AN ATLAS OF SHARED GENETIC ARCHITECTURE CONNECTING CELL BIOLOGY AND HUMAN DISEASE

While the genomes of two people are >99% identical(*1*), the genetic differences encode amazing diversity in human traits and disease susceptibility. Our understanding of the genetic architecture underlying human diversity in complex traits has been transformed by the application of genome-wide association studies (GWAS)(*2*). However, GWAS are just the first step in understanding how genetic variants contribute to disease risk by impacting genes that encode components of cellular physiology. Elucidating this chain of causality from SNP to gene to cell biology to disease can serve to not only functionally validate the genetic association with disease but also reveal potential therapeutic targets. Therefore, there is a need for approaches that can facilitate identification of the cellular pathways regulated by human genetic variants associated with disease.

Our long-term goal is to develop a cohesive understanding of how functional genetic differences impact molecular and cellular phenotypes to influence disease. **The objectives of this application are to systematically reveal the shared genetic architecture connecting cellular responses to disease susceptibility and to develop a database to facilitate hypothesis generation from these data.** We have the experimental and computational tools in place to carry this out successfully. For GWAS of cellular traits, we developed and extensively validated a cellular GWAS platform called Hi-HOST (High-throughput human in vitro susceptibility testing) that uses pathogens as probes to stimulate fundamental cellular pathways(*3-6*). Hi-HOST combines precise measurement of cellular phenotypes in lymphoblastoid cell lines (LCLs) from nearly a thousand people with genome-wide association using 15 million genetic variants to identify genetic differences that underlie the phenotypic variation. Building on the Hi-HOST platform, we recently completed data acquisition and genome-wide association for the Hi-HOST Phenome Project (H2P2), encompassing cellular responses of infectivity and replication, cytokine levels, and host cell death using 9 different pathogens and 148 cellular traits. Dozens of SNPs pass genome-wide significance for cellular traits induced by *Yersinia pestis*, *Chlamydia trachomatis,* and *Salmonella* species and have been incorporated into established workflows in the lab to validate their importance and determine mechanism. We will integrate this already collected H2P2 cellular GWAS dataset with human disease GWAS from the NHGRI-EBI catalog. Our central hypothesis is that the SNPs associated with each cellular phenotype and disease can serve as a "GWAS signature" and the similarity of these signatures can be used to connect different traits and quantify the contributions of individual cellular traits to heritability of diseases. We will apply two published methods to do this: 1) we will extend our own published method, CPAG ((*7*); Cross-Phenotype Analysis of GWAS), to calculate similarity and significance among GWAS based on overlap of associated SNPs; 2) we will use the LD-score regression approach of Bulik-Sullivann *et al.* (*8*) to partition heritability of diseases by SNPs associated with cellular traits. Thus, we are uniquely positioned to integrate cellular and disease GWAS to create **a re-interpretation of the human genome through the lens of cell biology.**

Aim 1. Characterize the shared genetic architecture between cellular phenotypes and human diseases through systematic identification of shared SNPs and partitioning of heritability. We hypothesize that similarity of GWAS signatures can be used to identify and quantify the contribution of cellular traits to human diseases. We will identify shared SNPs among 148 H2P2 cellular traits and 1123 human diseases/traits from the NHGRI-EBI GWAS catalog and identify overlap greater than expected by chance using CPAG. Furthermore, we will partition the heritability of human diseases to quantify the contribution of individual H2P2 cellular traits using LD-score regression. The revealed cell biology-disease connections will generate hypotheses that can be tested in the H2P2 LCLs, other cell types, animal models, and clinical samples.

Aim 2. Create a publicly available web database for exploring human genetic diversity in pathogen-induced cell biological phenotypes and its impact on human disease. The power of H2P2 and the connections between cell biology and disease from Aim 1 can only be fully realized if the cell biology, infectious disease, and human genetics communities can easily explore the data. We will build a relational web database (the Atlas of Integrated GWAS; AI-GWAS) to allow researchers to determine if genetic differences in their genes of interest are associated with specific traits and human diseases, and if the cell biological processes they study are regulated by human genetic variants that could point to a role in disease pathogenesis.

This project will demonstrate how genetic variation in cellular pathways play an essential role in human disease and will create a novel and useful community resource for generating testable hypotheses of how cellular pathways impact disease. The project leverages existing datasets into a hypothesis generating engine for researchers looking to explore new diagnostic and therapeutic possibilities.