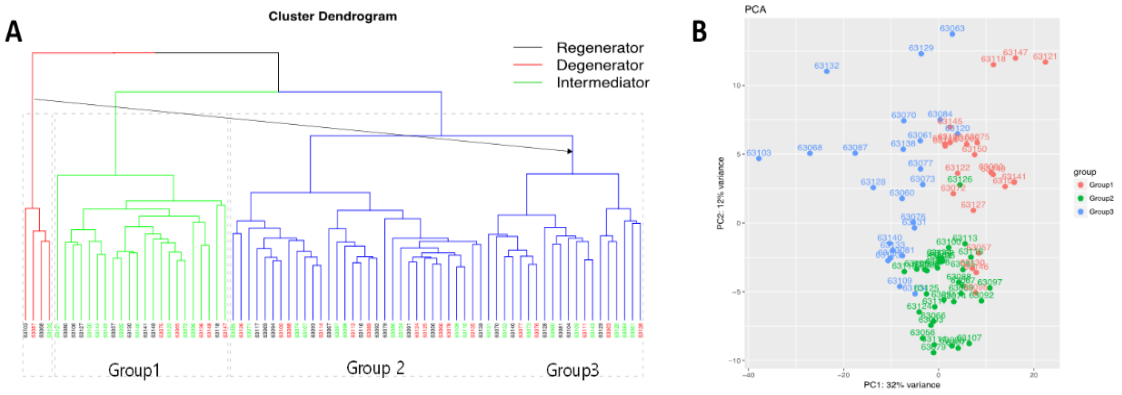

Figure 1. Data-driving grouping

Using multifactorial logistic regression analyses, we found that HbA1C was significantly different across these groups (ANOVA P-value = 0.03; Figure 2A). DESeq2 identified up to 1,400 DEGs between three groups (Figure 2B). When focusing on two groups (Group1 and Group 2, respectively with highest and lowest HbA1C levels), we identified 997 DEGs (adjusted P-value < 0.05). These DEGs included 60 genes with a minimum 2 fold-changes as shown in Figure 3. When we examined enriched biological functions in terms of Gene Ontology terms and pathways, we found that these DEGs were highly enriched with "cellular component/extracellular matrix organization", "phagosome; antigen processing and presentation" pathway, and "adaptive immune system" pathway.

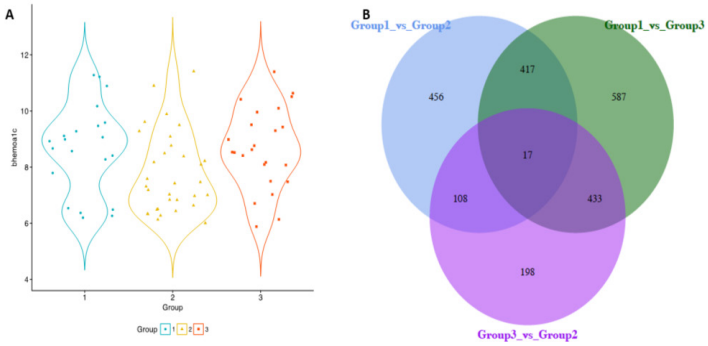

Figure 2. HbA1c levels and DEG Venn diagram.

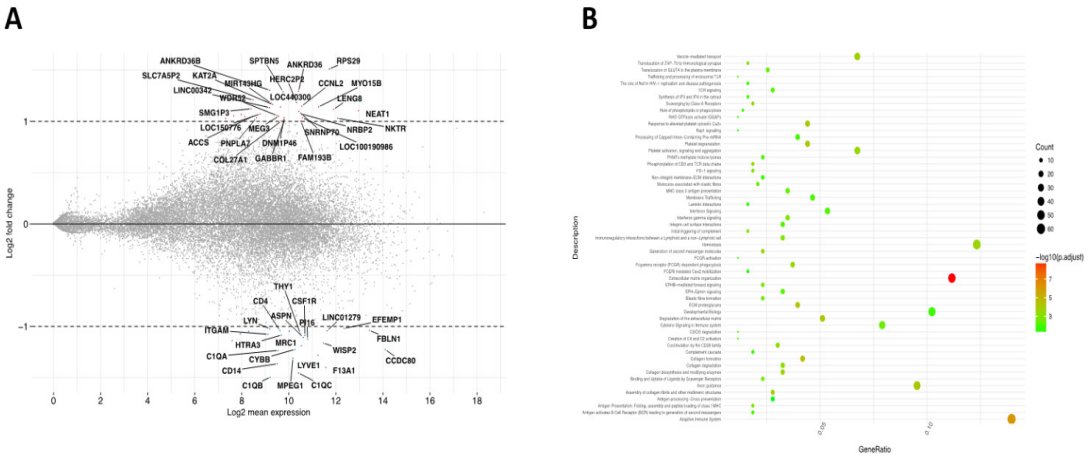

Figure 3. Group 1 vs Group 2 DEGs and enriched Reactome pathways

Conclusion and additional analyses: We performed a deep-sequencing to examine the genome-wide gene expression profiling in the peripheral nerves affected by diabetic neuropathy. The original sample classification didn't result in significantly differentially expressed genes. It should be noted that these samples obtained from human subjects with diabetic neuropathy among these groups; therefore, the differential gene expression signals may not be as strong as we expected between the regenerative and degenerative groups. Interestingly, our data-driven analysis using overall gene expression profiles to create new groups suggests that the glucose level, represented by HbA1c, may be a strong factor affecting the global gene expression in these samples. We have also found that O'Brien neuropathy composite score may be closely associated with the global gene expression profiles, and we are currently examining their correlation.

Data availability: The raw and normalized data are being prepared for deposition into the NCBI's Gene Expression Omnibus (GEO) database and the Sequence Read Archive (SRA) at the time of this report submission. As soon as we completed our further analyses and publish the results, the complete deposited data will be made publicly available.

Specific Aim 2. ***Identify commonly dysregulated DN-associated genes and pathways across species.***

Results: The initial goal was to use DEGs from Aim 1 as human DN signatures and compare them against murine DN-related microarray datasets at both gene and transcriptional network level to identify core-DN-associated genes and pathways shared across species. Due to the delay in the generation of the human DEGs using RNA-Seq in Aim 1, we used the previously published microarray data examining the differences between progressive vs non-progressive human DN as an alternative.

Datasets: We collected eight microarray datasets on peripheral nerve samples collected from type 1 diabetes (streptozotocin-treated) and type 2 diabetes (*db*/*db* and *ob*/*ob*) murine models of various ages and human subjects with non-progressive and progressive DN. DEGs were identified between non-diabetic and diabetic samples in murine models, and between non-progressive and progressive human samples using a unified analysis pipeline (Genomatix ChipInspector with a minimum probe number of 5 to define a transcript). Eight datasets of DEGs were generated based on significant transcripts between control and diabetic groups for each murine model (8, 16, and 24 week *db*/*db*; 5, 13, 26 week *ob*/*ob*; and 34 week STZ-treated DBA2/J) as well as between progressive and non-progressive groups for the human dataset.

DEGs: Over 11,000 DEGs were identified, and at least 2,100 were shared across a minimum of 3 datasets. Figure 4 is a heat-map demonstrating the expression patterns of shared DEGs across models (red: up-regulation, green: down-regulation).

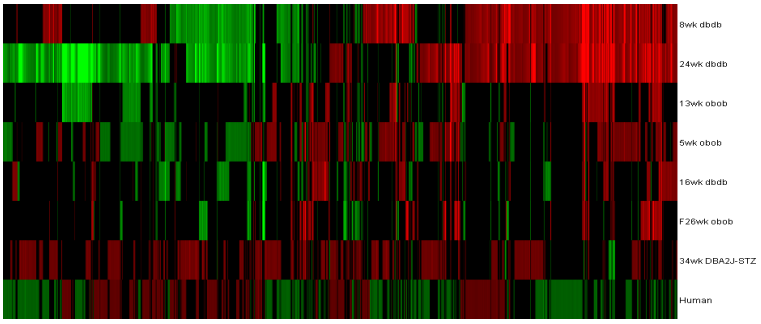

Figure 4. DEP pattern by DN models.

Transcriptional networks: A transcriptional network for each DEG set was constructed based on literature-derived gene-gene interaction information, identified by SciMiner, and shared sub-networks between the human and murine networks were identified by TALE, a network-comparison program. Seven pairwise human-vs-murine comparisons resulted in sub-networks including 46 to 396 genes, which was further merged into a single network of 688 genes (Figure 5) in Cytoscape. Node size is based on the degree of

connections and organized as a radial tree. Functional enrichment analyses using Ingenuity Pathway Analysis (IPA) suite on these shared networks revealed that genes involved in LXR/RXR activation, different rheumatoid arthritis pathways, and the adipogenesis pathways were highly affected by diabetes and conserved across species.

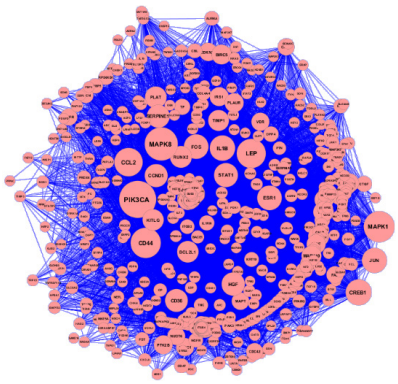

Figure 5. Shared transcriptional network

Gene Symbol	Description	Degree	8wk *db/db*	16wk *db/db*	24wk *db/db*	5wk *ob/ob*	13wk *ob/ob*	F26wk *ob/ob*	34wk DBA2J-STZ	Human
PIK3CA	phosphoinositide-3-kinase, catalytic, alpha polypeptide	304	1.43		1.24				1.28	1.10
MAPK8	mitogen-activated protein kinase 8	261	1.42		-1.35				1.29	-1.01
MAPK1	mitogen-activated protein kinase 1	234			-1.10					1.14
CD44	CD44 molecule (Indian blood group)	228	1.70		2.13	2.18	1.53	1.86	1.40	-1.16
LEP	leptin	216	2.57	2.25	4.14	2.23	2.77		5.86	-1.88
CCL2	chemokine (C-C motif) ligand 2	206			3.01	2.22	1.78		1.33	1.17
CREB1	cAMP responsive element binding protein 1	177	1.60		1.49	-1.21				1.12
IL1B	interleukin 1, beta	172				1.85			1.52	1.21
JUN	jun proto-oncogene	167			-1.37					1.16
ESR1	estrogen receptor 1	164	-1.67	-1.70	-1.77	-1.36		-1.84	1.27	-1.10
FOS	FBJ murine osteosarcoma viral oncogene homolog	163				1.39			1.49	1.39
STAT1	signal transducer and activator of transcription 1, 91kDa	157		-1.36	-1.45	-1.28				-1.12
HGF	hepatocyte growth factor (hepapoietin A; scatter factor)	149	1.81		1.58					-1.13
CCND1	cyclin D1	146	-1.49			-1.27				1.11
SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	144			5.13	1.88	2.01			-1.24
CD36	CD36 molecule (thrombospondin receptor)	143	1.71		2.06		1.90	1.61		-1.49
RUNX2	runt-related transcription factor 2	131	1.74		1.92	1.35			1.25	1.09
PLAT	plasminogen activator, tissue	130	-1.78		-1.59	-1.33				-1.15
BCL2L1	BCL2-like 1	129			1.39	1.35			1.35	1.11
KITLG	KIT ligand	125	1.43			-1.27			1.29	-1.10

Table 1. Highly connected DEGs

Centrality analysis: Centrality analysis identified the most highly connected genes which may play important roles in this cross-species DN shared network. Table 1 shows the fold changes of the most highly connected genes in each dataset with red coloring being an increased fold change and blue being a decreased fold change. A total of 688 genes were included in the network (Figure 5) with the degree, the number of connections between genes, ranging from 304 and 1. Each connection between genes were supported by a minimum of 3 citations as defined by SciMiner.

As shown in Table 1, the top most highly connected genes included PIK3CA, MAPK8, CD44, MAPK1, CREB1, LEP, CCL2, JUN, ESR1, FOS, CD36, IL1B, HGF, and PLAT. These genes indicate enrichment of many different pathways with the most significant including glucocorticoid receptor signaling, multiple cytokine pathways, and HMGB1 signaling.

Conclusion: Our systems biology approach, integrating multiple bioinformatics analyses, identified DN-associated pathways that are highly conserved across various murine models and across species. These conserved pathways are likely the key responses in DN and suggest an important role of dysregulated lipid metabolism in DN pathogenesis.

3. Publications:

We are currently preparing two original research manuscripts for peer-reviewed journal submission from this project. Some parts of our works have been and will be presented at multiple national and local conferences as listed below with the support of DiaComp acknowledged.

(1) Guo K, de-Anda-Jauregui G, McGregor BA, and Hur J. Deep RNA-Sequencing Reveals the Significant Impact of HbA1c and O'Brien Neuropathy Score on the Peripheral Nerves Affected by Diabetic Neuropathy. American Diabetes Association 77th Scientific Sessions, June 9-13, 2017. San Diego, CA (to be presented as a poster)
(2) McGregor BA, Porter J, Feldman EL, and Hur J. Transcriptional Signature of Diabetic Peripheral Neuropathy Shared Between Human and Mouse. The University of North Dakota Graduate Research Day. March 2, 2017. Grand Forks, ND (to be presented as a poster)
(3) McGregor BA, Porter J, Feldman EL, and Hur J. Systems Biology Approach to Identify Conserved Transcriptional Networks between Human and Murine Diabetic Neuropathy. American Diabetes Association 76th Scientific Sessions, June 10-14, 2016. New Orleans (Oral presentation)
(4) McGregor BA, Porter J, Feldman EL, and Hur J. Systems Pharmacology Approach to Identify Potential Therapeutic Small- Molecules for Treatment of Diabetic Peripheral Neuropathy. Experimental Biology (EB) American Society for Pharmacology & Experimental Therapeutics (ASPET) 2016 Meeting. April 2-6, 2016. San Diego. (Oral presentation; 2nd place – best abstract / presentation award)