Immune Checkpoint Inhibitor Induced Diabetes Mellitus:
A Unique Opportunity to Understand Beta-Cell Autoimmunity

Progress Report

Objective: To identify novel autoantibodies/autoantigens through use of proteome-wide programmable phage display technology (PhIP-Seq) in immune checkpoint inhibitor induced diabetes mellitus (CPI-DM) and Type 1 Diabetes Mellitus (T1DM) populations.

PhIP-Seq Cohort Type of Subject	Sample Size
CPI-DM	
Single Time Point	23
Longitudinal Time Points	33
T1DM	
Single Time Point	202
Longitudinal Time Points	2
T2DM (all single time point)	25
Healthy Control (all single time point)	158
CPI Control (all single time point)	77

Experimental Progress:

This project is composed of three stages: 1) completion of the PhIP-Seq experiment, 2) analysis of the PhIP-Seq data and 3) validation of any genes or peptides of interest.

STAGE 1: EXPERIMENT. The PhIP-Seq experiment was successfully completed on a group of 1152 samples, detailed in the table to the right. Recall that PhIP-Seq utilizes a custom designed phage library containing over 700,000 unique phage, each displaying a 49 amino acid proteome segment that tile the full protein-coding human genome including all isoforms with 25 amino acid overlap. This allows for identification of putative autoantigens at both the gene and peptide level relative to chosen control groups.

STAGE 2: ANALYSIS. I have primarily used three methods to analyze the PhIP-Seq data: 1) Shared gene among multiple patients, 2) longitudinal changes in genes across exposure to CPI and development of CPI-DM and 3) particularly high expression in unique patients. Review of candidate gene function and known associations was completed and survey of RNA and protein expression of candidate genes was done through use of BioGPS, GTex, and the Human Protein Atlas. The hope was that this would help to narrow the number of hits worth validating. This process is detailed in the figure below.

To maximize the ability to find novel hits that might be shared between CPI-DM, T1DM and auto-antibody negative DM by PhIP-Seq, I grouped these forms of diabetes together and then identified candidate genes that were present in at least 3% of the diabetes subjects and less than 3% of controls with a fold change of 25 over the background mock immunoprecipitate. This yielded 4 candidate genes: PTPRN (equivalent to IA2/ IC512, a known autoantibody), ZBTB26, KBTBD8, and NR5A1. I then loosened the criteria slightly, to allow more healthy controls (as they were blood bank samples and therefore could have had diabetes), and a lower fold change (15). This yielded 4 additional candidate genes: KLF7, TAS2R20, SYT10, C12orf42. I analyzed the longitudinal data by looking for genes that had at least a 10-fold change at CPI-DM relative to baseline, pretreatment. There was large variation in the number of candidate genes between different individuals. In total there were 978 genes that met this criterion. I then integrated these outputs to determine if any changes were present in more than one individual with CPI-DM. There were 27 candidate genes shared by two individuals and none shared by three or more individuals. Finally, I reviewed a subset of the top single genes that were present in each CPI-DM and T1DM subject that was not enriched in healthy controls for tissue specificity with the pancreas or islet cells depending on the database.

NAPPA Cohort Type of Subject	Sample Size
CPI-DM	
Single Time Point	4
Longitudinal Time Points	1
T1DM/LADA (all single time point)	5
Healthy Control (all single time point)	5
CPI Control (all single time point)	5

STAGE 3: VALIDATION. Due to the large number of candidate genes identified through the above methods, we attempted to use an initial alternative approach to validation through a disparate unbiased method called High Dimensional Nucleic Acid Programmable Protein Array (HD-NAPPA) through the Biodesign Institute of the University of Arizona. In contrast to PhIP-Seq, the cost per sample is quite high, and we therefore sent a limited number of samples, as detailed in the table to the right.

Interestingly, this platform did not corroborate the new findings from PhIP-Seq, possibly due to unique aspects of the platforms and possibly because these candidates will end up not replicating with additional methods.

I also used NAPPA to do explore for novel or known candidate proteins. Within this array, there is a suggestion that a Normalized IgG level should be considered "positive" at something between 3 to 5. Therefore, I identified proteins that had at least one individual within CPI-DM or T1DM that had a normalized IgG over 5 as long as there was no healthy control with a normalized IgG over 2 (as these controls were not from the blood bank, I definitively knew that none had diabetes). For CPI-DM, there were 96 hits in total, but after adjustment for 44 healthy controls with a shared hit, 52 remained. For T1DM there were 17 hits and for LADA there were 27 hits and after adjustment for healthy controls. I then took these candidate proteins back to the PhIP-Seq data and reviewed

Because of the lack of corroboration between the NAPPA and PhIP-Seq candidate antigens, I will now return to the original plan and use radio-ligand binding assays (RLBA) for further validation. This has been completed on one candidate gene, PGAP1 which has pancreas and pituitary specificity and was found to be high on PhIP-Seq in a subject with both CPI-DM and CPI induced hypophysitis. Unfortunately, the RLBA for PGAP1 was negative and did not validate. Overall, I am quite happy with the progress that we have made so far, although there were some delays for a newborn baby and additional COVID limitations. I expect that these RLBAs will be completed within the next couple of months.

Pending the results of the RLBA assays described in the figure below, an additional next step will be to allow for lower fold changes in PhIP-Seq or normalized IgG in NAPPA in "cases" as long as they are not present in controls. This may be especially high yield within the NAPPA platform as the current cutoff of 5 is likely higher than necessary. With results of RLBAs, we will also get a better sense of which hits are more likely to validate in further studies.

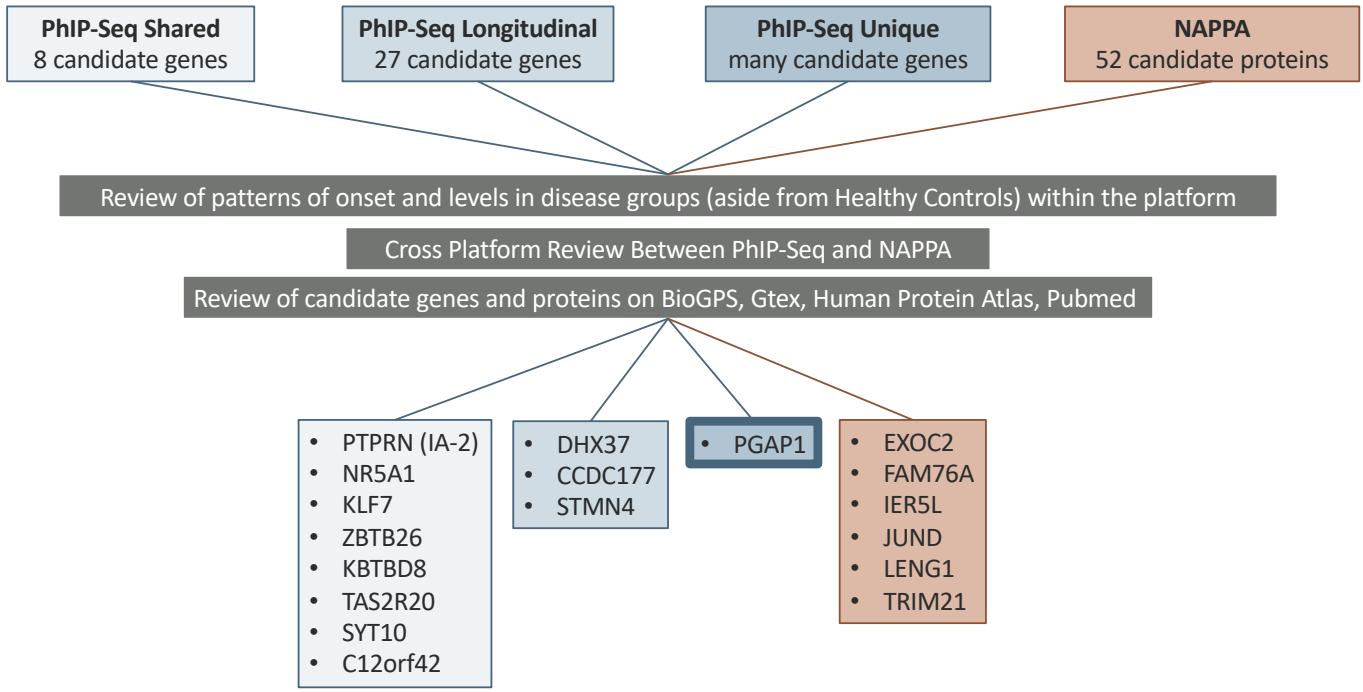

Figure 1. Analytic Process. Candidate genes and proteins identified through designated analytic method before and after investigation of biological plausibility, expression, and validation on the opposing platform. Thickened outlines indicate that further validation through an RLBA has been completed