ARIC Manuscript Proposal #2355

1.a. Full Title: Genome-wide association studies of susceptibility to childhood acute lymphoblastic leukemia (ALL) using exome chip genotypes

  b. Abbreviated Title (Length 26 characters): Exome chip and ALL

2.   Writing Group:
     Writing group members: ARIC writing group members include the individuals involved in generation and analysis of the exome chip data (alphabetical): Eric Boerwinkle, Megan Grove, and Alanna Morrison. Additional collaborators include Jun Yang (St. Jude) and Philip Lupo (Baylor College of Medicine).

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

     Senior **author**: Jun J. Yang, PhD
     Address: St. Jude Children's Research Hospital
               262 Danny Thomas Pl., MS313, Memphis, TN 38105

               Phone: 901-595-2517   Fax: 901-595-8869
               E-mail: jun.yang@stjude.org

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).
     Name: Alanna Morrison, PhD
               University of Texas Health Science Center at Houston
               School of Public Health
               1200 Pressler St.; Suite 453E; Houston TX 77030

               Phone: 713-500-9913   Fax: 713-500-0900
               E-mail: Alanna.C.Morrison@uth.tmc.edu

3. Timeline:
     Completion of analyses and draft manuscript(s) by the end of 2014. Dr. Yang will complete an ARIC DMDA, found on the ARIC study website:
http://www2.cscc.unc.edu/aric/ancillary-studies-pfg

4. Rationale:

Leukemia is the most frequent malignancy of childhood, accounting for one out of three cases of childhood cancer. In the United States, approximately 4,900 children develop leukemia per year [1]. Acute lymphoblastic leukemia (ALL) is the most common subtype of childhood leukemia, which accounts for 80% of all cases of leukemia [2]. Although ten-year survival is greater than 80%, there is concern over the long-term morbidities related to treatment. In fact, ALL survivors are at a greater risk of developing cardiovascular disease compared to their unaffected contemporaries [3,4]. Additionally, other treatment-related effects include obesity [5,6], metabolic syndrome [7], and diabetes [4,7]. Therefore, identifying risk factors for the development of ALL is important for understanding the biology of disease risk and therapy-related complications in this population (e.g., cardiovascular disease).

Until recently, contribution of inherited genetic variation to ALL susceptibility was poorly defined. Taking a genome-wide approach, Dr. Yang's group and others independently reported that genetic polymorphisms in *IKZF1*, *ARID5B*, *CEBPE*, *CDKN2A*, *PIP4K2A*, and *GATA3* are associated with risk of developing ALL in children [8-12]. However, almost all ALL susceptibility variants identified so far are located in non-coding regions of the genome and their functions are not clearly understood.

Variants located in coding regions have been linked to pathogenesis of a variety of diseases. For example, the majority of Mendelian diseases are driven by rare missense/nonsense variants. Common germline coding variants impacting protein function have been reported to underlie genome-wide association (GWA) association signals [13]. With the availability of commercial single nucleotide polymorphism (SNP) chips specifically designed for coding regions of the genome, exome-based GWAS has recently been applied to a number of diseases/traits with exciting novel discoveries [14].

We propose to utilize existing exome chip variation from ALL cases from St. Jude Children's Research Hospital and controls from the ARIC study to conduct a comprehensive GWAS to identify germline coding genetic variants associated with susceptibility to ALL in children.

5. Main Hypothesis/Study Questions:

The analytical plan outlined here encompasses exome chip variants genotyped for ARIC Whites and Blacks through the CHARGE consortium [15] and genotyped for ALL cases at St. Jude Children's Research Hospital. Analysis of exome chip will allow for: (1) identification of novel genes with common/rare variants that contribute to ALL susceptibility, and (2) identification of rare variation in known candidate genes (e.g., those identified by GWAS) influencing ALL susceptibility.

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).

Exome chip genotypes
Exome chip genotyping for the ARIC cohort has been described previously [7]. Exome chip genotyping for the ALL cases were performed at St. Jude using the exact same Illumina HumanExome array.

Analysis of ALL susceptibility will follow two main approaches:
1) Single variant association analyses
Single marker analyses will focus on variants that are relatively common (i.e., >1% minor allele frequency). We will apply principal component analysis to the combined set of ALL cases and ARIC subjects, to examine population structures and/or potential genotyping batch effects (or other technical factors). Genetic ancestry will be quantified by using STRUCTURE (European, African, East Asian, and Native American), as we described previously [4]. We will perform four GWAS: 1) for all subjects regardless of ancestry, 2) those of European descent (>90% European genetic ancestry), 3) those of African descent (>70% African genetic ancestry), 4) those of Hispanic ethnicity (>10% Native American genetic ancestry). The association of SNP genotype with ALL status will be evaluated by comparing genotype frequency between ALL cases and ARIC (as controls) in regression models, after adjusting for population structure. PLINK or R will be used to run the GWAS analyses.

2) Gene-based analyses
For variants of low frequency (i.e., <1% minor allele frequency), we will evaluate the rare variants in aggregate within a gene using the SKAT test. Rare variants may be further subset to only include those of possible functional consequence (e.g., nonsynonymous, splicing, stopgain, or stoploss). The SKAT test is implemented in R (package SKAT 2.13.0). We will specifically restrict this analysis to subjects of European descent (>90% European genetic ancestry) or those of African descent (>70% African genetic ancestry), due to population stratification.

Meta-analysis
Meta-analysis will be conducted across ethnicities for single variant tests and also for gene-based tests, using the weighted Z-score based METAL test.

Phenotypes:
ALL
ALL cases are from St. Jude Children's Research Hospital and the Children's Oncology Group. Approximately 5,000 subjects were included based on sample availability and genotyping quality. There is no notable sampling bias [3 and 4].

Non-ALL controls
Because the prevalence of adult survivors of childhood ALL (as well as the overall incidence of childhood ALL) is excessively low in the general population (1 in 10,000 in the US [1]), the ARIC cohort can serve as a comparison population in our study. This approach (e.g., using genetic data from an adult cohort) has been used successfully in other childhood cancer genome-wide association studies. For example, the Multi-Ethnic

Study of Atherosclerosis (MESA) and Genetic Association Information Network (GAIN) cohorts have successfully used non-ALL controls in other genome-wide studies [1, 3, 4].

Covariates: genetic ancestry and/or any observed genotyping batch effects

7.a. Will the data be used for non-CVD analysis in this manuscript?
_X_ Yes   __ 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?
_X_ Yes   ____ No
    (This file ICTDER 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)?

There are no manuscript proposals for the evaluation of exome chip data in relation to ALL susceptibility.

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

___   **A. primarily the result of an ancillary study (list number* _____)**
_X_   **B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* 2009.12)**

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

12a. 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

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PUBMED Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to Pubmed central.

Agree

13. References

1. Margolin JF (2011) Molecular diagnosis and risk-adjusted therapy in pediatric hematologic malignancies: a primer for pediatricians. Eur J Pediatr 170: 419-425.
2. Siegel R, DeSantis C, Virgo K, Stein K, Mariotto A, et al. (2012) Cancer treatment and survivorship statistics, 2012. CA Cancer J Clin 62: 220-241.
3. Oeffinger KC, Buchanan GR, Eshelman DA, Denke MA, Andrews TC, et al. (2001) Cardiovascular risk factors in young adult survivors of childhood acute lymphoblastic leukemia. J Pediatr Hematol Oncol 23: 424-430.
4. Oeffinger KC, Mertens AC, Sklar CA, Kawashima T, Hudson MM, et al. (2006) Chronic health conditions in adult survivors of childhood cancer. N Engl J Med 355: 1572-1582.
5. Tonorezos ES, Vega GL, Sklar CA, Chou JF, Moskowitz CS, et al. (2012) Adipokines, body fatness, and insulin resistance among survivors of childhood leukemia. Pediatr Blood Cancer 58: 31-36.
6. Oeffinger KC, Mertens AC, Sklar CA, Yasui Y, Fears T, et al. (2003) Obesity in adult survivors of childhood acute lymphoblastic leukemia: a report from the Childhood Cancer Survivor Study. J Clin Oncol 21: 1359-1365.
7. Gurney JG, Ness KK, Sibley SD, O'Leary M, Dengel DR, et al. (2006) Metabolic syndrome and growth hormone deficiency in adult survivors of childhood acute lymphoblastic leukemia. Cancer 107: 1303-1312.
8. Trevino LR, Yang W, French D, Hunger SP, Carroll WL, et al. (2009) Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nat Genet 41: 1001-1005.
9. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, et al. (2009) Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. Nat Genet 41: 1006-1010.

10. Sherborne AL, Hosking FJ, Prasad RB, Kumar R, Koehler R, et al. (2010) Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. Nat Genet 42: 492-494.

11. Xu H, Yang W, Perez-Andreu V, Devidas M, Fan Y, et al. (2013) Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. J Natl Cancer Inst 105: 733-742.

12. Perez-Andreu V, Roberts KG, Harvey RC, Yang W, Cheng C, et al. (2013) Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. Nat Genet 45: 1494-1498.

13. Fellay J, Thompson AJ, Ge D, Gumbs CE, Urban TJ, et al. (2010) ITPA gene variants protect against anaemia in patients treated for chronic hepatitis C. Nature 464: 405-408.

14. Huyghe JR, Jackson AU, Fogarty MP, Buchkovich ML, Stancakova A, et al. (2013) Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat Genet 45: 197-201.

15. Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, et al. (2013) Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. PLoS One 8: e68095.