Specific Aims

A better understanding of the population genetics processes and parameters that give rise to complex diseases and shape their genetic makeup is important to both human and evolutionary genetics. In particular, the genetic architecture of complex diseases – the number of mutations contributing to disease risk and the distributions of their allele frequencies and effect sizes (or odds ratios) – largely determines the power to map variants that contribute to disease risk, and misguided/incorrect expectations about genetic architecture account for many of the growing pains of genome wide association studies (GWAS) in humans (e.g. the missing heritability problem). Understanding how genetic architectures are shaped would lead to more accurate expectations, and in doing so would allow improvements on the design of mapping approaches and statistical models for phenotypic prediction (e.g. in personalized medicine). In addition, relating the findings from GWAS of human diseases with the population genetics processes that give rise to them, offers an unprecedented opportunity to learn about the processes that maintain genetic variation in complex, deleterious phenotypes.

To this end, I will develop models of the population genetic processes governing the evolution of complex diseases. I will rely on these models to develop statistical tools to infer the parameters that shape genetic architecture from the findings of GWAS.

Specific Aim 1: Develop Models Relating Population Genetic Processes with the Architecture of Complex Diseases I will extend a preliminary model of the population genetics of complex disease that I have developed in order to understand how natural selection, changes to the human environment (e.g. diet and lifestyle), pleiotropy, and demography affect the genetic architecture of complex diseases. Specifically, this modeling will elucidate how observable quantities, such as the number of segregating variants affecting disease risk, their frequencies and effect sizes, depend on the mutational target size, the strength of the selection against the disease, and the pleiotropic effects of variants.

Specific Aim 2: Inference of Model Parameters from GWAS of Complex Diseases Using the models developed in Aim 1, I will develop a likelihood based approach to estimate the evolutionary parameters governing disease susceptibility, e.g. the mutational target size, the fitness cost and prevalence of the disease, and the extent to which selection on variants affecting disease risk arises from their pleiotropic effects on other traits. I will begin with methods that rely on the frequencies and effect sizes of genome wide significant associations from GWAS, and will later extend them to use information about how genetic variance is distributed among alleles with effect too small to be detected in GWAS (which can be inferred by other means). I will apply the inference machinery to datasets from GWAS for at least 10 complex disease. In addition to estimating parameters based on a given model, I will compare the fit among models, in order to learn, for example, whether pleiotropic effects on disease associated alleles are primarily derived from effects on other diseases or on continuous traits, or whether disease prevalence has been substantially impacted by changes to the human environment. The inference tools I develop will be made available in a user-friendly software package.