ABSTRACT
DEVELOPMENT OF A METHOD TO IDENTIFY COMPENSATORY PATHWAYS USING HEURISTIC GENOMIC EPIDEMIOLOGY

KARL D. STAMM
Marquette University, 2016

Explaining the mechanism behind congenital disease involves two phases, collecting and analyzing genetic data, then interpretation in the context of biological systems. The objective of this work is to develop a method of integrating complementary datasets surrounding any single biological process, with the goal of presenting the response to a signal in terms of a set of downstream biological effects. This dissertation specifically tests the hypothesis that heuristic methods overlaid with domain expertise can direct research towards relevant systems-level signals underlying complex disease. To this end, I developed a software toolkit named GSEPD that can visualize multidimensional genetic expression to identify the biologically relevant gene sets that are altered in response to an experimental condition.

This dissertation highlights a problem facing the medical research community, and shows how computational sciences can help. By bringing large datasets together, a new analytical and software method was produced that helps unravel complicated experimental and biological data.

The dissertation shows four coauthored studies where the experts in their field have desired to annotate functional significance to a gene-centric experiment. Using GSEPD to show inherently high dimensional data as a simple colored graph, a subspace vector projection directly calculated how each sample behaves like each test condition. The end-user medical researcher understands their data as a series of somewhat-independent subsystems, and GSEPD provides the most understandable dimensionality reduction for high throughput experiments of limited sample size. Gene Ontology analyses are accessible on a sample-to-sample level, and this work aims to highlight not just the expected pathways, but the many annotated results that are currently going unseen and unappreciated in vast online databases.