Specific Aims

An understanding of the genetic architecture of complex disease (including e.g. the number of mutations contributing to disease risk and the distributions of their allele frequencies and effect sizes or odds ratios) is of particular importance in both human and evolutionary genetics. Such factors are major determinants of our ability to effectively map individual disease variants, and indeed many of the growing pains of the genome wide association (GWAS) approach (e.g. the missing heritability problem) stemmed from a mismatch between researchers expectations and the true nature of the genetic architecture of complex disease. A better understanding of the genetic architecture of disease would be valuable for guiding future study design, and for shaping intuition about the performance of polygenic prediction approaches when applied across evolutionarily diverged groups, and for the performance of multi-ethnic GWAS. Moreover, human disease GWAS currently represent one of the best opportunities to study the evolution of deleterious phenotypes and to understand how qunatitative genetic variation is maintained within populations.

To this end, I will develop models of the population genetic processes governing complex disease evolution. I will use these models to develop tools to infer the parameters that shape the genetic architecture of a range of individual diseases, and to make forecasts about how architecture will impact efforts to extend the GWAS approach outside of European ancestry populations.

Specific Aim 1: Development models of the evolution of complex disease. I will extend our preliminary model of complex disease evolution to understand how factors such as natural selection, pleiotropy, and demography affect the genetic architecture of genetic disease risk. The results of this aim will be an understanding of how observable factors such as the number of segregating variants that contribute to disease risk, their frequencies and their effect sizes depend on the mutational target size, the strength of the selection against the disease, and the nature of the pleiotropic effects of disease mutations.

Specific Aim 2: Statistical methods to infer the underlying parameters governing the evolution of complex disease. The effect sizes (odds ratios) and frequencies of GWAS variants contain information about the about the nature of pleiotropic effects and the selection coefficients they experience. Using the models developed in Aim 1, I will design likelihood based inference machinery to infer the parameters governing the evolution of disease susceptibility from GWAS data. In addition to taking significant GWAS hits as input, I will expand my inference machinery to use information about how heritable variation is distributed among minor allele frequency bins (which can be inferred by other means). This represents a rich source of information which will help constrain the space of possible models that are compatible with the architecture of a given disease. I will apply the inference tools I develop to at least 10 GWAS datasets for complex disease. For each disease, I will be able to learn about the size of the mutational target (i.e. the number of bases in the genome where a mutation would affect disease liability), the ancestral fitness cost and prevalence of the disease, and a measure of the extent of pleiotropy fitness effects of alleles associated with disease risk. The inference tools I develop will be made available in a user friendly software package.