Title

Using multiplexed functional data to reduce variant classification inequities in underrepresented populations.

Journal

Genome Medicine, 16(1)

Authors

Dawood, Moez
Fayer, Shawn
Pendyala, Sriram
et al.

Publication Date

2024-12-03

DOI

10.1186/s13073-024-01392-7

Peer reviewed

Genome Medicine

RESEARCH

Open Access

Using multiplexed functional data to reduce variant classification inequities in underrepresented populations

Moez Dawood[1,2,3*], Shawn Fayer[4,5], Sriram Pendyala[5,6], Mason Post[4], Divya Kalra[1], Karynne Patterson[5], Eric Venner[1,2], Lara A. Muffley[4,5], Douglas M. Fowler[4,5,7], Alan F. Rubin[8,9], Jennifer E. Posey[2], Sharon E. Plon[1,2], James R. Lupski[1,2,10,11], Richard A. Gibbs[1,2], Lea M. Starita[4,5], Carla Daniela Robles-Espinoza[12,13], Willow Coyote-Maestas[14,15*] and Irene Gallego Romero[16,17,18,19*]

Abstract

Background Multiplexed Assays of Variant Effects (MAVEs) can test all possible single variants in a gene of interest. The resulting saturation-style functional 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 increased 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.

*Correspondence:
Moez Dawood
Moez.Dawood@bcm.edu
Willow Coyote-Maestas
willow.coyote-maestas@ucsf.edu
Irene Gallego Romero
irene.gallego@svi.edu.au
Full list of author information is available at the end of the article

Keywords MAVE, Multiplexed assay of variant effects, Variants of uncertain significance, VUS, Pathogenic, Benign, Genetic ancestry, Equity, Inequity, *All of Us*, gnomAD, Missense

Background

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 populations [1–3]. 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 ancestry [4–6]. This lack of diversity in control population data has repeatedly led to incorrect diagnoses [7], missed diagnoses [8–11], and inappropriate medical management [7] for individuals of non-European-like genetic ancestry [7–9]. For example, in sequencing studies such as Deciphering Developmental Disorders (2166 non-European-like vs. 11,202 European-like) and NYCKidSeq (519 non-European-like vs. 126 European-like), probands of non-European-like genetic ancestry were less likely to receive a genetic diagnosis versus European-like probands [10, 11]. Compounding these challenges, genomic medicine can struggle to determine if identified genetic variants are of potential clinical impact (Pathogenic or Likely Pathogenic; P/LP) or have no apparent clinical impact (Benign or Likely Benign; B/LB) resulting in over 51% of short variants (defined as affecting 50 base pairs or less) in ClinVar classified as either a Variant of Uncertain Significance (VUS) or Conflicting Interpretation (CI) (1,362,519 VUS plus 126,009 CI out of 2,898,457 short variants as of July 2024). Further, VUS are more commonly reported in individuals of non-European-like ancestry [8, 12–15]. Currently, 22.5% of clinical exome or genome sequencing and 32.6% of multi-gene panels yield inconclusive results due to VUS [16]. While RNA sequencing has been shown to be effective at reclassifying VUS for individuals of non-European-like genetic ancestry [17] and increasing emphasis of diverse participant recruitment and engagement will expand the genetic diversity of biobanks, there is still an exponentially increasing number of VUS [12, 18]. 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 [18–21] (Fig. 1). When the clinical evidence strength for each MAVE was calibrated and the functional scores were systematically integrated into the guidelines for clinical variant interpretation, MAVE data drove reclassification for 50% in *BRCA1* [18], 69% in *TP53*, 75% in *MSH2* [22], and 93% in *DDX3X* [23] culminating in return of updated patient test results to providers. 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

Cohorts

Genomic data from 245,394 individuals enrolled in the *All of Us* v7 cohort were analyzed. Findings were validated using two independent datasets from the Genome Aggregation Database (gnomAD), specifically 123,709 exomes from gnomAD v2.1.1 and 51,535 genomes from gnomAD v3.1.2 (excluding individuals from gnomAD v2). To facilitate comparative analyses, individuals were stratified into two major superpopulation groups: European-like and non-European-like genetic ancestries. The final sample sizes were as follows: non-European-like vs. European-like groups: 122,322 vs. 123,072 (All of Us v7); 59,106 vs. 64,603 (gnomAD v2.1.1); and 25,547 vs. 25,988 (gnomAD v3.1.2). These cohorts and stratification were used for both statistical tests as well as variant reclassification. Further information regarding participant enrollment and sample collection and study origination can be found on the *All of Us* [9, 25–27] or gnomAD [28] website, respectively. Variants for each gene are available on the *All of Us* Public Data Browser [27] version 7 or gnomAD [28]. The clinical variant classifications in this study found in *All of Us* or gnomAD were originally sourced

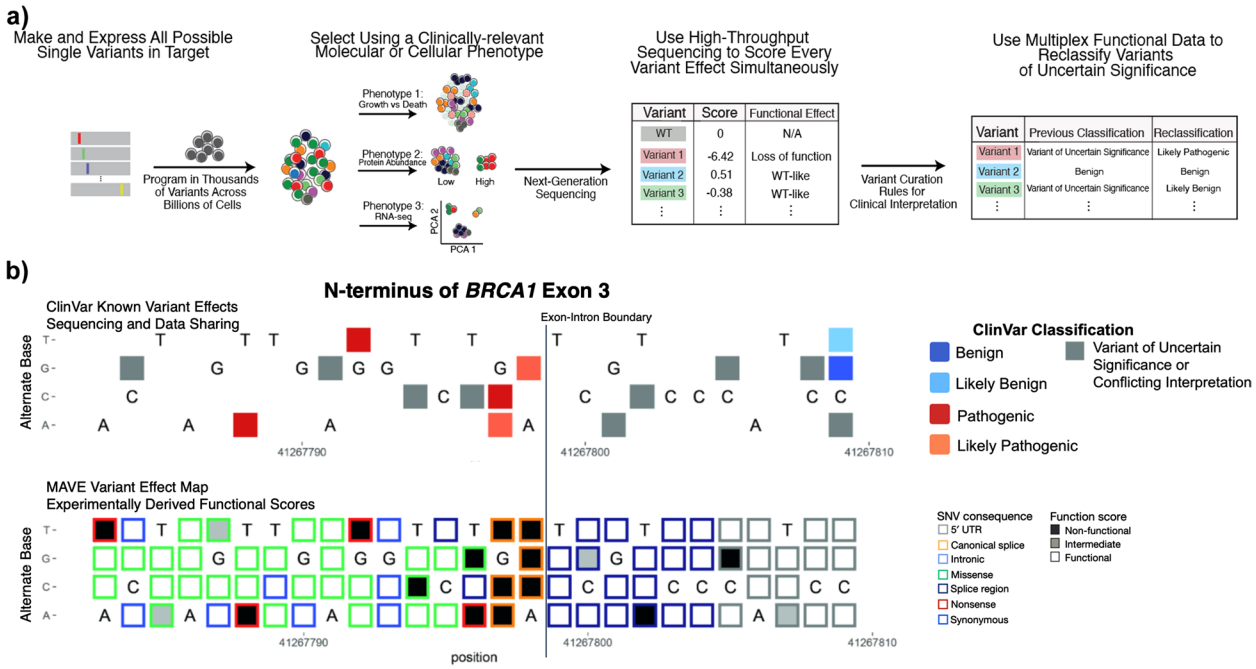

Fig. 1 Multiplexed Assays of Variant Effects (MAVEs) produce saturation-level variant effect maps containing functional scores for every variant in a target locus. **a** General scheme depicting the workflow of a MAVE starting with the design and construction of potentially every possible SNV or indel in a target locus. Next, the constructed variants are introduced into cells in vitro. MAVEs by their nature are able to test thousands of variants 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. [24] 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

from ClinVar [29, 30]. Variant calls, allele counts, population descriptors, and variant classifications were used as prescribed by *All of Us* or gnomAD.

Genetic ancestry

We use descriptors from *All of Us* and gnomAD for consistency, as we cannot reclassify individuals due to the public, deidentified nature of the databases. Full details are available in the respective website documentation and publications of gnomAD [28] and *All of Us* [9, 25–27]. All databases assign a single genetic ancestry to each individual based on projection to principal components built using reference populations [9, 25, 26]. We have appended "-like" to the labels to explicitly reflect that they primarily capture genetic similarity to reference groups used by the original publications to train their classifiers [31]. We acknowledge their imperfect and incomplete nature as descriptors of continuous human diversity. The non-European-like group encompassed individuals with genetic ancestries from the "African/African American," "Latino/Admixed American," "East Asian," "South Asian," and "Other" groups as prescribed by the genetic ancestry calculation done by *All of Us* or gnomAD. In all cases, individuals are assigned to a single genetic ancestry first by projection into a principal component space built from established population genetics resources. Principal

component loadings for each individual are then input into a random forest classifier, and the genetic ancestry label is assigned on the basis of the output from the classifier. Given the nature of random forest classifiers, this approach will struggle to assign a label to admixed individuals and to individuals whose genetic ancestry is poorly represented in the reference samples. These individuals, therefore, make up a significant fraction of the "Other" group, which is openly acknowledged by both *All of Us* and gnomAD.

Gene lists and calculating allele prevalence

Gene lists for medical specialties that commonly use genetic testing were compiled from genes known to be tested on next-generation sequencing tests of Invitae, Ambry Genetics, and Baylor Genetics. The ACMG78 gene list represents the 78 genes from the secondary findings list curated per the American College of Medical Genetics and Genomics Secondary Findings v3.2 standard. The GenCC gene list [32] represents all 4640 curated known clinical disease genes as of June 2023. The "Cancer" gene list represents 209 genes implicated in hereditary cancers and cancer syndromes across every major organ system. The "Cardiac" gene list represents 306 genes implicated in arrhythmias, cardiomyopathies, RASopathies, congenital heart diseases, lipidemias, and aortopathies. The "Hematology" gene list represents 240 genes implicated in benign and malignant blood disorders such as inherited platelet disorders and thrombocytopenias, anemias, enzymopathies, red blood cell membrane disorders, telomere disorders, bone marrow failure, and more. The "Newborn screening" gene list represents 1755 genes implicated in inherited metabolic disorders. The "Carrier Screening" gene list represents 568 genes commonly examined to understand if there is an increased risk of having a child affected with a genetic condition. The "Endocrinology" gene list represents 321 genes implicated in disorders of sex development, obesity, thyroid and parathyroid conditions, bone mineralization disorders, and glucose metabolism. The "Immunology" gene list represents 572 genes implicated in primary immunodeficiency, telomere biology disorders, antibody deficiencies, autoinflammatory syndromes, B and T cell deficiencies, phagocytic defects, hereditary angioedema, complement deficiencies, and congenital diarrhea. The "Nephrology" gene list represents 565 genes implicated in ciliopathies, nephrolithiasis, progressive renal disease, rare clinical syndromes with renal manifestations, atypical hemolytic uremic syndrome, and thrombotic microangiopathies. The "Neurology" gene list represents 1374 genes implicated in neuropathies, movement disorders, neurodegenerative disorders, neurovascular disorders, epilepsy disorders, seizure disorders, neurodevelopmental disorders, and neuromuscular disorders. The "Ophthalmology" gene list represents 514 implicated in blindness are rare disorders affecting vision, the eye, and/or the retina. The DDG2P gene list representing the curated list of 2307 genes reported to be associated with developmental disorders from the DECIPHER project was accessed in June 2023. The "SGE" gene list represents the 694 genes that are both essential in HAP1 cells and found in the GenCC gene list. The "VAMPseq" gene list represents the 394 genes that are both high priority for VAMPseq and found in the GenCC gene list. The high priority VAMPseq genes were selected because their proteins are not secreted extracellularly, thermostable, have previously been shown to be GFP tagged, and are monomeric. The "MAVERegistry" list was determined based on the 110 genes as of August 2023 that are either "Under Investigation" or in the "MAVE Data Collection" phases on the MAVERegistry [33]. When appropriate, the same gene may be found in more than one gene list (for example, *BRCA2* would be found in the Oncology, GenCC, ACMG78, SGE, and MAVERegistry lists). Overall, all gene lists and corresponding ENSG terms used in this study are available in Additional file 1: Table S1. 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).

Clinical significance classifications for variants

From gnomAD, allele prevalence for the individuals of European-like genetic ancestry was calculated from the "European-like (non-Finnish)" group. Due to the high degrees of consanguinity in the Finnish and Ashkenazi Jewish populations, these two populations were not included in our analysis. Allele counts, frequencies, population descriptors, ClinVar clinical significance calls, and number of individuals sequenced in each population were used as prescribed by *All of Us* [34] or gnomAD as of June 2023. As only approximately 2% of short variants (variants affecting less than 50 base pairs) are not assigned "one star" review status, we did not filter for review status or any other metric of clinical variant classification quality to prevent accidentally biasing against individual or smaller labs working with underrepresented communities. Further, for each variant in the *All of Us* v7 where the full set of unique submitted clinical classifications needed to be reconciled to just one clinical variant classification call, we took the most conservative approach per the aggregation of clinical variant germline classification approach used by ClinVar [35]. All clinical significance calls were mapped to one of six categories: "Pathogenic or Likely Pathogenic," "Benign

or Likely Benign," "Variant of Uncertain Significance," "Conflicting Interpretations," "Not Included," or "No Designation" based on their current ClinVar clinical significance designation as specified in *All of Us* or gnomAD. Trends were pinpointed if shown to be consistent across all three databases. Due to differences in extraction of ClinVar data between gnomAD and *All of Us*, there are systematic database level differences that potentially are unaccounted for. In these instances, the GenCC (The Gene Curation Coalition) list of all curated clinical genes being the biggest and most comprehensive list is used as the main indicator of a trend. gnomAD version 2.1.1 and version 3.1.2, non-v2 (removes individuals overlapping between v2 and v3) were treated as two independent population databases [25, 26].

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 thresholds [36]. 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. We ensured that the number of individuals of European-like genetic ancestry and non-European-like genetic ancestry was approximately equal in each database to prevent biased statistical analysis due to differences in group sizes. This is important because both orthogonal statistical methods are based on allele counts within genes or groups of genes (the matched pairs nature of the Wilcoxon test compares non-European-like allele prevalence to European-like allele prevalence and the chi-square test on unique variants).

Wilcoxon matched-pairs signed-rank test

We employed a matched pairs signed rank Wilcoxon test, with the matched pairs based on the gene itself and its allele prevalence between individuals of non-European-like versus European-like genetic ancestry. This gene-by-gene comparison mitigates any other confounders such as gene length, coverage during sequencing, and other gene-specific intricacies that are canceled out by comparing the allele prevalence within the non-European-like group to the allele prevalence in the European-like group within each gene. The ranking aspect of the test is crucial, as it does not presuppose a uniform trend of larger allele prevalence in the non-European-like group compared to the European-like group across all genes for every clinical significance allele type. By ranking the genes prior to the statistical test, we incorporate genes that have a higher number of alleles in Europeans into our analysis, ensuring a complete survey of the allele prevalence in all genes in the statistical test. The difference in allele prevalence and difference in unique VUS between the non-European-like group and European-like group was also used to rank the genes with the greatest VUS disparity between non-Europeans vs. Europeans.

While the p value informs us whether or not there is a difference, we then calculated the rank biserial coefficient (r) with a 95% confidence interval to quantify the magnitude of the statistically significant differences. This calculation was performed using Python-wrapped R code, employing the ggwithinstats function from the ggstatsplot library and the effectsize library, with settings based on thresholds outlined by Funder and Ozer [36]. The resultant coefficient categories are based on the magnitude of $r < 0.05$—tiny; $0.05 \leq r < 0.1$—very small; $0.1 \leq r < 0.2$—small; $0.2 \leq r < 0.3$—medium; $0.3 \leq r < 0.4$—large; and $r \geq 0.4$—very large. Additionally, we evaluated the statistical power of each Wilcoxon matched-pairs signed-rank using a simulation-based approach. The simulation iterates 50,000 times to generate matched pair samples under a normal distribution, with the first sample being the control and the second sample being offset by the defined effect size. Each iteration performs the Wilcoxon signed-rank test to assess the significance of the observed effect based on the Bonferroni-corrected alpha. The proportion of 50,000 iterations yielding significant results was the estimate of statistical power, reflecting the test's ability to correctly reject the null hypothesis for a specified effect size and sample size.

We also assessed the overlap in variants between the non-European-like and European-like groups. This involved calculating the number of variants present in both groups, as well as the number of variants unique to each group, and expressing these as percentage contributions. In contrast to the below orthogonal statistical method, all variants, including those shared between groups, were retained for the Wilcoxon matched-pairs signed-rank test to ensure that any unique variant's prevalence in both populations was duly considered in assessing potential differences.

Chi-square test for independence

Furthermore, we employed a chi-square test for independence to investigate the presence of unique variants in each population group. In contrast to the above

orthogonal statistical method, variants found in both groups were removed for the chi-square test for independence to examine prevalence differences of variants found exclusively in the European-like versus non-European-like genetic ancestry groups with an accompanying power estimate. Instead of the gene-by-gene approach, this approach allowed us to systematically assess three population databases, seeking to determine whether there is a consistent higher count of unique variants (not allele count) across different medical specialties and gene groups. Further, it helps to satisfy the requirement of independence of observations for the chi-square test as there are no relationships between the counts in the individual medical specialty groups and no pairing of the data between the super populations.

It is worth noting that neither the Wilcoxon test nor the chi-square test for independence necessitates an underlying distribution that approximates normality. Visual inspection of variant prevalence in the GenCC data revealed that the distributions of variants best resembles a chi-square distribution. Thus, the chi-square test, based on the chi-square data distribution, is particularly suitable for modeling our data.

For the analysis of ClinVar high confidence variants, 413,016 variants that were short variants (< 50 bp resulting in SNVs and indels) not haplotype entries and had multiple submitters in agreement on the clinical classification (2 stars or higher) were downloaded from ClinVar in September 2024 and annotated using gnomAD allele frequencies. In the same way as above, variants found in both individuals of European-like and non-European-like genetic ancestry were removed to examine the counts of different variants found exclusively in the European-like versus non-European-like genetic ancestry groups.

Bonferroni corrections

To counteract the potential for type I errors due to multiple comparisons, we apply a stringent Bonferroni correction to each statistical test. For testing the difference in allele prevalence of all coding variants of a particular clinical significance type across specialties, there are 14 specialties × 3 databases × 5 clinical significance groups = 210 total tests. For testing the difference in allele prevalence of all coding variants without missense variants of a particular clinical significance type across specialties, there are also 14 specialties × 3 databases × 3 clinical significance groups = 126 total tests. For testing the difference in allele prevalence of variant types for different clinical classifications for the GenCC curated genes list, there are 11 variant types × 3 databases × 5 clinical significance categories = 165 statistical tests. For testing the difference in allele prevalence of coding variants of a particular clinical significance type across population distributions for the GenCC curated genes list, there are 1 specialty × 3 databases × 5 clinical significance groups × 5 pairwise population comparisons = 75 total tests. For testing the difference in allele prevalence of noncoding variants of a particular clinical significance type across specialties, there are 14 specialties × 2 databases × 5 clinical significance groups = 140 total tests. For testing the difference in unique variants of a particular clinical significance found only in one population group via chi-square testing, there are 3 databases × 5 clinical significance groups = 15 total tests. Of particular note, because the three research questions are independent of each other (e.g., no nested hypotheses, no repeated measures, no sequential testing) and the underlying data distributions for each statistical test are very different for the three questions, each group of tests received its own Bonferroni correction.

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 [18, 24, 37–43]. Our pipeline follows the gene-specific criteria of the corresponding VCEP (*TP53* v1, *BRCA1* v1, *PTEN* v2) as closely as possible except for the functional data evidence code (PS3/BS3) where MAVE data was used. Initially, each variant was annotated using the 2015 ACMG (American College of Medical Genetics and Genomics) evidence codes through the Intervar API. During this process, we ensured that the correct reference genomes were used for the different databases (*All of Us* and gnomAD v3.1.2 utilized GRCh38; whereas gnomAD 2.1.1 utilized GRCh37). Following this initial annotation, each variant was further annotated with functional scores from MAVE data. The clinical curation and clinical strength assignment as per the ClinGen recommendations in Brnich et al. [44] for or against pathogenicity or benignity of each of these MAVE datasets utilized in this study were previously published in Fayer et al. [18]. In brief, for *BRCA1* variants, if a variant was categorized as FUNC (functional), it was assigned BS3 evidence and no PS3 evidence, whereas if it was categorized as LOF (loss of function), the variant was assigned PS3 evidence and no BS3 evidence. Variants categorized as INT (intermediate) were left unannotated. For the BRCA1 combining criteria, ≥ 1 criteria of strong benign evidence was enough to reclassify the VUS as Likely Benign. For *TP53*, we used the output of the Naïve Bayes classifier that synthesized data from four different *TP53* MAVEs in Fayer et al. If the

classifier predicted a variant to be "Functionally abnormal," the variant was assigned PS3 evidence and no BS3 evidence. If a variant was predicted to be "Functionally normal," BS3_moderate evidence was used with no PS3 evidence. For PTEN, two assays measuring activity and abundance were used. If the abundance was categorized as "wt-like" or "possibly wt-like," BS3_Supporting evidence was used. Furthermore, if the cumulative score was less than or equal to − 1.11, BS3_moderate evidence was used. All other evidence codes and combining criteria were adhered to as closely as possible based on the ClinGen gene-specific recommendations for *BRCA1*, *TP53*, and *PTEN*, respectively (Additional file 2: Fig. S42). The ClinGen VCEPs are highly regarded as the gold standard for gene-specific variant curation and are developed after extensive evaluation of the evidence by clinical and scientific experts for the particular gene to classify genomic variants on a spectrum from pathogenic to benign using the 2015 ACMG/AMP Variant Interpretation Guidelines as a backbone [43]. Reclassification of variants from gnomAD or *All of Us* focused only on variants originally classified as VUS.

We comprehensively reanalyzed the set of *BRCA1*, *PTEN*, and *TP53* VUS previously reclassified by Fayer et al. [18] (Supplemental Tables 7, 10, 11 in Fayer et al.) to benchmark our automated pipeline. The automated pipeline uses VCEP recommendations as of Fall 2023; however, the Fayer et al. VUS dataset was analyzed by hand with a mix of VCEP and ACMG/AMP 2015 recommendations prior to 2021. Using this dataset, we sought to establish a robust benchmark for the automated variant classification pipeline built for this project to ensure clinical variant classifications ascertained by the automated pipeline were concordant with the Fayer et al. reclassifications where MAVE data was also used for variant classification. We defined a concordant classification as a final clinical classification on the same side of pathogenicity as was found in the Fayer et al. dataset (the groups being Benign or Likely Benign versus Pathogenic or Likely Pathogenic versus remaining a VUS). Further, we used this dataset to follow-up on the essential code analysis with allele frequencies from gnomAD v4. We annotated all possible variants in the Fayer dataset with allele frequencies from gnomAD v4 (not using *All of Us* v7 nor gnomAD v3 nor gnomAD v2 to prevent accidental double-dipping).

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. To ensure reproducibility, transparency, and increased throughput, all the procedures for annotating variants and assigning evidence codes were codified using Python. All code has been made freely available and is linked in the "Availability of data and materials" Sect [45].

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 [9, 25, 26]. Even though other population databases may also contain a large number of individuals of non-European-like genetic ancestry, we chose these three population-scale databases, because the number of individuals sequenced in each was similar for both superpopulation groupings allowing for fair downstream statistical analyses predicated on allele counts (Additional file 2: 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)).

Overall, there are an average of 29.8 ClinVar VUS per individual of non-European-like genetic ancestry versus 24.3 ClinVar VUS per individual of European-like genetic ancestry (Table 1) across all curated clinical genes (GenCC) in all three databases. Further, individuals with non-European-like genetic ancestry have an average of 4.0 P/LP, 8232 B/LB, and 126.2 CI variants, and individuals of European-like genetic ancestry average 4.3 P/LP, 8016 B/LB, and 122.4 CI variants (Table 1).

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

Table 1 Average number of ClinVar alleles per individual in all curated clinical genes (GenCC)

	Non-European-like	European-like
Variant of Uncertain Significance	29.8	24.3
Pathogenic or Likely Pathogenic	4.0	4.3
Benign or Likely Benign	8232	8016
Conflicting Interpretation	126.2	122.4

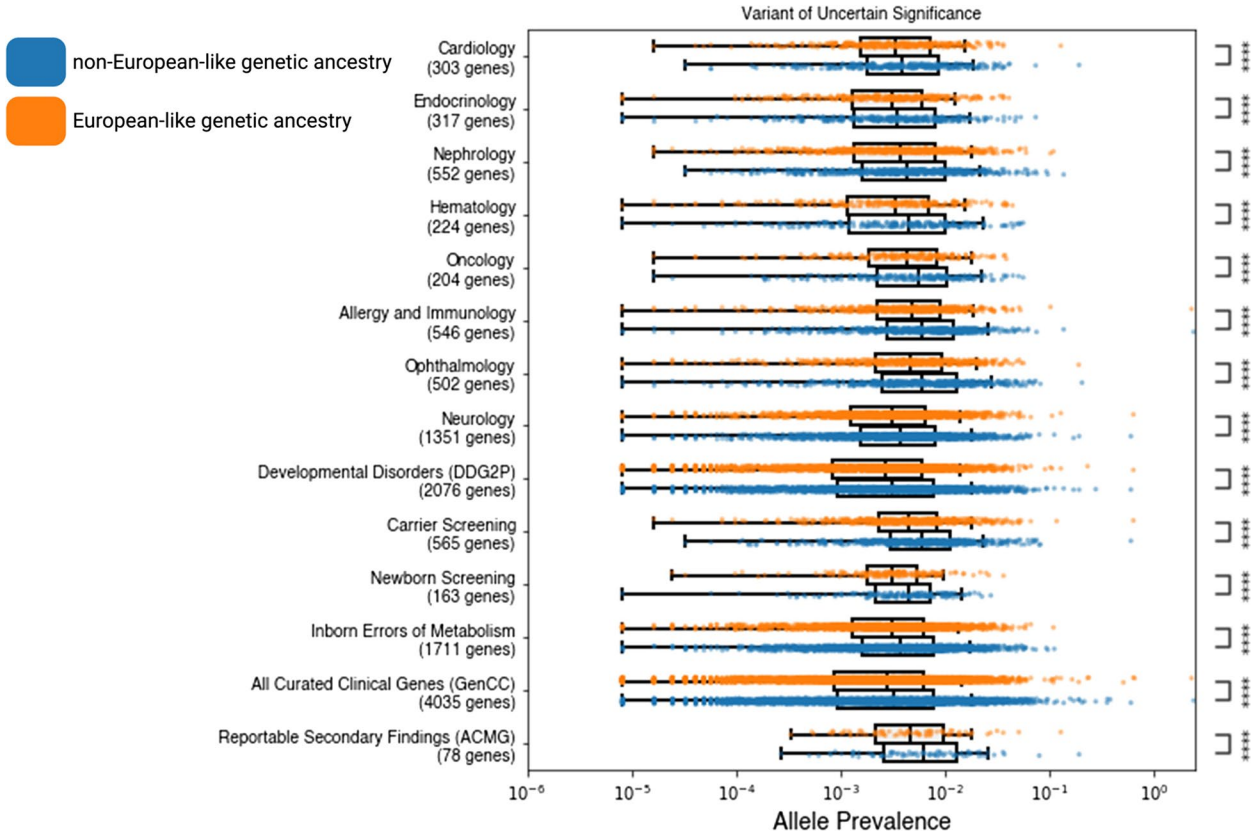

Fig. 2 Higher VUS prevalence found in individuals of non-European-like genetic ancestry across medical specialties. Box plots corresponding to VUS allele prevalence (x-axis) in each gene (dot) for individuals of non-European-like (blue) versus European-like (orange) genetic ancestry for the corresponding 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 $1.19e-04 < p \leq 2.38e-04$, ** for $5.95e-05 < p \leq 1.19e-04$, *** for $5.95e-06 < p \leq 5.95e-05$, and **** for $p \leq 5.95e-06$. 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

1.52e−211 to 1.4e−07; effect sizes ranging 0.35 to 0.76; Fig. 2, Additional file 1: Tables S2–4, Additional file 2: Fig. S2). 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; Additional file 1: Tables S2–4, Additional file 2: Fig. S3). Further, a significantly higher prevalence of B/LB and variants with **n**o clinical **d**esignation (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; Additional file 1: Tables S2–4, Additional file 2: Figs. S4–5), while only isolated significant differences that did not validate across all three databases were seen for Conflicting Interpretation (CI) or noncoding variants (Additional file 1: Tables S2–6, Additional file 2: Figs. S6–11).

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 (Additional file 1: Tables S7–9, Additional file 2: Figs. S12–17). Over 84% of VUS across each medical specialty for all three databases were missense variants (Fig. 3a, Additional file 2: Fig. S18). 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. 3b, Additional file 1: Tables S10–12, Additional file 2: Fig. S19), 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. 3c, Additional file 1: Tables S13–15,

Additional file 2: Fig. S22). All four of these variant types, missense, in-frame indels, splice region, and synonymous variants can be systematically cataloged 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. 4a–d, Additional file 1: Tables S13–15, Additional file 2: Figs. S29–30), while pathogenic variants were the sole clinical classification where the European-like genetic ancestry group showed significantly higher counts ($p = 1.05e-05$, Fig. 4e, Additional file 1: Tables S13–15, Additional file 2: Figs. S29–30). These trends of higher VUS and B/LB counts being found in individuals of non-European-like genetic ancestry versus higher counts of P/LP variants being found in individuals of European-like genetic ancestry are also corroborated when orthogonally examining all the "≥2 star" high confidence variants in ClinVar where the clinical classification is agreed upon by multiple independent submitters (Additional file 2: Fig. S31).

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 1e − 300 to 6.215e − 18, Additional file 2: Figs. S32–37). 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 ($p \leq 1.67e-05$; Additional file 2: Figs. S38–42).

Greater diversity of unique coding variants in individuals of non-European-like genetic ancestry at baseline

Our findings align with previous research, underscoring the greater diversity of unique coding variants present in non-European-like individuals when compared to an equivalent sized sample of individuals of European-like genetic ancestry. This observation is supported on a gene-by-gene basis by the significant increased allele prevalence in both B/LB variants and ND variants among non-European-like individuals when compared to Europeans for both coding and noncoding variants (Additional file 2: Figs. S4, S5, S9, S11). Moreover, using the orthogonal statistical method that focuses on comparing unique variants between individuals of

(See figure on next page.)

Fig. 3 disparity in VUS prevalence is present even in the absence of missense variants. **a** Pie charts representing the variant spectrum of VUS for all genes within the particular medical specialty in gnomAD v3.1.2. The most prevalent VUS variant type, missense variants (light blue), accounts for at minimum 84% of VUS in any given specialty across all three databases. **b** 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 is 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 [36] 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. **c** Box plots corresponding to VUS allele prevalence (x-axis) in genes (dots) for individuals of non-European-like (blue) versus European-like (orange) genetic ancestry for the corresponding variant type (y-axis) across gnomAD v3.1.2 (non v2) for all coding variants in the set of curated clinical genes (GenCC). The total number of alleles from individuals of non-European-like (right) versus European-like (left) genetic ancestry is indicated under each variant type in parentheses. Genes (y-axis) with zero alleles for the corresponding variant type for allele prevalence for either individuals of European-like or non-European-like genetic ancestry are omitted from the 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 $1.52e-04 < p \leq 3.03e-04$, ** for $7.58e-05 < p \leq 1.52e-04$, *** for $7.58e-06 < p \leq 7.58e-05$, and **** for $p \leq 7.58e-06$. Also refer to Additional file 1: Tables S13–15. Overall, we observe a statistically significant increase in VUS in individuals of non-European-like genetic ancestry compared to individuals of European-like genetic ancestry for missense, synonymous, splice region, and inframe variants

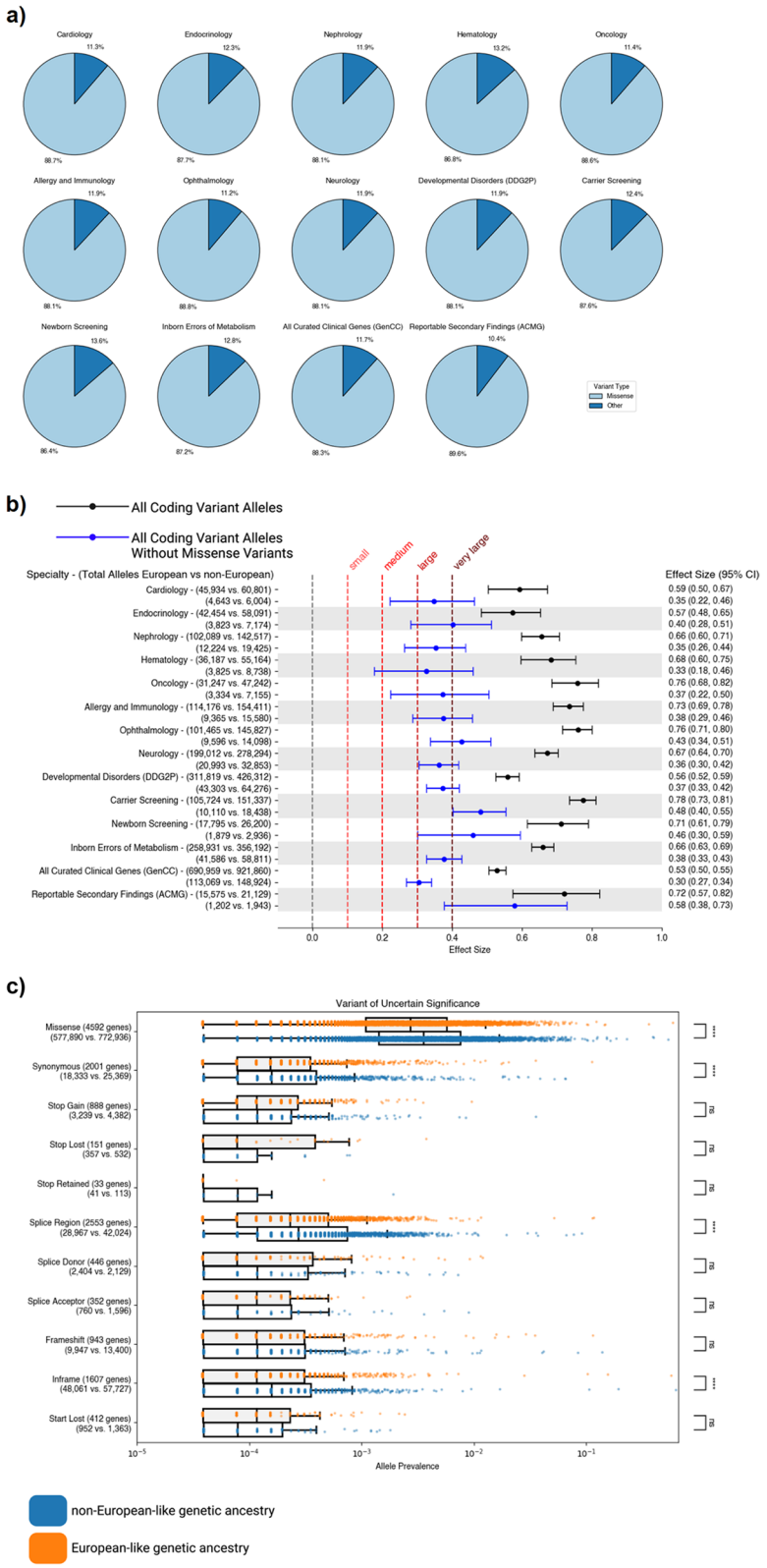

Fig. 3 (See legend on previous page.)

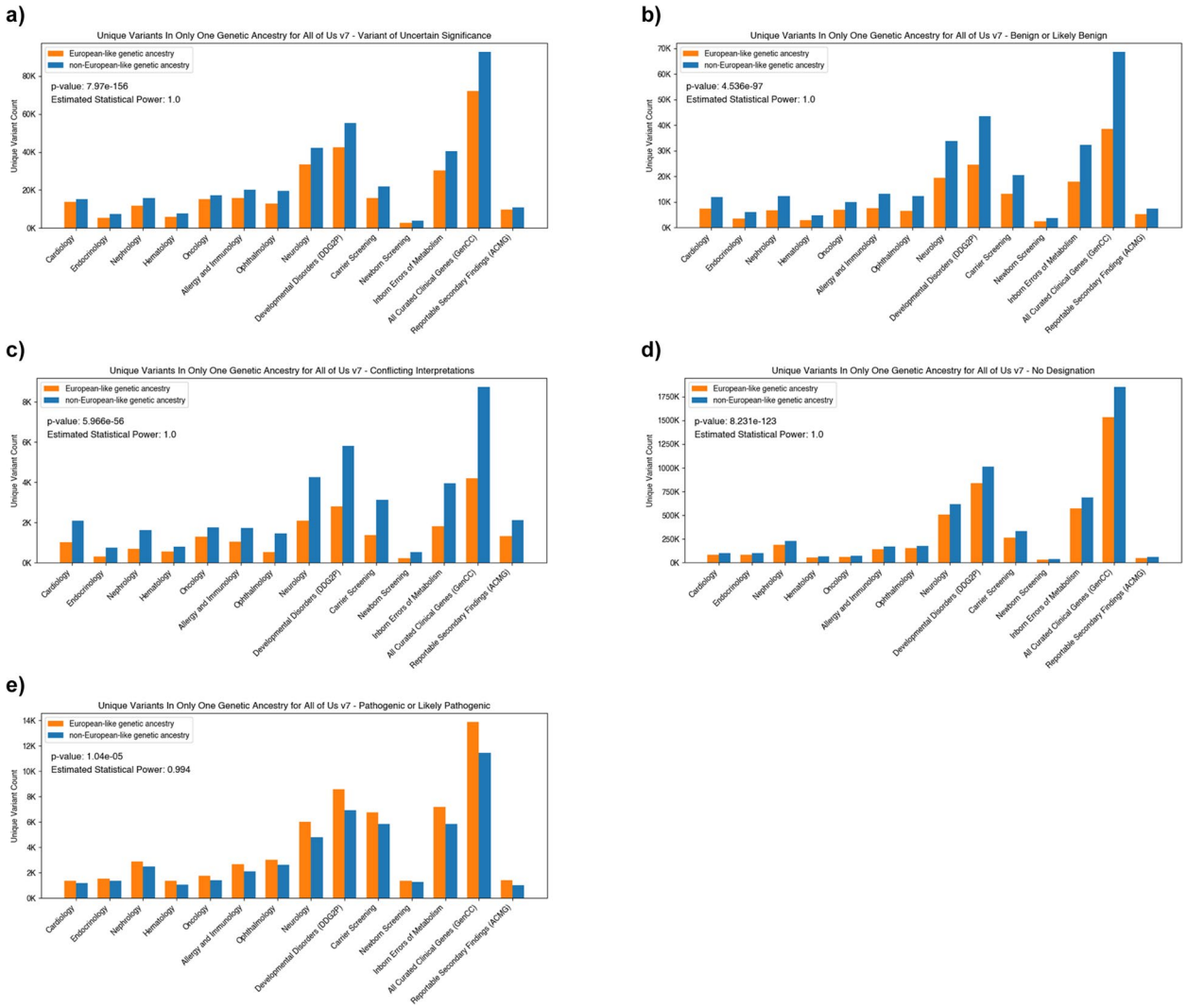

Fig. 4 Comparison of counts of unique variants found in only one genetic ancestry group. Grouped bar graphs corresponding to unique coding variant counts (y-axis) for **a** VUS, **b** B/LB, **c** CI, **d** ND, and **e** P/LP variants found either only in individuals of European-like (orange) genetic 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. Also refer to Additional file 1: Tables S2–4. Across all medical specialties and categories shown, VUS, B/LB, CI, and ND variants were found at a statistically significantly higher prevalence in individuals of non-European-like genetic ancestry. Conversely P/LP variants were found at a statistically significantly higher prevalence in individuals of European-like genetic ancestry

non-European-like versus European-like genetic ancestry, our study consistently reveals a significantly greater count of B/LB and ND unique variants in individuals of non-European-like genetic ancestry (Fig. 4, Additional file 1: Tables S13–15, Additional file 2: Figs. S30–31). Examining each of the five genetic ancestries (African/African-American, Latino/Admixed American, South Asian, East Asian, and Other) in pairwise comparisons with the European-like genetic ancestry group, each of these genetic ancestries displays a significant increased prevalence of variants with no designation, and several also show elevated prevalence of B/LB variants (Additional file 2: Figs. S40, S42). This trend is reinforced when examining the data by variant types. For non-designated (ND) variants, all coding and noncoding variant types exhibit significant increases in allele prevalence among non-European-like genetic ancestries, while several variant types also demonstrate heightened prevalence in non-European-like populations for benign variants (Additional file 2: Figs. S25, S27). These findings collectively establish a baseline depiction of the greater

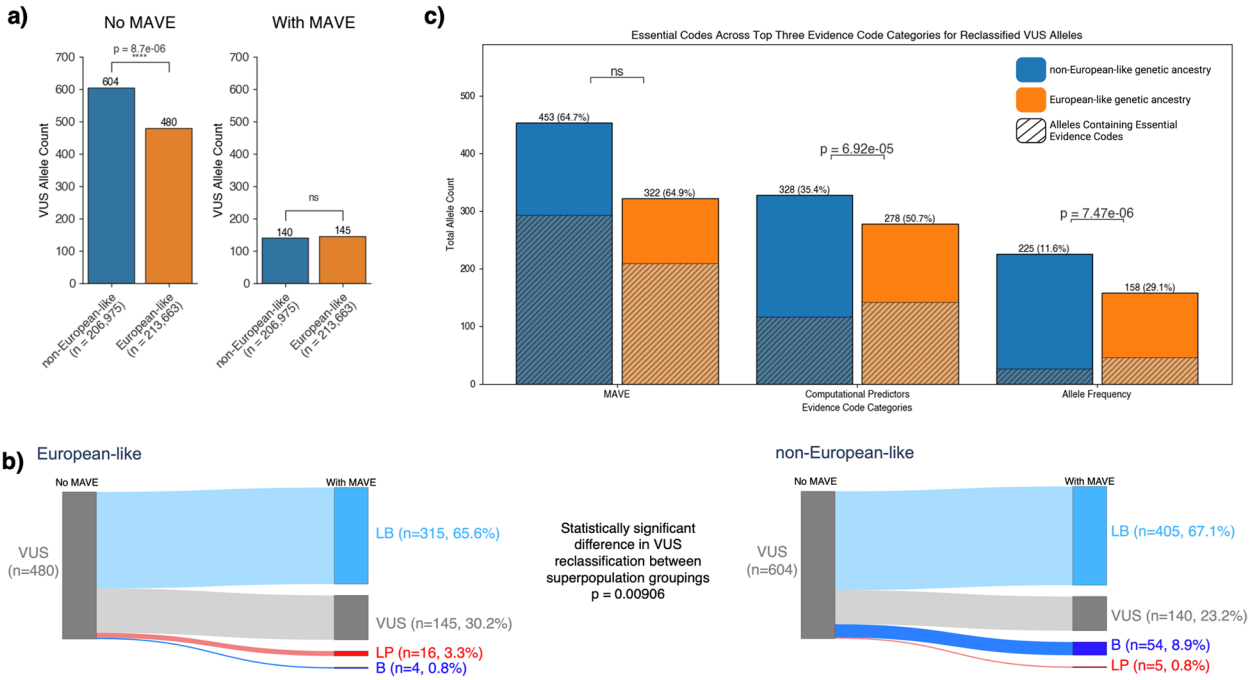

Fig. 5 MAVE data can reclassify non-European-like VUS at a statistically significant higher rate compared to European-like VUS. **a** The presence of VUS in individuals of non-European-like versus European-like genetic ancestry was statistically significantly higher in non-European-like superpopulation group. However, after using MAVE data for reclassification in the ClinGen VCEP frameworks, there was no statistically significant VUS disparity detected. **b** 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. **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, and BS3_Supporting corresponded to the MAVE evidence codes. BA1, BS1, and BS1_Supporting correspond to the allele frequency codes. The aggregate analysis for essential codes for the computational predictors is reflective of the cumulative contribution of several commonly used predictors as prescribed by the respective ClinGen VCEP (*BRCA1* relies on BayesDel no-AF, *TP53* relies on both aGVGD and BayesDel, and *PTEN* relies on REVEL)

diversity of unique coding variants among the non-European-like superpopulation compared to European-like.

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 (Additional file 2: Fig. S43). Thus, to validate the accuracy of our variant reclassifications, we benchmarked our pipeline against the Fayer et al. [18] 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. (Additional file 1: 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, Additional file 1: 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, Additional file 1: Table S18, Additional file 2: Fig. S44). 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, Additional file 1: Table S18, Additional file 2: Fig. S45). The statistically significant difference in reclassification rates ($p = 9.06e-03$; one-tail z proportions test; Fig. 5b, Additional file 1: 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

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, Additional file 1: Table S18, Additional file 2: Figs. S46–47). 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, Additional file 1: Table S18). We validated this finding in the Fayer et al. [18] 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; Additional file 1: Table S19, Additional file 2: Figs. S46–47). 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% [46] and 7.3% [12] with clinical evidence codes being most important for VUS resolution [12]. Our study incorporated MAVE data and achieved a cumulative VUS reclassification rate of 73.7% without clinical evidence codes. Clinical evidence codes drive the distinction between variant classification and interpretation, where classifications utilizes available public data, but interpretation involves a comprehensive evaluation of a variant in the context of an individual's unique genotypes and phenotypes. We hypothesize our VUS reclassification rate would have been even higher if clinical evidence codes were available in this study. Nonetheless, our VUS reclassification rate is similar to other single gene MAVE studies: 50% in *BRCA1* [18], 69% in *TP53*, 75% in *MSH2* [22], and 93% in *DDX3X* [23]. 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 variation [47]. 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 (Additional file 2: 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. Due to the more comprehensive picture of human genetic diversity represented by the non-European superpopulation, including population-specific mutations, 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 hindrance 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. Here, based on the ClinGen VCEP rules, akin to gold standard curation rules in the field, the computational predictors used for VUS reclassification in *BRCA1* rely on BayesDel no-AF, *TP53* rely on both aGVGD and BayesDel, and for *PTEN* rely on REVEL. Thus, the aggregate analysis we do is reflective of the cumulative contribution of these very commonly used predictors. When computational predictors are trained and tested against excerpts of current sequencing and clinical variant databases [48, 49], 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 potentially perpetuated forward as AI bias preventing equitable impact for VUS reclassification and contributing to VUS disparity as seen in this study. Likely hundreds of thousands of individuals of non-European-like genetic ancestry have had inequitable variant interpretations due to this bias in computational predictors. In the future, a systematic analysis should be undertaken to understand the potential bias of a variety of commonly used computational predictors individually [50, 51]. 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.

Current variant interpretation standards, focused on coding variants, still require expansion and refinement. For well-understood variant types such as stop gains, frameshifts, and canonical splice variants, our existing knowledge base is substantial enough that we do not observe a significant disparity in VUS classification between the non-Europeans-like and European-like groups. However, when classifying challenging synonymous, inframe indels, splice region, and missense variants, our current interpretation of coding variants falls short in preventing VUS disparities between population groups. This gap in knowledge could be potentially addressed by MAVEs which are able to systematically ascertain a functional effect for each of these coding variant types.

Commensurate with understanding which variant types contribute to these disparities is the importance of distinguishing our ability to classify variants that cause gain of function (GoF) versus loss of function (LoF). Our advanced understanding of LoF mechanisms, such as nonsense-mediated decay (NMD), NMD-escape [52, 53], nonstop decay, and more, make LoF variants easier to classify, while GoF variants remain less well understood. While MAVEs will enhance our ability to identify GoF variants, bridging the understanding gap to the level of LoF variants may still require more extensive mechanistic research. In the future, a study should examine whether LoF variants are more effectively classified than GoF variants and what disparities yield from the lack of mechanistic understanding of GoF on variant classification.

Conclusions

Calls for diversifying genomics have yielded a pangenome reference [54], H3Africa to equip Africa with genomics infrastructure [55], and diverse participant recruitment in *All of Us* [56]. Diversifying genomics via recruitment, engagement, and retention is just one approach to pursuing equity [57]. 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.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-024-01392-7.

Additional file 1: Supplemental Tables S1–S19.

Additional file 2: Supplemental results on noncoding variants and effect sizes; supplemental figures and legends for Figs. S1–46.

Acknowledgements
We thank Ryan D. Hernandez and Pradiptajati Kusuma for comments on the manuscript.

Authors' contributions
Conceptualization: SF, LMS, DF, MD, WCM, IGR; data curation: MD, MP, DK, KP; formal analysis: MD; funding acquisition: RAG, LMS, IGR; methodology: MD, SF, RAG, LMS, CDRE, WCM, IGR; project administration: MD, LAM; resources: MD, EV, RAG, LMS, IGR; software: MD; supervision: RAG, LMS, IGR; visualization: MD, SF, SP; writing—original draft: MD, IGR; writing—review and editing: MD, SF, SP, DMF, AFR, JEP, SEP, JRL, RAG, LMS, CDRE, WCM, IGR. All authors read and approved the final manuscript.

Funding
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.
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.

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 [46]. Additionally, all input data derived from both gnomAD v2.1.1 and gnomAD v3.1.2 (non v2) is linked at the GitHub above. For the *All of Us* analysis, 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. All variant data is publicly accessible as it has been released in a deidentified manner through either the gnomAD website [28] or the *All of Us* Public Data Browser v7 [27]. Complete rankings across all three databases of all clinically curated genes by VUS/PorLP/BorLB/CI/ND allele prevalence difference are also linked to the GitHub. All reclassified variants with evidence codes used can be found in Additional file 1: Tables S18–19. The ClinVar acession IDs associated with this paper are SCV005402472 to SCV005402658 and SCV005402681 to SCV005402746, and are available at https://www.ncbi.nlm.nih.gov/clinvar/?term=SUB14864172 and https://www.ncbi.nlm.nih.gov/clinvar/?term=SUB14788601.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
JRL has stock ownership in 23andMe, is a paid consultant for Regeneron Genetics Center, and is a coinventor on multiple US 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. The remaining authors declare that they do not have any competing interests.

Author details
[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. [19]Mary MacKillop Institute for Health Research, Australian Catholic University, Fitzroy, Australia.

Received: 20 May 2024 Accepted: 3 October 2024
Published online: 03 December 2024

References

1. Mata DA, Rotenstein LS, Ramos MA, Jena AB. Disparities according to genetic ancestry in the use of precision oncology assays. N Engl J Med. 2023;388:281–3.
2. Fatumo S, et al. A roadmap to increase diversity in genomic studies. Nat Med. 2022;28:243–50.
3. Borrell LN, et al. Race and genetic ancestry in medicine — a time for reckoning with racism. N Engl J Med. 2021;384:474–80.
4. Martin AR, et al. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51:584–91.
5. Collins FS, Doudna JA, Lander ES, Rotimi CN. Human molecular genetics and genomics — important advances and exciting possibilities. N Engl J Med. 2021;384:1–4.
6. Matalon DR, 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). Genet Med. 2023;25: 100812.
7. Manrai AK, et al. Genetic misdiagnoses and the potential for health disparities. N Engl J Med. 2016;375:655–65.
8. Cook S, 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: 101055.
9. Venner E, et al. The frequency of pathogenic variation in the all of us cohort reveals ancestry-driven disparities. Commun Biol. 2024;7:1–11.
10. Wright CF, et al. Genomic diagnosis of rare pediatric disease in the United Kingdom and Ireland. N Engl J Med. 2023;388:1559–71.
11. Abul-Husn NS, et al. Molecular diagnostic yield of genome sequencing versus targeted gene panel testing in racially and ethnically diverse pediatric patients. Genet Med. 2023;25: 100880.
12. Chen E, et al. Rates and classification of variants of uncertain significance in hereditary disease genetic testing. JAMA Netw Open. 2023;6: e2339571.
13. Giri, V. N., Hartman, R., Pritzlaff, M., Horton, C., Keith, S. W. Germline variant spectrum among African American men undergoing prostate cancer germline testing: need for equity in genetic testing. JCO Precis Oncol 2022:e2200234.https://doi.org/10.1200/PO.22.00234
14. Caswell-Jin JL, et al. Racial/ethnic differences in multiple-gene sequencing results for hereditary cancer risk. Genet Med. 2018;20:234–9.
15. Tatineni S, et al. Racial and ethnic variation in multigene panel testing in a cohort of BRCA1/2-negative individuals who had genetic testing in a large urban comprehensive cancer center. Cancer Med. 2022;11:1465–73.
16. Rehm HL, et al. The landscape of reported VUS in multi-gene panel and genomic testing: time for a change. Genet Med. 2023;25: 100947.
17. Horton C, et al. Diagnostic outcomes of concurrent DNA and RNA sequencing in individuals undergoing hereditary cancer testing. JAMA Oncol. 2023. https://doi.org/10.1001/jamaoncol.2023.5586.
18. Fayer S, 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;108:2248–58.
19. Fowler DM, et al. An Atlas of Variant Effects to understand the genome at nucleotide resolution. Genome Biol. 2023;24:147.
20. Fowler DM, et al. High-resolution mapping of protein sequence-function relationships. Nat Methods. 2010;7:741–6.
21. Macdonald CB, et al. DIMPLE: deep insertion, deletion, and missense mutation libraries for exploring protein variation in evolution, disease, and biology. Genome Biol. 2023;24:36.
22. Scott A, et al. Saturation-scale functional evidence supports clinical variant interpretation in Lynch syndrome. Genome Biol. 2022;23:266.
23. Radford EJ, et al. Saturation genome editing of DDX3X clarifies pathogenicity of germline and somatic variation. Nat Commun. 2023;14:7702.
24. Findlay GM, et al. Accurate classification of BRCA1 variants with saturation genome editing. Nature. 2018;562:217–22.
25. Karczewski KJ, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.
26. Chen S, et al. A genomic mutational constraint map using variation in 76,156 human genomes. Nature. 2024;625:92–100.
27. SNV/indel variants | All of Us Public Data Browser. https://databrowser.researchallofus.org/variants.
28. Gudmundsson S, et al. Variant interpretation using population databases: lessons from gnomAD. Hum Mutat. 2022;43:1012–30.
29. Landrum MJ, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46:D1062–7.
30. ClinVar. https://www.ncbi.nlm.nih.gov/clinvar/.
31. Coop, G. Genetic similarity versus genetic ancestry groups as sample descriptors in human genetics. Preprint at http://arxiv.org/abs/2207.11595 (2023).
32. The GenCC Home Page. https://thegencc.org/.
33. Kuang D, et al. MaveRegistry: a collaboration platform for multiplexed assays of variant effect. Bioinformatics. 2021;37:3382–3.
34. Harrison SM, et al. Harmonizing variant classification for return of results in the All of Us Research Program. Hum Mutat. 2022;43:1114–21.
35. Representation of classifications in ClinVar. https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/.
36. Funder DC, Ozer DJ. Evaluating effect size in psychological research: sense and nonsense. Adv Methods Pract Psycholog Sci. 2019;2:156–68.
37. Giacomelli AO, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. Nat Genet. 2018;50:1381–7.
38. Matreyek KA, et al. Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat Genet. 2018;50:874–82.
39. Mighell TL, Evans-Dutson S, O'Roak BJ. A saturation mutagenesis approach to understanding PTEN lipid phosphatase activity and genotype-phenotype relationships. American J Human Genet. 2018;102:943–55.
40. Parsons, M. T. et al. Evidence-based recommendations for gene-specific ACMG/AMP variant classification from the ClinGen ENIGMA BRCA1 and BRCA2 Variant Curation Expert Panel. 2024.01.22.24301588 Preprint at 2024. https://doi.org/10.1101/2024.01.22.24301588.
41. Fortuno C, et al. Specifications of the ACMG/AMP variant interpretation guidelines for germline TP53 variants. Hum Mutat. 2021;42:223–36.
42. Mester JL, et al. Gene-specific criteria for PTEN variant curation: recommendations from the ClinGen PTEN Expert Panel. Hum Mutat. 2018;39:1581–92.
43. Rehm HL, et al. ClinGen — the clinical genome resource. N Engl J Med. 2015;372:2235–42.
44. Brnich SE, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. 2019;12:3.
45. Dawood, M. MoezDawood/ReducingVariantClassificationInequities: v2. Zenodo. 2024. https://doi.org/10.5281/ZENODO.13777870.
46. Slavin TP, et al. Prospective study of cancer genetic variants: variation in rate of reclassification by ancestry. JNCI. 2018;110:1059–66.
47. Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research *et al. Using population descriptors in genetics and genomics research: a new framework for an evolving field*. 26902 (National Academies Press, Washington, D.C., 2023). https://doi.org/10.17226/26902.
48. Ioannidis NM, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99:877–85.
49. Feng B-J. PERCH: a unified framework for disease gene prioritization. Hum Mutat. 2017;38:243–51.
50. Pathak, A. K. et al. Pervasive ancestry bias in variant effect predictors. 2024.05.20.594987 Preprint at 2024. https://doi.org/10.1101/2024.05.20.594987.
51. Rastogi, R. et al. Critical assessment of missense variant effect predictors on disease-relevant variant data. 2024.06.06.597828 Preprint at 2024, https://doi.org/10.1101/2024.06.06.597828.
52. Coban-Akdemir Z, et al. Identifying genes whose mutant transcripts cause dominant disease traits by potential gain-of-function alleles. Am J Hum Genet. 2018;103:171–87.
53. Lindeboom RGH, Supek F, Lehner B. The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat Genet. 2016;48:1112–8.
54. Liao W-W, et al. A draft human pangenome reference. Nature. 2023;617:312–24.
55. Choudhury A, et al. High-depth African genomes inform human migration and health. Nature. 2020;586:741–8.
56. The "All of Us" Research Program. N England J Med. 2019;381:668–676. https://www.nejm.org/doi/full/10.1056/NEJMsr1809937.
57. Lee SS-J, Appelbaum PS, Chung WK. Challenges and potential solutions to health disparities in genomic medicine. Cell. 2022;185:2007–10.

Publisher's Note