

Defining and Reducing Variant Classification Disparities

Moez Dawood^{1-3*}, B.A., Shawn Fayer^{4,5}, M.S., Sriram Pendyala^{5,6}, B.A., Mason Post⁴, M.S., Divya Kalra¹, M.S., Karynne Patterson⁵, B.S., Eric Venner^{1,2}, Ph.D., Lara A. Muffley^{4,5}, B.S., Douglas M. Fowler^{4,5,7}, Ph.D., Alan F. Rubin^{8,9}, Ph.D., Jennifer E. Posey², M.D., Ph.D., Sharon E. Plon^{1,2}, M.D., Ph.D., James R. Lupski^{1,2,10,11}, M.D., Ph.D., Richard A. Gibbs^{1,2}, Ph.D., Lea M. Starita^{4,5}, Ph.D., Carla Daniela Robles-Espinoza^{12,13}, Ph.D., Willow Coyote-Maestas^{14,15*}, Ph.D., Irene Gallego Romero^{16-18*}, Ph.D.

1. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA.
2. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA.
3. Medical Scientist Training Program, Baylor College of Medicine, Houston, TX, USA.
4. Brotman Baty Institute for Precision Medicine, Seattle, WA, USA.
5. Department of Genome Sciences, University of Washington, Seattle, WA, USA.
6. Medical Scientist Training Program, University of Washington, Seattle, WA, USA.
7. Department of Bioengineering, University of Washington, Seattle, WA, USA.
8. Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia.
9. Department of Medical Biology, University of Melbourne, Melbourne, VIC, Australia.
10. Texas Children's Hospital, Houston, TX, USA.
11. Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA.
12. Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Campus Juriquilla, Querétaro, Qro, Mexico
13. CASM, Wellcome Sanger Institute, Hinxton, UK
14. Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, USA
15. Quantitative Biosciences Institute, University of California, San Francisco, USA
16. Human Genomics and Evolution, St Vincent's Institute of Medical Research, Fitzroy, 3065, Australia
17. School of BioSciences and Melbourne Integrative Genomics, The University of Melbourne, Royal Parade, Parkville, 3010, Australia
18. Center for Genomics, Evolution and Medicine, Institute of Genomics, University of Tartu, Riia 23b, 51010, Tartu, Estonia

*Corresponding Authors: Moez Dawood (mdawood@bcm.edu), Willow Coyote-Maestas (willow.coyote-maestas@ucsf.edu), Irene Gallego Romero (irene.gallego@svi.edu.au)

Abstract

Background:

Multiplexed Assays of Variant Effects (MAVEs) can test all possible single variants in a gene of interest. The resulting saturation-style data may help resolve variant classification disparities between populations, especially for variants of uncertain significance (VUS).

Methods:

We analyzed clinical significance classifications in 213,663 individuals of European-like genetic ancestry versus 206,975 individuals of non-European-like genetic ancestry from *All of Us* and the Genome Aggregation Database. Then, we incorporated clinically calibrated MAVE data into the Clinical Genome Resource's Variant Curation Expert Panel rules to automate VUS reclassification for *BRCA1*, *TP53*, and *PTEN*.

Results:

Using two orthogonal statistical approaches, we show a higher prevalence ($p \leq 5.95e-06$) of VUS in individuals of non-European-like genetic ancestry across all medical specialties assessed in all three databases. Further, in the non-European-like genetic ancestry group, higher rates of Benign or Likely Benign and variants with no clinical designation ($p \leq 2.5e-05$) were found across many medical specialties, whereas Pathogenic or Likely Pathogenic assignments were higher in individuals of European-like genetic ancestry ($p \leq 2.5e-05$).

Using MAVE data, we reclassified VUS in individuals of non-European-like genetic ancestry at a significantly higher rate in comparison to reclassified VUS from European-like genetic ancestry ($p = 9.1e-03$) effectively compensating for the VUS disparity. Further, essential code analysis showed equitable impact of MAVE evidence codes but inequitable impact of allele frequency ($p = 7.47e-06$) and computational predictor ($p = 6.92e-05$) evidence codes for individuals of non-European-like genetic ancestry.

Conclusions:

Generation of saturation-style MAVE data should be a priority to reduce VUS disparities and produce equitable training data for future computational predictors.

Introduction

Medicine faces challenges of unequal access and representation, particularly for individuals with non-European-like genetic ancestries, which results in a disproportionate number of inconclusive diagnostic outcomes for these populations¹⁻³. This inequity is exacerbated in genomic medicine since the vast majority of research and clinical genomic sequencing to date has prioritized individuals of European-like genetic ancestries resulting in a comparative deficiency in knowledge about disease risk associated with genetic variants for individuals of non-European-like genetic ancestry⁴⁻⁶. This lack of diversity in control population data has repeatedly led to incorrect diagnoses⁷, missed diagnoses⁸⁻¹¹ and inappropriate medical management⁷ for individuals of non-European-like genetic ancestry⁷⁻⁹. Compounding these challenges, genomic medicine can struggle to determine if identified genetic variants are harmful (Pathogenic or Likely Pathogenic; P/LP) or harmless (Benign or Likely Benign; B/LB), resulting in an exponentially increasing number of variants of uncertain significance (VUS) which are most commonly reported in individuals of non-European-like ancestry^{8,10-14}. Thus, medicine urgently requires a systematic, population-scale understanding of variant classification inequities across genetic ancestries and a solution for large-scale reclassification of VUS, especially for individuals of non-European-like genetic ancestry.

Recent advances in functional genomics are enabling systematic, high-throughput experimental testing via Multiplexed Assays of Variant Effect (MAVEs), a family of methods able to characterize every possible SNV (single nucleotide variant) or indel (insertion or deletion) in a target gene and are being used for reclassification of VUS at scale^{13,15-17} (**Fig.1**). Thus, we hypothesized that the saturation nature of MAVEs would produce variant effects for VUS in individuals of non-European-like genetic ancestry leading to a higher rate of VUS reclassification compared to individuals of European-like genetic ancestry by compensating for the original VUS disparity. Further, we posited there would be an inequitable impact of different evidence towards VUS reclassification, but MAVE data would be used equitably. MAVEs mark a pivotal experimental advance in rectifying variant classification disparities and contributing to more equitable health outcomes for diverse populations worldwide.

Methods

Data Curation and Allele Prevalence:

Variants for each gene are available on the *All of Us* Public Data Browser (v7; <https://databrowser.researchallofus.org/variants>) or the Genome Aggregation Database (gnomAD) (<https://gnomad.broadinstitute.org/>). All databases assign a single genetic ancestry to each individual based on projection to principal components built using reference populations^{9,18,19}. We have appended ‘-like’ to these descriptors to better reflect similarity to a particular set of reference populations²⁰. We acknowledge their imperfect and incomplete nature as descriptors of continuous human diversity. The clinical variant classifications in this study found in *All of Us* or gnomAD were originally sourced from ClinVar. Variant calls, allele counts, population descriptors and variant classifications were used as prescribed by *All of Us* or gnomAD. Allele prevalence was calculated by summing allele counts for variants of each clinical classification for examined genetic ancestries and dividing this sum by the number of individuals in the genetic ancestry group(s).

Two Orthogonal Statistical Methods:

Two orthogonal statistical methods were used to assess variant classification disparities. First, at the gene-level using a matched pairs, Wilcoxon signed-rank test with Bonferroni correction resulting in a p-value, estimate of statistical power, and rank biserial coefficient with 95% confidence interval to quantify the magnitude of the differences using pre-established thresholds²¹. The matched pairs were the same gene’s allele prevalence between ancestry groups. Second, unique variants (not allele counts) for each clinical classification were counted across a gene list that were exclusive to each superpopulation group. If alleles for a unique variant were found in both superpopulation groups that unique variant was excluded from the counts. Then a chi-square test for independence with Bonferroni correction was conducted.

Variant Reclassification and Essential Code Analysis:

We developed an automated pipeline to reclassify VUS in *BRCA1*, *TP53*, and *PTEN* found in gnomAD and *All of Us*. These three genes were selected because all three have clinically calibrated MAVE data and Clinical Genome Resource’s (ClinGen) Variant Curation Expert Panel (VCEP) guidelines^{13,22–29}. Our pipeline follows the gene-specific criteria of the corresponding VCEP as closely as possible except for the functional data evidence code (PS3/BS3) where MAVE data was used. To assess evidence code essentiality, we sequentially removed each code from the final set of codes for a reclassified VUS and observed if removal led to reversion of the reclassified variant back to VUS. Further details for any above methods can be found in the Supplementary Methods.

Results

Rationale for Selecting Databases

We analyzed genomes of 245,394 Americans in *All of Us* v7 and orthogonally validated our findings in two independent versions of the Genome Aggregation Database (123,709 exomes of gnomAD v2.1.1 and 51,535 genomes of gnomAD v3.1.2 (non v2)). We formed two superpopulation groupings: European-like and non-European-like genetic ancestry. Individual assignment was based on genetic ancestry labels reported by the respective database (Supplementary Methods)^{9,18,19}. We chose these three population-scale databases, because the number of individuals sequenced in each was similar for both superpopulation groupings (**Fig.S1**) (non-European-like vs. European-like: 122,322 vs. 123,072 *All of Us* v7; 59,106 vs. 64,603 gnomAD v2.1.1; 25,547 vs. 25,988 gnomAD v3.1.2 (non v2)).

Higher VUS Prevalence in Non-European-Like Genetic Ancestry

First, using the gene by gene statistical approach, we investigated allele prevalence differences of each clinical variant classification category between individuals of non-European-like versus European-like genetic ancestry at population scale. Individuals of non-European-like genetic ancestry exhibited significantly higher VUS prevalence across all medical specialties and gene groupings assessed in all three databases (p -values ranging 1.52e-211 to 1.4e-07; effect sizes ranging 0.35 to 0.76; **Fig.2, Fig.S2**, Tables S2-4). In contrast, P/LP classifications were significantly increased in individuals of European-like genetic ancestry (p -values ranging 2.3e-63 to 1.2e-04; effect sizes ranging -0.57 to -0.18; **Fig.S3**, Tables S2-4). Further, a significantly higher prevalence of B/LB and variants with no clinical designation (ND) was found in individuals of non-European-like genetic ancestry across several of the medical specialties (p -values ranging 2.9e-303 to 1.98e-05; effect sizes ranging 0.09 to 0.94; **Fig.S4-5**, Tables S2-4), while only isolated significant differences that did not validate across all three databases were seen for conflicting interpretation (CI) or noncoding variants (**Fig.S6-11**, Tables S2-6).

Next, to understand the magnitude and potential causes of VUS disparity, we ranked all curated clinical genes based on their difference in VUS allele prevalence and examined which genes were amenable to current MAVE techniques (**Fig.SF12-14** Tables S7-9). Over 83% of VUS across each medical specialty for all three databases were missense variants (**Fig.S15**). However, when missense VUS were excluded, the significant difference in VUS prevalence persisted (p -values ranging 2.78e-70 to 1.2e-05; effect sizes ranging 0.21 to 0.60;

Fig.3, Fig.S16, Tables S10-12), emphasizing the VUS disparity is not driven solely by missense variants. In-frame indels, splice region, and synonymous variants also drove the VUS disparity (p -values ranging 1.63e-194 to 1.63e-04; effect sizes ranging 0.11 to 0.51; **Fig.S19**, Tables S13-15). All four of these variant types, missense, in-frame indels, splice region, and synonymous variants can be systematically catalogued via MAVEs.

Increased P/LP Classifications for European-like Genetic Ancestry

Using a second orthogonal statistical approach based on unique variants exclusive to only one superpopulation group, we show similar patterns for each clinical variant classification category. Across all medical specialties and all three databases, the non-European-like genetic ancestry group exhibited significantly higher counts of unique VUS, B/LB, CI, and ND variants (p -values ranging 7.97e-156 to 6.215e-18, **Fig.S25**), while pathogenic variants were the sole clinical classification where the European-like genetic ancestry group showed significantly higher counts ($p=1.05e-05$, **Fig.4a**). Further, the overlap of unique variants shared between superpopulation groups for VUS, P/LP, B/LB, and especially for CI variants is significantly greater relative to ND variants across every medical specialty in all three databases (p -values ranging 7.97e-156 to 6.215e-18, **Fig.4b**, **Fig.S26**). Thus, our current understanding of clinical variation especially pathogenic variation for individuals of non-European-like genetic ancestry is heavily shaped and limited by our existing knowledge of clinical variation in individuals of European-like genetic ancestry.

Among all curated clinical genes (GenCC), all five genetic ancestries included in the non-European-like superpopulation group, African/African-American, Latino/Admixed American, South Asian, East Asian, and Other, demonstrated significantly decreased P/LP prevalence when compared to European-like genetic ancestry across all three databases ($p\leq 1.67e-05$; **Fig.4c**, **Fig.S27**). Overall, across all curated clinical genes and all three databases, there are an average of 29.8 VUS per individual of non-European-like genetic ancestry versus 24.3 VUS per individual of European-like genetic ancestry (**Fig.4d**).

Integration of MAVE Data Reduces VUS Disparity

Next, we tested our hypothesis that the saturation nature of MAVE data would produce functional scores for VUS from individuals of non-European-like genetic ancestry and reduce VUS disparity. We built an automated VUS reclassification pipeline based on ClinGen VCEP rules for *BRCA1*, *TP53*, and *PTEN* with the amendment that we incorporated clinically calibrated MAVE data for the functional evidence codes. Given both the *All of Us* Public Data Browser and gnomAD are public genomic resources with deidentified variant data, we did not

possess requisite individual-specific clinical histories to assess the clinically oriented evidence codes of the ClinGen VCEP criteria for gene-specific variant interpretation (**Fig.S28**). Thus, to validate the accuracy of our variant reclassifications, we benchmarked our pipeline against the Fayer et al. (2021)¹³ dataset where MAVE data was used for VUS reclassification. Our automated pipeline produced variant reclassifications that were 100% concordant for the 168 reclassified VUS in Fayer et al. (2021) (Table S18).

We found a significantly increased VUS prevalence ($p=8.7e-06$; one-tail z proportions test) for *BRCA1*, *TP53*, and *PTEN* across the three databases: 604 VUS across 206,975 non-European-like individuals assessed, compared to 480 VUS across 213,663 European-like individuals assessed (**Fig.5a**, Table S18). In individuals of European-like genetic ancestry, we reclassified 480 VUS as 315/480 (65.6%) Likely Benign, 4/480 (0.8%) as Benign, 16/480 (3.3%) as Likely Pathogenic, and 145/480 (30.2%) remained VUS (**Fig.5b**, Table S18, **Fig.SF29-30**). In individuals of non-European-like genetic ancestry, we reclassified the 604 VUS as 405/604 (67.1%) Likely Benign, 54/604 (8.9%) as Benign, 5/604 (0.8%) as Likely Pathogenic, and 140/604 (23.2%) remained VUS. MAVE evidence codes were used by most reclassified VUS alleles at 97.0% (775/799) compared to 75.8% (606/799) for computational predictors and 47.9% (383/799) for allele frequency (**Fig.5c**, Table S18). The statistically significant difference in reclassification rates ($p=9.06e-03$; one-tail z proportions test; **Fig.5b**, Table S18) between the two superpopulation groups resulted in nearly the same number of VUS remaining after reclassification in the non-European-like (140) and European-like (145) groups with no significant discernible disparity remaining (**Fig.5a**).

Inequitable Impact of Computational Predictor and Allele Frequency Evidence Codes

We quantified "essential codes." For each variant, we deemed an evidence code as essential if removal would revert the reclassified variant back to VUS. We did not observe any significant difference in essentiality of MAVE codes between individuals of European-like (64.9%) versus non-European-like genetic ancestry (63.9%) (**Fig.5c**, **Fig.S31**, Table S18). Surprisingly, we did observe a significant difference in essentiality of computational predictor (37.3% non-European-like versus 49.8% European-like; $p=1.65e-03$, one-tail z proportions test) and allele frequency codes (7.0% non-European-like versus 21.6% European-like; $p=1.13e-05$, one-tail z proportions test, **Fig.5c**, Table S18). We validated this finding in the Fayer et al. (2021)¹³ dataset and observed no significant difference in essentiality of MAVE codes but a significant difference for computational predictor codes (77.9% non-European-like versus 84.1% European-like; $p=5.73e-04$, one-tail z proportions test; **Fig.S32**, Table S19).

This suggests the impact of computational predictor and allele frequency evidence codes towards VUS reclassification is not equitable for the two superpopulation groupings and, at least in part, describes the gap contributing to VUS disparity for which MAVE evidence compensates.

Discussion

Our findings have important implications for ascertaining molecular diagnoses across medical specialties in patients of non-European-like genetic ancestry. Clinicians and genetic counselors should be aware when ordering next-generation sequencing (NGS) tests for non-European patients, there is a significantly higher pre-test probability of finding VUS or B/LB variants and significantly lower pre-test probability of finding P/LP variants relative to patients of European-like genetic ancestry. We show MAVE data reclassifies VUS at a significantly higher rate in individuals of non-European-like genetic ancestry compared to European-like compensating for the initial VUS disparity. Two prior studies reported VUS reclassification rates of 15.3%³⁰ and 7.3%¹² with clinical evidence codes being most important for VUS resolution¹². Our study incorporated MAVE data and achieved a cumulative VUS reclassification rate of 73.7% without clinical evidence codes. Our VUS reclassification rate is similar to other single gene MAVE studies: 50% in *BRCA1*¹³, 69% in *TP53*, 75% in *MSH2*³¹ and 93% in *DDX3X*³². These findings underscore the necessity of proactive engagement in saturation-style MAVE data production for VUS reclassification at scale to advance our understanding of clinical variation in a more inclusive manner.

Importantly, the genetic ancestry groupings dictating our sample classifications are artificially bounded and not reflective of continuous human genetic variation³³. We grouped individuals classified as non-European to improve statistical power due to limited sample sizes for each ancestry group. The non-European-like genetic ancestry group will contain a large number of admixed individuals, including many who are significantly admixed with individuals of European-like genetic ancestry. We hypothesize admixed individuals likely benefit from reduced VUS rates relative to more distantly related individuals, or those with reduced admixture proportions. Further, the population seen by clinical testing labs is significantly enriched in potential P/LP and VUS relative to population databases. Yet, we still identify consistent and significant trends across all three population databases independent of differences in genetic ancestry calculations, reference genome or NGS assay (**Fig.S1**).

Mechanistically, our findings suggest the increased VUS prevalence in individuals from non-European genetic ancestries is primarily due to the inability to interpret their genetic diversity. Non-Europeans had a

significantly greater number of unique variants with no clinical designation and B/LB variants compared to Europeans, leading to a higher baseline prevalence of VUS in non-Europeans, which effectively remains uninterpreted due to the lack of sufficient evidence to classify these variants as pathogenic or benign. In contrast, Europeans had a higher P/LP prevalence. This discrepancy is attributable to historical disparities in access to genetic testing for individuals of non-European-like genetic backgrounds. As shown here, this has resulted in clinical variant databases enriched in clinically-relevant and pathogenic variation from individuals of European-like genetic ancestry giving a biased representation of global human genetic variation that hinders the interpretation of non-European-like genetic diversity.

This hinderance can be directly observed by the inequitable impact of the allele frequency and computational predictor evidence codes towards VUS reclassification. Allele frequency is directly impacted by the quantity and disproportionate levels of sequencing across populations. When computational predictors are trained and tested against excerpts of current sequencing and clinical variant databases^{34,35}, there is a risk of overfitting on the distinctions between pathogenic and benign variations primarily within the European genetic ancestry group which may not always be translatable to other ancestry groups. Even though computational predictors produce saturation-style variant effects, we posit the lack of diverse training and testing data has perpetuated forward as AI bias preventing equitable impact for VUS reclassification and contributing to VUS disparity as seen in this study. Likely thousands of individuals of non-European-like genetic ancestry have had inequitable variant interpretations due to this bias in computational predictors. Thus, a systematic analysis should be undertaken to understand the potential bias of a variety of commonly used computational predictors. MAVES could mitigate AI bias by producing saturation-style training data for future computational predictors.

The forthcoming new standards, ACMG/AMP/CAP/ClinGen Sequence Variant Guidelines v4, for variant interpretation suggest returning VUS with a high likelihood of pathogenicity to providers for clinical follow-up. We suggest availability of saturation-style MAVE data may help to ensure equitable benefit of this VUS gradation across populations and mitigate any unintentional exacerbation of the current VUS disparity.

Calls for diversifying genomics have yielded a pangenome reference³⁶, H3Africa to equip Africa with genomics infrastructure³⁷, and diverse participant recruitment in *All of Us*³⁸. Diversifying genomics via recruitment, engagement, and retention is just one approach to pursuing equity³⁹. MAVES provide an orthogonal, experimental approach that can complement current sequencing efforts and benefit *All of Us* participants and

millions from non-European-like genetic ancestry in global biobanks. MAVEs can scale to the size of the VUS reclassification problem. The saturation-style of MAVE data can also produce equitable training and testing data for future computational predictors. Expansion of MAVE data can spearhead an equitable revolution in genomic medicine for populations previously left on the margins of genetic research.

Acknowledgements: This study originated within the Atlas of Variant Effects (AVE) and was further supported as a cross-consortia project via the Trans-Variant working group of the Impact of Genomic Variation on Function (IGVF) consortia of the United States National Human Genome Research Institute (NHGRI). Additional funding was provided in part by the NHGRI Genomics Research Elucidates Genetics of Rare Disease (BCM GREGoR Center, U01HG011758 for MD, JEP, JRL, RAG), NHGRI IGVF (University of Washington (UW) Center for Actionable Variant Analysis; UM1HG011969 for MD, SF, SP, MP, LAM, DMF, AFR, LMS), NHGRI Centers of Excellence in Genomic Sciences (UW Center for Multiplexed Assessment of Phenotypes; RM1HG010461 for MD, SF, SP, MP, LAM, DMF, AFR, LMS), NHGRI Clinical Genome (ClinGen) Resource (BCM/Stanford ClinGen Resource; U24HG009649 for SEP) and the NIH *All of Us* Program (The Baylor-Hopkins Clinical Genomics Center for *All of Us*; OT2OD002751 for MD, DK, KP, EV, RAG). MD was also supported by the Baylor College of Medicine Comprehensive Cancer Training Program of the Cancer Prevention Research Institute of Texas (CPRITRP210027). CDRE was supported by the Wellcome Trust through a Career Development Award (227228/Z/23/Z), the Melanoma Research Alliance (825924) and the Chan-Zuckerberg Initiative through the Ancestry Networks for the Human Cell Atlas grant program (CZI007). IGR was supported in part by Australian Research Council Discovery Project DP200101552, National Health and Medical Research Council Ideas Grant 2020501 and the European Union through the Horizon 2020 Research and Innovation Program under Grant No. 810645 and the European Union through the European Regional Development Fund Project No. MOBEC008. St Vincent's Institute acknowledges the infrastructure support it receives from the National Health and Medical Research Council Independent Research Institutes Infrastructure Support Program and from the Victorian Government through its Operational Infrastructure Support Program. We thank Ryan D. Hernandez and Pradiptajati Kusuma for comments on the manuscript.

The *All of Us* Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. In addition, the *All of Us* Research Program would not be possible without the partnership of its participants.

Conflicts of Interest: JRL has stock ownership in 23andMe, is a paid consultant for Regeneron Genetics Center, and is a coinventor on multiple U.S. and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, and bacterial genomic fingerprinting. JRL serves on the Scientific Advisory Board of Baylor Genetics. EV, JRL, and RAG declare that Baylor Genetics is a Baylor College of Medicine affiliate that derives revenue from genetic testing. BCM and Miraca Holdings have formed a joint venture with shared ownership and governance of Baylor Genetics which performs clinical microarray analysis and other genomic studies (exome sequencing and whole genome sequencing) for patient and family care. EV is a co-founder of Codified Genomics, a provider of genetic interpretation.

Code and Data Availability: All the code for replicating all the figures and tables in the main text and supplement is readily accessible at <https://github.com/MoezDawood/ReducingVariantClassificationInequities.git>. Additionally, all input data derived from both gnomAD v2.1.1 and gnomAD v3.1.2 (non v2) is linked at the GitHub above, and publicly available at <https://gnomad.broadinstitute.org/>. All of Us data used in this manuscript is publicly available through the All of Us Data v7 browser (<https://databrowser.researchallofus.org/variants>) but was accessed via the All of Us Researcher Workbench as Baylor College of Medicine is an approved institution. The project was declared on the All of Us Research Projects Directory in accordance with the All of Us data access model. This declaration can be publicly viewed at <https://allofus.nih.gov/protecting-data-and-privacy/research-projects-all-us-data/> ("MAVEComparison"). Both the input data and associated code are accessible through the All of Us workbench and will be promptly shared with requesters with approved

workbench access. The code used for analysis of the All of Us data is the same as in the above GitHub with minor modifications made that are specific to the All of Us Researcher Workbench. Complete rankings across all three databases of all clinically curated genes by VUS/Pathogenic/Benign/CI/ND allele prevalence difference are also linked to the GitHub. All variant reclassifications will be submitted to ClinVar and made publicly available following peer review.

Keywords: MAVE, Multiplexed Assay of Variant Effects, Variants of Uncertain Significance, VUS, Pathogenic, Benign, genetic ancestry, equity, inequity, *All of Us*, gnomAD, missense

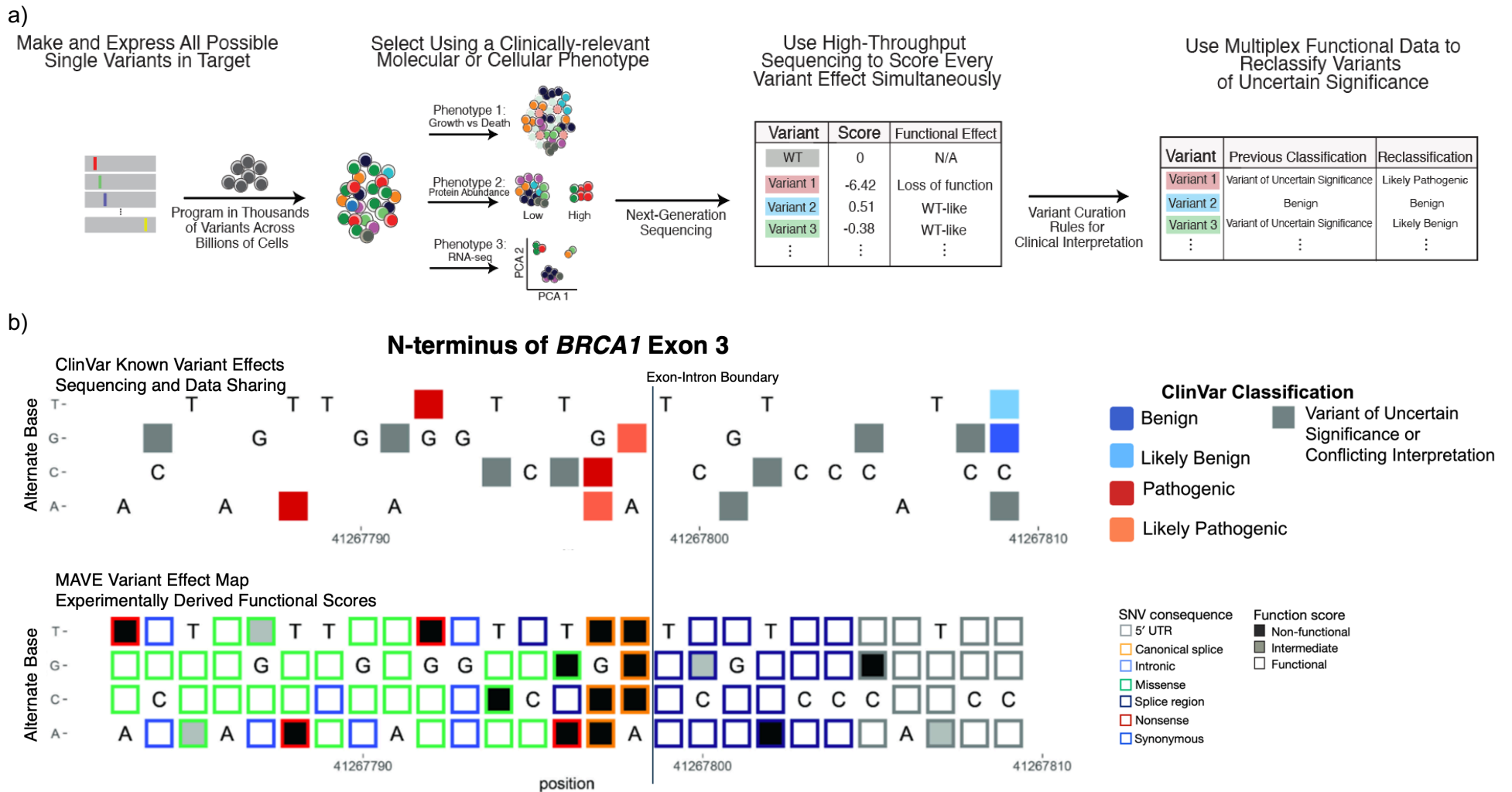


Figure 1: Multiplexed Assays of Variant Effects (MAVEs) Produce Saturation-level Variant Effect Maps Containing Functional Scores For Every Variant in a Target Locus. a) General scheme depicting the workflow of a MAVE starting with the design and construction of potentially every possible SNV or indel in a target locus. Next, the constructed variants are introduced into cells *in vitro*. MAVEs by their nature are able to test thousands of variant simultaneously across millions or potentially billions of cells ensuring each variant is programmed across thousands of cells for functional interrogation. After engineering the variants into the cells, a multiplexable phenotype such as cellular viability over time or fluorescence of an expressed

protein is measured. Changes in the measured molecular phenotype for each variant are then read out via next generation sequencing. Functional scores are then calculated from the sequencing data for each variant. When used within the standard ACMG/AMP clinical interpretation framework, potential PS3/BS3 evidence codes of varying strengths dependent on clinical calibration of the functional scores can reclassify VUS. **b)** Both the top and bottom maps show the N-terminus of *BRCA1* Exon 3 for comparison purposes. The top map represents all the known ClinVar classifications for this particular locus as of November 2023 in the style of a MAVE variant effect map. The bottom map is an excerpt and adaptation of the *BRCA1* MAVE variant effect map from Findlay et al. (2018)²² where the experimental functional scores are depicted by shading and mutational consequences by the outline color of each SNV box. For both maps, reference nucleotides are indicated by the letters based on their position in the GRCh37 reference genome (position numbers of x-axis), the alternate nucleotides are indicated by the row labels (y-axis) and missing data is represented by no boxes. Notably, the MAVE variant effect map exhibits significantly higher information content with no missing SNV functional effects, while the map of clinical significance data contains much sparser information with VUS and missing data dominating the map. Of note, *BRCA1* is one of if not the most well studied gene in medical genetics. Thus, for most other genes the difference in information content would be even more pronounced as there would be even sparser clinical information, but the MAVE map would still be saturated. Further, because of the saturation nature of the MAVE map, there is no bias in variant selection to include in the assay – all variants in the target locus receive a functional score.

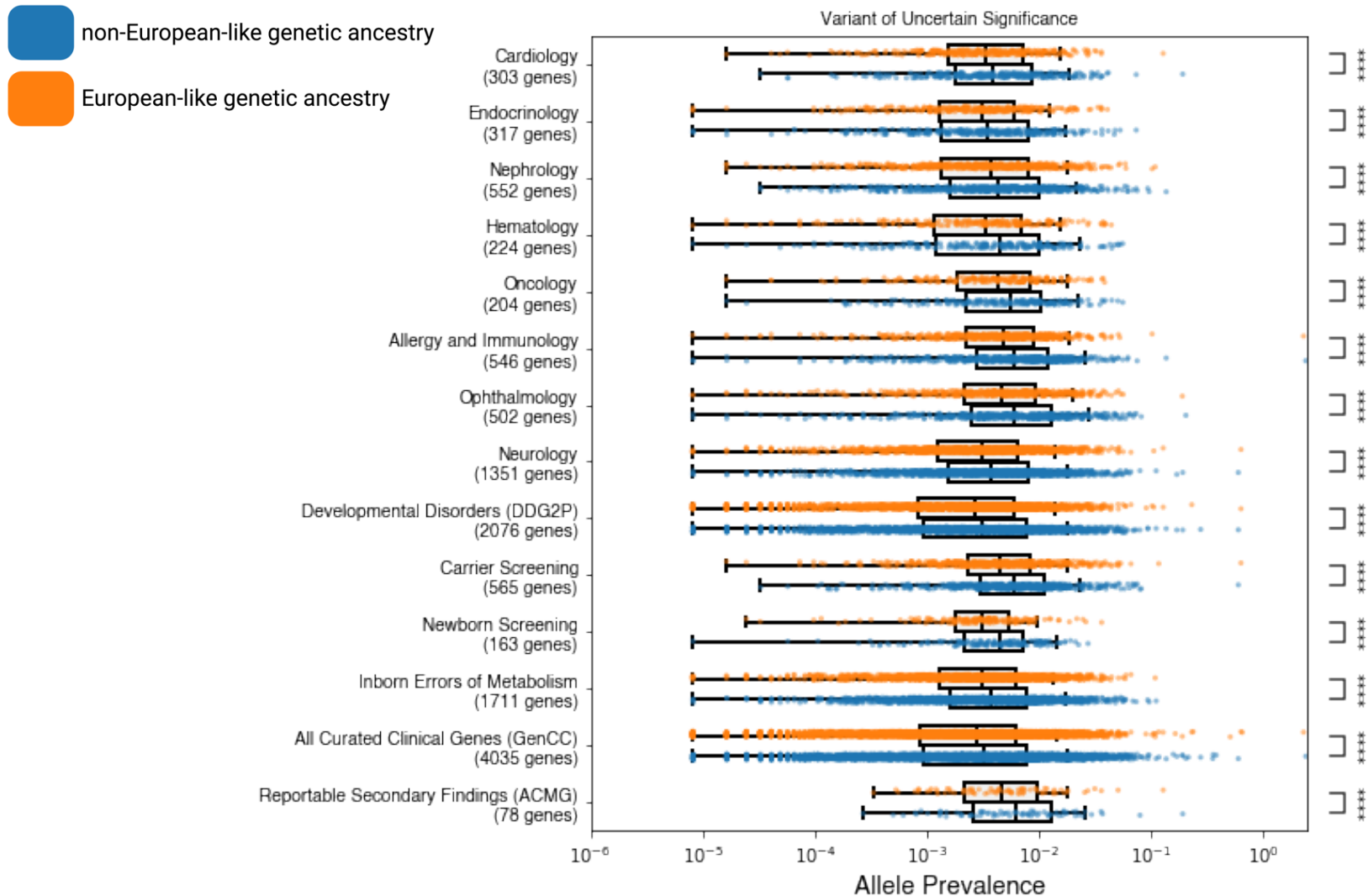


Figure 2: Higher VUS Prevalence Found in Individuals of Non-European-like Genetic Ancestry Across Medical Specialties

Box plots corresponding to VUS allele prevalence (x-axis) in each gene (dot) for individuals of non-European-like (blue) versus European-like (orange) genetic ancestry for the corresponding medical specialty (y-axis) as best visualized in *All of Us* v7 for all coding variants. Genes with zero alleles for allele prevalence for either individuals of European-like or non-European-like genetic ancestry are omitted from the above visualization to maintain a reasonable scale for data visualization. However, genes with zero alleles for only one category of either individuals of European-like or non-European-

like genetic ancestry are included in the Bonferroni-corrected, signed rank, matched pairs Wilcoxon statistical test. The Bonferroni corrected p-values associated with these comparisons are annotated as follows with "ns" indicating not significant, * for $1.19\text{e-}04 < p \leq 2.38\text{e-}04$, ** for $5.95\text{e-}05 < p \leq 1.19\text{e-}04$, *** for $5.95\text{e-}06 < p \leq 5.95\text{e-}05$, and **** for $p \leq 5.95\text{e-}06$. Across all medical specialties and categories shown, VUS are observed to be statistically significantly increased in individuals of non-European-like genetic ancestry compared to individuals of European-like genetic ancestry.

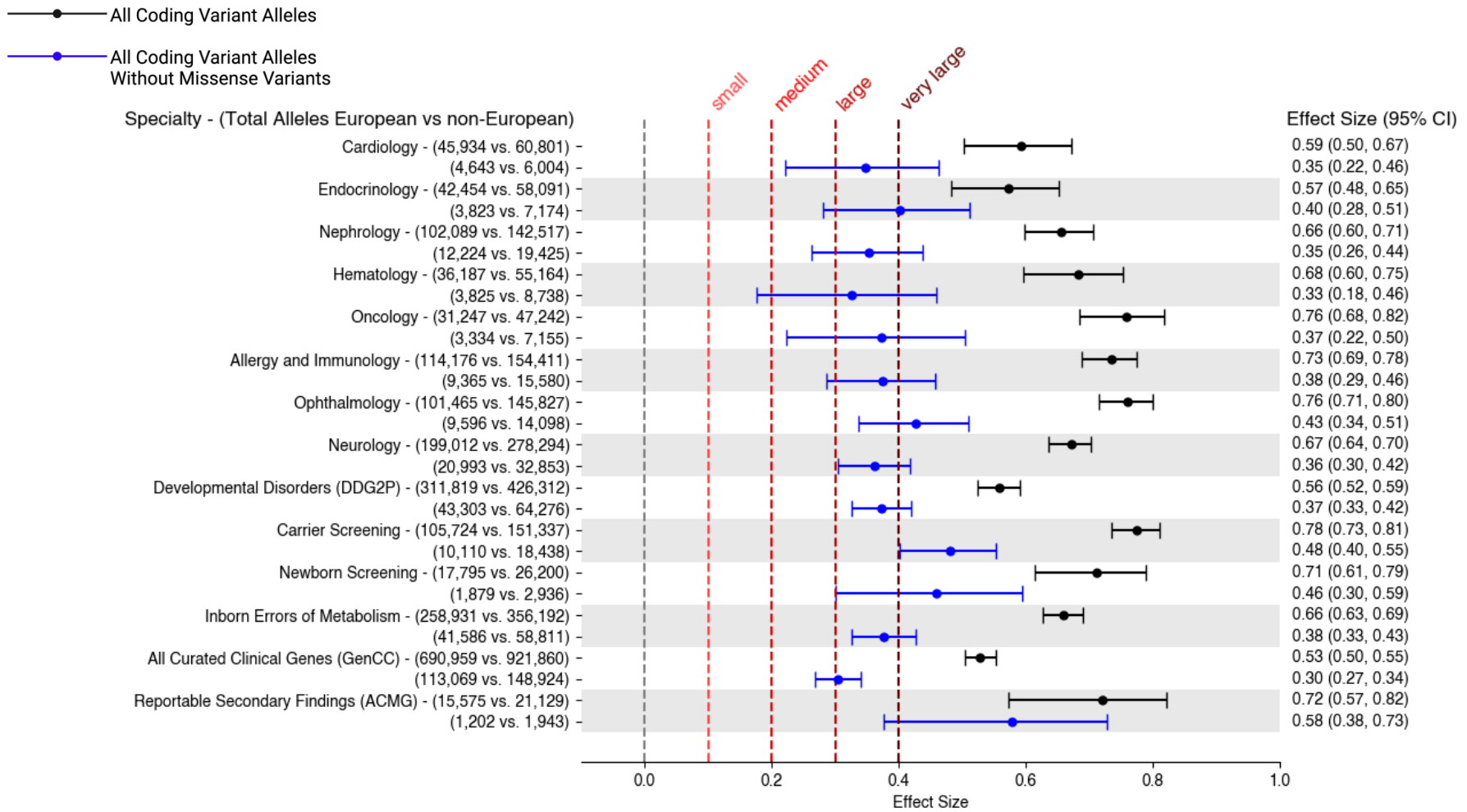
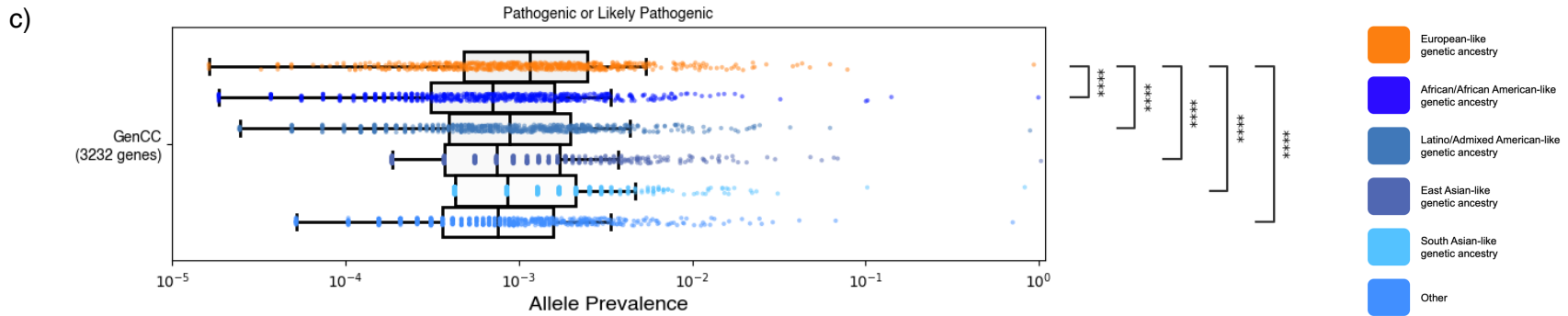
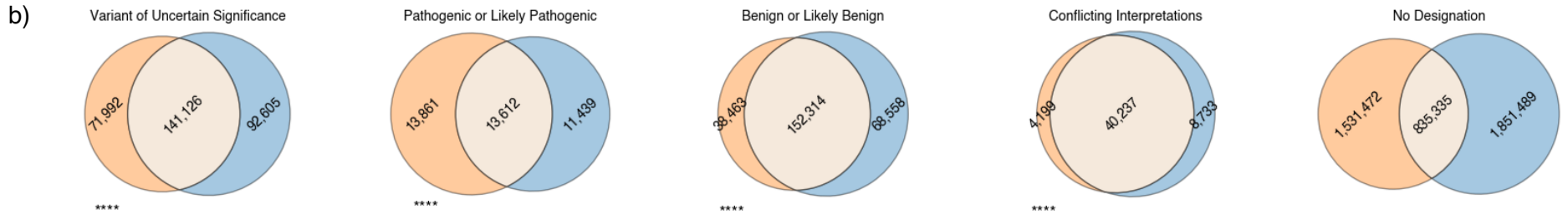
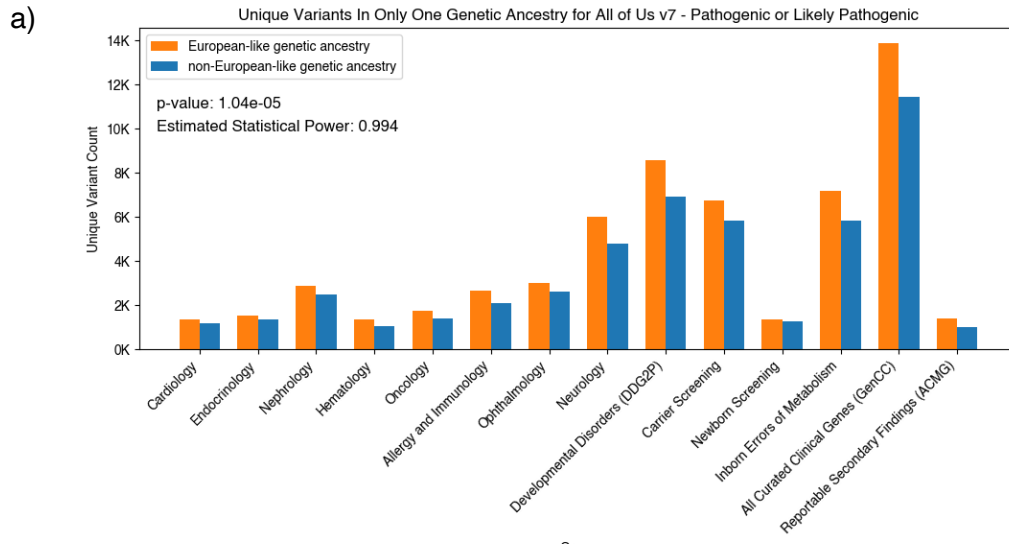


Figure 3: The Disparity in VUS Prevalence is Present Even in the Absence of Missense Variants

Effect size with 95% confidence interval (plotted and denoted on the right) shown for the differences between VUS prevalence in individuals of non-European-like versus European-like genetic ancestry as measured by the rank biserial coefficient from the signed rank, matched pairs, Wilcoxon test with a Bonferroni correction as best visualized in gnomAD v3.1.2 (non v2). The total number of alleles from individuals of non-European-like versus European-like genetic ancestry are indicated on the left. Effect sizes in black were calculated from all coding variants while effect sizes in blue were calculated from all coding variants excluding missense variants corresponding to the medical specialty (y-axis). Thresholds as determined by Funder

and Ozer (2019)²¹ for quantifying the magnitude of the effect size difference are plotted as vertical dashed lines. Across medical specialties and categories, the disparity in VUS prevalence between individuals of non-European-like versus European-like genetic ancestry is not just statistically significant but very large. Further, the statistically significant disparity in VUS prevalence is still intact and medium to large even with the exclusion of missense VUS (~85-90% of all VUS) across the medical specialties.



d) Average Number of Alleles in All Curated Clinical Genes with a Clinical Variant Classification per Individual

Genetic Ancestry	VUS	P/LP	B/LB	CI
Non-European-like	29.8	4.0	8,232	126.2
European-like	24.3	4.3	8,016	122.4

Figure 4: Lower Prevalence of Pathogenic or Likely Pathogenic Variants In Non-European-like Genetic Ancestry Group

a) Grouped bar graphs corresponding to unique coding variant counts (y-axis) for P/LP variants found either only in individuals of European-like (orange) genetic ancestry or only in individuals of non-European-like (blue) genetic ancestry across the medical specialties (x-axis) in *All of Us* v7. The Bonferroni corrected p-values from the chi-square test of independence associated with these comparisons are annotated along with the estimated statistical power. b) Venn Diagrams stratified by clinical variant classification (columns) for unique variants found either only in individuals of European-like (orange) genetic ancestry or only in individuals of non-European-like (blue) genetic ancestry across the medical specialties (y-axis) for all unique coding variant counts in *All of Us* v7 for the list of all curated clinical genes (GenCC). Across all medical specialties and categories shown, the overlap of unique variants for VUS, P/LP, B/LB, and CI variants is statistically significantly greater (chi-square test for independence) than the overlap between ND variants in all 3 databases across all medical specialties in both the presence and absence of missense variants. c) Box plots corresponding to Pathogenic or Likely Pathogenic allele prevalence (x-axis) in each gene (dot) for each non-European-like (shades of blue) genetic ancestry versus European-like (orange) genetic ancestry for all curated clinical genes (GenCC list; y-axis) across all three population databases (*All of Us* v7, gnomAD v2.1.1 and gnomAD v3.1.2 (non v2)) for all coding variants. Genes with zero alleles for allele prevalence for either individuals of European-like or non-European-like genetic ancestry are omitted from the above visualization to maintain a reasonable scale for data visualization. However, genes with zero alleles for only one category of either individuals of European-like or non-European-like genetic ancestry are included in the Bonferroni-corrected, signed rank, matched pairs Wilcoxon statistical test. The Bonferroni corrected p-values associated with these comparisons are annotated as follows with "ns" indicating not significant, * for $3.33e-04 < p \leq 6.67e-04$, ** for $1.67e-04 < p \leq 3.33e-04$, *** for $1.67e-05 < p \leq 1.67e-04$, and **** for $p \leq 1.67e-05$. There is a statistically significantly increased prevalence of Pathogenic or Likely pathogenic variants found in the European-like genetic ancestry relative to each non-European-like group. d) Average number of alleles per individual based on weighted average of all three databases in all curated clinical genes (GenCC) for the differing clinical variant classifications.

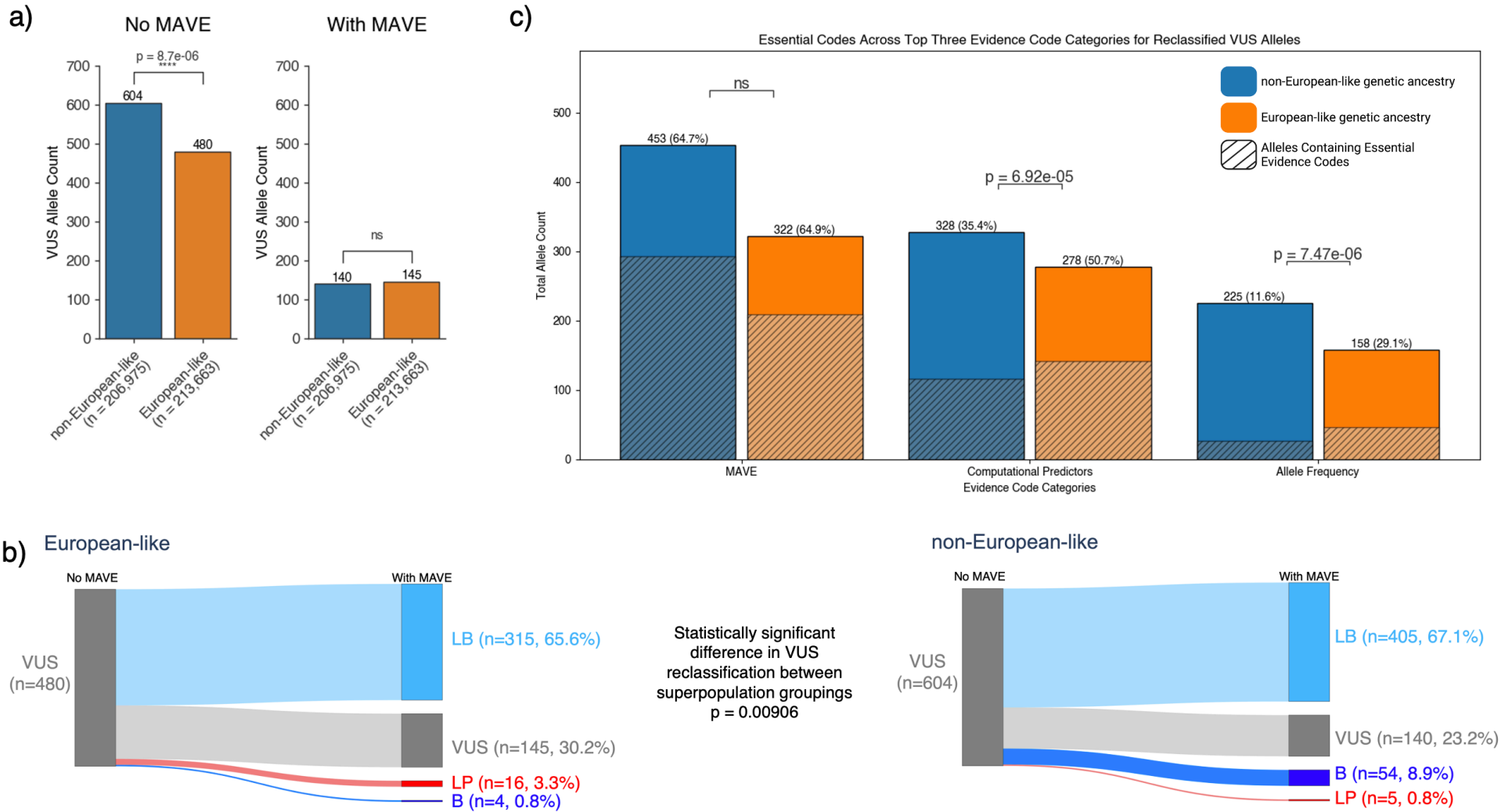


Figure 5: MAVE Data can Reclassify non-European-like VUS at a Statistically Significant Higher Rate Compared to European-like VUS

a) Sankey flow diagrams depicting VUS reclassification (read from left to right) for individuals of European-like (left) versus non-European-like (right) genetic ancestry before reclassification (No MAVE) and after reclassification (With MAVE). The examined VUS for *BRCA1*, *TP53* and *PTEN* are the total VUS alleles summed from all three databases *All of Us* v7, gnomAD v2.1.1, and gnomAD v3.1.2 (non v2) corresponding to the coding region saturated by the MAVE. The VUS were reclassified as either Likely Benign (LB; light blue), Benign (B; dark blue), Likely Pathogenic (LP; red), or

remained as Variants of Uncertain Significance (VUS; gray). Reclassification was conducted using an automated pipeline based on the ClinGen Variant Curation Expert Panel gene specific variant interpretation guidelines for each gene with the amendment of using clinically-calibrated MAVE data for the functional evidence codes. **b)** The presence of VUS in individuals of non-European-like versus European-like genetic ancestry was statistically significantly higher in non-European-like superpopulation group. However, after using MAVE data for reclassification in the ClinGen VCEP frameworks, there was no statistically significant VUS disparity detected. **c)** Bar graphs for each evidence code category (x-axis) used in VUS reclassification across *BRCA1*, *TP53*, and *PTEN* for all three databases, *All of Us* v7, gnomAD v2.1.1, and gnomAD v3.1.2 (non v2). Blue bars represent alleles from individuals of non-European-like genetic ancestry, whereas orange bars represent alleles from individuals of European-like genetic ancestry. Shading represents essential codes, codes which if removed from the set of evidence codes used to reclassify the VUS would cause the variant to regress back to VUS. MAVE evidence codes were used the most based on total allele count for both individuals of non-European-like and European-like genetic ancestry. However, computational predictor and allele frequency codes were more essential for individuals of European-like genetic ancestry. PP3, PP3_Moderate, and BP4 correspond to the computational predictor codes. PS3, PS3_Moderate, BS3, BS3_Moderate, BS3_Supporting corresponded to the MAVE evidence codes. BA1, BS1, BS1_Supporting correspond to the allele frequency codes.

References

1. Mata DA, Rotenstein LS, Ramos MA, Jena AB. Disparities According to Genetic Ancestry in the Use of Precision Oncology Assays. *New England Journal of Medicine* 2023;388(3):281–3.
2. Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. *Nat Med* 2022;28(2):243–50.
3. Borrell LN, Elhawary JR, Fuentes-Afflick E, et al. Race and Genetic Ancestry in Medicine — A Time for Reckoning with Racism. *New England Journal of Medicine* 2021;384(5):474–80.
4. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019;51(4):584–91.
5. Collins FS, Doudna JA, Lander ES, Rotimi CN. Human Molecular Genetics and Genomics — Important Advances and Exciting Possibilities. *New England Journal of Medicine* 2021;384(1):1–4.
6. Matalon DR, Zepeda-Mendoza CJ, Aarabi M, et al. Clinical, technical, and environmental biases influencing equitable access to clinical genetics/genomics testing: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine* 2023;25(6):100812.
7. Manrai AK, Funke BH, Rehm HL, et al. Genetic Misdiagnoses and the Potential for Health Disparities. *New England Journal of Medicine* 2016;375(7):655–65.
8. Cook S, Dunn E, Kornish J, et al. Molecular testing in newborn screening: VUS burden among true positives and secondary reproductive limitations via expanded carrier screening panels. *Genet Med* 2023;26(4):101055.

9. Venner E, Patterson K, Kalra D, et al. The frequency of pathogenic variation in the All of Us cohort reveals ancestry-driven disparities. *Commun Biol* 2024;7(1):1–11.
10. Wright CF, Campbell P, Eberhardt RY, et al. Genomic Diagnosis of Rare Pediatric Disease in the United Kingdom and Ireland. *New England Journal of Medicine* 2023;388(17):1559–71.
11. Abul-Husn NS, Marathe PN, Kelly NR, et al. Molecular diagnostic yield of genome sequencing versus targeted gene panel testing in racially and ethnically diverse pediatric patients. *Genetics in Medicine* 2023;25(9):100880.
12. Chen E, Facio FM, Aradhya KW, et al. Rates and Classification of Variants of Uncertain Significance in Hereditary Disease Genetic Testing. *JAMA Network Open* 2023;6(10):e2339571.
13. Fayer S, Horton C, Dines JN, et al. Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN. *Am J Hum Genet* 2021;108(12):2248–58.
14. Horton C, Hoang L, Zimmermann H, et al. Diagnostic Outcomes of Concurrent DNA and RNA Sequencing in Individuals Undergoing Hereditary Cancer Testing. *JAMA Oncology* [Internet] 2023 [cited 2024 Jan 8]; Available from: <https://doi.org/10.1001/jamaoncol.2023.5586>
15. Fowler DM, Adams DJ, Gloyn AL, et al. An Atlas of Variant Effects to understand the genome at nucleotide resolution. *Genome Biology* 2023;24(1):147.
16. Fowler DM, Araya CL, Fleishman SJ, et al. High-resolution mapping of protein sequence-function relationships. *Nat Methods* 2010;7(9):741–6.
17. Macdonald CB, Nedrud D, Grimes PR, Trinidad D, Fraser JS, Coyote-Maestas W. DIMPLE: deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology. *Genome Biology* 2023;24(1):36.

18. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581(7809):434–43.
19. Chen S, Francioli LC, Goodrich JK, et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* 2024;625(7993):92–100.
20. Coop G. Genetic similarity versus genetic ancestry groups as sample descriptors in human genetics [Internet]. arXiv.org. 2022 [cited 2024 Mar 18]; Available from: <https://arxiv.org/abs/2207.11595v2>
21. Funder DC, Ozer DJ. Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science* 2019;2(2):156–68.
22. Findlay GM, Daza RM, Martin B, et al. Accurate classification of BRCA1 variants with saturation genome editing. *Nature* 2018;562(7726):217–22.
23. Giacomelli AO, Yang X, Lintner RE, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. *Nat Genet* 2018;50(10):1381–7.
24. Matreyek KA, Starita LM, Stephany JJ, et al. Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat Genet* 2018;50(6):874–82.
25. Mighell TL, Evans-Dutson S, O’Roak BJ. A Saturation Mutagenesis Approach to Understanding PTEN Lipid Phosphatase Activity and Genotype-Phenotype Relationships. *The American Journal of Human Genetics* 2018;102(5):943–55.
26. Parsons MT, Hoya M de la, Richardson ME, et al. Evidence-based recommendations for gene-specific ACMG/AMP variant classification from the ClinGen ENIGMA BRCA1 and BRCA2 Variant Curation Expert Panel [Internet]. 2024 [cited 2024 Jan 29]; 2024.01.22.24301588. Available from: <https://www.medrxiv.org/content/10.1101/2024.01.22.24301588v1>

27. Fortuno C, Lee K, Olivier M, et al. Specifications of the ACMG/AMP variant interpretation guidelines for germline TP53 variants. *Human Mutation* 2021;42(3):223–36.
28. Mester JL, Ghosh R, Pesaran T, et al. Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Human Mutation* 2018;39(11):1581–92.
29. Rehm HL, Berg JS, Brooks LD, et al. ClinGen — The Clinical Genome Resource. *New England Journal of Medicine* 2015;372(23):2235–42.
30. Slavin TP, Van Tongeren LR, Behrendt CE, et al. Prospective Study of Cancer Genetic Variants: Variation in Rate of Reclassification by Ancestry. *JNCI: Journal of the National Cancer Institute* 2018;110(10):1059–66.
31. Scott A, Hernandez F, Chamberlin A, Smith C, Karam R, Kitzman JO. Saturation-scale functional evidence supports clinical variant interpretation in Lynch syndrome. *Genome Biology* 2022;23(1):266.
32. Radford EJ, Tan H-K, Andersson MHL, et al. Saturation genome editing of DDX3X clarifies pathogenicity of germline and somatic variation. *Nat Commun* 2023;14(1):7702.
33. Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research, Board on Health Sciences Policy, Committee on Population, Health and Medicine Division, Division of Behavioral and Social Sciences and Education, National Academies of Sciences, Engineering, and Medicine. *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field* [Internet]. Washington, D.C.: National Academies Press; 2023 [cited 2024 Mar 25]. Available from: <https://www.nap.edu/catalog/26902>

34. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 2016;99(4):877–85.
35. Feng B-J. PERCH: A Unified Framework for Disease Gene Prioritization. *Human Mutation* 2017;38(3):243–51.
36. Liao W-W, Asri M, Ebler J, et al. A draft human pangenome reference. *Nature* 2023;617(7960):312–24.
37. Choudhury A, Aron S, Botigué LR, et al. High-depth African genomes inform human migration and health. *Nature* 2020;586(7831):741–8.
38. The “All of Us” Research Program. *New England Journal of Medicine* 2019;381(7):668–76.
39. Lee SS-J, Appelbaum PS, Chung WK. Challenges and potential solutions to health disparities in genomic medicine. *Cell* 2022;185(12):2007–10.
40. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med* 2019;12(1):3.