My internship will take place under the supervision of Dr. Joe Song in New Mexico State University's computer science department. Dr. Song's research group is focused on the development and application of statistical tests to identify transcriptional regulatory relationships in complex networks. The group also maps biological pathway topologies in relation to expression data and relates them to differences found within promoter regions. The group includes graduate students from the computer science, biology, and chemistry departments.

This internship will involve computational analysis of gene regulatory networks (GRNs). An initial GRN analysis will be performed, with opportunities for follow-on analyses and further development of computational methods for GRN discovery and characterization.

A GRN is a collection of regulatory proteins and the transcription promoter sites to which they bind, represented (for our convenience) as a network with promoter sites as the nodes and regulator-promoter relationships as edges[1]. Intuitively, the level of activity at a network node corresponds to the expression level of the node's gene, and a network edge describes the excitatory or inhibitory effect of a source gene's product on a target gene's promoter site. Understanding GRN structure can help us interpret gene expression data[2].

During this internship, I will investigate whether and how GRN structure changes with varying biological circumstances. Does a particular GRN perform the same function and have the same regulatory relationships everywhere it appears? Or is the network context-specific, with the edges of the network graph reconfigured depending upon variables such as time, developmental stage, or surrounding tissue?

Song *et al* have developed a statistical test, the comparative chi-square test, and have demonstrated its ability to detect "rewiring" of network relationships, given data about activation levels at network nodes in different experimental conditions[3]. The comparative chi-square test evaluates the likelihood that observed promoter and gene expression levels would appear under different experimental conditions, under the null hypothesis that the regulatory relationships along the network edges are the same under all experimental conditions.

An upcoming paper by Dutta et al[4] presents transcriptome expression levels measured in the five anatomically distinct regions of the drosophila midgut (R1-R5)[5] for four cell types. The cell types profiled are enterocytes (EC) - epithelial cells that perform digestion; enteroendocrine cells (EE) – cells that participate in hormonal signaling; enteroblasts (EB) - precursor cells that differentiate into enterocytes and enteroendocrine cells; and intestinal stem cells (ISC) which replenish the gut lining by differentiating into enteroblasts and subsequently EE and EC cells. I will apply Song et al's techniques to this data in order to determine whether known GRN's extracted from the FlyBase database[6] exhibit structural differences across regions or cell types.

At the conclusion of this internship, I expect to have greatly improved knowledge of the biological principles of gene regulation, and I will have gained extensive practical experience in applying statistical inference techniques to large expression data sets.

References Cited

1. http://younglab.wi.mit.edu/regulatory_network/

2. Modelling and analysis of gene regulatory networks. Karlebach, G and Ron Shmir, Nature Reviews Molecular Cell Biology 9, 770-780 (October 2008) | doi:10.1038/nrm2503

3. Hunting complex differential gene interaction patterns across molecular contexts Song, Joe et al, Nucl. Acids Res. (2014)doi: 10.1093/nar/gku086

4. Regional cell-specific transcriptome mapping reveals the regulatory complexity of the adult Drosophila midgut. Devanjali Dutta, Adam J Dobson, Jerome Korzelius, Christine Gläßer, Philip L Houtz, Nicolas Buchon, Bruce A Edgar. Cell Reports. (under review).

5. FlyGut: an atlas of the *Drosophila* adult midgut. http://flygut.epfl.ch/anatomy

6. FlyBase: A Database of *Drosophila* Genes and Genomes. http://flybase.org/