**ARIC Manuscript Proposal # 2176 r**

| | | |
|---|---|---|
| **PC Reviewed: 11/08/16** | **Status: _____** | **Priority: 2** |
| **SC Reviewed: _____** | **Status: _____** | **Priority: ____** |

**1.a. Full Title**: **Addendum to MP #2176:** Identification and Characterization of Genetic Risk Associated with Measures of Thyroid Function and Disease

**2. Writing Group**:
Writing group members: Yong Li, Elizabeth Selvin, Mandy Li, Adrienne Tin, Nathan Bihlmeyer, Dan Arking, Anna Kottgen; other ARIC authors are invited. Different members of the writing group may participate in the different analyses.

**ARIC author**: **Adrienne Tin**
Address: Dept. of Epidemiology, JHSPH
615 N Wolfe Street, W6017
Baltimore, MD 21215
E-mail: atin1@jhu.edu

**3. Timeline**: Data analysis to start immediately, publication will likely occur in the setting of large consortia

**4. Rationale:**

This is an addendum to the ARIC MP #2176 (Identification and Characterization of Genetic Risk Associated with Measures of Thyroid Function and Disease).

Recently, genome-wide DNA methylation (DNAm) changes were studied in thyroid cancer.[1] However, the association between levels of DNAm and thyroid hormones in healthy population has not been explored. Epigenome-wide scans could increase our understanding on the regulatory mechanisms in thyroid hormonal metabolism. We propose to extend the original proposal to include the analysis of DNAm.

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

Some differentially methylated sites will be associated with thyroid hormone levels.

**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).**

1. Inclusion criteria: Participants with HM450K data passing quality control (n=2802),[2] with data in measures of thyroid function (FT3, FT4, TSH) and values in covariates.

Outcomes:
Quantile normalized beta value of DNA methylation at 473,788 CpG sites passing quality control[2]

Predictor: log transformed measures of thyroid function (TSH, FT3, FT4)

Other variable of interest at visit 2: age, gender, current smoking, center.

Data analysis:

The quality filtering criteria of the methylation data will be the same as those reported in Demerath et al.[2] To control for batch effect, we will use chip number as random effect and plate and chip row as fixed effect in a linear mixed effect model with quantile normalized DNAm beta value as outcome. Other covariates will include: age, age squared, sex, current smoking status.

**Model:**

DNAm_qqnormalized ~ ln(thyroid hormone) + age + age^2 + sex + as.factor(current_smoking) + plate + chip row + imputed neutrophils, lymphocytes, monocytes and eosinophils cell count + random_effect (chip number)

**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 ICTDER03 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 ICTDER03 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 ICTDER03 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; this is an extension of approved proposals _____ 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)?**

**#1343:** "Stage II of a Genome-Wide Association Study for Genetic Variants Associated with Uric Acid Levels and Gout" (approved addendum to conduct consortium-wide meta-analyses, 1343A)

#1379, "Genome-wide Association Study of Single Nucleotide Polymorphisms with Kidney Disease and Kidney Disease Related Traits"

**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*** 2009.24, 2006.03, 2007.02**)**

**___ 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/

**References**


1. Mancikova, V. *et al.* DNA methylation profiling of well-differentiated thyroid cancer uncovers markers of recurrence free survival. *Int J Cancer* **135**, 598-610 (2014).
2. Demerath, E.W. *et al.* Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet* (2015).

**ARIC Manuscript Proposal #2176**

| | | |
|---|---|---|
| **PC Reviewed: 7/9/13** | **Status: A** | **Priority: 2** |
| **SC Reviewed: _____** | **Status: _____** | **Priority: ____** |

**1.a. Full Title**: Identification and Characterization of Genetic Risk Associated with Measures of Thyroid Function and Disease

**b. Abbreviated Title (Length 26 characters)**: Thyroid Marker Genetics

**2. Writing Group**:
Writing group members: Yong Li, Elizabeth Selvin, Mandy Li, Adrienne Tin, Nathan Bihlmeyer, Dan Arking, Linda Kao, Anna Kottgen; other ARIC authors are invited. Different members of the writing group may participate in the different analyses outlined below (such as GWAS vs. exome chip).

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

**First author/analyst (publication will likely occur in the setting of large consortia):**

**Yong Li, MD**
Address: Dept. of Internal Medicine IV
Freiburg University Medical Center
Berliner Allee 29
79110 Freiburg
Germany

Phone: +49 (761) 270-78210 Fax: +49 (761) 270-78040
E-mail: yong.li@uniklinik-freiburg.de

**Corresponding/senior author (**must be an ARIC investigator for the proposal but can be different in the published paper; correspondence will be sent to both the first author & the corresponding author**)**:

**Senior author: Anna Kottgen, MD MPH**
Address: Dept. of Internal Medicine IV
Freiburg University Medical Center
Berliner Allee 29
79110 Freiburg
Germany

Phone: +49 761 270-78050 Fax: +49 761 270-78040
E-mail: anna.koettgen@uniklinik-freiburg.de, akottgen@jhsph.edu

**3. Timeline**:

Genotyping and measurement of thyroid function markers are complete. Data analysis can begin immediately after approval of the manuscript proposal.

**4. Rationale**:

Overt and subclinical thyroid disorders are common, with prevalence estimates around 5-10% for subclinical hypothyroidism in the general US population.[1] Previous genetic studies suggest a genetic component to serum concentration of thyroid hormones, thyroid disease and thyroid cancer. The heritability of thyroid hormone concentrations has been estimated as 40% to 60%,[2-4] and that of thyroid volume as between 60 and 80%.[5] Previous genome-wide association studies have successfully identified genetic variants associated with markers of thyroid function (thyroid stimulating hormone [TSH], free T4 [FT4], thyroid volume and goiter as well as thyroid cancer.[6-9] No genome-wide association studies for triiodothyronine [T3] and anti-thyroid peroxidase antibodies [TPOAb]) have yet been published.

As with other genetic studies of complex traits and disorders, currently identified genetic variants explain only a small proportion of the estimated heritability of thyroid function measures: 6% for TSH and 2% for FT4.[8] The identification of additional novel genes involved in the regulation of thyroid hormone secretion, metabolism, and action will be important for our understanding of thyroid physiology. In addition, identification of susceptibility genes for autoimmune thyroid disease will enhance our insights into this most common etiology of hypothyroidism.

As there has not any previous thyroid genetics research been conducted in the ARIC Study, this will be a multi-layer project, beginning with analyses of common to low frequency variants as available from a 1000 Genomes imputation of the genome-wide Affy 6.0 SNP chip data. Subsequently, we will analyze the exome chip and exome and whole genome data generated from CHARGE-S for association with thyroid function measures in the reference range as well as subclinical and overt thyroid dysfunction. Results from the analyses conducted within the ARIC Study will likely be combined with other cohorts from either the CHARGE Consortium or additional international partners for meta-analysis.

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

The main objectives of the proposed analyses are:

1. to identify common, low frequency and rare genetic variants that are associated with measures of thyroid function (T3, FT4, TSH, TPOAb) among euthyroid individuals (TSH in the normal range)
2. to identify common, low frequency and rare genetic variants that are associated with subclinical and overt hypo- and hyperthyroidism, and self-reported goiter and thyroid disease
3. to characterize significantly associated variants (arising within the ARIC Study alone or as a result of meta-analyses with other studies) across strata of sex, race, BMI, smoking, and thyroid medication intake

4. to relate significantly associated genetic variants to changes of thyroid function markers over time (from V2 to V5) as well as incident subclinical and overt thyroid disease once prospective data becomes available

**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).**

**Study design:**

**Inclusion and Exclusions:** Individuals who did not consent to genetic research will be excluded, as will be individuals with missing thyroid function measures or missing genotypes. We will further exclude individuals of self-reported race other than black or white because of small sample sizes. For analyses to detect genetic loci influencing thyroid function measures among euthyroid individuals, individuals with known thyroid pathologies or taking thyroid medications will be excluded from analyses, as will be individuals with TSH values outside the reference range (<0.4 mIU/L and >4 mIU/L). For analyses of subclinical or overt hypo- and hyperthyroidism, these individuals will be included.

**Outcome variables:**
TSH, T3 and FT4 were measured on a Roche Elecsys 2010 Analyzer using immunoassay methods, and TPOAb were measured on a a Roche e411 Analyzer using an immunoassay method. Quality control of all four markers at the JHSPH (L. Selvin) is completed. T3, FT4, TSH and TPOAb will be analyzed as continuous outcome variables, applying transformations if necessary. Subclinical and overt hypo- and hyperthyroidism will be defined based on commonly used clinical cutoffs and will be harmonized across studies participating in the genetic consortia. Information on self-reported thyroid disease was obtained at visit 3, and on self-reported thyroid disease and goiter at visit 4.

**Exposure variables:** Exposure variables are genetic markers from different sources: SNPs from genotyping of a genome-wide SNP chip (Affy 6.0) followed by genotype imputation using the latest 1000 Genomes all populations reference panel, variants identified from genotyping of a genome-wide exome chip (Illumina exome chip v1.0), and variants identified from sequencing of whole exomes or genomes.

**Covariables:** Age, sex, field center, principle components if necessary. Secondary analyses to characterize significantly associated variants will incorporate information on body mass index and smoking as well as investigate associations among individuals with known thyroid disease, taking thyroid medications, or with TSH values outside the reference range for loci detected among euthyroid inviduals.

**Summary of Statistical Analyses:** Analyses will be carried out separately for European ancestry and African American ancestry individuals. Quality-controlled genotype datasets will be used.

For the association analyses of single common variants (MAF>1%), we will use linear (thyroid function measures) or logistic (hypo-/hyperthyroidism) regression for each outcome on additive coding of SNP genotype as well as relevant covariates (age, sex, study center, principle components if necessary). We will use standard softwares (R, snptestv2) and procedures to ensure high quality of results such as correction for potential inflation of the results. There is plenty of prior experience with these procedures from standard association analyses of genome-wide SNP data.

The analyses of the exome chip data will be conducted in two steps: in a first step, single SNP analyses will be conducted by regressing the outcomes on single genetic variants at a time, common or rare. As in GWAS, additive allele coding will be used and the same covariates will be included in the model. In a second step, additional analyses for rare variants on the exome chip will be conducted. Variants are already provided on the forward strand for the ARIC Study as a result of the joint CHARGE calling effort of the exome chip data. Variants will be recoded such that the CHARGE-wide minor allele is coded. Gene-based aggregate analyses will then be preformed for the four thyroid function measures using the skatCohort function of the skatMeta package in R. Aggregation of variants into genes will be done in accordance with the centrally distributed SNP information file. All variants, including monomorphic ones, will be included in the analysis. This way, SNP selection criteria for burden test and SKAT test, including MAF threshold and SNP function, can later be decided on at the meta-analysis stage. For exploratory analyses within ARIC only, we will test different thresholds such as MAF of 0.01 and 0.05. For processing of the data and formatting and exchange of the results, we will adhere to established CHARGE standards as detailed in the skatMeta_Cookbook.pdf developed by the CHARGE Analysis Committee.

As known sex-specific differences in thyroid function are known to exist, secondary analyses will be conducted stratified by sex for all variants if power calculations indicated sufficient power. We will attempt to follow up variants of high interest through bioinformatic characterization and initiation of functional collaborations. Variants will be prioritized based on their association p-value, their function and predicted effect, their presence in public databases, and their genomic localization within functional elements.

**Limitations/Challenges:** No thyroid ultrasound available: it is therefore not possible to assess thyroid size and volume as well as echotexture. Goiter and thyroid disease are only available from self-report at different visits from the ones when thyroid hormones were measured. The number of these endpoints may be too small to allow for analyses, especially of rare variants.

**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 ICTDER03 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 ICTDER03 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 ICTDER03 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)?**
None yet, the thyroid function measures have just become available and there are no manuscript proposals in this area yet.

**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*)** 2009.24, 2006.03, 2007.02
**___ 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. Agree**

## References


1. Cooper, D.S. & Biondi, B. Subclinical thyroid disease. *Lancet* **379**, 1142-54 (2012).
2. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet* **2**, e132 (2006).
3. Panicker, V. *et al.* Genetic loci linked to pituitary-thyroid axis set points: a genome-wide scan of a large twin cohort. *J Clin Endocrinol Metab* **93**, 3519-23 (2008).
4. Alul, F.Y. *et al.* The heritability of metabolic profiles in newborn twins. *Heredity (Edinb)* **110**, 253-8 (2013).
5. Hansen, P.S. *et al.* Genetic and environmental causes of individual differences in thyroid size: a study of healthy Danish twins. *J Clin Endocrinol Metab* **89**, 2071-7 (2004).
6. Arnaud-Lopez, L. *et al.* Phosphodiesterase 8B gene variants are associated with serum TSH levels and thyroid function. *Am J Hum Genet* **82**, 1270-80 (2008).
7. Teumer, A. *et al.* Genome-wide association study identifies four genetic loci associated with thyroid volume and goiter risk. *Am J Hum Genet* **88**, 664-73 (2011).
8. Porcu, E. *et al.* A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. *PLoS Genet* **9**, e1003266 (2013).
9. Gudmundsson, J. *et al.* Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat Genet* **44**, 319-22 (2012).