ARIC Manuscript Proposal # 1238

1.a. Full Title: DNA-DAMAGE PATHWAY AND GENETIC SUSCETIBILITY TO TYPE 2 DIABETES MELLITUS AND INSULIN RESISTANCE STATES

b. Abbreviated Title (Length 26 characters): Checkpoint 2 gene, diabetes and insulin resistance.

2. Writing Group: Kari North and Nora Franceschini
   Writing group members: Eric Boerwinkle, Christy Avery, James Pankow, Christy Ballantyne, Linda Kao (any other interested ARIC investigator)

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

First author: Kari E North, PhD
Address: Department of Epidemiology
         University of North Carolina Chapel Hill
         Bank Of America Center
         137 E. Franklin St., Suite 306
         CB #8050
         Chapel Hill, NC 27514

Phone: (919) 966-2148    Fax: (919) 966-9800
E-mail: kari_north@unc.edu

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

Phone:    Fax:
E-mail:
3. **Timeline:** Undergoing analysis, gene SNPs already genotyped by Dr Boerwinkle’s laboratory on the ARIC cohort. Ancillary study related to this manuscript is approved. This work will lead to a manuscript within six to 12 months.

4. **Rationale:**

Type 2 diabetes is associated with a large burden of morbidity and mortality worldwide and accounts for a large proportion of health care cost (Hogan, Dall et al. 2003). Hyperglycemia can result from impaired insulin secretion by pancreatic β-cells, increased hepatic gluconeogenesis (Magnusson, Rothman et al. 1992), and skeletal muscle insulin resistance (Petersen and Shulman 2006). Understanding the metabolic pathways that lead to type 2 diabetes is important and may lead to identification of new therapeutic targets and preventive measures for the disease. In this proposal, we suggest that type 2 diabetes and DNA damage pathway may be linked in humans.

Exposure to environmental and metabolic agents can damage DNA, e.g., DNA double-strand breaks are triggered by the generation of reactive oxygen species (ROS) (O'Driscoll and Jeggo 2006). Most of the ROS is produced via mitochondrial respiration. Another source of ROS is endoplasmatic reticulum (ER) stress, occurring in cells with high secretory capacity such as immunoglobulin-secreting cells and insulin-secreting pancreatic β-cells. The β-cells sense the glucose level by first transporting glucose into the cell through glucose transporter 2. Transported glucose is then metabolized to ATP through glycolysis and oxidative phosphorylation. ATP in turn triggers depolarization of the membrane, calcium influx and secretion of insulin. Thus, high glucose level in the blood leads to high ATP production, which results in ROS production. In addition, high secretory rate of insulin must be supported by high translation rate of insulin mRNA in the ER, which results in increased ER stress and ROS generation. Pancreatic β-cells have low levels of antioxidants to buffer ROS, which leads to ROS toxicity. DNA breaks and, in susceptible individuals, β-cells apoptosis will occur as the burden for insulin production increases. When β-cells can no longer sustain the demand for insulin production, either because of dysfunction or reduction in mass of β-cells, type 2 diabetes occurs.

DNA damage promotes genomic instability through mutations and chromosomal rearrangements. In turn, DNA breaks trigger a coordinated system of diverse cellular responses leading to apoptosis, cell cycle checkpoints and DNA repair. These response pathways are called DNA-damage or DNA break pathways (Niida and Nakanishi 2006; O'Driscoll and Jeggo 2006). DNA breaks are sensed by a family of PI-3-like kinases: DNA-PK, ATM and ATR. These kinases, in turn, activate effector kinases such as CHEK1 and CHEK2. CHEK2 is an effector kinase that activates p53, which can lead to cell arrest and apoptosis. Chung et al. showed that DNA-PK null mice and the CHEK2 null mice are glucose intolerant, due to a defect in the inability to produce and secrete adequate amount of insulin to maintain normal glucose homeostasis (unpublished data). Thus, CHEK2 and other genes in the DNA-break response pathway are potential candidates for type 2 diabetes susceptibility.

Genetic and environmental factors have been implicated in the susceptibility to type 2 diabetes. Although the progress to identify susceptibility genes has been slow,

Genome-wide scans for linkage have identified additional susceptibility genes on chromosomes 1q21-24, 2q37, 8q11.23, 12q24, 20q and 22q12.1. In HyperGEN study participants, we previously detected a QTL influencing diabetes variance on chromosome 22 (robust LODs = 2.7, 1.7 and 1.0 in all samples, Caucasians only, and African Americans only, respectively) (Avery, Freedman et al. 2004). Interestingly, the microsatellite marker nearest our peak linkage signal, GATA21F03, is just 0.22 MB from the *CHEK2* gene. Therefore, we tested six SNPs (four tag SNPs selected by Dr. North and two SNPs selected by Dr. Chung) in the *CHEK2* gene for evidence of association with type 2 diabetes in hypertensive siblings and their offspring and/or parents in HyperGEN.

The diabetic phenotype was adjusted for sex, study center, age, and age$^2$ using a maximum likelihood method that modeled affection status by a liability threshold model as implemented in SOLAR. All analyses were stratified by race. The additive genetic effect of each SNP was assessed independently using the measured genotype approach in 1531 European American and 1584 African American participants (SOLAR). Two of the four tag SNPs and one of the other two SNPs selected by Dr. Chung demonstrated evidence of association with type 2 diabetes in Caucasians in SOLAR ($P < 0.05$). A fourth SNP (one of the tagSNPs) was borderline significant in SOLAR ($p = 0.10$) but became significant using FBAT software. No significant findings were detected in African Americans. Given the role of *CHEK2* in glucose homeostasis in mice and our preliminary association results among European American HyperGEN participants, we propose to investigate the role of *CHEK2* variants on genetic susceptibility for type 2 diabetes and insulin resistance states in humans by replicating the findings in African American and European American participants of the Atherosclerosis Risk in Communities (ARIC) study.

5. **Main Hypothesis/Study Questions:**

We submit that type 2 diabetes and DNA damage pathway may be linked in humans. Given the role of *CHEK2* in glucose homeostasis in mice and our preliminary association results among European American HyperGEN participants, we propose to investigate the role of *CHEK2* variants on genetic susceptibility for type 2 diabetes and insulin resistance states in humans by replicating the findings in African American and European American participants of the Atherosclerosis Risk in Communities (ARIC) study.
We propose to implement the following aims:

**AIM 1.** Evaluate the association of SNPs/haplotypes of the *CHEK2* gene with type 2 diabetes in African American and European American participants of the ARIC study.

**AIM 2.** Evaluate the association of SNPs/haplotypes of the *CHEK2* gene with prediabetes phenotypes including fasting glucose, fasting insulin, impaired fasting glucose (IFG), impaired glucose tolerance (IGT), HbA1c, and HOMA-IR.

**AIM 3.** Evaluate the association of SNPs/haplotypes of the *CHEK2* gene with type 2 diabetes complications of the eye, kidney and heart, including diabetic retinopathy, microalbuminuria, and impaired (high frequency) cardiac autonomic tone.

**AIM 4.** Assess the putative modification of the aforementioned associations by biological and environmental attributes, including gender, physical activity, and body mass index (BMI).

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).**

**Design:** cross-sectional analysis of prevalent type 2 diabetes and longitudinal analysis of incident type 2 diabetes. Some design for other phenotypes.

**Inclusion/Exclusion criteria:** We plan to include all participants without missing data on the outcome variables and Chek2 SNP data. Ethnic minorities in field centers will be excluded. For incident analysis, all prevalent cases will be excluded.

**Outcomes:** type 2 diabetes, prediabetes phenotypes (fasting glucose, fasting insulin, HbA1c and HOMA-IR) and diabetes-related complications (diabetic retinopathy, chronic kidney disease (defined as MDRD estimated glomerular filtration rate less than 60 ml/min/m² and/or microalbuminuria) and and impaired (high frequency) cardiac autonomic tone.

**Analytical Strategy to Address Specific Aims**

For each Aim, a multiplicative, dominant (XX OX vs. OO), and general model (XX vs. OO and OX vs. OO) will be utilized to analyze the gene-disease association. All analyses will be stratified by race to account for population stratification. Haplotype analysis will be performed as described below. Type 2 diabetes will be defined using the WHO definition.

For **AIM 1.** Prevalence odds ratios (OR) will be calculated from logistic regression models with prevalent type 2 diabetes as the response variable and *CHEK2* SNP/haplotypes as the main exposure of interest. Potential confounders will include
BMI, gender, physical activity, cigarette smoking, center, age at baseline, family diabetes history, and hypertension.

To estimate the 9-year relative risk of type 2 diabetes incidence associated with variants of the CHEK2 gene, individuals with diabetes at baseline will be excluded from the time-to-event analyses. Since almost all the type 1 diabetes cases occur before 45 years old and prevalent cases will be excluded, cases will be classified as incident type 2 diabetes cases. The hazard ratios (HRs) will be calculated from proportional hazards models. For incident type 2 diabetes patients, time-to-event will be measured in years from the entry date into the ARIC cohort to the date of diagnosis. Plots of the log (-log) survival curves and the Cox test will be utilized to assess violations of proportional hazard assumptions.

For AIM 2. We will use analysis of variance to evaluate the hypothetical association of SNP/haplotypes of the CHEK2 gene with each continuous prediabetes phenotypes, including fasting glucose, fasting insulin, HbA1c and HOMA-IR in a cross-sectional analysis. We will use logistic regression models as described in Aim 1 to evaluate the effect of CHEK2 variants on IFG and IGT.

For AIM 3. We will use proportional hazards models and logistic regression models as described in aims 1 and 2 to evaluate the effect of CHEK2 variants on incident and prevalent diabetic retinopathy, chronic kidney disease (defined as MDRD estimated glomerular filtration rate less than 60 ml/min/m² and/or microalbuminuria) and and impaired (high frequency) cardiac autonomic tone.

For AIM 4. The effect measure modification of the association between SNP/haplotypes of the CHEK2 gene and type 2 diabetes will be assessed in logistic regression models stratified by BMI, gender, and physical activity. The interaction term by BMI, gender, or physical activity will be fit into the models and the likelihood ratio test with alpha = 0.15 will be utilized to assess effect measure modifications. BMI, gender, and physical activity will be evaluated as potential categorical covariates (using the same categories for stratification) if the stratified analysis indicates they are not significant effect measure modifiers.

Additive interaction will be also be assessed using the interaction contrast ratios (ICR) (Rothman and Greenland 1998). For proportional hazards analysis, ICR = HRR_AB – HRR_A – HRR_B + 1, where HRR_AB represents the joint effect of exposure of interest and the SNP/haplotype and HRR_A and HRR_B represent the main effects of exposure and the SNP/haplotype, respectively. Departures from zero suggest additive interaction. The OR and variance covariance matrix is used to calculate values for ICR and the confidence intervals (Hosmer and Lemeshow 1992; Assmann, Hosmer et al. 1996). For all tests, a two sided hypothesis is adopted and a P value < 0.15 will be regarded as being statistically significant.

The effect measure modification of the association of SNP/haplotypes of the CHEK2 gene with each continuous phenotype will be assessed using a product interaction term.
Haplotype analysis

An important aspect of this project will be the ability to test multiple genotypes in predicting CVD and associated traits. One possible approach is to consider haplotypes. It is important to account for phase uncertainty in the analysis of haplotype-disease associations. Recently, maximum likelihood methods have been developed. These methods are based on the maximization of the observed-data likelihood and produce valid and efficient estimation of haplotype effects and haplotype-environment interactions. Epstein MP and Satten (Epstein and Satten 2003) developed the maximum likelihood method for case-control studies without environmental factors. Dr. Danyu Lin extended this work to accommodate environmental factors (Lin, Zeng et al. 2005) and he provides flexible and user-friendly software, HAPSTAT, on his website: http://www.bios.unc.edu/~lin.

Limitations:

One potential limitation of this study is the lower coverage of the Chek2 gene in African American since our SNP selection was based on HapMap CEPH population. We plan to use haplotype analysis to estimate the association between Chek2 and diabetes in the African American population.

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

8.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”? __X__ 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.cscc.unc.edu/ARIC/search.php

__X____ Yes _______ No

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)?
MP# 1141: TCF7L2 and diabetes mellitus

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

11.b. If yes, is the proposal

__X__  A. primarily the result of an ancillary study (list number* 2006.12_)
___    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/

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.