Diabetic Complications Consortium

Application Title: Novel American Indian exomic variants and diabetic nephropathy.

Principal Investigator: Nora Franceschini, MD MPH

1. Project Accomplishments:

During the one-year funding period of this project, we provided updated annotation to existing exome sequencing data from American Indians of the Strong Heart Family Study (SHFS) to select novel variants not available in repositories including dbSNP, 1000 Genomes Project and Exome Sequencing Project. We used this information to design a custom iSelect Illumina panel for genotyping. We then selected chronic kidney disease (CKD) case-control samples from American Indians of the SHFS for genotyping. The genotyping of the single nucleotide variants (SNVs) was performed at the Texas Biomedical Research Institute Genetic Laboratory, followed by quality control of the data, as described below. We provide a summary of the preliminary results for association analyses.

2. Specific Aims:

Specific Aim 1. Assess the associations of novel SNVs in known kidney loci identified in exome sequencing of American Indians with CKD-related diabetes complications.
Hypothesis: Rare and common American Indian novel exonic SNVs in kidney loci confer risk to CKD.

Specific Aim 2. Examine the associations of novel American Indian SNVs in known T2D loci with CKD-related diabetes complications.
Hypothesis: Rare and common American Indian novel exonic SNVs in validated T2D loci confer risk to CKD.

Results:

1. Study population and selection of CKD cases and controls. This study uses samples from the Strong Heart Study, a large population-based cohort study of cardiovascular disease in American Indians, which recruited individuals from tribes in Arizona, Oklahoma, and South and North Dakota [1, 2]. We selected individuals from the family study, which began in 1998 and examined 3,776 individuals in 94 multigenerational families (mean family size of 40 individuals, range 5 to 110) in two clinical visits (Phase 4 and Phase 5). The study has been approved by the Institutional Review Boards of the participating Institutions, and by the Indian tribes [3, 4]. All participants gave informed consent for genetic testing. Standardized anthropometric, clinical and 12-hour fasting laboratory measures were collected at each visit (including serum creatinine) as well as spot urine samples for albumin and creatinine measurement [3, 4]. Diabetes is defined by a fasting glucose ≥126 mg/dl, A_{1c} >6.5% or use of diabetic medications [5].

For this pilot study, CKD cases were selected based on a diagnosis of kidney failure (dialysis or transplant), reduced estimated glomerular filtration rate (eGFR) of less than 60 ml/min/1.73 m^2 and/or increased UACR ≥30 mg/g in any of the two longitudinal clinic visits. Individuals with an eGFR>80 ml/min/1.73 m^2 and an UACR<30 mg/g in both clinical visits were considered controls. To better match the age of cases, we excluded controls aged < 40 years.

2. Selection of SNVs and custom genotyping panel. Using whole exome sequencing data from 94 American Indian participants of the Strong Heart Family Study, we selected 2,709 SNVs considered novel based on comparison with dbSNP, 1000 Genome Project and Exome Sequencing Project. Additional selection criteria included presence of the SNV in at least two individuals, and predicted function priority based on Genome Variant Server (GVS): frameshift, splice-3, splice-5, stop-gain of function, stop-loss of function, and missense variants. These SNVs were included in a custom iSelect Illumina panel for genotyping of 1,152 samples (including 24 blind duplicates).

3. Genotyping quality control (QC) and filters/exclusions.
SNVs: 144 SNVs failed manufacturing and 1,357 SNVs were not polymorphic. Of the remaining 1,208 heterozygous SNVs, 1,206 had a call rate > 95%.
Participant samples: After exclusion of blind duplicates, data from an additional sample was excluded due to low call rate, resulting in data from a total of 1,127 individuals included in analyses.

4. Description of case-control samples. Among 1,127 genotyped individuals, 555 are CKD cases and 572 are controls. These individuals were selected from 24 pedigrees within the Dakotas and Oklahoma centers. CKD cases are individuals with kidney failure (dialysis or kidney transplant, N=28), an eGFR < 60 ml/min/1.73 m^2 clinic visits (N=233), or an UACR ≥ 30 mg/g (N=322). Additional demographic and clinical data for cases and controls are shown in Table 1. CKD cases are slightly older, more often have diabetes and hypertension at baseline compared to controls. The average eGFR is lower and median UACR is higher in cases than controls.

Table 1. Characteristics of CKD case-control samples

Characteristics	Case (555)	Control (572)
Mean age (SD)	53.8 (17.5)	44.4 (8.8)
Sex % women	62.3	63.8
Type 2 diabetes	42.3	15.4
Hypertension	57.0	30.2
Mean eGFR (SD)	82.8 (30.4)	97.5 (15.4)
Median UACR (interquartiles)	26.1 (9.2-91.4)	6.5 (4.5-10.2)

5. Preliminary results for association with novel variants. We performed a case-control association analysis for each SNV using mixed models to account for family relatedness (variance component using SOLAR software). We model allelic dosage (additive genetic model) and additionally adjusted for age, sex, and diabetes. The main findings from these analyses are shown in Table 2. Given the genetic correlation among SNVs due to linkage disequilibrium, we used a p-value threshold for significance of < 4.6×10^{-5}, which accounts for 1,104 independent tests. None of the associations reached the significance threshold, but five low frequency and

rare SNVs were associated with CKD but p-values did not reach the threshold for significance (Table 2). These include two missense variants associated with 35% and 22% reduction in the odds of CKD, respectively, at the *NID2* and *PREPL* genes, and three SNVs (a missense and two frameshift) associated with 22% to 65% increased odds of CKD at the *CAPN12, ZNF845* and *OR2H1* genes. Mutations at the *PREPL* gene have been associated with hypotonia-cystinuria syndrome, a disease that includes nephrolithiasis and CKD[6]. *NID2* encodes nidogen-2, an extracellular matrix protein ubiquitous in basement membranes, which links the laminin and collagen IV networks.

Table 2. Main association findings for the CKD case-control analysis (sorted by p-values)

SNV (Chr:position)	A1/A2	Freq A1	Beta	P	OR	N	SNV function	Gene
19:39229112*	G/A	0.005	0.5037	$7.4x10^{-4}$	1.65	1124	Missense	*CAPN12*
14:52509623*	A/G	0.005	-0.4281	$3.0x10^{-3}$	0.65	1126	Missense	*NID2*
2:44554030	A/G	0.01	-0.2502	$6.5x10^{-3}$	0.78	1126	Missense	*PREPL*
19:53854457-del	TC/T	0.02	0.2087	$6.6x10^{-3}$	1.23	1126	Frameshift	*ZNF845*
6:29430260-del*	CT/C	0.02	0.1987	$6.9x10^{-3}$	1.22	1126	Frameshift	*OR2H1*

*ExAC

Three of these variants are included in the Exome Aggregation Consortium (ExAC) browser (http://exac.broadinstitute.org/). The SNV at *CAPN12* (p.Leu279Pro) and *NID2* (p.Glu486Lys) are reported only in Hispanic/Latinos samples in ExAC (allele frequencies 0.0025 and 0.0016, respectively), and the frameshift at *OR2H1* (p.Cys239AlafsTer52) is rare in European and African samples (allele frequencies of 0.00019 and 0.00033, respectively) but has low frequency in Hispanic/Latino samples (0.027).

6. Summary of findings and conclusions. In summary, our pilot study adds to the literature reporting essential information on (1) the important contributions to be gained by conducting studies using American Indian samples and data to uncover genes associated with complex traits, and (2) presence of low frequency Amerindian exonic variants influencing the risk of CKD that are also segregating in Hispanic/Latino populations. Our preliminary findings strongly support further studies of Amerindian variants associated with CKD risk.

None of the identified association were for variants within genes previously associated with kidney traits or diabetes in GWAS, as hypothesized in this pilot study. These findings are of interest as variants identified in GWAS are usually common in populations and are markers of functional variants which likely have a regulatory function. In contrast, our genotyped variants were selected because of their protein-coding changes, and more often were low frequency variants. These variants are unlikely to be identified in GWAS or in populations that don't have Amerindian admixture.

6. Future work:

Our future plans include linking gene expression data to our genotypes to assess if the predicted damaging variants affect gene expression, and incorporating information from pedigrees in the family dataset. For example, we identified several complex (multi-generational) pedigrees with one or more individuals carrying a copy of these alleles. We plan to also genotype ancestry informative markers to adjust for population stratification in analyses, and additional participants within pedigrees segregating these variants. This expanded study will allow for analyses within strata of diabetes, although the power for these analyses may be low. Additional follow-up could

be done using experimental models to investigate allele-specific or combined effects of the variant *in vitro* or *in vivo*.

References

1. Lee ET, Welty TK, Fabsitz R, Cowan LD, Le NA, Oopik AJ, Cucchiara AJ, Savage PJ and Howard BV. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol*. 1990;132:1141-55.
2. North KE, Williams JT, Welty TK, Best LG, Lee ET, Fabsitz RR, Howard BV and MacCluer JW. Evidence for joint action of genes on diabetes status and CVD risk factors in American Indians: the strong heart family study. *Int J Obes Relat Metab Disord*. 2003;27:491-7.
3. Lee ET, Welty TK, Fabsitz R, Cowan LD, Le NA, Oopik AJ, Cucchiara AJ, Savage PJ and Howard BV. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol*. 1990;132:1141-55.
4. North KE, Howard BV, Welty TK, Best LG, Lee ET, Yeh JL, Fabsitz RR, Roman MJ and MacCluer JW. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. *Am J Epidemiol*. 2003;157:303-14.
5. Maples BK, Gravel S, Kenny EE and Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet*. 2013;93:278-88.
6. Jaeken J, Martens K, Francois I, Eyskens F, Lecointre C, Derua R, Meulemans S, Slootstra JW, Waelkens E, de Zegher F, Creemers JW and Matthijs G. Deletion of PREPL, a gene encoding a putative serine oligopeptidase, in patients with hypotonia-cystinuria syndrome. *Am J Hum Genet*. 2006;78:38-51.