

RelA as a Potential Regulator of Inflammation and Tissue Damage in
Streptozotocin-Induced Diabetic STAT5 Knockout Mice

A Thesis Presented to the Honors Tutorial College, Ohio University

In Partial Fulfillment of the Requirements for Graduation from the Honors Tutorial
College with the degree of Bachelor of Science in Biological Sciences

By Emilee R. Holtzapple

Contents

I. Abstract.....	3
II. Introduction	4
A. Type 1 Diabetes	4
1. General	4
2. Long Term Effects	5
3. Affected Metabolic Pathways	6
B. Kidney.....	7
C. Diabetic Nephropathy	9
D. The Immune System in DN	11
E. The JAK/STAT Pathway	12
F. NF κ B.....	14
G. Bioinformatics.....	15
H. Diabetic Nephropathy Animal Model.....	16
III. Hypothesis and Specific Aims	18
IV. Materials and Methods.....	18
A. Datasets	18
B. DAVID.....	21
C. STRING	21
D. GO Term Analysis	21
E. Ingenuity Pathway Analysis	22
1. Diseases and Functions	22
2. Causal Network Analysis	22
F. Selection of RelA/p65 for <i>in vivo</i> Binding Activity Measurements	23
1. Target Gene Identification	23
2. Motif Identification	23
3. Motif Scanning.....	23
G. Chromatin Immunoprecipitation.....	24
1. Fixation of Tissue.....	25
2. Disruption of Tissue.....	25
3. Sonication.....	26

4.	Input DNA Preparation	27
5.	Immunoprecipitation	28
6.	Reversal of Cross-links and DNA Purification	30
H.	Binding Site Quantitation	30
1.	Primer Design and Determination of Amplification Efficiency	31
2.	qPCR Analysis	32
V.	Results	34
A.	DAVID.....	34
B.	STRING.....	39
C.	GO Term Analysis	44
D.	IPA	47
1.	Diseases and Functions	48
2.	Causal Network Analysis	51
E.	Selection of RelA/p65 for <i>in vivo</i> binding activity measurements	64
F.	Measurement of <i>in vivo</i> p65/RelA DNA Binding Events.....	68
1.	Selection of Potential RelA-Dependent Gene Targets.....	68
2.	<i>In silico</i> Identification of Potential p65/RelA Binding Sites	69
3.	ChIP-qPCR Results of p65/RelA Binding Event Measurements.....	71
VI.	Discussion	74
VII.	References.....	79

I. Abstract

Type 1 Diabetes (T1D) affects 1.25 million Americans, and that number is expected to increase to 5 million by 2050. Failure to properly control blood glucose levels in T1D can result in life-threatening side effects such as kidney damage, also known as diabetic nephropathy (DN), and end-stage renal disease (ESRD). As the incident rate of T1D continues to rise worldwide, understanding DN becomes more important. This can be accomplished by examining the molecular mechanisms of damage in DN. It has been shown that the loss of STAT5 in diabetic mice exacerbates diabetic kidney damage. In this study, we used pathway analysis software to analyze gene expression results previously obtained from a microarray experiment using this diabetic STAT5 knockout (DB SKO) mouse model. We found that expression of many immune system pathways was significantly altered in the kidneys of DB SKO mice, as compared to nondiabetic and wildtype control mice. A number of different immune cell functions were also predicted to be altered. The RelA gene encoding the p65 subunit of NF κ B was predicted to be a common or “master” regulator of many of the differentially expressed genes within our dataset. Using chromatin immunoprecipitation, we found altered numbers of p65-DNA binding interactions in the promoters of differentially expressed genes within the DB SKO kidney, again in comparison to the nondiabetic and wildtype control kidneys. Therefore, our analyses indicate that STAT5 may act through RelA to affect immune system signaling pathways, resulting in an increase in inflammation and tissue damage in the absence of STAT5.

II. Introduction

A. Type 1 Diabetes

1. General

More than 18,000 children are diagnosed yearly with T1D (Centers for Disease Control and Prevention, 2014). The overall incidence rate of T1D increases annually, and it increases even more for young children (Bjornstad et al., 2010). T1D is a complex disease that involves metabolism and immunity and is thought to be caused by both genetic and environmental factors (Belle et al., 2011).

The effects of T1D arise from the inability of the body to produce adequate amounts of insulin. (Belle et al., 2011) In normal tissue, insulin stimulates glucose absorption in response to high blood glucose levels by recruiting glucose transporters to the surface of muscle and fat cells. The cells responsible for insulin production are the β -cells of the pancreas. In T1D, the body produces antibodies against the β -cells, which leads to infiltration of immune cells, such as macrophages and T cells. These infiltrating cells then secrete cytotoxic factors, causing apoptosis of the β -cells. This disturbs the body's ability to both secrete insulin and regulate blood glucose levels.

Abnormally high blood glucose levels, or hyperglycemia, resulting from insulin deficiency is treated through either subcutaneous insulin injections or insulin delivery through a pump (Bjornstad et al., 2010) The amount of insulin needed daily

can be calculated from the amount of carbohydrates ingested. This requires constant supervision to keep blood glucose levels within an acceptable range. Otherwise, chronic hyperglycemia damages tissues and leads to long-term side effects. (Shaw et al., 2006)

2. Long Term Effects

Currently, the majority of Americans afflicted with T1D struggle with keeping their glucose levels within the proper range, or “glycemic control” (Bjornstad et al., 2010). This may not be imminently life threatening for younger patients but, due to the difficulty of glycemic control, the disease gradually damages organs over time. Retinopathy (eye damage), neuropathy (nerve damage), and nephropathy (kidney damage) are all hyperglycemia-induced chronic complications. The best predictor of mortality in diabetic patients is the presence of DN (Reidy et al., 2014). Being afflicted with T1D is correlated with a life expectancy loss of over ten years. Treatment of T1D is not foolproof or easy, and chronic complications can still occur. The overall incidence and mortality rate of T1D patients with DN makes this an important pathology to research.

3. Affected Metabolic Pathways

In T1D, hyperglycemia is present throughout the body. However, only a few subsets of specialized cells are blatantly damaged (Brownlee, 2005). This is because most cells can control their intracellular glucose levels. When a certain type of cell, such as the proximal tubule epithelial cell of the kidney, cannot alter the rate of glucose transport across the cell membrane, then that cell is adversely affected. Hyperglycemia affects several metabolic pathways within the cells, leading to oxidative stress.

One pathway influenced by high glucose levels is the polyol pathway (Brownlee, 2001). High intracellular glucose levels increase the activity of aldose reductase, a key member of the polyol pathway. Aldose reductase is an NADPH-dependent oxidoreductase that catalyzes carbonyl compound reduction. In normal cells, aldose reductase is responsible for reducing aldehydes to alcohols, and has a low affinity for glucose. In a hyperglycemic cell, aldose reductase also reduces glucose to sorbitol, consuming the cofactor NADPH. NADPH is needed to produce reduced glutathione, which is an antioxidant. High concentrations of intracellular glucose thus lead to a deficit of reduced glutathione and increased susceptibility to oxidative stress.

High levels of intracellular glucose can also lead to increased production of advanced glycation end product (AGE) precursors (Brownlee, 2004). These precursors can react with amino groups of proteins, altering the protein's function. For example, modification of signaling molecules such as transcription factors can alter gene expression. Diffusion of precursors out of the cell can lead to modification of

extracellular matrix molecules, such as collagen and laminin, which in turn can alter matrix-cell interactions. Finally, AGE-modified proteins can bind and activate AGE receptors (RAGEs), causing an increase in the production of cytokines and other pro-inflammatory molecules.

B. Kidney

The kidney serves several important functions, such as removal of waste and excess fluid; regulation of levels of electrolytes such as sodium, potassium, and phosphate; and production of hormones that help with blood pressure regulation, red blood synthesis and bone strength (National Kidney and Urologic Diseases Information Clearinghouse, 2014). While it removes excess liquid and metabolites, the kidney also reabsorbs essential substances from this filtrate, therefore stabilizing the balance of chemical compounds within the body. It filters massive amounts of water and solutes daily. The component of the kidney where this filtration and reabsorption occurs is called the nephron.

The nephron itself is composed of two parts: the glomerulus and the tubule (Figure 1).

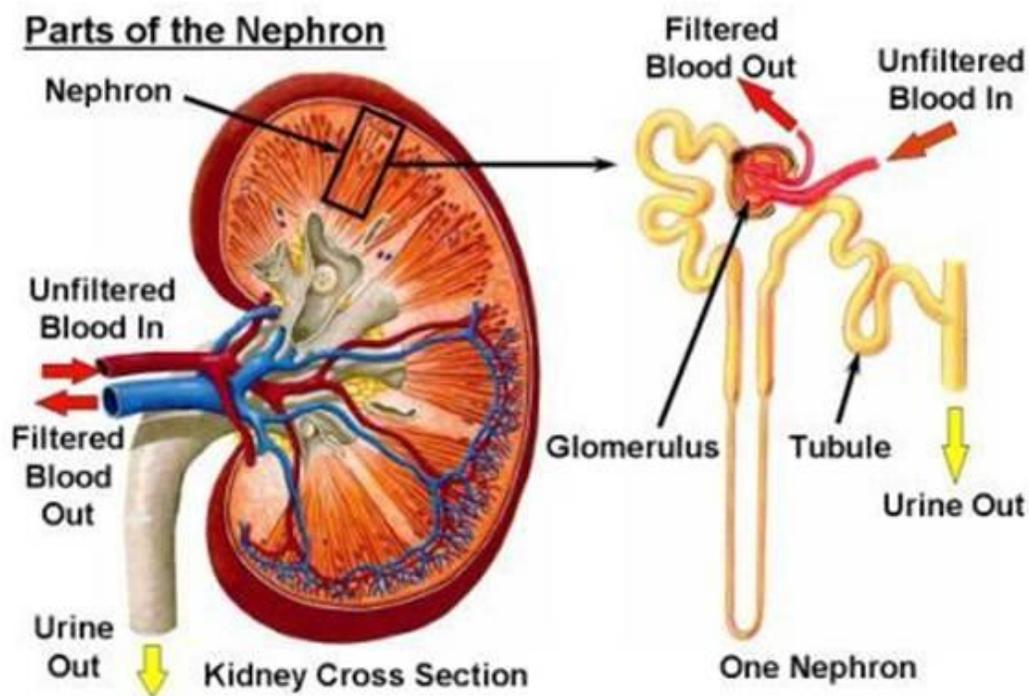


Figure 1: Parts of the Nephron. Obtained from <http://unckidneycenter.org/images/kidney-health-library-pictures/parts-of-the-nephron>

The glomerulus is a capillary tuft within the nephron (Valtin and Schafer, 1995). It is surrounded by a membranous structure called the Bowman's capsule. The glomerulus filters water and waste out of the blood vessels. Glomerular capillaries are especially suited for the task of ultrafiltration because they differ from systemic capillaries in several ways. Glomerular capillaries are more permeable to water and crystalloids in plasma, and they are less permeable to proteins due to the different layers of the glomerular barrier. Blood cells and large proteins are unable to pass through the glomerular barrier, which consists of three layers: the endothelium, basement membrane, and epithelium. Water and crystalloids pass through the glomerular

barrier, between the cells in the fenestrated endothelium, and into the Bowman's space (Satchell and Braet, 2009). This structure surrounds the capillaries and begins the route to urine formation.

From the Bowman's space, the filtrate passes on to the tubule. The tubule reabsorbs solutes from the filtrate that are needed to maintain the proper balance of water and salts between the filtrate and the blood. It spans the length from the glomerulus to the collecting duct system, which connects the nephrons to the ureter. Once the liquid reaches the ureter, it is excreted out of the body as urine. The healthy kidney is capable of filtering out waste and maintaining electrolyte and water balances. (Valtin and Schafer, 1995)

C. Diabetic Nephropathy

Several functional and structural changes are observed in the diabetic kidney (Figure 2). In the early stages of T1D, glomerular filtration rate (GFR) is typically increased. This is caused by increased intraglomerular pressure, which in turn compromises the integrity of the glomerular barrier. (Bjornstadt et al., 2014) The basement membrane of the glomerular capillaries, along with the tubules, starts thickening due to the deposition of excess collagen, fibronectin, and other extracellular matrix components (Reidy et al., 2014). Podocyte damage and loss, which also affect the integrity of the glomerular filtration barrier, can occur as well.

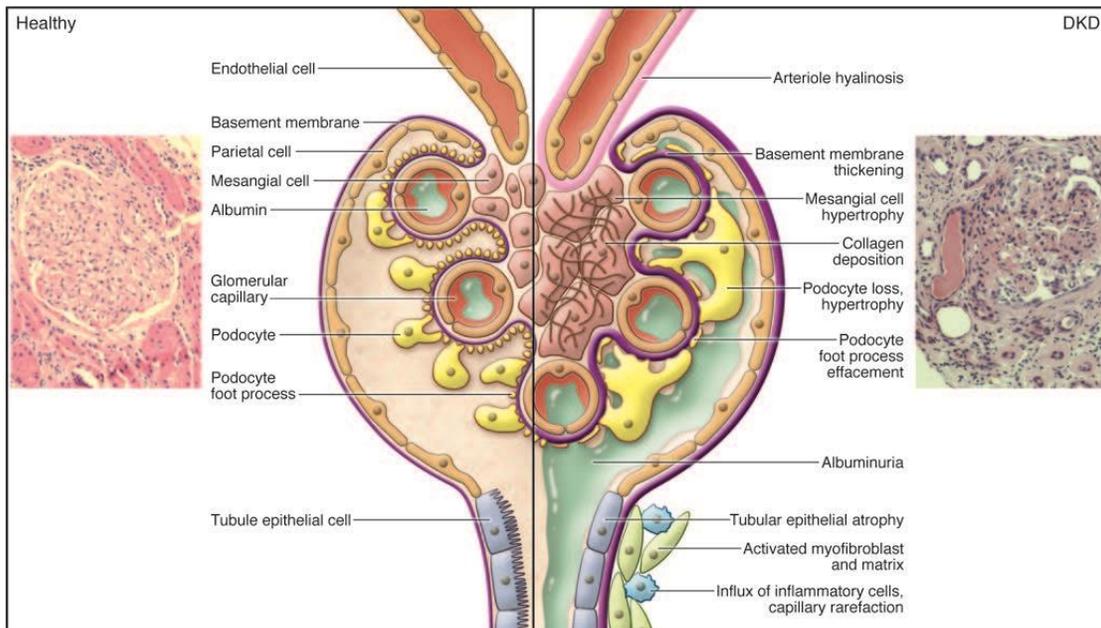


Figure 2: Glomerular diabetic damage (right) as compared to normal structure (left). Republished with permission from the Journal of Clinical Investigation.

Eventually, the changes affect the blood flow, filtration and reabsorption functions of the nephron. As the disease progresses, the GFR decreases, dropping well below the normal GFR. The presence of small proteins in the urine, or albuminuria, is a common symptom of DN. This is a sign that the filtration barrier of the glomerulus has been compromised, and that larger molecules can now pass through into the filtrate. The severity of damage can usually be predicted by glucose levels, hypertension, and the presence of albuminuria (Reidy et al., 2014).

D. The Immune System in DN

The immune system has been implicated in the progression of DN. As just described, due to the cellular damage from hyperglycemia, there are changes in the structure and function of the different cell types within the kidney. The damaged cells can release danger signals (Miyaki, 2007). These danger signals are endogenous products which are released from cells during an inflammatory response or tissue damage. These molecules can act as ligands for Toll-like receptors, and binding of these receptors can stimulate an immune response. Oxidative stress can also increase the production of pro-inflammatory cytokines within damaged kidney cells, causing migration of immune cells (Elmarakby and Sullivan, 2012). In DN, the inflammation caused by damaged cells has been shown to contribute to renal damage (Ohga et al., 2007). Different immune cells, such as dendritic cells, macrophages, and T-cells, have also been shown to be active in DN (Zheng and Zheng, 2015).

Macrophages are involved in the initiation of renal remodeling (Duran-Salgado and Rubio-Guerra, 2014). There are two subtypes, M1 and M2, which carry out separate functions. M1 macrophages are pro-inflammatory, and their secretion of inflammatory cytokines contributes to renal damage. On the other hand, M2 macrophages contribute to renal repair by emitting anti-inflammatory cytokines. Part of the drastic differences in these two macrophage subtypes is their metabolism. M1 macrophages initiate a pathogen-defense mechanism by metabolizing arginine to nitric oxide (NO). NO is cytotoxic, effectively halting cell proliferation and adding to the

kidney damage. M2 initiates healing by producing ornithine from arginine, encouraging cell proliferation. Both subtypes have been implicated in DN pathogenesis.

E. The JAK/STAT Pathway

The JAK/STAT pathway is a diverse and well-studied pathway with a multitude of functions (Lim and Cao, 2006). The pathway is named for the Janus kinase (JAK) and Signal Transducers and Activators of Transcription (STAT) proteins.

The STAT proteins are a family of transcription factors (Lim and Cao, 2006). However, as the name suggests, STATs can act both as a transcription factor and a cytoplasmic signaling molecule. There are seven mammalian STAT proteins: STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, and STAT6. The STAT protein family is activated by a range of different molecules, including interferon, hormones, interleukins, and growth factors. STAT1, STAT3, STAT5A, and STAT5B can all be activated by a wider range of ligands than STAT2, STAT4, and STAT6.

JAK initially stood for “Just another Kinase” (Dawson and Bannister, 2012). When the importance of the protein family was recognized, JAK was renamed “Janus Kinase”, due to the two kinase domains of the enzyme. In mammals, there are four proteins within the JAK family: JAK1, JAK2, JAK3, and TYK2. JAK proteins are essential for cell signaling pathways in a number of different cell types. In addition to

the JAK/STAT pathway, JAK proteins are also involved in the MAPK and PI3K-AKT signaling pathways. The JAK family has also been implicated in chromatin remodeling.

JAK and STAT proteins interact to form pathways that are essential for cell growth and proliferation, immunity, and other important functions (Lim and Cao, 2006). Cell signaling in the JAK/STAT pathways follow a general mechanism. First, a signaling ligand binds with a receptor, causing a conformational change within the receptor. This activates a JAK, which phosphorylates the receptor, creating a binding site for a STAT protein. Upon binding, JAK phosphorylates the STAT with its second kinase domain, which activates STAT and releases it from the receptor. STAT then dimerizes with another phosphorylated STAT and translocates into the nucleus where it regulates gene expression.

STAT5, of special interest in this thesis, refers to two protein isoforms, STAT5A and STAT5B, which are encoded by separate genes (Teglund et al., 1998). Due to their high amino acid similarity and their similar expression throughout cell types, they are somewhat “functionally redundant”. One difference between them is that STAT5A is involved in prolactin signaling while STAT5B is active in growth hormone signaling (Lim and Cao, 2006). STAT5A and STAT5B have also been implicated in immune cell signaling and play crucial roles in the differentiation of hematopoietic stem cells (Nivarthi et al., 2012).

F. NF κ B

The nuclear factor kappa-light-chain-enhancer of activated B cells (NF κ B) protein complex refers to a family of five genes in mammals: RelA/p65, RelB, c-Rel, NF κ B1/p50/p105 and NF κ B2/p52/p100 (Gilmore, 2006; Figure 3). All five NF κ B proteins have a conserved DNA-binding/dimerization domain called the Rel homology domain (RHD). The C-termini of the Rel proteins have a transcriptional activation domain (TAD). NF κ B1 (p50/p105) and NF κ B2 (p52/p100) both have ankyrin repeats in their C-termini, acting as a transcriptional repressor.

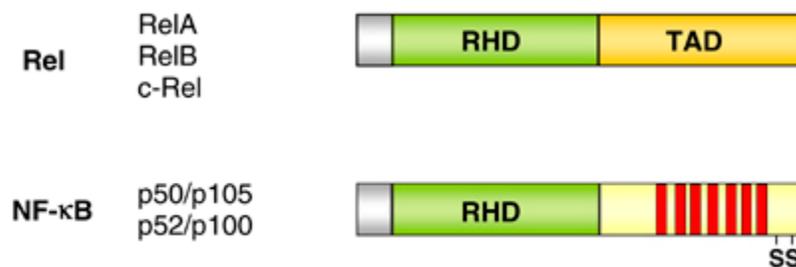


Figure 3: NF κ B family of proteins. The gene names are listed on the left and a diagram of their protein structure is shown on the right. RHD: Rel homology domain; TAD: transcriptional activation domain; ankyrin repeats are indicated as red stripes. Republished with permission from the Nature Publishing Group. (Modified from Gilmore 2006)

The NF κ B family controls gene transcription in response to a ligand binding a cell surface receptor, such as a Toll-like receptor (Gilmore, 2006). NF κ B signaling is usually accomplished through two widely recognized signaling pathways, the canonical and non-canonical NF κ B signaling pathways. In the canonical signaling

pathway, a p50/RelA heterodimer is formed to induce transcription. In the non-canonical signaling pathway, a p52/RelB heterodimer forms and induces transcription. It has also been shown that p50 can form a homodimer to repress transcription.

G. Bioinformatics

Within the 21st century, biologists have seen an exponential increase in the accumulation of genomic data. With the sequencing of the human genome and advances in next-generation sequencing methods, data analytics have become even more important. Bioinformatics is the practice of using modern computational methods to analyze biological data. Common analyses include identification of transcription factor binding sites, phylogenetic tree construction, data and literature mining, and generation of gene expression profiles (Luscombe et al., 2001). The tools for these analyses allow the understanding of complex biological systems.

The result of gene expression analyses such as microarrays, which survey RNA expression of an entire genome under different circumstances, is a large dataset of differentially expressed genes. To uncover the inner workings of the phenotype studied, different bioinformatics tools can be used. Instead of manually researching every molecule with differential expression, these tools can predict between molecules in the dataset. Some of these tools include pathway analysis, protein-protein interaction tools, and Gene Ontology term analysis. Pathway analysis tools like Ingenuity Pathway Analysis and the DAVID functional annotation tool can predict

enriched pathways. STRING can be used to predict interactions within a set of gene products. Gene Ontology Term analysis can identify biological functions, processes, and cell components associated with a set of genes. These tools can be used to predict affected molecules, pathways, and functions within a model.

H. Diabetic Nephropathy Animal Model

To examine the effect of loss of STAT5 gene expression within the diabetic kidney, the SKO mouse model was used. This mouse line was created in the lab of Dr. Jim Ihle by homologous recombination, deleting the exons encoding the N-terminal regions of the STAT5A and STAT5B genes (Teglund et al., 1998). Truncated STAT5 proteins are expressed at low to undetectable levels, resulting in double knockout mice that are smaller than normal and defective in T cell proliferation. To induce diabetes, streptozotocin (STZ) was administered to the mice (Brosius, 2015). STZ destroys the pancreatic β -cells, effectively inducing T1D. Four mouse groups were compared: a non-diabetic wildtype group (ND WT), a diabetic wildtype group (DB WT), a non-diabetic SKO group (ND SKO), and a diabetic SKO group (DB SKO) (Coschigano et al., 2013).

The diabetic SKO mouse group was characterized as having increased tubulointerstitial kidney damage as compared to the diabetic wildtype and nondiabetic SKO groups after 11 weeks of diabetes (Coschigano et al., 2013). This was measured

as an increase in tubular pathology (Figure 4), CD68⁺ macrophage infiltration and fibronectin protein expression. Urinary albumin excretion and mRNA levels of inflammation-related genes were also increased. This research demonstrated that experimental induction of diabetes in mice lacking the transcription factors STAT5A and STAT5B produced greater kidney damage than in mice with intact STAT5 signaling. Taken together, this suggests that STAT5 may play a role in ameliorating damage in the diabetic kidney.

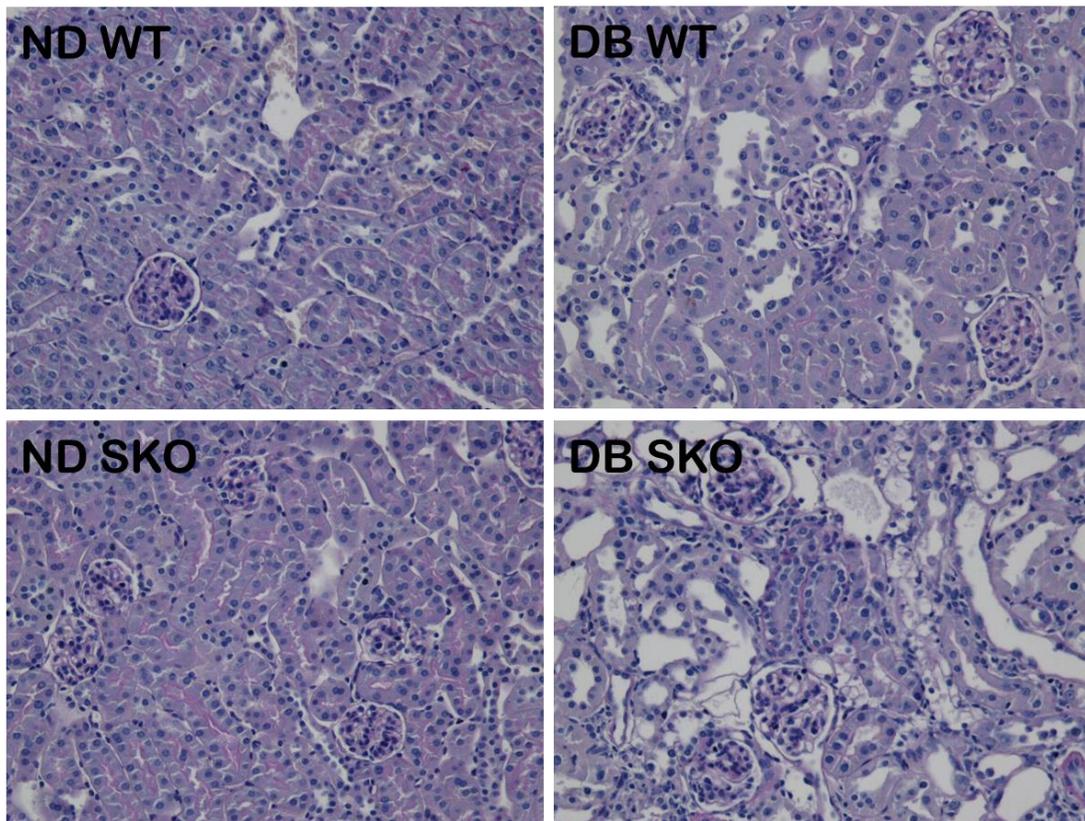


Figure 4: Representative periodic acid Schiff (PAS) staining in kidney sections of the four comparison groups. The DB SKO (bottom right) shows an increase in tubule damage as evidenced by dilation, atrophy, and necrosis. (Coschigano et al., unpublished results)

III. Hypothesis and Specific Aims

In the DB SKO mouse model, increased kidney damage and differential gene expression was observed. The aim of this thesis was to identify how the loss of STAT5 caused differential gene expression in the DS SKO kidney, and to explore how the changes in gene expression increased diabetic kidney damage. By identifying pathways, molecules, and biological functions that are affected by the absence of STAT5 in the diabetic kidney, the role of STAT5 in diabetes can indirectly be studied.

IV. Materials and Methods

A. Datasets

Genome-wide RNA expression was previously assessed in four groups of mice (ND WT, DB WT, ND SKO, and DB SKO) using Agilent microarrays (Coschigano et al., unpublished results). RNA was isolated from the kidneys of three mice per group and each kidney sample was analyzed on a separate microarray (12 microarrays total). The expression level data was then normalized and prepared for further analysis. This included calculating the log fold change, p-value, and other statistical tests. From this normalized microarray data, pair-wise comparisons were made between groups to identify differentially expressed genes (DEGs):

- WT: significantly up- or down-regulated in DB WT as compared to ND WT – i.e. anything altered by diabetes in the WT, but not necessarily exclusive to the WT. (38 up, 7 down)
- KO: significantly up- or down-regulated in DB SKO as compared to ND SKO- i.e. anything altered by diabetes in the SKO, but not necessarily exclusive to the SKO. (955 up , 365 down)
- ND: significantly up- or down-regulated in ND SKO as compared to ND WT- i.e. anything changed in the non-diabetic KO as compared to the non-diabetic WT, but not necessarily exclusive to the non-diabetic status (137 up, 32 down)
- DB: significantly up- or down-regulated in DB SKO as compared to DB WT- i.e. anything changed in the diabetic KO as compared to the diabetic WT, but not necessarily exclusive to the diabetic status. (634 up, 129 down)

Each of the four comparison lists was checked for duplicates. If duplicates were found, the gene with the highest level of expression was kept, and any other duplicates were removed. Each of these four unique lists was then prepared for further analysis. While these four comparisons revealed differential expression profiles for a specific genotype or disease, the model of interest was the SKO with diabetes. Thus, more specific lists were made by subtracting one list from another as follows:

- KO-WT: Significantly up- or down-regulated genes present in the KO comparison that were not present in the WT comparison- i.e. anything

altered by diabetes in the SKO that is unique to the SKO genotype.

(929 up, 361 down)

- DB-ND: Significantly up- or down-regulated genes present in the DB comparison that were not present in the ND comparison- i.e. anything altered in the diabetic SKO that is unique to the SKO genotype. (532 up, 121 down)

There was a possibility that a gene could be upregulated in one comparison, but be downregulated in another. If two groups had opposite expression profiles for a common gene, then that gene was not considered to be common to both groups. To prevent confusing these instances as just differential expression and removing them from the subtraction, the up- and down-regulated genes were compared separately and then combined.

To examine the overlap between the two subtractions, a final list was created as an intersection of the two subtractions:

- Intersection ($KO-WT \cap DB-ND$): This list contains all genes that are present in both the KO-WT and DB-ND lists- i.e. anything that is altered in the SKO by diabetes. (364 up, 74 down)

These three lists, KO-WT, DB-ND, and Intersection, all portray the combined effect of the loss of STAT5 and diabetes. These three lists were used for analysis using a host of bioinformatics tools.

B. DAVID

All three datasets were analyzed using DAVID 6.7 (<https://david.ncifcrf.gov/>). The Functional Annotation tool was used. Each list of DEGs was uploaded separately using the OFFICIAL_GENE_SYMBOL identifier. The “gene list” parameter was chosen. The annotations were limited to *Mus musculus*. The pathways from KEGG were then exported.

C. STRING

All three datasets were analyzed using STRING 9.1 (<http://string-db.org/>). Each list of DEGs was searched using the “Multiple Proteins” function. The organism was set to *Mus musculus*. The resulting graph was then set to evidence-mode. The parameters were changed to high confidence (0.9). All disconnected nodes were then removed. All resulting graphs from the three datasets were exported and examined.

D. GO Term Analysis

Gene Ontology (GO) term analysis was performed on the three datasets using the GOTermFinder tool (<http://go.princeton.edu/cgi-bin/GOTermFinder>). Each list of DEGs was uploaded, and analyzed using the Process tool. The organism was restricted to *Mus musculus*.

E. Ingenuity Pathway Analysis

All three datasets were uploaded into Ingenuity Pathway Analysis, Summer Release 2015, and were analyzed through the use of QIAGEN's Ingenuity Pathway Analysis (IPA® , QIAGEN Redwood City, www.qiagen.com/ingenuity). The lists of DEGs, along with the normalized expression data (expressed as log fold change base 2) were analyzed. The Advanced Analytics function was enabled.

1. Diseases and Functions

All three datasets were visualized using the Diseases and Functions tool in IPA, which produces a mosaic map. The associated categories were sorted from high to low significance, as determined by the p-value. The categories were colored by z-score.

2. Causal Network Analysis

All three datasets were examined using the Causal Network Analysis (CNA) feature of IPA. Only predicted upstream regulators that were capable of interacting downstream of STAT5 were included. The molecule type was restricted to transcription factors. The predicted master regulators were sorted from high to low significance (expressed as a corrected-value).

F. Selection of RelA/p65 for *in vivo* Binding Activity Measurements

1. Target Gene Identification

To identify target genes for ChIP-qPCR, a list of DEGS that are known p65/relA targets, as curated in the IPA software package, was generated using IPA. The potential target genes were limited to DEGs that were bound directly by p65, as predicted by IPA. All direct interactions (phosphorylation, ubiquitination, protein-DNA binding, etc.) were based on current literature where a direct relationship had been experimentally confirmed. The genes were also restricted to those that shared a protein-DNA binding relationship with p65, as predicted by IPA.

2. Motif Identification

The JASPAR (<http://jaspar.genereg.net/>) and MotifMap (<http://motifmap.ics.uci.edu/>) databases were searched for a mouse p65 binding motif. The resulting motif logo was exported and represented as a consensus sequence.

3. Motif Scanning

To identify potential RelA binding sites within the promoters of RelA-dependent genes, the nucleotide sequence 1kb upstream of the transcription start site of the IL1 β , TLR2, CXCL12, BIRC5, and ICAM1 genes was first extracted using the Ensembl (<http://useast.ensembl.org/index.html>) BioMart tool. A 1 kb region is the standard length used for these searches. FIMO (Find Individual Motif Occurrences) within the MEME suite (<http://meme-suite.org/>) was then used, importing the identified p65 binding consensus sequence.

G. Chromatin Immunoprecipitation

Chromatin Immunoprecipitation (ChIP) was performed using Active Motif's ChIP-IT High Sensitivity Kit essentially as recommended by the manufacturer as described in detail below (Carlsbad, California). This kit was chosen for its capability to detect binding at low levels, such as for transcription factors.

A single kidney from each of the four groups (ND WT, DB WT, ND SKO, DB SKO) was assayed for RelA/p65 chromatin binding activity (see Table 1). The tissues were weighed upon collection, frozen in liquid nitrogen and then stored at -80°C prior to ChIP. Since multiple tissue samples needed to be prepared, each sample was individually processed through the first centrifugation and resuspension in PBS Wash Buffer described in Step 2 before proceeding with the next sample.

Table 1: Samples used for ChIP.

Genotype	Animal Identification	Tissue Weight (g)
ND WT	2071	0.1051
DB WT	2118	0.1275
ND SKO	2072	0.1124
DB SKO	2100	0.1168

1. Fixation of Tissue

The frozen sample was placed in a petri dish on ice. A 10 ml aliquot of Complete Tissue Fixation Solution (CTFS) was measured out, and then approximately 1-2 ml was added to the sample. Tissue was diced into a slurry using a razor blade, and the remainder of the CTFS was added. The solution containing the tissue was then transferred to a 15 ml conical tube and rotated at room temperature for 10 minutes. A 515 μ l aliquot of Stop Solution was added after the incubation and the samples rotated at room temperature for an additional 5 minutes.

2. Disruption of Tissue

To disrupt the tissue, the sample was passed through a 3 ml syringe without a needle 10 times and then through an 18 gauge needle 20 times, collecting the sample in a

conical tube. The sample was then centrifuged at 4°C at 1,250 x g for 3 minutes. The supernatant was poured off and discarded. The pellet was resuspended in 10 ml of PBS Wash Buffer. After this point, the samples were kept on ice for the remainder of the protocol.

The above procedure was performed individually for all samples. The samples were then centrifuged at 4°C at 1,250 x g for 3 minutes. The supernatants were poured off, and then each pellet was resuspended a second time in 10 ml of PBS Wash Buffer. The samples were then centrifuged at 4°C at 1,250 x g for 3 minutes and the supernatant discarded. Each pellet was resuspended in 5 ml of Chromatin Prep Buffer supplemented with 5 µl Protease Inhibitor Cocktail (PIC) and 5 µl 100 mM PMSF. Samples were incubated on ice for 10 minutes. Each sample was then transferred to chilled 7 ml dounce homogenizer on ice. Using the tight “A” pestle, the sample was homogenized for 30 strokes. The samples were transferred to a new 15 ml conical tube, then centrifuged at 4°C at 1,250 x g for 3 minutes. The supernatant was then discarded. The pellets were then resuspended in 500 µl ChIP Buffer supplemented with 5 µl PIC and 5 µl 100 mM PMSF. The samples were transferred to new 1.5 ml microcentrifuge tubes and incubated on ice for 10 minutes.

3. Sonication

For this procedure, a Sonic Dismembrator probe sonicator (FB-120; Fisher Scientific, Waltham, Massachusetts) was used. Prior to sonication, a 25 µl aliquot was

transferred to a PCR tube for each sample for input DNA preparation and stored on ice. Samples were sonicated at 30% amplitude in cycles of 30 seconds on, 30 seconds off on ice for 16 minutes (32 minutes total time). After sonication, another 25 μ l aliquot was transferred to a new PCR tube for input DNA preparation. The remainder was split between three tubes and stored at -80°C for immunoprecipitation.

4. Input DNA Preparation

For Input DNA preparation, 175 μ l of TE, pH 8.0 and 2 μ l RNase A were added to the reserved 25 μ l aliquots. The samples were then heated in a thermocycler for one hour at 37°C . Then 5 μ l of Proteinase K was added to the samples and incubated at 37°C for three hours. A 10 μ l aliquot of 5M NaCl was then added to each sample and incubated for 16 hours at 65°C . The samples were then transferred to 1.5 ml microcentrifuge tubes and 250 μ l of phenol, pH 8.0 and 125 μ l of chloroform:isoamyl alcohol (24:1) added to the samples. The samples were vortexed for approximately 10 seconds, and then centrifuged for two minutes at maximum speed ($\sim 14,000 \times g$). The upper aqueous layer was transferred to a new 1.5 ml microcentrifuge tube. A 250 μ l aliquot of chloroform:isoamyl alcohol (24:1) was then added to the each sample. After vortexing for approximately 10 seconds, the samples were centrifuged for two minutes at maximum speed. Each upper aqueous layer was then moved to a new 1.5 ml microcentrifuge tube. A 1 μ l aliquot glycogen carrier (20 mg/ml) and 900 μ l absolute ethanol were then added to each sample. The tubes were vortexed and chilled at -80°C

for several hours. Samples were then centrifuged at maximum speed for 15 minutes at 4°C. The supernatant was poured off and discarded. A 500 µl aliquot of 70% ethanol was then added to the pellets, the samples vortexed to dislodge the pellets and then centrifuged at maximum speed for 5 minutes at 4°C. The supernatant was then removed and the pellets left to dry at room temperature for 15 minutes. When the pellets had dried, 25 µl of DNA Purification Elution Buffer was added. The samples were then vortexed and spun down three times. This is the input DNA. The concentration of each sample was measured on a NanoDrop. To assess the size of the sheared DNA, 1.5 µg of DNA was electrophoresed on a 1.5% agarose, 1X Tris-Acetate-EDTA gel and visualized by ethidium bromide staining.

5. Immunoprecipitation

Sonicated chromatin was thawed and centrifuged at 15000 rpm for 2 minutes at 4°C. 30 µg of chromatin was used for each ChIP reaction, based on the measured concentration of the corresponding sheared input DNA. Antibodies used were either the NFκB p65 antibody (sc-8008 X, (F-6) X, mouse monoclonal antibody, Santa Cruz Biotechnology, Dallas, Texas), a negative IgG antibody or a positive Pol II antibody (ChIP-IT Mouse qPCR Analysis kit, Active Motif, Carlsbad, California). A bridging antibody was also used (ChIP-IT Mouse qPCR Analysis kit, Active Motif). The appropriate antibody was incubated with proprietary Blocker reagent from the kit for 1

minute at room temperature before adding to the chromatin, ChIP buffer and PIC (see Table 2). The tubes were then capped and rotated at 4°C overnight.

To pull down the antibody-bound DNA, Protein G agarose beads were used. The Protein G agarose beads were washed twice with TE pH 8.0. A 30 µl aliquot of beads was then added to each ChIP reaction and the samples rotated at 4°C for three hours. A 600 µl aliquot of ChIP Buffer was added to each tube and the ChIP reactions transferred to separate proprietary ChIP Filtration Columns, allowing the liquid to flow through by gravity while retaining the ChIP-agarose bead sample on the column. The samples were then washed with 900 µl of Wash Buffer five times, allowing the buffer to flow through the column by gravity. The columns were then spun for 3 minutes at 1250 x g at room temperature to remove any residual Wash Buffer. The samples were then eluted twice with 50 µl 37°C Elution Buffer and centrifuged in a room temperature microcentrifuge at 1250 x g for 3 minutes. The flow-through contains the ChIP DNA.

Table 2: ChIP reaction composition.

Sample	Chromatin (µl)	ChIP Buffer (µl)	PIC (µl)	p65 (µl)	Bridging Antibody (µl)	IgG (µl)	Pol II (µl)	Blocker (µl)
ND-WT	58.0	142.0	5	2	4	-	-	5
ND-SKO	57.3	142.7	5	2	4	-	-	5
DB-ND	62.1	137.9	5	2	4	-	-	5
DB-SKO	76.4	123.6	5	2	4	-	-	5
ND-WT	58.0	142.0	5	-	-	20	-	5
ND-WT	58.0	142.0	5	-	-	-	20	5

6. Reversal of Cross-links and DNA Purification

Each ChIP reaction was then transferred to a 250 μ l PCR tube. A 2 μ l aliquot of Proteinase K was added to each sample, the tubes vortexed, and heated at 55°C for 30 minutes, then 80°C for 2 hours in a thermocycler. Samples were transferred to new 1.5 ml microcentrifuge tubes and 500 μ l of DNA Purification Binding Buffer added to each tube. The pH was then adjusted with 5 μ l 3 M Sodium Acetate. After achieving the correct pH (samples were bright yellow), the samples were transferred to individual DNA Purification Columns inside collection tubes. The samples were centrifuged for 1 minute at 14,000 rpm and the flow-through discarded. A 750 μ l aliquot of DNA Purification Wash Buffer was then added to each tube, followed again by centrifugation at 14,000 rpm for 1 minute. After the flow through was discarded, samples were spun again for two minutes at 14,000 rpm to remove residual buffer. The DNA purification column was then transferred to a new 1.5 ml microcentrifuge tube and DNA eluted twice with 100 μ l DNA Purification Buffer and centrifuged at 1250 x g for 3 minutes. The flow-through contained the purified ChIP DNA.

H. Binding Site Quantitation

To quantify p65 binding to the predicted target genes, quantitative PCR (qPCR) was performed.

1. Primer Design and Determination of Amplification Efficiency

Primer pairs for qPCR were designed using Primer-BLAST

(<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) to amplify a region within 1 kb of the predicted p65 binding site. Primer pairs were chosen using default settings for amplicon length, GC content, and T_m. Primer pairs used (purchased from Integrated DNA Technologies, Coralville, Iowa) were the following:

Table 3: Primer pair characteristics.

Primer Pair	GC%	T _m	Product length
mgICAM1+ 5'-GCT TGG ATC GCT GCT TCA TC-3'	55.0	59.69	105
mgICAM1- 5'-AGG GCT CCC TGG AAT CGT TA-3'	55.0	60.33	
mgTLR2+ 5'-ATA AGG GCA GGG GGA CAA AG-3'	55.0	59.37	171
mgTLR2- 5'-TGC CAA GAC CCA GGT CCT AA-3'	55.0	60.47	
mgBIRC5+ 5'-CTT CGG GAC GTG ACA AAA CC-3'	55.0	59.42	164
mgBIRC5- 5'-TGC AAA CGA GTG CCC TGA TA-3'	50.0	59.68	
mgCXCL2+ 5'-GGA CAT CCC AGG GTC CCA TA-3'	60.0	60.40	118
mgCXCL2- 5'-CCC TCA TCA GGA AGC ACA GA-3'	55.0	59.09	
mgIL1B+ 5'-AGA GCA CAG AAG CAC CAT CC-3'	55.0	60.04	130
mgIL1B- 5'-AGG ATG TGC GGA ACA AAG GT-3'	50.0	59.89	

Amplification efficiency and specificity of the primers was tested. Each primer was diluted to 2.5 μ M with Tris pH 8.0. Five-fold serial dilutions of input DNA (starting at 518 ng/ μ l) were used to test primer efficiency. A 10 μ l aliquot of 2X iTaq Universal SYBR Green Supermix (Bio-Rad, Hercules, California), 5 μ l 2.5 μ M primer pair mix, and 5 μ l of the appropriate DNA concentration were added to each well of a 96 well

PCR plate. Duplicate reactions were performed. Plates were sealed with optical tape and centrifuged at 3750 x g for 3 minutes. The PCR plate was then placed in a MyiQ Real-Time PCR machine (Bio-Rad) and cycled using the following program programmed in iQ5 software package (Bio-Rad): 95°C for 3 minutes; 50 cycles of (95°C for 15 seconds, 55°C for 30 seconds, 72°C for 30 seconds with fluorescence measured during this step); 95°C for 1 minute; 55°C for 1 minute; 81 steps of 0.5°C increase every 10 seconds with fluorescence measured at each increase.

2. qPCR Analysis

For qPCR analysis, p65 association with seven binding regions was measured using the ChIP-IT qPCR Analysis kit (Active Motif). A primer pair for each of the five predicted p65 binding regions (IL1B, CXCL2, ICAM1, BIRC5, TLR2) was used (see above). A positive primer pair (Catalog No. 71015; Actb1) and a negative primer pair (Catalog No. 71012; Negative-2 qPCR primer pair; gene desert) supplied with the ChIP-IT Control qPCR kit for mouse (Active Motif) were also used. Each DNA Standard reaction contained 5 µl of the appropriate DNA standard (AM1, AM2, Am3), 5 µl of the provided 2 µM Standard Curve Primer Pair and 10 µl 2X iTaq Universal SYBR Green Supermix (Bio-Rad, Hercules, California). ChIP DNA and Input DNA reactions contained 10 µl 2X iTaq Universal SYBR Green Supermix (Bio-Rad), 5 µl primer pair mix (2.5 µM of each primer), and 5 µl DNA sample (ChIP DNA or pooled input DNA adjusted to 2.5 ng/µl). Each of the qPCR reactions was performed in

triplicate using the plate setups below (Tables 7 and 8). Plates were sealed with optical tape and centrifuged at 3750 x g for 3 minutes. The PCR plate was then placed in a MyiQ Real-Time PCR machine (Bio-Rad) and cycled using the following program programmed in iQ5 software package (Bio-Rad): 95°C for 3 minutes; 50 cycles of (95°C for 15 seconds, 55°C for 30 seconds, 72°C for 30 seconds with fluorescence measured during this step); 95°C for 1 minute; 55°C for 1 minute; 81 steps of 0.5°C increase every 10 seconds with fluorescence measured at each increase.

Table 7: qPCR plate template 1.

	Std AM1	Std AM1	Std AM1	Std AM2	Std AM2	Std AM2	Std AM3	Std AM3	Std AM3		
ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1
ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2
ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3
ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4
ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5
ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6
Input	Input	Input	Input	Input	Input	Input	Input	Input	Input	Input	Input
Pos.	Pos.	Pos.	Neg.	Neg.	Neg.	IL1B	IL1B	IL1B	CXCL2	CXCL2	CXCL2

Table 8: qPCR plate template 2.

	Std AM1	Std AM1	Std AM1	Std AM2	Std AM2	Std AM2	Std AM3	Std AM3
ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1	ChIP1
ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2	ChIP2
ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3	ChIP3
ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4	ChIP4
ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5	ChIP5
ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6	ChIP6
Input	Input	Input	Input	Input	Input	Input	Input	Input
ICAM1	ICAM1	ICAM1	BIRC5	BIRC5	BIRC5	TLR2	TLR2	TLR2

The PCR plate was then placed in a MyiQ Real-Time PCR machine (Bio-Rad) and cycled using the following program programmed in iQ5 software package (Bio-Rad): 95°C for 3 minutes; 50 cycles of (95°C for 15 seconds, 55°C for 30 seconds, 72°C for 30 seconds with fluorescence measured during this step); 95°C for 1 minute; 55°C for 1 minute; 81 steps of 0.5°C increase every 10 seconds with fluorescence measured at each increase.

V. Results

A. DAVID

The Database for Annotation, Visualization and Integrated Discovery (DAVID) is a free collection of bioinformatics tools provided by the National Institutes of Health

(NIH). The DAVID knowledgebase is based on a graph theory evidence-based method that uses data from a collection of different public bioinformatics resources such as KEGG, NCBI, and OMIM (Huang et al., 2009). The tools provided by DAVID can be used to perform analyses of gene enrichment, pathway mapping, and other functions that are beneficial for pathway analysis.

DAVID generated a list of overrepresented and thus predicted affected pathways for the three datasets: KO-WT, DB-ND, and Intersection ($KO-WT \cap DB-ND$) (Tables 1-3). These provided insight into biological functions or pathways possibly affected in the DB SKO mouse kidney. For example, the chemokine signaling pathway, which contains genes that are crucial for immune cell chemoattraction, was overrepresented in all three datasets. The differential expression of this pathway indicates that immune cell migration may be affected in the DB SKO kidney. In addition, DAVID predicted enrichment of a number of KEGG pathways that are essential immune cell pathways. Pathways such as the NOD-like receptor pathway, Toll-like receptor signaling pathway, and the cytokine/chemokine signaling pathways are known to be vital for immune cell function. DAVID also predicted enrichment of the B cell receptor signaling pathway, the natural killer cell mediated cytotoxicity, and other pathways specific to individual hematopoietic cells. Even the hematopoietic cell lineage pathway was affected (Figure 15). This indicates that it is not one specific cell type or pathway that has been affected, but it is a combination of cell types and pathways that are affected in the DB SKO kidney.

Table 6: Significantly enriched pathways in the DB-ND dataset.

Affected Pathways*

Chemokine signaling pathway	Primary Immunodeficiency
Cytokine-cytokine receptor interaction	Fc epsilon RI signaling pathway
Natural killer cell mediated cytotoxicity	JAK-STAT signaling pathway
Fc gamma R-mediated phagocytosis	T cell receptor signaling pathway
B cell receptor signaling pathway	Cytosolic DNA-sensing pathway
Hematopoietic cell lineage	Complement and coagulation cascades
Leukocyte transendothelial migration	NOD-like receptor signaling pathway
Toll-like receptor signaling pathway	Cell adhesion molecules (CAMs)

* The database used for pathway enrichment was KEGG. (p<0.05)

Table 7: Enriched pathways in the Intersection dataset.

Affected Pathways*

Chemokine signaling pathway	Toll-like receptor signaling pathway
Complement and coagulation cascades	Jak-STAT signaling pathway
ECM-receptor interaction	Cell cycle
Cytokine-cytokine receptor interaction	Ascorbate and aldarate metabolism
Fc gamma R-mediated phagocytosis	Fc epsilon RI signaling pathway
Hematopoietic cell lineage	NOD-like receptor signaling pathway
Cell adhesion molecules (CAMs)	Regulation of actin cytoskeleton
Leukocyte transendothelial migration	Small cell lung cancer
B cell receptor signaling pathway	Cytosolic DNA-sensing pathway
p53 signaling pathway	T cell receptor signaling pathway
Focal adhesion	Starch and sucrose metabolism
Natural killer cell mediated cytotoxicity	Adipocytokine signaling pathway
Tryptophan metabolism	Retinol metabolism
Primary immunodeficiency	PPAR signaling pathway
Prion diseases	

* The database used for pathway enrichment was KEGG. (p<0.05).

Table 8: Significantly enriched pathways in the KO-WT dataset.

Affected Pathways*

Cytokine-cytokine receptor interaction	Prion diseases
Chemokine signaling pathway	Cytosolic DNA-sensing pathway
Natural killer cell mediated cytotoxicity	Systemic lupus erythematosus
Fc gamma R-mediated phagocytosis	Jak-STAT signaling pathway
B cell receptor signaling pathway	Type II diabetes mellitus
Hematopoietic cell lineage	Tryptophan metabolism
Leukocyte transendothelial migration	Regulation of actin cytoskeleton
Toll-like receptor signaling pathway	Drug metabolism
Primary immunodeficiency	Phosphatidylinositol signaling system
Fc epsilon RI signaling pathway	Intestinal immune network for IgA production
T cell receptor signaling pathway	Starch and sucrose metabolism
Cell adhesion molecules (CAMs)	Small cell lung cancer
NOD-like receptor signaling pathway	Insulin signaling pathway

* The database used for pathway enrichment was KEGG. ($p < 0.05$).

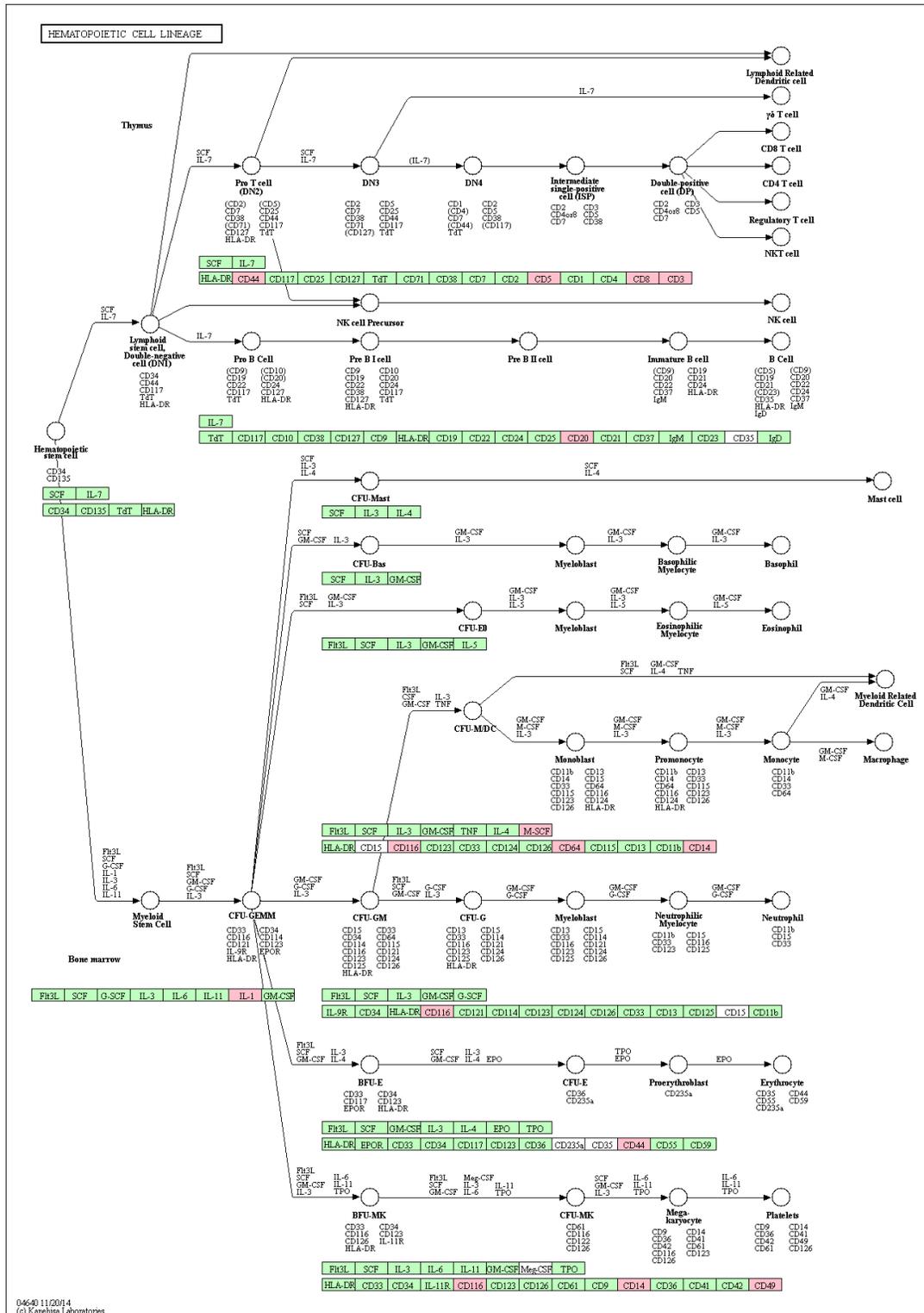


Figure 5: Hematopoietic cell lineage pathway from KEGG. Green objects are gene products present in the KEGG database. Red objects are DEGs present within the Intersection dataset.

B. STRING

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is a tool for finding and predicting protein-protein interactions (Jensen et al., 2009). STRING takes a list of genes (or gene products) and generates a visual of all proven and predicted interactions. STRING was used to visualize which gene products interact within the three datasets. From there, molecules and hubs of interest were established.

In Figures 3-5, the interacting gene products within the datasets were visualized. Every node (sphere) is a gene from the dataset, and every connection (edge) is an established or predicted interaction. For different types of interactions, a different color connection is used, and two gene products can share more than one kind of interaction. Pink and light blue represent known interactions from databases and experimental data, respectively. A green connection indicates a gene neighborhood, or genes that have adjacent genomic locations which are conserved. Deep blue represents co-occurrence, or conserved co-expression among species. Red indicates a gene fusion event that has been conserved. Yellow represents a textmining association from journal abstracts. Black represents co-expression within gene expression data. Grey represents homologous protein domains. The evidence view in string encompasses a wealth of meaningful interactions.

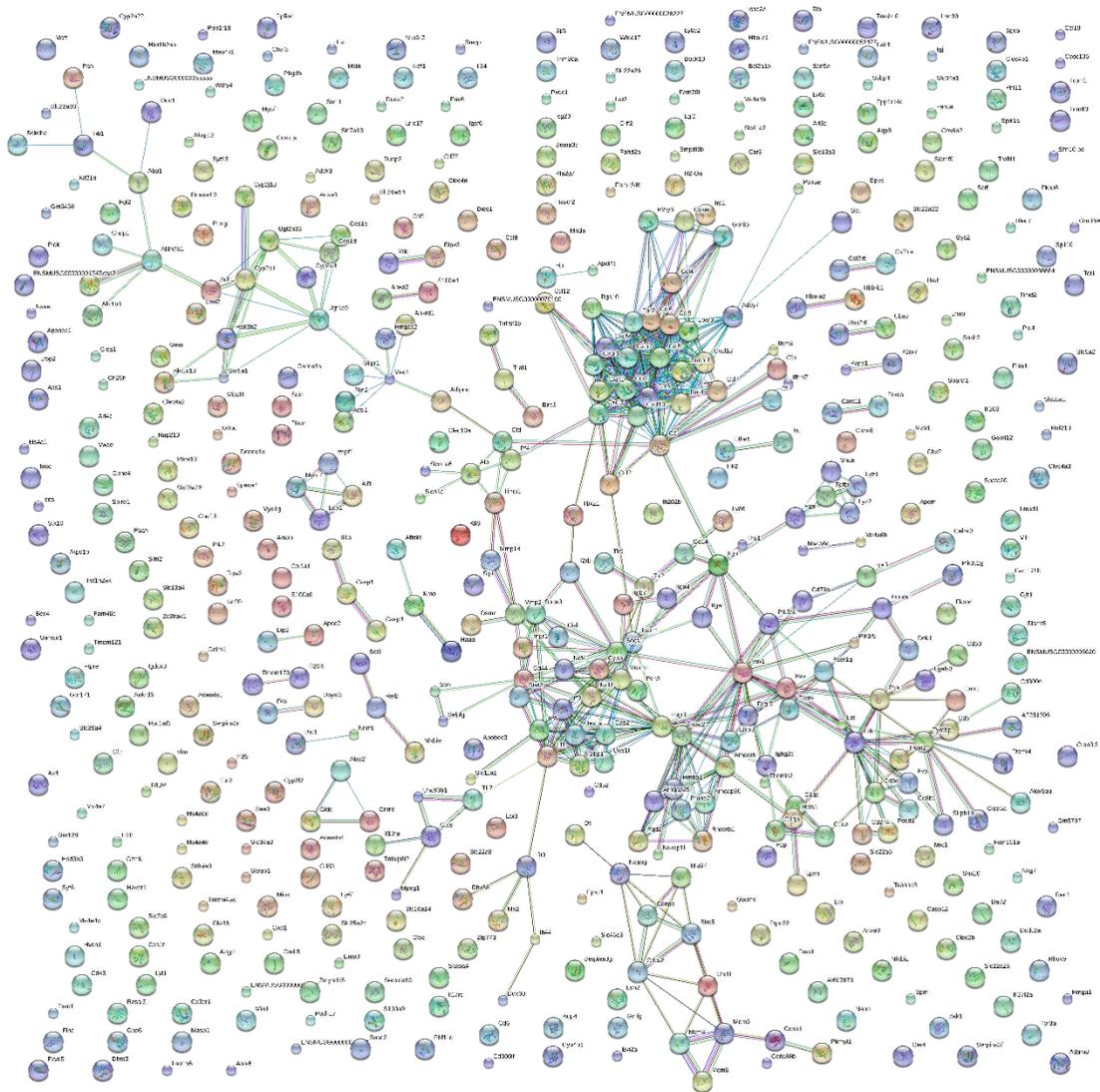


Figure 6: Protein-protein interactions visualized in STRING for the DB-ND list. Each edge represents a different association. Pink and light blue represent known interactions from databases and experimental data, respectively. A green connection indicates a gene neighborhood, or genes that have adjacent genomic locations which are conserved. Deep blue represents co-occurrence, or conserved co-expression among species. Red indicates a gene fusion event that has been conserved. Yellow represents a textmining association from journal abstracts. Black represents co-expression within gene expression data. Grey represents homologous protein domains. The colors of the nodes are irrelevant.

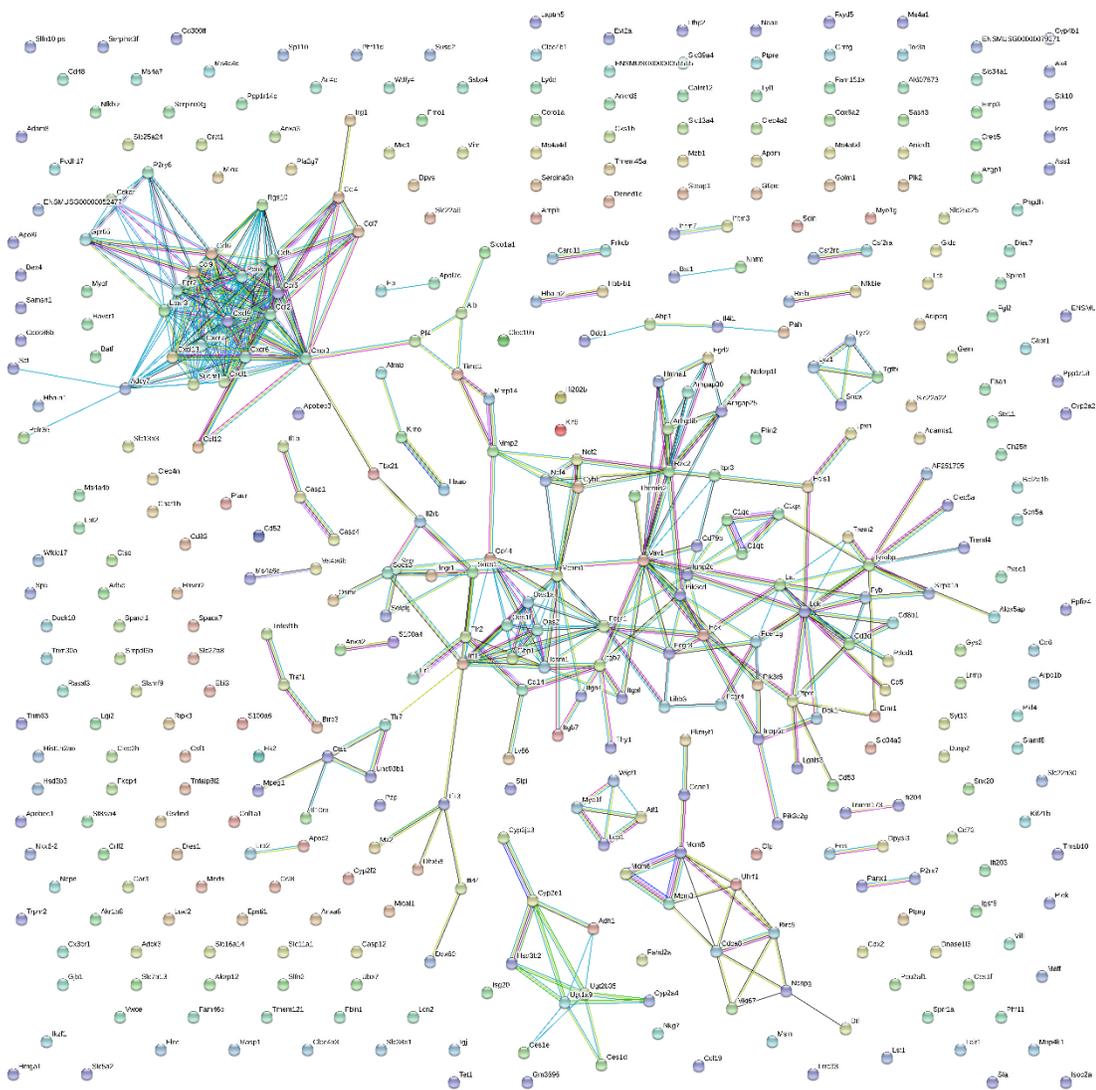


Figure 7: Protein-protein interactions visualized in STRING for the Intersection list. See Figure 6 legend for key.

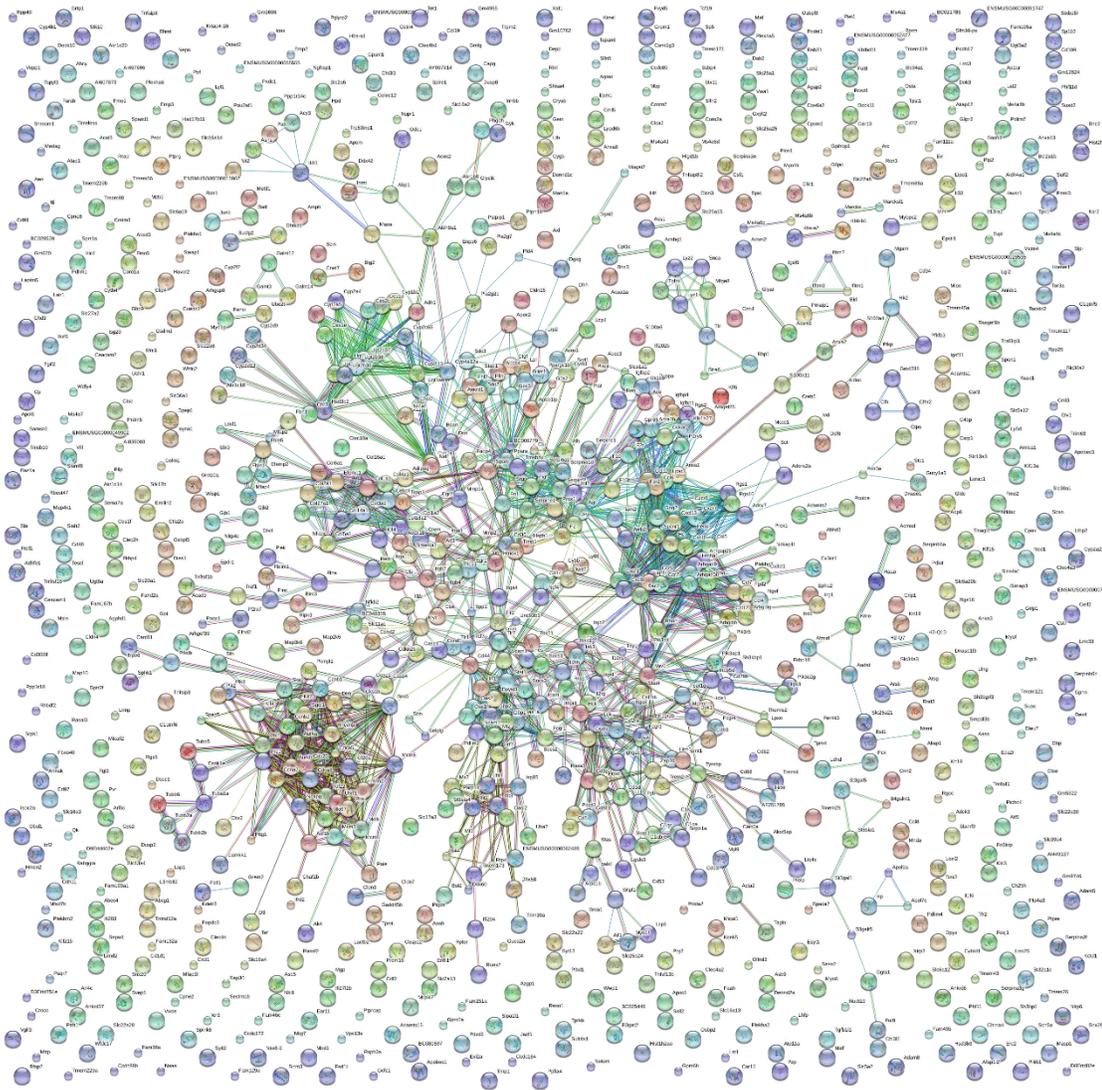


Figure 8: Protein-protein interactions visualized in STRING for the KO-WT list. See Figure 6 legend for key.

For Figures 6-8, similar results were obtained. Even at the highest level of confidence, hundreds of interactions were observed within each of the datasets. In addition to generating interaction networks, STRING also calculated the significance of the results. For the number of interactions found between the input genes, the probability of these genes sharing that many connections by chance is calculated.

(Table 9) For each dataset, the DEGs share a significant number of connections. This indicates that the DEGs within the datasets share a biological function.

Table 9: Statistical analysis of STRING results.

<i>Dataset:</i>	Number of nodes:	Number of edges:	Expected edges:	p-value:
<i>KO-WT</i>	112	1700	198	0
<i>DB-ND</i>	529	574	54	0
<i>Intersection</i>	411	424	38	0

In every interaction network produced in STRING, one centralized hub was observed. That hub is pictured in Figure 9. Figure 9 shows a close-up of the interactions between chemokine receptors and their ligands from the DB-ND list. However, this hub was also observed in the Intersection and KO-WT lists. By focusing on the nodes in the datasets that interact with multiple proteins (centralized hubs), it was found that the chemokine and cytokine families were likely to be important within the DB SKO kidney. Many gene products in the cytokine interaction hub had multiple connections, and they displayed many interactions with other gene products. The number of connections between these gene products indicate a relevant biological connection.

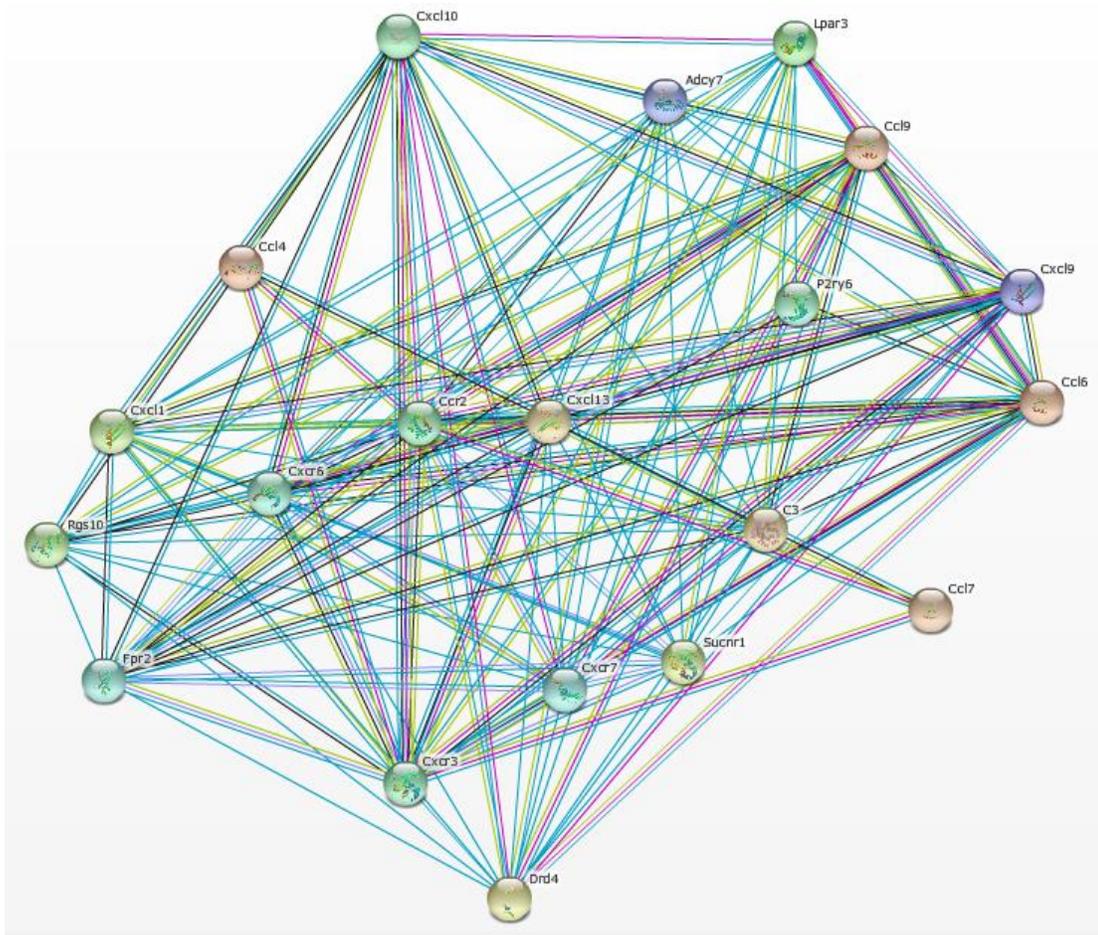


Figure 9: A close-up of a hub of interactions between affected chemokines and ligands produced using STRING for the DB-ND list. See Figure 6 legend for key.

C. GO Term Analysis

The tool used was the GoTermFinder, developed at the Lewis-Sigler Institute at Princeton (Boyle et al., 2004). Gene Ontology (GO) Terms are biological functions, processes, and cell components associated with a gene. The GoTermFinder calculates the statistical significance of how many input genes are associated with a particular GO Term. Statistically significant GO Terms are those that are associated with more

genes than would be predicted by random chance. The results are consistent across the datasets (Tables 10-12). The DEGs in the datasets are significantly associated with immune cell processes. These results indicate that in the DB SKO, the changes in gene expression may be affecting immune cell activation.

Table 10: KO-WT GO Term Analysis (Process) most significant results.

GO ID	Term	Corrected p-value
GO:0002376	immune system process	1.32E-50
GO:0006955	immune response	7.77E-43
GO:0006952	defense response	1.04E-38
GO:0006950	response to stress	7.97E-35
GO:0022610	biological adhesion	5.16E-32
GO:0007155	cell adhesion	6.21E-31
GO:0048518	positive regulation of biological process	6.90E-31
GO:0002682	regulation of immune system process	3.51E-30
GO:0002684	positive regulation of immune system process	4.96E-30
GO:0001775	cell activation	1.58E-29
GO:0044699	single-organism process	1.21E-28
GO:0006954	inflammatory response	1.70E-27
GO:0048522	positive regulation of cellular process	8.75E-26
GO:0045087	innate immune response	9.00E-26
GO:0044763	single-organism cellular process	6.37E-25
GO:0048583	regulation of response to stimulus	8.34E-25
GO:0045321	leukocyte activation	8.85E-25
GO:0048584	positive regulation of response to stimulus	2.07E-24
GO:0002252	immune effector process	1.70E-23
GO:0050896	response to stimulus	1.89E-23
GO:0001816	cytokine production	5.04E-22
GO:0030155	regulation of cell adhesion	9.58E-22
GO:0001817	regulation of cytokine production	2.47E-21
GO:0016477	cell migration	5.34E-21
GO:0098602	single organism cell adhesion	7.47E-21

Table 11: Intersection GO Term analysis (Process) most significant results.

GO ID	Term	Corrected p-value
GO:0006955	immune response	1.06E-55
GO:0002376	immune system process	2.34E-55
GO:0006952	defense response	3.61E-40
GO:0002684	positive regulation of immune system process	1.00E-33
GO:0002682	regulation of immune system process	9.02E-33
GO:0045087	innate immune response	1.82E-31
GO:0045321	leukocyte activation	2.09E-30
GO:0001775	cell activation	8.25E-29
GO:0006950	response to stress	1.76E-23
GO:0002252	immune effector process	2.40E-23
GO:0006954	inflammatory response	2.57E-23
GO:0046649	lymphocyte activation	4.81E-22
GO:0034097	response to cytokine	8.84E-22
GO:0007159	leukocyte cell-cell adhesion	4.66E-21
GO:0009605	response to external stimulus	1.04E-20
GO:0050776	regulation of immune response	2.40E-20
GO:0050900	leukocyte migration	1.44E-19
GO:0071593	lymphocyte aggregation	4.78E-19
GO:0070486	leukocyte aggregation	9.25E-19
GO:0002250	adaptive immune response	9.63E-19
GO:0050896	response to stimulus	2.44E-18
GO:0048584	positive regulation of response to stimulus	2.99E-18
GO:0042110	T cell activation	3.06E-18
GO:0070489	T cell aggregation	3.06E-18

Table 12: DB-ND GO Term analysis (Process) most significant results.

GO ID	Term	Corrected p-value
GO:0002376	immune system process	9.68E-60
GO:0006955	immune response	1.06E-59
GO:0006952	defense response	1.38E-47
GO:0045087	innate immune response	1.19E-37
GO:0002682	regulation of immune system process	3.70E-36
GO:0002684	positive regulation of immune system process	1.09E-34
GO:0045321	leukocyte activation	4.31E-29
GO:0001775	cell activation	1.92E-28
GO:0034097	response to cytokine	1.42E-26
GO:0006950	response to stress	1.46E-26
GO:0002252	immune effector process	3.50E-25
GO:0006954	inflammatory response	1.37E-24
GO:0009605	response to external stimulus	1.71E-23
GO:0043207	response to external biotic stimulus	2.44E-22
GO:0051707	response to other organism	2.44E-22
GO:0071345	cellular response to cytokine stimulus	3.84E-22
GO:0007159	leukocyte cell-cell adhesion	1.46E-21
GO:0046649	lymphocyte activation	2.41E-21
GO:0009607	response to biotic stimulus	3.86E-21
GO:0050776	regulation of immune response	4.45E-21
GO:0050900	leukocyte migration	2.63E-20
GO:0070486	leukocyte aggregation	1.31E-19
GO:0071593	lymphocyte aggregation	3.91E-19
GO:0031349	positive regulation of defense response	4.78E-19
GO:0048584	positive regulation of response to stimulus	5.60E-19
GO:0042110	T cell activation	2.17E-18

D. IPA

Ingenuity Pathway Analysis (IPA) is a commercial pathway analysis program. It uses normalized expression level data to predict biochemical systems within a certain phenotype, disease, or cell line. This software has several features such as the Diseases

and Functions tool and Causal Network Analysis (CNA) that help visualize the affected molecules and pathways. IPA also curates information on millions of genes and proteins, and the interactions between them, also known as the Ingenuity Knowledgebase. By using an extensive, regularly updated database, all results are consistent with current literature findings. The number of different features within IPA and the extensive database make it useful to analyze gene expression data.

1. Diseases and Functions

The Diseases and Functions Tool within IPA analyzes the predicted biological functions that are affected from the datasets. IPA uses the expression data from the DEGs to determine what biological functions are associated with the dataset. In Diseases and Functions output, the most significantly affected biological functions are pictured graphically. Each rectangle represents a biological function (e.g. hematological system development and function). This contains subsets of that function (e.g. cell movement). Within these rectangles are more specific subsets (e.g. cell movement of lymphocytes). The size of each rectangle represents the significance of the enrichment, as measured by a p-value. The intensity of the color represents the degree to which that function is affected, as measured by the z-score. Orange represents activation, while blue represents inhibition. White represents little to no statistically measurable effect, and grey means that no z-score could be calculated.

In Figure 10A, the most affected biological functions in the DB SKO from the Intersection dataset can be observed: hematological system development and function, inflammatory response, immune cell trafficking, cellular movement, and cell-to-cell signaling and interaction. In Figure 10B, a close-up of the hematological system development and function is shown. In the green square, cell movement of phagocytes is enriched. In the red square, cell movement of leukocytes is enriched. All three datasets showed consistent results that indicated overactivation of immune system processes. Similar to the DAVID results, this suggests that a wide variety of immune cell functions are affected and, more specifically, are upregulated in the DB SKO kidney.

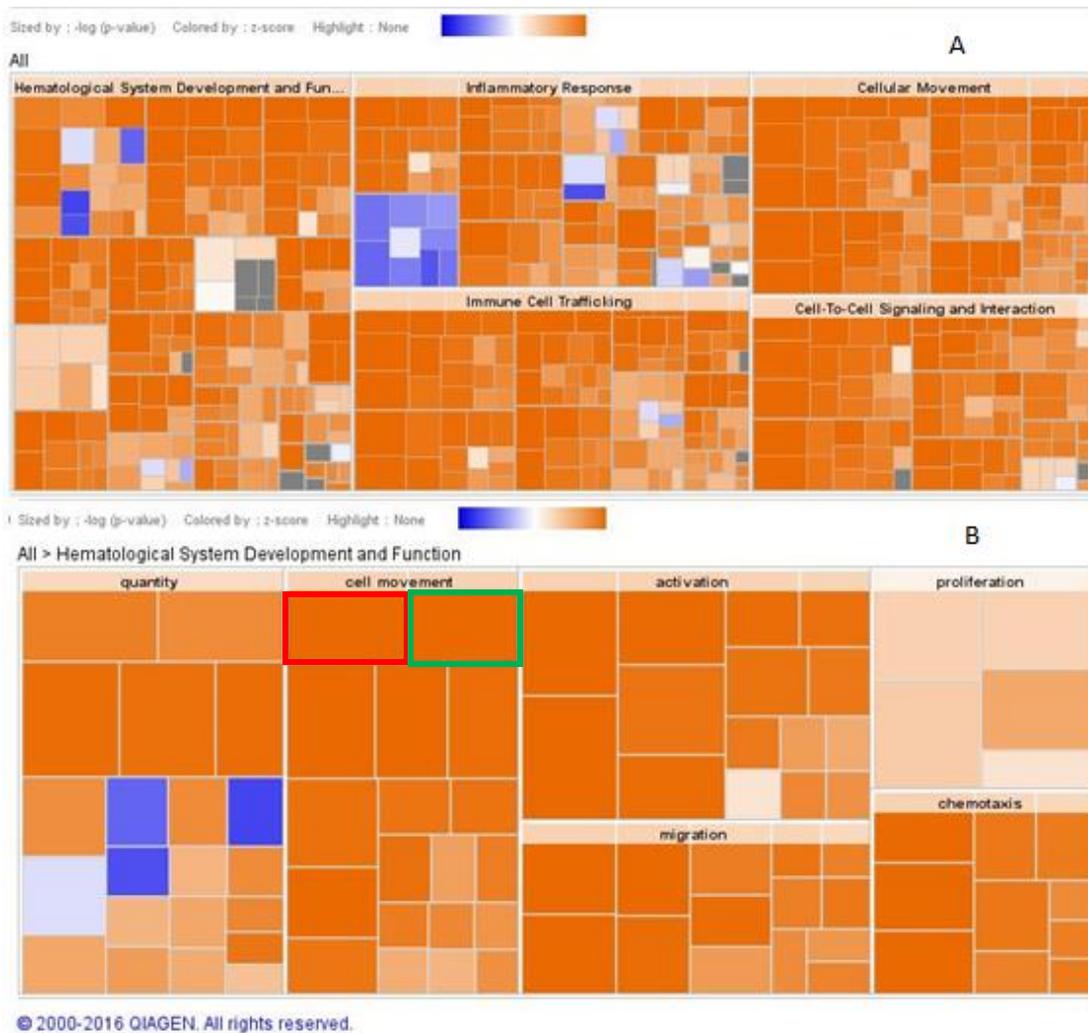


Figure 10: IPA Diseases and Function graphical results. A: Enrichment of diseases and functions within the intersection dataset. B: Close-up of the functions within the umbrella term “Hematological System Development and Function”. Each smaller rectangle represents an immune cell subtype function. For example, the cell movement of leukocytes (red box) and cell movement of NK cells (green box) are both affected. The color represents the degree to which that function is affected, as measured by the z-score. Orange represents activation, while blue represents inhibition. White represents little to no statistically measurable effect, and grey means that no z-score could be calculated.

2. Causal Network Analysis

Causal Network Analysis (CNA) was used to identify upstream molecules potentially acting on, or regulating, the differentially expressed genes within our dataset. By examining the predicted upstream regulators of the DEGs within the dataset, a link between the loss of STAT5 and the gene expression changes can be found. This proposed mechanism is visualized in Figure 11. In the simplest model explaining the dramatic changes in gene expression upon loss of STAT5, STAT5's normal function could be inhibition of a key regulator. Thus, in the absence of STAT5, this inhibition would be relieved and the regulator would be free to act. This regulator could act directly or indirectly as an "upstream regulator" of DEGs in the DB SKO kidney. By identifying an influential molecule, or "master regulator", upstream of the DGEs but downstream of STAT5, the link between the loss of STAT5, drastic changes in gene expression, and the increased kidney damage in the DB SKO mice can be explored.

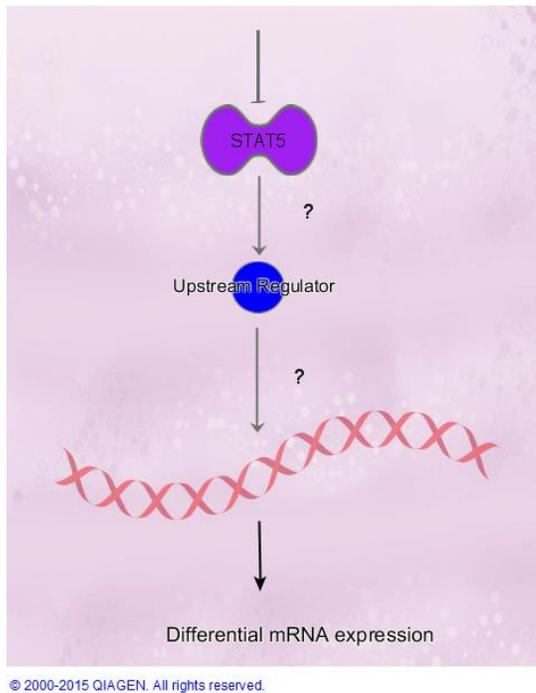


Figure 11: Proposed interaction between STAT5, an intermediate upstream regulator, and the affected genes. Identification of a “master regulator” could solidify the link between the loss of STAT5 and the increased kidney damage.

Using CNA produced hundreds of implicated upstream regulators for each dataset. Each predicted upstream regulator had an associated interaction network with the DEGs. To identify a master regulator whose direct interactions with DEGs could be assessed in mouse kidneys at a later point, all interaction networks were examined for a predicted upstream regulator that interacted directly with the DEGs.

The previously described criteria were used to identify the best candidate upstream regulator. Any predicted upstream regulator had to be a transcription factor that interacted downstream of STAT5. It also had to be predicted during the CNA results for all three datasets. The following molecules met all three criteria:

1.	EGR3
2.	GATA3
3.	NKX2-5
4.	SMAD5
5.	TBX5
6.	JUND
7.	TAL1

This list of potential regulators obtained from IPA could affect a large number of genes in each dataset and, therefore, account for many of the changes observed in the DB SKO kidneys. The potential master regulator interaction networks were then examined individually.

After narrowing down the list of potential master regulators, the interaction networks for each transcription factor were examined more closely. The proposed approach was to determine how many differentially expressed genes within our dataset could be affected. However, a similarity between the figures soon became apparent. TAL1, TBX5, JUND, and EGR3 were all depicted as directly upstream of the NF κ B complex (Figures 7-13). Furthermore, the TAL1 and TBX5 interaction networks showed NF κ B directly interacting with the majority of dataset genes within the predicted interaction networks (Figures 7-10).

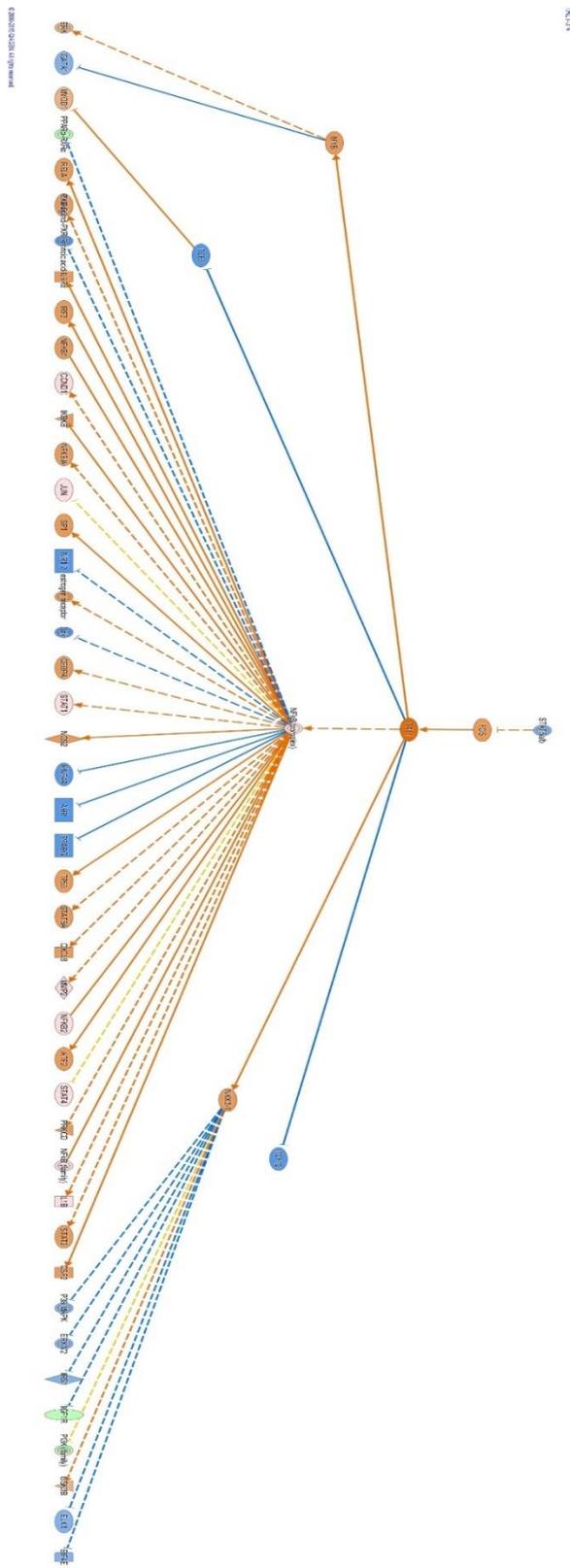


Figure 12: Interaction map of TAL1 predicted in CNA. A red node indicates an activated dataset gene, while a green node represents an inhibited dataset gene. An orange node represents a predicted activated co-regulator, and blue represents an inhibited co-regulator. The orange connections show predicted upregulation. Dark blue shows predicted downregulation. Yellow identifies a conflict between a predicted regulator effect and gene expression.

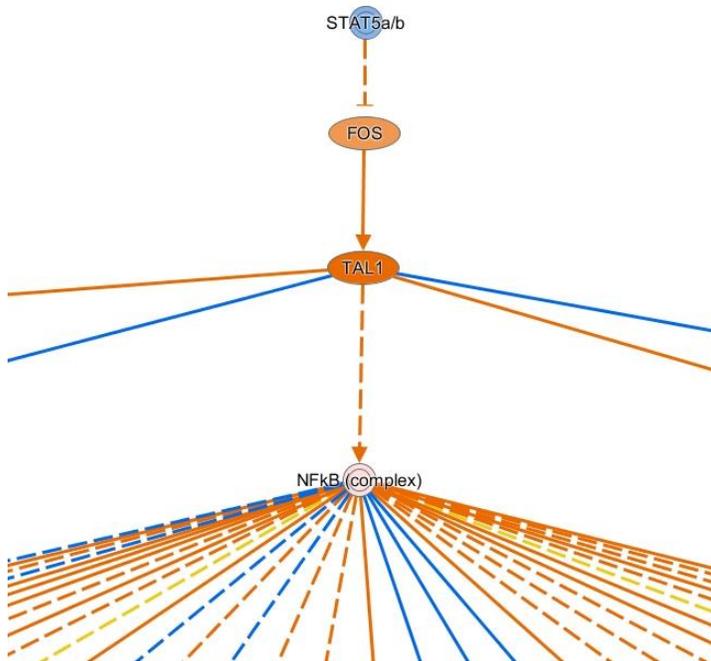


Figure 13: Close-up of the TAL1 interaction network. See Figure 12 legend for key.

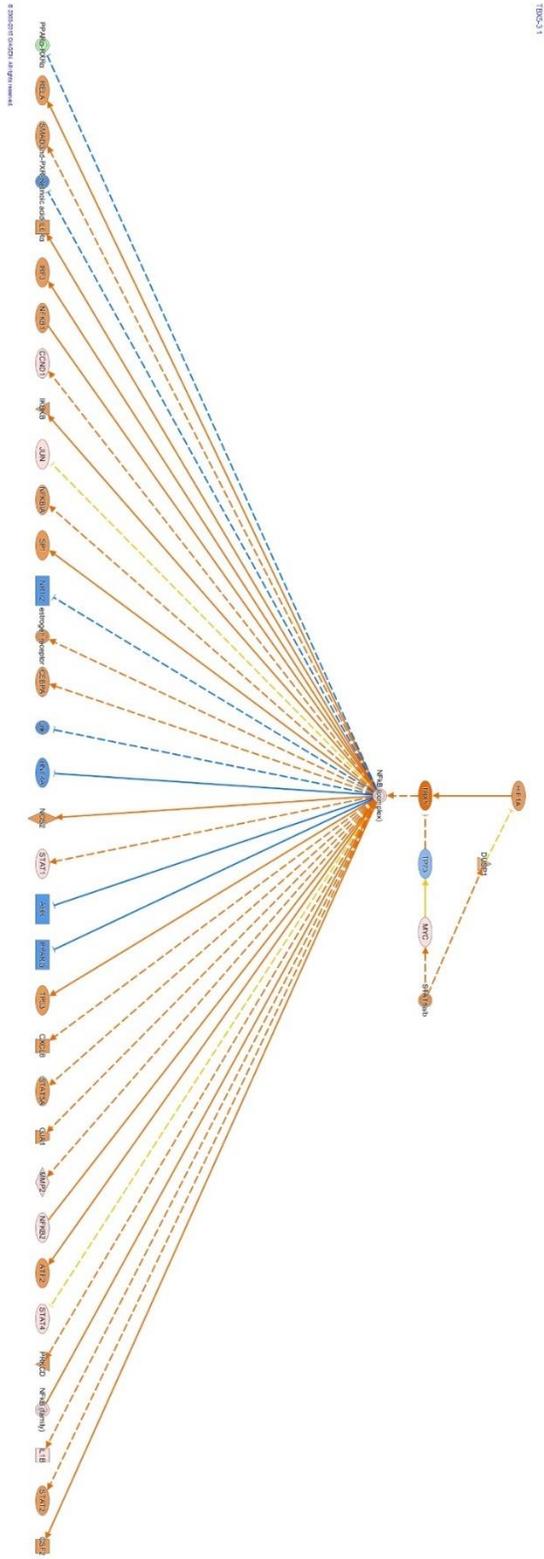


Figure 14: Predicted TBX5 interaction network. See Figure 12 legend for key.

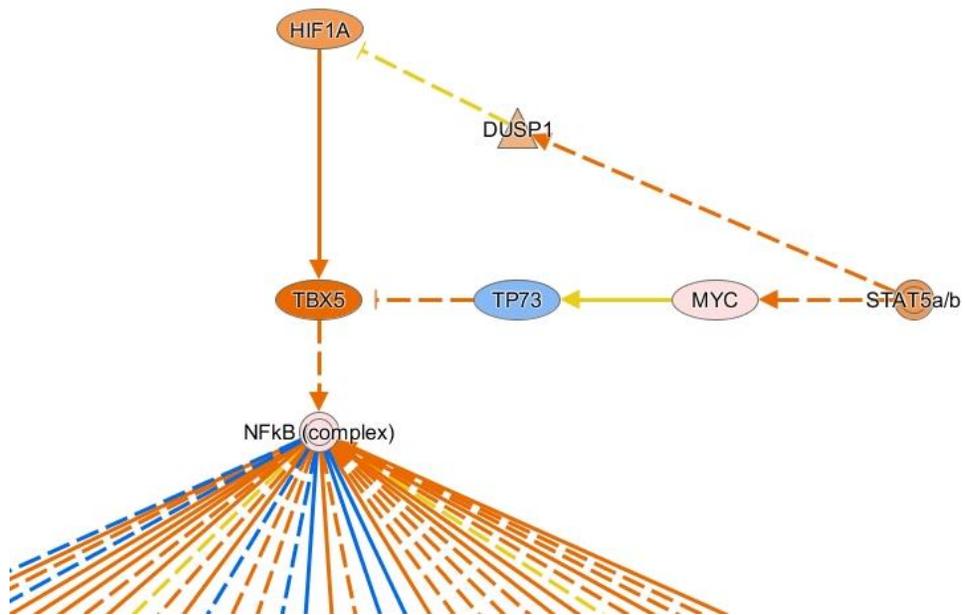
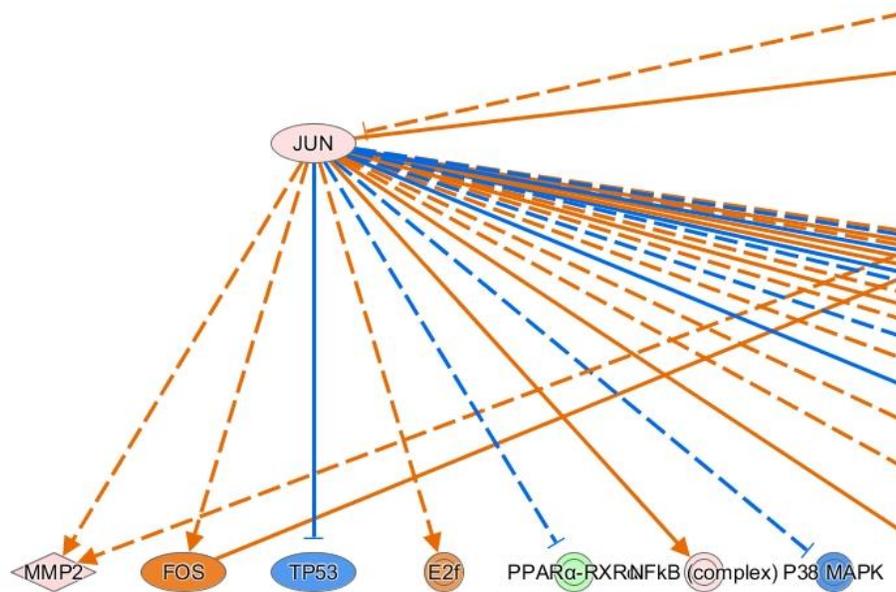


Figure 15: Close-up of the TBX5 interaction network. See Figure 12 legend for key.



© 2000-2015 QIAGEN. All rights reserved.

Figure 17: Close-up of a portion of the JunD interaction network. See Figure 12 legend for key.

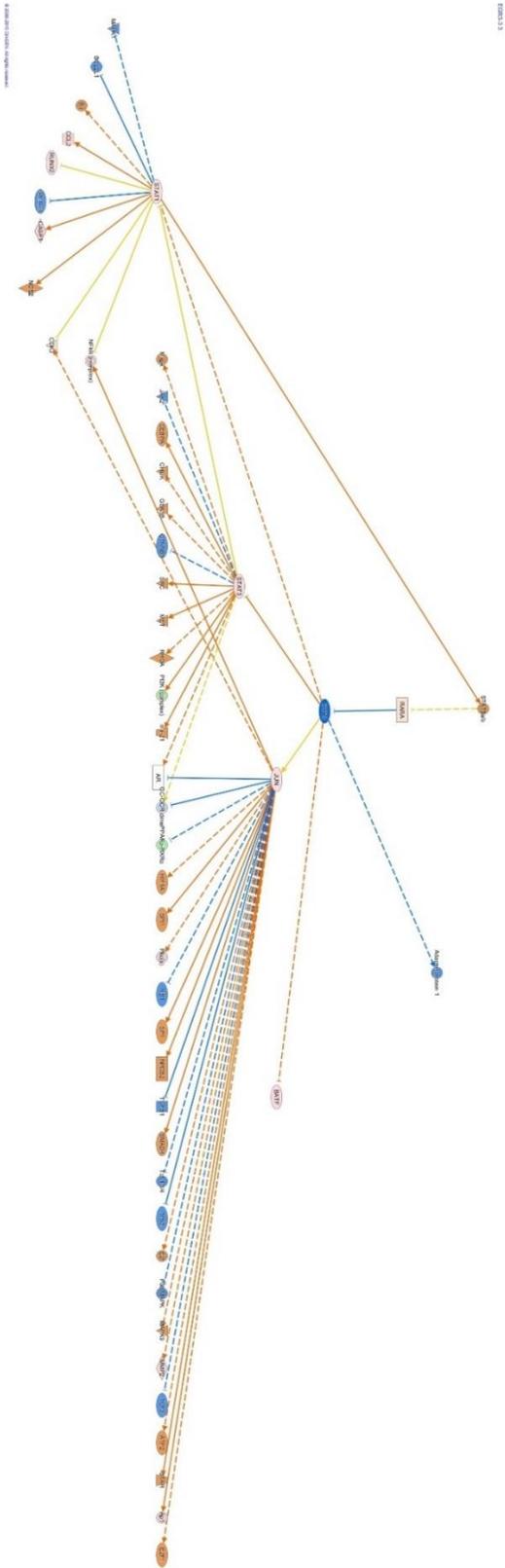


Figure 18: Predicted EGR3 interaction network. See Figure 12 legend for key.

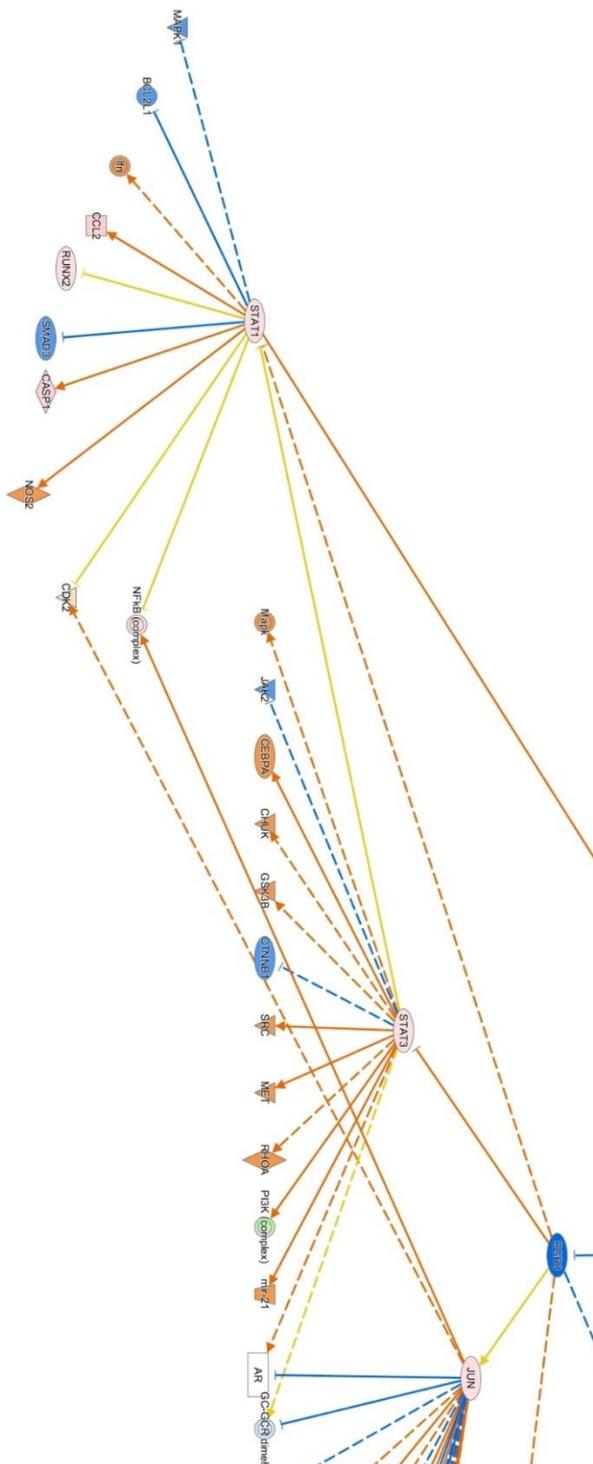


Figure 19: Close-up of a portion of the EGR3 interaction network. See Figure 12 legend for key.

After examining the interaction networks, the following general pattern was established:

Stat5 → Regulator(s) → NFκB → dataset genes

This pointed to NFκB as a potential master regulator in the DB SKO kidney. To fully determine the scope of NFκB interactions within the various datasets, NFκB was viewed in the Upstream Regulator function in IPA (Figures 20 and 21). In Figure 20, the entire predicted interaction network for NFκB is shown. In addition to directly interacting with a large number of dataset genes, NFκB was predicted to indirectly interact with additional dataset genes by affecting the activity of several molecules outside of the dataset. An example of this can be seen in Figure 21, where affected genes in the Intersection dataset are seen in green and red, and molecules outside of the dataset are seen in blue and orange.

Overall, the NFκB complex was predicted to either directly or indirectly regulate activity of 192 out of 439 genes within the Intersection dataset (Figure 15). Through regulation of such a large number of genes, an increase in NFκB binding activity could potentially explain much of the increased tissue damage and transcriptome changes observed in the DB SKO kidney. While the absence of STAT5 likely affects multiple co-regulators that up- or down-regulate the DEGs within the datasets, the large interaction network predicted for NFκB in the DB SKO kidney

positions NF κ B to be a major regulator. Examination of NF κ B activity within the DB SKO kidney could help support this major role.

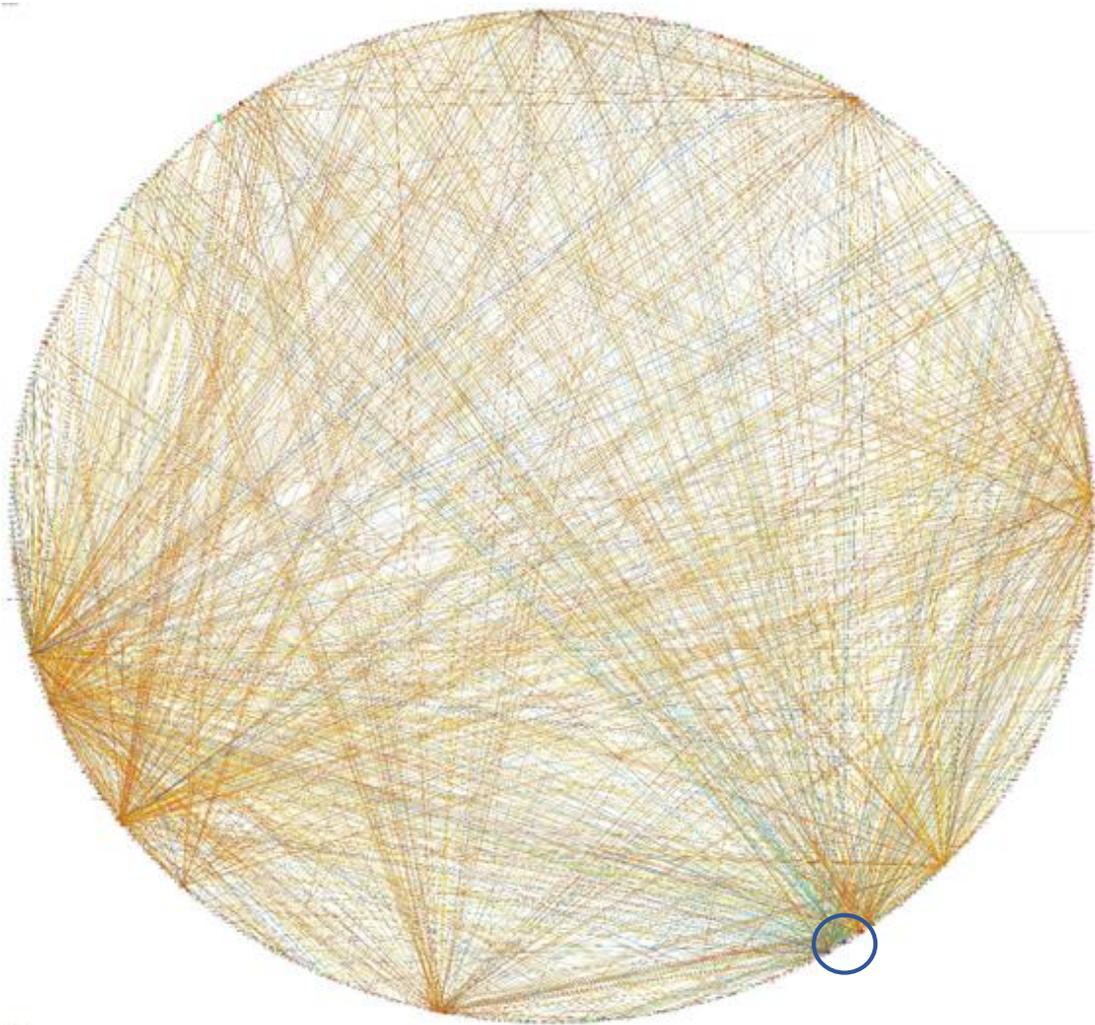


Figure 20: Radial view of NF κ B interactions within the intersection dataset. NF κ B is circled in blue. This figure shows the complexity of predicted NF κ B activity within the dataset. Each node is either a DEG within the Intersection dataset or a predicted co-regulator. A pink node indicates an activated dataset gene, while a green node represents an inhibited dataset gene. An orange node represents a predicted activated co-regulator, and blue represents an inhibited co-regulator. The orange connections show predicted upregulation. Dark blue shows predicted downregulation. Yellow connections identify a conflict between a predicted regulator effect and gene expression. Light blue identifies direct NF κ B interactions. Solid lines represent an observed direct interaction, while a dashed line represents an observed indirect interaction.

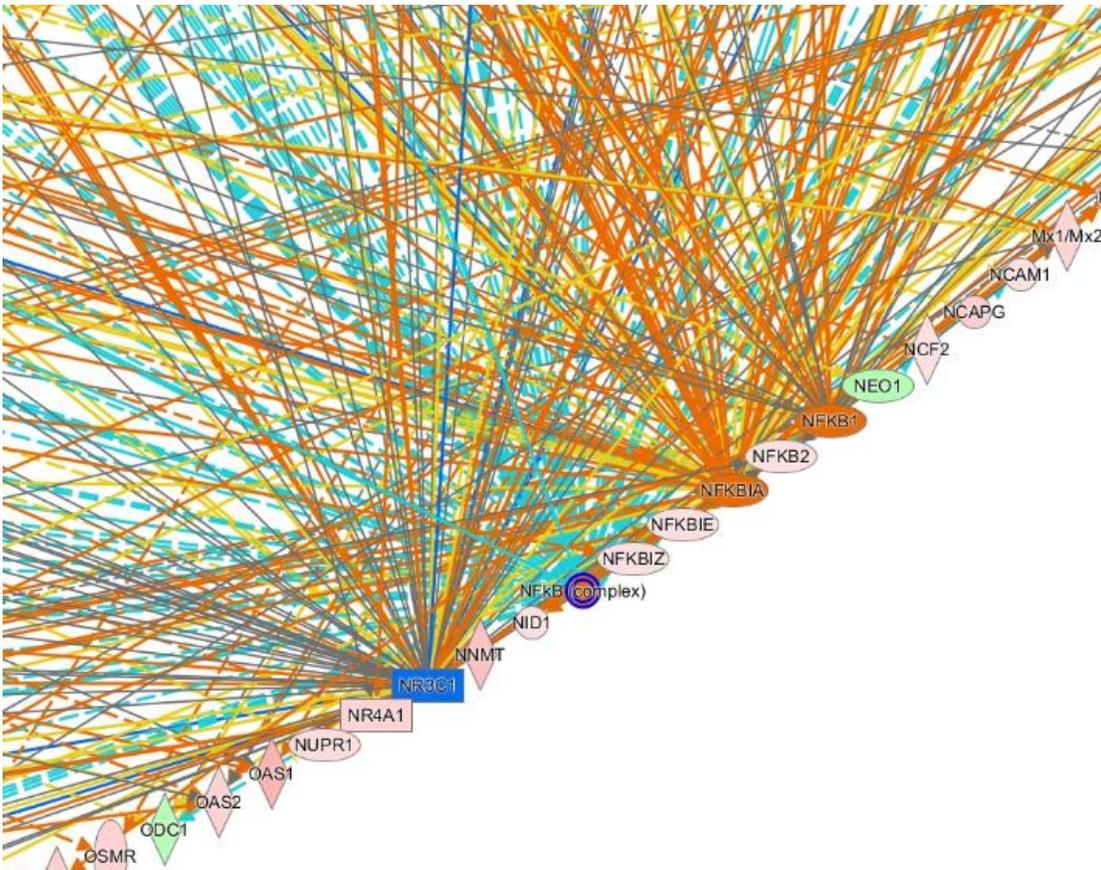


Figure 21: Close-up of blue circled area in Figure 20. See Figure 20 key for legend.

E. Selection of RelA/p65 for *in vivo* binding activity measurements

CNA showed that NFκB could account for the expression of a large number of the differentially expressed genes within the DB SKO kidney. This supported earlier findings from DAVID that showed enrichment in NFκB signaling (Figure 22).

However, CNA predicted activation and interaction of the general NFκB complex, not a specific transcription factor within the family. To confirm NFκB-DNA binding *in vivo*, the specific proteins themselves must be assayed.

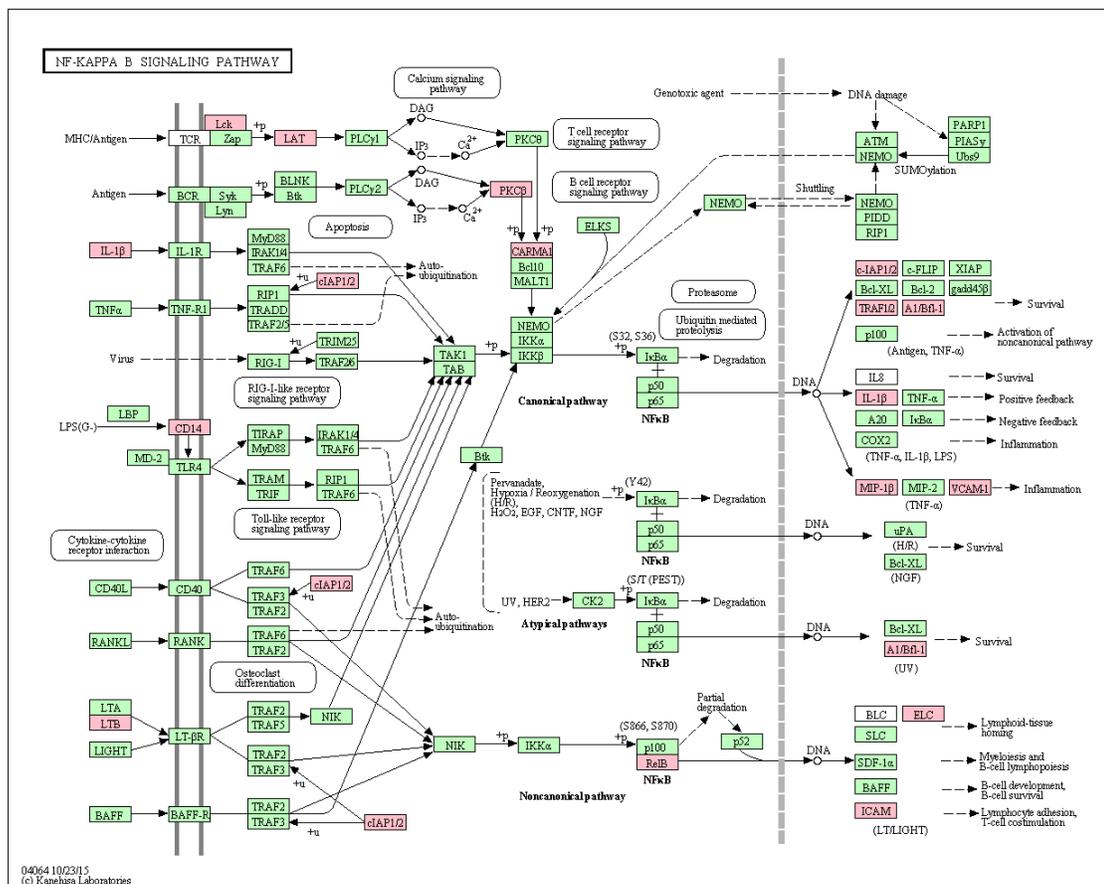


Figure 22: NF κ B signaling KEGG pathway. Gene products in pink are upregulated in the intersection dataset.

Due to time restraints, one NF κ B family member was chosen for *in vivo* measurement of binding interactions with dataset genes. RelA, RelB and c-Rel are the NF κ B family members with both DNA binding and transcriptional activation domains (Rel homology domain (RHD) and transcriptional activation domain (TAD)); Figure 3). While c-Rel can induce transcription, the most established NF κ B is the canonical and noncanonical signaling that occurs through RelA or RelB (Gilmore, 2006). Activation or inhibition of either RelA or RelB could explain the changes in gene

expression in the DB SKO kidney. Therefore, the transcription factor with the most predicted interactions with dataset genes was chosen for ChIP. In the Intersection dataset, RelA was predicted to interact with a total of 166 genes (Figure 23) as compared to 88 for RelB. Therefore, RelA was chosen for measurement of DNA binding using a chromatin immunoprecipitation (ChIP) protocol.

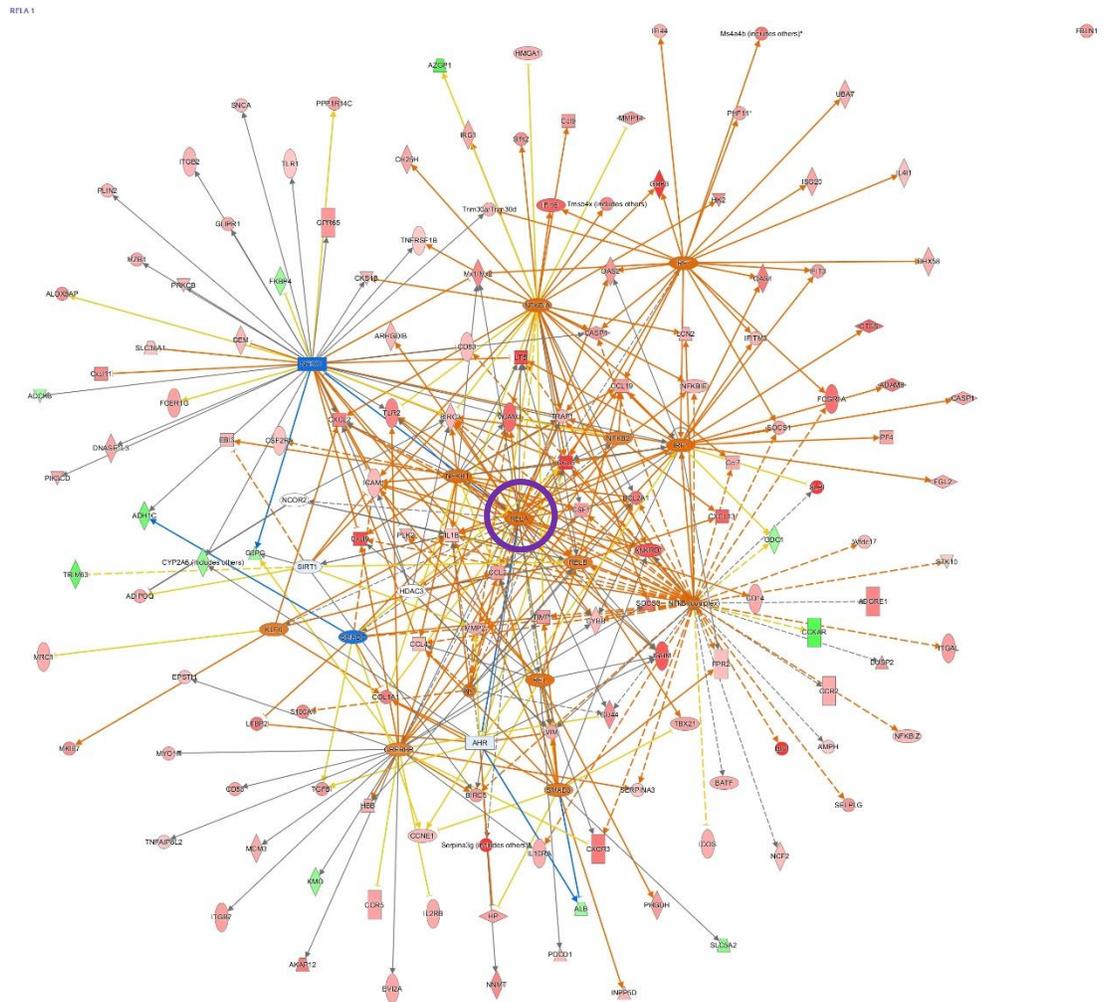


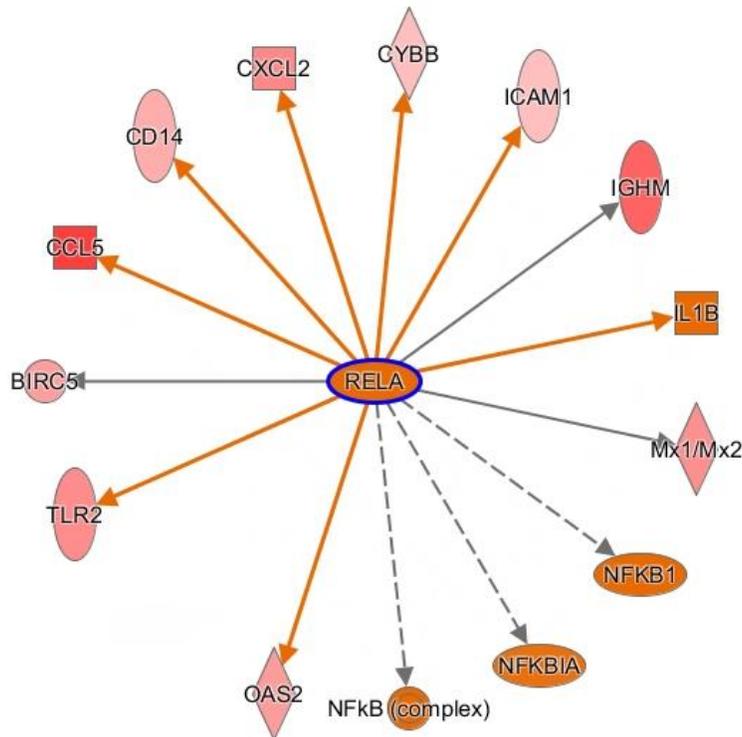
Figure 23: RelA and potential co-regulators. RelA is pictured at the center, circled in purple. A red node indicates an activated dataset gene, while a green node represents an inhibited dataset gene. An orange node represents a predicted activated co-regulator, and blue represents an inhibited co-regulator. The orange connections show predicted upregulation. Dark blue shows predicted downregulation. Yellow identifies a conflict between a predicted regulator effect and gene expression. Solid lines represent an observed direct interaction, while a dashed line represents an observed indirect interaction.

F. Measurement of *in vivo* p65/RelA DNA Binding Events

1. Selection of Potential RelA-Dependent Gene Targets

To identify DEGs that could be targets of *in vivo* P65/RelA DNA binding in the DB SKO mouse kidney, the interaction network in Figure 11 was refined. First, all genes that were not present in both mouse and humans were removed. This was to prevent choosing a target that did not exist in our mouse model or would have little relevance to human DN. Second, all genes that did not interact directly with RelA were removed. This was because ChIP can only determine direct protein-DNA binding events. Lastly, all genes that did not have a predicted protein-DNA binding interaction with RelA were disqualified. If RelA affected transcription of these genes in a different manner, then ChIP would not be able to assay for that. The final list of potential targets is shown in Figure 24. These are all of the possible ChIP targets that met the criteria.

RELA 1



© 2000-2016 QIAGEN. All rights reserved.

Figure 24: Direct RelA interactions in *mus musculus*. All red nodes indicate a DEG within the Intersection dataset, along with IL1B, which was designated orange, due to its predicted effect upon other DEGs. All orange nodes are non-dataset predicted affected genes. Other members of the NF κ B family were also predicted to be affected.

2. *In silico* Identification of Potential p65/RelA Binding Sites

Transcription factor binding site scanning was performed using FIMO (Grant et al., 2011). This was used to predict sites of specific protein-DNA interactions. This information can then be used to target measurements of protein-DNA interactions. The binding motif for RelA was identified as 5'-BGGRNWTYCC-3' (Figure 25).

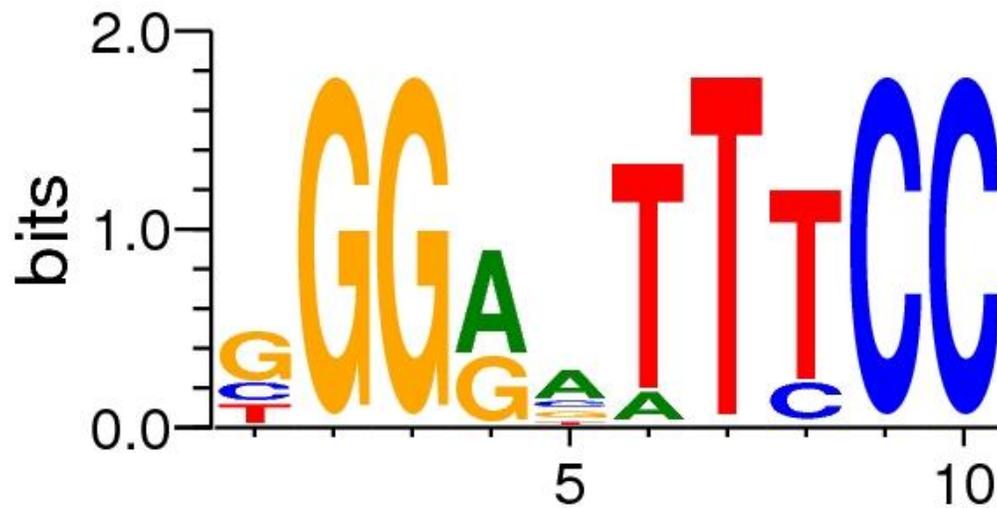


Figure 25: RelA binding motif obtained from the JASPAR database.

Binding site scanning for this motif in the upstream regions of the 11 potential RelA-target genes identified at least one site in 5 of the genes (Table13). Primers were then designed using Primer-BLAST (Ye et al., 2012). This was to amplify regions near these binding sites for measurement of p65/RelA binding events by quantitative PCR following ChIP.

Table 13: The RelA motif sequences and locations.

RelA Dependent Gene	Sequence	Location (bp upstream of TSS)
TLR2	GGGGGTTTCC	958
	GGGGAATTCC	4
BIRC5	GGGACTTTCC	48
ICAM1	TGGAAATTCC	164
IL1B	GGGGGTTTCC	679
CXCL2	AGGACATCCC	191
	GGGAATTTC	59

3. ChIP-qPCR Results of p65/RelA Binding Event Measurements

Chromatin Immunoprecipitation (ChIP) is an experimental technique that can be used to study the DNA-binding interactions of a protein *in vivo*. This requires freezing the interactions in place by cross-linking the chromatin and any associated proteins within the samples of interest (in this case kidneys of the four mouse groups previously studied: ND WT, DB WT, ND SKO and DB SKO), isolating chromatin-protein complexes from the nuclei of the cells, shearing the chromatin into practically sized fragments (ideally 200-1200 bp), isolating specific protein-chromatin complexes using an antibody specific for the protein of interest (p65/RelA and a positive and negative control), reversing the protein-chromatin cross-links and collecting the released DNA for quantitative analyses. Primers specific for the dataset genes of interest can then be used in quantitative polymerase chain reaction (qPCR) assays, comparing the number

of binding events obtained for a given gene between the kidney samples used for the ChIP.

To first verify that the ChIP reactions worked, several values were examined (Table 9). First, for each ChIP reaction, the binding events of the negative control qPCR primer set was <0.09 binding events per 1,000 cells, well below the manufacturer's recommended value of <2 binding events per 1,000 cells, indicating extremely low background. Second, the fold-enrichment of Pol II binding events measured by the gene-specific primers over the negative primers ranged from 1 to 81 fold while the fold-enrichment of p65 binding events ranged from 0.8 to 52 fold. Fold enrichment greater than 5 fold is highlighted in blue in Table 14.

Table 14: qPCR Results.

Genomic Site	Binding events detected per 1,000 cells*					
	ND-WT + p65	ND-SKO + p65	DB-WT + p65	DB-SKO + p65	ND-WT + Pol II	ND-WT + IgG
Actb-1	0.41 (6.8)	0.51 (8.5)	0.49 (5.4)	0.38 (7.6)	0.33 (33)	0.05 (2.5)
Negative	0.06	0.06	0.09	0.05	0.01	0.02
Il1B	0.05 (0.8)	0.07 (1.2)	0.09 (1.0)	0.07 (1.4)	0.01 (1.0)	0.01 (0.5)
CXCL2	0.09 (1.5)	0.12 (2.0)	0.11 (1.2)	0.07 (1.4)	0.01(1.0)	0.01 (0.5)
ICAM1	2.66 (44)	2.74 (46)	1.20 (13)	2.52 (50)	0.81 (81)	0.34 (17)
BIRC5	1.42 (24)	2.67 (45)	1.67 (19)	2.62 (52)	0.69 (69)	0.30 (15)
TLR2	0.19 (3.2)	0.26 (4.3)	0.29 (3.2)	0.30 (6.0)	0.05 (5.0)	0.03 (1.5)

* Fold-enrichment of specific binding over background (negative control) is shown in parentheses. All gene-specific and positive primers with a fold-enrichment ≥ 5 are highlighted in blue.

In Figure 26, the binding frequencies are visualized. In regards to the gene-specific primer pairs, only two showed any differences in RelA binding between the four sample groups. ICAM1-RelA binding was observably lower in the DB WT group as compared to the other three. BIRC5 also showed higher RelA binding in the SKO groups as compared to the WT groups. Because each group had a sample size of one, it could not be determined if there was a statistical difference between the groups.

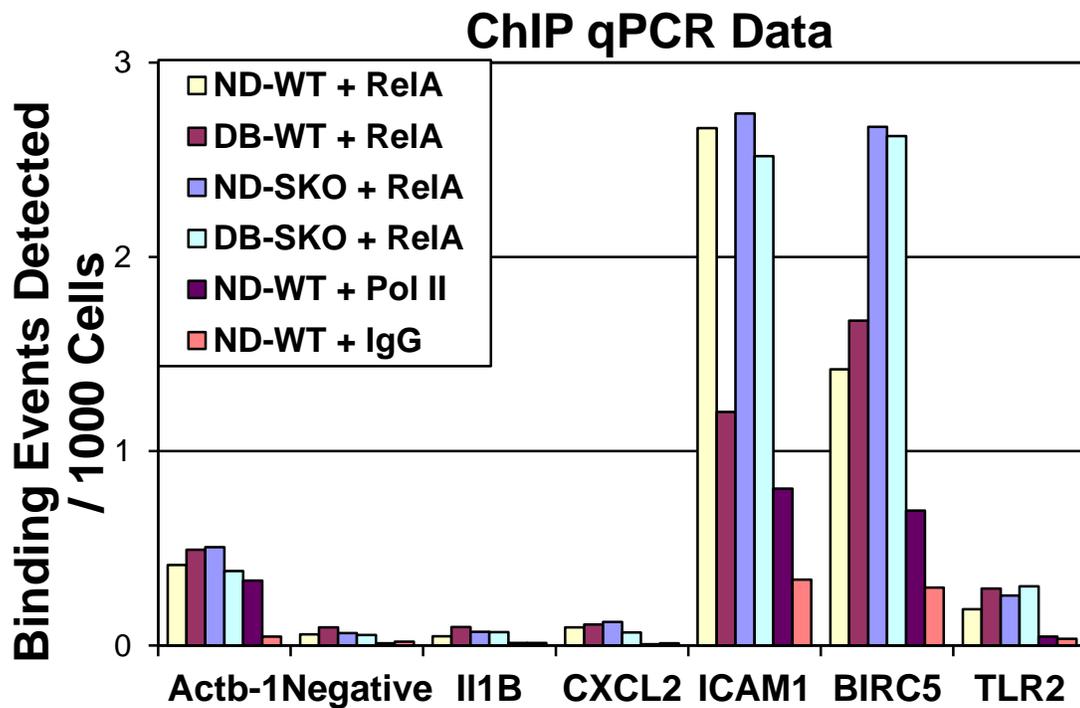


Figure 26: ChIP qPCR Results. The relative binding frequency of RelA is measured on the Y-axis (binding events/1000 cells). The seven primer pairs used are shown on the X-axis.

VI. Discussion

In T1D, diabetic kidney damage is a common secondary complication. This is due to elevated glucose levels, resulting in increased flux through a number of metabolic pathways (Brownlee, 2004). These end products cause cellular damage, resulting in inflammation and tissue damage. Due to the high progression rates to end-stage renal disease and the high mortality rates of end-stage renal disease, preventing kidney damage in T1D is of utmost importance (Reidy et al., 2014). In previous diabetes studies of the SKO mouse line, the DB SKO mouse showed increased kidney damage, inflammation, and differential gene expression as compared to the ND SKO and DB WT. The pathology of the DB SKO kidney indicated an essential role for STAT5 in protecting against DN.

To study the molecular effects of the absence of STAT5 within the diabetic kidney, gene expression results from the DB SKO kidney were examined. Using a variety of *in silico* tools, results were obtained that indicated the link between the loss of STAT5 and the exacerbated damage in the DB SKO kidney was predominantly the upregulation of pro-inflammatory genes, indicating possible immune system upregulation. In support of this finding, the pathway analysis tools showed enrichment of pathways crucial for a host of immune cell functions. The IPA Diseases and Functions tool also confirmed the consistent dysregulation of the immune system in the DB SKO kidney.

The proposed master regulator, NF κ B, is a well-studied, widely-expressed, pro-inflammatory transcription factor that is crucial for immune cell function (Gilmore, 2006). In the DB SKO kidney scenario, upregulation of NF κ B could explain many of the pathological changes observed. Increased p65 DNA binding activity was seen in protein lysates made from DB SKO kidney as compared to ND SKO and DB SKO (Coschigano et al., 2013). Activation of pro-inflammatory NF κ B-dependent genes in a variety of immune cells could explain the increase in immune cell movement, quantity, and activation. Widespread activation of NF κ B could lead to the indiscriminate activation of immune cell pathways and functions. However, *in vivo* confirmation of increased NF κ B activity in the DB SKO kidney is needed to prove the validity of this claim.

ChIP-qPCR was performed to test the DNA-binding activity of RelA. Due to the small sample size, statistical analyses could not be performed to determine significant differences between the DB SKO and the other groups. While no statistical inferences can be drawn from the ChIP-qPCR experiment, future experiments can draw on it. The RelA binding activity in the promoters of ICAM1 and BIRC5 should serve as a starting point for assessing NF κ B activity.

The *in silico* analyses also indicated that RelB may be involved in the DB SKO pathology since RelB was overexpressed in the datasets. Also, certain proteins, such as RelB, VCAM1, and LAT in the non-canonical pathway were also overexpressed. (Figure 22) Out of the three members of the NF κ B family that can induce transcription, only one was assayed *in vivo*. If multiple members of the NF κ B family

are responsible for increasing transcription of pro-inflammatory genes, that could explain the large scope of the affected pathways and functions.

However, there is a flaw in depending entirely upon expression data to predict pathway enrichment. Increased mRNA transcription does not guarantee an increase at the protein level. To get a complete picture of the changes in the DB SKO, other factors must be taken into account. Measuring protein levels in the DB SKO could portray expression changes more accurately.

Protein activation also could be dependent on factors besides increased mRNA transcription. Microarray data doesn't measure the level of phosphorylation of proteins. Phosphorylation is a common mechanism for activating proteins in signaling pathways (Chattopadhyay and Sen, 2014). It is possible that in the DB SKO, protein phosphorylation data could enhance the pathway analysis.

The methods used rely heavily on observed interactions from public databases. Only STRING predicts novel interactions, through the use of literature mining. IPA was used for the entirety of the identification of RelA as a master regulator. It is possible that the analyses missed a significant novel interaction between STAT5 and another protein, which in turn, could explain the differential transcription in the DB SKO kidney. In the future, STRING could be used for predicting novel interactions between RelA and the DEGs.

This project highlighted the importance of confirming results from *in silico* tools. Without the results from the ChIP-qPCR experiment, identifying RelA as a

master regulator would not have been completely biologically accurate. It is quite possible that multiple members of the NF κ B family interact to affect transcription, upregulate the immune system, and increase kidney damage in the DB SKO kidney.

In the future, this project could be continued in Dr. Coschigano's lab. While the results obtained were not expected, they revealed a potential mechanism of the increased tissue damage and inflammation in the DB SKO kidney. The experiment could be repeated, after changing a few parameters. First of all, the ChIP protocol used in the experiment could be further refined. Dr. Coschigano's lab had never performed ChIP before. The time needed for thorough troubleshooting exceeded the time allotted for completing this thesis. The sample sizes should also be larger. In this experiment, a single kidney was used for each group. This limits the statistical analyses that can be performed on the obtained qPCR data. Therefore, our results could not be evaluated based on statistical significance. By using multiple tissue samples for ChIP (biological replicates), the binding frequency for RelA could be determined with more confidence. Finally, ChIP-seq could be used instead of ChIP-qPCR. A high throughput sequencing technique would be more suited for a larger, broader experiment. This method would be used to sequence all of the DNA fragments that were bound to RelA, not just targets that were assessed using qPCR. This could help identify other targets. Other targets could possibly be identified. By changing some parts of the ChIP protocol, more reliable results could be obtained.

The majority of the analyses indicate immune system overactivation. A large number of pathways that are critical to immune cell signaling are predicted to be

affected. This corresponds to a drastic change in multiple functions across immune cell subtypes. The identification of RelA as a master regulator agreed with the previous results- that a major signaling pathway, or multiple pathways, had to be affected to cause such drastic changes in the DB SKO kidney. While increased binding and involvement of RelA was not yet able to be confirmed *in vivo* in the DB SKO kidney, the involvement of NF- κ B can continue to be studied. This analysis indicates that STAT5 may play a protective role in T1D by regulating the immune response to renal damage.

VII. References

- Belle, T. L. V., Coppieters, K. T. & Herrath, M. G. V. Type 1 Diabetes: Etiology, Immunology, and Therapeutic Strategies. *Physiological Reviews* **91**, 79–118 (2011).
- Bjornstad, P., Cherney, D. & Maahs, D. M. Early Diabetic Nephropathy in Type 1 Diabetes – New Insights. *Curr Opin Endocrinol Diabetes Obes* **21**, 279–286 (2014).
- Boyle, E. I. *et al.* GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**,3710–3715 (2004).
- Brosius F. Low-Dose Streptozotocin Induction Protocol (Mouse). *Diabetic Complications Consortium*. (2015)
- Brownlee, M. Biochemistry and molecular cell biology of diabetic complications. *Nature* **414**, 813–820 (2001).
- Brownlee, M. The Pathobiology of Diabetic Complications A Unifying Mechanism. *Diabetes* **54**,1615–1625 (2005).
- Centers for Disease Control and Prevention. National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, 2014. Atlanta, GA: U.S. Department of Health and Human Services (2014).

- Chattopadhyay, S. & Sen, G. C. Tyrosine phosphorylation in Toll-like receptor signaling. *Cytokine Growth Factor Rev.* **25**, 533–541 (2014).
- Coschigano, K.T., Heine, E., Malgor, R. Increased Tubulointerstitial Damage, Macrophage Infiltration, and Nuclear Factor- κ B Activity in the Kidneys of Diabetic STAT5 Knockout Mice. American Society of Nephrology Renal Week 2013, Atlanta, GA, TH-PO399 (poster)
- Dawson, M.A. & Bannister, A.J. Nuclear Functions of the Janus Kinases. *JAK-STAT Signaling: From Basics to Disease*. Springer-Verlag Wien. 27-46. (2012)
- Duran-Salgado, M. B. & Rubio-Guerra, A. F. Diabetic nephropathy and inflammation. *World J Diabetes* **5**, 393–398 (2014).
- Elmarakby, A. A. & Sullivan, J. C. Relationship between Oxidative Stress and Inflammatory Cytokines in Diabetic Nephropathy. *Cardiovascular Therapeutics* **30**, 49–59 (2012).
- Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
- Gilmore, T. D. Introduction to NF- κ B: players, pathways, perspectives. *Oncogene* **25**, 6680–6684 (2006).
- Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).

- Jensen, L. J. *et al.* STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**, D412–D416 (2009).
- Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014).
- Lim, C. P. & Cao, X. Structure, function, and regulation of STAT proteins. *Mol Biosyst* **2**, 536–550 (2006).
- Luscombe, N. M., Greenbaum, D. & Gerstein, M. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med* **40**, 346–358 (2001).
- Mills, C. D. Anatomy of a Discovery: M1 and M2 Macrophages. *Front Immunol* **6**, (2015).
- National Kidney and Urologic Diseases Information Clearinghouse. The kidneys and how they work. NIH Publication No. 14–3195. National Institute of Diabetes and Digestive and Kidney Diseases. (2014).
- Nivarthi, H., Friedbichler, K. & Moriggl, R. STAT5 as a Hematopoietic Master Regulator for Differentiation and Neoplasia Development. *JAK-STAT Signaling: From Basics to Disease*. Springer-Verlag Wien. 153-168 (2012).
- Ohga, S. *et al.* Thiazolidinedione ameliorates renal injury in experimental diabetic rats through anti-inflammatory effects mediated by inhibition of NF-kappaB activation. *Am. J. Physiol. Renal Physiol.* **292**, F1141–1150 (2007).

- Reidy, K., Kang, H. M., Hostetter, T. & Susztak, K. Molecular mechanisms of diabetic kidney disease. *J Clin Invest* **124**, 2333–2340 (2014).
- Satchell, S. C. & Braet, F. Glomerular endothelial cell fenestrations: an integral component of the glomerular filtration barrier. *Am J Physiol Renal Physiol* **296**, F947–F956 (2009).
- Shaw, K. & M. Cummings. *Diabetes Chronic Complications*. Wiley. 2nd Ed. (2006).
- Teglund, S. *et al.* Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. *Cell* **93**, 841–850 (1998).
- Valtin, H. & Schafer, J.A. *Renal Function*. Lippincott Williams & Wilkins. 3rd Ed. (1995).
- Wada, J. & Makino, H. Inflammation and the pathogenesis of diabetic nephropathy. *Clin. Sci.* **124**, 139–152 (2013).
- Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).