Diabetic Complications Consortium

Application Title: snoRNAs in Complications of Type 1 Diabetes

Principal Investigator: Jean E. Schaffer, MD

1. Project Accomplishments:

This study sequenced snoRNAs in the DNA of 473 subjects from the Joslin Medalist Study. We focused our investigation on 267 of the known 305 box C/D snoRNAs in the human genome for which it was possible to design custom amplicons for targeted sequencing using the Illumina TruSeq Custom Amplicon platform. We obtained high quality sequence for more than 90% of the targeted snoRNAs. Although we hypothesized that sequence variation with the potential to alter expression or function of snoRNAs might be associated with susceptibility to diabetic retinopathy, we found no association in this sample set.

2. Specific Aims:

This project had a single Specific Aim: to systematically examine snoRNA sequences in a well-phenotyped cohort of subjects with longstanding diabetes to determine the relationship between variation in snoRNA sequences and diabetic retinopathy.

Results:

We employed the Illumina Tru-seq Custom Amplicon-based sequencing platform to perform a case-control study using samples from subjects participating in the Joslin Medalist Study, all of whom have had type 1 diabetes for 50 years or more. We initially selected 253 participants who had no or mild non-proliferative diabetic retinopathy as cases and 127 participants with proliferative diabetic retinopathy as controls. Cases and controls were otherwise matched for age, duration of diabetes, glycosylated hemoglobin, total cholesterol, LDL, HDL, triglycerides, and blood pressure. We obtained high quality sequence data from these 380 samples — more than 95% had sufficient coverage of greater than 95% of the snoRNA targets. Using a threshold of > 20 reads per target snoRNA, we identified box C/D snoRNA sequences that differed from the Hg19 human reference sequence and were minor alleles.

We observed a total of 281 minor alleles in 147 different snoRNAs. Minor alleles, when present, were generally heterozygous. Overall, 6% of alleles in cases were minor, compared to 5% in controls ($p < 2.2e-16$). We initially observed 17 minor alleles that were significantly overrepresented in cases compared to controls with p-values for the relationship between case status and these minor alleles, after correcting for multiple comparisons (Bonferroni), from 5.1e-22 to 5.5e-25. To validate these observations, we analyzed this subset of snoRNAs in an independent set of Joslin Medalist samples. These replication samples were from another 62 cases and 31 controls that were similarly defined and matched as our discovery set. Unexpectedly, none of the initially significant minor alleles was over-represented in the cases in the replication set.

To understand the lack of validation, we revisited the analysis of our initial discovery set of samples. We found that most of the variants in cases derived from a single 96-well plate of sequencing samples, raising concern

regarding potential artifact. We also repeated our analysis of the discovery sample set data, raising the threshold for minor alleles to 40% of reads (which would be close to the expected frequency for alleles present in heterozygous dose). On the basis of this more stringent analysis, several of the initial variant calls no longer met criteria. Furthermore, in our re-analysis, we took into consideration the generally lower quality of sequence found near the end of reads by requiring that variants be present in the sequence obtained in both directions. A number of the initially called variants were near the ends of snoRNAs, but were present at the end of a read from one-direction only. We no longer considered such events to meet criteria for a variant call. With the 15 remaining variants that survived these stricter criteria, our repeat analyses showed no variants were significantly over-represented in cases compared to controls. Nor did we find a general overrepresentation of snoRNA variants in the cases in collapsing analyses, in which all observed variants within a particular snoRNA were considered as a single outcome variable. Preliminary comparison to the 1000 Genomes database suggests that some of the variants we observed could be novel, and we are extending these comparisons to other available databases.

In conclusion, our study failed to identify significant differences in minor allele frequencies between the cases and control Joslin Medalist subjects. One important factor that may have contributed to the lack of positive findings is the relatively small number of samples we were able to analyze due to cost-considerations. A second factor may relate to limitations of the sequencing platform we chose. While the TruSeq Amplicon approach was well suited for targeted sequencing of multiple small stretches of genomic sequence and cost-effective over alternative capture approaches, it is possible that the primer ligation step at the end of the initial round of reverse transcription was error prone. Nonetheless, studies in our laboratory have clearly demonstrated that alterations in snoRNA expression have profound effects on the response to metabolic stress in a number of tissues in mouse models. In the future, as more human whole genome sequence data becomes available, we will leverage existing data from large cohort studies to assess the relationship between variant snoRNA sequences and metabolic disease phenotypes.

3. Publications:

None as yet