Diabetic Complications Consortium

Application Title: snoRNAs in Complications of Type 1 Diabetes

Principal Investigator: Jean E. Schaffer, MD

Project Accomplishments:

small nucleolar RNAs (snoRNAs) play a critical role in metabolic and oxidative stress-induced cell death in cultured cells and animal models. However, little is known regarding potential human snoRNA sequence variation, because most mammalian snoRNAs are encoded within the introns of genes that are not assessed in exomic sequencing approaches. We hypothesize that rare variation in snoRNA sequences in the human genome contribute to variable expression of end-organ diabetic complications from metabolite-induced tissue damage. In this project we have systematically examined box C/D snoRNA sequences in a well-phenotyped cohort of diabetic subjects to determine the relationship between variation in snoRNA sequences and the development of diabetic retinopathy.

Specific Aims:

We used Illumina's Tru-seq CustomAmplicon based sequencing system to design amplicons for sequencing box C/D snoRNAs across the human genome. Of the 305 known and predicted snoRNAs, we successfully designed 330 amplicons to cover 267 snoRNAs. However, for 38 snoRNAs, we were unable to design unique amplicons, because these sequences lie within highly repetitive regions of the genome.

We obtained high quality genomic DNA from subjects who are participating in the Joslin Medalist Study, all of whom have had type 1 diabetes for 50 years or more. We selected 253 participants who have 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, HbA1c, total cholesterol, LDL, HDL, triglycerides, and blood pressure. DNA from each individual was sequenced using the above described amplicons for box C/D snoRNAs and Illumina's MiSeq high throughput platform. We considered that for each individual, 20 reads per snoRNA would be sufficient to determine the presence of variant alleles. Using a threshold of > 20 reads per target snoRNA, we obtained sufficient coverage of more than 95% of the box C/D snoRNAs. Of the 380 DNA samples we studied, more than 95% had sufficient coverage of more than 95% of the snoRNA targets.

We considered that box C/D snoRNA sequences that differed from the Hg19 human reference sequence were minor alleles. There were a total of 281 minor alleles in 147 different snoRNAs. Among these 147 snoRNAs, we detected only one minor allele in many, but there were 18 minor alleles in SNORD13. Minor alleles, when present were generally heterozygous. Thus, we observed many variant sequences in many snoRNAs. Preliminary analysis indicates that a number of these variants have not previously been reported in the 1000 Genomes database.

We next assessed if minor alleles were differentially present between cases and controls. Overall 6% of alleles in cases were minor, compared to 5% in controls ($p < 2.2e-16$). We observed 17 minor alleles that were significantly overrepresented in cases compared to controls (or were absent in controls): SNORD114-5 (2 alleles), SNORD114-9 (3 alleles), SNORD114-10 (1 allele), SNORD114-11 (2 alleles), SNORD114-16 (1 allele), SNORD114-17 (2 alleles), SNORD114-21 (2 alleles), SNORD54 (1 allele) SNORD77 (2 alleles), and SNORD77.2 (1 allele). P-values for the relationship between case status and these minor alleles, after correcting for multiple comparisons (Bonferroni), ranged from 5.1e-2 to 5.5e-25, suggesting that these observations are highly significant. The SNORD114 species are encoded within a locus on chromosome 14 that contains 40 snoRNAs, and it is likely that the different minor alleles we observed are linked. More than 50% of cases had 9 or more significant minor alleles, whereas 50% of

controls had 3 or fewer significant minor alleles. Together these observations indicate that minor alleles are overrepresented in cases, who are relatively spared from diabetic retinopathy.

Minor alleles contained variant nucleotides in a range of positions relative to the mature snoRNA sequence that is processed from host introns. A number of variants were observed in the several nucleotides 3′ to the D box motif. We hypothesize that variant nucleotides in this location may impact the ability of cells to process the mature snoRNA out of precursor intron lariats. A number of variants were observed in the predicted antisense element 5′ to the D box motif. We hypothesize that variant nucleotides in this location may impact the ability of the mature snoRNA to bind and act upon its target RNA.

Future Directions

Ongoing analysis includes comparison with recently updated data from the 1000 Genomes project. We are also performing analyses in which all rare variants within a particular snoRNA are collapsed into a single burden variable that will be compared between cases and controls. Moreover, studies are underway to validate our observations in an additional 99 Medalist subjects: 66 with no/minimal non-proliferative diabetic retinopathy and 33 with proliferative diabetic retinopathy. These subjects are matched for other characteristics in a similar fashion as the discovery sample set described above.

Our study was performed in an unusual group of type 1 diabetics who have had diabetes for more than 50 years. It will be important to extend our findings to diabetics who have had disease of shorter duration. While the Joslin Medalist subjects are well-phenotyped, there is limited data on their longitudinal diabetic control. As we validate our findings in other diabetics, one goal will be to leverage studies whose resources include more extensive data on diabetic control.

Although we have observed minor alleles that are significantly associated with variation in disease phenotype, it is not possible to discern whether the variant sequences are causative or simply represent nucleotide variants linked to a nearby causative genetic element (i.e., SNP). To test directly the variant snoRNA sequences, we will express genomic constructs for variant versus wild type snoRNAs in cell lines and assess whether variant sequences affect the production or function of the encoded snoRNAs.

Significance

Our study provides the first systematic examination of sequence variation in box C/D snoRNAs in diabetes. Our findings suggest that snoRNA variants have potential to serve as biomarkers of risk of a diabetic complication. If future studies indicate that the snoRNA variants themselves affect the response to metabolic stress and diabetes complications, then our findings could provide the basis for future therapeutic approaches leveraging antisense technology.

Publications:

None as yet