PI: **SCHRODI, STEVEN Jon**	Title: Multi-Omics Prediction of Eight Common Diseases in the Central Wisconsin Population	
Received: 07/03/2018	FOA: PA18-484 Clinical Trial:Not Allowed	Council: 01/2019
Competition ID: FORMS-E	FOA Title: NIH Research Project Grant (Parent R01 Clinical Trial Not Allowed)	
1 R01 GM129598-01A1	Dual: AI	Accession Number: 4193576
IPF: 4843101	Organization: MARSHFIELD CLINIC RESEARCH FOUNDATION	
Former Number:	Department:	
IRG/SRG: GHD	AIDS: N	Expedited: N
Subtotal Direct Costs (excludes consortium F&A) Year 1: 250,000 Year 2: 250,000 Year 3: 225,000 Year 4: 225,000 Year 5: 225,000	Animals: N Humans: Y Clinical Trial: N Current HS Code: 20 HESC: N	New Investigator: Y Early Stage Investigator: N
Senior/Key Personnel:	*Organization:*	*Role Category:*
Steven Schrodi Ph.D.	Marshfield Clinic Research Institute	PD/PI
Shicheng Guo Ph.D.	Marshfield Clinic Research Institute	Post Doctoral
Mehdi Maadooliat Ph.D.	Marquette University	Co-Investigator
Judith Smith M.D.	University of Wisconsin - Madison	Consultant
Krista Trappl-Kimmons Ph.D.	Antigen Discovery, Inc.	Co-Investigator
Xiaowu Liang Ph.D.	Antigen Discovery, Inc.	Co-Investigator

OMB Number: 4040-0001
Expiration Date: 10/31/2019

APPLICATION FOR FEDERAL ASSISTANCE
SF 424 (R&R)

3. DATE RECEIVED BY STATE	**State Application Identifier**

1. TYPE OF SUBMISSION*

❍ Pre-application ❍ Application ● Changed/Corrected Application

4.a. Federal Identifier
GM129598

b. Agency Routing Number

2. DATE SUBMITTED | **Application Identifier**

c. Previous Grants.gov Tracking Number
GRANT12497653

5. APPLICANT INFORMATION **Organizational DUNS*:** 0747760300000

Legal Name*: Marshfield Clinic Research Institute
Department:
Division:
Street1*: 1000 North Oak Avenue
Street2:
City*: Marshfield
County:
State*: WI: Wisconsin
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 54449-5790

Person to be contacted on matters involving this application
Prefix: Ms. First Name*: Jordon Middle Name: Last Name*: Ott Suffix:
Position/Title: Director, Office of Research Support Services
Street1*: 1000 North Oak Avenue
Street2:
City*: Marshfield
County:
State*: WI: Wisconsin
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 54449-5790
Phone Number*: 715-387-9192 Fax Number: Email: sponsored.programs@marshfieldresearch.org

6. EMPLOYER IDENTIFICATION NUMBER *(EIN) or (TIN)** 39-0452970

7. TYPE OF APPLICANT* M: Nonprofit with 501C3 IRS Status (Other than Institution of Higher Education)

Other (Specify):
Small Business Organization Type ❍ Women Owned ❍ Socially and Economically Disadvantaged

8. TYPE OF APPLICATION*
❍ New ● Resubmission
❍ Renewal ❍ Continuation ❍ Revision

If Revision, mark appropriate box(es).
❍ A. Increase Award ❍ B. Decrease Award ❍ C. Increase Duration
❍ D. Decrease Duration ❍ E. Other *(specify)* :

Is this application being submitted to other agencies?* ❍Yes ●No What other Agencies?

9. NAME OF FEDERAL AGENCY*
National Institutes of Health

10. CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER
TITLE:

11. DESCRIPTIVE TITLE OF APPLICANT'S PROJECT*
Multi-Omics Prediction of Eight Common Diseases in the Central Wisconsin Population

12. PROPOSED PROJECT
Start Date* Ending Date*
05/01/2019 04/30/2024

13. CONGRESSIONAL DISTRICTS OF APPLICANT
WI-007

SF 424 (R&R) APPLICATION FOR FEDERAL ASSISTANCE

14. PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR CONTACT INFORMATION

Prefix: Dr. First Name*: Steven Middle Name: Last Name*: Schrodi Suffix: Ph.D.
Position/Title: Associate Research Scientist
Organization Name*: Marshfield Clinic Research Institute
Department:
Division:
Street1*: 1000 North Oak Avenue
Street2:
City*: Marshfield
County:
State*: WI: Wisconsin
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 54449-5790
Phone Number*: 715-221-6443 Fax Number: Email*: schrodi.steven@marshfieldresearch.org

15. ESTIMATED PROJECT FUNDING

a. Total Federal Funds Requested*	$1,854,536.00
b. Total Non-Federal Funds*	$0.00
c. Total Federal & Non-Federal Funds*	$1,854,536.00
d. Estimated Program Income*	$0.00

16.IS APPLICATION SUBJECT TO REVIEW BY STATE EXECUTIVE ORDER 12372 PROCESS?*

a. YES ○ THIS PREAPPLICATION/APPLICATION WAS MADE AVAILABLE TO THE STATE EXECUTIVE ORDER 12372 PROCESS FOR REVIEW ON:
DATE:

b. NO ● PROGRAM IS NOT COVERED BY E.O. 12372; OR
○ PROGRAM HAS NOT BEEN SELECTED BY STATE FOR REVIEW

17. By signing this application, I certify (1) to the statements contained in the list of certifications* and (2) that the statements herein are true, complete and accurate to the best of my knowledge. I also provide the required assurances * and agree to comply with any resulting terms if I accept an award. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. (U.S. Code, Title 18, Section 1001)

● I agree*

** The list of certifications and assurances, or an Internet site where you may obtain this list, is contained in the announcement or agency specific instructions.*

18. SFLLL or OTHER EXPLANATORY DOCUMENTATION File Name:

19. AUTHORIZED REPRESENTATIVE

Prefix: Dr. First Name*: Amit Middle Name: Last Name*: Acharya Suffix: Ph.D.
Position/Title*: Executive Director
Organization Name*: Marshfield Clinic Research Institute
Department:
Division:
Street1*: 1000 North Oak Avenue
Street2:
City*: Marshfield
County:
State*: WI: Wisconsin
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 54449-5790
Phone Number*: 715-221-6423 Fax Number: 715-389-3131 Email*: acharya.amit@marshfieldresearch.org

Signature of Authorized Representative*
Brian Nikolai

Date Signed*
07/03/2018

20. PRE-APPLICATION File Name:

21. COVER LETTER ATTACHMENT File Name:1252-CoverLetter_Schrodi_RO1_02Jul2018.pdf

424 R&R and PHS-398 Specific
Table Of Contents

SF 424 R&R Cover Page....1
Table of Contents....3
Performance Sites....4
Research & Related Other Project Information....5
Project Summary/Abstract(Description)....6
Project Narrative....7
Facilities & Other Resources....8
Research & Related Senior/Key Person....13
PHS398 Cover Page Supplement....40
PHS 398 Modular Budget....42
Personnel Justification....48
Consortium Justification....51
PHS 398 Research Plan....53
Introduction to Application....54
Specific Aims....55
Research Strategy....56
PHS Human Subjects and Clinical Trials Information....68
Study 1: Multi-Omics Prediction of Eight Common Diseases in the Central Wisconsin Population....70
Inclusion Enrollment Reports....76
Bibliography & References Cited....83
Consortium/Contractual Arrangements....88
Letters of Support....89
Resource Sharing Plan(s)....91

OMB Number: 4040-0010
Expiration Date: 10/31/2019

Project/Performance Site Location(s)

Project/Performance Site Primary Location

❍ I am submitting an application as an individual, and not on behalf of a company, state, local or tribal government, academia, or other type of organization.

Organization Name: Marshfield Clinic Research Institute
Duns Number: 0747760300000
Street1*: 1000 North Oak Avenue
Street2:
City*: Marshfield
County:
State*: WI: Wisconsin
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 54449-5790
Project/Performance Site Congressional District*: WI-007

Project/Performance Site Location 1

❍ I am submitting an application as an individual, and not on behalf of a company, state, local or tribal government, academia, or other type of organization.

Organization Name: Marquette University
DUNS Number: 0469296210000
Street1*: 1324 West Wisconsin Avenue
Street2: Holthusen Hall, Suite 341
City*: Milwaukee
County:
State*: WI: Wisconsin
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 53233-0000
Project/Performance Site Congressional District*: WI-004

Additional Location(s) File Name:

Funding Opportunity Number: PA-18-484. Received Date: 2018-07-03T18:44:57.000-04:00

OMB Number: 4040-0001
Expiration Date: 10/31/2019

RESEARCH & RELATED Other Project Information

1. Are Human Subjects Involved?* ● **Yes** ❍ **No**

1.a. If YES to Human Subjects

Is the Project Exempt from Federal regulations? ❍ Yes ● No

If YES, check appropriate exemption number: _ 1 _ 2 _ 3 _ 4 _ 5 _ 6 _ 7 _ 8

If NO, is the IRB review Pending? ● **Yes** ❍ **No**

IRB Approval Date:

Human Subject Assurance Number 00000873

2. Are Vertebrate Animals Used?* ❍ **Yes** ● **No**

2.a. If YES to Vertebrate Animals

Is the IACUC review Pending? ❍ Yes ❍ No

IACUC Approval Date:

Animal Welfare Assurance Number

3. Is proprietary/privileged information included in the application?* ❍ **Yes** ● **No**

4.a. Does this project have an actual or potential impact - positive or negative - on the environment?* ❍ **Yes** ● **No**

4.b. If yes, please explain:

4.c. If this project has an actual or potential impact on the environment, has an exemption been authorized or an environmental assessment (EA) or environmental impact statement (EIS) been performed? ❍ Yes ❍ No

4.d. If yes, please explain:

5. Is the research performance site designated, or eligible to be designated, as a historic place?* ❍ **Yes** ● **No**

5.a. If yes, please explain:

6. Does this project involve activities outside the United States or partnership with international collaborators?* ❍ **Yes** ● **No**

6.a. If yes, identify countries:

6.b. Optional Explanation:

	Filename
7. Project Summary/Abstract*	1242-PROJECT_SUMMARY_Schrodi_RO1_02Jul2018.pdf
8. Project Narrative*	1243-PROJECT_NARRATIVE_Schrodi_RO1_02Jul2018.pdf
9. Bibliography & References Cited	1244-Bibliography References_Schrodi_02Jul2018.pdf
10.Facilities & Other Resources	1245-Facilities_Resources_Schrodi_RO1_02Jul2018.pdf
11.Equipment	

Funding Opportunity Number: PA-18-484. Received Date: 2018-07-03T18:44:57.000-04:00

PROJECT SUMMARY

The early prediction of disease is key goal of precision medicine. Using broad molecular data, this study investigates the machine learning-based prediction of rheumatoid arthritis (RA), systemic lupus erythematosus, relapsing-remitting multiple sclerosis, premature myocardial infarction, chronic lymphocytic leukemia, obsessive compulsive disorder, autoimmune thyroid disease and axial spondyloarthritis. Additionally, we will test for the existence of significant molecular subgroups within each disease phenotype using unsupervised clustering approaches and a suite of advanced dimensional reduction methods. Multiple lines of evidence support the notion that numerous molecular and cellular factors drive common disease susceptibility and trajectory, including genetic variants, chronic systemic inflammation, metabolic dysfunction, compromised immune tolerance, and exposure to pathogens. We hypothesize that interrogating diverse biomarkers that play a pivotal role in these molecular processes in samples obtained from individuals prior to clinical diagnosis will provide informative features for the generation of classifiers useful for early diagnosis of the diseases studied. Currently, prediction algorithms for these diseases are inadequate for highly accurate determination of any of these diseases. We will use biobanked samples from the 20,000-individual Marshfield Clinic Personalized Medicine Research Project (PMRP) with linked longitudinal electronic medical records to study these diseases against matched controls. The Central Wisconsin population—the source population for the PMRP—is highly homogeneous, reducing potential confounding artifacts. DNAs from this biobank have been subjected to the Illumina exome beadchip, generating excellent protein-coding coverage of common polymorphisms and low frequency variants. Plasma from cases of each disease and matched controls will be interrogated for (i) 42 inflammatory cytokines and chemokines and 6 metabolic proteins using two multiplexed Luminex assays and (ii) antibodies to 800 autoantigens and pathogen epitopes on an Antigen Discovery platform (only RA and SLE). Concurrently using all of these molecular data types, we will use sure independence screening (SIS) to extract informative signals in each disease sample set. An Elastic Net is then used to select features for use in the classifiers. Bayesian networks and logistic regression with adaptive lasso will then be employed to classify each pre-disease group against controls. Averaged AUC values will be used to evaluate the predictive performance of our models. We will test the top performing RA classifier in an independent RA sample set. Additionally, to investigate molecular-based subgroups in each disease, we will use logistics PCA and biclustering approaches. Knowledge of molecular profiles that are informative for classification of diseases can be essential for understanding the pathogenesis of these diseases and associated biomarkers can serve as targets for studies of targeted therapeutics. Prediction of common diseases using a combination of exome genetics, circulating biomarkers and antibodies can have a transformative impact on disease diagnosis and enable early clinical interventions.

PROJECT NARRATIVE

The early prediction of chronic disease can enable timely intervention that can often circumvent more severe outcomes. This study uses genetic factors, proteins that can be indicative of systemic inflammation and metabolic dysfunction, and circulating antibodies to predict diseases very early in their development using machine learning algorithms. Understanding the relationship between these molecular markers in specific pathways not only may enable early diagnosis and timely treatment to remediate severe complications, but also provides insight into the causes of the diseases studied.

RESOURCES

Marshfield Healthcare System
The Marshfield medical complex consists of four components including Marshfield Clinic, Marshfield Clinic Research Institute (MCRI), Marshfield Laboratories and Marshfield Clinic Health System Hospital.

Marshfield Clinic
Marshfield Clinic employs approximately 780 physicians representing 84 different specialties and 7,227 additional staff working on the main campus in Marshfield or one of 52 regional clinics serving the population of northern, central, and western Wisconsin, and the Upper Peninsula of Michigan. The Clinic records over 3.4 million patient encounters annually. The Clinic, a not-for-profit 501(c) (3) tax-exempt organization, has developed this regional structure to address the health needs of the rural population. Marshfield Clinic's Marshfield Hospital (a 525-bed acute facility) providing primary, secondary and tertiary care for individuals regardless of financial status. Marshfield Clinic maintains a joint electronic medical record system with the Hospital, including computerized diagnostic files dating back to 1963.

Marshfield Clinic Research Institute (MCRI)
MCRI is the nonprofit research division of Marshfield Clinic. The Research Institution employs approximately 181 staff, including 30 PhD and MD scientists. The Research Institute has five research centers, which are described below. It also provides a full range of research administration and support services through its offices of Scientific Writing and Publication, Sponsored Programs and Fiscal Affairs, Core Laboratory, and Research Integrity and Protections. It maintains an Institutional Review Board, which is responsible for the protection of human research participants. The Research Institute publishes *Clinical Medicine and Research*, a nationally and internationally indexed professional medical journal. The Research Institute is part of the Select Agent Program Administered by the Centers for Disease Control and Prevention (CDC) and is one of only two private research institutes in the country participating in the CDC Laboratory Response Network. Recently, MCRI was awarded an NIH grant for participation in the All of Us precision medicine study. Scientists and clinicians at MCRI and Marshfield Clinic are currently involved in over 400 research projects the scope of which includes agriculture-related safety and health, veterinary research, food safety, oncology, infectious diseases, vaccine safety, genetic basis for disease and drug interactions, treatment-based outcomes studies, development of models for the study of disease processes, mental health research, women's health issues, pediatric health and safety and health services research.

MCRF Research Facilities
The research laboratories are located in state-of-the-art facilities at the Lawton Center (total square footage: 54,000 sq. ft.) and the Laird Center (total square footage: 52,000 sq. ft. for Laird South and 106,000 sq ft for Laird North), which are linked to the Lawton Center by a skywalk. The Lawton Center houses Health Services Research, the Biochemistry Laboratory, Cancer Genetics Laboratory, Environmental Health Laboratory, a Laboratory Commons area, a portion of the Epidemiology Research Center, and the Clinical Research Center. Together these laboratories occupy 21% of the building's available space. The Laird Center houses the National Farm Medicine Center (NFMC), Center for Human Genetics, a portion of the Clinical Research Center, the Personalized Medicine Research Project's population-based genetics research facilities, the Epidemiology Research Center, the Infectious Diseases Laboratory, the Molecular Diagnostics laboratory, and the Biomedical Informatics Research Center. Together these laboratories and Centers occupy 40% and 30% of the Laird South and Laird North available space, respectively. The Laird North building is also occupied by Marshfield Laboratories.

University of Wisconsin Institute for Clinical and Translational Research
The UW Institute for Clinical and Translational Research, funded under NIH, represents a novel partnership between UW-Madison and the Marshfield Clinic. Bringing together the strong and distinct resources of these institutions, the union enhances clinical and translational research in Wisconsin.

LABORATORY:
Marshfield Clinic Research Institute Integrated Research & Development Laboratory

The Core Laboratory, with 2,800 sq. ft. of workspace, is under the directorship of Dr. Jennifer Meece who oversees clinical research studies. This is a service laboratory staffed by a team of research scientists and technicians with broad-based expertise who can assist in the design and execution of laboratory-based testing for research projects. It is outfitted with equipment necessary to support tissue culture, bacteriology, virology, molecular microbiology, and general molecular biology research.

The Integrated Research and Development Laboratory instrumentation is capable of supporting tissue culture, bacteriology, virology, molecular microbiology, and general molecular biology, genetics, and genotyping research; including:

- Cell culture and analysis: MoFlo flow cytometer and cell sorter, incubators (standard and CO2) and microscopes (inverted and conventional, fluorescent and visible)
- Analytical equipment: Gamma counter, scintillation counters, fluorimeters, spectrophotometers, ELISA readers and electrophoresis
- Imaging: Nikon Eclipse 600 digital image capture and Metamorph software, Gel Dock 2000 electrophoretic imager and Zeiss model 9 electron microscope
- Molecular analyses: DNA sequencing capability with DNAsis and DNAStar software, LightCycler real time PCR, DNA imaging and analysis software and GeneAmp 9700 thermocycler
- Chromatography: Amino acid analyzer, GC, HPLC and LC capabilities
- Biosafety level-3 (BSL-3) containment capabilities
- Various support equipment such as autoclaves, Milli-Ro water systems, vented laminar-flow hoods, fume hoods, refrigerated and high-speed centrifuges, ultra-cold freezers, water baths, shakers, pH meters, and analytical balances
- Genotyping platforms including Affymetrix and Sequenom mass spectrophotometer
- Massively Paralleled Sequencing platform: Illumina MiSeq with Center for Human Genetics
- Single Molecule Sequencing: Oxford Nanopore MinIon platform with Dr. Schrodi's laboratory

Marshfield Labs
Marshfield Labs is the clinical laboratory serving the health care providers and patient population at Marshfield Clinic. The laboratory is a joint venture serving both the Marshfield Clinic Health System Hospital, a 525-bed trauma I facility, and Marshfield Clinic, a 780 physician primary and sub-specialty clinic system. The 50,000 sq ft laboratory is located in the Melvin R. Laird Center on the Marshfield Campus. Marshfield Labs performs over 2.4 million diagnostic clinical laboratory tests and reports over 22 million results annually. The laboratory is supported by a staff of over 350 and led by physicians and clinical consultants (PhD level scientists). In addition to having state-of-the-art facilities and informatics systems, Marshfield Labs holds accreditation and deemed status from the clinical Laboratory Improvement Act (CLIA), the College of American Pathologists (CAP), the Joint Commission for Accreditation of Healthcare Organizations (JCAHO), is an OSHA-approved laboratory for blood lead level testing, and accredited by the College of American Pathologists (CAP-FDT) for forensic drug testing.

COMPUTER
Information Systems and Databases:
The Marshfield medical complex is served by an integrated information system. The information system supports high quality health care, research and education at Marshfield's main campus and regional centers. The information system is built upon a 4,000-node wide area network that connects Marshfield Clinic and its regional centers, Marshfield Clinic Health System Hospital, Marshfield Laboratories, and MCRI. Fiberoptic backbones, T3, T1, and 56K circuits provide access to over 50 Unix, Novell, and Microsoft NT servers. The

Biomedical Informatics Research Center maintains a high performance computing cluster and a research data warehouse. Custom-built applications provide seamless access to an integrated database from any networked workstation. A key component of this database is the clinical repository that stores the electronic medical record for more than one million patients. In addition, an on-line information systems support staff is available to assist in troubleshooting of computer-related questions in order to minimize computer down time and loss of productivity. Data collected tracking patient encounters with the healthcare systems are captured in a sophisticated system of integrated databases Clinical data are transferred daily or weekly to the Marshfield Clinic Data Warehouse.

Databases include:

- Marshfield Clinic Clinical and Administrative (transaction) Processing System (MARS): is a highly integrated system automating the financial, practice management, clinical, and real-time decision support processes of the Clinic. MARS supplies a mature, fully featured electronic medical record. This system is the source of most of the clinical data entering the data warehouse. Some of the clinical data captured in the MARS system include: 1) laboratory results, 2) diagnosis, 3) procedures, 4) immunizations, 5) medications (under development), 6) vitals and 7) demographics.
- Medical Event Charting and Code Acquisition (MECCA): This database stores point of care applications between patients and providers. MECCA captures the detail of visits, providers, diagnoses, procedures and vitals. It uses a clinically relevant lexicon developed by Marshfield Clinic physician specialists.
- Laboratory Information System. Marshfield Laboratories has a sophisticated laboratory management system that automates the processes within the clinical laboratory. This system currently supports high volume testing for the main lab at Marshfield and 22 regional labs. The laboratory serves over 10,000 clients and performed over 20 million lab tests last year.
- Clinical Registries. The Clinical Registry systems are developed to support the evaluation of clinical practice and health outcomes in targeted clinical populations. The clinical registries contain either data that is collected at the time of patient care or data retrospectively collected by abstracting the medical chart of CMR. The majority of the clinical registry systems collect data for the purpose of evaluating patient outcomes, benchmarking, accreditation, or management reporting.
- Documents System. Healthcare providers generate over 50,000 clinical textual documents weekly and have been captured in a database since 1992. These documents can be searched to provide coded information for research utilizing key word search capabilities
- Financial System. Marshfield Clinic Business information systems include processing charges, third-party charges, statement/collections, payment/remittances, managed care, accounting and payroll/personnel. These systems are tightly integrated with many other systems in the Marshfield Clinic including MECCA, Appointment/Event, Laboratory, Radiology, Security Health Plan and several purchased packages.
- Web Portals: Marshfield Clinic is developing Internet-based applications (portals) that enable its customers to directly interact with the Clinic's information systems. Patients will be able to access and update portions of their electronic medical record, refill prescriptions, manage healthcare finances, and take more control of their healthcare through wellness and disease management programs. The patient portal will be expanded to collect useful environmental, and lifestyle information.

CENTERS:

Marshfield Epidemiology Research Center (MERC):

The mission of the Marshfield Epidemiology Research Center (MERC) is to contribute to the medical and scientific knowledge base by conducting high-quality population-based and other epidemiologic research. Marshfield Clinic is unique in the country in its support as a physician group practice of an epidemiology research center of this size. MERC is a member of several national research networks including the NIH-funded Cancer Research Network and the Cardiovascular Disease Research Network, the CDC-funded Vaccine Safety Datalink, and the Health Maintenance Organization Research Network. Current research concentrations include vaccine safety, influenza vaccine effectiveness, cancer surveillance, obesity prevention, cardiovascular care and outcomes, and antibiotic resistance.

The Marshfield Epidemiologic Study Area (MESA), a 24-Zip code region within the Clinic's primary service area housed within MERC, is one of the Research Foundations unique population-based research resources. It provides outstanding opportunities for epidemiologic studies of chronic diseases, infectious diseases, and genetic determinants of disease. MERC staff support researchers in epidemiological and clinical database

interrogation required for clinically- and epidemiologically-based research projects and have the capacity to interface epidemiological and clinical data.

Biomedical Informatics Research Center (BIRC):
BIRC provides informatics support to researchers throughout the Marshfield Clinic Research Institute as well as conducts original research in clinical informatics. The BIRC supports over 215 PCs. It includes five Windows servers that provide file sharing, research databases, electronic data forms, and the Personalized Medicine genetic repository. In addition there are two dedicated Apple Xserves for the development and evaluation of Human Computer Interaction prototypes for healthcare. BIRC data analysts have access to the Marshfield Clinic data warehouses which provide electronic records on over a million clinical encounters, covering multiple decades. Including the Biostatistics and Bioinformatics Core, BIRC has over 25 research staff, including informaticians, statisticians, programmers, and data analysts. BIRC supports a high performance computing cluster.

Center for Human Genetics (CHG):
The Research Foundation became recognized nationally and internationally for research in human genetics in the early 1990s with the discovery by Dr. Jim Weber of short tandem repeat polymorphisms, which revolutionized the study of human genetics. Marshfield genetic scientists focused their research on discovering the structure of the human genome, and the Marshfield Maps became the most reliable and widely used maps of the human genome in the world until it was sequenced in 2003. Its National Institutes of Health funded Mammalian Genotyping Service contributed to over 200 genetic research projects nationally and internationally between 1994 and 2006.

The vision of Center for Human Genetics today is one of excellence in translational genetic research that improves human health. It is home to the Personalized Medicine Research Project (PMRP), which includes the largest population-based biobank in the United States. Since PMRP was launched in September 2002, nearly 20,000 participants have enrolled. Currently it is supporting multiple studies of the genetic basis of disease and in pharmacogenetics. The Wisconsin Genomics Initiative (WGI), announced at Marshfield by Governor Jim Doyle in October 2008, builds upon PMRP. WGI is a historic collaboration of Wisconsin's three academic medical centers and its major urban university to advance predictive medicine and Personalized Health Care.

Clinical Research Center (CRC):
The Clinical Research Center (CRC) supports and oversees clinical trials conducted at the Marshfield campus and the Clinic's Regional Centers. Led by a physician Director of Clinical Research and Regional Assistant Directors, the CRC is staffed by over 50 research nurses, study coordinators, and other personnel. More than 21,000 patients are currently participating in clinical trials. CRC has clinical research offices at Clinic locations in Marshfield, Weston, Eau Claire, Chippewa Falls, and Minocqua.

National Farm Medicine Center (NFMC):
The National Farm Medicine Center (NFMC) was established in 1981. Its mission is to conduct high quality research addressing human health and safety associated with rural and agricultural work, life, and environments. Its has one of the longest running and most successful agricultural health and safety research programs in the country, and it has helped to shape national policy in many areas. Major research priorities include agro-medicine; childhood injury prevention; environmental health, such as food and water safety; and health and safety in the dairy industry. It also provides professional training and outreach services and publishes the nationally indexed *Journal of Agromedicine*. NFMC is one of eight CDC-funded Centers for Agricultural Disease and Injury Research, Education, and Prevention in the nation. NFMC is partnering with the Wisconsin agricultural producers, including the Progressive Dairy Producers of Wisconsin and Dairy Business Group and UW-Madison College of Agricultural and Life Sciences, to develop the Wisconsin Dairy Worker Initiative. This effort which will focus upon research, outreach and provision of services directed towards the changing work force in expanding Wisconsin dairies, primarily the Hispanic work force. NFMC is a very active charter member of the Agricultural Safety and Health Council of America (ASCHA), a national collaborative effort of agricultural producers, NIOSH, and researchers to improve translational research in agricultural health and safety.

OTHER

Office of Research Integrity and Protections:

The Office of Research Integrity and Protections (ORIP) is responsible for protecting the rights and welfare of participants in the research process and for preserving the integrity of the research process. To these ends, the ORIP has established and maintains the following institutional committees to ensure that:

- research is appropriately designed *(Research Committee)*
- the safety and welfare of human subjects is considered *(Institutional Review Board)*
- the safety and welfare of animal subjects is considered *(Institutional Animal Care and Use Committee)*
- employees working with Recombinant DNA or Biosafety Level 3 agents are protected *(Biosafety Committee);*
- any conflicts of interest are eliminated or managed *(Conflict of Interest Committee);* and
- research activities are monitored, education is provided and corrective action is taken when necessary (*Research Compliance Committee).*

ORIP is also responsible for policy development and enforcement of regulations related to scientific misconduct and the research aspects of the HIPAA Privacy rule.

Office of Sponsored Programs and Fiscal Affairs:

This department assists with submission and management of external funding awarded to Institutional initiatives. The staff has extensive experience in the administration of competitive, federally funded research grants and other award monies.

Office of Scientific Writing and Publication:

The mission of the Office of Scientific Writing and Publication (OSWP) is to support scientists and clinician investigators in developing, writing, and submitting external and internal grants as well as to assist in the publication of results of medical research by providing Marshfield Clinic investigators support in manuscript writing, editing, and submission. The OSWP seeks to establish a synergistic partnership with researchers from project conception through successful publication. In addition, this office also assists *Clinical Medicine & Research* in an editorial capacity.

G. E. Magnin Medical Library:

The Medical Library is an on-site, full service health science library with about 3,000 textbooks, 200 core print journal titles as well as over 8,000 online periodicals available to all Marshfield Clinic staff. The staff of six, including professional librarians, is available to assist researchers and health care professionals' retrieval of information from a wide variety of print and online resources. Online resources are available to all staff via network connections, but computer workstations are also available on-site. The Medical Library participates in interlibrary loan systems including the National Network of Libraries of Medicine. In addition, the library is available 24/7 for study and research through a wireless key-card security system.

Telecommunications Service:

This department facilitates operations involving successful executions of audio and video teleconferencing and with creation of audio-visual presentations.

Graphic Arts Department:

Department members assist in the creation of graphic art, posters, photography and artwork for publications and presentations.

OMB Number: 4040-0001
Expiration Date: 10/31/2019

RESEARCH & RELATED Senior/Key Person Profile (Expanded)

PROFILE - Project Director/Principal Investigator

Prefix: Dr. First Name*: Steven Middle Name Last Name*: Schrodi Suffix: Ph.D.

Position/Title*: Associate Research Scientist
Organization Name*: Marshfield Clinic Research Institute
Department:
Division:
Street1*: 1000 North Oak Avenue
Street2:
City*: Marshfield
County:
State*: WI: Wisconsin
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 54449-5790

Phone Number*: 715-221-6443 Fax Number:

E-Mail*: schrodi.steven@marshfieldresearch.org

Credential, e.g., agency login: SCHRODIS

Project Role*: PD/PI Other Project Role Category:

Degree Type: PhD Degree Year: 2001

Attach Biographical Sketch*: File Name: 1236-NIH_Biosketch_SchrodiSJ_02Jul2018.pdf

Attach Current & Pending Support: File Name:

Tracking Number: GRANT12668618 Funding Opportunity Number: PA-18-484 . Received Date: 2018-07-03T18:44:57.000-04:00

PROFILE - Senior/Key Person

Prefix: Dr. First Name*: Shicheng Middle Name Last Name*: Guo Suffix: Ph.D.

Position/Title*: Postdoctoral Fellow
Organization Name*: Marshfield Clinic Research Institute
Department: Human Genetics
Division:
Street1*: 1000 North Oak Avenue
Street2:
City*: Marshfield
County:
State*: WI: Wisconsin
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 54449-5790

Phone Number*: 715-389-3508 Fax Number:

E-Mail*: guo.shicheng@marshfieldresearch.org

Credential, e.g., agency login: SHICHENGGUO

Project Role*: Post Doctoral Other Project Role Category:

Degree Type: PhD Degree Year: 2015

Attach Biographical Sketch*: File Name: 1237-NIH_Biosketch_Guosc.pdf

Attach Current & Pending Support: File Name:

PROFILE - Senior/Key Person

Prefix: Dr. First Name*: Mehdi Middle Name Last Name*: Maadooliat Suffix: Ph.D.

Position/Title*: Assistant Professor
Organization Name*: Marquette University
Department: Statistics
Division:
Street1*: Katharine R. Cudahy Hall, Room 311
Street2: 1313 W. Wisconsin Avenue
City*: Milwaukee
County:
State*: WI: Wisconsin
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 53201-1881

Phone Number*: 414-288-7573 Fax Number:

E-Mail*: mehdi@mscs.mu.edu

Credential, e.g., agency login:

Project Role*: Co-Investigator Other Project Role Category:

Degree Type: PhD Degree Year: 2011

Attach Biographical Sketch*: File Name: 1238-Biosketch_Mehdi-N.pdf

Attach Current & Pending Support: File Name:

PROFILE - Senior/Key Person

Prefix: Dr. First Name*: Judith Middle Name Last Name*: Smith Suffix: M.D.

Position/Title*: Associate Professor
Organization Name*: University of Wisconsin - Madison
Department: Pediatrics
Division:
Street1*: 4159 Microbial Sciences Building
Street2: 1550 Linden Drive
City*: Madison
County:
State*: WI: Wisconsin
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 53792-4108

Phone Number*: 608-263-1251 Fax Number:

E-Mail*: jsmith27@wisc.edu

Credential, e.g., agency login: JSMITH27

Project Role*: Consultant Other Project Role Category:

Degree Type: MD Degree Year: 1999

Attach Biographical Sketch*: File Name: 1239-Smith_biosketch 062818.pdf

Attach Current & Pending Support: File Name:

PROFILE - Senior/Key Person

Prefix: Dr. First Name*: Krista Middle Name Last Name*: Trappl-Kimmons Suffix: Ph.D.

Position/Title*: Project Manager
Organization Name*: Antigen Discovery, Inc.
Department:
Division:
Street1*: 1 Technology Drive #E309
Street2:
City*: Irvine
County:
State*: CA: California
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 92618-2319

Phone Number*: 949-679-4068 Fax Number:

E-Mail*: ktrappl@antigendiscovery.com

Credential, e.g., agency login: ktrappl

Project Role*: Co-Investigator Other Project Role Category:

Degree Type: PhD Degree Year: 2011

Attach Biographical Sketch*: File Name: 1240-NIH_biosketch_Trappl_July2018_MCRI.pdf

Attach Current & Pending Support: File Name:

PROFILE - Senior/Key Person

Prefix: Dr. First Name*: Xiaowu Middle Name Last Name*: Liang Suffix: Ph.D.

Position/Title*: Chief Executive Officer
Organization Name*: Antigen Discovery, Inc.
Department:
Division:
Street1*: 1 Technology Drive #E309
Street2:
City*: Irvine
County:
State*: CA: California
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 92618-2319

Phone Number*: 949-679-4068 Fax Number:

E-Mail*: xliang@antigendiscovery.com

Credential, e.g., agency login: xliangso

Project Role*: Co-Investigator Other Project Role Category:

Degree Type: PhD Degree Year: 1990

Attach Biographical Sketch*: File Name: 1241-XIaowu_NIH_BioSketch_2018Schrodi2.pdf

Attach Current & Pending Support: File Name:

OMB No. 0925-0001 and 0925-0002 (Rev. 10/15 Approved Through 10/31/2018)

BIOGRAPHICAL SKETCH
DO NOT EXCEED FIVE PAGES.

NAME: Steven J. Schrodi, PhD

eRA COMMONS USER NAME (credential, e.g., agency login): SCHRODIS

POSITION TITLE: Tenure-Track Associate Research Scientist–Human Genetics

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
University of California, Davis	B.S.	06/1995	Genetics
University of California, Irvine	M.S.	12/1998	Biological Sciences
University of California, Irvine	Ph.D.	03/2001	Biological Sciences

A. Personal Statement

My primary research interests lie in complex disease gene mapping in the areas of systemic inflammatory diseases and statistical genetics/disease genetics theory. My work has synthesized disease gene mapping approaches and statistical methods with the aim of identifying variants that underlie inflammatory disease phenotypes. Current areas of investigation include the development of probability-based methods to analyze genetic data, linkage disequilibrium patterns from disease-susceptibility variants, and shared chromosomal region analyses. My experimental research focuses on the discovery of alleles and biomarkers that predispose individuals to autoimmunity and dysfunction in immune tolerance, aberrant immune responses to pathogens, and chronic inflammation. The ultimate goals of my research are to develop widely-applicable methods of disease genetics analysis, gain an understanding of molecular pathogenesis of immunological diseases, and provide results useful for remediation of disease.

1. Bansal, N.K., Maadooliat, M., **Schrodi, S.J.** (2018) Empirical Bayesian approach to testing multiple hypotheses with separate priors for left and right alternatives. *Stat Appl Genet Mol Biol* 17(3):20180002.

2. Maadooliat M., Bansal N.K., Updhya J., Farazi M.R., Li X., He M., Hebbring S.J., Ye Z., **Schrodi S.J.*** (2016) The decay of disease association with declining linkage disequilibrium: A fine mapping theorem. *Frontiers in Genetics*, 7:217. (*****Corresponding Author)

3. Carter, T.C., Rein, D., Padberg, I., Peter, E., Rennefahrt, U., David, D.E., McManus, V., Stefanski, E., Martin, S., Schatz, P., **Schrodi S.J.*** (2016). Validation of a metabolite panel for early diagnosis of type 2 diabetes. *Metabolism*, 65(9):1399-1408. (*****Corresponding Author)

4. Hebbring SJ, Slager SL, Epperla N, Mazza JJ, Ye Z, Zhou Z, Achenbach SJ, Vasco DA, Call TG, Rabe KG, Kay NE, Caporaso NE, Camp NJ, Strom SS, Goldin LR, Cerhan JR, Brilliant MH, **Schrodi, S.J.*** (2012) Genetic evidence of PTPN22 effects on chronic lymphocytic leukemia. (2012). *Blood*, 121(1):237-238. (*****Corresponding Author)

5. Cargill, M.‡, **Schrodi, S.J.**‡, Chang, M., Garcia, V.E., Brandon, R., Callis-Duffin, K.P., Matsunami, N., Ardlie, K.G., Civello, D., Catanese, J.J., Leong, D.U., Panko, J.M., McAllister, L.B., Papenfuss, J., Prescott, S.M., White, T.J., Leppert, M.F., Krueger, G.G., Begovich, A.B. (2007) A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *American Journal of Human Genetics*, 80(2):273-290. (‡Contributed equally)

B. Positions and Honors

Positions and Employment

2000-2001	Scientist, DNA Sciences, Fremont, CA
2001-2006	Senior Scientist, Celera, Alameda, CA
2006-2008	Staff Scientist, Celera, Alameda, CA
2008-2010	Senior Staff Scientist, Celera, Alameda, CA
2013-Pres	Faculty Trainer, Computation and Informatics in Biology and Medicine, University of Wisconsin-Madison, Madison, WI
2010-Pres	Tenure-Track Associate Research Scientist, Center for Human Genetics, Marshfield Clinic Research Institute, Marshfield, WI

Other Experience and Professional Memberships

1992-1993	Internship, Theoretical Space Science Division, NASA Ames
1995	Visiting Researcher, Institute for Theoretical Dynamics
2000-2001	Scientific Advisory Board, DNA Sciences
2008-2010	Statistical Genetics Consultant, Biotechnology and Pharmaceutical Companies
2012	Critical Assessment of Massive Data Analysis, Scientific Committee
2014	Multiple Sclerosis Research Australia Grants Review Panel
2006-2014	American Association for the Advancement of Science, Member
2000-2014	American Society of Human Genetics, Member
2014-2015	Marshfield Clinic Research Institute Strategic Planning Committee
2014-2015	Institute of Clinical and Translational Research Grant Review Panel
2013-2016	Marshfield Clinic Research Compliance Committee
2004-2016	Twenty-one US Patent Applications, Inventor
2005-2017	International Society of Bayesian Analysis, Member
2018	Clinician Scientist Collaborative Research Award Grant Review Panel
2013-2018	Marshfield Clinic Research Institute Seminar Series Committee
1993-Pres	Thirty-two Invited Oral Presentations
1998-Pres	Fifty Accepted Scientific Conference Abstracts
1999-Pres	*Ad hoc* Reviewer for Nineteen Scientific Journals
2015-Pres	Associate Editor, *Frontiers in Genetics, Statistical Genetics and Methodology*

Honors

2004	Top 10 Arthritis Advances of 2004, Arthritis Foundation
2005	UCSF Frontiers in Neurology & Neuroscience, Keynote Speaker
2007	Applera Demonstrated Noteworthy Achievement Award
2010	US Patent 7,833,706; Inventor
2011	US Patent 7,863,021; Inventor
2011	US Patent 7,947,451; Inventor
2011	US Patent 7,993,833, Inventor
2015	US Patent 8,975,022; Inventor
2016	US Patent 9,371,565; Inventor
2016	Mathematics, Statistics and Computer Science Colloquium Speaker, Marquette Univ.

C. Contribution to Science

1. **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.

    a. **Schrodi, S.J.** (2001) *Mathematical models in population genetics, molecular evolution and genomics.* UMI Dissertation Services, Ann Arbor, MI.

2. **First Exome-wide Association Scan for a Common Disease: Discovery of *PTPN22* and *TRAF1* Rheumatoid Arthritis Susceptibility Genes.** In 2003, I led the design and analysis and interpreted results from the first large-scale exome-wide SNP association disease scan using 30,000 putatively functional coding variants, interrogating RA susceptibility. 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. R620W is a key factor in autoimmunity and has now been significantly associated with 11 common, chronic inflammatory diseases. Further, I led a fine-mapping effort 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 describing this work.
    a. Begovich, A.B., Carlton, V.E., Honigberg, L.A., **Schrodi, S.J.**, et al. (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *American Journal of Human Genetics*, 75(2), 330-337.
    b. Carlton, V.S., Hu, X., Chokkalingam, A.P., **Schrodi, S.J.**, et al. (2005) PTPN22 genetic variation: evidence for multiple variants associated with rheumatoid arthritis. *American Journal of Human Genetics*, 77(4), 567-581.
    c. Chang, M., Rowland, C.M., Garcia, V.E., **Schrodi, S.J.**, et al. (2008) A large-scale rheumatoid arthritis genetic study identifies association at chromosome 9q33.2. *PLoS Genetics*, 4(6), e1000107.
    d. Begovich, A.B., Carlton, V.E.H., **Schrodi, S.J.**, Alexander, H.C. (Filed Jan 30, 2003; Awarded Nov 16, 2010) *United States Patent* 7,833,706. Genetic polymorphisms associated with rheumatoid arthritis, methods of detection and uses thereof.
    e. **Schrodi, S.J.** and Begovich, A.B. (Filed Sept 5, 2007; Awarded Jan 4, 2011) *United States Patent* 7,863,021. Genetic polymorphisms associated with rheumatoid arthritis, methods of detection and uses thereof.

3. **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 four United States Patents describing these psoriasis and autoinflammatory disease results.
    a. Cargill, M.[‡], **Schrodi, S.J.**[‡], Chang, M., Garcia, V.E., et al. (2007) A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *American Journal of Human Genetics*, 80(2):273-290. ([‡]Equal contributions)
    b. Garcia, V.E., Chang, M., Brandon, R., Li, Y.J., Matsunami, N., Callis-Duffin, K.P., Civello, D., Rowland, C.M., Bui, N., Catanese, J.J., Krueger, G.G., Leppert, M.F., Begovich, A.B., **Schrodi, S.J.*** (2008) Detailed genetic characterization of the interleukin-23 receptor in psoriasis. *Genes & Immunity*, 9(6):546-555. (*Corresponding Author)
    c. Chang, M., Li, Y.J., Yan, C., Callis-Duffin, K.P., Matsunami, N., Garcia, V.E., Cargill, M., Civello, D., Bui, N., Catanese, J.J., Leppert, M.F., Krueger, G.G., Begovich, A.B., **Schrodi, S.J.** (2008) Variants in the 5q31 cytokine gene cluster are associated with psoriasis. *Genes & Immunity*, 9(2):176-181.
    d. Nair, R.P., Duffin, K.C., Helms, C., …, **Schrodi, S.J.**, et al. (2009) Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways*. Nature Genetics*, 41(2):199-204.
    e. Tsoi, L.C., Spain, S.L., Knight, J., …**Schrodi, S.J.**, et al. (2012) Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nature Genetics*, 44(12):1341-1348.

4. **Statistical Genetics**. I have an ongoing interest in developing novel statistical genetics methods and approaches. In 2000, I was recruited by Dr. Ray White to DNA Sciences where I 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 founder populations. Over the past 17 years, I have worked on methods of selecting tagging SNPs, TDT statistical approaches, Bayesian hypothesis testing for genetic association, Bayesian estimation applied to genetics, permutation approaches, and multiple testing.

a. **Schrodi, S.J.*** (2005) A probabilistic approach to large-scale association scans: a semi-Bayesian method to detect disease-predisposing alleles. *Stat Appl Genet Mol Biol*, 4, Article 31. (*Corresponding Author)
b. **Schrodi, S.J.***, DeBarber, A., He, M., Ye, Z., et al. (2015) Prevalence estimation for monogenic autosomal recessive disease using population-based genetic data. *Human Genetics*, 134(6):659-669. (*Corresponding Author)
c. **Schrodi, S.J.*** (2016) The use of multiplicity corrections, order statistics and generalized family-wise statistics with application to genome-wide studies. *PLoS One* 11(4):e0154472. (*Corresponding Author)
d. **Schrodi, S.J.*** (2017) The impact of diagnostic code misclassification on optimizing the experimental design of genetic association studies. *J Healthc Eng* 2017:7653071. (*Corresponding Author)
e. Bansal, N.K., Maadooliat, M., **Schrodi, S.J.** (2018) Empirical Bayesian approach to testing multiple hypotheses with separate priors for left and right alternatives. *Stat Appl Genet Mol Biol* 17(3):20180002.

5. **Disease Genetics Theory and Prediction of Disease Traits.** My research on theoretical models of disease genetics has shown how linkage disequilibrium with a causal site varies with mode of inheritance, including a mathematical formulation for precisely how disease association statistics decays as linkage disequilibrium declines from a causal site. My colleagues and I show 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.

   a. **Schrodi, S.J.***, Garcia, V.E., Rowland, C.M., Jones, H.B. (2007) Pairwise linkage disequilibrium under disease models. *Eur J Human Genetics*, 15(2), 212-220. (*Corresponding Author)
   b. **Schrodi, S.J.***, Mukherjee, S., Shan, Y., Tromp, G., et al. (2014) Genetic-based prediction of disease traits: prediction is very difficult, especially about the future. *Frontiers in Genetics*, 5:162. (*Corresponding Author)
   c. **Schrodi, S.J.*** (2016) Reflections on the field of human genetics: A call for increased disease genetics theory. *Frontiers in Genetics*, 7:106. (*Corresponding Author)
   d. Carter, T.C., Rein, D., Padberg, I., Peter, E., Rennefahrt, U., David, D.E., McManus, V., Stefanski, E., Martin, S., Schatz, P., **Schrodi, S.J.*** (2016). Validation of a metabolite panel for early diagnosis of type 2 diabetes. *Metabolism*, 65(9):1399-1408. (*Corresponding Author)
   e. Maadooliat, M., Bansal, N.K., Updhya, J., Farazi, M.R., Li, X., He, M., Hebbring, S.J., Ye, Z., **Schrodi, S.J.*** (2016) The decay of disease association with declining linkage disequilibrium: A fine mapping theorem. *Frontiers in Genetics*, 7:217. (*Corresponding Author)

Complete List of Published Work:
http://www.ncbi.nlm.nih.gov/sites/myncbi/steven.schrodi.1/bibliography/47248234/public/?sort=date&direction=descending

D. Additional Information: Research Support and/or Scholastic Performance

Ongoing Research Support

Marshfield Clinic Research Institute Schrodi (PI) 07/15/2018 – 07/14/2020
Detecting Shared Chromosomal Regions and Compound Heterozygous Effects for Diseases
Role: Principal Investigator

UL1 TR000427 Drezner (PI) 06/01/2015 - 05/31/2018
University of Wisconsin
NIH-NCATS/UW-Institute for Clinical & Translational Research
The major goal of this study is to create an environment that facilitates the transformation or research at the University into a continuum extending from investigation through discovery to translation into practice, thereby linking even the most basic research to practical improvements in human health.
Role: Scientist

1RO1GM114128 Hebbring (PI) 09/01/2014 - 08/31/2019
PheWAS of Loss-of-Function Variants
The hypothesis being tested in this project is that loss-of-function variants – a class of variation with the highest probability of being clinically relevant – may cause disease phenotypes described in EMRs.
Role: Co-Investigator

5R01MH097464-03 Lainhart (PI) 04/01/2013 - 09/30/2018
The Biological Basis of Variations in Brain Structure and Function in Autism
The overarching hypothesis of this study is: Looking at variability and extremes of autism-associated imaging data as phenotypic measures of autism we will be able to find new genetic variants that lead to biological pathways that explain the imaging findings and advance understanding of clinical impairing manifestations of the disorder.
Role: Site Principal Investigator

Completed Research Support

Clinical Scientist Development Award Shelef (PI) 07/01/2017 - 06/30/2018
Doris Duke Charitable Foundation
Genetic variants, immune dysregulation, & rheumatoid arthritis
The major goals of this study are to examine how exome variants drive dysfunction in NET formation and autoantibody production in rheumatoid arthritis.
Role: Consultant

Second Genome Schrodi (PI) 08/09/2015 - 08/09/2016
Inflammasome Host and Microbiome Genetics
Role: Principal Investigator

Rheumatology Research Foundation Smith (PI) 07/01/2013 - 12/31/2015
WIC - Analysis of Causal Variants in the IL-23/IL-17 Pathway Genes in Axial Spondyloarthritis
Axial Spondyloarthritis, including the prototypic ankylosing spondylitis is a slowly progressive debilitating disease that may be caused by excess production of inflammatory mediators, in particular IL-17. Although previous genetic studies have implicated multiple genes in an “IL-23/IL-17” pathway that regulates pathogenic immune cells, the proposed research is necessary to explain how variations within the individual genes alter immune responses.
Role: Site Principal Investigator

Metanomics Health Schrodi (PI) 08/01/2012 - 07/13/2013
Metanomics Health and Marshfield Type 2 Diabetes Prediction
As a collaboration between Metanomics Health and MCRI, this study evaluates and develops predictive models for T2D using a panel of metabolites and additional biological markers as applied to PMRP samples.
Role: Principal Investigator

UL1 TR000427 Drezner (PI) 06/01/2012 - 05/31/2013
University of Wisconsin
NIH-NCATS/UW-Institute for Clinical & Translational Research
The major goal of this study is to create an environment that facilitates the transformation or research at the University into a continuum extending from investigation through discovery to translation into practice, thereby linking even the most basic research to practical improvements in human health.
Genomics of IL-23/IL-17-Mediated Chronic Inflammation
Role: T1 Pilot Grant Principal Investigator

OMB No. 0925-0001 and 0925-0002 (Rev. 10/15 Approved Through 10/31/2018)

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Guo, Shicheng, PhD

eRA COMMONS USER NAME (credential, e.g., agency login): SHICHENGGUO

POSITION TITLE: Postdoctoral Fellow–Human Genetics

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Northeast Agricultural University, Harbin, China	B.S.	06/2009	Biology
Fudan University, Shanghai, China	Ph.D.	01/2015	Human Genetics
University of Texas Health Science Center at Houston, Houston, TX	Postdoc	04/2015	Genetic Epidemiology
University of California, San Diego, CA	Postdoc	10/2017	Human Genetics

A. Personal Statement

My research has focused on the development and analysis of the epigenomic architecture assembly of human cells/tissue and other important model organisms using epigenetic- (DNA methylation and miRNA profiles) and genetic variant- (single-nucleotide polymorphisms (SNPs) and copy number variant (CNV) screens) based approaches. Through this work, I have discovered susceptibility factors associated with the development and progression of various diseases. These disease-susceptibility factors can be used as diagnostic and prognostic biomarkers to further clinical research in human complex diseases, such as lung cancers, thyroid cancer, bladder cancer, liver cancer, ankylosing spondylitis (AS), gout, and systemic sclerosis (SSc). My previous work includes (i) the identification of SSc and RA-predisposing SNPs and CNVs using case-control approaches, (ii) identification of diagnostic biomarkers for solid tissue human cancers, and (iii) origin-tissue mapping for cell-free DNA based on tissue-specific methylation panels. Current areas of investigation include disease susceptibility screening using genome-wide association studies (GWAS) and phenome-wide association studies (PheWAS) approaches and assessing the genetic-epigenetic interactions in the identification, etiology, and treatment of various human diseases. The ultimate goal of my research is to develop widely-applicable biomarker-based methods for disease diagnosis, disease subtype identification, and/or prognosis. I have experience with different bioinformatics-based analyses for genetic variation, epigenetic data (methylation sequencing, ChIP-seq data), text-mining, and machine learning analysis using Perl, R, and Python programs which will serve me well in a broad array of projects utilizing bioinformatics and biostatistics analyses.

B. Positions and Honors

Positions and Employment

2015-2015 Postdoctoral Fellow, University of Texas Health Science Center at Houston, TX
2015-2017 Postdoctoral Fellow, University of California, San Diego, CA
2017-Present Postdoctoral Fellow, Center for Human Genetics, Marshfield Clinic, WI

Other Experience and Professional Memberships

2011-2014 Internship, Institute of Rheumatology, Immunology and Allergy, Shanghai, China
2012-2013 Internship, CAS-MPG Partner Institute for Computational Biology, Shanghai, China
2012-2013 Visiting Scholar, University of Texas Health Science Center at Houston, Houston, TX
2013-2015 Research Assistant, University of Texas Health Science Center at Houston, Houston, TX

Honors

2007 Second prize of National Mathematical Modeling Contest in Heilongjiang province, Harbin, China
2012 Silver award of "Cup of Challenge" for College Students' Innovative Undertaking Contest, Shanghai
2014 First Place Poster, 17th Annual Human and Molecular Genetics Program Symposium, GSBS, TX

C. Contribution to Science

1. **Identification of autoimmune disease susceptibility genetics**

Early in my career, I investigated genetic variants involved in systemic sclerosis (SSc) and rheumatoid arthritis within the Chinese Han population. Applying a multiple candidate pre-selection method (SNP and CNV screens), I identified multiple susceptibility genes, such as an important CNV within *HLA-DQA1* and *APOBEC3A/3B* for SSc, *CFH* for age-related macular degeneration, and *FOXE1* for thyroid cancer. I also conducted a large association study interrogating genetic variants in miRNA for human cancer and identified miR-4293 as being significantly associated with non-small cell lung cancer, and miR-196a2/miR-499 involved in esophageal squamous cell carcinoma. These findings have provided much needed molecular insight into the role of miRNA regulation and genetic variants involved in these cancer etiologies.

a. Huang, L[‡], Y. Li[‡], **S. Guo**[‡], Y. Sun, C. Zhang, Y. Bai, S. Li, F. Yang, M. Zhao, B. Wang, W. Yu, C.C. Khor, and X. Li, Different hereditary contribution of the CFH gene between polypoidal choroidal vasculopathy and age-related macular degeneration in Chinese Han people. Invest Ophthalmol Vis Sci, 2014. 55(4): p. 2534-8. ([‡]Contributed equally)
b. Shen, F., J. Chen, **S. Guo**, Y. Zhou, Y. Zheng, Y. Yang, J. Zhang, X. Wang, C. Wang, D. Zhao, M. Wang, M. Zhu, L. Fan, J. Xiang, Y. Xia, Q. Wei, L. Jin, and J. Wang, Genetic variants in miR-196a2 and miR-499 are associated with susceptibility to esophageal squamous cell carcinoma in Chinese Han population. Tumour Biol, 2016. 37(4): p. 4777-84.
c. **Guo, S**., Y. Li, Y. Wang, H. Chu, Y. Chen, Q. Liu, G. Guo, W. Tu, W. Wu, H. Zou, L. Yang, R. Xiao, Y. Ma, F. Zhang, M. Xiong, L. Jin, X. Zhou, and J. Wang, Copy Number Variation of HLA-DQA1 and APOBEC3A/3B Contribute to the Susceptibility of Systemic Sclerosis in the Chinese Han Population. *J Rheumatol*, 2016. 43(5): p. 880-6.
d. L., L. Chen, X. Ni, **S. Guo**, Y. Zhou, C. Wang, Y. Zheng, F. Shen, V.K. Kolluri, M. Muktiali, Z. Zhao, J. Wu, D. Zhao, Z. He, X. Feng, Z. Yuan, J. Zhang, L. Jin, J. Wang, and M. Wang, Genetic variant of miR-4293 rs12220909 is associated with susceptibility to non-small cell lung cancer in a Chinese Han population. *PloS one*, 2017. 12(4): p. e0175666.

2. **Epigenome architecture assembly to normal and disease tissues**

Starting in 2015, I investigated the epigenetics of human disease with a particular focus on DNA methylation. I participated in several large projects to build a model of the epigenome architecture for human cells and tissues under normal and disease conditions. Notable work includes evaluating the genomic methylation profiles (methylomes) for normal human blood cells, animal model 'silk', CD4+ T-cells of patients with rheumatoid arthritis, pancreatic cancer cells, and hepatocellular carcinoma cells with different methylation methods, such as BS-seq and MBD-seq.

a. **Guo, S**[ǂ]., Q. Zhu[ǂ], T. Jiang, R. Wang, Y. Shen, X. Zhu, Y. Wang, F. Bai, Q. Ding, X. Zhou, G. Chen, and D.Y. He, Genome-wide DNA methylation patterns in CD4+ T cells from Chinese Han patients with rheumatoid arthritis. *Mod Rheumatol*, 2017. 27(3): p. 441-447. ([ǂ]Contributed equally)
b. Zhao, Y[ǂ]., F. Xue[ǂ], J. Sun[ǂ], **S. Guo**[ǂ], H. Zhang, B. Qiu, J. Geng, J. Gu, X. Zhou, W. Wang, Z. Zhang, N. Tang, Y. He, J. Yu, and Q. Xia, Genome-wide methylation profiling of the different stages of hepatitis B virus-related hepatocellular carcinoma development in plasma cell-free DNA reveals potential biomarkers for early detection and high-risk monitoring of hepatocellular carcinoma. *Clin Epigenetics*, 2014. 6(1): p. 30. ([ǂ]Contributed equally)
c. Zhao, Y[ǂ]., J. Sun[ǂ], H. Zhang[ǂ], **S. Guo**[ǂ], J. Gu, W. Wang, N. Tang, X. Zhou and J. Yu, High-frequency aberrantly methylated targets in pancreatic adenocarcinoma identified via global DNA methylation analysis using methylCap-seq. *Clin Epigenetics*, 2014. 6(1): p. 18. ([ǂ]Contributed equally)
d. Zhao, Y[ǂ]., **S. Guo**[ǂ], J. Sun[ǂ], Z. Huang, T. Zhu, H. Zhang, J. Gu, Y. He, W. Wang, K. Ma, J. Wang, and J. Yu, Methylcap-seq reveals novel DNA methylation markers for the diagnosis and recurrence prediction of bladder cancer in a Chinese population. *PloS one*, 2012. 7(4): p. e35175. ([ǂ]Contributed equally)

3. **Epigenetic variations and their use in diagnosing and treating complex diseases.**
DNA methylation is known to be aberrant in the early stages of cancer. We identified a large number of methylation-based markers with diagnostic and prognostic implications for non-small cell lung cancer, bladder cancer, and pancreatic cancer. Since DNA methylation has different patterns for different tissue types, we proposed a prediction model to map the origin of cell-free DNA fragments based on tissue-specific methylation signals. This model provides a potential non-invasive approach for the diagnosis of solid cancers. In my current investigation, I am assessing the interaction effects of genetic variants with epigenetic variations in human complex diseases and applying these findings to the diagnosis and identification of disease subtypes.
    a. **Guo, S**., F. Yan, J. Xu, Y. Bao, J. Zhu, X. Wang, J. Wu, Y. Li, W. Pu, Y. Liu, Z. Jiang, Y. Ma, X. Chen, M. Xiong, L. Jin, and J. Wang, Identification and validation of the methylation biomarkers of non-small cell lung cancer (NSCLC). *Clin Epigenetics*, 2015. 7: p. 3.
    b. Geng, X., W. Pu, Y. Tan, Z. Lu, A. Wang, L. Tan, S. Chen, **S. Guo**, J. Wang, and X. Chen, Quantitative assessment of the diagnostic role of FHIT promoter methylation in non-small cell lung cancer. *Oncotarget*, 2017. 8(4): p. 6845-6856.
    c. Pu, W., C. Wang, S. Chen, D. Zhao, Y. Zhou, Y. Ma, Y. Wang, C. Li, Z. Huang, L. Jin, **S. Guo**, J. Wang, and M. Wang, Targeted bisulfite sequencing identified a panel of DNA methylation-based biomarkers for esophageal squamous cell carcinoma (ESCC). *Clin Epigenetics*, 2017. 9: p. 129.
    d. **Guo, S**.[ǂ], D. Diep[ǂ], N. Plongthongkum, H.L. Fung, and K. Zhang, Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet*, 2017. 49(4): p. 635-642. . ([ǂ]Contributed equally)

Complete List of Published Work:

https://www.ncbi.nlm.nih.gov/myncbi/browse/collection/45297273/?sort=date&direction=descending

D. Research Support

Ongoing Research Support
None at this time.

Completed Research Support
None at this time.

OMB No. 0925-0001 and 0925-0002 (Rev. 11/16 Approved Through 10/31/2018)

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Mehdi Maadooliat

eRA COMMONS USER NAME (credential, e.g., agency login):

POSITION TITLE: Assistant Professor

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Sharif University of Technology, Tehran, Iran	BS	12/03	Applied Mathematics
Marquette University	MS	08/06	Mathematics, Statistics and Computer Science
Texas A&M University	PhD	08/11	Statistics
Texas A&M University (2 years Postdoc)	Postdoctoral	06/13	Bioinformatics

A. Personal Statement

My primary research interests lie in the developments of the statistical models in high-dimensional data structures with application to biological sciences, including, but not limited to genomics and proteomics. In particular my doctoral research focuses on the dimension reduction and the functional data analysis in non-Gaussian frameworks, which has been applied in the context of analysis of high-dimensional gene expression data; and the main focus of my current research is on the modeling of the large spherical data structures with an application in protein structure prediction and classification.

The goal of the proposed project is to develop molecular-based predictive models for eight chronic and complex diseases using an extensive longitudinal electronic health records linked to the Marshfield Clinic Biobank. In this project Bayesian Networks will be used to coalesce the prognostic signals from genetic variants and protein biomarkers into disease classifiers. Also, a careful consideration will be given for appropriate modeling of the potential correlation structure and controlling the Type-I error for multiple hypotheses testing as well.

I have experience in statistical consulting for such data analysis. I have done statistical consulting for graduate students in Texas A&M University since 2007 for many projects mostly related to bioinformatics. I have a sound experience in R, C++, and Matlab to perform modeling and data analyses as proposed in this project.

B. Positions and Honors

2004-2006	Teaching Assistant, Marquette University, WI
2006-2009	Teaching Fellow, Texas A&M University, TX
2008-2009 (6 months)	Summer Internship, MD Anderson Cancer Center, TX
2009-2011	Research Assistant, Texas A&M University, TX
2011-2013	IAMCS Postdoctoral Fellow, Texas A&M University, TX
2013-	Assistant Professor, Marquette University, WI
2015-	Associate Research Scientist, Marshfield Clinic Research Institute, WI

Honors

2016 Way Klingler Young Scholar Award, Marquette University, Milwaukee, WI

C. Contributions to Science

1. **Bioinformatics:** My main area of research is on modeling the large data structures with an application in protein structure prediction and statistical genetics using machine learning techniques and functional data analysis.
    a. S. M. Najibi, **M. Maadooliat**, L. Zhou, J. Z. Huang and X. Gao, "Protein structure classification and loop modeling using multiple Ramachandran distributions", *Computational and Structural Biotechnology Journal*, Vol. 15, 243-254 (2017)
    b. **M. Maadooliat**, N. K. Bansal, J. Upadhya, M. R. Farazi, Z. Ye, X. Li, and S. J. Schrodi, "The decay of disease association with declining linkage disequilibrium: A fine mapping theorem", *Frontiers in Genetics: Statistical Genetics and Methodology*, Vol. 7, A217, (2016)
    c. **M. Maadooliat**, L. Zhou, S. M. Najibi, X. Gao and, J. Z. Huang, "Collective estimation of multiple bivariate density functions with application to angular-sampling-based protein loop modeling", *Journal of the American Statistical Association*, Vol. 111, No. 513, 43-56, (2016)
    d. **M. Maadooliat**, X. Gao and J. Z. Huang, "Assessing protein conformational sampling methods based on bivariate lag-distributions of backbone angles", *Briefings in Bioinformatics*, Vol.14, No.6, 724-736, (2013)

2. **Statistical methodology:** Second part of my research focuses on developing the statistical methodology for modeling the dependency (correlation) structure and dimension reduction in non-Gaussian framework.
    a. H. M. Kim, **M. Maadooliat**, M. Genton, and R. B. Arellano-Valle, "Skewed Factor models using selection mechanism", *Journal of Multivariate Analysis*, Vol. 145, 162–177, (2016)
    b. **M. Maadooliat**, J. Z. Huang and J. Hu, "Integrating data transformation in principal components analysis", *Journal of Computational and Graphical Statistics*, Vol.24, No.1, 84-103, (2015)
    c. **M. Maadooliat**, M. Pourahmadi and J. Z. Huang, "Robust estimation of the correlation matrix of longitudinal data", *Statistics and Computing*, Vol.23, No.1, 17-28, (2013)
    d. **M. Maadooliat**, J. Z. Huang and J. Hu, "Analyzing multiple-probe microarray: estimation and application of gene expression indexes", *Biometrics*, Vol.68, No.3, 784-792, (2012)

3. **Collaboration:** I have experienced collaborations with a nursing department, a group of statisticians in studying directional multiple hypothesis, and a group of offshore engineers for developing a new technique to study the reliability of mooring line systems.
    a. N. K. Bansal, **M. Maadooliat**, and S. J. Schrodi, "Empirical Bayesian approach to testing multiple hypotheses with separate priors for left and right alternatives." *Statistical Applications in Genetics and Molecular Biology*, Accepted, (2018)
    b. M. Bull, L. Boaz, **M. Maadooliat**, M. Hagle, L. Settrust, M. Greene, S. Holmes and J. Saczynski. "Preparing family caregivers to recognize delirium symptoms in older adults following elective hip or knee arthroplasty", Journal of the American Geriatrics Society, Vol. 65, e13-e17, (2017)
    c. N. K. Bansal, G.G. Hamedani, and **M. Maadooliat**, "Testing multiple hypotheses with skewed alternatives", *Biometrics*, Vol. 72, No. 2, 494-502, (2016)
    d. G.G. Hamedani, Z. Javanshiri, **M. Maadooliat** and A. Yazdani, "Remarks on Characterizations of Malinowska & Szynal", *Applied Mathematics and Computation*, Vol.246, 377-388, (2014)
    e. M. E. Mousavi, P. Gardoni and **M. Maadooliat**, "Progressive reliability method and its application to offshore mooring systems", *Engineering Structures*, Vol.56, 2131-2138, (2013)

D. Additional Information: Research Support and/or Scholastic Performance

Ongoing Research Support:

R01 NIH: R. Fitts, S. Hunter, A. Ng, S. W. Trappe, C. Konersman, **M. Maadooliat**, 09/01/15-04/30/20
"Fatigability of limb muscle in older adults: Protective effects of exercise"
This project will provide important new information that will translate into clinically important exercise programs to improve quality of life and reduce health care costs with aging.
Role: Co-Investigator

Marshfield Clinic Research Institute: S. J. Schrodi, **M. Maadooliat**, S. Guo, 07/15/18-07/14/20
"Detecting Shared Chromosomal Regions and Compound Heterozygous Effects for Diseases within PMRP"
This project will focus into discovering novel genes harboring genetic variants that contribute to diseases.
Role: Co-PI

Completed Research Support

Retirement Research Foundation: M. Bull, L. Boaz, M. Maadooliat, L. Gettrust, M. Hagle, J. Saczynski, M. Greene, S. Holmes, 09/01/14-08/31/15
"Preparing family carers to recognize symptoms of acute confusion (Delirium) in older adults following elective arthroplasty of the knee or hip"
Role: Statistician

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.

Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: Smith, Judith Anne

eRA COMMONS USER NAME (agency login): JSMITH27

POSITION TITLE: Associate Professor (tenured)

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.)*

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Yale University, New Haven, CT	BS	06/1991	Biology
University of Chicago, Chicago, IL	PHD	08/1997	Immunology
University of Chicago Pritzker School of Medicine, Chicago, IL	MD	06/1999	
University of Chicago, Chicago, IL	NIH training grant	1997	NIH Pulmonary Training Grant
University of Chicago, Chicago, IL	NIH training grant	1999	NIH Medical Scientist Training Program
Cincinnati Children's Hospital Medical Center, Cincinnati, OH	Resident	2002	Pediatrics
Cincinnati Children's Hospital Medical Center, Cincinnati, OH	Fellow	2005	Pediatric Rheumatology

A. Personal Statement

The long-term goal of my research is to gain a better understanding of the pathogenesis of ankylosing spondylitis (AS) and related conditions in the hopes of improving patient care. This research trajectory grew out of my excellent Pediatric Rheumatology fellowship training in Dr. Robert Colbert's lab and I have continued in this field independently. AS appears to be a largely genetic disease, with >90% heritability. A strong link to the HLA-B27 gene has been known for decades, though exactly how this gene contributes to disease remains unclear. HLA-B27 accounts for less than half the genetic risk. GWAS studies have indicated other immunomodulatory loci, including multiple "hits" in genes that regulate the development and activity of T helper 17 cells. Through my experience treating pediatric rheumatology patients, I have witnessed the surprising efficacy of cytokine blockade. As a researcher, I have become intrigued by the signals (such as ER stress) that regulate inflammatory cytokine production. Current efforts are underway to determine the genetics regulating cytokine production and resulting functional abnormalities in AS. Beyond cytokine production and innate immune macrophage responses, I also have a solid basic science foundation in adaptive immunology (T cells in particular) from my Immunology PhD. I am fortunate to be collaborating with Dr. Steven Schrodi at the Marshfield clinic. He has tremendous expertise in the genetics underlying Rheumatologic diseases. Our two very different backgrounds and areas of expertise provide a highly unusual synergistic opportunity to understand how genetic variation translates into immune functional alteration. I am very excited to contribute my medical and immunologic expertise to Dr. Schrodi's current proposal.

B. Positions and Honors

Positions and Employment

2005 - 2006	Research Associate, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, Cincinnati, OH
2006 - 2014	Assistant Professor, University of Wisconsin-Madison, Department of Pediatrics, Madison, WI
2014 –	Associate Professor (tenured), University of Wisconsin-Madison, Department of Pediatrics, Madison, WI
2016	Affiliate, Department of Medical Microbiology and Immunology, UW-Madison

Other Experience and Professional Memberships

2008 - 2011	Member, Pediatric Research Society (PRS)
2009 -	Member, SPARTAN (Spondyloarthropathy research and therapy network)
2010 -	Voting Member, CARRA (Childhood Arthritis and Rheumatology Research Alliance)
2011 -	Member, American Association of Immunology
2014 -	Board of Directors, SPARTAN
2014 - 2016	Mentor, AMIGO (Rheumatology mentoring program)
2014 - 2017	Rheumatology Research Foundation study section
2016	NIH R13 study section
2016 - 2018	Treasurer, SPARTAN
2018 -	Vice Chair, SPARTAN
2018-	Rheumatology Research Foundation scientific advisory committee
2018	NIH P30 study section

Honors

1990	Elected to National Honor Society, Phi Beta Kappa
1991	Graduated Summa Cum Laude, with honors in Biology, Yale University
2008-2009	CTSA KL2 Dean's Scholar, University of Wisconsin, Madison
2008,2011, 2013, 2015, 2017	Graduate School Fall Competition Award, UW Division of Biological Sciences
2010	Notable Poster Award, American College of Rheumatology
2011	Young Investigator Award, Spondylitis Association of America
2011-2012	Beverly Howland Memorial Fellow, Arthritis National Research Foundation
2012	Research and Education Foundation Pilot Grant Award, American College of Rheumatology
2013	Gerald B. Odell Department of Pediatrics Research Award, University of Wisconsin

C. Contribution to Science

1. **Partial T cell receptor signals:** Early in my career, as a student in the Bluestone lab, I embarked on a study of how T cell receptor antibodies (anti-CD3) modulate immune responses. Although transplantation related immunosuppression was the initial impetus, the findings took on a broader significance. Historically, during this time, the potential for partial or suboptimal signal delivery through the T cell receptor (TCR) was gaining increasing appreciation. The TCR was no longer conceived as a purely “on-off” switch. We found that decreased cross-linking of the TCR resulted in a partial signal very reminiscent of the signal conveyed by “altered peptide” ligands at the biochemical level. Further, these partial signals had different consequences for different T helper subsets, specifically resulting in Th1 anergy. This work had implications for the development of non-activating immunosuppressive anti-CD3 therapies, as well as for understanding T cell activation.
   a. Madrenas J, Chau LA, **Smith J**, Bluestone JA, Germain RN. The efficiency of CD4 recruitment to ligand-engaged TCR controls the agonist/partial agonist properties of peptide-MHC molecule ligands. J Exp Med. 1997 Jan 20;185(2):219-29. PubMed PMID: 9016871; PubMed Central PMCID: PMC2196122.
   b. **Smith JA**, Tso JY, Clark MR, Cole MS, Bluestone JA. Nonmitogenic anti-CD3 monoclonal antibodies deliver a partial T cell receptor signal and induce clonal anergy. J Exp Med. 1997 Apr 21;185(8):1413-22. PubMed PMID: 9126922; PubMed Central PMCID: PMC2196281.
   c. **Smith JA**, Bluestone JA. T cell inactivation and cytokine deviation promoted by anti-CD3 mAbs. Curr Opin Immunol. 1997 Oct;9(5):648-54. PubMed PMID: 9368773.
   d. **Smith JA**, Tang Q, Bluestone JA. Partial TCR signals delivered by FcR-nonbinding anti-CD3 monoclonal antibodies differentially regulate individual Th subsets. J Immunol. 1998 May 15;160(10):4841-9. PubMed PMID: 9590231.

2. **Unfolded protein response and type I IFN:** In ankylosing spondylitis, the link between the highly associated HLA-B27 and aberrant cytokine production has remained unclear. One major hypothesis is that

HLA-B27 misfolding leads to an endoplasmic reticulum stress response called the Unfolded Protein Response (UPR). Over the past decade it has become increasingly apparent that the UPR can modify cytokine responses. We were among the first to identify the synergy between ongoing UPR and pattern recognition receptor (PRR) stimulation in inducing pro-inflammatory cytokines. However one of the most strikingly synergistic interactions between UPR and PRR stimulation is in the generation of type I IFN. We further elucidated the mechanism underlying synergistic IFN-β production. As a fellow, I was the primary investigator, and then as faculty, principal investigator. These findings have implications for handling of intracellular infections (e.g. virus) as well as inflammatory conditions involving type I IFN and ER stress.

a. **Smith JA**, Turner MJ, DeLay ML, Klenk EI, Sowders DP, et al. Endoplasmic reticulum stress and the unfolded protein response are linked to synergistic IFN-beta induction via X-box binding protein 1. Eur J Immunol. 2008 May;38(5):1194-203. PubMed PMID: 18412159; PubMed Central PMCID: PMC2838478.

b. Zeng L, Liu YP, Sha H, Chen H, Qi L, **Smith JA.** XBP-1 couples endoplasmic reticulum stress to augmented IFN-beta induction via a cis-acting enhancer in macrophages. J Immunol. 2010 Aug 15;185(4):2324-30. PubMed PMID: 20660350; PubMed Central PMCID: PMC2916979.

c. Liu YP, Zeng L, Tian A, Bomkamp A, Rivera D, Gutman D, Barber GN, Olson JK, **Smith JA.** Endoplasmic reticulum stress regulates the innate immunity critical transcription factor IRF3. J Immunol. 2012 Nov 1;189(9):4630-9. PubMed PMID: 23028052; PubMed Central PMCID: PMC3478468.

d. **Smith JA**. A new paradigm: innate immune sensing of viruses via the unfolded protein response. Front Microbiol. 2014;5:222. PubMed PMID: 24904537; PubMed Central PMCID: PMC4032990.

3. **Pathogenesis of ankylosing spondylitis**: Cytokine dysregulation drives multiple autoimmune and autoinflammatory processes, including ankylosing spondylitis (AS). Through direct interrogation of macrophages from AS subjects, we have made several cardinal observations: 1) AS subject macrophages produce decreased IFN-γ, and reveal evidence for a decrease in IFN-γ regulated genes. This data is consistent with, and extends earlier observations from peripheral blood. The study also inspired investigations from a group in France investigating the animal model. 2) AS macrophages produce excess IL-23 in response to LPS in the absence of an overt UPR. This finding suggested dysregulated Th17 promoting cytokine responses to environmental stimuli as a potential underlying pathogenic driver. The study also suggested there must be other mechanisms besides UPR that promote IL-23 production. I served as primary or principal investigator for these studies. I have also synthesized related information on pathogenesis in several invited reviews.

   a. **Smith JA**, Barnes MD, Hong D, DeLay ML, Inman RD, Colbert RA. Gene expression analysis of macrophages derived from ankylosing spondylitis patients reveals interferon-gamma dysregulation. Arthritis Rheum. 2008 Jun;58(6):1640-9. PubMed PMID: 18512784; PubMed Central PMCID: PMC2888278.

   b. Zeng L, Lindstrom MJ, **Smith JA**. Ankylosing spondylitis macrophage production of higher levels of interleukin-23 in response to lipopolysaccharide without induction of a significant unfolded protein response. Arthritis Rheum. 2011 Dec;63(12):3807-17. PubMed PMID: 22127699; PubMed Central PMCID: PMC3228355.

   c. **Smith JA,** Colbert RA. Review: The interleukin-23/interleukin-17 axis in spondyloarthritis pathogenesis: Th17 and beyond. Arthritis Rheumatol. 2014 Feb;66(2):231-41. PubMed PMID: 24504793; PubMed Central PMCID: PMC4058712.

   d. Yiping Liu, Zhan Ye, Xiang Li, Jennifer L. Anderson, Mike Khan, Douglas DaSilva, Marissa Baron, Deborah Wilson, Vera Bocoun, Lynn C. Ivacic, Steven J. Schrodi, **Judith A. Smith**, Genetic and functional associations with decreased anti-inflammatory Tumor Necrosis Factor Alpha Induced Protein 3 in macrophages from subjects with axial spondyloarthritis, Front Immunol, 24 July https://doi.org/10.3389/fimmu.2017.00860, 2017. PMCID: PMC5523649

4. **Host-Pathogen interaction in *Brucella* infection**. Brucellosis is the most frequent zoonosis worldwide, with over 500,000 infections per year and no available human vaccine. My interest in brucellosis began with the finding that this ER replicating organism induces the UPR (publication a below). This study was a natural outgrowth from my interest in how UPR shapes inflammatory responses. *Brucella* infection is also

frequently complicated by spondyloarthritis. At UW-Madison, I developed a productive collaboration with Dr. Splitter, a leading expert in Brucella pathogenesis. Since this first foray into host-pathogen interaction, my lab has expanded its scope of *Brucella* investigation to cover adaptive immune responses (publication b), and other regulators of bacterial physiology (publication c) and host response in collaboration with Dr. Oliveira's lab in Brazil. There are few groups in the United States investigating the human pathogen *B. melitensis*, and thus this work addresses an important global need.

a. **Smith JA,** Khan M, Magnani DD, Harms JS, Durward M, Radhakrishnan GK, Liu YP, Splitter GA. *Brucella* induces an unfolded protein response via TcpB that supports intracellular replication in macrophages, PLos Pathog, 2013; 9(12):e1003785. PMCID: 3855547.

b. Durward-Diioia M, Harms JS, Khan M, Hall C, **Smith JA**, Splitter GA. CD8+ T cell exhaustion, suppressed IFN-γ production, and delayed memory response induced by chronic *Brucella melitensis* infection, Infect Immun, 2015; 83(12):4759-71. PMID: 26416901; PMCID: PMC4645381.

c. Khan M, Harms J, Marim F, Armon L, Hall C, Liu YP, Banai M, Oliveira SC, Splitter GA, **Smith JA**. The bacterial second messenger cyclic-di-GMP regulates Brucella pathogenesis and leads to altered host immune response, Infect Immun, Nov 18:84(12): 3458-3470, 2016. PMCID: 5116723.

d. Jerome Harms, Mike Khan, Cherisse Hall, Jane Homan, Robert Bremel, Gary A Splitter, **Judith A Smith**, Brucella-peptide cross-reactive presentation activates the Ovalbumin-specific TCR, accepted to Infection and Immunity, June 21;86(7): e00281-18, 2018. PMCID in process

Complete List of Published Work in My Bibliography:
http://1.usa.gov/1MTNtmF

D. Research Support

Ongoing Research Support

P01 HL070831, Lemanske (PI) 09/26/13 – 06/30/18
NIH/NHLBI
Rhinovirus Infections and Asthma in Children and Adolescents
Project I of this Program Project Grant will address the following hypothesis: HRV wheezing illnesses, working through at least two independent mechanistic pathways, lead to the development of distinct asthma phenotypes that can be characterized immunologically and physiologically and be further modulated over time based on gender and developmental stage of the host.
Role: Co-Investigator (Project 1)

R21 AI121808, Smith (PI)
NIH/NHLBI 09/01/16-08/31/18
Regulation of Human Immune Function by 17q21 asthma risk polymorphism
The goal of this project is to determine how altered levels of ORMDL3 related to 17q21 risk polymorphism alter B-cell and eosinophil functional responses.

R01 AI116453, Oliveira *(*PI) 01/20/16 – 12/31/20
NIH-NIAID & Federal University of Minas Gerais
Brucella survival strategy requires endoplasmic reticulum restructuring and interferes with innate immunity
The goals of this grant are to unravel how Brucella activates innate immune components that result in host resistance versus bacterial subversion of the immune response.
Role: Site PI, Co-Investigator

- NB: this grant provides supply money, which will be leveraged towards what is proposed here, but no lab personnel support for the Smith lab.

Completed Research Support (last 3y)

R01 AI073558-08, Splitter, Smith, Harms (MPIs) 05/01/14 – 04/30/17
NIH-NIAID

Brucella epitope recognition by CD8+T cells
The purpose of this grant is to determine CD8 T cell response during acute and chronic brucellosis, identify CD8 memory T cells and evaluate the effect of Brucella TcpB protein on CD8 T cell effector function.
Role: MPI

135-AAA7943, Smith (PI) 01/01/16 – 06/30/17
University of Wisconsin-Madison, Graduate School
Immunogentics of Axial Spondyloarthritis
The aims of this project are 1) to identify causal genetic variants in a panel of 28 deep-sequenced Spondyloarthritis (SpA) and autoimmunity implicated genes by associating gene expression; and 2) use the RNAseq as a discovery platform to identify other dysregulated immune pathways in SpA.
Role: PI

UL1 TR000427, Drezner (PI) 08/01/15 – 07/31/16
NIH-NCATS/UW-Institute for Clinical & Translational Research
Development of IL-17 Targeting Small Molecule Therapeutics for Autoimmune Diseases
The pro-inflammatory cytokine IL-17 plays a critical role in multiple autoimmune conditions. Although anit-IL-17 monoclonal antibody therapy has proven effective for psoriasis and ankylosing spondylitis, there is a pressing need for less expensive and invasive therapies. This collaboration will examine the cellular effects, molecular structure, and targets of two lead compounds discovered through a program sponsored by Eli Lilly. Together these approaches will begin to define the mechanism of action by which these two compounds inhibit cytokine production.
Role: Pilot Grant PI

133-PRJ73ZD, Smith (PI) 07/01/13 – 04/30/16
Rheumatology Research Foundation
Analysis of Causal Variants in the IL-23/IL-17 pathway
We will deep sequence genes and regulatory regions in genome wide association studies, and determine consequences for gene expression, and function, correlating with disease phenotype.
Role: PI

135-PRJ73RZ, Smith (PI) 07/01/13 – 09/30/15
University of Wisconsin-Madison, Graduate School
The role of the host Unfolded Protein Response in *Brucella* replication
Supplemental funding with aim of generating a competitive NIH grant application (gained F31 for lab).
Role: PI

OMB No. 0925-0001 and 0925-0002 (Rev. 10/15 Approved Through 10/31/2018)

BIOGRAPHICAL SKETCH

NAME: Trappl-Kimmons, Krista

eRA COMMONS USER NAME: ktrappl

POSITION TITLE: Project Manager

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	Completion Date MM/YYYY	FIELD OF STUDY
Leopold Franzens University Innsbruck, Austria	M.Sc.	10/2006	Microbiology
Medical University Innsbruck, Austria	Ph.D.	11/2011	Molecular Biology
University of California San Diego	Postdoctoral	12/2015	Chemistry & Biochemistry

A. Personal Statement

I have the expertise and the necessary training, motivation, and institutional support to successfully carry out the proposed research project titled, Multi-Omics Prediction of Eight Common Diseases in the Central Wisconsin Population. I have a broad background in medically relevant infectious disease, bacterial genomics and biochemistry. Together with Dr. Liang I will be responsible for overseeing administrative activities and the design and implementation of the scientific activities. As a project manager at Antigen Discovery I will oversee the day-to-day activities related to probing of samples, data acquisition and analysis.

Antigen Discovery, Inc. (ADI) is dedicated to the development, production, and use of highly sensitive and specific protein microarrays for detection of immune response to pathogens in human specimens. In the past, ADI has developed genome-wide protein microarrays for more than 30 medically important infectious microorganisms such as bacteria, protozoa, and viruses. ADI already has already developed various protein microarrays for *Escherichia coli (ETEC, EHEC, EPEC, EIEC, EAEC)*. ADI has a proven track record of successfully analyzing and characterizing the immune response in human serum, plasma and fecal samples for large numbers of samples. As a project manager at ADI, I will oversee the probing and analysis of ETEC infected samples and case control samples. With my history of successfully administering projects and coordinating collaborations with other researchers, I will be well-suited to bring success to the proposed project.

B. Positions and Honors

Positions and Employment

2016- Project Manager, ImmPORT Therapeutics dba Antigen Discovery Inc., Irvine, CA
2012-2015 Postdoctoral Fellow, Division of Chemistry and Biochemistry, University of California San Diego, San Diego, CA
2007-2011 Research Associate, Medical University Innsbruck, Austria
2005-2006 Research Associate, Leopold Franzens University Innsbruck, Austria

Other Experience and Professional Memberships

2008,2015- Member, The RNA Society
2012- Member, ASCINA Austrian Scientists & Scholars in North America
2014-2015 Member, AWIS Association for Women in Science

Awards

2011 Max Kade Fellowship, awarded by the Austrian Science Foundation

2014 Erwin Schrödinger Fellowship, awarded by the Austrian Science Fund (FWF)

C. Contribution to Science

1. Thermodynamic, Kinetic and Structural Insights into Stop Codon Recognition by Release Factor 1

The ribosome produces all proteins in every living cell. This process is called translation as it translates the genetic information from messenger RNA (mRNA) into proteins. The importance of the ribosome is reflected in the fact that most antibiotics target the ribosome. With sophisticated fast-kinetic experiments, I was able to observe the event of RF1 binding to the ribosome during translation-termination in a nano-second range. The thermodynamic and kinetic analysis of the RF1 mutants showed that the mutations inhibited the binding of RF1 to the ribosome. This is significant for the development of antibiotics in the future. Further, I aimed to identify the conditions, properties, and chronology of a conformational change in RF1. It was not clear whether only the open form of RF1 binds to the ribosome or the closed form of RF1 binds to the ribosome and then undergoes a conformational change into the open state. Using transition metal ion FRET experiments (tmFRET), I precisely monitored the conformation of RF1 in the absence and in the presence of the ribosome. In the publication, we propose that high termination fidelity is achieved by linking stop codon recognition by RF1 to a change in conformation from closed to open state, which increases the binding affinity of RF1 to the ribosome and induces peptide release.

a. Trappl K., Mathew MA., Joseph S. (2014) Thermodynamic and Kinetic Insights into Stop Codon Recognition by Release Factor 1. PLoS One. 9(4):e94058. doi: 10.1371/journal.pone.0094058
b. Trappl K. and Joseph S. (2016) Ribosome induces conformational change in Release Factor 1. Journal of Molecular Biology 428(6):1333-44. doi: 10.1016/j.jmb.2016.01.021.

2. Probing the ribosomal exit tunnel with tRNA-peptide conjugates

Each nascent peptide has to pass the ribosomal exit tunnel on its way out of the ribosome. The ribosomal exit tunnel is a favored target of macrolide antibiotics. Novel RNA-peptide conjugates, carrying these so-called resistance peptides were bound to the ribosome in the presence of the antibiotic. It was shown that the antibiotic was not able to bind to the ribosome to the same extent when those resistance peptides were present. It furthermore showed conformational changes in the exit tunnel on an atomic level due to the presence of the antibiotic and suggests a “talk-back” mechanism driven by a single nucleotide in the ribosomal exit tunnel that eventually leads to the resistant mechanism. My work shed light on the development of antibiotic resistance in bacteria and is the foundation for future research on this topic. Antimicrobial resistance is resistance of a microorganism to an antimicrobial drug that was originally effective for treatment of infections caused by it. Antimicrobial resistance threatens the effective prevention and treatment of an ever-increasing range of infections caused by bacteria. It is an increasingly serious threat to global public health that requires action. A high percentage of hospital-acquired infections are caused by highly resistant bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA) or multidrug-resistant Gram-negative bacteria. Patients with infections caused by drug-resistant bacteria are generally at increased risk of worse clinical outcomes and death, and consume more health-care resources than patients infected with the same bacteria that are not resistant.

a. Graber D, Moroder H, Steger J, Trappl K, Polacek N, Micura R. (2010) Reliable semi-synthesis of hydrolysis-resistant 3’-peptidyl- tRNA conjugates containing genuine tRNA modifications. Nucleic Acids Res. (19):6796-802
b. Moroder H., Steger J., Graber D., Fauster K., Trappl K., Marquez V., Polacek N., Wilson D.W., Micura R. (2009) Non-hydrolyzable RNA-peptide conjugates: a powerful advance in the synthesis of mimics for 3’-peptidyl tRNA termini. Angew Chem Int Ed Engl.:48(22):4056-60
c. Trappl K. and Polacek N. (2011) Metal Ions Life Science: The ribosome: A molecular machine powered by RNA, 9 (Structural and Catalytic Roles of Metal Ions in RNA)

3. Effect of platelets and antibiotics on disease-causing aspergilla

In my research project, I investigated the additive effect of platelets and commonly used antimycotics (Amphotericin B, Caspofungin, Posacanazol and Voriconazol) against the four main disease-causing *Aspergillus* strains. My study showed that platelets and the antimycotic substances support each other to fight the fungus. A potential synergistic effect of antimycotic substances in combination with human platelets against *Aspergillus* was suggested by the gained data. The study suggests that a low platelet count is related to duration of an *Apergillus* infection. The in vitro data suggest that a normal platelet count contributes to overcome fungal infections and that platelets are capable of enhancing the efficacy of antimycotics. The gained knowledge from my work will influence how patients with low platelet counts will be treated for *Apergillus* infections, as this connection was unknown before. A presumably higher concentration of antimycotics will have to be used, which will lead to a faster cure. This gives the fungi less time to develop a resistance mechanism, which is amongst other factors promoted by to low concentrations of antimycoticum.

a. Perkhofer S, Trappl K, Striessnig B, Nussbaumer W, Lass-Flörl C. (2011) Platelets enhance activity of antimycotic substances against non-Aspergillus fumigatus Aspergillus species in vitro. Medical Mycology; 49(2):157-66.
b. Perkhofer S, Trappl K, Nussbaumer W, Dierich MP, Lass-Flörl C. (2010) Potential synergistic activity of antimycotic substances in combination with human platelets against Aspergillus fumigatus. J Antimicrob Chemother; 65(6):1309-11.

Complete List of Published Work in MyBibliography:

http://www.ncbi.nlm.nih.gov/pubmed/?term=trappl+k

D. Research Support

Ongoing Research Support

1 R43 AI132078-01 Camerini D (PI) 02/03/2017 – 1/31/2019
Project Title: Specific Detection of Antibodies to Emerging and Established Arboviruses Including Zika Virus
Role: Co-Investigator

1 R43 AI136197-01 Camerini D (PI) 01/16/2018 – 12/31/2019
Project Title: Discovery of Yellow Fever Virus-Specific Epitopes for Development of an Accurate Serodiagnostic Assay
Role: Co-Investigator

Completed Research Support

J3597-B22 Trappl (postdoctoral fellow) 09/01/2014-10/31/2015
Erwin Schrödinger Fellowship, awarded by the Austrian Science Fund (FWF)
Identification of conformational changes in RF1

Max Kade Trappl (postdoctoral fellow) 04/01/2012-11/31/2014
Max Kade Fellowship, awarded by the Austrian Science Foundation
RF1 mutational studies reveal molecular insights into translation termination

BIOGRAPHICAL SKETCH

NAME	POSITION TITLE
Xiaowu Liang	Chief Executive Officer
ECOMMONS USERNAME: xliangso	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY	FIELD OF STUDY
Peking University, Peking, China	BS	07/82	Biochemistry
University of California, San Diego / Salk Institute,	PhD	03/90	Molecular Biology
University of California, Berkeley, CA	Postdoc	04/90	Plant Molecular
Harvard Business *School*, MA	Exec. Ed.	05/09	Leadership & Strategy in Pharma and Biotech

A. Personal Statement

My role is to serve as co-investigator at Antigen Discovery Inc in support of Dr. Steven Schrodi for the R01 proposal entitled “Multi-Omics Prediction of Eight Common Diseases in the Central Wisconsin Population”. I will be responsible for coordinating and carrying out day to day operations to ensure successful execution of experimental plans and timely delivery of the human autoantigen protein microarrays, sample probing and data analysis. I do look forward to a pleasant and productive collaboration with Dr. Schrodi on this exciting project.

I have worked in biotechnology and drug development for over 20 years. Much of the fundamental technology used by Antigen Discovery, Inc. (ADI, previously, ImmPORT Therapeutics, Inc) was the product of my hands-on work. Since 2004, I have assumed various executive positions at ADI while retaining an active role in managing its growing suite of high throughput proteomics projects. Under my leadership and through team effort, ADI has quickly become a recognized leader in the development and application of high throughput proteomics platform technology for antigen/biomarker discovery. The innovative nature of the technology is detailed in over 80 peer reviewed publications in the past 9 years from ADI and our collaborators, as well as multiple patent applications covering over 2,000 novel, immunodominant, antigens from 30 microorganisms for use in the development of serodiagnostic tests, vaccines and therapeutics. Besides several phase I and phase II SBIR/STTR grants, ADI has also been awarded numerous R&D and service contracts from the US Government, the Gates Foundation, and from domestic as well as international biotech/pharmaceutical companies. Before joining ADi, I was the Lead Scientist in the Biodefense Division at Invitrogen Corp. where my team successfully developed, validated and launched the first multiplex-PCR and microfluidic analysis based system for the detection of biothreat agents, PathAlert^{TM}. Prior to Invitrogen I was with Gene Therapy Systems as Scientific co-founder and Molecular Biology Departmental Manager, where I was involved in the development of various novel tools for molecular biology and proteomics research, including the patented Transcriptionally Active PCR (TAP) system for high throughput gene expression. I also worked at Vical Inc. from 1990 to 1994 and had a major role in the development of DNA vaccines against cancer and infectious diseases.

B. Professional Positions and Honors

Professional Background

1994-1998	Senior Research Scientist	*Vical Inc., San Diego, CA*
1998-2002	Departmental Manager, Molecular Biology	*Gene Therapy Systems, Inc., San Diego, CA*
2002-2004	Lead Scientist, Biodefense Division	*Invitrogen, Carlsbad, CA*
2004 - 2008	Chief Technical Officer, & COO (Acting)	*ImmPORT Therapeutics, Inc., Irvine, CA*
2008 - present	President & CEO	*ImmPORT dba Antigen Discovery Inc., Irvine, CA*

Honors & Organizations

1979 – 1982	Outstanding Student Award, Peking University
1985 - 1990	Scholarship for Graduate Studies in Biology, University of California, San Diego
1990-present	Scientific Advisory Board, Rockefeller-Asia Advisory Group Ltd; Athenese Therapeutics

C. Contributions to Science (1-5)

1. I created and demonstrated the first single promoter-based *in vivo* inducible and repressible mammalian expression vector.
   - *Liang, X et al (1996) Novel, high expressing and antibiotic-controlled plasmid vectors designed for use in gene therapy. Gene Ther. 1996 Apr;3(4):350-6;* ***PMID: 8732167***
   - *US Patent #5,891,718 (1999)*

2. Created and demonstrated the first use of a peptide nucleic acid (PNA) "clamp" to directly and irreversibly modify plasmid DNA, without affecting either its supercoiled conformation or its ability to be efficiently transcribed.
   This detection method provides a way to simultaneously monitor the intracellular localization and expression of plasmid DNA in living cells, and to elucidate the mechanism of plasmid delivery and its nuclear import with synthetic gene delivery systems.
   - *Zelphati O, Liang X, Hobart P, Felgner PL. (1999) Gene chemistry: functionally and conformationally intact fluorescent plasmid DNA. Hum Gene Ther. 1999 Jan 1;10(1):15-24.****PMID:10022527***

3. I created the patented technology called *Transcriptionally Active PCR fragments (TAP)* for high throughput cell-free activation and expression of genes.
   The methodology has been adapted to a robotic work station enabling the high throughput generation of transcriptionally active genes. This technology offers a practical approach to directly utilize genome sequence data to generate functional proteomes.
   - *Liang, X, Teng, A., Braun, DM., et al. (2002) Transcriptionally active polymerase chain reaction (TAP): high throughput gene expression using genome sequence data. J Biol Chem. ;277(5):3593-8* ***PMID:11713261***
   - *US Patent #6,280,977 (2001)*

4. First to demonstrate using a high throughput proteomics approach to generate whole proteome microarrays for biomarker/antigen discovery.
   The platform provides a rapid way to comprehensively scan humoral immunity against microbial proteomes for diagnostic and vaccine antigen discovery
   - *Davies, H., Liang, X., et al (2005) Profiling the humoral immune response to infection using proteome microarrays. Proc Natl Acad Sci U S A. 102(3):547-52 PMCID:* ***PMID: 15647345***
   - *de la Maza, L et al (2014) Whole genome identification of C. trachomatis immunodominant antigens after genital tract infections and effect of antibiotic treatment of pigtailed macaques. J. Proteomics 28;108:99-109.* ***PMID:24862987***
   - *Baum E, et al (2013) Protein microarray analysis of antibody responses to Plasmodium falciparum in western Kenyan highland sites with differing transmission levels. PLoS One. 2013 Dec 2;8(12):e82246.* ***PMID: 24312649***
   - *Tan X, et al, (2012) Failure of the Smallpox Vaccine To Develop a Skin Lesion in Vaccinia Virus-Naive Individuals Is Related to Differences in Antibody Profiles before Vaccination, Not After. Clin Vaccine Immunol. 2012 Mar;19(3):418-28.* ***PMCID: 22258709***
   - US Patent #9,297,803 (2016)

5. My lab created the world first whole proteome microarray of *Mycobacterium tuberculosis* that was successfully used to identify serodiagnostic antigens for active TB disease.
   - *Kunnath-Velayudhan S, et al. (2010) Dynamic antibody responses to the Mycobacterium tuberculosis proteome. Proc Natl Acad Sci U S A. 2010 Aug 17;107(33):14703-8.* ***PMCID: 20668240***
   - US patents #7,927,818 (2011), #8,114,614 (2012), #8,883,431 (2014)

Issued and Pending Patents

- US Patent #9,297,803 (2016) Composition and methods for immunodominant antigens
- US Patent #8,883,431 (2014) Composition and immunodominant antigens of *M tuberculosis*.
- US Patent #8,114,614 (2012) Composition and immunodominant antigens of *M tuberculosis*.
- US Patent #7,927,818 (2011) Composition and immunodominant antigens of *M tuberculosis*.
- US Patent #6,936,470 (2005) Rapid and enzymeless cloning of nucleic acid fragments
- US Patent #6,451,769 (2002) Compositions and methods for administering Borrelia DNA vaccine
- US Patent #6,280,977 (2001) Method for generating transcriptionally active DNA fragments
- US Patent #5,891,718 (1999) Tetracycline inducible/repressible systems
- US Patent #5,846,946 (1998) Compositions and methods for administering Borrelia DNA vaccine
- Pending (2006) Protein Microarrays and antigens identified therewith
- Pending (2005) Methods, systems and products for generating plasmids, making microarray chips and performing immunological screening
- Pending (2011) Methods and compositions of protein antigens for the diagnosis and treatment of *Toxoplasma gondii* infections and toxoplasmosis
- Pending (2012) Methods and compositions of protein antigens for the diagnosis and treatment of Herpes Simplex Viruses type 1 and 2

D. Research Support (PI or Co-I in following awarded grants)

Current active

1R43AI136197 Camerini, D. (PI) (01/01/2018 – 12/31/2019) 1.0 Calendar Month
NIAID: Discovery of Yellow Fever Virus-Specific Epitopes for Development of an Accurate Serodiagnostic Assay
Role: Co-I; No Overlap.

1R43AI124911-01 Liang (PI) (08/01/16 – 07/31/18) 1.2 Calendar Months
NIH/NIAID A Novel Immuno-Proteomic Approach To a Genital Herpes Vaccine
Using proteomic approach to identify target antigens for the development of genital Herpes vaccine.
Role: PI; No Conflict

R43 AI132078-01 Camerini, D. (PI) (02/06/2017 – 01/31/2018) 1.2 Calendar Months
NIAID: Specific Detection of Antibodies to Emerging and Established Arboviruses Including Zika Virus
Role: Co-I; No Overlap.

Completed in last 3 years

1R43DE025440 Camerini (PI) (08/20/15 – 08/29/17) 1.2 Calendar Months
NIH/NIDCR Point of Care Detection of Oral Pathogens
Create a proteomic microarray containing all proteins from all common types of HIV-1, HPV and HSV, expand the microarray to include other oral pathogens, and develop a protocol for sensitive and specific screening for antibodies to oral pathogens in saliva using a proteomic microarray. Role: Co-Investigator. No Overlap

1R43DE025165 Camerini (PI) (09/16/14 – 12/31/15) 1.2 Calendar Months

NIH/NIDCR Development of a Pan-HIV Proteomic Chip
Development of a Pan-HIV Chip that will allow identification of the type, group and subtype of an HIV infection, facilitate rapid characterization of humoral immune response to nearly all HIV infections and vaccination regimens, identify reactivity with some broadly neutralizing epitopes and differentiate natural from vaccine-induced humoral immunity. Role: Co-Investigator

1R43AI102288-01 Liang (PI) (08/01/12 – 07/31/15) 1.2 Calendar Months
NIH/NIAID Integrated Microfluidic Platform for Point-of-Care Serodiagnostics
Development of microfluidic chip based multiplex detection system for infectious disease targets. Role: PI

2R44AI066791 Felgner (PI) (08/01/12 – 07/31/15)
NIH/NIAID Protective biomarkers for the development of vaccines against malaria 1.0 Calendar Months
Use P falciparum proteome microarrays to screen sera from malaria vaccine clinical trials to identify protective protein antigen markers. Role: Co-Investigator

PHS 398 Cover Page Supplement

OMB Number: 0925-0001

Expiration Date: 03/31/2020

1. Vertebrate Animals Section

Are vertebrate animals euthanized? ○ Yes ● No

If "Yes" to euthanasia

Is the method consistent with American Veterinary Medical Association (AVMA) guidelines?

○ Yes ○ No

If "No" to AVMA guidelines, describe method and provide scientific justification

2. *Program Income Section

*Is program income anticipated during the periods for which the grant support is requested?

○ Yes ● No

If you checked "yes" above (indicating that program income is anticipated), then use the format below to reflect the amount and source(s). Otherwise, leave this section blank.

*Budget Period	*Anticipated Amount ($)	*Source(s)

Tracking Number: GRANT12668618

Funding Opportunity Number: PA-18-484 . Received Date: 2018-07-03T18:44:57.000-04:00

PHS 398 Cover Page Supplement

3. Human Embryonic Stem Cells Section

*Does the proposed project involve human embryonic stem cells? ○ Yes ● No

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s) from the following list: http://grants.nih.gov/stem_cells/registry/current.htm. Or, if a specific stem cell line cannot be referenced at this time, check the box indicating that one from the registry will be used:

☐ Specific stem cell line cannot be referenced at this time. One from the registry will be used.

Cell Line(s) (Example: 0004):

4. Inventions and Patents Section (Renewal applications)

*Inventions and Patents: ○ Yes ○ No

If the answer is "Yes" then please answer the following:

*Previously Reported: ○ Yes ○ No

5. Change of Investigator/Change of Institution Section

☐ Change of Project Director/Principal Investigator

Name of former Project Director/Principal Investigator

Prefix:

*First Name:

Middle Name:

*Last Name:

Suffix:

☐ Change of Grantee Institution

*Name of former institution:

Tracking Number: GRANT12668618 Funding Opportunity Number: PA-18-484 . Received Date: 2018-07-03T18:44:57.000-04:00

PHS 398 Modular Budget

OMB Number: 0925-0001
Expiration Date: 03/31/2020

Budget Period: 1

Start Date: 05/01/2019 End Date: 04/30/2020

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium Indirect (F&A)*	250,000.00
Consortium Indirect (F&A)	0.00
Total Direct Costs*	250,000.00

B. Indirect (F&A) Costs

	Indirect (F&A) Type	Indirect (F&A) Rate (%)	Indirect (F&A) Base ($)	Funds Requested ($)
1.	Modified Total Direct Costs	60.00	250,000.00	150,000.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number): Dept of Health and Human Services; Arif M. Karim, 214-767-3,261

Indirect (F&A) Rate Agreement Date: 12/01/2016

Total Indirect (F&A) Costs: 150,000.00

C. Total Direct and Indirect (F&A) Costs (A + B)

Funds Requested ($): 400,000.00

Funding Opportunity Number: PA-18-484. Received Date: 2018-07-03T18:44:57.000-04:00

PHS 398 Modular Budget

Budget Period: 2			
Start Date: 05/01/2020 End Date: 04/30/2021			

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium Indirect (F&A)*	250,000.00
Consortium Indirect (F&A)	6,007.00
Total Direct Costs*	256,007.00

B. Indirect (F&A) Costs

	Indirect (F&A) Type	Indirect (F&A) Rate (%)	Indirect (F&A) Base ($)	Funds Requested ($)
1.	Modified Total Direct Costs	60.00	250,000.00	150,000.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number): Dept of Health and Human Services; Arif M. Karim, 214-767-3,261

Indirect (F&A) Rate Agreement Date: 12/01/2016

Total Indirect (F&A) Costs: 150,000.00

C. Total Direct and Indirect (F&A) Costs (A + B)

Funds Requested ($): 406,007.00

Funding Opportunity Number: PA-18-484. Received Date: 2018-07-03T18:44:57.000-04:00

PHS 398 Modular Budget

Budget Period: 3

Start Date: 05/01/2021 End Date: 04/30/2022

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium Indirect (F&A)*	225,000.00
Consortium Indirect (F&A)	20,270.00
Total Direct Costs*	245,270.00

B. Indirect (F&A) Costs

	Indirect (F&A) Type	Indirect (F&A) Rate (%)	Indirect (F&A) Base ($)	Funds Requested ($)
1.	Modified Total Direct Costs	60.00	225,000.00	135,000.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number): Dept of Health and Human Services; Arif M. Karim, 214-767-3,261

Indirect (F&A) Rate Agreement Date: 12/01/2016

Total Indirect (F&A) Costs: 135,000.00

C. Total Direct and Indirect (F&A) Costs (A + B)

Funds Requested ($): 380,270.00

Funding Opportunity Number: PA-18-484. Received Date: 2018-07-03T18:44:57.000-04:00

PHS 398 Modular Budget

Budget Period: 4

Start Date: 05/01/2022 End Date: 04/30/2023

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium Indirect (F&A)*	225,000.00
Consortium Indirect (F&A)	21,314.00
Total Direct Costs*	246,314.00

B. Indirect (F&A) Costs

	Indirect (F&A) Type	Indirect (F&A) Rate (%)	Indirect (F&A) Base ($)	Funds Requested ($)
1.	Modified Total Direct Costs	60.00	225,000.00	135,000.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number): Dept of Health and Human Services; Arif M. Karim, 214-767-3,261

Indirect (F&A) Rate Agreement Date: 12/01/2016

Total Indirect (F&A) Costs: 135,000.00

C. Total Direct and Indirect (F&A) Costs (A + B)

Funds Requested ($): 381,314.00

Funding Opportunity Number: PA-18-484. Received Date: 2018-07-03T18:44:57.000-04:00

PHS 398 Modular Budget

Budget Period: 5			
Start Date: 05/01/2023 End Date: 04/30/2024			

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium Indirect (F&A)*	225,000.00
Consortium Indirect (F&A)	21,953.00
Total Direct Costs*	246,953.00

B. Indirect (F&A) Costs

	Indirect (F&A) Type	Indirect (F&A) Rate (%)	Indirect (F&A) Base ($)	Funds Requested ($)
1.	Modified Total Direct Costs	60.00	225,000.00	135,000.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number): Dept of Health and Human Services; Arif M. Karim, 214-767-3,261

Indirect (F&A) Rate Agreement Date: 12/01/2016

Total Indirect (F&A) Costs: 135,000.00

C. Total Direct and Indirect (F&A) Costs (A + B)

Funds Requested ($): 381,953.00

Funding Opportunity Number: PA-18-484. Received Date: 2018-07-03T18:44:57.000-04:00

PHS 398 Modular Budget

Cumulative Budget Information

1. Total Costs, Entire Project Period

Section A, Total Direct Cost less Consortium Indirect (F&A) for Entire Project Period ($)	1,175,000.00
Section A, Total Consortium Indirect (F&A) for Entire Project Period ($)	69,544.00
Section A, Total Direct Costs for Entire Project Period ($)	1,244,544.00
Section B, Total Indirect (F&A) Costs for Entire Project Period ($)	705,000.00
Section C, Total Direct and Indirect (F&A) Costs (A+B) for Entire Project Period ($)	1,949,544.00

2. Budget Justifications

Personnel Justification	1234-Marshfield_Multiomics_Budget_Justification_070218.pdf
Consortium Justification	1235-BudgetandJustificationMaadooliat.pdf
Additional Narrative Justification	

Tracking Number: GRANT12668618
Funding Opportunity Number: PA-18-484. Received Date: 2018-07-03T18:44:57.000-04:00

Budget Justification: Marshfield Clinic Research Institute

Multi-Omics Prediction of Eight Common Diseases in the Central Wisconsin Population

FRINGE BENEFITS
Fringe benefits applicable to direct salaries and wages for Marshfield Clinic Research Institute (MCRI) personnel are treated as direct costs. The fringe benefit rate for each position varies depending upon the employee's salary level and the benefits chosen. Fringe benefits are included in the total personnel figure and are shown in the Fringe Benefit column under Personnel in the budget breakdown.

MERIT INCREASE
All salaries in all years, except those at the NIH salary cap, include a 3% merit increase. Historically, 3% has been our institutional practice for merit increases.

PERSONNEL
Steve Schrodi, PhD – Principal Investigator: Dr. Schrodi, from the Marshfield Clinic Research Institute (MCRI), will serve as the PI for the project. In this role, Dr. Schrodi will perform oversee the timing and implementation of the study. Dr. Schrodi will develop, conduct and oversee statistical analyses, genetics analyses, and biomarker analyses. Dr. Schrodi will oversee the development of machine learning approaches to create predictive models. Dr. Schrodi will develop diagnostic utility measures designed to assess the predictive capacity of machine learning classifiers. Dr. Schrodi will oversee the database queries and data management. Dr. Schrodi will work with Drs. Mehdi Maadooliat, Judy Smith, and MCRI Personnel in preparing manuscripts for publication. Dr. Schrodi will devote 6 C/M in Year1; 3.6 C/M in Year 2; 4.8 C/M in Years 3 & 4; and 5.52 C/M in Year 5.

Jennifer Meece, PhD – Director of the Integrated Research and Development Laboratory: Dr. Meece will provide oversight on laboratory assays development. Dr. Meece will devote 0.12 C/M in Years 1 and 2.

TBN Biostatistican: The TBN Biostatistician will provide expert statistical programming support and bioinformatics during the final phase of the project. The TBN Biostatistician will work with Dr. Schrodi in implementing statistical and programming algorithms to analyze exome variants, biomarkers and antibody data. The TBN Biostatistician will devote 1.2 C/M in Year 5.

Shicheng Guo, PhD – Postdoctoral Fellow: Dr. Guo will aid with statistical programming and data analyses for both the protein and exome data. Dr. Guo will aid in the interpretation of molecular results. Under the mentorship of Dr. Schrodi, Dr. Guo will be involved with the manuscript preparation. Dr. Guo will devote 3.84 C/M in Year 2; 9.6 C/M in Years 3 & 4; and 12 C/M in Year 5.

Research Programmer/Analyst: Within MCRI's Biomedical Informatics Research Center (BIRC) service center, the Research Programmer/Analyst will extract patient records from the electronic medical records. The Research Programmer/Analyst will devote 1.2 C/M in Year 1.

Robert Strenn – Database Analyst/Programmer: Within the BIRC service center, Mr. Strenn will merge biological data files with clinical data files. Mr. Strenn will devote 0.36 C/M in Year 1 and 0.12 C/M in Year 5.

Elisha Stefanski – Sr. Research Associate: Within the IRDL service center, Ms. Stefanski will be responsible for pulling and plating samples. Ms. Stefanski will devote 1.2 C/M in Year 1 and 1.32 C/M in Year 2.

Terrie Kitchner – Research Coordinator Associate: Ms. Kitchner will handle administrative and regulatory paperwork submissions. Ms. Kitchner will devote 1.2 C/M in Year 1.

OTHER EXPENSES:

Laboratory Supplies: $3,396 in funds are requested in Years 1 & 3 for LIMS-based pulling of samples, plates, reagents, tips, gloves, etc.

Biomarker Professional Services: Funds are requested in Years 1 & 2 to support assays and laboratory work. The major expenses are listed below:

Eve Technologies	HD42	Human Cytokine Array /Chemokine Array 42-Plex with IL-18	$75,200.50
Eve Technologies	HDMET9	Human Metabolic Hormone Array 9-Plex	$33,299.50

Antibody Professional Services: Funds to the amount of $61,000 in Year 2 and $30,500 in each of Years 3 & 4 are requested for the Antigen Discovery human autoimmune/pathogen antigen screen, labor, analyses, and scientific interpretation.

Xiaowu Liang, PhD – Co-Investigator, Antigen Discovery, Inc.: Dr. Liang will aid in the coordination of sample processing, experimental design for the Antigen/Antibody screen, and aid in analyses and scientific interpretation. Dr. Liang will aid in manuscript preparation. Dr. Liang's effort will be cost-shared by Antigen Discovery, Inc.

Krista Trappl-Kimmons, PhD – Co-Investigator, Antigen Discovery, Inc.: Dr. Trappl-Kimmons will aid in the coordination of sample processing, experimental design for the Antigen/Antibody screen, and aid in analyses and scientific interpretation. Dr. Trappl-Kimmons will aid in manuscript preparation. Dr. Trappl-Kimmons's effort will be cost-shared by Antigen Discovery, Inc.

Subcontractor/Contractual: Funds are requested in Years 2-5 for a subcontract with Marquette University, Milwaukee, Wisconsin for Dr. Mehdi Maadooliat as a consultant who will work with Dr. Schrodi providing statistical and machine learning Expertise.

Mehdi Maadooliat, PhD – Assistant Professor of Mathematics, Statistics and Computer Science, Marquette University: Dr. Maadooliat also has an Associate Research Scientist position at the Center for Human Genetics, MCRI where he works closely with Dr. Schrodi. Dr. Maadooliat will provide expertise in statistical approaches to analysis of high dimensional data, dimensional reduction methods, and machine learning analyses. Dr. Maadooliat will develop novel methods using binary PCA and generalized linear mixed models to analyze the data for this study. Dr. Maadooliat will mentor the Marquette University graduate student in the Department of Mathematics, Statistics and Computer Science. Dr. Maadooliat will assist with the preparation of manuscripts for publication.

Subcontractor/Contractual: Funds are requested in all Years for a subcontract with UW-Madison, Madison, Wisconsin for Dr. Judith A. Smith as a consultant who will work with Dr. Schrodi providing clinical and immunology expertise.

Judith Smith, MD, PhD – Tenured Associate Professor/Consultant: Dr. Smith will provide mentorship for Dr. Schrodi, particularly in the areas of clinical phenotypes and immunology. Dr. Smith work with Dr. Schrodi in interpreting ICD9 codes, clinical laboratory data, imaging data and

relevant patient data. She will also participate in interpretation of inflammatory marker expression results and genetic findings. Dr. Smith will provide clinical and immunological expertise for the study. Dr. Smith will meet with Dr. Schrodi several times a year and contribute above expertise through regular phone calls and e-mail exchange.

INDIRECT COST

On **December 1, 2016**, our institution entered into an agreement with the Department of Health and Human Services authorizing the use of a modified total indirect cost rate for research grants, contracts, and other agreements with the Federal Government. This predetermined rate of **60%** is effective **10/01/16 through 09/30/19**, which is also our provisional rate until our current agreement is amended. The indirect cost base is total direct costs excluding equipment, capital expenditures, charges for patient care, student tuition remission, rental costs of off-site facilities, scholarships, and fellowships as well as the portion of each subcontract in excess of $25,000.

MARQUETTE UNIVERSITY

Statement of Work
Mehdi Maadooliat, Ph.D., Assistant Professor of Mathematics, Statistics, and Computer Science at Marquette University, Co-Investigator (8.3% effort; 1 summer month per year) will provide expertise on the statistical design of the study in year 2. In years 3-5, Dr. Maadooliat's focus will be on modeling and data analysis.

Budget

Budget Item	5/1/20-4/30/21	5/1/21-4/30/22	5/1/22-4/30/23	5/1/23-4/30/24	Total
Marquette University Co-Investigator, Dr. Mehdi Maadooliat – 8.3% effort; 1 summer month	10,153.18	10,457.78	10,771.51	11,094.66	42,477.13
Graduate Research Assistant		27,135.51	28,788.06	29,651.70	85,575.27
Fringe Benefits-16% on faculty summer salary	1,624.51	1,673.24	1,723.44	1,775.15	6,796.34
Fringe Benefits-7.65% on summer salary for student		479.05	508.22	523.47	1,510.74
Graduate Tuition, 14 credits		17,290.00	17,808.70	18,342.96	53,441.66
Total Direct Costs	$11,777.69	$57,035.58	$59,599.93	$61,387.94	$189,801.14
Indirect Costs, 51% of MTDC	6,006.62	20,270.25	21,313.53	21,952.94	69,543.34
Total Request	$17,784.31	$77,305.83	$80,913.46	$83,340.88	$259,344.48

BUDGET JUSTIFICATION

Senior/Key Personnel
Mehdi Maadooliat, Ph.D., Assistant Professor of Mathematics, Statistics, and Computer Science at Marquette University, Co-Investigator (8.3% effort; 1 summer month per year) will provide expertise on the statistical design of the study in year 2. In years 3-5, Dr. Maadooliat's focus will be on modeling and data analysis. Dr. Maadooliat has a nine-month academic base salary. Each full-time summer month is budgeted at one-ninth of his nine-month academic base salary.

Other Personnel
Graduate Research Assistant (GRA) – A GRA will be recruited in years 3-5 to assist Dr. Maadooliat with the statistical design, testing, modeling and data analysis. This individual will devote 20 hours per week throughout each year. Tuition is a part of the standard compensation package and is budgeted in the Other category.

Fringe Benefits
Marquette University's fringe benefits package includes a rate of 28.5% of full-time 9 or 12-month salaries, 16% of summer salary for 9-month employees and 8% of part-time salaries and summer salaries for students. The full package includes health, life, dental and disability insurance; TIAA-CREF; social security and workers compensation payroll taxes.

Other Direct Costs
Tuition – As part of the customary compensation package for GRAs at Marquette University, 14 graduate tuition credits are budgeted in years 3-5 and will increase by 3% each year.

Indirect Costs
Marquette University's federally approved indirect cost rate is 51% of Modified Total Direct Costs (MTDC). MTDC is total direct cost excluding capital expenditures, the portion of each subaward or subcontract in excess

of $25,000 and student tuition. Marquette University's indirect cost rate was established with the U.S. Department of Health and Human Services (DHHS) on March 28, 2014.

PHS 398 Research Plan

OMB Number: 0925-0001
Expiration Date: 03/31/2020

Introduction	
1. Introduction to Application (for Resubmission and Revision applications)	1246-INTRODUCTION TO APPLICATION Schrodi 02Jul2018.pdf
Research Plan Section	
2. Specific Aims	1247-SpecificAims_SchrodiSJ_02Jul2018.pdf
3. Research Strategy*	1248-RESEARCH STRATEGY_Schrodi_02Jul2018.pdf
4. Progress Report Publication List	
Other Research Plan Section	
5. Vertebrate Animals	
6. Select Agent Research	
7. Multiple PD/PI Leadership Plan	
8. Consortium/Contractual Arrangements	1249-Consortium Contractual Arrangement.pdf
9. Letters of Support	1250-Letters of Support.pdf
10. Resource Sharing Plan(s)	1251-Resource_Sharing_Schrodi_RO1_02Jul2018.pdf
11. Authentication of Key Biological and/or Chemical Resources	
Appendix	
12. Appendix	

Tracking Number: GRANT12668618
Funding Opportunity Number: PA-18-484. Received Date: 2018-07-03T18:44:57.000-04:00

INTRODUCTION TO APPLICATION

We appreciate the review of our previous RO1 proposal, 1 RO1 GM129598-01 on 03/01/2018. Our application is focused on the measurement of exome genotype data, circulating cytokines and metabolic proteins, and antigen/antibody profiles on several diseases obtained from the Marshfield Clinic Personalized Medicine Research Project biobank for the purpose of predicting disease states early in their course using statistical and machine learning approaches. The application was reviewed by the Biostatistical Methods and Research Design Study Section. We found the review to be useful in improving the proposal however a large number of the concerns centered on the lack of development of novel statistical methods. The original proposal and the resubmitted proposal were designed to rely heavily on established statistical methods to serve as a tool for analyzing molecular omics data for the purpose of developing predictive models useful for classifying individuals into disease states. Our overarching goal is to measure exome genotype data and dynamic biomarkers in individuals prior to disease diagnosis to create a multi-omics-based predictive model for disease.

Several other weakness were noted which we addressed in the resubmission. We clarified the feature selection procedures and added additional details on the Bayesian network and penalized logistic regression classifiers. We have used a Bayesian network to successfully generate a highly accurate model for type 2 diabetes (AUC=0.912), which was published in *Metabolism*. Further, we have used both classifiers to predict inflammatory arthritis cases. Dr. Maadooliat is an Assistant Professor in statistics at Marquette University and is well-trained in Sure Independence Screening and Elastic Net Procedures. Additionally, Dr. Schrodi has used Bayesian Networks and Elastic Net procedures for a decade at Celera to generate proprietary genetic-based diagnostics. These approaches are well-studied in the literature and commonly yield effective predictive models in application. We do wish to avoid a wide screen of numerous machine learning methods to the same data sets as that approach can generate overfit models lacking robustness.

An additional concern was the limited collaboration among team members. Dr. Guo is a new postdoctoral fellow in Dr. Schrodi's lab and they have worked closely for the past six months. They have recently generated a manuscript on a recessive diplotype scan using the PMRP biobanked samples. Dr. Maadooliat has been a faculty member in the same department as Dr. Schrodi for the past three years and they have been collaborators on several studies during his tenure at the Marshfield Clinic Research Institute. Indeed, in 2016 Dr. Maadooliat was awarded the Way Klinger Young Scholar Award for the purpose of collaborating with Dr. Schrodi at the Marshfield Clinic. Drs. Maadooliat and Schrodi have co-authored two published research manuscripts since 2016:

Maadooliat M, Bansal NK, Upadhya J, Farazi MR, Li X, He MM, Hebbring SJ, Ye Z, **Schrodi SJ** (2016) The decay of disease association with declining linkage disequilibrium: A fine mapping theorem. *Front Genet* 7:217.

Bansal NK, **Maadooliat M**, **Schrodi SJ** (2018) Empirical Bayesian approach to testing multiple hypotheses with separate priors for left and right alternatives. *Stat Appl Genet Mol Biol* 17(3):20180002.

Additionally, Drs. Maadooliat and Schrodi have conducted a genome-wide scan of shared chromosomal regions using the PMRP biobanked samples and are collaborating on a pQTL study of cytokines (manuscripts in preparation). Dr. Smith (UW-Madison) and Dr. Schrodi are also close collaborators and have worked together since 2013. Dr. Schrodi has a faculty appointment at UW-Madison. Dr. Schrodi regularly travels to UW-Madison to meet with Dr. Smith. Drs. Smith and Schrodi were awarded a grant from the Rheumatology Research Foundation and have completed the funded study on axial spondyloarthritis, yielding a scientific abstract and a publication:

Liu Y, Ye Z, Li X, Anderson JL, Khan M, DaSilva D, Baron M, Wilson D, Bocoun V, Ivacic LC, **Schrodi SJ**, **Smith JA.** (2017) Genetic and functional associations with decreased anti-inflammatory Tumor Necrosis Factor Alpha Induced Protein 3 in macrophages from subjects with axial spondyloarthritis. *Front Immunol* 8:860.

In summary, in response to this useful critique of our initial submission, we have produced an improved proposal with rigorous, established statistical methods with a fully collaborative scientific team with complementary expertise.

SPECIFIC AIMS

Early clinical interventions can often dramatically curtail the trajectory of diseases, effectively reducing morbidity in a wide array of diseases. However, frequently clinical diagnoses occur following significant, often irreversible disease progression. Early disease prediction from genetics and molecular biomarkers offers to improve the management of individuals with pre-clinical disease states. The overarching goals of this study are to develop robust, molecular-based predictive models for eight chronic, complex diseases using a combination of exome genetic data, key circulating inflammatory and metabolic proteins, and the results from an antigen screen. By doing so, we will gain insight on how these biomarkers and genetic variants contribute to the pathogenesis of these diseases. The diseases studied were selected for having considerable genetic and inflammatory/metabolic components and were well-defined in the Marshfield Clinic Biobank as part of a previous study. Using extensive longitudinal electronic health records linked to this biobank, we are able to assay biomarkers in samples obtained prior to clinical diagnosis of the diseases studied. From these multi-omics data, we will employ and develop a Sure Independence Screening approach followed by an Elastic Net to select informative signals from large molecular datasets. Bayesian Networks will be used to coalesce the predictive signals from genetic variants and protein biomarkers into disease classifiers. The diseases studied are rheumatoid arthritis, systemic lupus erythematosus, relapsing-remitting multiple sclerosis, premature myocardial infarction, chronic lymphocytic leukemia, obsessive compulsive disorder, autoimmune thyroid disease, and axial spondyloarthritis. Further, we will validate the classifer built for rheumatoid arthritis in an independent set of individuals. As the definitive clinical diagnosis of rheumatoid arthritis is typically prolonged, delaying treatment and risking irreversible impairment of mobility, chronic pain and structural damage to joints, an effective molecular prognostic will serve as a useful tool for rheumatologists. **Importantly, the methods and approaches developed for this study will inform and enable other researchers to perform complementary studies.**

Specific Aim 1: To systematically characterize the levels of 48 key circulating inflammatory and metabolic proteins and exome variants in eight common diseases and matched control groups. Additionally screen antibodies to 800 autoantigens and pathogen epitopes in rheumatoid arthritis and systemic lupus erythematosus. To understand the level of baseline systemic inflammatory activity, metabolic dysfunction and immune tolerance in eight sets of individuals with disease and controls by measuring (i) exome variants and GWAS-significant polymorphisms (ii) circulating levels of 42 widely-studied inflammatory cytokines characteristic of NFkB-signaling, T_H17 signaling, Treg activity, innate immunity, complement activity and Th1/Th2 balance, (iii) antibodies to 800 selected autoantigens and viral/bacterial antigens, and (iv) six metabolic proteins in the insulin-glucose axis in biobanked plasma collected on a 20,000-person cohort selected from a homogeneous Central Wisconsin population. For the disease groups, we will measure all genetic polymorphisms and molecular biomarkers in samples stored prior to a clinical diagnosis. We will also determine which genetic polymorphisms and molecular biomarkers exhibit statistically significant effects in distinguishing between cases and controls. ***We hypothesize that significant differences in the genetics and protein biomarkers exist between each set of cases and their respective controls.***

Specific Aim 2: To optimally classify cases and controls using genetics and circulating proteins in DNA and plasma obtained prior to clinical diagnosis. Using the genetic and protein features exhibiting disease signals, we will select informative sets of these genetic and protein biomarkers and develop a Bayesian Network classifier and Penalized Logistic Regression classifier for each of the diseases listed. Robustness of each of the disease classifiers will be assessed using cross validation procedures. ***We hypothesize that the combination of selected (i) triggering event signals from antibodies to antigens, (ii) exome variants, (iii) GWAS-significant SNPs, (iv) levels of inflammatory cytokines in plasma, and (v) levels of metabolic proteins in plasma are capable of predicting disease states in the Central Wisconsin cohort for the eight phenotypes studied.***

Specific Aim 3: To validate the predictive model for rheumatoid arthritis in an independent sample set. For the top performing rheumatoid arthritis predictive model developed in Aim2, we will measure all component genetic polymorphisms and protein biomarkers in an independent set of samples from 97 physician-identified rheumatoid arthritis individuals and 200 matched controls. We will calculate the classifier and assess the performance using standard measures of diagnostic utility. ***We hypothesize that the rheumatoid arthritis classifier has the predictive capacity and robustness to effectively delineate cases from matched controls in an independent sample set, thereby validating the classifier.***

RESEARCH STRATEGY

A. SIGNIFICANCE

Molecular-based early prediction of complex diseases is a major goal of medical research as it both provides an avenue to curtail pathologies through timely interventions and may elucidate underlying disease mechanisms. Numerous studies have demonstrated beneficial alterations of disease trajectories with early medical interventions in diseases such as systemic inflammatory conditions[1-5], diabetes[6-7], cancers[8-10], cardiovascular disease[11-13], and obesity.[14-16] The use of biomarkers, such as genetic variants, circulating cytokines, metabolic markers, and antibodies to pathogens, has shown increasing utility for disease prediction and prognosis.[17-20] Furthermore, careful examination of predictive sets of molecular biomarkers can highlight underlying perturbations in critical pathways involved in pathogenesis.

The overarching goal of this study is to use powerful machine learning techniques to create predictive models for eight common diseases using a combination of exome data, circulating cytokines, metabolic markers, and endogenous antibodies to selected epitopes (for two diseases). Our long-term aim is to produce clinically useful, molecular-based predictive models of diseases and to show the utility of using large biobanks linked to electronic medical record information. By screening multiple diseases in our population-based biobank, we increase the likelihood of identifying one or more diseases that exhibit excellent predictive accuracy that will serve as the basis for subsequent replication and translational investigations. The exome data will capture signals from coding variants across the genome that confer increased inherited susceptibility to the diseases studied. Pivotal signaling in the immuno-metabolic axis will be measured in the circulating proteins selected. Lastly, autoantibodies representing disruption in immune tolerance and antibodies to infectious agents that may trigger pathogenesis will be interrogated in the multiplexed antigen assay. The diseases interrogated in this study are: **rheumatoid arthritis**, **systemic lupus erythematosus**, **relapsing-remitting multiple sclerosis**, **premature myocardial infarction**, **chronic lymphocytic leukemia**, **obsessive compulsive disorder**, **autoimmune thyroid disease**, and **axial spondyloarthritis**. Importantly, we are proposing to interrogate biological samples collected prior to clinical diagnosis. Hence, this study has the strong potential for aiding wide array of individuals through providing predictive information early in disease processes where medical interventions are the most efficacious.

A1. Importance of results from this study: The development of molecular-based predictive models for common diseases would represent a dramatic advancement in medical diagnostics, enabling prompt therapeutic intervention and potentially curtailing morbidity. High throughput molecular screens coupled with advanced statistical and computational approaches have set the stage for assessment of the viability of such predictive models. Several valuable results will be generated from this study: (i) we will determine the diagnostic efficacy of classifiers for eight common diseases using the combined signals from genetic data, circulating cytokines, metabolic markers, and antibodies to autoantigens and pathogens, (ii) we will gain understanding of the role of these various molecular markers in each disease studied, (iii) we will elucidate the proportion of disease-affected individuals who harbor predictive molecular signals, (iv) by assaying molecular panels prior to clinical diagnoses using biobanked samples, we will help delineate the molecular changes that precede clinical diagnosis, and (v) the study results may suggest specific biological pathways that play a role in the etiology of these diseases.

A2. Critical barriers: The large majority of common diseases have been relatively resistant to prognostic efforts. Four important impediments have conspired to hinder advances in the development of viable predictive models for risk of the diseases proposed to be studied: 1) The use of dynamic biomarkers assayed in samples collected after clinical diagnosis has produced signals that can be driven by advanced disease processes and therapeutic interventions; 2) The use of a single type of biomarker (e.g., solely genetic markers) has generally produced classifiers with insufficient diagnostic utility; 3) The heterogeneity in clinically-defined phenotypes reduces the efficacy and applicability of predictive models and 4) The use of genetic-based predictive models that are restricted to GWAS-significant markers undermines the performance of such models for common diseases.[21]

Prediction of disease onset using banked samples and interrogating a large combined set of inflammatory cytokines, metabolic markers, exome variants, and circulating antibodies has not been performed due to the lack of availability of such data, previous technological impediments and complexity of analyses. The difficult access to large biorepositories with linked, longitudinal electronic medical records, high-throughput protein-based assays, and exome genotyping data has also substantially hindered advancement of disease prediction studies. In addition, the analysis of data generated by a large-scale study using samples with complicated medical record information and high density genomic/protein/antibody measurements requires

expertise in complex statistical and machine learning methods as well as highly developed informatics skills. Assembling a seasoned team with expertise in these areas is difficult.

A3. Overcoming the critical barriers: (1) To better determine biomarkers that are antecedent to disease onset, we will use biobanked DNA and plasma collected prior to clinical diagnosis. The use of the Personalized Medicine Research Project (PMRP)—a 14 year-old biobank of DNA, plasma and serum from 20,000 adults with linked electronic medical records—at the Marshfield Clinic enables us to determine sets of individuals with biological samples collected before a physician diagnoses an individual with disease. (2) To harness orthogonal signals across molecular data types, we will interrogate the combined signals from (i) exome-wide genotype data, (ii) circulating inflammatory cytokines and metabolic markers, and (iii) the presence of antibodies to a variety of autoantigens and pathogen epitopes; (3) Using electronic medical record information consisting of laboratory test results, diagnostic codes and medications along with the genetic, protein and antibody data, we will perform dimensional reduction techniques such as PCA and binary PCA designed to identify significant subgroups within a heterogeneous disease state. (4) To capture both GWAS-significant and sub-GWAS-significant predictive signals from the genome, we will use the entire exome genotype array data measured on individuals in the study.

The Marshfield Clinic Research Institute (MCRI) has collected, maintained, and enhanced the PMRP, which has played an integral role in numerous disease susceptibility studies.[17,22-27] The Integrated Research & Development Laboratory, a service center at MCRI under the direction of Dr. Meece, has implemented a high-throughput pipeline for utilization of plasma samples from the PMRP. Dr. Schrodi has exploited this resource in quantifying inflammatory cytokines on >2,000 samples and immune-metabolic markers on >600 samples, resulting in a high utility Bayesian Network classifier for type 2 diabetes in samples obtained prior to clinical diagnosis (AUC=0.91). The MCRI houses the Biomedical Informatics Research Center, a service center with considerable computational infrastructure consisting of a research data warehouse built to store health and biological data and the High Performance Computing Cluster designed to perform complex analyses of high dimensional data. Preliminary queries have been performed by this Center under the guidance of Dr. Schrodi to identify individuals satisfying a rigorous set of criteria involving ICD codes and clinical laboratory measurements for each disease studied. Dr. Schrodi has 17 years of post-PhD experience performing human genomic experiments and statistical analyses on genetic and biological data using Bayesian methods, machine learning algorithms, permutation routines, Monte Carlo simulations, and various frequentist tests. He has designed and managed several large-scale genetic mapping studies of inflammatory diseases (~$1.6M/study) in the biotechnology industry for over a decade. Recently, using the PMRP samples, he has conducted studies of genetic variants underlying cytokine expression and recessive diplotypes across the diseases proposed in this study.

Dr. Judy Smith, a collaborator with Dr. Schrodi on several studies including a Rheumatology Research Foundation-funded study on ankylosing spondylitis, has considerable clinical expertise and expert knowledge of immunology. Dr. Smith is a tenured associate professor in the University of Wisconsin-Madison School of Medicine & Public Health. Drs. Smith and Schrodi recently published the results of a study showing how genetic variants and expression of Tnfaip3 contribute to differential macrophage responses in individuals with ankylosing spondylitis compared to controls.[28]

Dr. Shicheng Guo has also collaborated with Dr. Schrodi. Dr. Guo has training and expertise in human genetics and computational genetics. Dr. Guo has conducted research in genetic analyses, epigenetics, computational biology, and bioinformatics. Dr. Guo is currently a postdoctoral fellow in Dr. Schrodi's laboratory.

Drs. Trappl-Kimmons and Liang have considerable expertise in developing, assaying and analyzing high-throughput antigen/antibody assays.

Dr. Maadooliat, a tenure-track assistant professor in Mathematics, Statistics and Computer Science at Marquette University, is also a close collaborator with Dr. Schrodi. Dr. Maadooliat has spent the last three years working from the Center for Human Genetics at the Marshfield Clinic Research Institute where he is a faculty member, sponsored by Dr. Schrodi. Dr. Maadooliat has considerable expertise in statistical/computational methods applied to protein structure and functional data analysis. Drs. Maadooliat and Schrodi have jointly conducted research in the areas of linkage disequilibrium patterns from disease loci[29], Bayesian multiple testing[30], and identifying disease susceptibility genes through shared chromosomal regions. Two published manuscripts have resulted from the collaboration.

A4. Scientific knowledge gained: The development of predictive models for one or more common diseases would provide substantial insight into how molecular intermediate phenotypes can be used to predict disease states and could significantly facilitate early treatment of those conditions. Importantly, (1) This study will

evaluate each disease against matched controls and test for significant signals from genetic markers, circulating proteins and antigen/antibody reactivity profiles. Doing so will provide evidence for or against known and novel molecular markers in correlation with each disease entity and identify specific pathways that contribute to disease etiologies. (2) This study will characterize the level of classification performance for molecular-based predictive models. (3) This study will show the relative contributions of genetics, key cytokines and metabolic proteins, and antibodies to the prediction of diseases; and (4) We will assess whether the panel of molecular markers partitions patients into significant subgroups within any of the diseases investigated. Importantly, we will validate an initial cross-validated predictive model for rheumatoid arthritis in an independent set of rheumatoid arthritis and matched controls. In potential situations where we cannot develop a viable, cross-validated predictive model, then we will have generated evidence of either the lack of classifier signal from the combined action of exome genetics, dynamic biomarkers and antibodies for our sample sizes and disease definitions and/or the machine learning algorithms employed. Such negative results would suggest that new approaches to disease prediction are needed.

A5. Potential implications: There has been a rapid adoption of the use of large patient populations with linked electronic medical records for the purpose of understanding the molecular pathogenesis and pathophysiology across a spectrum of disease phenotypes. This work will demonstrate how studies utilizing these resources can serve to better understand how to create predictive models for diseases. Once predictive models for early disease states are validated, implementation in clinical settings will enable physicians to identify individuals early in their disease course and provide timely therapeutic interventions.

B. INNOVATION

The proposed study has several important aspects that, in combination, result in a highly innovative study. Our group is well-positioned and experienced to harness of three key advances to create molecular-based predictive models for the diseases studied here. (i) We have access to 20,000 biobanked samples with linked longitudinal electronic medical records (PMRP). As the patients for the biobank were recruited over 14 years ago and the average individual has in excess of 34 years of electronic medical record data, these samples enable us to utilize molecular profiles from samples obtained prior to disease diagnosis. (ii) The recent and explosive growth of high-throughput genomic and proteomic technologies has enabled cost-effective large-scale screens of important types of molecular markers. The exome genotyping array, applied to the PMRP samples, interrogates the large majority of genome-wide significant SNPs, identify-by-descent polymorphisms, ancestry informative SNPs, and ~250k nonsynonymous coding variants. The Luminex multiplex immunoassay technology enables the concurrent measurement of dozens of key inflammatory cytokines, chemokines and metabolic proteins with high sensitivity and wide dynamic range. The Antigen Discovery platform provides a method to interrogate circulating antibodies to hundreds of different autoantigens and pathogen epitopes. Lastly, (iii) the recent advances in computational statistics and machine learning algorithms allows for the rigorous development of robust and accurate predictive models from high dimensional data sets. The team assembled will consolidate all three aspects into a study that dives deeper into extensive molecular profile data to produce predictive models of early disease using data obtained prior to diagnosis.

B1. Shift in current research paradigm through novel approaches

Often, with dynamic biomarkers such as the inflammatory cytokine and metabolic protein panels proposed here, retrospective studies can face difficulties in delineating the effects of molecular pathogenesis from effects generated by the disease process and/or medical treatment. This issue hinders the mechanistic interpretation of biomarkers and may reduce the robustness of predictive models using these markers. The experimental design proposed here is quasi-prospective in that plasma samples from the biobank were obtained ~14 years ago and subsequent disease diagnosis occurred in a substantial fraction of those with the diseases studied. Hence, we are restricting our analyses to those cases that had samples collected prior to clinical diagnosis. While not truly a prospective design, this design reduces identifying case/control molecular differences that result from disease treatment or the molecular perturbations that occur with advanced disease states. By adopting such an experimental design, we are better positioned to interrogate molecular effects in dynamic biomarkers more likely to underlie pathogenesis. While still correlative, we argue that results from this quasi-prospective approach yield more reliable evidence for causal relationships than purely retrospective designs. We have generated data for metabolic dysfunction using this design, interrogating immuno-metabolic markers. Our results show that, for example, C-peptide is dramatically increased (54%, on average) in individuals who were undiagnosed with type 2 diabetes at the time of measurement, but have subsequently transitioned to type 2 diabetes following enrollment in PMRP (P=1.18E-13).

Another advance that this study employs is the concurrent measurement of 'multi-omics' data on individuals studied. For complex diseases, sole reliance on single types of molecular data can produce limited results for predictive models. For example, reliance on GWAS-significant SNPs for disease prediction has repeatedly shown limited diagnostic utility.[21] By expanding the investigation to include key inflammatory cytokines, critically important metabolic markers, antibody binding to hundreds of antigens, and exome data, we harness predictive signal from disparate biochemical processes and susceptibility factors. This technique has recently produced several insights into various clinical traits, including hepatocellular carcinoma[32], Alzheimer's disease[33], and optimal nutrition.[34] As preliminary data for our ability to assay protein levels and measure genome-wide genetics, we measured 9 inflammatory cytokines in stored plasma samples of over 2,000 PMRP individuals. Our results show, for example, that (i) a polymorphism in the *LAIR1* gene, an inhibitory receptor expressed on PBMCs and reportedly involved in systemic lupus erythematosus[35], is strongly associated with IL-17A levels (P=1.33E-08). Further, we have shown interesting co-regulatory cytokine patterns, such as IL-10 levels are highly correlated with IL-12p70 levels (P-value Spearman's rho = 1.91E-67).

Moreover, our predictive models will use all exome array data instead of focusing on simply the GWAS-significant markers. Theoretical and empirical research in livestock and crop phenotypic traits and breeding values had suggested that using genome-wide SNP data with generalized linear mixed models (e.g., best linear unbiased prediction) approach has considerable merit.[36-38] Building on this work, researchers in human genetics have provided evidence that the prediction of traits in humans is well-informed by using all SNPs across the genome.[39-41] Logistic regression with adaptive Lasso (one of the two classification methods proposed) is an approach that can utilize all of the data on the exome array in combination with the circulating protein and antibody data.

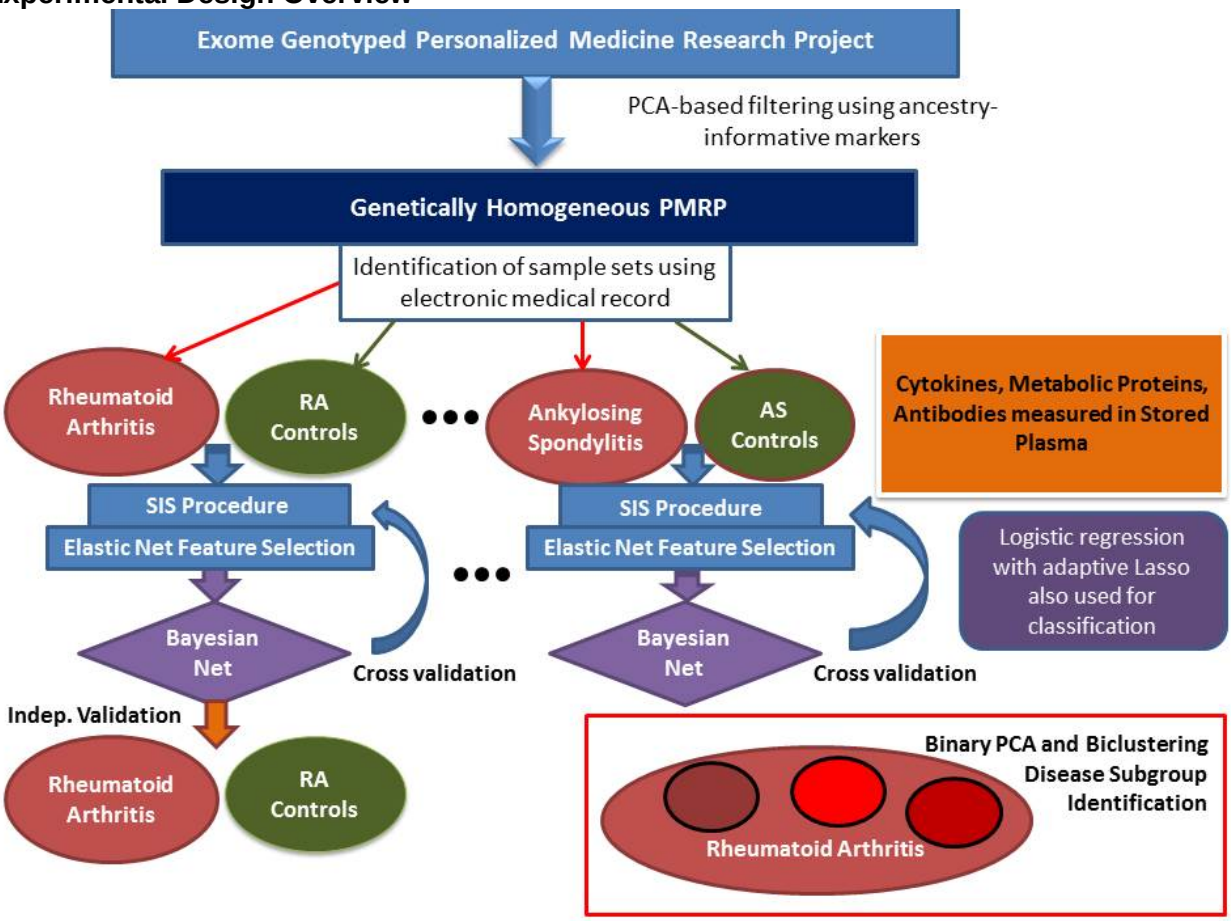
B2. Shift in current research paradigm through methodology

Several types of statistical methods and machine learning algorithms will be used in this study. Three researchers involved in the study, Drs. Maadooliat, Schrodi and Guo, have expertise in the development of a variety of quantitative methods including Bayesian networks and regression methods, multiplicity approaches, transforms, permutation methods, functional data analysis, Monte Carlo simulations, clustering methods, resampling methods, and general linear mixed models.[29,30,42-52] Aside from the use of standard methods across the study, three areas of methodology research will be investigated: 1) Due to heavy reliance on observations of gross physiology in a clinical setting, the correspondence between molecular pathogenic effects and clinical disease states is not a one-to-one mapping for common, complex diseases. Indeed, there are multiple examples of molecular etiologies that were previously classified as single disease entities (e.g., breast cancer can meaningfully subdivided into estrogen receptor positive/negative disease, with profound implications for prognosis, trajectory, and treatment decisions).[53-55] To better identify molecular-based disease subgroups within each disease examined, we will use and extend binary PCA and biclustering methods.[56,57] 2) The use of multiple 'omics' data types in predictive models necessitates sophisticated machine learning approaches. We will refine and tailor a Sure Independence Screening (SIS) procedure for priming data for feature selection from these molecular data. SIS is a simple and elegant tool that is widely used to reduce the dimensionality of data to a manageable scale.[58] It is well known that the screening and commonly used penalties (e.g., adaptive Lasso, Elastic net, and SCAD) can produce valuable results for variable selection in high dimensional frameworks. SIS and its extensions are useful techniques to improve the computational efficiency of these feature selection algorithms, as well as the accuracy in extracting informative signals from very large biological datasets. Using the SIS/Elastic Net-determined features, we will construct Bayesian Network-based classifiers[59,60] for use on the generated exome genotype data, protein biomarkers, and antibody data. Bayesian Networks are widely considered to produce highly competitive predictive performance when compared to other approaches. In addition, Bayesian Networks can (i) be easily motivated by first principles of probabilistic laws applied to graphs, and (ii) produce highly transparent results showing the level at which each feature drives the classification and the correlative structure between features. In addition to Bayesian Networks, we will also evaluate high dimensional logistic regression with adaptive Lasso as an alternative classifier. 3) Characterizing results from machine learning classifiers in a way that both captures the performance of the prediction as well as the clinical utility is challenging. While we will use the standard area under the Receiver Operating Characteristic curve (AUC)[61] (by averaging AUC values over 10-fold cross validations) to evaluate classifier performance, we will also explore the use of the posterior probability of disease (PPD) (probability of disease conditioned on the molecular features used in the predictive model), where we investigate metrics that compare the probability mass of the PPD that exceeds a high threshold and the probability mass of PPD below a low threshold.[21] We believe this type of analysis method both measures the performance of a classifier and

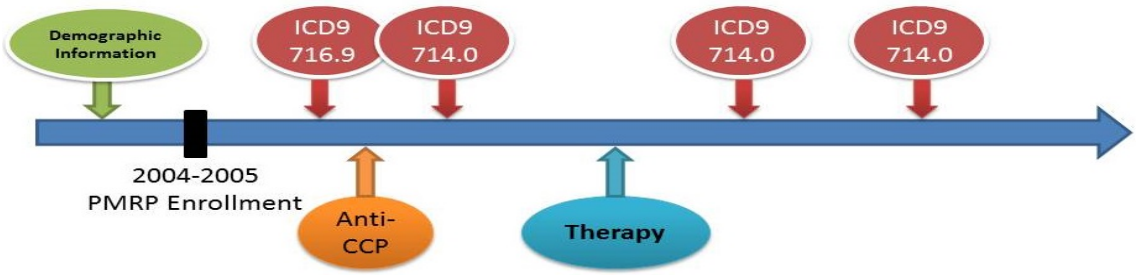
has direct applicability in a clinical setting as physicians often tailor clinical decisions based on whether they believe an individual's probability of disease exceeds a particular value.

C. APPROACH

Experimental Design Overview

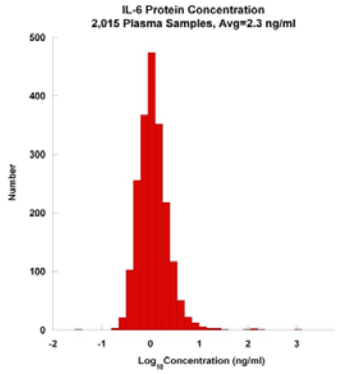


The first step in the experimental design is to utilize the panel of ancestry informative markers on the exome beadchip to conduct a PCA analysis removing genetic-background outliers within the set of 20,000 PMRP individuals. From the resulting set of individuals, we will identify individuals from each of the diseases studied using the electronic medical record information. We have found that for many diseases using solely ICD9 codes can often generate inaccurate identification (i.e., low concordance rates with manual adjudication from clinical professionals). Therefore, multiple incidences of ICD9 codes (not simply one instance of a code) coupled with laboratory test, prescription data, and/or imaging data is used to identify cases. The impact of the number of ICD9 instances on the power of genetic association studies has been previously investigated and has informed this work.[62] Much of the identification of cases from the eight diseases investigated has been performed by as part of a different study from Schrodi, Maadooliat and Guo. Manual adjudication of cases defined in this manner has seen very high validation rates (100% for multiple sclerosis and >96% in other phenotypes). Further, these have been used in genetic studies where known positive controls (e.g., HLA-DRB1 in RA) are significantly associated. Controls will be defined as those individuals without any evidence of the disease studied or similar diseases and matched based on age and sex. Dr. Smith will provide clinical expertise aiding in determining case/control definitions. Below is an example timeline of the medical trajectory of an individual diagnosed with RA:

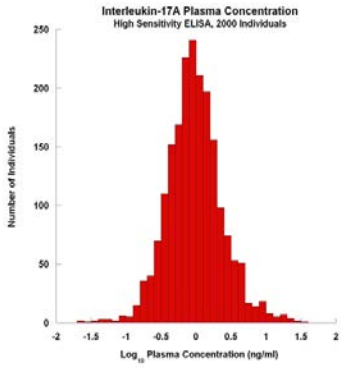


Importantly, we will restrict the case definitions to those with clinical diagnoses occurring after sample collection/enrollment in PMRP. Once case/control sets are identified, stored plasma will be pulled and plated for interrogation for cytokines/chemokines and metabolic proteins on two Luminex multiplex assays by Eve Technologies (the 42-plex Human Cytokine Array/Chemokine Array and the Human Metabolic Hormone array 9-plex). To correct for total protein level variation across samples, we will run an inexpensive BSA-based total protein assay. This has already been accomplished for 1,000 PMRP samples where moderate differences were observed for results obtained under total protein-adjusted cytokine concentration values compared to unadjusted cytokine levels. Genes identified using the BSA-adjusted data were more likely to fall in pathways that corresponded to what is known about the biology of these cytokines. Preliminary data on IL-6 and IL-17A cytokine concentrations for 2,015 PMRP samples are shown below.

IL-6 distribution



IL-17A distribution



In addition, stored plasma for rheumatoid arthritis, systemic lupus erythematosus and controls will be sent to Antigen Discovery Inc. (ADI) to interrogate antibody reactivity to 800 autoantigens and pathogen antigens. ADI has produced a high throughput protein microarray for the identification of immune signatures of humoral response to hundreds of proteins on a single assay. The microarray is produced by coupled *in vitro* transcription/translation of pathogen and human genetics, followed by printing onto nitrocellulose bonded to glass slides. ADI has demonstrated the utility of this platform. Only a few microliters of plasma are necessary for screening. Proteins include established autoantigens & pathogen epitopes. Proteins with no known immune reactivity serve as technical controls. Using a fluorescence scanner, data from the microarrays is acquired and checked for quality, followed by statistical analysis. Drs. Liang and Trappl-Kimmons from ADI will be Co-Investigators on this study.

Central Wisconsin Population and the PMRP

The genetically homogeneous population in rural Central Wisconsin is the source population for the PMRP.[63] All PMRP participants are adults. This highly stationary population is largely derived from Bavarian migrants from the late 1800s and carried high utility for genetic and biomarker studies through avoidance of confounding by population stratification and reduction in allelic heterogeneity. Environmental exposures are through to be relatively uniform across this population and migration rates are low. The PMRP has been effectively used in numerous genetic and clinical research studies including the NHGRI-funded eMERGE network.[64,65]

Plasma Collection

30mL (3x10mL K_3EDTA purple-top tubes) of blood was collected for plasma from each PMRP participant. Tubes were immediately placed on ice or refrigerated before processing. Within 24 hours, plasma cryovials were stored at -80^{o}C.

Selection of Study Individuals

PMRP samples were collected ~14 years ago. ~19,000 individuals were recently genotyped on the Illumina Human Exome + Human Core Exome Beadchips. This genotyping platform has excellent coverage of coding variants across the exome and also has GWAS-significant markers. Sample randomization was used for plating samples for genotyping. The exome genotyping beadchips used have ancestry informative marker (AIM) content. The Schrodi Lab used this set of AIMs to run a standard PCA on the full set of 19,000 individuals and genetic background outliers (~180 individuals) based on the first 3 PCs were removed from subsequent analyses. In an attempt to diminish effects from transient inflammation stimulated from acute infection, only individuals without evidence of acute infection +/- 2 weeks of PMRP enrollment will be included. Evidence of acute infection consisted of either a temperature reading of >37.7 ^{o}C, an hs-CRP lab test >3.0, or an abnormal white blood cell count of <4.3 or >10.8 within this time duration. 249 individuals satisfied one or more of these acute infection conditions and will be excluded from the study. Similarly, a record of immunization within 2 weeks of PMRP enrollment will be an exclusion criterion. Record of immuosuppressive drug prescription within 3 months prior to enrollment will also be an exclusion criterion.

Disease Phenotypes

Predictive models each of eight diseases will be developed for this study. Cases and Controls will be selected from the PMRP biobank using the electronic medical record, inclusive of ICD9/ICD10 codes, laboratory tests and medications. Over the past 4 years, Dr. Schrodi has consulted with several physicians at the Marshfield Clinic and UW-Madison to refine definitions of these 8 diseases. Using the CDC statistics, we estimate that over 4% of the United States population suffers from one or more of the diseases examined. The table below gives the results of informatics queries on the estimated number of individuals satisfying the disease criteria:

Disease	Estimated Number with Samples Prior to Dx
Rheumatoid Arthritis	**194 total (97 for Classifier Development / 97 for Validation)**
Systemic Lupus Erythematosus	**86**
Multiple Sclerosis	**31**
Premature Myocardial Infarction	**88**
Chronic Lymphocytic Leukemia	**84**
Obsessive Compulsive Disorder	**112**
Autoimmune Thyroid Disease	**95**
Axial Spondyloarthritis	**30**

An *a priori* guess at the effect size from the combined effects of the multi-omics data investigated here is extremely difficult. Hence, power calculations may not be informative. We note that for our pilot study of type 2 diabetes, 56 pre-T2D cases and 445 matched controls yielded an average cross-validated AUC of 0.912 using our elastic net logistic regression feature selection coupled with a Bayesian Network classifier (feature selection was included in the cross validation procedure). This impressive result suggests that the sample sizes for this study should be sufficient to produce accurate predictive models in at least 6 of the 8 diseases.

Rheumatoid arthritis. Rheumatoid arthritis is a debilitating systemic autoimmune disease targeting the synovium, but affecting numerous organ systems. There is evidence that rheumatoid arthritis may be triggered in susceptibility individuals by exposure to one or more pathogens. To identify individuals with a diagnosis of rheumatoid arthritis within the PMRP, we selected patients satisfying all of the following criteria: (i) three or more instances of ICD9 714.0 for rheumatoid arthritis, (ii) a positive rheumatoid factor or anti-CCP antibody test, and (iii) having received at least one prescription of methotrexate and/or anti-TNF medication. Manual spot-checking of the medical records of the selected individuals validated the approach. In addition, we extracted exome genotype data from these individuals and compared against controls. A gene-based test was applied to these genetic data to test for compound heterozygosity effects as part of a related pilot study. The positive control *HLA-DRB1* was clearly significant, following correction for multiple testing. Controls for rheumatoid arthritis will be composed of PMRP individuals without any instances of ICD9 codes for systemic inflammatory diseases, arthropathies, and similar conditions. We expect to use 194 RA cases and 400 controls, to be split evenly into a 97/200 training set and 97/200 test set (to be used as verification of the top performing classifier, SA3).

Systemic lupus erythematosus. Multiple instances of the 710.0 code will be used in conjunction with a positive lupus anticoagulant, ACA, anti-ENA, or ANA test results will define cases. We will use controls matched on age and sex and without any autoimmune disease diagnoses or chronic inflammatory diseases. We will select twice the number of cases as controls.

Multiple sclerosis. Multiple sclerosis is also a systemic autoimmune disease characterized by demyelination of the central nervous system. Motility impairment, vision loss, pain, muscle weakness and fatigue are all frequent symptoms of the disease. Within the Marshfield Clinic system, we have found excellent accuracy of the ICD9 codes and imaging data for multiple sclerosis. Indeed, of a sample of 20 PMRP individuals determined to have multiple sclerosis from their ICD9 code information, all 20 were verified by a Marshfield Clinic neurologist to have relapsing-remitting multiple sclerosis from chart review. We will use 100 controls matched on age and sex for multiple sclerosis.

Myocardial infarction. Any 410 ICD9 code (410, 410.x, 410.xx) coupled with a high-sensitivity cardiac troponin T (hs-cTnT) test greater than 51 ng/L within 1 week of the ICD9 code. Earlier age of onset will be prioritized for inclusion in the case group. The top priority group for inclusion will be males ages <50 and females <55. Between 400 and 500 controls, matched on age and sex, will be used for myocardial infarction.

Chronic Lymphocytic Leukemia. CLL cases are defined as PMRP individuals with two or more instances of the 204.1 ICD9 code and prescriptions for relevant chemotherapies/immunotherapies (e.g., rituximab).

Controls will be matched on age and sex and not have any neoplasm diagnoses or systemic inflammatory diseases. We will select twice the number of cases as controls.

Obsessive compulsive disorder. Two or more instances of the 300.3 ICD9 code will serve at the criterion for case inclusion. We will use 200 age and sex-matched controls without any psychiatric disease diagnosis.

Autoimmune thyroid disease. Multiple instances of the 245.2 or 242.0 ICD9 code coupled with abnormal TSH, anti-TPO, anti-Tg, FT4 or TSI laboratory tests will serve as the criterion for cases. We will use 200 age and sex-matched controls without any autoimmune diagnosis or systemic inflammatory disease diagnosis.

Axial spondyloarthritis. Cases will be defined as having the 720.2 ICD9 code or multiple instances of the 720.0 ICD9 code coupled with a positive HLA-B27 test and having received a prescription for anti-TNF medication or methotrexate. We will use 100 age and sex-matched controls, without any systemic inflammatory disease diagnosis.

Specific Aim 1

To systematically characterize the levels of 48 key circulating inflammatory and metabolic proteins and exome variants in eight common diseases and matched control groups. Additionally screen antibodies to 800 autoantigens and pathogen epitopes in rheumatoid arthritis and systemic lupus erythematosus.

Our approach to SA1 will be to measure (1) 42 cytokines and chemokines on an Eve Technologies (Alberta, Canada) Luminex multiplex assay in 50uL of plasma from all cases and controls selec0ted from the PMRP biobank; (2) 6 metabolic proteins on an Eve Technologies Luminex multiplex assay on 50uL of plasma from all cases and controls selected from the PMRP biobank; Eve technologies Luminex multiplex arrays have been successfully used on human samples to investigate numerous different inflammatory and infectious diseases[66-68]; and (3) All individuals studied will have the Illumina exome beadchip data previous collected and stored within the PMRP database at MCRI. Extraction of exome variants involves identifying case and control individuals, then merging the phenotype files with the exome variant file by the BIRC center at MCRI.

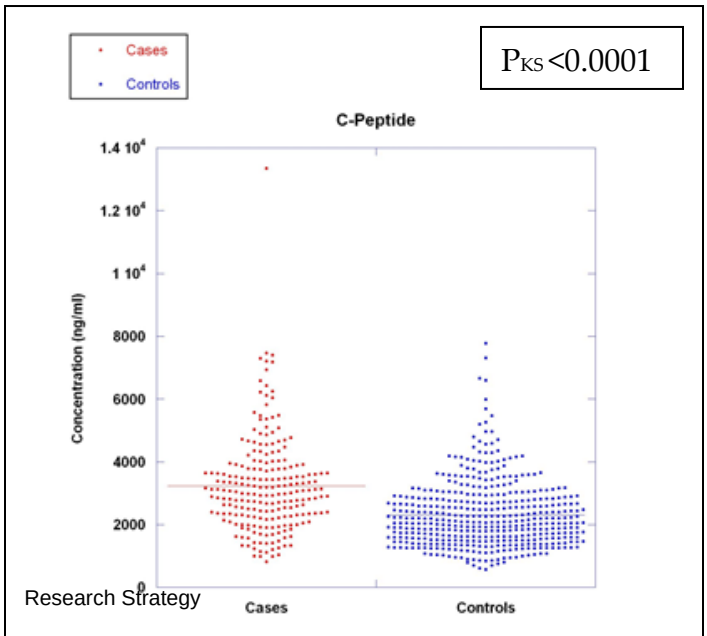
Eve Technology Human Cytokine array/Chemokine Array 42-plex: EGF, Eotaxin-1, FGF-2, Flt-3L, Fractalkine, G-CSF, GM-CSF, GRO(alpha), IFNalpha2, IFNgamma, IL-1alpha, IL-1beta, IL-1ra, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12 (p40), IL-12 (p70), IL-13, IL-15, IL-17A, IL-18, IP-10, MCP-1, MCP-3, MDC, MIP-1alpha, MIP-1beta, PDGF-AA, PDGF-AB/BB, RANTES, sCD40L, TGFalpha, TNFalpha, TNFbeta, VEGF-A

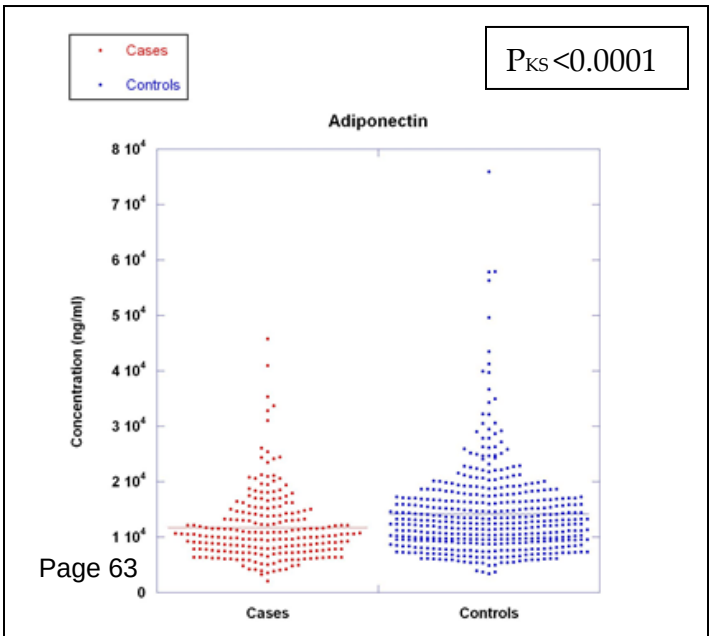
Eve Technology Human Metabolic Hormone Array: C-Peptide, GIP, Insulin, Leptin, MCP-1, PYY (total)

Lastly, (4) a custom 800 antigen high-density protein microarray panel from Antigen Discovery using plasma from 120 rheumatoid arthritis cases, 80 systemic lupus erythematosus cases and 100 shared controls from the PMRP biobank. This antigen/antibody protein microarray has been successfully used to interrogate human humoral immune response to infectious agents and autoantigen activity.[69-72]

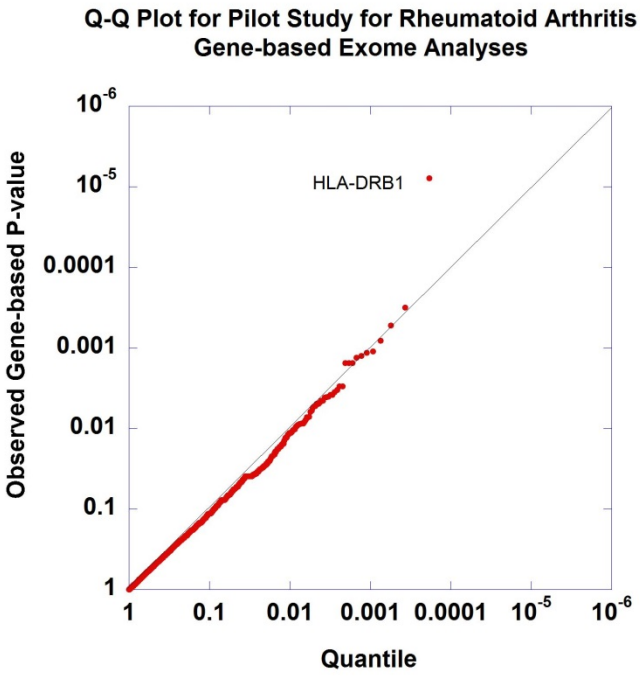
Peptide Quantification Preliminary Data

We measured C-peptide and adiponectin in a cohort of individuals without type 2 diabetes at the time of blood draw, then partitioned individuals who had met the formal criteria for diabetes in the years following the blood draw. These immune-metabolic markers were strongly correlated with prospective disease as determined by a Kolmogorov-Smirnov test.





Exome Genotyping Array Preliminary Data



We have analyzed the exome beadchip data as part of several PMRP-based studies. ~250k nonsynonymous coding variants, >5k GWAS-significant SNPs, >5k grid of common SNPs, AIM SNPs, HLA tagging SNPs, and IBD SNPs are included on the beadchip. Various QC procedures have been applied to these data including PCA on AIM SNPs to identify genetic background outliers, determination of variants that pass quality control criteria (PASS), calculation of Hardy-Weinberg exact p-values, and kinship coefficient calculations to identify cryptic relatedness. In addition, we have phased all of these data using Beagle[73] and used the resulting haplotypes to conduct scans of disease association. This Quantile-Quantile plot shows one such scan investigating rheumatoid arthritis under a recessive mode of inheritance using gene-based data. The positive control, HLA-DRB1 showed significant association.

Specific Aim 2

To optimally classify cases and controls using genetics and circulating proteins in DNA and plasma obtained prior to clinical diagnosis.

Predictive Modeling: Machine learning approaches have revolutionized the analysis of high dimensional data sets and this is particularly true with regard to developing robust classifiers. Bayesian Networks are an important, well-studied class of predictive modeling algorithms which, in general, produce accurate and transparent results. As other types of predictive models may perform well on these data, we will be flexible in our classifier approaches and also perform logistic regression with adaptive Lasso. Dr. Maadooliat has considerable expertise in employing high-dimensional logistic regression with L1 penalty procedures.

Feature Selection: To prime the molecular data for feature selection and greatly improve computational efficiency on these large data sets, we will apply Sure Independence Screening to the full set of molecular variables. Following SIS, we will use an Elastic Net logistic regression for feature selection across exome variants, GWAS-significant SNPs, cytokines, chemokines, metabolic proteins and antibodies.

Bayesian Network Classifier: Following feature selection, we will primarily use Bayesian Network classifiers to generate predictive models. Bayesian Networks are directed acyclic graphs, often used to model and analyze biological data. In essence, within the context of disease classification, Bayesian Networks calculate the PPD through Bayes theorem:

$P(D|\bigcap_{i=1}^{n} F_i) = \frac{P(\bigcap_{i=1}^{n} F_i|D)P(D)}{P(\bigcap_{i=1}^{n} F_i)}$; where F_i is the i[th] feature in the classifier and D is an indicator variable for disease status. The degree to which the joint densities are factorized depends on the covariance structure of the set of features. Applying a given threshold for the PPD enables the classification decision. Drs. Maadooliat, Guo and Schrodi have experience using Bayesian Networks through the machine learning software Weka, Mathematica, as well as the R programming language.

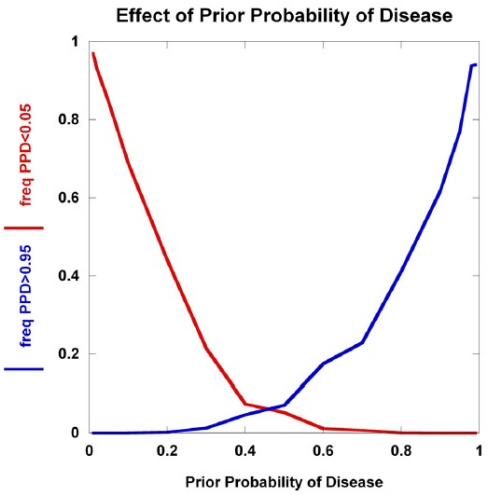
Penalized Logistic Regression with Adaptive Lasso: As an alternative classification approach, we will use penalized logistic regression with adaptive lasso. High dimensional logistic regression with adaptive lasso has several compelling statistical features for analysis of these data. Logistic regression is a likelihood-base predictive model that is commonly used across numerous fields including genetics and bioinformatics for analysis of binary outcomes. A penalized logistic regression is incorporated for instances with a large number of features and small number of subjects.[74] A popular penalty for model selection and estimation is lasso.[75] In some scenarios with high-dimensional input, such as is the case with this study, the lasso variable selection

can be inconsistent.[76] To avoid this potential inconsistency, we will adopt an adaptive lasso, which is consistent under very general conditions.[74,76]

Control for Overfitting: A common pitfall encountered in predictive modeling is overfitting the data. 10-fold cross validation will be employed to reduce the effect of overfitting. The feature selection procedure will be embedded in the cross validation routine.

Incorporation of Antigen/Antibody Profiles: For rheumatoid arthritis and systemic lupus erythematosus sample sets, we will incorporate the results derived from the ADI antigen *in vitro* transcription/translation system as features for the Bayesian network and penalized logistic regression classifiers.

Evaluating Diagnostic Utility of Classifiers: Proper and informative evaluation of the performance of classifiers is critically important to understanding the potential utility of disease predictive models. The primary method that will be employed to do so will be the area under the ROC curve, or AUC. That said, we will explore methods to evaluate the diagnostic utility of the Bayesian Network classifier results that both capture the predictive capacity of the classification and have interpretability in a clinical setting. One such approach is to calculate a posterior probability of disease (PPD) using a prior density that reflects the risk in a population from which potential patients may be derived for initial evaluation by physicians (e.g., those with symmetrical joint pain being evaluated for possible rheumatoid arthritis). Setting thresholds for the density mass that exceeds the threshold (τ_{pos}) and does not exceed a lower threshold (τ_{neg}) is one such approach:



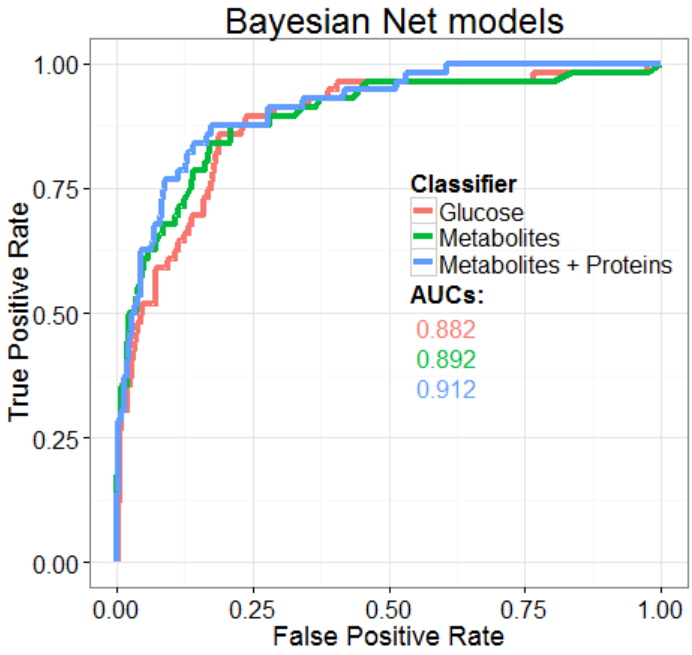
$$C_1: P(Disease \mid Genotype\ Data, other\ features) > \tau_{pos}$$
$$C_2: P(Disease \mid Genotype\ Data, other\ features) < \tau_{neg}$$

Example: Using the molecular data, we can calculate the posterior probability of RA for each individual. Thus, we would generate a distribution of posterior probabilities for RA. With perfect information, the PPD mass is concentrated at 1 and 0 (everyone either has or does not have RA). Suppose that clinicians would likely intervene therapeutically if their assessment of the probability of RA exceeds 0.95. That is, suppose a common clinical threshold for employing a particular therapeutic (e.g., methotrexate) is 0.95 probability of RA. Then, the proportion of individuals with PPD>0.95 should be close to the frequency of RA in the population evaluated for a classifier with high predictive capacity. Similarly, one can argue for the lower threshold. The plot to the left shows the impact of the expectation of the prior probability of disease on the PPD mass.

Testing for Significant Clustering Within Each Disease: To test for significant molecular subgroups within each disease, we will use a generalization of logistics PCA[77] and biclustering approaches.[78] The probabilistic nature of the logistic PCA model will allow us to extend the methodology to obtain a dimension reduction tool that handles both binary and continuous variables in a unified framework.[77] Dr. Maadooliat is an expert in PCA methods and dimensional reduction.

Preliminary Data on Type 2 Diabetes: The Schrodi lab has recently conducted a predictive study of type 2 diabetes (T2D) using a panel of metabolites, immunological and metabolic markers, and T2D-associated genetic markers. The samples were obtained from PMRP and all case individuals (n=56) were selected to have a T2D diagnosis within 18 months following enrollment in the biobank. T2D diagnosis was determined by the Marshfield Clinic electronic medical record information and was defined as fasting glucose > 125 mg/dL, random plasma glucose > 200 mg/dL, and/or HbA1c exceeding or equal to 6.5%. Controls totaled 445 individuals who did not have any ICD9 codes for diabetes and all measurements of fasting glucose, random glucose and HbA1c were within the normal range. Using the stored plasma samples on all T2D cases and controls, a metabolite panel was measured by Metanomics Health. We measured concentrations of 6 inflammatory and metabolic proteins: C-peptide, adiponectin, IL-6, insulin, glucagon, and leptin. Using a combination of (i) forward selection that maximizes orthogonality between features, and (ii) an elastic net logistic regression model selecting features with the highest correlation with residuals, we select features to be

used in the classifier. A Bayesian Network was generated for the selected features within each cross-validation run and an AUC calculated (avg AUC 0.912).[17]



The results of the Bayesian Network classifier for early T2D prediction shows high sensitivity and specificity using an elastic net for feature selection, followed by calculation of PPD incorporating correlative structure across the features. The addition of other data types, in this example proteins and additional metabolites, improves the averaged cross validated AUC values.

Specific Aim 3

To validate the predictive model for rheumatoid arthritis in an independent sample set.

The PMRP houses sufficient RA patients with diagnoses following sample collection and controls to split the sample into training and test sample sets. The same predictive modeling procedure will be used for the initial set of RA and controls as will be used for the remaining 7 diseases (i.e., SIS, elastic net feature selection, classifier development, and 10-fold cross validation incorporating the SIS/feature selection steps). The predictive model with the highest average cross validation AUC values will be tested in the independent hold-out sample set of 97 cases and 200 controls. A portion of these RA samples (n=60) and controls (n=100) will be classified by the predictive model utilizing the exome variants, circulating cytokines/metabolic proteins as well as the ADI protein microarray features. This will further test if the molecular panels measured and machine learning/statistical procedures have the capacity to produce a robust, accurate classifier for RA.

Timeline

Year 1: Obtain Marshfield IRB approval for the study. Partition exome data for the PMRP into data types (exome variants, GWAS-significant SNPs, AIMs). Define the PMRP cohort to be studied by applying exclusion criteria and performing the PCA analysis on the exome beadchip data to remove outliers. Identify cases and controls for all diseases with aid from Dr. Smith. Send plasma from 50% of the samples to Eve Technologies to run the Human Cytokine Array/Chemokine Array 42-Plex (HD42) and the Human Metabolic Hormone Array 9-Plex. Send plasma from 50% of the samples to Antigen Discovery for interrogation on the 800 antigen protein microarray platform.

Year 2: Send plasma from 50% of the samples to Eve Technologies to run the Human Cytokine Array/ Chemokine Array 42-Plex with IL-18 (HD42) and the Human Metabolic Hormone Array 9-Plex. Send plasma from 50% of the samples to Antigen Discovery for interrogation on the 800 antigen microarray platform. Retrieve and merge data from Eve Technologies and Antigen Discovery with exome variant data and phenotype data. With aid from Dr. Maadooliat and Dr. Guo, set up computational software and algorithms for running the association analyses, SIS, feature selection, and classifier development on the high performance computing cluster.

Year 3: Perform statistical and machine learning analyses on the merged data files. Set up binary PCA and biclustering algorithms for analysis of disease subgroups through clustering. Dr. Guo will also receive training from Drs. Schrodi and Maadooliat in analysis methods.

Year 4: Complete disease prediction for eight diseases and evaluate diagnostic utility measures. Select predictive model for validation sample set of rheumatoid arthritis. Perform analysis to identify molecular-based disease subgroups. Mentor and train Dr. Guo (postdoctoral fellow on the study). Development of novel methods for prediction, clustering through dimensional reduction, and investigate new diagnostic utility metrics. Work with Dr. Smith to interpret biological implications of results.

Year 5: Complete analysis of disease subgroups. Write manuscripts and present at national conferences. Coordinate with Drs. Maadooliat, Guo, Trappl-Kimmons and Liang on analysis results, the performance of novel methods. Identify new hypotheses that are generated from these results and apply for funding to continue investigation in new direction. Work with Dr. Smith to interpret clinical implications of results.

PHS Human Subjects and Clinical Trials Information

OMB Number: 0925-0001 and 0925-0002

Expiration Date: 03/31/2020

Are Human Subjects Involved	● Yes ○ No
Is the Project Exempt from Federal regulations?	○ Yes ● No
Exemption Number	❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8
Other Requested Information	

Human Subject Studies

Study#	Study Title	Clinical Trial?
1	Multi-Omics Prediction of Eight Common Diseases in the Central Wisconsin Population	No

Section 1 - Basic Information (Study 1)

OMB Number: 0925-0001 and 0925-0002

Expiration Date: 03/31/2020

1.1. Study Title *

Multi-Omics Prediction of Eight Common Diseases in the Central Wisconsin Population

1.2. Is this study exempt from Federal Regulations * ❍ Yes ● No

1.3. Exemption Number ❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8

1.4. Clinical Trial Questionnaire *

1.4.a. Does the study involve human participants?	● Yes	❍ No
1.4.b. Are the participants prospectively assigned to an intervention?	❍ Yes	● No
1.4.c. Is the study designed to evaluate the effect of the intervention on the participants?	❍ Yes	● No
1.4.d. Is the effect that will be evaluated a health-related biomedical or behavioral outcome?	❍ Yes	● No

1.5. Provide the ClinicalTrials.gov Identifier (e.g. NCT87654321) for this trial, if applicable

Section 2 - Study Population Characteristics (Study 1)

2.1. Conditions or Focus of Study

- Rheumatoid Arthritis
- Systemic Lupus Erythematosus
- Relapsing-remitting multiple sclerosis
- premature myocardial infarction
- chronic lymphocytic leukemia
- obsessive-compulsive disorder
- autoimmune thyroid disease
- axial spondyloarthritis

2.2. Eligibility Criteria

see Research Strategy

2.3. Age Limits	Min Age: 20 Years	Max Age: 110 Years
2.4. Inclusion of Women, Minorities, and Children	1253-Inclusion Children_Women_Minorities.pdf	
2.5. Recruitment and Retention Plan	1254-Recruitment Retention.pdf	
2.6. Recruitment Status	Completed	
2.7. Study Timeline	1255-Timeline.pdf	
2.8. Enrollment of First Subject	05/01/2019	Anticipated

Inclusion of Children:

All Personalized Medicine Research Project (PMRP) participants were over 18 at enrollment. We note that there is a small population between 18-21 years of age in the PMRP. Only one individual is between 18-21 years of age is in the set that is exome genotyped.

Inclusion of Women and Minorities:

We are not recruiting any new subjects for this study. The existing database that we will use, the Personalized Medicine Research Project (PMRP) is representative of the population of Central Wisconsin. There are marginally more females than males and the population is overwhelmingly (98%) self-reported white, primarily of Northwestern European origin. This population is ideal for the proposed study in that confounding by population stratification is minimized, and that the underlying genetic architecture of inflammation is likely to be less complex than that in more heterogeneous populations.

Recruitment and Retention

This study uses biobanked samples from previously recruited subjects.

TIMELINE

Year 1: Obtain Marshfield IRB approval for the study. Partition exome data for the PMRP into data types (exome variants, GWAS-significant SNPs, AIMs). Define the PMRP cohort to be studied by applying exclusion criteria and performing the PCA analysis on the exome beadchip data to remove outliers (completed but will verify results). Identify cases and controls for all diseases with aid from Dr. Smith. Send plasma from 50% of the samples to Eve Technologies to run the Human Cytokine Array/Chemokine Array 42-Plex (HD42) and the Human Metabolic Hormone Array 9-Plex. Send plasma from 50% of the samples to Antigen Discovery for interrogation on the 800 antigen protein microarray platform.

Year 2: Send plasma from 50% of the samples to Eve Technologies to run the Human Cytokine Array/ Chemokine Array 42-Plex with IL-18 (HD42) and the Human Metabolic Hormone Array 9-Plex. Send plasma from 50% of the samples to Antigen Discovery for interrogation on the 800 antigen microarray platform. Retrieve and merge data from Eve Technologies and Antigen Discovery with exome variant data and phenotype data. With aid from Dr. Maadooliat and Dr. Guo, set up computational software and algorithms for running the association analyses, SIS, feature selection, and classifier development on the high performance computing cluster.

Year 3: Perform statistical and machine learning analyses on the merged data files. Set up binary PCA and biclustering algorithms for analysis of disease subgroups through clustering. Dr. Guo will also receive training from Drs. Schrodi and Maadooliat in analysis methods.

Year 4: Complete disease prediction for eight diseases and evaluate diagnostic utility measures. Select predictive model for validation sample set of rheumatoid arthritis. Perform analysis to identify molecular-based disease subgroups. Dr. Schrodi will continue to mentor and train Dr. Guo (postdoctoral fellow on the study). Drs. Maadooliat and Schrodi will develop/implement methods for prediction, clustering through dimensional reduction, and investigate new diagnostic utility metrics. Work with Dr. Smith to interpret biological implications of results.

Year 5: Complete analysis of disease subgroups. Write manuscripts and present at national conferences. Coordinate with Drs. Maadooliat, Guo, Trappl-Kimmons and Liang on analysis results, the performance of methods. Identify new hypotheses that are generated from these results and apply for funding to continue investigation in new direction. Work with Dr. Smith to interpret clinical implications of results. Draft and submit manuscripts describing the study and communicate analysis results.

Inclusion Enrollment Reports

IER ID#	Enrollment Location Type	Enrollment Location
Study 1, IER 1	Domestic	Wisconsin

Inclusion Enrollment Report 1

Using an Existing Dataset or Resource* : ● Yes ❍ No

Enrollment Location Type* : ● Domestic ❍ Foreign

Enrollment Country(ies): USA: UNITED STATES

Enrollment Location(s): Wisconsin

Comments:

Planned

Racial Categories	Ethnic Categories				Total
	Not Hispanic or Latino		Hispanic or Latino		
	Female	Male	Female	Male	
American Indian/ Alaska Native	22	14	0	0	36
Asian	48	32	0	0	80
Native Hawaiian or Other Pacific Islander	0	0	0	0	0
Black or African American	11	28	0	0	39
White	11099	8196	0	0	19295
More than One Race	0	0	0	0	0
Total	11180	8270	0	0	19450

Cumulative (Actual)

Racial Categories	Ethnic Categories									Total
	Not Hispanic or Latino			Hispanic or Latino			Unknown/Not Reported Ethnicity			
	Female	Male	Unknown/ Not Reported	Female	Male	Unknown/ Not Reported	Female	Male	Unknown/ Not Reported	
American Indian/ Alaska Native	22	14	0	0	0	0	0	0	0	36
Asian	48	32	0	0	0	0	0	0	0	80
Native Hawaiian or Other Pacific Islander	0	0	0	0	0	0	0	0	0	0
Black or African American	11	28	0	0	0	0	0	0	0	39
White	11099	8196	0	0	0	0	0	0	0	19295
More than One Race	0	0	0	0	0	0	0	0	0	0
Unknown or Not Reported	0	0	0	104	94	0	0	0	0	198
Total	11180	8270	0	104	94	0	0	0	0	19648

Section 3 - Protection and Monitoring Plans (Study 1)

3.1. Protection of Human Subjects	1256-ProtectionHumanSubjects_Schrodi_RO1_02Jul2018.pdf
3.2. Is this a multi-site study that will use the same protocol to conduct non-exempt human subjects research at more than one domestic site?	❍ Yes ● No ❍ N/A
If yes, describe the single IRB plan	
3.3. Data and Safety Monitoring Plan	
3.4. Will a Data and Safety Monitoring Board be appointed for this study?	❍ Yes ● No
3.5. Overall structure of the study team	

Additional Review Criteria

Protections for Human Subjects

Risks to Subjects

Human Subjects Involvement, Characteristics, and Design

- Human subjects will be used for this study. Human subjects will come from Marshfield Clinic's Personalized Medicine Research Project (PMRP) consisting of nearly 20,000 Marshfield Clinic patients. Plasma biospecimens will be used to measure circulating levels of proteins and antibodies. Genotype data on PMRP individuals will be used and analyzed.
- The PMRP cohort is primarily white North Americans (98.4%) of German ancestry (76.7%) consistent with the region sampled. The PMRP has a slight enrichment for females (57%). The age range is no less than 18 and some are older than 100 with mean and median age over 50. All patients are linked to the electronic medical record (EMR) and have extensive medical histories available.
- No additional recruitment is required.
- This is an existing cohort population with no individuals younger than 18 years of age. There was no selection for vulnerable populations.
- No collaborating sites where human subject research will be performed are expected.

Sources of Materials

- Research material obtained from living individuals includes DNA and plasma.
- Data collected from human subjects will include patient data in the EMR and genetic data.
- Only a select few individuals in Marshfield Clinic's BIRC facility have access to individually identifiable private information. Patient identifiers will be converted to de-identified study identifiers for this project. Only staff involved in this project will have access to de-identified data.
- All biological specimens are de-identified and stored in a locked facility and locked freezers. Clinical and genetic data is stored in a secured database and is password protected. Patient clinical phenotype data and genetic data cannot be viewed together. Individuals who have access to clinical data do not have access to genetic data. Those with access to genetic data do not have access to clinical data. Only when patient identifiers are converted to a coded study identifier can the two data sets be combined.

Potential Risks

- The largest potential risk is breach of confidentiality. Procedures have been established to minimize this risk. This includes use of de-identified samples/data, maintaining samples/data in a controlled and secure location, and restricting access of samples/data to pertinent research personnel.
- No treatments or procedures are expected.

Adequacy of Protection Against Risks

Recruitment and Informed Consent

- All human subjects have been recruited within PMRP and have provided signed informed consent that allows their clinical records and biological samples to be used for research purposes. All studies using PMRP samples must get IRB approval. No children have been recruited into PMRP.
- A pilot study working with 2015 individuals from PMRP doing the same work has been reviewed and approved by the Marshfield Clinic Research Foundation IRB.
- In PMRP, informed consent has been obtained. A request for a waiver of informed consent may be needed because many patients will have moved or are deceased and attaining consent would be a significant burden to the project when there is a minimal risk to the patient.

Protections Against Risks

- All data used in this study will be de-identified with a study identifier and only a few have access to the patient identifier and study identifier manifest. Clinical and genetic data are stored in a

secured database that is password protected. Patient clinical phenotype data and patient genetic data cannot be viewed together unless de-identified. Individuals who have access to clinical data do not have access to genetic data. Those with access to genetic data do not have access to clinical data. Only those involved in this project will have access to study data and are professionally trained in the Protection of Human Subjects.

- This project does not involve vulnerable populations
- Human fetuses and neonates are not part of this study. Pregnant women were not targeted during PMRP recruitment, but could be participants since this is a cohort population.
- Prisoners are not part of this study.
- Children are not part of this study.
- This study is not a clinical trial and does not propose any medical or professional intervention.
- Individual results will not be returned to participants

Potential Benefits of the Proposed Research to Human Subjects and Others

- There are no direct benefits to be gained for the research subjects in this study. The potential benefits of the proposed research are to society as a whole. This project may have a tremendous public health impact by describing the potential genetic contribution and inflammatory protein contribution to many diseases. If an unexpected association is found during the course of this experiment, there could be no resulting action taken as the principal investigator will not be able to match up any genetic results with an individual. Further, no clinically actionable results will arise from this study as (1) results may not definitively demonstrate causation, only association, (2) samples and/or data were not collected in a CLIA-approved manner, and (3) **PMRP consent states no return of results.**
- The largest potential risk is breach of confidentiality. Significant efforts are made to minimize this risk. MCRI has safeguards in place within the Biomedical Informatics Research Center to verify that identifying information is not released and that all activities are governed by HIPAA regulations.

Importance of the Knowledge to be Gained

- This is a study that links disease states with genetic and protein variation. In combination, Identifying strong associations can be used to develop genetic and/or protein-based tests for disease prediction and/or treatment. Identification of specific constellations of inflammatory markers, metabolic markers, genetics and antibodies for chronic diseases will enable us to use molecular panels to better elucidate the molecular changes and signaling pathways responsible for disease pathogenesis. Discovery of coding-region genetic variants that drive disease susceptibility will illuminate novel pathways and mechanisms. The study will assess the utility of prediction of chronic diseases prior to clinical diagnosis using genetics, circulating biomarkers and antibody profiles. Risks to subjects are reasonable since a great deal of clinically relevant information can be gained from this project with only minimal risk to the participants.

Section 4 - Protocol Synopsis (Study 1)

4.1. Brief Summary

4.2. Study Design

4.2.a. Narrative Study Description

4.2.b. Primary Purpose

4.2.c. Interventions

Type	Name	Description

4.2.d. Study Phase

Is this an NIH-defined Phase III Clinical Trial? ❍ Yes ❍ No

4.2.e. Intervention Model

4.2.f. Masking ❍ Yes ❍ No

❑ Participant ❑ Care Provider ❑ Investigator ❑ Outcomes Assessor

4.2.g. Allocation

4.3. Outcome Measures

Type	Name	Time Frame	Brief Description

4.4. Statistical Design and Power

4.5. Subject Participation Duration

4.6. Will the study use an FDA-regulated intervention? ❍ Yes ❍ No

4.6.a. If yes, describe the availability of Investigational Product (IP) and Investigational New Drug (IND)/ Investigational Device Exemption (IDE) status

4.7. Dissemination Plan

Delayed Onset Studies

Delayed Onset Study#	Study Title	Anticipated Clinical Trial?	Justification
The form does not have any delayed onset studies			

BIBLIOGRAPHY & REFERENCES CITED

1. Danese S, Fiorino G, Peyrin-Biroulet L. (2017) Early intervention in Crohn's disease: toward disease modification trials. *Gut* Sep. 5. [Epub ahead of print]

2. Jakimovski D, Weinstock-Guttman B, Ramanathan M, Kolb C, et al. (2017) Ocrelizumab: a B-cell depleting therapy for multiple sclerosis. *Expert Opin Biol Ther* 17(9):1163-1172.

3. Zink A, Albrecht K. (2017) Rheumatoid arthritis: The benefits of early treatment after decades. *Nat Rev Rheumatol.* 13(8):458-459.

4. Smolen JS, Landewe R, Bijlsma J, Burmester G, et al. (2016) EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 76(6):960-977.

5. Barkham N, Keen HI, Coates LC, O'Connor P, et al. (2009) Clinical and imaging efficacy of infliximab in HLA-B27-Positive patients with magnetic resonance imaging-determined early sacroiliitis. *Arthritis Rheum* 60(4):946-954.

6. Romanelli RJ, Chung S, Pu J, Nimbal V, et al. (2015) Comparative effectiveness of early versus delayed metformin in the treatment of type 2 diabetes*. Diabetes Res Clin Pract* 108(1):170-178.

7. Frydkjaer-Olsen U, Soegaard Hansen R, Simo R, Cunha-Vaz J, et al. (2016) Correlation between retinal vessel caliber and neurodegeneration in patients with type 2 diabetes mellitus in the European Consortium for the Early Treatment of Diabetic Retinopathy (EUROCONDOR). *Ophthalmic Res* 56(1):10-6.

8. Chaudhuri AA, Chabon JJ, Lovejoy AF, Newman AM, et al. (2017) Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discov* Sep 12 [Epub ahead of print]

9. Nair B, Wilt T, MacDonald R, Rutks I. (2002) Early versus deferred androgen suppression in the treatment of advanced prostatic cancer. *Cochrane Database Syst Rev* 1:CD003506.

10. Neal RD, Tharmanathan P, France B, Din NU, et al. (2015) Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer* 112(Suppl a):S92-S107.

11. Ford I, Murray H, Packard CJ, Shepherd J, et al. (2007) Long-term follow-up of the West of Scotland Coronary Prevention Study. *N Engl J Med* 357:1477-1486.

12. Lindstrom J, Peltonen M, Eriksson JG, Aunola S, et al. (2008) Determinants for the effectiveness of lifestyle intervention in the Finnish Diabetes Prevention Study. *Diabetes Care* 31(5):857-862.

13. Sacre JW, Jellis CL, Jenkins C, Haluska BA, et al. (2014) A six-month exercise intervention in subclinical diabetic heart disease: Effects on exercise capacity, autonomic and myocardial function. *Metabolism* 63(9):1104-1114.

14. Campbell F, Conti G, Heckman JJ, Moon SH, et al. (2014) Early childhood investments substantially boost adult health. *Science* 343(6178):1478-1485.

15. Berentzen NE, van Rossem L, Gehring U, Koppelman GH, et al. (2016) Overweight patterns throughout childhood and cardiometabolic markers in early adolescence*. Int J Obes* 40(1):58-64.

16. Hu Y, Bhupathiraju SN, de Koning L, Hu FB. (2014) Duration of obesity and overweight and risk of type 2 diabetes among US women. *Obesity* 22(10):2267-2273.

17. Carter TC, Rein D, Padberg I, Peter E, et al. (2016) Validation of a metabolite panel for early diagnosis of type 2 diabetes. *Metabolism* 65(9):1399-1408.

18. Herbst RS, Soria J-C, Kowanetz M, Fine GD, et al. (2014) Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* 515(7528):563-567.

19. Zethelius B, Berglund L, Sundstrom J, Ingelsson E, et al. (2008) Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. *N Engl J Med* 358:2107-2116.

20. Schrag A, Siddiqui UF, Anastasiou Z, Weintraub N, Schott JM (2017) Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study. *Lancet Neurol* 16(1):66-75.

21. Schrodi SJ, Mukherkee S, Shan Y, Tromp G, et al. (2014) Genetic-based prediction of disease traits: prediction is very difficult, especially about the future. *Front Genet* 5:162.

22. Hebbring SJ, Slager SL, Epperla N, Mazza JJ, et al. (2013) Genetic evidence of PTPN22 effects on chronic lymphocytic leukemia. *Blood* 121(1):237-238.

23. Hebbring SJ, Schrodi SJ, Ye Z, Zhou Z, et al. (2013) A PheWAS approach in studying HLA-DRB1*1501. *Genes Immun* 14(3):187-191.

24. O'Brien SE, Schrodi SJ, Ye Z, Brilliant MH, et al. (2015) Differential lipid response to statins is associated with variants in the BUD13-APOA5 gene region. *J Cardiovasc Pharmacol* 66(2):183-188.

25. Marouli E, Graff M, Medina-Gomez C, Lo KS, et al. (2017) Rare and low-frequency coding variants alter human adult height. *Nature* 542(7640):186-190.

26. Schmidt AF, Swerdlow DI, Holmes MV, Patel RS, et al. (2017) PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomization study. *Lancet Diabetes Endocrinol* 5(2):97-105.

27. Fritsche LG, Igl W, Bailey JN, Grassman F, et al. (2016) A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet* 48(2):134-143.

28. Liu Y, Ye Z, Li X, Anderson JL, et al. (2017) Genetic and functional associations with decreased anti-inflammatory tumor necrosis factor alpha induced protein 3 in macrophages from subjects with axial spondyloarthritis. *Front Immunol* 8:860.

29. Maadooliat M, Bansal NK, Upadhya J, Farazi MR, et al. (2016) The decay of disease association with declining linkage disequilibrium: a fine mapping theorem. *Front Genet* 7:217.

30. Bansal NK, Maadooliat M, Schrodi SJ. Empirical Bayesian approach to multiple hypotheses with separate priors for left and right alternatives. (Manuscript Under review)

31. Hasin Y, Seldin M, Lusis A. (2017) Multi-omics approaches to disease. *Genome Biol* 18:83.

32. Miao R, Luo H, Zhou H, Li G, et al. (2014) Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. *J Hepatol* 61(4):840-849.

33. Gjoneska E, Pfenning AR, Mathys H, Quon G, et al. (2015) Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* 518:365-369.

34. Zeevi D, Korem T, Zmora N, Israeli D, et al. (2015) Personalized nutrition by prediction of glycemic responses. *Cell* 163:1079-1094.

35. Colombo BM, Canevali P, Magnani O, Rossi E, et al. (2012) Defective expression and function of the leukocyte associated Ig-like receptor 1 in B lymphocytes from systemic lupus erythematosus patients. *PLoS One* 7(2):e31903.

36. Zhang Z, Liu J, Ding X, Bijma P, et al. (2010) Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS One* 5(9):e12648.

37. Meuwissen THE, Hayes BJ, Goddard ME. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.

38. VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91(11):4414-4423.

39. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, et al. (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 43:519-525.

40. de Los Campos G, Vazquez AI, Fernando R, Klimentidis YC, Sorensen D. (2013) Prediction of complex human traits using genomic best linear unbiased predictor. *PLoS Genet* 9(7):e1003608.

41. Yang J, Benyamin B, McEvoy BP, Gordon S, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42:565-569.

42. Maadooliat M, Zhou S, Najibi SM, Gao X, Huang JZ. (2016) Collective estimation of multiple bivariate density functions with application to angular-sampling-based protein loop modeling. *J Am Stat Assoc* 11(513):43-56.

43. Kim HM, Maadooliat M, Arellano-Valle RB, Genton MG. (2016) Skewed factor models using selection mechanism. *J Multivariate Analysis* 145:162-177.

44. Maadooliat M, Pourahmadi M, Huang JZ. (2013) Robust estimation of the correlation matrix of longitudinal data. *Stat Comp* 23(1):17-28.

45. Maadooliat M, Huang JZ, Hu J. (2012) Analyzing multiple-probe microarray: estimation and application of gene expression indexes. *Biometrics* 68(3):784-792.

46. Maadooliat M, Huang JZ, Hu J. (2015) Integrating data transformation in principal components analysis. *J Computational and Graphical Statistics*. 24(1):84-103.

47. Schrodi SJ. (2005) A probabilistic approach to large-scale association scans: a semi-Bayesian method to detect disease-predisposing alleles. *Stat Appl Genet Mol Biol* 4(1):Article 31.

48. Schrodi SJ, DeBarber A, He M, Ye Z, et al. (2015) Prevalence estimation for monogenic autosomal recessive diseases using population-based genetic data. *Hum Genet* 134(6):659-669.

49. He M,, Person TN, Hebbring SJ, Heinzen E, et al. (2015) SeqHBase: a big data toolset for family based sequencing data analysis. *J Med Genet* 52(4):282-288.

50. Schrodi SJ. (2016) The use of multiplicity corrections, order statistics and generalized family-wise statistics with application to genome-wide studies. *PLoS One* 11(4):e0154472.

51. Ye Z, Tafti AP, He KY, Wang K, He MM. (2016) Sparktext: Biomedical text mining on big data framework. *PLoS One* 11(9):e0162721.

52. Mosley JD, Witte JS, Larkin EK, Bastarache L, et al. (2016) Identifying genetically driven clinical phenotypes using linear mixed models. *Nat Commun* 7:11433.

53. Fisher B, Redmond C, Fisher ER, Caplan R. (1988) Relative worth of estrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node negative breast cancer patients: findings from National Surgical Adjuvant Breast and Bowel Project Protocol B-06. *J Clin Oncology* 6:1076-1087.

54. Fisher B, Costantino J, Redmond C, Poisson R, et al. (1989) A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. *N Engl J Med* 320: 479-484.

55. Paik S, Shak S, Tang G, Kim C, et al. (2004) A multigene assay to predict recurrence of tamoxifen-treated node-negative breast cancer. *N Engl J Med* 351:2817-2826.

56. Madeira S, Oliveira AL (2004) Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM transactions on Computational Biology and Bioinformatics (TCBB)* 1(1):24-45.

57. Pang Y, Tao D, Yuan Y, Li X. (2008) Binary two-dimensional PCA. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 38(4):1176-1180.

58. Fan J, Lv J. (2008) Sure independence screening for ultrahigh dimensional feature space. *J R Statist Soc B* 70(5):849-911.

59. Friedman N, Geiger D, Goldszmidt M (1997) Bayesian network classifiers. *Machine Learning* 29(2-3):131-163.

60. Jordan MI, Ghahramani Z, Jaakkola TS, Saul LK (1999) An introduction to variational methods for graphical models. *Machine Learning* 37(2):183-233.

61. Hastie T, Tibshirani R, Friedman JH (2009) *The elements of statistical learning: data mining, inference, and prediction* (2nd ed.). Springer-Verlag, New York.

62. Schrodi SJ. The impact of diagnostic code misclassification on optimizing the experimental design of genetic association studies *J Healthc Eng* (in Press)

63. McCarty CA, Wilke RA, Giampietro PF, Wesbrook SD, Caldwell MD. (2005) Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank. *Per Med* 2:49-79.

64. https://emerge.mc.vanderbilt.edu/

65. Holzinger ER, Verma SS, Moore CB, Hall M, et al. (2017) Discovery and replication of SNP-SNP interactions for quantitative lipid traits in over 60,000 individuals. *BioData Min* 10:25.

66. El-Gabalawy H, Robinson D, Smolik I, Hart D, et al. (2012) Familial clustering of the serum cytokine profile in the relatives of rheumatoid arthritis patients. *Arthritis Rheum* 64(6):1720-1729.

67. Egli A, Santer DM, O'Shea D, Barakat K, et al. (2014) IL-28B is a key regulator of B- and T-cell vaccine responses against influenza. *PLoS Pathog* 10(12):e1004556.

68. Heard BJ, Rosvold JM, Fritzler MJ, El-Gabalawy H, et al. (2014) A computational method to differentiate normal individuals, osteoarthritis and rheumatoid arthritis patients using serum biomarkers. J R Soc Interface 11(97):20140426.

69. Sundaresh S, Doolan DL, Hirst S, My Y, et al. (2006) Identification of humoral immune responses in protein microarrays using DNA microarray data analysis techniques. *Bioinformatics* 22:1760-1766.

70. Vigil A, Davies DH, Felgner PL (2010) Defining the humoral immune response to infectious agents using high-density protein microarrays. *Future Microbiol* 5:241-251.

71. Lee SJ, Liang L Juarez S, Nanton MR, et al. (2012) Identification of a common immune signature in murine and human systemic Salmonellosis. *Proc Natl Acad Sci USA* 109(13):4998-5003.

72. Kunnath-Velayudhan S, Salamon H, Wang HY, Davidow AL, et al. (2010) Dynamic antibody responses to the Mycobacterium tuberculosis proteome. *Proc Natl Acad Sci USA* 107(33):14703-14708.

73. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am J Hum Genet* 81:1084-1097.

74. Ma S, Huang J. (2008) Penalized feature selection and classification in bioinformatics. *Briefing in Bioinformatics*, 9(5):392–403.

75. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Royal Stat Society. Series B (Methodological)*. 58(1):267-288.

76. Hui Zou. (2006) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476):1418-1429.

77. Lee S, Huang JZ and Hu J. (2010) Sparse logistic principal component analysis for binary data. *Annuals of Applied Statistics*, 4(3):1579-1601.

78. Cheng Y, Church GM. (2000) Biclustering of expression data, *ISMB 8*, 93-103.

Consortium/Contractual Arrangements

This grant application is collaboration amongst two investigators from two institutions. The involved investigators from these institutions bring specific expertise and experience to the success of the project. Dr. Steven Schrodi is from the Marshfield Clinic Research Institute while Dr. Mehdi Maadooliat is from Marquette University.

The Office of Sponsored Programs at Marshfield Clinic Research Institute will initiate and monitor the consortium and contractual arrangement with Marquette University. The fiscal and administrative arrangements will be defined in the consortium agreement.

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

Department of Pediatrics

June 28th, 2018

Re: Schrodi R01 submission on prediction of diseases using exome data, cytokines and antibodies

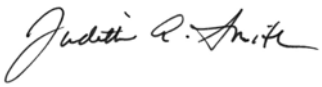
Dear Dr. Schrodi,

I am very excited to be working with you as a consultant on this project examining the genetics, expression of proinflammatory cytokines, autoantibodies and other proteins, and how these genetic derangements and biomarkers ultimately inform inflammatory and other diseases. I have been pursuing a complementary line of investigation with you in the pathogenesis of ankylosing spondylitis. This other project has been a real testament to our ability to work together as a team to pursue translational research. In regards to this current proposal, I am happy to lend my medical expertise, with my particular focus on inflammatory diseases, as well as my immunology expertise in inflammatory cytokine regulation.

My hourly consulting rate is $160/hour, not to exceed 75 hours per year. This consulting rate has not been increased based on the amount being provided for the agreement.

I look forward to many more fruitful years in partnership.

Sincerely,



Judith A Smith MD PhD
Associate Professor
Department of Medical Microbiology and Immunology
Department of Pediatrics
Division of Allergy, Immunology and Rheumatology
4159 Microbial Sciences Building
1550 Linden Dr.
Madison WI 53706
Phone: 608 263-1251
e-mail: jsmith27@wisc.edu

KLINGLER
College of Arts & Sciences
MARQUETTE UNIVERSITY

Mathematics, Statistics and Computer Science

www.mscs.mu.edu

Steven J. Schrodi, Ph.D.
Center for Human Genetics,
Marshfield Clinic Research Institute,
Marshfield, WI 54449

Dear Dr. Schrodi,

I would like to express my strong support for your research proposal, "Multi-Omics Prediction of Eight Common Diseases in the Central Wisconsin Population" This proposal builds well on your prior works.

I would be delighted to serve as a consulting statistician on your proposed research project as it is related on my own research in statistical modeling. I look forward to participating in this role.

I will be able to offer my expertise in statistics and machine learning to develop molecular-based predictive models for eight chronic and complex diseases using an extensive longitudinal electronic health records linked to the Marshfield Clinic Biobank. Furthermore I can provide help for appropriate modeling of the potential correlation structure and controlling the Type-I error for multiple hypotheses testing as well.

As a research scientist and an experienced reviewer of research proposals, I believe this research project is important, feasible, and consistent with the goals of the National Institutes of Health. I am hopeful that this proposal will be a success.

Sincerely,

Mehdi Maadooliat

Assistant Professor
Katherine Reed Cudahy Hall, Room 311
Dept. of Mathematics, Statistics and Computer Science
Marquette University
Milwaukee, WI 53201-1881, USA
Phone: (414) 288-6341
Fax: (414) 288-5472 email: mehdi@mscs.mu.edu
http://www.mscs.mu.edu/~mehdi

RESOURCE SHARING PLAN:

1. Data Sharing Plan

Marshfield Clinic has been committed since the outset to make tools and data available from the Personalized Medicine Research Project (PMRP), as evidenced by the fact that the consent forms, questionnaire and other study materials are readily available on the Clinic internet site (http://www.marshfieldclinic.org/chg/pages/default.aspx?page=chg_pers_med_res_prj).

Marshfield's demographic and clinical data reside in Marshfield's Oracle Data Warehouse. The PMRP genetic data reside in an Oracle relational database on a Linux Server. Marshfield will develop SQL queries to select and merge the demographic and clinical data with the SNP probe data and then format the data to meet NCBI data specifications. Marshfield will submit data to the NCBI via secure file transfer protocol.

Data transfer protocols, de-identification and data formats, and encryption project:
Data transfer protocols: Marshfield Clinic has approved the use of several data transfer protocols for sharing data outside of the Marshfield network. These protocols include:

File Transfer Protocol (FTP), an industry standard or Secure File Transfer Protocol (SFTP), this method includes encryption of the data being transmitted;

Bulletin Board System (BBS), this method involves the creation of an account with the "sharing" organizations. Files can be placed or received by either Marshfield or the "sharing. Organization. This method is also an industry standard that works on the Internet. The "sharing" organization can pick up or place a file on the BBS using a web browser. This protocol uses Secure Soccet Layer (SSL), which encrypts data over the Internet.

Instant Virtual Extranet (IVE) is an appliance that will replace the current vendor and employee Virtual Private Network (VPN) and dial-up connections. This protocol is similar to the BBS and is accessible with a Web browser.

An example of Marshfield's use of FTP is the Vaccine Safety Datalink study. The CDC has developed a secure FTP web site and files are transferred there when needed. The BBS is utilized by several Wisconsin-related research studies including the Wisconsin Provider Office Visit data collection project. Marshfield has not utilized the IVE method to date. IVE has been tested and implementation is planned in the near future.

Record de-identification and export data formats: Marshfield is experienced in de-identifying and exporting standardized patient data files to other organizations in a HIPAA compliant manner. Developing a common repository for data from the various funded sites under this RFA will require that unique repository codes be assigned to unique patient data with encryption of patient identifiers. Unique repository codes will prevent duplication of data sets to prevent over-count of the actual number of unique subjects with a defined phenotype available for research.

Encryption: Marshfield currently uses Pretty Good Privacy Encryption (PGP) a Public Key Infrastructure (PKI) or shared key cryptology system. Another available option is GPG, the free-ware version of PGP. In some situations, we have used a combination of encryption and FTE to ensure the privacy of transmitted data.

Marshfield has a long standing interest and plans to actively participate in the process used to define standards for the data transfer protocol, encryption, de-identification and data export strategies.

2. Sharing Model Organisms

Not applicable.

3. Genome Wide Association Studies (GWAS)

The proposed study will use pre-existing exome genotype data. The proposed study will generate molecular trait information in the form of protein levels in plasma samples. These trait data and exome genotype data, in a de-identified form, will be made available to the NIH-designated GWAS data repository.