# Kirk Gosik 4188 Roundtop Road Elizabethtown, PA 17022 (717) 329-2455 [kdgosik@gmail.com](mailto:kdgosik@gmail.com)

<https://github.com/kdgosik> ; <https://sites.google.com/view/kdgosik>

As a PhD candidate at the Penn State College of Medicine, my research has focused on ultra-high dimensional data, where p >> n, specifically with model selection and epistatic screening for the purposes of genetic mappings. I have applied the screening and modeling techniques on both theoretical and real datasets in aims to find a more complete understanding of the relationship between the genotypes and phenotypes of interest.

In my current work I have applied and extended a form of forward selection that dynamically includes interaction effects throughout the selection process. By treating genetic effects as a biological system, several issues arise from the large amount of predictors that could be considered in the model. Along with the main effects of each genetic effect and then extending to include epistatic effects of interaction between genetic effects grows the set of possibilities to an even greater extent. In my dissertation work I incorporated model selection techniques to help screen for significant effects that would create the network of effects which would impact a phenotype.

My first paper focused on a single static phenotype and using main genetic effects and second order epistatic effects as possible predictors. Using the selection procedure, it would screen through the effects and determine which effects were important to the phenotype of interest. Extending this slightly the next paper focused on including high-order interaction effects to show the epistatic impact on the phenotype. The phenotype was also single growth parameters in hopes to identify important predictors of areas of the functional part of a phenotype. Currently I am extending this further to include the entire growth function as a phenotype while incorporating the selection procedure with higher-order epistatic effects. I hope to submit my third paper in late spring around this topic.

Continuing on, my goal would be to work with large datasets and incorporate these types of statistical methods and machine learning techniques to aid in analysis and gain larger insights into the genetic/epigenetic architecture of biological systems. On top of the importance of a functional component to the phenotype, considering other types of multivariate responses would be interesting to study in context of such a system.

Incorporating different level of omics data and the challenges that arise with such complicated and large datasets has interested me throughout my PhD work. Translating such a complex system to actual usable information that can shared in order to prevent and fight disease would be ideal research for me. This type of research would need both statistical methods as well as machine/deep learning techniques to handle the magnitude of the problem.

In general, dynamic biological systems and the interactions between the complexities of a system is an area where I would like to focus on for my future research. Considering genetic and environmental interactions and to find underlying control mechanisms of phenotypes is a very important area. The results could help in furthering scientific research and enable deeper levels of understanding. Implementing and improving statistical methods to handle such analyses will help to continue and expand a systems mapping to build on previous knowledge and complement new and ongoing research avenues.