The NEW ENGLAND JOURNAL of MEDICINE

Please review the Supplemental Files folder to review documents not compiled in the PDF.

Genetic Misdiagnoses and the Potential for Health Disparities

Journal:	*New England Journal of Medicine*
Manuscript ID	15-07092.R4
Article Type:	Special Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Manrai, Arjun; Harvard Medical School, Department of Biomedical Informatics; Harvard-MIT, Health Sciences and Technology Funke, Birgit; Partners HealthCare, Laboratory for Molecular Medicine Rehm, Heidi; Partners Healthcare Center for Personalized Genetic Medicine, Laboratory for Molecular Medicine; Brigham & Women's Hospital & Harvard Medical School, Department of Pathology Olesen, Morten; Rigshospitalet, University of Copenhagan, Department of Cardiology Maron, Bradley; Brigham and Women's Hospital, Cardiology Szolovits, Peter; MIT, Electrical Engineering and Computer Science Margulies, David; Harvard Medical School, Department of Biomedical Informatics Loscalzo, Joseph; Brigham and Women's Hospital, Department of Medicine Kohane, Isaac; Harvard Medical School, Department of Biomedical Informatics
Abstract:	BACKGROUND—Risk stratification for hypertrophic cardiomyopathy (HCM) has been enhanced by targeted genetic testing for over a decade. Using sequencing results, clinicians routinely assess risk for an HCM patient's relatives and diagnose HCM in ambiguous clinical presentations. However, the benefits of genetic testing come with the risk that variants may be misclassified. METHODS—Using publicly accessible exome data, we identified variants previously considered causal of HCM that were overrepresented in the general population. We studied these variants in diverse populations, and reevaluated their initial ascertainments in the medical literature. We reviewed patient records at a leading genetic testing laboratory for variant occurrences during the near decade-long history of the laboratory. RESULTS—Multiple patients, notably all of African or unspecified ancestry, received positive reports with variants misclassified as pathogenic based on prior understanding. Subsequently, all reported variants were recategorized as benign. The mutations that were most common were significantly more common in African Americans than European Americans ($P < 0.001$). Simulations show that even small numbers of African Americans included in control cohorts would likely have prevented these

	misclassifications. We identify methodological shortcomings that contributed to these errors in the medical literature. CONCLUSIONS—These findings illustrate the need for sequencing the genomes of diverse populations, both in asymptomatic controls and the tested population. These results expand upon current guidelines, which recommend using ancestry matched controls to interpret variants. As additional populations of different ancestry backgrounds are sequenced, we expect variant reclassifications to increase, particularly for ancestry groups that have historically been less well studied.

Genetic Misdiagnoses and the Potential for Health Disparities

Arjun K. Manrai
Division of Health Sciences and Technology, Harvard-MIT, Cambridge, MA
Department of Biomedical Informatics, Harvard Medical School, Boston, MA

Birgit H. Funke
Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA
Department of Pathology, Harvard Medical School and Massachusetts General Hospital, Boston, MA

Heidi L. Rehm
Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA
Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

Morten S. Olesen
Laboratory for Molecular Cardiology, Department of Cardiology, the Heart Centre,
Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

Bradley A. Maron
Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School,
Boston, MA

Peter Szolovits
Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology,
Cambridge, MA

David M. Margulies
Department of Biomedical Informatics, Harvard Medical School, Boston, MA

Joseph Loscalzo
Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

Isaac S. Kohane
Division of Health Sciences and Technology, Harvard-MIT, Cambridge, MA
Department of Biomedical Informatics, Harvard Medical School, Boston, MA
10 Shattuck St., Boston, MA, 02115
isaac_kohane@harvard.edu
617-432-2145 (phone)
617-432-6039 (fax)

BACKGROUND

Although hypertrophic cardiomyopathy (HCM) is best known as a fatal disease of young athletes, it causes significant morbidity and mortality in patients of all ages and lifestyles.[1,2] The defining feature of HCM is unexplained left ventricular hypertrophy (LVH), but its clinical presentation is variable, manifesting as severe heart failure in some patients yet being asymptomatic in others.[3] In over one-third of patients, causal genetic lesions are identified, enabling clinicians to risk stratify the patient's relatives[4] and in rare circumstances, tailor therapy for a patient found to have a tractable disorder such as Fabry disease.[5] Additionally, in patients with clinical features but not a definitive diagnosis of HCM, identification of a pathogenic variant may be used to help establish a diagnosis.

The provision of false genetic information to a patient, such as when a patient is incorrectly informed that one of his or her variants is causal when in fact it is benign, can have far-reaching adverse consequences within the family. First, relatives possessing the non-causal variant receive prolonged at-risk screening and are advised about lifestyle modifications (e.g., cessation of certain sports and activities) and may suffer stress and economic burden consequent to an incorrect diagnosis. Second, relatives who lack the non-causal variant are given false reassurance that further surveillance is unnecessary. Third, for patients with clinical features but without a definitive diagnosis of HCM, such as young athletes with modest hypertrophy and a family history of sudden cardiac death, misclassification of a benign variant as pathogenic may lead to overestimation of the benefits of implanting a cardioverter-defibrillator to prevent sudden cardiac death. And, when

a variant's status is changed from pathogenic to benign, the sequencing laboratory often (but not obligatorily) re-contacts the referring physician who, in turn, re-contacts the patient and their tested family members, engendering confusion and distrust.

Much effort has gone into developing standards for correct interpretation of genetic variants.[6–10] The principal challenge is to separate truly pathogenic variants from the historically underappreciated amount of background variant noise in the genome.[8,11] Expert guidelines generally recommend classifying variants using ancestry-matched control sequence data.[4,6]

When large-scale control sequence data from the NHLBI Exome Sequence Project[12] (ESP) were systematically reviewed for HCM-associated variants labeled "disease-causing" or "pathogenic" in an expert-curated database,[13,14] many more "HCM" variants were found than were expected in the general population (given the prevalence of HCM), implying reduced penetrance or misclassification errors in prior HCM-variant associations, or both. We observed that five high-frequency variants (>1% minor allele frequency in either ESP African Americans or European Americans) account for the majority of this overabundance of misclassified HCM variation, and that these variants occur disproportionately in African American individuals.

We hypothesized that the identification of HCM-associated high-frequency variants in the general population implied historical errors in patient reports, and that most or all individuals affected would be of African ancestry. We further posited that these variant associations stemmed from ascertainment bias and other

methodological shortcomings in the original studies. In order to test these hypotheses, we searched patient records for occurrences of these variants at a premier genetic testing laboratory and reviewed the literature implicating these variants. We describe here a cautionary tale of broad relevance to genetic diagnosis. Although this tale is well known to geneticists, physicians may be less familiar with it.

METHODS

Study populations

We used publicly-accessible sequence data from the NHBLI Exome Sequence Project (ESP),[12] 1000 Genomes Project (1000G),[15] and Human Genome Diversity Project (HGDP).[16] For estimating minor allele frequency, the NHBLI ESP had exome data for 4,300 European Americans and 2,203 African Americans; the 1000 Genomes Project Phase 1 had whole-genome data for 1,092 individuals from 14 worldwide populations; and the HGDP had whole-genome SNP data for 938 individuals from 51 worldwide populations. Clinical records for HCM patients were reviewed at the Laboratory for Molecular Medicine (LMM), Partners HealthCare, Boston, MA. This population has been described in detail previously.[17] All HCM patient reports with an originally reported variant status of "Pathogenic," "Presumed Pathogenic," "Unknown Significance," and "Pathogenicity Debated" were included (Table 1). Demographic data were acquired from the genetic testing requisition form. The LMM patient population is a mixed population of 64% white/European American and 8% black/African American individuals, with the remaining individuals of other or unspecified ancestry.[17] This study was conducted in accordance with a waived-consent, institutional review board-approved protocol.

Variant ascertainment

A targeted search was performed for initial disease-variant associations for all HCM-associated high-frequency variants in the medical literature using PubMed. All Human Genome Variation Society (HGVS) names for the variants (e.g., K247R and

Lys247Arg) were used as well as all possible transcript variants obtained from NCBI dbSNP Build 140.[18] All original reports of disease-variant associations were in agreement with those listed in the Human Gene Mutation Database (HGMD) Professional Version 2016.1.[14] "HCM-associated high-frequency variants" were defined as variants with minor allele frequency (MAF) greater than 1% in either NHLBI subpopulation.

Statistical and bioinformatics analyses

P values were computed using the chi-squared test and Mann-Whitney *U*-test. The HGDP Selection Browser[19] was used to display allele frequencies in worldwide populations. The "penetrance" of a genetic variant is defined as the proportion of individuals with the variant who have HCM. Unless otherwise specified, all analyses were performed using the R statistical package.[20]

RESULTS

HCM gene variation in the general population

The NHLBI Exome Sequence Project (ESP) has previously been searched for any variant labeled a "Disease causing mutation" ("DM") for HCM in the Human Gene Mutation Database (HGMD Version 2012.2).[13,14] Although 94 distinct variants previously associated with HCM were found in the ESP data, we observed that relatively few variants account for the bulk of the genotype prevalence signal (Figure 1A). Five of the ninety-four HGMD HCM-associated variants identified in the ESP data met our threshold to be "HCM-associated high-frequency variants" (MAF > 1% in either NHLBI subpopulation), and accounted for nearly 75% of the overall genotype prevalence signal.

HCM-associated high-frequency variants in African Americans

All five HCM-associated high-frequency variants had significantly greater frequencies in African Americans than in European Americans (Figure 1B, Chi-squared $P < 0.001$ for each comparison). The minor allele frequency for these five variants ranged from 1.5% to 14.9% in African Americans, 0.01% to 1.5% in European Americans, and 0.5% to 6.0% in the combined population. The genotype frequency, defined as (the number of individuals with at least one copy of the minor allele)/(the total number of individuals), ranged from 2.9% to 27.1% in African Americans, 0.02% to 2.9% in European Americans, and 1.0% to 11.1% in the combined population. The summed genotype frequency of the remaining 89 variants was not statistically different between African Americans and European Americans.

Penetrance of HCM-associated high-frequency variants in African Americans

We computed the penetrance for each variant across several clinical contexts (Figure S1). Because HCM occurs rarely in the general population (Panel A) with a prevalence of 1 in 500 individuals,[2] even variants with minor allele frequency as small as 1% have a theoretical maximal penetrance of 0.2, but likely much lower, given the high allelic heterogeneity of HCM,[4] and are likely benign. Even if the *TNNT2* K247R variant were present in all African Americans with HCM (which is certain not to be the case), *K247R* would have a penetrance of less than 1%. Penetrance may take on rather different values in other clinical contexts (Panels B, C). Notably, for first-degree relatives, benign high-frequency variants may have large apparent penetrance owing to the high prior probability of disease in this population.

HCM-associated high-frequency variants are benign

Applying the clinical classification algorithm in use at the Laboratory for Molecular Medicine (LMM), Partners HealthCare Personalized Medicine,[9] we classified all high-frequency variants unambiguously as "Benign," given their elevated frequency in control populations as well as the mix of patient and functional data available for these variants. By contrast, in the HGMD database version 2016.1, four of the five variants remain classified in the most pathogenic category, "Disease causing mutation." Only one variant (*OBSCN* R4344Q) was downgraded from "disease-causing" to "disease-causing?" in September, 2012.

Benign variants misclassified as pathogenic in genetic reports

Seven patients, all of African or unspecified ancestry, received reports between 2005 and 2007 that one of the two benign variants *TNNI3* (P82S) or *MYBPC3* (G278E) was misclassified as "Pathogenic" or "Presumed Pathogenic" (Table 1). In five of the seven reports, P82S or G278E was the most pathogenic variant reported to the patient. Six additional inconclusive and positive cases reported later listed one of these two variants as of "Unknown Significance" or "Pathogenicity Debated." Nine patients (of 13 total) had a clinical diagnosis of HCM, two had clinical features of HCM, and one had clinical symptoms of HCM. Five of 13 patients had a documented family history of HCM. From the records available, it was not possible to determine whether the families affected by these reclassifications were re-contacted.

Sample size and representativeness of original studies

All high-frequency variants were examined for their initial association in the medical literature (Table 2).[21–25] In the initial studies of *TNNI3* P82S and *MYBPC3* G278E, control sample sizes were 85 and 100, which are below and equal to, respectively, a minimum currently accepted (but likely still inadequate) standard needed to corroborate pathogenicity.[4] Furthermore, none of the studies implicating these variants were undertaken in persons of African ancestry explicitly; however, several studies might have sequenced or genotyped persons of African ancestry during the discovery stage (Table 2). Generally, the original study that established the variant-HCM association consisted of three steps. First, a handful of genes

previously connected with HCM were sequenced in DNA samples obtained from HCM patients. Second, discovered variants were examined in ostensibly ancestry-matched unrelated controls and, where available, family members. Third, functional analyses were conducted in a subset of studies to assess causality of the variant.

Variation in *MYBPC3* and *TNNI3* in African Americans

We used the 1000 Genomes Project (1000G) data to compare sequence variation between African Americans and European Americans, using as proxies the populations ASW (Americans of African Ancestry in SW USA) and CEU (Utah Residents (CEPH) with Northern and Western European ancestry), respectively. As shown in Figures 2A and 2B, African Americans harbor significantly more segregating loci (variable genomic sites within a population) than European Americans in both genes. These "private sites," where the minor allele frequency (MAF) is non-zero in one population but zero in the other population, are represented for ASW by the red points in Figures 2A and 2B. There are 66 (ASW) compared to 15 (CEU) private sites for *MYBPC3* and 45 (ASW) compared to 6 (CEU) private sites for *TNNI3*.

Genetically diverse populations reduce the risk of false positives

As shown in Figure 2C, even small studies of diverse populations are comparatively well-powered to avoid misclassifying the five HCM-associated high-frequency variants. Conservatively, we used the lower-frequency variant of the two that were misclassified in patients (*MYBPC3* G278E, MAF 0.0157 in African Americans, 0.000122 in European Americans). At these frequencies, even if African Americans

constituted just 10% of the control cohort, we would have a 50% chance of correctly ruling out pathogenicity with a control cohort of only 200 individuals.

We documented how the HCM-associated high-frequency variants could be studied in worldwide populations (Figure 2D, Table S1). For example, the highest-frequency HCM variant (*TNNT2* K247R) was a locus in the Human Genome Diversity Project (HGDP)[16] (Figure 2D), and has non-zero minor allele frequency in many worldwide populations.

Errors from paucity of diverse control data

Table 3 shows the probability of ruling out pathogenicity for truly benign variants using several sequencing resources that have been in use for the past several years. For example, using the 1000G population "Mexican Ancestry from Los Angeles" (MXL), which consists of 66 individuals, we have only a 50% chance of ruling out pathogenicity when the MAF is 0.5%. If MAF is 0.1%, such as for a rare variant discovered on high-coverage exome sequencing, the probability of ruling out pathogenicity is only 12% using the MXL population. By contrast, if new variant association studies use a resource like the recently created ExAC database,[26] they are well-powered even for rarer variation.

DISCUSSION

We hypothesized that genetic variants common in African Americans were previously misclassified in patients receiving genetic testing for hypertrophic cardiomyopathy. We identified multiple individuals, all of African or unspecified ancestry, who had received positive reports based on misclassified benign variants. Such misclassifications invalidate risk assessments undertaken in relatives, requiring a chain of amended reports and management plans. Our findings suggest that false-positive reports are an important and perhaps underappreciated component of the "genotype-positive/phenotype-negative" subset of tested individuals.[27] These findings demonstrate how health disparities may arise from genomic misdiagnosis. Disparities may result from errors that are neither related to access to care nor to posited "physiological differences," but, rather, the historical dearth of diverse individuals used in control populations. Future work is needed to assess the extent to which this pattern holds across other variants, types of misclassifications, and diseases.

Minimizing misclassifications by sifting through genomic noise for causal variants is closely related to assessing penetrance, the proportion of individuals with the variant who express disease. However, estimating penetrance is often difficult because it is sensitive to clinical context (Figure S1) and because many studies are designed for variant discovery, not unbiased estimation of true effect sizes, a pattern not limited to HCM.[28] This approach is due, in part, to historically limited sequencing data. Fortunately, recent large-scale sequencing efforts are mitigating this aspect of the variant annotation challenge,[12,15] although also

introducing an unprecedented scale of novel variants and genes to consider.[8,29] Even when penetrance cannot be computed precisely, we may still be able to bound the quantity in order to infer that a variant is benign, given a sufficiently elevated minor allele frequency in the general population.[9] While the NHLBI Exome Sequence Project is a powerful resource in its supply of exome sequence data for African Americans and European Americans, analyses using comparable resources of genomic data from Native Americans and Asian Americans are urgently needed (Table 3). Large-scale sequencing resources such as the NHLBI ESP and ExAC are not only well-powered to "rule out" benign variants and reduce false positives (Figure 2C), but also allow pathogenicity to be corroborated for truly pathogenic variants (help "rule in" variants in addition to reducing false negatives). Indeed, diverse population sequencing data are necessary to find ancestry-specific pathogenic variants.[30] Additionally, diverse population sequencing data have been used successfully in "admixture mapping" to find risk loci by detecting deviations in local ancestry.[31]

Large-scale sequencing data from the general population also enable systematic reassessments of previous disease-variant associations.[13,32,33] For variant interpretations in HCM, expert guidelines generally recommend using ancestry-matched controls.[4] Ironically, insistence on using only ancestry-matched controls may delay proper annotation if matching is imperfect. As an example, consider *MYBPC3* G278E, an HCM-associated high-frequency variant that was first discovered in a Parisian cohort,[24] also identified in subsequent studies,[34] and misclassified in several African ancestry individuals (Table 1). We were able to verify that although

most cases were of European origin in the original study,[24] the cases also included individuals of African ancestry (personal communication, P. Richard). If only or primarily European-ancestry persons were used in the original study's control sample or in follow-up studies, analyses would be underpowered to classify the variant as non-pathogenic (Figure 2C). These findings suggest how current guidelines might be extended—sequencing data from diverse ancestry groups may be used to refute the possibility that novel variants are pathogenic, and to reevaluate the status (with respect to pathogenicity) of known variants.[9,35]

Several steps are recommended to improve care going forward. First, sequencing data from diverse populations should be used to evaluate novel variants and reevaluate known variants. Second, if researchers and genetic testing laboratories adopt a probabilistic framework by computing relative risks explicitly, much of the confusion in pathogenicity assessments would likely be reduced. Third, reevaluating the fragmented disease-variant literature depends on continued data-sharing and reporting standardization that are the aims of centralized databases like ClinVar.[36] In line with President Obama's Precision Medicine Initiative, we support the development of an Information Commons[37] that stores anonymized genetic testing results from patients and family members.[38] This effort would demonstrate the value of responsible data sharing, as detailed in a recent Institute of Medicine report.[39,40] Lastly, as variant annotations are updated, agile clinical systems would ideally make this information available in near real time to physicians and genetic testing laboratories.[41] Many physicians express a desire for additional guidance on genetic testing.[42] Point-of-care support for physicians and genetic testing

professionals such as risk calculators for each HCM variant, incorporating family history and ethnicity, would assist in decision making and help minimize misclassification errors. We expect that many "variants of uncertain significance"[8] will be recategorized in the near future with ongoing efforts to sequence individuals from diverse populations. Far from being a clear binary decision, variant classification is an evolving process. A synergy of clinical, genetic, statistical and political perspectives is needed to ensure that genomic medicine benefits all populations equally.

References

1. Maron BJ. Hypertrophic Cardiomyopathy. JAMA 2002;287(10):1308–20.
2. Maron BJ, Gardin JM, Flack JM, Gidding SS, Kurosaki TT, Bild DE. Prevalence of Hypertrophic Cardiomyopathy in a General Population of Young Adults : Echocardiographic Analysis of 4111 Subjects in the CARDIA Study. Circulation 1995;92(4):785–9.
3. Maron BJ, Maron MS. Hypertrophic cardiomyopathy. Lancet 2013;381(9862):242–55.
4. Maron BJ, Maron MS, Semsarian C. Genetics of hypertrophic cardiomyopathy after 20 years: clinical perspectives. J Am Coll Cardiol 2012;60(8):705–15.
5. Weidemann F, Niemann M, Breunig F, et al. Long-term effects of enzyme replacement therapy on fabry cardiomyopathy: evidence for a better outcome with early treatment. Circulation 2009;119(4):524–9.
6. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. Genet Med 2008;10(4):294–300.
7. Richards S, Aziz N, Bale 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 2015;17(5):405–23.
8. Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. Nat Rev Genet 2013;14(4):295–300.
9. Duzkale H, Shen J, McLaughlin H, et al. A systematic approach to assessing the clinical significance of genetic variants. Clin Genet 2013;84(5):453–63.
10. Norton N, Robertson PD, Rieder MJ, et al. Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. Circ Cardiovasc Genet 2012;5(2):167–74.
11. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science 2012;335(6070):823–8.
12. NHLBI GO Exome Sequencing Project (ESP). Exome Variant Server [Internet]. [cited 2015 Jul 25];Available from: http://evs.gs.washington.edu/EVS/
13. Andreasen C, Nielsen JB, Refsgaard L, et al. New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. Eur J Hum Genet 2013;21(9):918–28.
14. Stenson PD, Mort M, Ball E V, et al. The Human Gene Mutation Database: 2008 update. Genome Med 2009;1(1):13.
15. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491(7422):56–65.
16. Li JZ, Absher DM, Tang H, et al. Worldwide human relationships inferred from

genome-wide patterns of variation. Science 2008;319(5866):1100–4.

17. Alfares AA, Kelly MA, McDermott G, et al. Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity. Genet Med 2015;
18. Sherry ST. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29(1):308–11.
19. Pickrell JK, Coop G, Novembre J, et al. Signals of recent positive selection in a worldwide sample of human populations. Genome Res 2009;19(5):826–37.
20. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. J Comput Graph Stat 2012;
21. Garcia-Castro M. Hypertrophic Cardiomyopathy: Low Frequency of Mutations in the -Myosin Heavy Chain (MYH7) and Cardiac Troponin T (TNNT2) Genes among Spanish Patients. Clin Chem 2003;49(8):1279–85.
22. Arimura T, Matsumoto Y, Okazaki O, et al. Structural analysis of obscurin gene in hypertrophic cardiomyopathy. Biochem Biophys Res Commun 2007;362(2):281–7.
23. Niimura H. Sarcomere Protein Gene Mutations in Hypertrophic Cardiomyopathy of the Elderly. Circulation 2002;105(4):446–51.
24. Richard P, Charron P, Carrier L, et al. Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. Circulation 2003;107(17):2227–32.
25. Matsushita Y, Furukawa T, Kasanuki H, et al. Mutation of junctophilin type 2 associated with hypertrophic cardiomyopathy. J Hum Genet 2007;52(6):543–8.
26. Exome Aggregation Consortium (ExAC) [Internet]. [cited 2015 May 5];Available from: http://exac.broadinstitute.org
27. Maron BJ, Yeates L, Semsarian C. Clinical challenges of genotype positive (+)-phenotype negative (-) family members in hypertrophic cardiomyopathy. Am J Cardiol 2011;107(4):604–8.
28. Beutler E, Felitti VJ, Koziol JA, Ho NJ, Gelbart T. Penetrance of 845G--> A (C282Y) HFE hereditary haemochromatosis mutation in the USA. Lancet (London, England) 2002;359(9302):211–8.
29. Keinan A, Clark AG. Recent explosive human population growth has resulted in an excess of rare genetic variants. Science 2012;336(6082):740–3.
30. Struewing JP, Hartge P, Wacholder S, et al. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 1997;336(20):1401–8.
31. Seldin MF, Pasaniuc B, Price AL. New approaches to disease mapping in admixed populations. Nat Rev Genet 2011;12(8):523–8.
32. Pugh TJ, Kelly MA, Gowrisankar S, et al. The landscape of genetic variation in

dilated cardiomyopathy as surveyed by clinical DNA sequencing. Genet Med 2014;16(8):601–8.

33. Bick AG, Flannick J, Ito K, et al. Burden of rare sarcomere gene variants in the Framingham and Jackson Heart Study cohorts. Am J Hum Genet 2012;91(3):513–9.
34. Morita H, Rehm HL, Menesses A, et al. Shared genetic causes of cardiac hypertrophy in children and adults. N Engl J Med 2008;358(18):1899–908.
35. Ioannidis JPA, Ntzani EE, Trikalinos TA. "Racial" differences in genetic effects for complex diseases. Nat Genet 2004;36(12):1312–8.
36. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 2014;42(Database issue):D980–5.
37. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington, D.C.: National Academies Press; 2011.
38. Manrai AK, Ioannidis JPA, Kohane IS. Clinical Genomics: From Pathogenicity Claims to Quantitative Risk Estimates. JAMA 2016;315(12):1233–4.
39. Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. JAMA 2015;313(8):793–4.
40. Drazen JM. Sharing Individual Patient Data from Clinical Trials. N Engl J Med 2015;372(3):201–2.
41. Wilcox AR, Neri PM, Volk LA, et al. A novel clinician interface to improve clinician access to up-to-date genetic results. J Am Med Inform Assoc 2014;21(e1):e117–21.
42. Klitzman R, Chung W, Marder K, et al. Attitudes and practices among internists concerning genetic testing. J Genet Couns 2013;22(1):90–100.

Table 1: Clinical findings for HCM-associated high-frequency variants

Age	Ethnicity	Report Year	Report	Variant	Originally Reported Status	Current Status	Most Significant?	Indication for Test
46	Unavailable	2005	Positive	Pro82Ser	Pathogenic	Benign	Y	Clinical Diagnosis of HCM
75	Unavailable	2005	Positive	Pro82Ser	Pathogenic	Benign	Y	Family History and Clinical Symptoms of HCM
32	Black or African American	2005	Positive	Pro82Ser	Presumed Pathogenic	Benign	N	Clinical Diagnosis of HCM
34	Black or African American	2005	Positive	Pro82Ser	Pathogenicity Debated	Benign	N	Clinical Diagnosis and Family History of HCM
12	Black or African American	2006	Inconclusive	Pro82Ser	Unknown Significance	Benign	Y	Family History of HCM
40	Black or African American	2007	Inconclusive	Pro82Ser	Unknown Significance	Benign	Y	Clinical Diagnosis of HCM
45	Black or African American	2007	Inconclusive	Pro82Ser	Unknown Significance	Benign	Y	Clinical Features of HCM
16	Asian	2008	Positive	Pro82Ser	Unknown Significance	Benign	N	Clinical Diagnosis and Family History of HCM
59	Black or African American	2006	Positive	Gly278Glu	Presumed Pathogenic	Benign	Y	Clinical Features of HCM
15	Black or African American	2007	Positive	Gly278Glu	Presumed Pathogenic	Benign	Y	Clinical Diagnosis of HCM
16	Black or African American	2007	Positive	Gly278Glu	Presumed Pathogenic	Benign	Y	Clinical Diagnosis of HCM
22	Black or African American	2007	Positive	Gly278Glu	Presumed Pathogenic	Benign	N	Clinical Diagnosis and Family History of HCM
48	Black or African American	2008	Positive	Gly278Glu	Unknown Significance	Benign	N	Clinical Diagnosis of HCM

*The "Most Significant?" column indicates whether the variant was unequivocally the most pathogenic variant on the original report provided to the patient.

Table 2: Studies that initially implicated HCM-associated high-frequency variants

Gene (Variant)	*Reference*	Discovery	Cases	Controls	*In vitro*	*In vivo*	Country	LMM Clinical Panel
TNNT2 (K247R)	Garcia-Castro (2003)[21] *Clin Chem* 49, 1279	Targeted gene sequencing of unrelated cases (Asturias)	30	200 (Asturias)	-	-	Spain	+
OBSCN* (R4344Q)	Arimura (2007)[22] *Biochem Biophys Res Commun* 362,281	Targeted gene sequencing of unrelated cases (Japanese)	144	288 (Japanese)	+	-	Japan	-
TNNI3 (P82S)	Niimura (2002)[23] *Circulation* 105, 446	Targeted gene sequencing of unrelated cases**	31	85**	-	-	USA	+
MYBPC3 (G278E)	Richard (2003)[24] *Circulation* 107, 2227	Targeted gene sequencing of unrelated cases***	197	100***	-	-	France	+
JPH2* (G505S)	Matsushita (2007)[25] J *Hum Genet* 52, 543	Targeted gene sequencing of cases (Japanese)	195	236 (Japanese)	+	-	Japan	-

* *OBSCN* and *JPH2* have never been included in cardiomyopathy testing at the LMM.
** No specific ethnicity provided, but "informed consent was obtained in accordance with human subject committee guidelines at Brigham and Women's Hospital, St. George's Hospital Medical School [U.K.], and Minneapolis Heart Institute Foundation."
*** "Patients were recruited in France, and most of them were of European origin." Cases included individuals of African ancestry (personal communication, P. Richard).

Table 3: Control sequencing resources for ruling out pathogenicity in several worldwide populations

Cohort	Population	N	MAF = 1%	MAF = 0.5%	MAF = 0.1%
NHLBI ESP	European Americans	4300	100%	100%	100%
NHLBI ESP	African Americans	2203	100%	100%	99%
1000 Genomes	Mexican Ancestry from Los Angeles, USA	66	73%	48%	12%
1000 Genomes	Han Chinese in Beijing, China	97	86%	62%	18%
1000 Genomes	Southern Han Chinese	100	87%	63%	18%
ExAC	African/African American	5203	100%	100%	100%
ExAC	Non-Finnish European	33370	100%	100%	100%
ExAC	East Asian	4327	100%	100%	100%

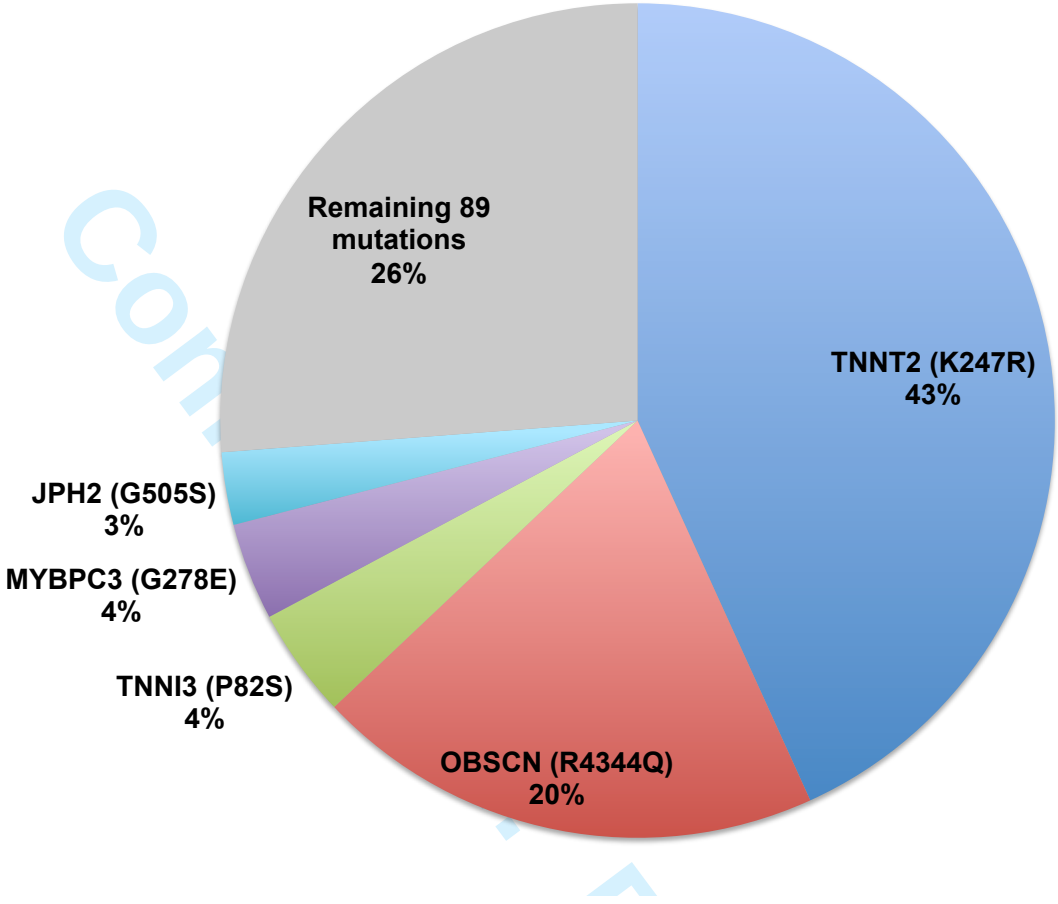
The probability of ruling out pathogenicity for a benign variant with minor allele frequency (MAF) values of 1%, 0.5%, and 0.1% is shown for several example worldwide populations with population size N.

Figure 1A: Overrepresented HCM variants in the general population. The five highest-frequency variants account for 74% of the misclassified HCM variation in the general population. **Figure 1B: All HCM-associated high-frequency variants are significantly more common in African Americans than European Americans.** Chi-squared $P < 0.001$ for the five HCM-associated high-frequency variants.

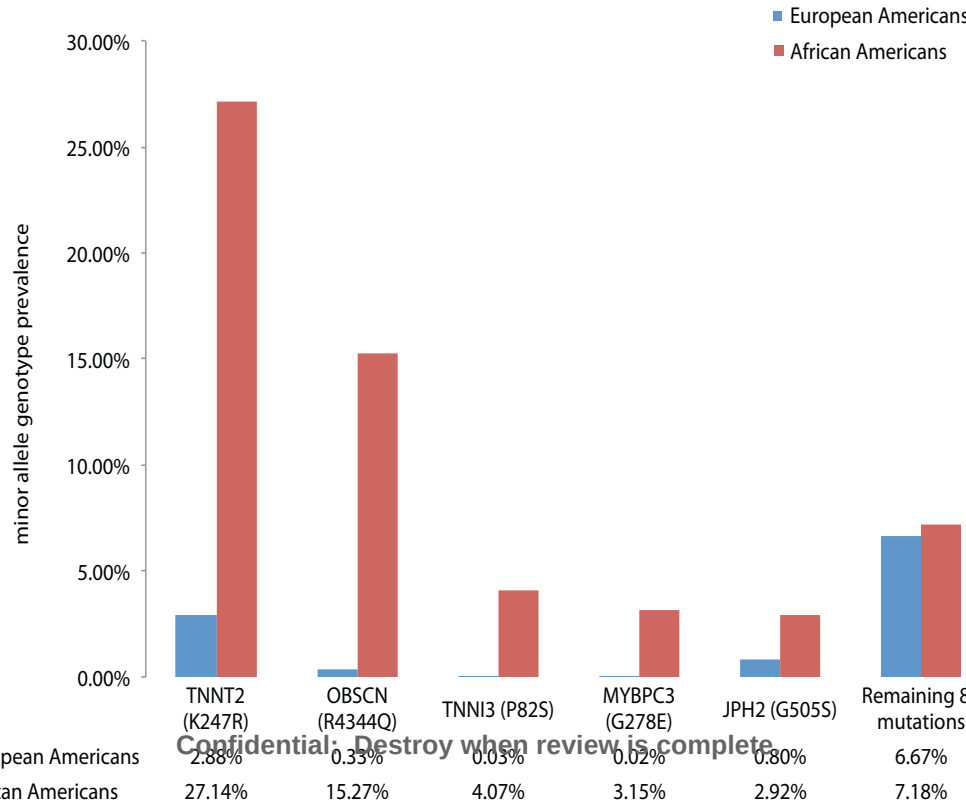
Figure 2: Diverse population sequencing data help prevent variant misclassifications (A/B) Non-reference allele frequency for 1000G populations ASW (y-axis, 61 individuals) and CEU (x-axis, 85 individuals) for the HCM genes *MYBPC3* and *TNNI3*. Each point represents a distinct variant. African Americans have significantly more private variants (CEU MAF = 0% and ASW MAF > 0%, colored in red) than European Americans (ASW MAF = 0% and CEU MAF > 0%). **(C)** For a variant predominantly found in one ancestry group, the chance of correctly ruling out pathogenicity for a truly benign variant generally increases with the fraction of the control cohort comprised of that ancestry group and the number of controls (control chromosomes shown in legend). These simulations use the allele frequencies of the *MYBPC3* G278E variant, which has an African American minor allele frequency (MAF) of 0.0157 and a European American MAF of 0.000122. **(D)** *TNNT2* (K247R) was a variant genotyped in the HGDP. Most populations around the world have non-zero minor allele frequency.

Figure 1

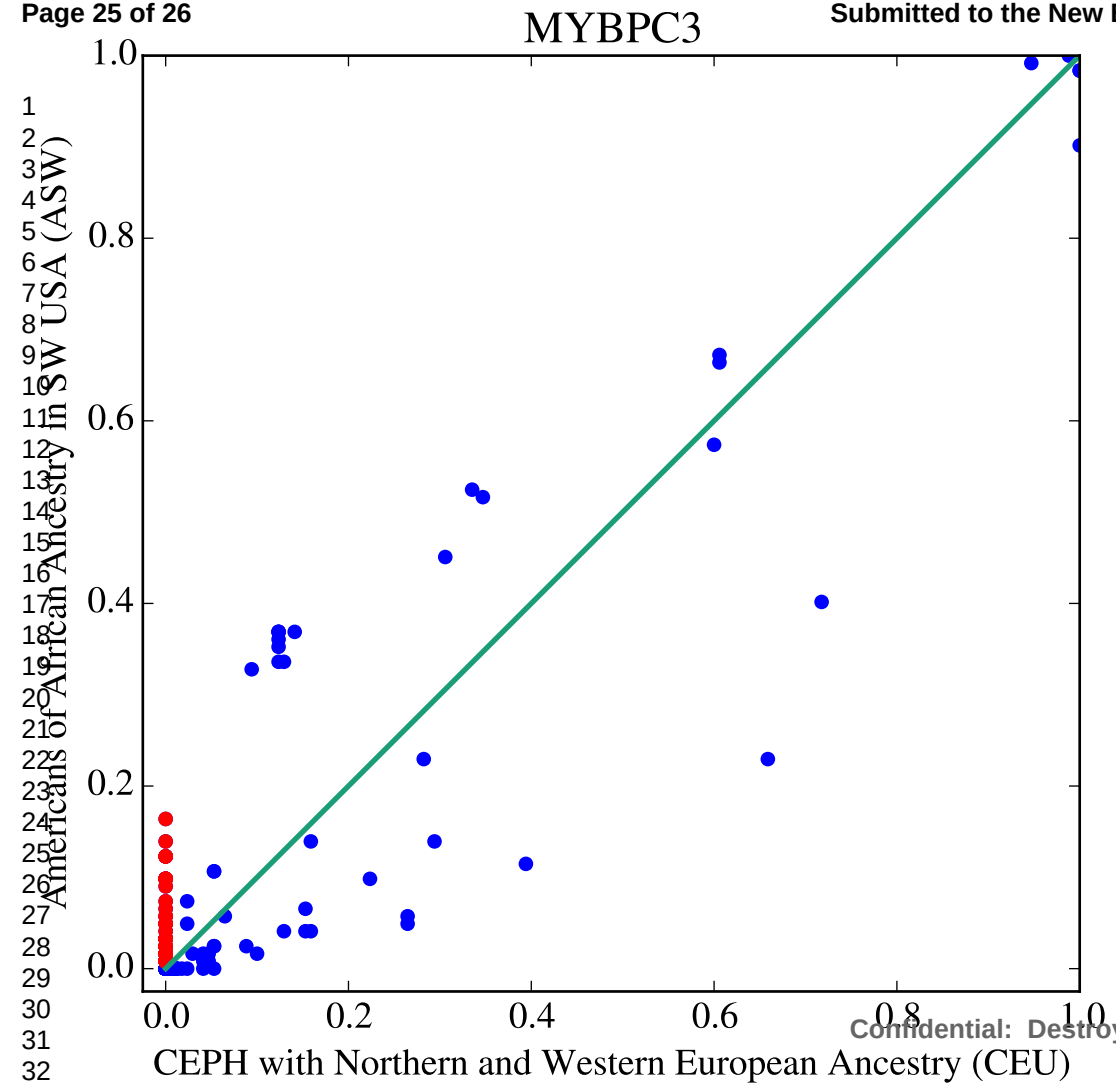
(A)

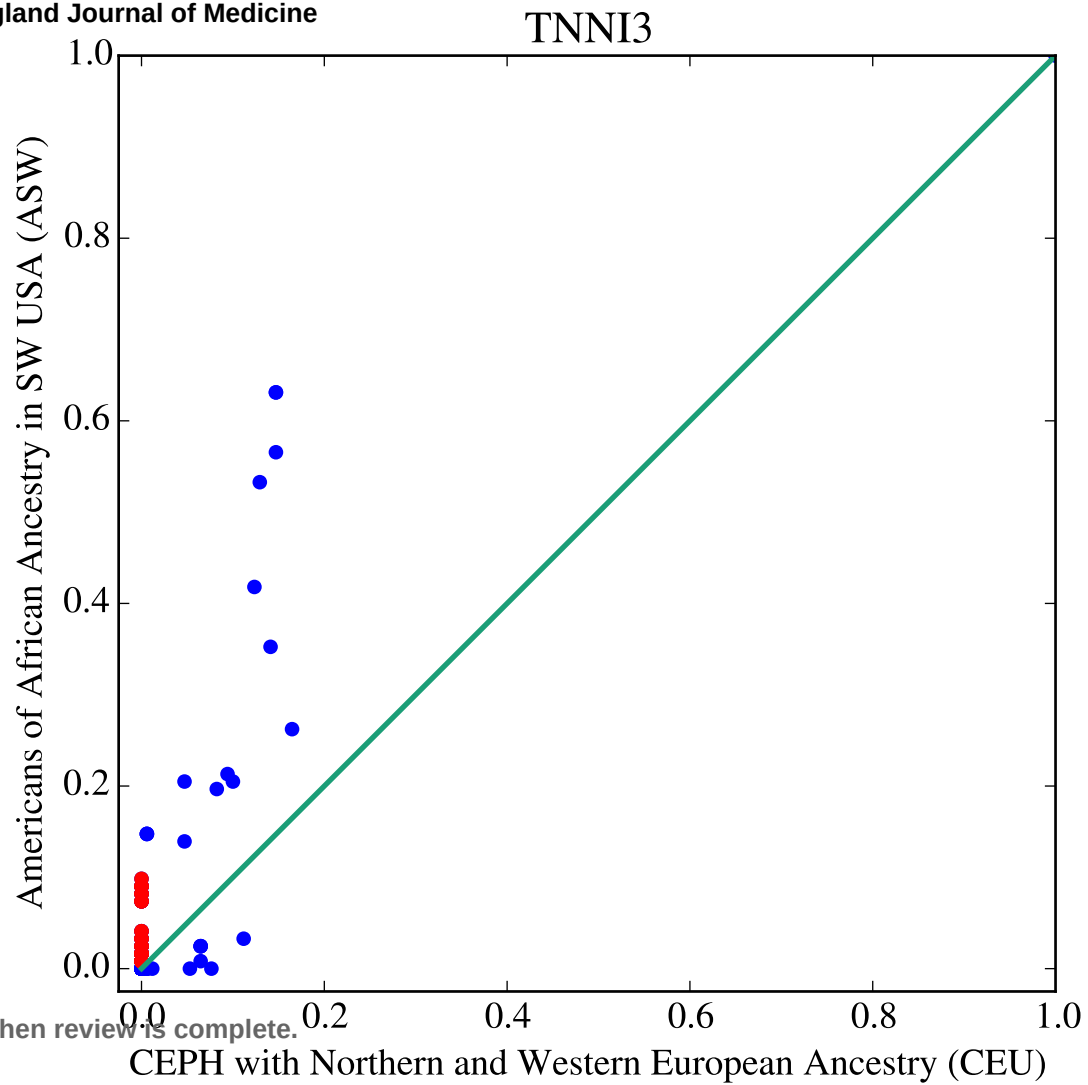


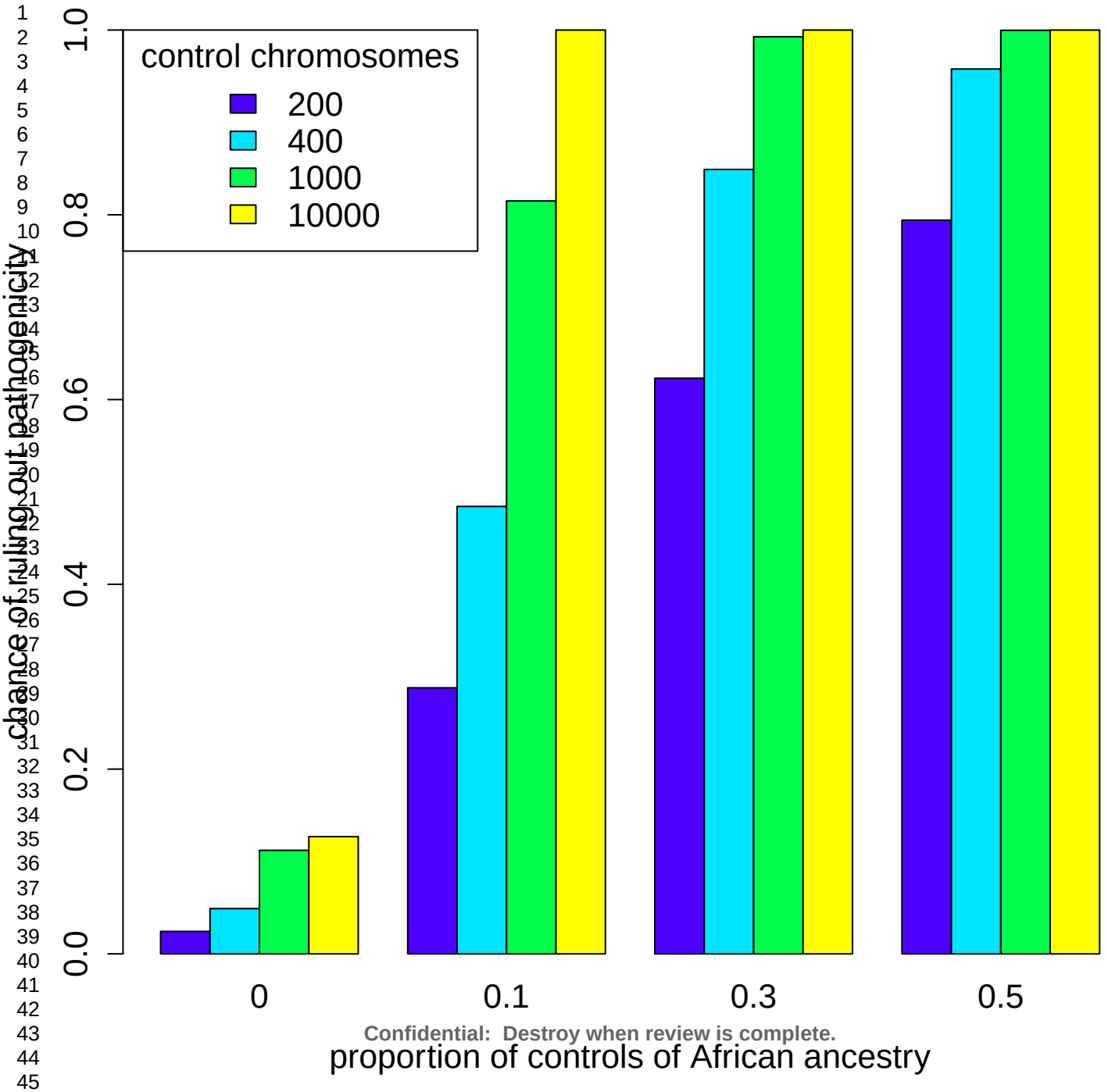
(B)



	TNNT2 (K247R)	OBSCN (R4344Q)	TNNI3 (P82S)	MYBPC3 (G278E)	JPH2 (G505S)	Remaining 89 mutations
European Americans	2.88%	0.33%	0.03%	0.02%	0.80%	6.67%
African Americans	27.14%	15.27%	4.07%	3.15%	2.92%	7.18%







Submitted to the New England Journal of Medicine