Usher Institute
Population Health Sciences & Informatics

Nine Edinburgh BioQuarter
9 Little France Road
Edinburgh EH16 4UX
Usher.Bioquarter@ed.ac.uk

Old Medical School
Teviot Place
Edinburgh EH8 9AG
Usher.Teviot@ed.ac.uk

THE UNIVERSITY of EDINBURGH

August 8, 2016

ARIC Publication Committees
THE JOHNS HOPKINS UNIVERSITY
Department of Epidemiology
615 North Wolfe Street; Room W6009
Baltimore, Maryland 21205
410-955-4380
410-955-8086 fax

Dear Dr. Coresh,

Thank you very much for your mail. This letter contains specific responses to the questions raised concerning our manuscript proposal #2793.

Question. Please clarify the key definitions in section 6 – for example, the data used to define male pattern baldness and how age at assessment will be incorporated into the analysis (at visit 4 ages are 54-74 which may be sufficiently old to express all early male pattern baldness but older individuals may have later onset thinning hair). The analysis itself is unclear.

Response. We have included a revised description of the analysis protocol. In particular, we will use a multivariate logistic regression where the phenotype will be defined as cases for individuals with baldness greater than IV grade on the Northwood-Hamilton scale and controls designated as individuals with grade I (no sign of baldness). Age at assessment will be used as a covariate in the analysis together with 10 principal components and any cohort specific covariates.

Thank you for the opportunity to clarify our proposal.

Sincerely yours,

Nicola Pirastu

ARIC Manuscript Proposal #2793r

PC Reviewed: 09/13/16	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title:
42 novel associations for male pattern baldness provide new insights into aetiology and genetic correlations

b. Abbreviated Title (Length 26 characters):
42 new loci for Male Baldness

2. Writing Group:
Alanna C. Morrison, Paul S. de Vries, Kari North, Nora Franceschini

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

First author: Nicola Pirastu
Address: Usher Institute of Population Health Sciences and Informatics
Nine Edinburgh BioQuarter
9 Little France Road
Edinburgh
EH16 4UX

Phone: +44 131 651 7112
E-mail: nicola.pirastu@ed.ac.uk

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
Address: Human Genetics Center
1200 Pressler Street, Suite W114
Houston, TX 77030

Phone: 713-500-9058 Fax: 713-500-9020
E-mail: alanna.c.morrison@uth.tmc.edu

3. Timeline:
Data are available, and analyses are to start as soon as possible. Manuscript is in progress and ARIC data is contributing to a replication effort.

4. Rationale:
Male pattern baldness or androgenetic alopecia is the most common cause of hair loss in men, with a prevalence ranging between 50% and 100%, depending on age. Its importance is not limited to physical appearance and related distress, baldness has also been associated with increased risk of cardiovascular disease and type 2 diabetes. Understanding the physiology of male pattern baldness may potentially help not only in developing new treatments but also could potentially give new insight in the etiopathology of other associated disorders. Despite its high heritability (~80% estimated in twins), and its apparent dominant inheritance, up to now a limited number of associated loci have been identified suggesting a very high polygeny. The main limitation up to now has been the relatively low sample size in previous studies, the largest of which was comprised of less than 20,000 samples in total. We thus aim at bridging this gap by using ~45.000 samples for the discovery phase and ~30.000 more for replication. Furthermore we will employ extensive bioinformatics analysis to help understand which pathways are involved in the etiopathology of this common disorder.

5. Main Hypothesis/Study Questions:
Identify genetic variation associated with male pattern baldness to better understand its physiology.

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).

ARIC data will be included as a replication cohort.

The analysis is a simple logistic regression using age and at least 10 PC to correct for population stratification. Add to the model al the covariates used in the specific cohort (ie. array batch etc..)

y~ age+PC1+PC2…PC10+cohort specific covariates.

Define as cases people with at least IV grade on the Norwood-Hamilton scale.

Controls should be defined as people with grade I on the Norwood-Hamilton scale at the time of assessment.

Analysis should be run on the most recent available imputation using dosages

Genotype data includes ~12.000 SNPs from 1000 Genomes imputation.

The analysis involves a logistic regression model, including age at visit 4 and at least 10 PCs to correct for population stratification.

Results reported for replication include
SNP: SNP name
CHR: Chromosome
BP: Position

Beta: regression coefficient on the logistic scale (log(OR))
seBeta: standard error for the regression coefficient
A1: coded allele
A0: non coded allele
EAF: coded allele frequency
P: p-value

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)?

ARIC Manuscript Proposal # 1035
Title: Relation of Male Pattern Baldness to Myocardial Infarction, Prevalent CHD, and Stroke

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 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. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __X__ No.

ARIC Manuscript Proposal #2793r

PC Reviewed: 09/13/16	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title:
42 novel associations for male pattern baldness provide new insights into aetiology and genetic correlations

b. Abbreviated Title (Length 26 characters):
42 new loci for Male Baldness

2. Writing Group:
Alanna C. Morrison, Paul S. de Vries, Kari North, Nora Franceschini

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

First author: Nicola Pirastu
Address: Usher Institute of Population Health Sciences and Informatics
Nine Edinburgh BioQuarter
9 Little France Road
Edinburgh
EH16 4UX

Phone: +44 131 651 7112
E-mail: nicola.pirastu@ed.ac.uk

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
Address: Human Genetics Center
1200 Pressler Street, Suite W114
Houston, TX 77030

Phone: 713-500-9058 Fax: 713-500-9020
E-mail: alanna.c.morrison@uth.tmc.edu

3. Timeline:
Data are available, and analyses are to start as soon as possible. Manuscript is in progress and ARIC data is contributing to a replication effort.

4. Rationale:
Male pattern baldness or androgenetic alopecia is the most common cause of hair loss in men, with a prevalence ranging between 50% and 100%, depending on age. Its importance is not limited to physical appearance and related distress, baldness has also been associated with increased risk of cardiovascular disease and type 2 diabetes. Understanding the physiology of male pattern baldness may potentially help not only in developing new treatments but also could potentially give new insight in the etiopathology of other associated disorders. Despite its high heritability (~80% estimated in twins), and its apparent dominant inheritance, up to now a limited number of associated loci have been identified suggesting a very high polygeny. The main limitation up to now has been the relatively low sample size in previous studies, the largest of which was comprised of less than 20,000 samples in total. We thus aim at bridging this gap by using ~45.000 samples for the discovery phase and ~30.000 more for replication. Furthermore we will employ extensive bioinformatics analysis to help understand which pathways are involved in the etiopathology of this common disorder.

5. Main Hypothesis/Study Questions:
Identify genetic variation associated with male pattern baldness to better understand its physiology.

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).

ARIC data will be included as a replication cohort.

The analysis is a simple logistic regression using age and at least 10 PC to correct for population stratification. Add to the model al the covariates used in the specific cohort (ie. array batch etc..)

y~ age+PC1+PC2…PC10+cohort specific covariates.

Define as cases people with at least IV grade on the Norwood-Hamilton scale.

Controls should be defined as people with grade I on the Norwood-Hamilton scale at the time of assessment.

Analysis should be run on the most recent available imputation using dosages
~~Phenotype data to be used includes information on pattern baldness. Cases are defined as individuals with at least IV grade on the Norwood-Hamilton scale. Controls are defined as individuals with grade I on the Norwood-Hamilton scale at the time of assessment.~~

Genotype data includes ~12.000 SNPs from 1000 Genomes imputation.

The analysis involves a logistic regression model, including age at visit 4 and at least 10 PCs to correct for population stratification.

Results reported for replication include

SNP: SNP name
CHR: Chromosome
BP: Position
Beta: regression coefficient on the logistic scale (log(OR))
seBeta: standard error for the regression coefficient
A1: coded allele
A0: non coded allele
EAF: coded allele frequency
P: p-value

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)?

ARIC Manuscript Proposal # 1035
Title: Relation of Male Pattern Baldness to Myocardial Infarction, Prevalent CHD, and Stroke

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 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. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __X__ No.