Oklahoma Nathan Shock Center
—of EXCELLENCE IN AGING RESEARCH—

Geroinformatics Resource Services

Bioinformatics & Statistical Analysis

The Geroinformatics Resource has a Ph.D.-level statistician (Dr. Constantin Georgescu) to assist with statistical issues, such as pre-experimental considerations (eg, power analysis) and also with post-experimental interpretation. We can handle the QC & processing of high-throughput sequencing data of various forms (e.g., RNA-seq, ATAC-seq, etc.).

Mining & Analysis of Literature Networks

We have developed large-scale text-mining software that has processed all 34 million MEDLINE records to create a weighted network of linked entities, e.g., genes, diseases, phenotypes, chemicals, lipids, metabolites, aging-related concepts, etc. (Wren et al., 2004; Wren & Garner, 2004). This network can be used to understand: 1) How two things are related to each other (in the literature), 2) What a set of things (eg, genes) has in common in PubMed, and 3) Implied relationships (what interesting, yet unpublished, connections might exist for a gene, disease or drug of interest).

References

Wren, J.D., Bekeredjian, R., Stewart, J.A., Shohet, R.V., and Garner, H.R. (2004). Knowledge discovery by automated identification and ranking of implicit relationships. *Bioinformatics* 20, 389-398.

Wren, J.D. and Garner, H.R. (2004), Shared relationship analysis: ranking set cohesion and commonalities within a literature-derived relationship network. *Bioinformatics* 20,191-198.

Wren, J.D. (2004). Extending the mutual information measure to rank inferred literature relationships. *BMC bioinformatics* 5,145.

Transcriptional Network Analysis and Gene Function Prediction

About 31% of human gene names do not appear in MEDLINE, nor do approximately 90% of human non-coding RNAs. And existing literature is heavily skewed towards well-known genes with 75% of effort focusing on 10% of genes (Edwards et al., 2011). **We can predict what individual genes do and find genes related to phenotypes/diseases of interest** (e.g., synaptic transmission, sarcopenia). We have developed software that analyzes over 75,000 human microarray experiments to identify genes correlated across heterogeneous conditions (Dozmorov & Wren, 2011a,b; Wren, 2009) that can then be linked back to what they have in common in MEDLINE. Phenotype prediction has been 85% accurate in wet-lab experiments (48 positive out of 56 tested) and several formerly uncharacterized genes successfully validated (Towner et al, 2013a; Towner et al., 2013b; Clemmensen et al., 2012; Lupu et al., 2011; Daum et al., 2009).

Miscellaneous

Genomic analysis: We have also developed software to identify genomic commonalities among a set of genomic regions, such as might be identified in sequencing experiments (eg, CHiP-seq) (Dozmorov et al., 2012).

Epigenetic clocks: We have developed and applied epigenetic clocks for experimental problems.

Artificial Intelligence and Machine Learning: We use AI/ML in a number of different experimental contexts such as deep neural nets (DNNs) and conditional variational autoencoders (CVAEs) to predict, classify and for generative applications.

Services

Service	Cost per Unit
Pre-experimental Stastical analysis	$200/hour
Post-experimental Stastical analysis	$200/hour
Sequencing data analysis	$200/hour
Uncharacterized gene function	$200/hour
Literature-based commonality analysis	$200/hour
Consultation on Artificial Intelligence and Machine Learning (AI/ML) approaches	$200/hour
Help with development of Artificial Intelligence and Machine Learning (AI/ML) approaches	$200/hour