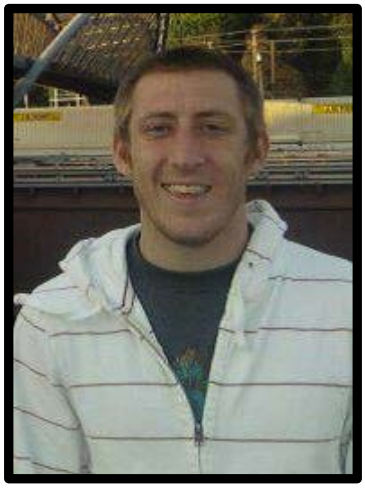
THE GRADUATE COLLEGE OF THE
UNIVERSITY OF OKLAHOMA HEALTH SCIENCES CENTER

ANNOUNCES THE FINAL EXAMINATION OF

Cory Giles

FOR THE DEFENSE OF THE DOCTOR OF PHILOSOPHY DEGREE
GRADUATE COLLEGE
Department of Biochemistry & Molecular



Wednesday, August 16, 2017 | 2:00 p.m.
Robert M. Bird Library Building, Auditorium 299

A Computational System for Automated Meta-analysis of Gene Expression Data

COMMITTEE IN CHARGE: Jonathan Wren, Ph.D., Hiroyuki Matsumoto, Ph.D., Courtney Montgomery, Ph.D., Blaine Mooers, Ph.D., Ann-Louise Olson, Ph.D., William Sonntag, Ph.D.

ABSTRACT: A number of open biological problems can potentially be answered with the data that exists now within publicly available databases. There are millions of freely available high-throughput datasets cntaining gene expression, methylation and genomic alteration data. But the data exists in an unstructured format, and processing is needed before it can be analyzed. In particular, rather than a single effort to answer one question, a more generic system is needed to answer question regarding as many identifiable aspects of the data as possible.

We constructed an automated computational system focused on preprocessing, normalization and analysis of this data, one that can also test multiple machine-learning and statistical methods to identify the most effective one. It is able to extract important sample parameters such as sex, age, and tissue type, when available. These parameters not only serve as covariates, but also enable more specific questions to be asked regarding how they affect experimental data.

To enable cross-experiment comparisons, we tested several alternative preprocessing pipelines for probe collapsing, imputation, and normalization. The result of these steps is a system containing a combined database of computer-readable expression data and metadata.

Using this system, we performed an automated meta-analysis gene expression changes that occur with age in humans across a variety of transcriptional profiling platforms and experiments to determine a "transcriptional aging signature". We compare our automatically-determined signature with previous meta-analyses and show substantial overlap. We primarily found a strong upregulation in immune-related genes, specifically those related to neutrophil function, and a downregulation in components related to the translation machinery. We also demonstrate the versatility of this system by showing its ability to identify genes with tissue-specific expression patterns, to predict function of uncharacterized transcripts on the basis of coexpressed genes with known function, and to identify modules of coexpressed genes.