ARIC Manuscript Proposal #3544 (Amended)

PC Reviewed: 1/12/21 **Status: _____** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1.a. Full Title:

Identifying genes with pleiotropic effects in rhabdomyosarcoma and developmental disorders

b. Abbreviated Title (Length 26 characters):

RMS pleiotropic genes

2. Writing Group:

Writing group members: Drs. Melissa Richard, He Li, Aniko Sabo, Philip Lupo, Sharon Plon, Eric Boerwinkle

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

First author: Melissa Richard
Address: One Baylor Plaza
Houston, TX. 77030
Phone: 713-798-2971 Fax:
E-mail: melissa.richard@bcm.edu

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: Eric Boerwinkle
Address: UTHealth School of Public Health
1200 Pressler St., Suite W114A
Houston, Texas 77030
Phone: 713-500-9041 Fax:
E-mail: eboerwin@bcm.edu

3. Timeline:

We expect to submit a manuscript in early 2021.

4. Rationale:

Rhabdomyosarcoma (RMS) is the most common soft-tissue sarcoma in children, yet has one of the poorest 5-year survival rates among all childhood cancers. For those with metastatic disease, the 5-year survival rate is <30%, even with aggressive therapy. Sarcomas are understudied cancers for which there are few established risk factors, including an incompletely characterized genetic basis. While about 8% of RMS cases are caused by a known genetic syndrome, >90% are considered to be sporadic. Therefore, much work remains to characterize epidemiologic and germline genetic risk factors in RMS. There are ongoing questions about the connection between childhood cancers and developmental disorders, however the research to date has been limited. For example, small sample sizes have prevented specific examination of a connection between RMS and other developmental disorders in previous assessments, despite multiple lines of evidence that suggest a shared etiology. First, there is evidence that germline *de novo* variants are associated with common developmental diseases, including autism spectrum disorders and congenital heart defects, as well as RMS. Second, developmental disorders and RMS share some risk factors, including advanced paternal age and *in vitro* fertilization. Third, pathogenic variants in Ras-MAPK signaling confer risk for developmental disorders collectively known as RASopathies. RAS genes and RASopathies are associated with RMS, autism, and congenital heart defects. Our preliminary analysis of germline sequencing identified 15/347 individuals (4.3%) with eRMS who carry pathogenic variants in RAS-related genes also associated with autism and congenital heart defects. Collectively, these data suggest a better understanding of the connection between RMS and developmental disorders, such as autism and congenital heart defects, could provide substantial insight in the etiology of RMS.

5. Main Hypothesis/Study Questions:

The objectives of this proposal are to advance understanding of the genetic underpinnings of pediatric RMS, and to ultimately improve prevention, treatment, and surveillance efforts. Our central hypotheses are: 1) autism and congenital heart defects occur more frequently among children with RMS than the general population, and 2) shared biological mechanisms underlie these associations.

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).

To better understand the disease risk associated with germline mutations in cancer predisposition genes, we obtained a set of 1,200 well-annotated RMS samples leveraging the Children's Oncology Group Soft Tissue Sarcoma Biology and Banking Protocol during the past years. We additionally have obtained a set of >2,000 cases of autism spectrum disorder in the Simons Foundation Powering Autism Research (SPARK) program and >1,000 cases with congenital heart defects. All samples have written informed consent for this proposed research.

We proposed to use the ARIC cohort as population-based controls in the proposed analyses because: 1) Majority of the RMS samples are of European or African descent, which matches the self-reported ethnicity composition of the ARIC cohort; 2) the RMS, autism, and congenital heart defect samples have been sequenced at the Human Genome Sequence Center (HGSC) at Baylor College of Medicine (BCM) where the ARIC samples were sequenced, using the same

analytical pipelines to perform alignment, variant calling, and quality controls. This strategy will minimize the inter-laboratory batch effect for performing case-control analysis. We will only include ARIC samples that provide broad informed consent.

The analysis plan will first match each of the cases (RMS, autism, congenital heart defect) with 4-9 ARIC controls based on sex and self-reported ethnicity. We will remove variants with discrepant quality metrics and genotyping rates between the cases and ARIC datasets to further minimize batch effect. An additive genetic model will be adopted to test the associations between each genetic variant and the susceptibility to RMS, autism, or congenital heart defects via case-control analysis using logistic regression, adjusting for sex and the first 10 principle components of population structure. We will then test for pairwise pleiotropic effects between the conditions using LD regression, or similar. The results from this analysis will expand our knowledge toward a better understanding of genetic architectures underlying RMS susceptibly.

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/mantrack/maintain/search/dtSearch.html

___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)?

We propose to use ARIC cohort as a population control to study pleiotropic effect of genes in rhabdomyosarcoma and developmental disorders and are not aware of any manuscript proposals with similar topics.

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ____ Yes _X___ 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 https://www2.cscc.unc.edu/aric/approved-ancillary-studies

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.