Medical Student Research Program – Full Option Application

Please develop this application with your mentor.

Send completed documents and final application in PDF form to Marjorie West (mwest@umc.edu) and your mentor by Monday, October 12, 2015. Incomplete forms, lack of additional documents, and deviations from the application formatting requirements may result in application denial.

ADDITIONAL DOCUMENTS TO BE SUBMITTED WITH APPLICATION

☒ Mentor NIH Biosketch ☒ Transcript ☒ Letter of Good Standing (Academic Affairs) ☒ Student CV

1. **Applicant Name** Subhi Talal Younes
2. **Project Title** Identification of functional genetic variants affecting disease phenotype in the JHS cohort
3. **Mentor Information**

   a. Mentor Name James Wilson, MD

   b. Department Physiology and Biophysics

   c. Email jgwilson2@umc.edu

   d. Phone 601-984-2855

Funding Source(s)	☐None	☐ AHA	☒NIH	☐ Other
IRB/IACUC Approval	☒ Approved	☐Pending	☐N/A	
Protocol Name	Metabolomic Predictors of Insulin Resistance and Diabetes			
Protocol Number	IRB File #2015-0120			

4. **Personal Statement (previous accomplishments, goals for MSRP, short-term career goals, long-term plans in academic medicine, etc.) Limit: 500 words.**

The world in which we live is a marvelously complex place. The field of science is dedicated to understanding this world on a more fundamental level. In turn, the pursuit of this knowledge produces practical results which many times can be used to reshape how we see the world, how we live, or how we treat disease. It is the latter which truly fascinates me and motivates me to pursue the career of a physician-scientist.

Advances in scientific understanding have brought us from living in caves to landing on other worlds; from being unable to treat even the simplest of diseases to transplanting entire organs. Yet, we still have much further to go. One need only look around to see people, individuals still suffering from incurable diseases which rob them of a normal life and, sometimes, of their very humanity. In order to adequately attack these diseases, we must know more about them. Though occasionally discoveries of treatments are made entirely by accident with no real understanding of why they work, other treatments are developed by a thorough knowledge of the underlying molecular mechanism of the disease.

I do not wish to paint a utopian idea of scientific research - for the life of a scientist is not without sacrifice. Certainly on occasion a scientist will make a truly revolutionary discovery or cure, but such discoveries require years, if not an entire lifetime of painstaking research. One may ask, then, why would anyone choose this life of meticulous work whose results are all but clear? Why would one reject a “normal” life free of uncertain results, funding shortages, and painstaking work? For me, the answer is remarkably simple. While there is still one poor soul who suffers from an incurable disease, one man who struggles to maintain his very humanity in the face of unbelievable adversity, I choose to reject a life of normality. I refuse to do nothing while others suffer with no way to ease their suffering. And yet, the sacrifice turns out to be minimal. The journey, though arduous, is full of wonder, challenge, and triumph.

I have experienced all this while working on my undergraduate thesis on Parkinson Disease and during my summer research with the MSRP program. My undergraduate thesis consisted of extensive review of the available literature on Parkinson Disease, which allowed me to construct a model for the pathogenesis of the disease, thereby informing potential approaches to novel treatments. I hope to continue this type of work during my training. By careful examination of the underlying molecular mechanisms of diseases, I hope to be able to reveal novel treatment methods. Therein lies my hope for my career - a life of absolute abnormality as I unravel the fabric of medicine in order to truly understand disease and thus, truly cure it.

5. **Project Description – Use the next 3 pages with the embedded headings for your proposal. Do not exceed 2 pages for your proposal description, including figures and tables. The 3rd page is for references only. Do not change margins, spacing or font size.**

Background
The sequencing of the human genome promised to provide unprecedented insight into both the normal functioning of the human body and the etiology and pathogenesis of disease. By establishing the normal sequence and studying its perturbations, it was hoped that physicians would be able to predict disease before it occurred and tailor medical interventions to those treatments which would provide the most benefit while minimizing potential adverse effects. This vision, referred to as precision or personalized medicine, remains alive - most notably, in the recently launched precision medicine initiative. The completion of the human genome project unveiled the incredible complexity of the human genome and its regulation. With the understanding that only 1.5% of the human genome is coding, the search began to establish the function, if any, of the remaining 98.5%. As exemplified by such studies as the ENCODE project, it is now believed that the function of much of these noncoding sequences is regulatory in nature. Furthermore, numerous whole genome sequencing, exome sequencing, and genome-wide association studies (GWAS) have found that the majority of single nucleotide polymorphisms (SNPs) that are associated with medically important traits lie outside of the coding region (1) and are thus believed to impact regulatory networks. However, most of these non-coding regulatory networks remain uncharacterized, and the impact of these SNPs on human health and disease is unclear.

Objectives/Hypothesis
We will utilize metabolomics data being produced through Dr. Wilson's (as joint PI with Drs. Robert Gerszten [Massachusetts General Hospital] and Thomas Wang [Vanderbilt University]) recently funded R01DK081572 "Metabolomic predictors of insulin resistance and diabetes", genome wide genotyping and whole genome sequencing from the Jackson Heart Study (JHS), and bioinformatic analysis to seek causal variants for metabolite profiles that are associated with incident type 2 diabetes (T2D). Furthermore, we will annotate the catalog of coding and non-coding SNPs and structural variations (SV) with emphasis on those that are associated with T2D and related traits. Building upon Dr. Wilson's previous work, we will specifically focus on those alleles believed to play a role in the increased prevalence of T2D in the African American population.

Statement of Relevance
GWAS have identified thousands of genetic variants that are statistically associated with medically relevant traits. However, a mechanistic description of these associations is lacking. More than 80% of SNPs identified by GWAS are intronic or intergenic (1). Bioinformatic tools and analyses are needed for functional annotation of these loci in order to identify causal variants and elucidate mechanistic descriptions.

It has been well established that genetic factors play a significant role in the risk of T2D, coronary heart disease (2) and chronic kidney disease (3) - diseases with higher prevalence in African-ancestry populations as compared to European-ancestry. However, due to sub-optimal genome-wide tagging panels based largely on genetic variation in Europeans, shorter blocks of linkage disequilibrium in persons of African ancestry, and smaller sample sizes in African-ancestry studies, there is a relative paucity of robust genetic association signals for these diseases among African Americans. The JHS cohort, specifically established to explore factors related to diseases in African Americans, with its extensive catalog of meticulous phenotypic information, provides a unique opportunity to identify genotype-phenotype correlations and establish functional annotations of loci relevant to disease in African Americans. Ultimately, the goal of these studies is to inform clinical intervention to improve health in these populations.

Methods
Identification of causal genetic variants by bioinformatic analyses requires two datasets: one which identifies those variants that are statistically associated with a given phenotype, and a catalog of known functional elements. This project aims to contribute to both datasets. First, single-variant association studies which seek to identify SNPs or other genetic variations (e.g. structural variations) that are statistically associated with a disease phenotype are underway in our laboratory. This project will build upon and continue this work in an effort to focus functional analysis (see below) on those variants most strongly associated with human disease. Second, in order to postulate the effect of the identified variants, it is necessary to establish the functional element(s) affected. To this end, we will utilize publicly available data and data from our collaborators to construct a catalog of functional elements for those tissues of interest (e.g. pancreatic beta cells, muscle, adipose tissue, kidney, and heart). For example, using data from the ENCODE project, we will identify tissue-specific enhancers and transcription factor binding sites across the genome. Cross-referencing this data with

alleles identified by single-variant association studies described above can distinguish between non-causal variants in linkage disequilibrium with the causal variant and suggest a possible mechanism for the observed association.

A second, related method to identify causal alleles is to examine the number of all variants that are singletons - an approach that assumes disease-causing variants will be selected against in the population. Expanding this approach to include other genomic and epigenomic data (e.g. chromatin state), it is possible to identify noncoding SNPs that are negatively selected, presumably due to disease-causing or other detrimental effects. For example, in sequences that exist in open chromatin in brain, lie <20 kb from brain-expressed genes, and contain sequences of known brain transcription factors such as NeuroD, we find that the fraction of all variants that are singletons exceeds the corresponding rate for missense variants in protein-coding sequence, indicating the importance of noncoding SNPs on human disease. Importantly, software developed by our collaborators at the Broad Institute already exists to perform such computations (4).

Anticipated Results/Potential Problems

We anticipate identification of candidate functional variants involved in the development of multiple disease phenotypes with emphasis on those variants of particular relevance to disease in African American populations and the Jackson Heart Study cohort, namely T2D and coronary heart disease. In addition, we anticipate deposition of cataloging data into scientific databases.

Notably, all methods proposed are well established in our laboratory or those of our collaborators at Massachusetts General Hospital and the Broad Institute. Furthermore, the computational resources necessary for the proposed studies are already established at the Jackson Heart Study. Therefore, we anticipate straightforward implementation of the proposed work. That being said, we do not exclude the emergence of unforeseen pitfalls. In such a circumstance, we believe that our lab has the expertise and skill in genetic research necessary to confront such difficulties.

Timeline

We expect that much of the proposed work is achievable over the course of three years. Specifically, with regards to Mr. Younes's involvement and the MSRP program, we plan to use the coming year (medical student year 2) to familiarize Mr. Younes with genetic research tools and parlance through review of the relevant literature and preliminary use of genetic analysis software. This will allow us to launch without delay into the heart of the project during the third and fourth year research electives.

References

1. Hindorff LA, Sethupathy P, Junkins HA et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc.Natl.Acad.Sci.U.S.A 2009;106:9362-9367.
2. Kullo I, Ding K. Mechanisms of Disease: The genetic basis of coronary heart disease. Nat.Clin.Pract.Card. 2007;4(10):558-569.
3. Naik RP, Derebail VK, Grams ME, et al. Association of Sickle Cell Trait With Chronic Kidney Disease and Albuminuria in African Americans. *JAMA.* 2014;312(20):2115-2125.
4. Trynka G, Westra HJ, Slowikowski K et al. Disentangling the effects of colocalizing genomic annotations to functionally prioritize non-coding variants within complex-trait loci. Am.J.Hum.Genet. 2015;97:139-152.

6. **Please list purchases/expenses and provide a short justification for each. Total budget not to exceed $5,000. Budget for sections "a" through "e" are not to exceed $4,000. Budget for section "f" (travel) may exceed $1,000 if justified.**
   a. **Supplies**
   b. **Equipment**
   c. **Computer** $1500
   d. **Software**
   e. **Other**
   f. **Travel** $3500 - Justification for travel expense: Mr. Younes will travel to Boston during each year of the project to spend 2-3 weeks working with Dr. Robert Gerszten at Massachusetts General Hospital and Drs. Sekar Kathiresan and Soumya Raychaudhuri at the Broad Institute. He will also attend analysis workshops of NHLBI's Trans-Omics for Precision Medicine (TOPMed) project.

   Total **$5,000**

-All computers and software purchased belong to the lab, not the individual student.

7. **Education and Training Plan**
   a. **Lab Meetings, Seminar Series, and Didactic Coursework**
   Mr. Younes will attend weekly project meetings with Dr. Wilson and participate in monthly JHS Genetics Working Group meetings, as well as teleconferences of the TOPMed project and its working groups. The Department of Physiology weekly seminar series offers an excellent educational opportunity to provide a broad knowledge base of research techniques, experiment interpretation, and presentation skills. In addition, it provides a venue for discourse on the current state of research both at the medical center and abroad, providing the student with a broad background in many areas of research. To this end, Mr. Younes will endeavor to attend this seminar series when feasible, especially those seminars with direct relevance to the proposed research.

   b. **Anticipated Abstracts, Presentations, and Publications**
   We anticipate at least two manuscripts:
   Identification of candidate functional variants affecting T2D-associated metabolites
   Identification of candidate functional variants affecting T2D and glycemic traits

   c. **Other Educational Opportunities (e.g., grant writing, ethics, statistics, etc.)**
   Mr. Younes will have a strong introduction to bioinformatics through his association with Drs. Gerszten, Kathiresan, and Raychaudhuri and their groups, and his participation in TOPMed analysis workshops.

   d. **Mentorship Plan**
   Both Dr. Wilson and Mr. Younes agree to meet weekly in order to stay abreast of the research topic (in accordance with the MSRP guidelines) and to explore potential opportunities for contribution during the year. Dr. Wilson will provide assigned reading, and Mr. Younes will work with Dr. Hao Mei (Center of Biostatistics) on computational approaches to genetic analysis.