1.a. Full Title: Genetic risk score for type 2 diabetes

b. Abbreviated Title (Length 26 characters): GRS for diabetes

2. Writing Group:
   Writing group members:

   Jim Pankow  
   Linda Kao  
   Kari North  
   Jim Peacock  
   Aaron Folsom  
   Eric Boerwinkle  
   Others

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __JP__ [please confirm with your initials electronically or in writing]

First author: Jim Pankow  
Address: Division of Epidemiology and Community Health  
          University of Minnesota  
          1300 South Second Street, Suite 300  
          Minneapolis, MN 55454

Phone: 612-624-2883  
Fax: 612-624-0315  
E-mail: pankow@epi.umn.edu

Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author):
Address:

Phone:  
Fax:  
E-mail:

3. Timeline:
4. Rationale:

A series of large-scale genome-wide association (GWA) studies have recently identified variants in ten genes or chromosomal regions that are reproducibly associated with type 2 diabetes (see table below) (1-5). The strongest and most robust of these is the TCF7L2 rs7903146 T allele (6, 7). A meta-analysis of 17,202 cases and 29,195 controls from 28 study populations found a pooled odds ratio of 1.46 (p=5 x 10^{-140}) for this allele (8). GWA studies have also implicated SNPs in the CDKN2A/2B region of chromosome 9, IGF2BP2 on chromosome 3, CDKAL1 on chromosome 6, HHEX on chromosome 10, SLC30A8 on chromosome 8, FTO on chromosome 16, PPARG on chromosome 3, KCNJ11 on chromosome 11, and an intergenic region on chromosome 11.

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDKN2A/2B</td>
<td>rs10811661</td>
<td>9</td>
</tr>
<tr>
<td>IGF2BP2</td>
<td>rs4402960</td>
<td>3</td>
</tr>
<tr>
<td>CDKAL1</td>
<td>rs7754840</td>
<td>6</td>
</tr>
<tr>
<td>HHEX</td>
<td>rs1111875</td>
<td>10</td>
</tr>
<tr>
<td>SLC30A8</td>
<td>rs13266634</td>
<td>8</td>
</tr>
<tr>
<td>TCF7L2</td>
<td>rs7903146</td>
<td>10</td>
</tr>
<tr>
<td>FTO</td>
<td>rs12255372</td>
<td>16</td>
</tr>
<tr>
<td>PPARG</td>
<td>rs1801282</td>
<td>3</td>
</tr>
<tr>
<td>KCNJ11</td>
<td>rs5219</td>
<td>11</td>
</tr>
<tr>
<td>Unknown</td>
<td>rs9300039</td>
<td>11</td>
</tr>
</tbody>
</table>

The functional relevance of most of these genes / regions is unclear. The FTO gene is associated with greater adiposity and this appears to completely explain its association with diabetes (9). However, it has been hypothesized that many, if not most, of the genes influence insulin secretion because of putative effects on pancreatic beta cell function. For example, KCNJ11 encodes the islet ATP-sensitive potassium sodium channel Kir6.2, SLC30A8 is a pancreatic beta-cell specific zinc transporter, and CDKAL1 shares strong homology to CDK5, which has been implicated in the regulation of beta cell function. If some of these genes operate through a common pathway (i.e., pancreatic beta cell function), then it is plausible that disease-causing alleles in these genes may act synergistically (i.e., gene-gene interaction).

Although the relative risks associated with these gene variants are relatively modest, generally ranging from 1.1-1.4 per copy of each risk allele, their combined effect may be substantially larger. Using data from the FUSION study, Scott et al. (3) ranked individuals according to their genetic risk based on a logistic regression model containing the ten risk variants listed in the table above. Subjects in the highest 5% of predicted risk had a prevalence of type 2 diabetes that was nearly four times higher than subjects in the lowest 5% of genetic risk. However, the authors acknowledged that prediction in FUSION may be biased because of overestimation of relative risks due to post hoc
selection of SNPs, enrichment for familial type 2 diabetes cases, and exclusion of non-cases with impaired glucose tolerance or impaired fasting glucose. Therefore, it is unclear whether or not a genetic risk score will provide similar levels of prediction in other population-based samples. It is also unknown whether prediction of diabetes can be substantially improved by adding genetic variants to models with established risk factors.

5. Main Hypothesis/Study Questions:

This manuscript will address three study questions:

(1) Can the genetic risk score for diabetes developed in the FUSION study be replicated in the ARIC cohort?

(2) Is there interaction between diabetes-related SNPs in their association with diabetes?

(3) Does a genetic risk score containing the ten SNPs described above, possibly supplemented with other SNPs typed in the full cohort [rs780094 in GCKR (nominally associated with fasting glucose, HOMA-IR, and diabetes in subjects from Finland and Sweden (2)) and rs3792267 in CAPN10 (associated with diabetes in ARIC blacks (10))], add to the prediction of incident diabetes beyond that provided by established risk factors for diabetes?

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

All analyses described below will be stratified by race and subjects who are not white or blacks will be excluded. Analysis of blacks is of particular interest because most previous studies of these SNPs have been limited to whites. If data on ancestry informative markers or genome-wide association markers are available, correction for hidden population substructure in Aim 2 (i.e., white subjects) will be performed using principal components analysis as implemented in the EIGENSTRAT algorithm (11).

For study question 1, separate analyses will be conducted for prevalent diabetes at visit 1 (cross-sectional analysis using logistic regression) and incident diabetes (longitudinal analyses using Cox proportional hazards regression). Similar to Scott et al., predicted risk of diabetes will be estimated for each subject using a model containing the ten variants listed in the table above and subjects will then be placed into categories of predicted risk. The actual pattern of risk will then be compared across categories of predicted risk.

For study question 2, we propose to combine prevalent diabetes at visit 1 and incident diabetes in order to improve power to test gene-gene interactions. Logistic regression will be used. Each SNP will be modeled assuming additive allelic effects (variables coded 0, 1, or 2) and interactions will be evaluated by testing the significance of a product term between SNP variables. Only two-way interactions will be considered, and
the level of significance will be adjusted by a Bonferroni correction for the 45 possible pairwise combinations of SNPs (=0.05/45).

For study question 3, only incident diabetes will be considered and all subjects with prevalent diabetes or missing diabetes information at visit 1 will be excluded. Diabetes prediction will utilize a diabetes genetic risk score (GRS) similar to the CHD genetic risk score used by Morrison et al. (12) for coronary heart disease. Unless prior analyses provide strong evidence of gene-gene interaction, the GRS will be derived by assigning value of 1 for the risk homozygote, 0 for the heterozygote, and –1 for the nonrisk homozygote. The GRS for each individual will be the sum of these values for each SNP. The area under the curve will be compared between a model containing traditional risk factors alone and a model containing traditional risk factors and the GRS. Based on the risk functions developed by Schmidt et al. in the ARIC cohort (13), the following traditional risk factors will be included: age, parental history of diabetes, fasting glucose, systolic blood pressure, waist circumference, HDL cholesterol, and triglycerides. However, given that Schmidt et al. defined incident diabetes based on the visit 4 oral glucose tolerance test rather than the standard ARIC definition as proposed here, predictors identified in other study populations (e.g., sex, height BMI, smoking, physical activity, coffee, and alcohol consumption) will also be considered. As part of this analysis, we will also evaluate whether the magnitude of association between parental history of diabetes and incident diabetes is attenuated when the diabetes GRS is included in the model. Such a pattern of results would suggest that the SNPs included in the GRS account at least partially account for the familial aggregation of diabetes.

7.a. Will the data be used for non-CVD analysis in this manuscript?  ___ Yes  ___ No

   b. If Yes, is the author aware that the file ICTDER02 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES_DNA = “CVD Research” would be used?  ___

      Yes  ___ No

      (This file ICTDER02 has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript?  ___ Yes  ___ No

   b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER02 must be used to exclude those with value RES_DNA = “No use/storage DNA”?  ___ Yes  ___ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at:  http://www.csec.unc.edu/ARIC/search.php
10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

Manuscript #357D (Garrant et al., *Diabetes* 2002; 51:231-237) investigated associations between SNP 43 in *CAPN10* and prevalent or incident diabetes in a subset of ARIC blacks, but did not evaluate combinations of diabetes-related genes.

Manuscript #1141 (Yen et al., in preparation) is investigating associations of *TCF7L2* SNPs and prevalent or incident diabetes in the full ARIC cohort, but will not evaluate combinations of diabetes-related genes.


11. a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ____ Yes ___ No

11.b. If yes, is the proposal
___ A. primarily the result of an ancillary study (list number* __________)
___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __________ __________ __________)

*ancillary studies are listed by number at [http://www.cscc.unc.edu/aric/forms/](http://www.cscc.unc.edu/aric/forms/)

12. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

References:

1. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.