

ARTICLE

Received 17 Dec 2015 | Accepted 12 Jul 2016 | Published 11 Oct 2016

DOI: 10.1038/ncomms12521

OPEN

Challenges and disparities in the application of personalized genomic medicine to populations with African ancestry

Michael D. Kessler¹, Laura Yerges-Armstrong^{2,3}, Margaret A. Taub⁴, Amol C. Shetty¹, Kristin Maloney³, Linda Jo Bone Jeng³, Ingo Ruczinski⁴, Albert M. Levin⁵, L. Keoki Williams^{6,7}, Terri H. Beaty⁸, Rasika A. Mathias^{8,9}, Kathleen C. Barnes^{8,9,10}, Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA)[†] & Timothy D. O'Connor^{1,2,3}

To characterize the extent and impact of ancestry-related biases in precision genomic medicine, we use 642 whole-genome sequences from the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) project to evaluate typical filters and databases. We find significant correlations between estimated African ancestry proportions and the number of variants per individual in all variant classification sets but one. The source of these correlations is highlighted in more detail by looking at the interaction between filtering criteria and the ClinVar and Human Gene Mutation databases. ClinVar's correlation, representing African ancestry-related bias, has changed over time amidst monthly updates, with the most extreme switch happening between March and April of 2014 ($r = 0.733$ to $r = -0.683$). We identify 68 SNPs as the major drivers of this change in correlation. As long as ancestry-related bias when using these clinical databases is minimally recognized, the genetics community will face challenges with implementation, interpretation and cost-effectiveness when treating minority populations.

¹Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. ²Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. ³Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. ⁴Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21287, USA. ⁵Department of Public Health Sciences, Henry Ford Health System, Detroit, Michigan 48202, USA. ⁶Center for Health Policy & Health Services Research, Henry Ford Health System, Detroit, Michigan 48202, USA. ⁷Department of Internal Medicine, Henry Ford Health System, Detroit, Michigan 48202, USA. ⁸Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA. ⁹Department of Medicine, Johns Hopkins University, Baltimore, Maryland 21224, USA. ¹⁰Department of Medicine, University of Colorado, Aurora, Colorado 80045, USA. Correspondence and requests for materials should be addressed to T.D.O. (email: timothydoconnor@gmail.com).

[†]A full list of consortium members appears at the end of the paper.

The idiom ‘searching for a needle in a haystack’ is frequently used in genomics, and is especially apt for describing the search for causal alleles in patients with non-canonical diseases of likely genetic origin. As a field, we tend to be singularly focused on the needle and forget that the complexity of the haystack is actually a highly rate-limiting step of this search. The motivation of this project is to characterize the complex interaction between variant prioritization and ancestry, often believed to be largely affected by the predominance of European-based data within clinical databases^{1–3}, to better understand the application of clinical genomics to minority populations. Any ancestry-related biases that exist when using typical filters and databases to implement variant prioritization and other similar precision genomic medicine techniques can have profound confounding effects, as most methodological biases do. Therefore, here we quantify the extent of ancestry-related biases inherent to approaches and databases typically used for precision genomic medicine, and we present how such biases have changed over time. We also show how these biases translate to the level of the individual and their proportion of African ancestry, with implications for diagnostic accuracy and cost.

To explore the role ancestry plays in variant prioritization approaches often implemented in genomic medicine, we utilize whole-genome sequencing data from 642 study individuals in the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA). The CAAPA project represents a diverse group of admixed individuals of African descent with no suspected Mendelian conditions. It has been shown that there is a strong correlation between the overall number of variants found per individual and African ancestry^{4–7}. Furthermore, significant differences exist between populations in the number of variants per individual considered disease causing by the two popular clinical databases, Human Gene Mutation Database (HGMD) and ClinVar⁶. On the basis of annotations from HGMD, individuals with predominantly African ancestry have by far the most variants considered disease causing, whereas variants prioritized as disease causing based on annotations from ClinVar are most abundant in individuals with predominantly European ancestry and are of intermediate to below-average abundance in predominantly African-ancestry individuals⁶. These population-based discrepancies reflect differences between databases, and suggest that the interplay between database and sample ancestry is important. The CAAPA cohort utilized here serves as an appropriate sample, with representative quantities of variation (that is, similar-sized haystacks), for evaluating whether biases exist when applying precision genomic medicine to African-ancestry individuals. Any biases and/or population specificities for African-ancestry patients that inflate the number of prioritized variants (that is, make the haystack bigger), would result in increased effort (that is, time and money) to identify a causative variant (that is, find the needle) in African-ancestry patients.

Results

Variant classification. We initially classified variants into two main groups, with pathogenic annotated variants (PAVs) comprising those identified as disease-causing in the Online Mendelian Inheritance in Man (OMIM)⁸, HGMD⁹ or ClinVar¹⁰ databases, and non-annotated variants (NAVs) consisting of those not annotated as disease-causing in these databases. Unless otherwise noted, we used an allele frequency filter, and excluded common variants with a minor allele frequency (MAF) >5% from our analyses (Methods). Each category was then sub-classified as deleterious or non-deleterious based on computational predictions (Fig. 1a)^{11–20}, which we consider as a

type of filter based on deleteriousness and note when used for categorization. Since there is evidence for all PAVs (deleterious and non-deleterious) and deleterious NAVs to be further evaluated as higher priority, variants in these categories often require time-consuming and costly follow-up review by a clinical team^{1,21,22} to identify causative variants with a low false-negative rate.

Correlations with African ancestry and variant counts. We find significant correlations between estimated African ancestry (Supplementary Fig. 1) and the number of variants per individual in all variant sets except deleterious PAVs (Fig. 1b–d). Both deleterious and non-deleterious NAVs show similar levels of correlation with African ancestry as does all genomic variation pooled together⁷. When we remove the aforementioned MAF and deleteriousness filters, as well as a filter on stop/splice sites, and identify PAVs from either HGMD or ClinVar databases separately, we find a strong positive correlation between estimated African ancestry and variants identified in HGMD ($r = 0.992$, $P = 6.12 \times 10^{-14}$) and a modest positive correlation between African ancestry and variants in ClinVar ($r = 0.539$, $P = 0.031$). The correlation becomes less positive or even negative (Supplementary Table 1) as we re-add our two main filters: (1) inclusion of variants with MAF <5% (MAF filter); and (2) inclusion of variants called deleterious by at least 2 of 11 *in silico* predictions (deleterious filter).

One possible explanation for this general reduction of the positive correlation with ancestry is that these filters effectively remove functionally neutral variants, of which there are more in persons of African ancestry. Assuming this, one would predict a reduction in the positive correlation with African ancestry, as long as the filters remove a higher number of functionally neutral variants, relative to causative variants, from African populations compared with European populations. Given recent studies showing that African populations have more genetic variation than European populations^{5,23,24}, but that the number of deleterious alleles in an individual is independent of demography or lower in Africans, depending on the level of deleteriousness of these alleles^{25–29}, one would expect all filters to remove higher numbers of non-causal variants from individuals with greater African ancestry, as is consistent with what we report here. Specifically, as we apply the MAF filter and exclude all common variants, we are eliminating variants that have been misidentified in databases as disease causing²², of which there are more among individuals of African ancestry. Similarly, as we use *in silico* predictors to filter out putatively non-deleterious variants, we remove more functionally neutral variants from Africans than from Europeans. For instance, the number of non-deleterious PAVs per individual increases with African ancestry, whereas the number of deleterious PAVs per individual does not. Furthermore, because the number of deleterious mutations in African individuals is not greater than in European individuals^{25–28}, these filters do not remove more deleterious variants from Africans. This disproportionate removal of functionally neutral variants will more effectively reduce the number of incorrectly characterized variants in each class in African-ancestry individuals, and explains the reduction of the positive correlation with African ancestry as filters are applied.

Deleterious predictors are different depending on annotation.

While filtering significantly reduces the correlation between the number of deleterious PAVs and African ancestry, it does not impact the correlation between the number of deleterious NAVs and African ancestry. One possible explanation is that the effects of the filters differ between the two categories of variants, with functionally neutral variation filtered out more efficiently for

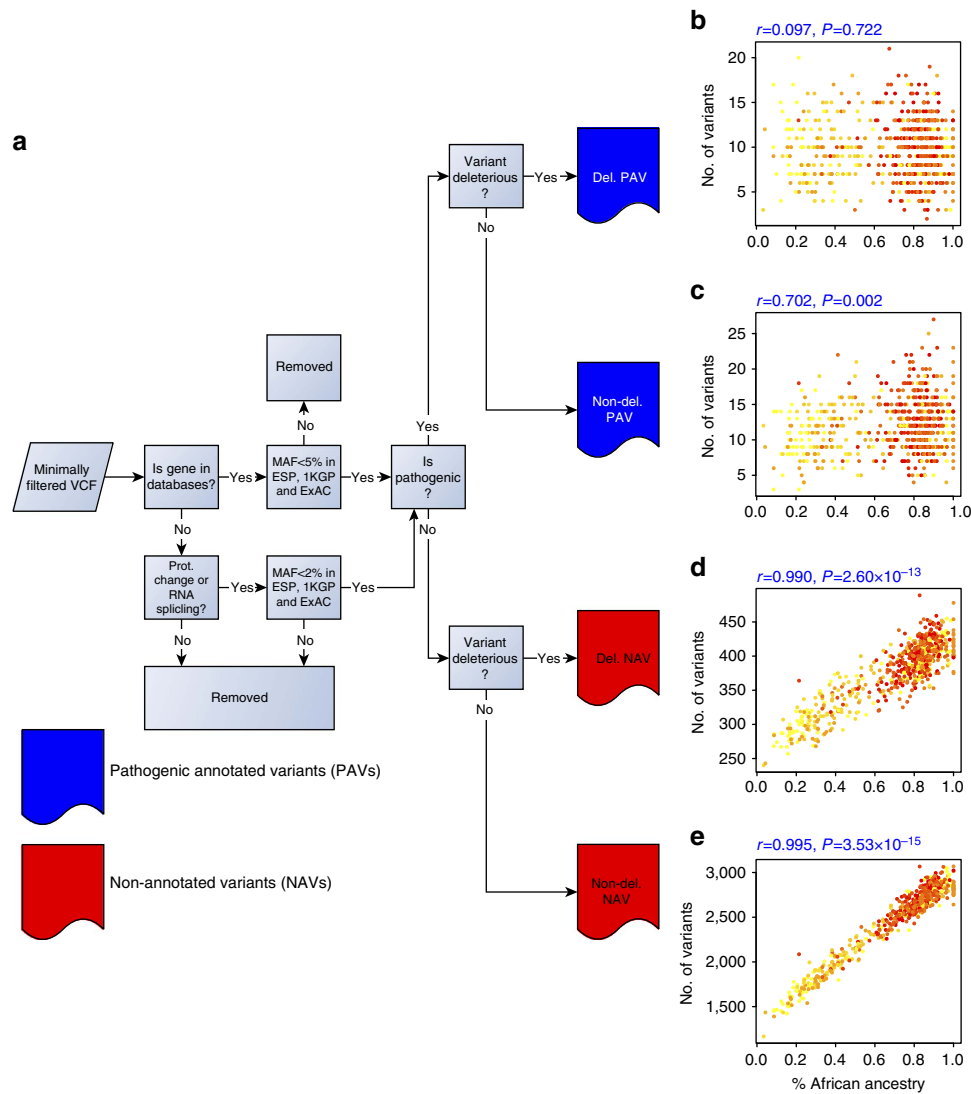


Figure 1 | Variant classification workflow and correlation with African ancestry. (a) The pipeline used to categorize variation into four groups (deleterious (Del.) PAVs, non-deleterious (Non-del.) PAVs, deleterious NAVs and non-deleterious NAVs) each with different levels of clinical relevance (see Methods for further explanation). The three key filters used in separating variants are (1) MAF from multiple databases, (2) pathogenic annotation (as defined by the ClinVar and/or HGMD) and (3) deleterious prediction. For **b–e** the x axis is the proportion of African ancestry as estimated by ADMIXTURE. The corresponding y axes represent the total number of variants per individual for the following groups: (b) deleterious PAVs, (c) non-deleterious PAVs, (d) deleterious NAVs and (e) non-deleterious NAVs. Colours of each individual reflect the population sampling location.

PAVs. Because we require at least 2 of 11 predictors to call a variant putatively deleterious, it is possible that predictors calling PAVs deleterious are consistently different than those calling NAVs deleterious. This is what we observe ($P \leq 10^{-15}$, χ^2 -test of independence), with predictors that use clinical databases to train their algorithms being over represented in deleterious PAV calls, and algorithms that are agnostic to clinical databases making up a larger percentage of the deleterious NAV calls (Supplementary Table 5). One possibility is that the machine learning-based algorithms preferentially optimize for patterns within the PAVs, and can thus inherit an ancestry-specific bias. Supporting this is the notion that most new African-specific causal variants will initially be identified as NAVs and may thus be less likely to be called by the currently trained predictors. Alternatively, though not mutually exclusively, conservation algorithms may be better able to remove background variants from the NAVs if the conservation score range for NAVs is significantly larger than PAVs. Given that NAVs are not

annotated and are less processed than PAVs, they are more likely to be sampled equally across the entire distribution of conservation scores, and to therefore represent a wider range of conservation scores than PAVs. This is consistent with what others have observed², and might explain why conservation algorithms predominate in the separation of deleterious NAVs from non-deleterious NAVs, compared with PAVs. While this differences in the type of predictors used in distinguishing deleterious and non-deleterious variants of different classification may represent the potential extension of ancestry related biases to deleterious predictors, this needs to be studied in more detail.

ClinVar correlation with African ancestry over time. To explore the historical context of recognized PAVs, and evaluate how ancestry related biases may have impacted the reproducibility of previous clinical applications relying on ClinVar, we conducted an analysis of how biases in archived versions of ClinVar have

changed over time. ClinVar, a developing database of pathogenic variation officially released in April 2013, was chosen for this analysis as it has monthly updates that allow us to easily track changes over time. As of March 2015, the number of known pathogenic variants has almost doubled from 14,697 to 26,409. In Fig. 2, we show the correlation over time between African-ancestry proportion in our CAAPA individuals and counts of ClinVar-based pathogenic variants in these same individuals for each update between 16 June 2012 (pre-official release) and 5 March 2015. As seen from this figure, the content of the database is highly susceptible to ancestry-related biases, which affects the interpretation of results. Furthermore, these biases can change over time, further complicating the ability to interpret results and account for ancestry-related biases. The largest change happens over a single month, from March to April of 2014, when a significant positive correlation ($r = 0.733$, $P = 0.001$) switches to a significant negative correlation ($r = -0.683$, $P = 0.004$). An analysis of differences between the March and April 2014 releases identifies 68 single-nucleotide polymorphisms (SNPs) that drive this marked change, and more details are presented in the supplement (Supplementary Note 1 and Supplementary Table 3).

The red line near the bottom of Fig. 2 shows the same correlation over time after filtering the data, again by MAF, mutation type and deleterious predictions (Fig. 1a). Similar to the unfiltered data, the filtered data show the first major shift in correlation from March to April of 2014, but the shift is in the opposite direction, with April showing a significantly less negative correlation (stats test) compared with March. The filtered data continue to show a less negative correlation for 3 months, before the pattern returns to a more significant negative correlation in July 2014, which is again similar in trend to but opposite in direction from the pattern seen in the unfiltered analysis. These simultaneous similarities and differences in the shift of the correlation between ancestry and pathogenic variation across database releases and filtering procedures reflect the precariousness of the current clinical databases, particularly when prioritizing variants of individuals with significant non-predominantly European ancestry. In contacting ClinVar about any possible curation differences for the March to July 2014 releases, we learned that ClinVar received a large deposit in April 2014 from the Breast Cancer Information Core database³⁰ with significant amounts of non-European data. While further information about this deposit is unavailable and exactly why it caused a marked change from positive to negative correlation is currently unclear, these observations further support our message that database content reflects ancestry-related biases and can impact overall reproducibility.

Analysis of ancestral biases at the gene level. To explore ancestral biases at a gene level, we evaluated the correlation between the number of PAVs per gene and African ancestry using the March 2015 release of ClinVar. After correcting for multiple testing, we found a significant negative correlation with African ancestry for 10 genes (Supplementary Table 2). These genes represent a subset with the strongest bias, and while we suspect this negative correlation with African ancestry is likely due to some type of technical or ascertainment bias, it is nevertheless possible that this bias could have some biological basis. These genes will require particular care in clinical analysis and represent an interesting set for follow-up investigation, as African causal variants in these genes are more likely to be labelled as NAVs and require a greater identification effort. We find, in general, that the subset of genes with significant positive or negative correlations ($P < 0.05$, uncorrected) are not enriched for those genes

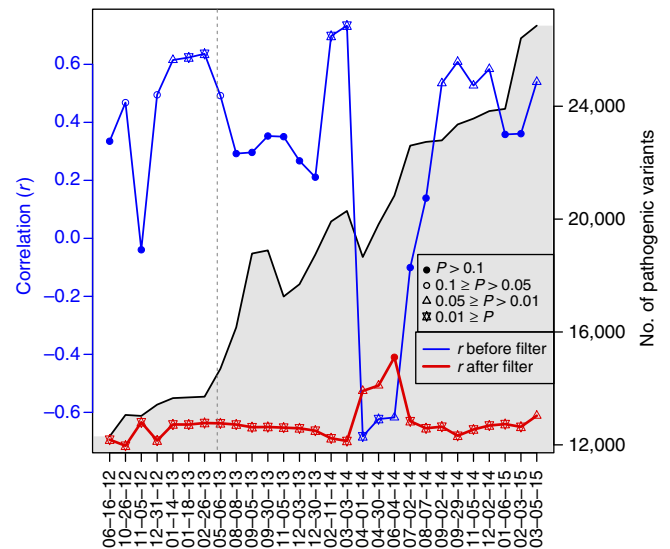


Figure 2 | Historical view of African-ancestry biases in ClinVar. The x axis represents the various archived versions of ClinVar. In the black y axis on the right, we see the number of PAVs recorded from each version of the database. There are a few decreases in numbers, but overall this number shows continuous growth. In the blue y axis on the left, we see the correlation coefficient estimated between the number of PAVs per CAAPA individual and their proportion of African ancestry. The dotted grey line represents the date of the first official release of ClinVar. The blue trend line shows the instability across different ClinVar releases of the correlation of African-ancestry proportion with average number of pathogenic variants per individual. The change in correlation is particularly notable for sequential releases between March and April 2014, after which the correlation remains significantly negative for 3 months (April–July 2014) before once again becoming significantly positive. The red trend line represents the same relationship between ancestry–pathogenicity correlation and ClinVar release over time after applying filters, and shows a significant change in correlation during the same 3-month period of April–July 2014 despite an overall reduction in movement of the correlation across time.

associated with known Mendelian diseases or those found in the GWAS catalogue³¹ (Methods).

Discussion

The ability to accurately report whether a genetic variant is responsible for a given disease or phenotypic trait depends in part on the confidence in labelling a variant as pathogenic. Such determination can often be more difficult in persons of predominantly non-European ancestry, as there is less known about the pathogenicity of variants that are absent from or less frequent in European populations. A key part of this are the differences between pathogenic variants, deleterious variants and prioritized variants, which are merely members of the proverbial haystack with differing levels of evidence for potential disease causality. It is important to note that a deleterious variant will only be labelled as pathogenic if its effect size is large enough to directly cause disease and this effect has been seen and annotated, and that a pathogenic variant will only be deleterious if it negatively impacts reproductive fitness. These terms are not the same, nor synonyms of true causality, but the use of deleteriousness as evidence for true disease causality is predicated on the fact that deleteriousness and pathogenicity should be correlated. While we cannot be sure which of these variants are truly disease-causing (actual ‘needles’ rather than haystack members) without additional functional or association-based evidence, we believe

that discrepancies between true pathogenicity and annotated pathogenicity are a major source of the biases we report. A likely contributor to this incongruity is that databases are missing population-specific pathogenicity information, and with regard to the results we report here, African-specific pathogenicity data. Therefore, true causal variants for predominantly non-European patients are likely to fall into the NAV categories. Since NAVs have the highest degree of positive correlation with African ancestry (that is, bias), causal variants falling into this group are more difficult to distinguish, as they exist amongst a larger number of high-priority background variants (that is, larger haystack). This problem is compounded in individuals of substantial African ancestry, as their larger amount of overall genetic variation^{5,23,24} results in an even greater number of deleterious NAVs requiring adjudication.

As a consequence, review of genomic test results for persons of predominantly non-European ancestry could be both more challenging and costly. Positive correlations of African-ancestry proportion with non-deleterious PAVs and deleterious NAVs result in more variants to evaluate for African-ancestry individuals (that is, larger haystack), which leads to higher costs and longer turnaround times. Assuming a cost of \$500 per variant for Sanger confirmation in a CLIA-certified laboratory (see Supplementary Table 6 for the range of costs found in clinical laboratories), and given gene candidate prioritization approaches that use phenotype to gene mapping³² and limit variants receiving follow-up confirmation to those in about 1% of the genome (that is, about 200 genes), we estimate an African-ancestry patient would have about 4.5 prioritized variants needing validation compared with 2.8 in an individual of European ancestry. This translates to a 1.6-fold increase in the number of variants prioritized, and represents a confirmation cost difference of over \$800 per patient. Notably, these estimates are simplified and conservative, as we do not consider the substantial cost of having each of these variants reviewed by a clinician.

A potential solution would be to reserve follow-up confirmation for deleterious PAVs, which are uncorrelated to African ancestry and should therefore not be more common in individuals of African ancestry. However, doing this would limit the diagnostic landscape for both Europeans and non-Europeans to only previously found variation, and would greatly undermine the promise that sequencing technology holds for clinical genomics. Furthermore, this would limit the field to Euro-centric databases that would frequently miss causal variants in minority populations. In these situations, the missed causal variants would only be represented among the NAVs, which underlines the importance of not excluding prioritized NAVs from follow-up analysis.

These limitations translate into serious challenges, and despite the increased costs, provide good reason to cast a wider net for variant prioritization and confirmation when applying genomic testing to patients of African ancestry, and likely other predominantly non-European ancestries. As long as ancestry-related biases are not addressed, and most studies continue to predominantly sample from European populations, the genetics community will face challenges with implementation, interpretation and cost-effectiveness when treating minority populations.

Methods

Filtering pipeline. We annotate all variation using ANNOVAR³³, a programme that facilitates the comprehensive and integrative annotation of multiple data types for each variant. Variants are divided into two main classes, each with two subgroups, for a total of four categories. PAVs consist of variants annotated as pathogenic in clinically annotated genetic databases, and are subdivided into deleterious and non-deleterious subgroups as determined by *in silico* predictions. NAVs include all variants not annotated as pathogenic (labelled as disease mutations), as well as those entirely absent from clinically annotated databases, and

are also subdivided into deleterious and non-deleterious subgroups. Using customized ANNOVAR index tables, we annotate variants with 11 *in silico* predictors of function^{11–20} (Supplementary Table 4), functional information about protein-coding effect, clinical variation knowledge from ClinVar¹⁰ (all archived versions from 2012 to March 2015), the professional version of HGMD⁹ (fourth quarter version of 2014) and allele frequencies from multiple population sequencing projects, including the 1000 Genomes Project (phase 3)²³, the ExAC database (<http://exac.broadinstitute.org>) and the Exome Sequencing Project⁵. We also integrate the final output as a list of variants belonging to each of the variant classes described above (Fig. 1a)¹.

Filtering criteria. For variants from the OMIM⁸, HGMD⁹ and ClinVar¹⁰ databases, only those found in protein-coding genes are included. We also remove variants with MAF > 5% in any of the 1000 Genome super-populations, ExAC populations or Exome Sequencing Project populations. With regard to the analysis portrayed in Fig. 1, if a variant is not found in any of the clinical databases, we use an allele frequency cutoff of 2% (Fig. 1a) and include only protein-altering variants found in the three following gene annotation databases: ENSEMBL GENE; KnownGene; or RefSeq. We also filter variants on the basis of *in silico* prediction, and require that at least 2 of 11 *in silico* prediction methods identify variants as deleterious (see Supplementary Table 4 for individual predictor cutoffs). An exception to this is that nonsense and splice site variants are called deleterious irrespective of their *in silico* predictors. Situations where these predicted deleteriousness filters are not applied are identified as exceptions in the text.

Variant classes. The first variant class, deleterious PAVs, are defined as variants with exact matches in genes in the OMIM⁸, HGMD⁹ or ClinVar¹⁰ databases, and are known to be associated with disease phenotypes. In addition, this class has to meet the above *in silico* prediction filter. The second class of variants is non-deleterious PAVs, and they only differs from the first category in that the requirement of being deleterious is removed. Deleterious NAVs make up the third class. This class is not annotated as pathogenic in any of the clinical databases, but these variants are identified by at least two *in silico* predictors as being deleterious. Finally, variants neither previously annotated as pathogenic nor predicted to be deleterious by at least two *in silico* predictors are classified as non-deleterious NAVs; they are seen as the least likely to be causative for a known disorder. Non-deleterious NAVs are also filtered by the frequency filters described above. Overall, <1% of NAVs are found in databases but not annotated as disease-causing; the remaining NAVs are not identified in any database.

Whole-genome sequencing data from the CAAPA Project. CAAPA consists of high coverage (~30×) whole-genome sequence data ($N=642$) and provides a catalogue of genetic diversity from multiple populations of African descent. These populations include individuals from North America, South America, the Caribbean and continental Africa⁷, and study individuals are categorized as being cases and controls for asthma (sampling and variant calling are presented in more detail in Mathias *et al.*⁷) However, we do not suspect an atypical number of clinically relevant pathogenic variants among cases with such a complex disease phenotype as asthma. We have sampled 16 populations, including 8 different African American populations. Assembly of individual genomes, as well as variant calls are done using the Consensus Assessment of Sequence and Variation (CASAVA) package³⁴. Using probabilistic models to build probability distributions over all diploid genotypes at every genomic site, genotypes are called after numerous quality control filtering steps. For each genomic position, a set of candidate SNPs becomes output. Multi-sample VCFs are generated at Knome Inc. (Cambridge, MA, USA), using VCFtools v0.1.11 (ref. 35) and custom scripts for additional data processing. While everything we report is based on a genome sequencing data set containing 642 samples, when we repeated our analyses on an expanded set of samples (total $N \sim 950$) containing additional currently unpublished CAAPA data, our results were unchanged.

Estimation of ancestry proportions. To estimate ancestry we combine the CAAPA data with phase 1 of the 1000 Genomes Project, and data from previously published studies, which genotyped Hispanic and Native American samples on an Affymetric 6.0 chip^{36,37}. All A–T and G–C SNPs are removed, and a missingness filter of 5% and a MAF filter of < 5% are applied. The resulting SNPs are then LD pruned with plink³⁸ using windows of 50 SNPs and removing SNPs with an $r^2 > 0.25$, then iterating by 5 SNPs (that is, plink command—*indep-pairwise* 50 5 0.25). This results in 167,987 SNPs for admixture analysis.

We estimate ancestry proportions using the software package ADMIXTURE³⁹. After performing 30 replicates modelling four clusters, we select the parameter values with the highest negative log likelihood. We identify the cluster that represents African ancestry by using the African groups from the 1000 Genomes Project as a reference (that is, the cluster where they have >99% membership), and we extract the proportion estimates for each of our CAAPA samples from this cluster. These become the values used to estimate the correlations. We present them as a bar plot in Supplementary Fig. 1.

Statistics to accommodate sampling structure. Owing to our population sampling approach, the full cohort does not represent an unstructured selection of individuals of African ancestry. To account for this when performing correlation analysis, we use the approach implemented in the R package ‘psych’⁴⁰. The approach estimates correlations within each single population, which represent the pillars of the population substructure, and then combines these estimates weighted by sample size. Reported correlation coefficients and the *P* values are from the ‘weights’⁴¹ package, and significance is reported with a false discovery rate approach to correct for multiple testing.

Data availability. The whole-genome sequence data referenced in this study were generated by the CAAPA⁷ and have been deposited in dbGAP with the accession code phs001123.v1.p1.

References

- Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *N. Engl. J. Med.* **369**, 1502–1511 (2013).
- Lee, H. *et al.* Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* **312**, 1880–1887 (2014).
- Yang, Y. *et al.* Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* **312**, 1870–1879 (2014).
- Kidd, J. M. *et al.* Population genetic inference from personal genome data: impact of ancestry and admixture on human genomic variation. *Am. J. Hum. Genet.* **91**, 660–671 (2012).
- Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
- 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Mathias, R. A. *et al.* A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat. Commun.* **7**, 12522 (2016).
- Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
- Stenson, P. D. *et al.* The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum. Genet.* **133**, 1–9 (2014).
- Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–D985 (2014).
- Siepel, A., Pollard, K. S. & Haussler, D. in *Research in Computational Molecular Biology* 190–205 (Springer, 2006).
- Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
- Garber, M. *et al.* Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**, i54–i62 (2009).
- Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
- Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
- Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6**, e1001025 (2010).
- Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
- Shihab, H. A., Gough, J., Cooper, D. N., Day, I. N. & Gaunt, T. R. Predicting the functional consequences of cancer-associated amino acid substitutions. *Bioinformatics* **29**, 1504–1510 (2013).
- Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
- Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
- Saunders, C. J. *et al.* Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci. Transl. Med.* **4**, 154ra135 (2012).
- Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
- 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
- Fu, W., Gittelman, R. M., Bamshad, M. J. & Akey, J. M. Characteristics of neutral and deleterious protein-coding variation among individuals and populations. *Am. J. Hum. Genet.* **95**, 421–436 (2014).
- Simons, Y. B., Turchin, M. C., Pritchard, J. K. & Sella, G. The deleterious mutation load is insensitive to recent population history. *Nat. Genet.* **46**, 220–224 (2014).
- Do, R. *et al.* No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat. Genet.* **47**, 126–131 (2015).
- Henn, B. M., Botigue, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nat. Rev. Genet.* **16**, 333–343 (2015).
- Henn, B. M. *et al.* Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proc. Natl Acad. Sci. USA* **113**, E440–E449 (2016).
- Szabo, C., Masiello, A., Ryan, J. F. & Brody, L. C. The breast cancer information core: database design, structure, and scope. *Hum. Mutat.* **16**, 123 (2000).
- Hindorf, L. A. *et al.* A Catalog of Published Genome-Wide Association Studies. (*European Bioinformatics Institute*) Available at: www.genome.gov/gwastudies (Date accessed 14 October 2015).
- Groza, T. *et al.* The human phenotype ontology: semantic unification of common and rare disease. *Am. J. Hum. Genet.* **97**, 111–124 (2015).
- Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
- CASAVA v1.8.2 (Illumina Inc., 2014).
- Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
- Bigham, A. *et al.* Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* **6**, e1001116 (2010).
- Wall, J. D. *et al.* Genetic variation in Native Americans, inferred from Latino SNP and resequencing data. *Mol. Biol. Evol.* **28**, 2231–2237 (2011).
- Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- Revelle, W. *psych: Procedures for Personality and Psychological Research*. R package version 1 (Northwestern University, Evanston, Illinois, USA, 2014).
- Pasek, J., Tahk, Alex, Culter, Gene & Marcus, Schwemmler. *Weights: Weighting and Weighted Statistics*. Computer software. CRAN. Version 0.80. CRAN, 04 March 2014 <https://cran.r-project.org/web/packages/weights/index.html> (accessed on 14 October 2015) (2014).

Acknowledgements

We acknowledge the contributions of Paul Levett, Anselm Hennis, P. Michele Lashley, Raana Naidu, Malcolm Howitt and Timothy Roach (BAGS); Audrey Grant, Eduardo Viera Ponte, Alvaro A. Cruz and Edgar Carvalho (BIAS); Susan Balcer-Whaley, Maria Stockton-Porter and Mao Yang (GRAAD); Mario Meraz, Jaime Nuñez and Eileen Fabiani Herrera Mejía (HONDAS); Deanna Ashley (JAAS); Silvia Jimenez, Nathalie Acevedo and Dilia Mercado (PGCA); Ann Jedlicka (REACH); Addison K. May, Caroline Gilmore and Patricia Minton (Vanderbilt University); Qun Niu (University of Chicago); and Adeyinka Falusi and Abayomi Odetunde (University of Ibadan, Nigeria). We also acknowledge the support of John Jay Shannon (Cook County Health Systems) and Kevin Weiss (Northwestern University); Regina Miranda and the Indians Zenues guards (San Basilio de Palenque, Bolivar, Colombia); Ulysse Ateba Ngoa (Leiden University); and Charles Rotimi, Adeyemo Adebowale, Floyd J Malveaux and Elena Reece (Howard University). We thank the numerous health-care providers and community clinics and co-investigators who assisted in the phenotyping and collection of DNA samples, and the families and patients for generously donating DNA samples to BAGS, BIAS, BREATHE, CAG, GRAAD, HONDAS, REACH, SAGE II, VALID, SAPPHIRE, SARP, COPDGene, JAAS, GALA II, PGCA and AEGS. Special thanks to community leaders, teachers, doctors and personnel from health centres at the Garifuna communities for organizing the medical brigades and to the medical students at Universidad Católica de Honduras, Campus San Pedro y San Pablo for their participation in the fieldwork related to HONDAS; study coordinator Sandra Salazar, and the recruiters in SAGE and GALA: Duanny Alva, MD; Gaby Ayala-Rodriguez; Ulysses Burley; Lisa Caine; Elizabeth Castellanos; Jaime Colon; Denise DeJesus; Iliana Flexas; Blanca Lopez; Brenda Lopez, MD; Louis Martos; Vivian Medina; Juana Olivo; Mario Peralta; Esther Pomares, MD; Jihan Quraishi; Johanna Rodriguez; Shahdad Saeedi; Dean Soto; Ana Taveras; Emmanuel Viera; Dr Michael LeNoir; Dr Kelley Meade; Mindy Jensen; and Adam Davis; and health liaisons and public health officers of the main Conde office, Adaliudes Conceição, Luciana Quintela, Ivanice Santos, Analú Lima, Benivaldo Valber Oliveira Silva and Iraci Santos Araujo, and students from the Federal University of Bahia who assisted in data collection in BIAS: Rafael Santana; Roberta Barbosa; Ana Paula Santana; Charlton Barros; Marcelle Brandão; Ludmila Almeida; Thiago Cardoso; and Daniela Costa. We are grateful for the support from the international state governments and universities from Honduras, Colombia, Brazil, Gabon, Nigeria, Netherlands, Jamaica, Barbados and the United States who made this work possible. We also thank Robert Guenaro for invaluable assistance in the whole-genome sequencing at Illumina, Inc.; Gonçalo Abecasis, William Cookson and Miriam Moffatt for helpful discussions; Pat Oldewurtel and Murali Bopparaju for technical support; Shuai Yuan for software support; and Kit

Rees and Cate Kiefe for artistic contributions. We thank Steven Salzberg and Alex Szalay for computing and data storage resources available on the Data-Scope instrument at the Institute for Data Intensive Science (IDIES), Johns Hopkins University. We also thank Aksinija A. Shamah for artistic assistance. We acknowledge the support from James Kiley, Susan Banks-Schlegel and Weiniu Gan at the National Heart, Lung, and Blood Institute. Funding for this study was provided by the National Institutes of Health (NIH) R01HL104608 and the Center for Health Related Informatics and Bioimaging at the University of Maryland (M.D.K., A.C.S. and T.D.O.). Additional NIH funding includes the following: NCI: R21CA178706 (R.D.H.), U01CA161032 and P50CA125183 (O.I.O.); NCCR: G12RR003048 (G.M.D.) and RR24975 (TH); NHGRI: R01HG007644, R21HG007233 (R.D.H.), R21HG004751 (H.R.J., J.G. and Z.S.Q.) and T32HG000044 (C.R.G.); NHLBI: R01HL087699 (K.C.B.), R01HL118267 (L.K.W.), R01HL117004, R01HL088133, R01HL004464 (E.G.B.), HL081332, HL112656 (L.B.W.), R01HL69167, U01HL109164 (E.B. and D.M.), RC2HL101651, RC2HL101543, U01HL49596, R01HL072414 (C.O.), R01HL089897, R01HL089856, K01HL092601 (M.G.F.), R01HL51492, R01HL/AI67905 (J.G.F.), HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C and HHSN268201300050C (J.G.W.); NIAID: K08AI01582 (T.H.), R01AI079139 (L.K.W.) and U19AI095230 (C.O.); NIEHS: R01ES015794 (E.G.B.); NIGMS: S06GM08016 (M.U.F.) and T32GM07175 (C.R.G.); NIMHD: P60MD006902 (E.G.B.), 8U54MD007588 and P20MD0066881 (M.G.F.); NSFGRF #1144247 (R.T.). Additional sources of funding include the following: American Asthma Foundation (L.K.W. and E.G.B.); American Lung Association Clinical Research Grant (T.H.); Colombian Government (Colciencias) 331-2004 and 680-2009 (L.C.); EDCTP:CT.2011.40200.025 (A.A.A.); EU-IDEA HEALTH-F3-2009-241642 and EU-TheSchistoVac HEALTH-Fe-2009-242107 (M.Y.); Ernest Bazley Fund (P.C.A., R.K., L.G. and R.S.); and the Fund for Henry Ford Hospital (L.K.W.). The Jamaica 1986 Birth Cohort Study was supported by grants from the Caribbean Health Research Council, Caribbean Cardiac Society, National Health Fund (Jamaica) and Culture Health Arts Sports and Education Fund (Jamaica). Study nurses were supported by the University Hospital of the West Indies (T.F. and J.K.M.), Ralph and Marion Falk Medical Trust (C.O.O., O.O. and G.A.), UCSF Dissertation Year Fellowship (C.R.G.), Universidad Católica de Honduras, San Pedro Sula (E.H.P.), University of Cartagena (J.M.) and Wellcome Trust 072405/Z/03/Z, 088862/Z/09/Z (P.J.C.). The Jackson Heart Study is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C and HHSN268201300050C from the NHLBI and the NIMHD. E.G.B. was funded by Flight Attendant Medical Research Institute, RWJF Amos Medical Faculty Development Award and the Sandler Foundation; the Sloan Foundation to R.D.H.; C.R.G. was supported in part by the UCSF Chancellor's Research Fellowship and Dissertation Year Fellowship. K.C.B. was supported in part by the Mary Beryl Patch Turnbull Scholar Program. R.A.M. was supported in part by the MOSAIC Initiative Awards from Johns Hopkins University. M.P.-Y. was funded by a Postdoctoral Fellowship from Fundación Ramón Areces. M.I.A. is an investigator

supported by National Council for Scientific and Technological Development (CNPq). T.V.H. was supported in part by K24 AI 77930, UL1 TR00445 and U19 AI95227. R.O. was funded by NHLBI Diversity Supplement R01HL104608. Funding for the cohorts was provided by the following: AEGS, BAGS, BIAS, BREATHE (K08AI001582 and RR24975), CAG, COPDGene, GALA II, GRAAD, HONDAS, JAAS (The Jamaica 1986 Birth Cohort Study was supported by grants from the Caribbean Health Research Council, Caribbean Cardiac Society, National Health Fund (Jamaica) and Culture Health Arts Sports and Education Fund (Jamaica)). The study nurses were supported by the University Hospital of the West Indies), PGCA (University of Cartagena and Colciencias Contracts 183-2002, 680-2009), REACH, SAGE II, SAPPHIRE, SARP, SCAALA and VALID.

Author contributions

T.D.O. conceived the project; M.D.K. and T.D.O. designed and performed experiments, analysed data and wrote the manuscript; L.Y.-A., M.A.T., A.C.S., K.M., L.J.B.J., I.R., A.M.L., L.K.W., T.H.B., R.A.M. and K.C.B. provided technical assistance and assisted with the writing and review of the manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Kessler, M. D. *et al.* Challenges and disparities in the application of personalized genomic medicine to populations with African ancestry. *Nat. Commun.* 7:12521 doi: 10.1038/ncomms12521 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA)

Meher Preethi Boorgula⁹, Monica Campbell⁹, Sameer Chavan⁹, Jean G. Ford^{8,11}, Cassandra Foster⁹, Li Gao⁹, Nadia N. Hansel⁹, Edward Horowitz⁹, Lili Huang⁹, Romina Ortiz⁹, Joseph Potee⁹, Nicholas Rafaels⁹, Alan F. Scott⁹, Candelaria Vergara⁹, Jingjing Gao¹², Yijuan Hu¹³, Henry Richard Johnston¹³, Zhaohui S. Qin¹³, Badri Padhukasahasram^{6,7}, Georgia M. Dunston^{14,15}, Mezbah U. Faruque¹⁵, Eimear E. Kenny^{16,17}, Kimberly Gietzen¹⁸, Mark Hansen¹⁸, Rob Genuario¹⁸, Dave Bullis¹⁸, Cindy Lawley¹⁸, Aniket Deshpande¹⁹, Wendy E. Grus¹⁹, Devin P. Locke¹⁹, Marilyn G. Foreman²⁰, Pedro C. Avila²¹, Leslie Grammer²¹, Kwang-Youn A. Kim²², Rajesh Kumar^{23,24}, Robert Schleimer²⁵, Carlos Bustamante¹⁶, Francisco M. De La Vega¹⁶, Chris R. Gignoux¹⁶, Suyash S. Shringarpure¹⁶, Shaila Musharoff¹⁶, Genevieve Wojcik¹⁶, Esteban G. Burchard^{26,27}, Celeste Eng²⁷, Pierre-Antoine Gourraud²⁸, Ryan D. Hernandez^{26,29,30}, Antoine Lizee²⁸, Maria Pino-Yanes^{27,31}, Dara G. Torgerson²⁷, Zachary A. Szpiech²⁶, Raul Torres³², Dan L. Nicolae^{33,34}, Carole Ober³⁵, Christopher O. Olopade³⁶, Olufunmilayo I. Olopade³³, Oluwafemi Oluwole³³, Ganiyu Arinola³⁷, Wei Song^{1,2,3}, Goncalo Abecasis³⁸, Adolfo Correa³⁹, Solomon Musani³⁹, James G. Wilson⁴⁰, Leslie A. Lange⁴¹, Joshua Akey⁴², Michael Bamshad⁴³, Jessica Chong⁴³, Wenqing Fu⁴², Deborah Nickerson⁴², Alexander Reiner⁴⁴, Tina Hartert⁴⁵, Lorraine B. Ware^{45,46}, Eugene Bleecker⁴⁷, Deborah Meyers⁴⁷, Victor E. Ortega⁴⁷, Maul R.N. Pissamai⁴⁸, Maul R.N. Trevor⁴⁸, Harold Watson^{49,50}, Maria Ilma Araujo⁵¹, Ricardo Riccio Oliveira⁵², Luis Caraballo⁵³, Javier Marrugo⁵⁴, Beatriz Martinez⁵³, Catherine Meza⁵³, Gerardo Ayestas⁵⁵, Edwin Francisco Herrera-Paz^{56,57,58}, Pamela Landaverde-Torres⁵⁶, Said Omar Leiva Erazo⁵⁶, Rosella Martinez⁵⁶, Alvaro Mayorga⁵⁷, Luis F. Mayorga⁵⁶, Delmy-Aracely Mejia-Mejia^{57,58}, Hector Ramos⁵⁶, Allan Saenz⁵⁵, Gloria Varela⁵⁵, Olga Marina Vasquez⁵⁸, Trevor Ferguson⁵⁹, Jennifer Knight-Madden⁵⁹, Maureen Samms-Vaughan⁶⁰, Rainford J. Wilks⁵⁹, Akim Adegnika^{61,62,63}, Ulysse Ateba-Ngoa^{61,62,63} & Maria Yazdanbakhsh⁶³

¹¹Department of Medicine, The Brooklyn Hospital Center, Brooklyn, New York 11201, USA; ¹²Data and Statistical Sciences, AbbVie, North Chicago, Illinois 60064, USA; ¹³Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia 30322, USA; ¹⁴Department of Microbiology, Howard University College of Medicine, Washington, DC 20059, USA; ¹⁵National Human Genome Center, Howard University College of Medicine, Washington, DC 20059, USA; ¹⁶Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA; ¹⁷Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA; ¹⁸Illumina, Inc., San Diego, California 92122, USA; ¹⁹Knome Inc., Cambridge, Massachusetts 02141, USA; ²⁰Pulmonary and Critical Care Medicine, Morehouse School of Medicine, Atlanta, Georgia 30310, USA; ²¹Department of Medicine, Northwestern University, Chicago, Illinois 60611, USA; ²²Department of Preventive Medicine, Northwestern University, Chicago, Illinois 60611, USA; ²³Department of Pediatrics, Northwestern University, Chicago, Illinois 60611, USA; ²⁴The Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois 60611, USA; ²⁵Department of Medicine, Northwestern Feinberg School of Medicine, Chicago, Illinois 60611, USA; ²⁶Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California 94158, USA; ²⁷Department of Medicine, University of California, San Francisco, San Francisco, California 94143, USA; ²⁸Department of Neurology, University of California, San Francisco, San Francisco, California 94158, USA; ²⁹Institute for Human Genetics, University of California, San Francisco, San Francisco, California 94143, USA; ³⁰California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, California 94143, USA; ³¹CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, 28029 Madrid, Spain; ³²Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, California 94158, USA; ³³Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA; ³⁴Department of Statistics, University of Chicago, Chicago, Illinois 60637, USA; ³⁵Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA; ³⁶Department of Medicine and Center for Global Health, University of Chicago, Chicago, Illinois 60637, USA; ³⁷Department of Chemical Pathology, University of Ibadan, 900001 Ibadan, Nigeria; ³⁸Department of Biostatistics, SPH II, University of Michigan, Ann Arbor, Michigan 48109, USA; ³⁹Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA; ⁴⁰Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA; ⁴¹Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA; ⁴²Department of Genetic Sciences, University of Washington, Seattle, Washington 98195, USA; ⁴³Department of Pediatrics, University of Washington, Seattle, Washington 98195, USA; ⁴⁴Department of Epidemiology, University of Washington, Seattle, Washington 98195, USA; ⁴⁵Department of Medicine, Vanderbilt University, Nashville, Tennessee 37232, USA; ⁴⁶Department of Pathology, Microbiology and Immunology, Vanderbilt University, Nashville, Tennessee 37232, USA; ⁴⁷Center for Human Genomics and Personalized Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA; ⁴⁸Genetics and Epidemiology of Asthma in Barbados, The University of the West Indies, Bridgetown BB11115, Barbados; ⁴⁹Faculty of Medical Sciences Cave Hill Campus, The University of the West Indies, Bridgetown BB11000, Barbados; ⁵⁰Queen Elizabeth Hospital, The University of the West Indies, St Michael BB11115, Barbados; ⁵¹Immunology Service, Universidade Federal da Bahia, 401110-170 Salvador, Brazil; ⁵²Laboratório de Patologia Experimental, Centro de Pesquisas Gonçalo Moniz, 40296-710 Salvador, Brazil; ⁵³Institute for Immunological Research, Universidad de Cartagena, Cartagena 130000, Colombia; ⁵⁴Instituto de Investigaciones Inmunológicas, Universidad de Cartagena, Cartagena 130000, Colombia; ⁵⁵Faculty of Medicine, Universidad Nacional Autónoma de Honduras en el Valle de Sula, 21102 San Pedro Sula, Honduras; ⁵⁶Facultad de Medicina, Universidad Católica de Honduras, 21102 San Pedro Sula, Honduras; ⁵⁷Centro de Neumología y Alergias, 21102 San Pedro Sula, Honduras; ⁵⁸Faculty of Medicine, Centro Médico de la Familia, 21102 San Pedro Sula, Honduras; ⁵⁹Tropical Medicine Research Institute, The University of the West Indies, St Michael BB11115, Barbados; ⁶⁰Department of Child Health, The University of the West Indies, Kingston 7, Jamaica; ⁶¹Centre de Recherches Médicales de Lambaréné, 13901 Libreville, Gabon; ⁶²Institut für Tropenmedizin, Universität Tübingen, 72074 Tübingen, Germany; ⁶³Department of Parasitology, Leiden University Medical Center, 02333 Leiden, Netherlands.