Statement of Research Interests

Steven J. Schrodi

RESEARCH BACKGROUND

Scientific progress results from a robust interplay between theoretical models, well-designed experimental studies, and statistical methods. I have actively pursued research in these three areas of human genetics. Much of my experimental research has been devoted to mapping genes underlying complex diseases with an emphasis on systemic inflammatory diseases. These studies have resulted in the discovery of specific variants involved in the molecular pathogenesis of those diseases, both verifying known disease pathways and revealing novel pathways. My work on statistical genetics methods has aimed to provide robust, accurate assessment of disease association using high-throughput and high-density genotype data. Lastly, I have also concentrated on the development of disease genetics theory to better understand the patterns of genetic variation and results expected in Mendelian, oligogenic and complex diseases for the purpose of designing more powerful methods to interrogate the genetics of these phenotypes.

Temporal Variation in DNA Substitution Processes. Early in my career, I investigated theoretical models in population genetics and molecular evolution where I developed a novel method for testing competing models of DNA substitution processes through measuring temporal patterns of DNA substitution variation. Applying this to mammalian protein-coding sequence data, I discovered that leading models of molecular evolution were rejected in favor of models where selection coefficients vary slowly over time.

First Exome-wide Association Scan for a Common Disease: Discovery of *PTPN22* and *TRAF1* Rheumatoid Arthritis Susceptibility Genes. In 2003, I led the design, analysis and interpreted results from the first large-scale, exome-wide SNP association scan of any disease using 30,000 putatively functional coding variants. This landmark study was the subject of *A Machine to Make a Future*, 2005, by Rabinow and Dan-Cohen, published by Princeton University Press. I tested initial findings in a replication sample set of severe rheumatoid arthritis. The study resulted in the discovery of the R620W polymorphism in the protein tyrosine phosphatase, *PTPN22*, being strongly correlated with RA susceptibility. The 620W allele was subsequently found to confer profound effects on T-cell activation, B-cell pruning, NK cell stimulation, and impact numerous other innate and adaptive immune responses. This work was deemed one of the top 10 arthritis advances of 2004 by the Arthritis Foundation. Further, I led a fine-mapping effort as part of the same study which discovered *TRAF1* haplotypes as critically important RA susceptibility alleles. I was placed as chief architect for all Applied Biosystems and Celera fine mapping studies. I was awarded two United States Patents and have eight pending United States Patent Applications describing this work.

First Large-Scale Genetics Association Scan for Psoriasis: Discovery of *IL23R*, *IL12B* and *IL13* Psoriasis Susceptibility Genes. Starting in 2005, I designed, managed and analyzed the first exome-wide association scan of psoriasis. I developed a novel, pooled, multi-staged experimental design to interrogate 30,000 putatively functional coding variants to study psoriasis etiology. The study confirmed the *IL12B*-association with psoriasis and was the first investigation to discover the involvement of *IL23R* variants in disease. The findings solidified the view that Th17 signaling plays a fundamental role in autoinflammatory conditions. In addition, the study discovered polymorphisms segregating at *IL13* playing a role in psoriasis-predisposition. The *IL12B/IL23R* findings provided evidence supporting the use of anti-IL-23 and anti-IL-17 monoclonal antibodies as targeted therapies for autoinflammatory diseases. I was awarded five United States Patents and have 13 pending United States Patent Applications describing these psoriasis and autoinflammatory disease results.

Statistical Genetics. I have developed several novel statistical genetics methods and approaches to analyzing human genetics data. In 2000, I was recruited by Dr. Ray White to DNA Sciences where I developed and led a very large Monte Carlo simulation study involving several scientific programmers and a genetic epidemiologist to simulate disease genetics in extended kinships in an effort to inform the development of powerful mapping methods in homogeneous populations. Over the past 18 years, I have worked on methods of selecting tagging SNPs, TDT statistical approaches, Bayesian hypothesis testing, a Bayesian estimator for the prevalence of rare Mendelian diseases using population-based deep sequencing data, software for next-generation sequence analysis, and multiple testing approaches.

Disease Genetics Theory and Prediction of Disease Traits. My research on theoretical models of disease genetics has shown how LD with a causal site varies with mode of inheritance, including a mathematical formulation for precisely how disease association statistics decays as LD declines from a causal site. My colleagues and I have shown the utility of this work for developing new fine mapping approaches. Additionally, we have applied machine learning techniques to utilize molecular markers for disease prognosis and information theory metrics for characterizing the predictive capacity of such models.

CURRENT RESEARCH

My research program at the Center for Human Genetics at the Marshfield Clinic Research Institute is composed of several studies designed to elucidate disease genetics. Several of these studies have taken the approach of using cellular, molecular, or imaging intermediate phenotypes as an effort to better refine the relationship between genetics and disease traits. My studies have also developed and employed novel genetic approaches such as exome-wide, gene-based

recessive diplotype scans (single site recessive effects and compound heterozygosity effects) and genome-wide shared chromosomal regions among disease-affected individuals to uncover new susceptibility genes. My laboratory is currently conducting the following studies:

- A UW-Madison ICTR-funded study (PI: Schrodi) to identify genetic variants underlying variability in the IL-23/IL-17 axis cytokines and develop machine learning and probabilistic methods for prognosis and classification of dysfunctional Th17 activity.
- A Rheumatology Research Foundation-funded study to probe the individualized molecular mechanisms behind axial spondyloarthritis with my UW-Madison collaborator, Professor Judy Smith (PI). This study performed targeted sequencing on genes within the IL-23/IL-17 pathway and correlated variants with cytokine expression from stimulated macrophages obtained from axial spodyloarthritis patients and controls. Thus far, we have identified novel Tnfaip3 mechanisms driving differences between these important cellular phenotypes in cases and controls.
- A BASF-funded study (PI: Schrodi) to use machine learning applied to genetics, circulating proteins, and metabolites to predict metabolic dysfunction prior to clinical diagnoses. My collaborators and I have genotyped type 2 diabetes SNPs, measured a panel of circulating metabolites, and inflammatory/metabolic proteins in samples obtained from individuals within 18 months prior to a clinical diagnosis of type 2 diabetes. A Bayesian Network using these molecular features attained an AUC of 0.91 for prediction of future T2D.
- A pilot study (PI: Schrodi) to conduct a high-throughput serological screen for antigens that prospectively predict the onset of autoimmune disease driven by aberrant antibody response to pathogen epitopes within the context of specific class II HLA alleles.
- An MCRI-funded study (PI: Schrodi) with my Marquette University collaborator, Professor Mehdi Maadooliat, to use exome-wide genotype data to map genes in several diseases using shared chromosomal regions within distantly related samples from the Central Wisconsin population (extended kinship). This work has resulted in the identification of an excess of chromosomal sharing in seropositive rheumatoid arthritis patients in a region containing the gene encoding for the follistatin-like protein 1. This approach has also revealed an early B-cell protein for relapsing-remitting multiple sclerosis. A user friendly software package conducting these analyses is being developed.
- An MCRI-funded study (PI: Schrodi) to map disease genes in iron overload disorders, obesity, type 2 diabetes, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, premature myocardial infarction, and obsessive compulsive disorder using a novel approach of gene-based recessive diplotype analysis applied to exome-wide genotype data. This study has resulted in the discovery of a novel mechanism for maintenance of iron homeostasis, the fibroblast growth factor 6, which regulates hepcidin and modulates the uptake of ferrous iron in hepatocytes. Significant novel findings have also been made in obesity, type 2 diabetes, premature myocardial infarction and rheumatoid arthritis.
- The statistical genetics component for an NIH-funded study to investigate the sequencing genetics of neuroimaging data (PI: Professor Janet Lainhart, Waisman Center)
- The development of Bayesian and frequentist methods to address multiplicity in genome-wide studies. My work demonstrated a connection between order statistics and experiment-wise Fisherian approaches to multiple hypothesis testing which resulted in a generalization of experiment-wise hypothesis testing. My work with Professors Maadooliat and Bansal has presented an Empirical Bayesian approach to multiplicity where priors on left and right alternative models are allowed to vary.
- A theoretical study (PI: Schrodi) unifying the mathematical description of the decay of disease association with declining linkage disequilibrium from a causal site. We are currently using this work to develop fine-mapping approaches.

I would like to emphasize that the large majority of my work is highly collaborative and I would like to continue with my current collaborations and build new synergistic collaborations at UW-Madison.

Proposed Research Program

While inflammation is essential for effective response to infection, general immunocompetence and overall health, diverse studies have demonstrated that chronic dysfunction of inflammatory pathways and aberrant immune tolerance contribute to numerous pathologies. Our understanding of the molecular processes that lead to pathogenic inflammation is poor. Large-scale genetic association scans have provided some insight into the genes and pathways responsible for immune dysfunction in inflammatory diseases, but much of the heritability of these diseases remains enigmatic. This state of affairs presents significant hurdles for the advancement of genomic medicine. There are several substantial impediments to uncovering the etiological mechanisms of these impactful diseases including (i) standard analyses applied to GWAS and population-based sequencing studies of complex disease susceptibility carry low power to detect disease genes under complicated, but nonetheless realistic modes of inheritance, (ii) the vast reservoir of rare variants in humans, and (iii) determination of variant function in the context of specific diseases.

The discovery of new genetics, epigenetics and environmental insults that conspire to generate these disorders will enable a more complete understanding of these disorders at the molecular level and motivate the development of targeted therapies and clinical treatments. **The overarching goal of my research will be to identify the specific genetic variants, genes and functional motifs that give rise to systemic inflammatory diseases through the combined efforts of human genetics experiments, disease genetics theory and novel statistical genetics methods.** I intend to conduct studies that 1) will reveal novel genes underlying seropositive rheumatoid arthritis, ankylosing spondylitis, and relapsing-remitting multiple sclerosis using my approaches of gene-based, exome-wide recessive diplotype scans and shared chromosomal regions, 2) use a combination of genetics, cytokine profiling, and antibody reactivity profiling to predict disease occurrence in these three systemic inflammatory diseases using a multi-omics approach with Bayesian networks and other machine learning methods, 3) perform a large-scale human antigen and pathogen epitope/antibody screen on plasma from these diseases stratified by genotypes obtained from targeted sequencing on class I and class II HLA genes, and 4) continue my research in statistical genetics and disease genetics theory. Samples to be used for the experimental portion of my research program will be obtained from the Marshfield Clinic Personalized Medicine Research Project (PMRP) through the ICTR collaboration. I have eight years of experience using these samples. Additional sample sets for RA and AS will be interrogated as part of ongoing collaborations with Dr. Dongyi He, Guanghua Rheumatology Hospital; Professor Minghua Wang from Soochow University; and Professor Jiucun Wang, Fudan University.

Disease Gene Discovery using Novel Approaches

Analytic and Monte Carlo-based power calculations clearly show that typical analysis methods used in GWAS and NGS-based association studies are underpowered to detect recessive diplotype effects (carrying at least one pathogenic variant on each homolog at a gene. i.e., combination of compound heterozygosity and single site recessive effects). However, rare missense, nonsense, splice site, and UTR variants conferring functional effects are not uncommon. The absence of at least one copy of a properly functioning gene has been shown to generate disease in numerous Mendelian disorders, oligogenic diseases, and many complex diseases, but this mode of inheritance has not been examined in RA, AS, or MS. We have recently discovered a novel mechanism of iron homeostasis through fibroblast growth factor regulation of hepcidin using this approach. I intend to interrogate seropositive RA, AS and MS. Additionally, I propose to utilize the extended kinship structure of the PMRP to map regions housing susceptibility alleles through shared chromosomal regions, identical by descent. This design borrows both linkage and association signal to identify causal regions and has been successful in a pilot study of RA where the gene encoding for follistatin-related protein 1 has been identified by my lab and my colleague, Dr. Mehdi Maadooliat.

Multi-Omics Prediction of Early Disease Onset

As the PMRP has obtained DNA, plasma and sera samples ~14 years ago, a substantial fraction of individuals have since transitioned to a clinical diagnosis of RA, AS or MS. I propose measuring genetics, inflammatory cytokines, and antibody profiles in these samples and controls for the purpose of predicting disease diagnosis using molecular markers. My lab has used this approach to predict type 2 diabetes using a classifier developed on genetic, metabolite and protein features yielding a cross validation averaged AUC of 0.91 in a test set. Given that early treatment of autoimmune and autoinflammatory conditions is efficacious in reducing severity, such classifiers may have high clinical utility.

Extensive Antigen/Antibody Screens in HLA Genotype Classes

Aberrant antigen presentation and reduced pruning of B-cell clones is posited to be key mechanisms for autoimmune pathogenesis. Further, there is accumulating evidence for nonspecific immune response to microbial epitopes as autoreactivity triggering mechanisms. Professor Mark Craven (UW-Madison) and I are developing a study to interrogate thousands of human and microbial antigens for reactivity with circulating antibodies in disease-affected individuals. Stratifying these results by Class I and Class II HLA genotypes is hypothesized to refine any antibody signal in these data.

Statistical Genetics of Fine Mapping and Disease Genetics Architecture

Concurrently, I intend to continue my work on statistical genetics methods and theoretical disease genetics. Within the area of statistical methods, I will advance my investigation of fine mapping approaches. The era of GWAS has produced a multitude of highly significant, replicated genetic markers for complex diseases. However, the set of functional variants driving these results remains largely enigmatic. Statistical genetics analyses can be used to aid in the refinement of disease associated regions by placing a posterior probability of causal involvement in disease susceptibility for each variant in a region. My colleagues and I are currently developing a variety of techniques to obtain these posterior probabilities utilizing permutation and resampling routines and the expected patterns of linkage disequilibrium under disease models. Further, models of genetic architecture of disease are understudied in human genetics. I therefore intend to start a large-scale Monte Carlo-based research program to simulate the effects of various models of disease genetics architecture. This work will use population genetics of standing variation in humans, a variable number of causal loci, effects of *de novo* mutations, a variable number of alleles segregating at each locus, variable penetrance parameters for diplotypes at each locus, and interaction effects between loci. Results from this simulation will be compared to known data from linkage studies and GWAS to focus on plausible models. This work will enable the design and implementation of more powerful methods to uncover disease-predisposing genes.