PROJECT SUMMARY/ABSTRACT

More than 27 million Americans suffer from Type 2 diabetes (T2D). GWAS have identified 128 lead SNPs associated with T2D and/or fasting hyperglycemia, but little is known about how these variants contribute to T2D pathogenesis. A major challenge in functionally characterizing variants found in GWAS is that each lead SNP directly associated with a trait is in LD with a collection of additional variants, and thus identifying the precise variant(s) underlying the association requires extensive computational and experimental analyses. Additionally, the majority of the associated SNPs are located within non-coding regions, where inferring functional consequences of sequence variants remains challenging. Finally, when associated SNPs are identified as candidate regulatory variants, functional testing is frequently hampered by a lack of appropriate experimental models. To address these challenges we have assembled a team of highly accomplished researchers in genomics (Frazer), epigenomics (Ren) and T2D biology (Sander). We propose to combine state-of-the-field computational methods, high throughput molecular assays, and disease modeling in human embryonic stem cells to comprehensively annotate T2D GWAS data and test variants for their gene regulatory function. In Aim 1 we will analyze 5,150 whole-genomes for variants in T2D and fasting hyperglycemia GWAS risk-associated loci. We estimate that these analyses will identify ~1,000,000 variants with a MAF > 1% in the intervals of interest. Additionally, we will identify rare variants that are enriched in T2D patients. We estimate that intersecting these data with existing epigenomic datasets will identify ~99,000 variants in putative regulatory elements in T2D-relevant tissues. In Aim 2 we will use three high throughput molecular assays to characterize these 99,000 candidate regulatory variants in T2D-relevant cell types. First, we will carry out massively parallel reporter assays (MPRA) to test the potential of each SNP-harboring sequence element to act as a transcriptional enhancer, and if so, whether enhancer activity is affected by the candidate variant. Second, we will carry out a high throughput *in vitro* binding assay (SELEX) to determine whether the candidate variants affect DNA binding of relevant transcription factors. Third, we will predict target genes of the candidate variants using high throughput chromosome conformation capture (Hi-C) assays. In Aim 3, results from Aim 1 and Aim 2 will be integrated to prioritize 20 beta cell-relevant SNPs for functional validation. Key criteria include: (1) the variant resides in an active beta cell enhancer, (2) disrupts transcription factor binding, and (3) targets T2D-relevant genes in Hi-C assays. We will validate these variants by (1) genetic engineering of an embryonic stem cell-derived cell model of human beta cells, testing how deletion of the *cis*-regulatory element or introduction of the risk variant affects target gene expression; (2) genetic engineering of mouse models, testing whether candidate enhancer/target gene pairs control glucose metabolism *in vivo*. The proposed study will provide key insights into the underpinnings of regulatory variants identified through GWAS in T2D etiology.