Instructions

Research Letters are concise, focused reports of original research. These should not exceed **600 words of text** and **6 references** and may include up to **2 tables or figures**. Online supplementary material is not allowed. Research Letters may have no more than 7 authors. The text should include the **full name, academic degrees, and a single institutional affiliation for each author and the email address for the corresponding author.** Other persons who have contributed to the study may be indicated in an Acknowledgment, with their permission, including their academic degrees, affiliation, contribution to the study, and an indication if compensation was received for their role. Letters must not duplicate other material published or submitted for publication. In general, Research Letters should be divided into the following sections: Introduction, Methods, Results, and Discussion. They should not include an abstract, but otherwise should follow all of the guidelines in [Manuscript Preparation and Submission Requirements](http://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecManuscriptPreparationandSubmissionRequirements). Letters not meeting these specifications are generally not considered

Manuscripts should be prepared in accordance with the [AMA Manual of Style](http://www.amamanualofstyle.com/), 10th edition,[2](http://jamanetwork.com/journals/jama/pages/instructions-for-authors#Ref2) and/or the [ICMJE Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals](http://icmje.org/recommendations/).[4](http://jamanetwork.com/journals/jama/pages/instructions-for-authors#Ref4)

Include in the manuscript file a title page, abstract, text, references, and as appropriate, figure legends and tables. Start each of these sections on a new page, numbered consecutively, beginning with the title page. Figures should be submitted as separate files (1 file per figure) and not included in the manuscript text.

**Cover Letter**

*Include a cover letter and complete contact information for the corresponding author (affiliation, postal/mail address, email address, and telephone number) and whether the authors have published or submitted any related papers from the same study (see* [*Duplicate/Previous Publication or Submission*](http://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecDuplicate/PreviousPublicationorSubmission)*).*

Dear Dr. Howard Bauchner,

This submission joins the recent excitement over the release of new control sequence datasets. It aims to bring a population genetics approach to the estimation of penetrance, a persistent problem with secondary findings. We specifically chose the updated set of recommended reporting genes from ACMG-59, in hopes that the sensitivity analysis can advise clinicians on their understanding of genetic test results.

We do not claim to have pinned down precise penetrance values for the ACMG-59, but argue that the population model of penetrance serves is useful for a more fine-grained understanding of outside effects on the penetrance estimates derived from classical family studies. We have attached a web application tool which can take prevalence and allelic heterogeneity inputs specific to a particular region or demographic and project new penetrance estimates, which may demonstrate differences.

All of the data, code, and figures are provided on github and are fully reproducible from scratch and from pre-loaded data.

We hope that this can introduce a quick, reproducible, and semi-quantitative framework for evaluating the sensitivity of penetrance to ancestry and other effects.

Corresponding author: Arjun K. Manrai, PhD.

Department of Biomedical Informatics, Harvard Medical School, Boston, MA

10 Shattuck St., Boston, MA, 02115

Arjun\_Manrai@hms.harvard.edu

302-593-3453 (phone)

The authors have not published or submitted any related papers from the same study.

Title Page

Penetrance Estimates for Secondary Genomic Findings using Allele Frequencies from Diverse Populations (97/100)

James A. Diao

I’m not sure what my affiliation would be- my home institution, or the institution at which I did the work?

Isaac S. Kohane, MD, PhD

Department of Biomedical Informatics, Harvard Medical School, Boston, MA

Arjun K. Manrai, PhD

Department of Biomedical Informatics, Harvard Medical School, Boston, MA

10 Shattuck St., Boston, MA, 02115

Arjun\_Manrai@hms.harvard.edu

302-593-3453 (phone)

Manuscript word count: \_\_\_

Text

Introduction

The dropping costs and rising popularity of next-generation sequencing has introduced the possibility of personalizing medical treatments and screening for genetic diseases. Still, there is limited consensus on the proper interpretation for many genetic variants. Complications include insufficient diversity in sequencing data used to justify pathogenicity, as well as reduced penetrance in findings unrelated to the diagnostic indication (secondary findings). The American College on Medical Genetics and Genomics (ACMG) now recommends that clinical sequencing laboratories seek and report secondary findings in a minimum list of 59 genes (ACMG-59). This study explores the effect of applying these recommendations to healthy cohorts from reference sequencing data, and the sensitivity of population-based penetrance estimates to ancestral effects.

Methods

Quantitative risk estimates were derived by modeling the penetrance of variants at the disease-level. Each penetrance value is expressed as a function of disease prevalence, allele frequency, and allelic heterogeneity using Bayes’ rule. Plausible estimates for these parameters were determined from ExAC and the medical literature. Confidence intervals for point estimates are imputed using the average range.

Allele frequency data was collected for five continental super-populations (African, American, East Asian, European, and South Asian) from phase 3 of the 1000 Genomes Project, and the Exome Aggregation Consortium (ExAC). 1000 Genomes contains individual-level sequence data for 2,504 individuals, while ExAC contains population-level sequence data for 60,706 individuals (including those in 1000 Genomes). Classifications of variant significance were derived from the November 2016 update of ClinVar, the public central repository for interpretations of variants that have been classified by researchers.

Penetrance values were computed for each superpopulation separately, as well as overall.

Results

Pathogenic variants were found to be distributed unevenly across ancestral groups in this cohort, and incidental findings were found to be inflated relative to empirical disease prevalence.

We find that the average person has \_\_\_ genetic disorders and that x% of people have at least 1.

Under the most generous assumptions, maximum penetrance estimates for the majority of diseases fall under 50%, with many under 5%. There is remarkable variation across ancestral groups, with some overall penetrance estimates derived largely or entirely from just a few populations. For example, catecholaminergic

retinoblastoma variants are only sufficiently rare to be clinical useful in European populations, and Juvenile

Discussion

Because of the vast uncertainty around the parameters, penetrance values themselves should be considered to have very large error bars. However, the disappointing values under optimistic conditions suggest caution with secondary findings.

The allele frequencies are estimated more precisely, with strong differences between ancestral groups suggesting that the actual penetrance of the ACMG-59 also varies by ancestry. This effect is amplified by the lack of knowledge around prevalence and allelic heterogeneity values, which are also recognized to vary greatly between ancestral populations.

The population genetics approach to computing penetrance has only been possible in recent years, with the release of reference datasets such as ExAC. Other work has already taken advantage of ExAC to derive maximum credible allele frequencies to reclassify pathogenic variants as benign. Although the population model of penetrance fails to define tight penetrance bounds, it provides a quick, reproducible, and semi-quantitative framework for evaluating the sensitivity of penetrance to ancestry and other effects.

for medically-actionable genes

References

1. Manrai AK, et al. Genetic misdiagnoses and the potential for health disparities. *N Engl J Med*. 2016;375(7):655-665
2. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538(7624): 161-164.
3. Lek M, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291.
4. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42:980-985.
5. sdfsdf

Acknowledgements

Acquisition and analysis of data; drafting of manuscript: James Diao

Concept and design; interpretation of data; critical revisions of manuscript: James Diao and Arjun Manrai.

Administrative support: Isaac Kohane.

James Diao and Arjun Manrai had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Figure legends and tables

All symbols, indicators, line styles, and colors in statistical graphs should be defined in a key or in the figure legend. Axes in statistical graphs must have labels. Units of measure must be provided for continuous data

#### Figures

Number all figures (graphs, charts, photographs, and illustrations) in the order of their citation in the text. The number of figures should be limited. Avoid complex composite or multipart figures unless justified. See [Categories of Articles](http://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecCategoriesofArticles) for limits on the number of figures and/or tables according to article type.

For initial manuscript submissions, figures must be of sufficient quality and may be embedded at the end of the file for editorial assessment and peer review. If a revision is requested and before a manuscript is accepted, authors will be asked to provide figures that meet the requirements described in [Figure File Requirements for Publication](http://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecFigureFileRequirementsforPublication).

Graphs, charts, some illustrations, titles, legends, keys, and other elements related to figures in accepted manuscripts will be re-created and edited according to JAMA Network style and standards prior to publication. Online-only figures will not be edited or re-created (see [Online-Only Supplements and Multimedia](http://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecOnline-OnlySupplementsandMultimedia)).

#### Figure Titles and Legends (Captions)

At the end of the manuscript, include a title for each figure. The figure title should be a brief descriptive phrase, preferably no longer than 10 to 15 words. A figure legend (caption) can be used for a brief explanation of the figure or markers if needed and expansion of abbreviations. For photomicrographs, include the type of specimen, original magnification or a scale bar, and stain in the legend. For gross pathology specimens, label any rulers with unit of measure. Digitally enhanced images must be clearly identified in the figure legends as enhanced or manipulated, eg, computed tomographic scans, magnetic resonance images, photographs, photomicrographs, x-ray films.

#### General Figure Guidelines

* Primary outcome data should not be presented in figures alone. Exact values with measure of variability should be reported in the text or table as well as in the abstract.
* All symbols, indicators (including error bars), line styles, colors, and abbreviations should be defined in a legend.
* Each axis on a statistical graph must have a label and units of measure should be labeled.
* Do not use pie charts, 3-D graphs, and stacked bar charts as these are not appropriate for accurate statistical presentation of data and should be revised to another figure type or converted to a table.
* Error bars should be included in both directions, unless only 1-sided variability was calculated.
* Values for ratio data—odds ratios, relative risks, hazard ratios—should be plotted on a log scale. Values for ratio data should not be log transformed.
* For footnotes, use letters (a, b, c, etc) not symbols.

For images featuring patients or other identifiable persons, it is not acceptable to use black bars across the eyes in an attempt to deidentify. Cropping may be acceptable as long as the condition under discussion is clearly visible and necessary anatomic landmarks display. If the person in the image is possibly identifiable (not only by others but also by her/himself), permission for publication is required (see [Patient Identification](http://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecPatientIdentification)).