ARTICLE | Genetics inMedicine

A novel statistical method for interpreting the pathogenicity of rare variants

Jun Wang, PhD[1,2], Hehe Liu, MMed[1,2], Renae Elaine Bertrand, BSc[1,3], Alejandro Sarrion-Perdigones, PhD[3], Yezabel Gonzalez, MSc[3], Koen J. T. Venken, PhD[3,4] and Rui Chen, PhD[1,2]

Purpose: To achieve the ultimate goal of personalized treatment of patients, accurate molecular diagnosis and precise interpretation of the impact of genetic variants on gene function is essential. With sequencing cost becoming increasingly affordable, the accurate distinguishing of benign from pathogenic variants becomes the major bottleneck. Although large normal population sequence databases have become a key resource in filtering benign variants, they are not effective at filtering extremely rare variants.

Methods: To address this challenge, we developed a novel statistical test by combining sequencing data from a patient cohort with a normal control population database. By comparing the expected and observed allele frequency in the patient cohort, variants that are likely benign can be identified.

Results: The performance of this new method is evaluated on both simulated and real data sets coupled with experimental validation. As a result, we demonstrate this new test is well powered to identify benign variants, and is particularly effective for variants with low frequency in the normal population.

Conclusion: Overall, as a general test that can be applied to any type of variants in the context of all Mendelian diseases, our work provides a general framework for filtering benign variants with very low population allele frequency.

Genetics in Medicine (2021) 23:59–68; https://doi.org/10.1038/s41436-020-00948-3

Keywords: variant interpretation; allele frequency; Mendelian diseases; statistical test; clinical genomics

INTRODUCTION

The advancement of high throughput sequencing technology significantly facilitates the identification of genetic variations in individuals and populations. However, the determination of the pathogenicity of genetic variants upon sequencing remains a major challenge for precision medicine. Over the last two decades, many in silico variant functional prediction tools have been developed to distinguish pathogenic from likely benign genetic variants, but they are far from perfect. For example, the methods based on evolutionary sequence conservation might be prone to both false positives and false negatives as many benign variants also occur in the evolutionarily conserved regions and vice versa. In parallel, many computational prediction methods that apply machine learning algorithms are limited by the training data set and the type of information incorporated into the model. Furthermore, most of the computational prediction methods have focused on missense variants, leaving other types of variants, such as INDEL and noncoding variants, largely unexplored. Recently, with genome sequences of large populations becoming available, a significant proportion of benign variants can be identified based on population allele frequency (AF) and disease prevalence, given that variants with high AF in the normal population tend to be benign.[1,2] Similarly, by comparing AF between a patient cohort and a normal control population, variants that are enriched in the patient cohort and therefore likely pathogenic can be identified.[3] However, these AF-based methods have limited power when the AF of the variant is low in the normal population.

In this study, we developed a novel statistical method to identify likely benign variants. Briefly, in contrast to previous methods that primarily use the AF in the normal population, our method calculates the expected frequency of an allele in the patient cohort by taking into account the disease prevalence and the AF in the normal population. By comparing the expected and the observed frequency of the variant in the patient cohort, the probability that the variant is pathogenic can be calculated. To test the new statistics, we have applied it to both simulated and real data sets and evaluated its performance based on literature and other variant prediction methods. A subset of the variants with prediction contradictory to previous reports were further examined with experimental functional assays. Based on our results, the model has significantly improved power to

[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]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX, USA; [4]Department of Pharmacology and Chemical Biology, Baylor College of Medicine, Houston, TX, USA. Correspondence: Rui Chen (ruichen@bcm.edu)
Deceased: Alejandro Sarrion-Perdigones

Submitted 22 January 2020; revised 11 August 2020; accepted: 12 August 2020
Published online: 4 September 2020

annotate rare variants, and it can be applied for all variant types such as single-nucleotide polymorphisms (SNPs), INDELs, and noncoding variants. Overall, we show that our new method is effective and robust as a general framework in evaluating the pathogenicity of variants in the context of Mendelian diseases.

MATERIALS AND METHODS

Ethics statement

The studies have approval from the institutional review board (IRB) of Baylor College of Medicine.

Derivation of the combined test

The combined test is composed of two binomial tests.

Let Q = the disease prevalence due to the variants in gene g in a population.

q_k = the AF of a single pathogenic variant a_k of the gene g in a population. Assume that n patients are randomly sampled from the patients attributable to the variants in gene g.

For a recessive disease gene g, the expected number of pathogenic alleles in n patients would be 2^*n. Among the 2^*n pathogenic alleles, the expected occurrence count of a pathogenic allele a_k should follow a binomial distribution with $N = 2^*n$ trials and the occurrence frequency (success rate) of $\frac{q_k}{\sqrt{Q}}$.

The first test (test1) is a left-tailed $Binomial.test\,(X = x, N = 2n, p = \frac{q_k}{\sqrt{Q}})$:

H0 of the test1: The allele is pathogenic, thus its observed occurrence in patient cohort follows a $Binomial\,(N = 2n, p = \frac{q_k}{\sqrt{Q}})$ distribution.

H1 of the test1: The observed occurrence of the allele in patient cohort does not follow $Binomial\,(N = 2n, p = \frac{q_k}{\sqrt{Q}})$, significantly lower than $2n \times \frac{q_k}{\sqrt{Q}}$. Therefore, it is unlikely to be pathogenic (detailed derivation in Supplementary Methods).

Additionally, if an allele is likely benign, its AF in the patient cohort should be similar to its AF in the normal population, given it has equal association with patients and normal population.

Thus, the second test (test2) is a right-tailed $Binomial.test\,(X = x, N = 2n, p = q_k)$:

H0 of the test2: The allele is benign, thus its observed occurrence in the patient cohort follows a $Binomial\,(N = 2n, p = q_k)$ distribution.

H1 of the test2: The observed occurrence of the allele in the patient cohort does not follow $Binomial\,(N = 2n, p = q_k)$, significantly higher than $2n \times q_k$. Therefore, it is unlikely to be benign.

The combined test result is based on the results of test1 and test2.

H0 of the combined test: The allele is pathogenic, thus test1 H0 is true (test1 p value > 0.05) or test2 H0 is rejected (test2 p value ≤ 0.05).

H1 of the combined test: The test1 H1 is true (test1 p value ≤ 0.05) and test2 H0 is true (test2 p value > 0.05), thus the allele is unlikely to be pathogenic (i.e., likely benign).

For a dominant disease gene g, the expected number of pathogenic alleles in n patients (with the rare dominant disease) ≈n. Among the n pathogenic alleles, the expected occurrence count of a pathogenic allele a_i should follow a binomial distribution with $N = n$ trials and the occurrence frequency (success rate) of $\frac{q_i}{1-\sqrt{1-Q}}$.

Test1 is a left-tailed $Binomial.test\,(X = x, N = n, p = \frac{q_k}{1-\sqrt{1-Q}})$:

H0 of the test1: The allele is likely pathogenic, thus its observed occurrence in the patient cohort follows a $Binomial\,(N = n, p = \frac{q_k}{1-\sqrt{1-Q}})$ distribution.

H1 of the test1: The observed occurrence of the allele in the patient cohort does not follow $Binomial\,(N = n, p = \frac{q_k}{1-\sqrt{1-Q}})$, significantly lower than $n \times \frac{q_k}{1-\sqrt{1-Q}}$. Therefore, it is unlikely to be pathogenic (Supplementary Methods).

Additionally, if an allele is benign, its AF in the patient cohort should be similar to its AF in the normal population, given it has equal association with patients and normal population.

Test2 is a right-tailed $Binomial.test\,(X = x, N = n, p = 2 \times q_k)$:

H0 of the test2: The allele is likely benign, thus its observed occurrence in the patient cohort follows a $Binomial\,(N = n, p = 2 \times q_k)$ distribution.

H1 of the test2: The observed occurrence of the allele in the patient cohort does not follow $Binomial\,(N = n, p = 2 \times q_k)$, significantly higher than $n \times 2 \times q_k$. Therefore, it is unlikely to be benign.

The combined test result is based on the results of test1 and test2.

H0 of the combined test: The allele is likely pathogenic, thus test1 H0 is true (test1 p value > 0.05) or test2 H0 is rejected (test2 p value ≤ 0.05).

H1 of the combined test: The test1 H1 is true (test1 p value ≤ 0.05) and test2 H0 is true (test2 p value > 0.05), thus the allele is unlikely to be pathogenic (i.e., likely benign).

The treatment for X-linked genes and for population stratification is in the Supplementary materials.

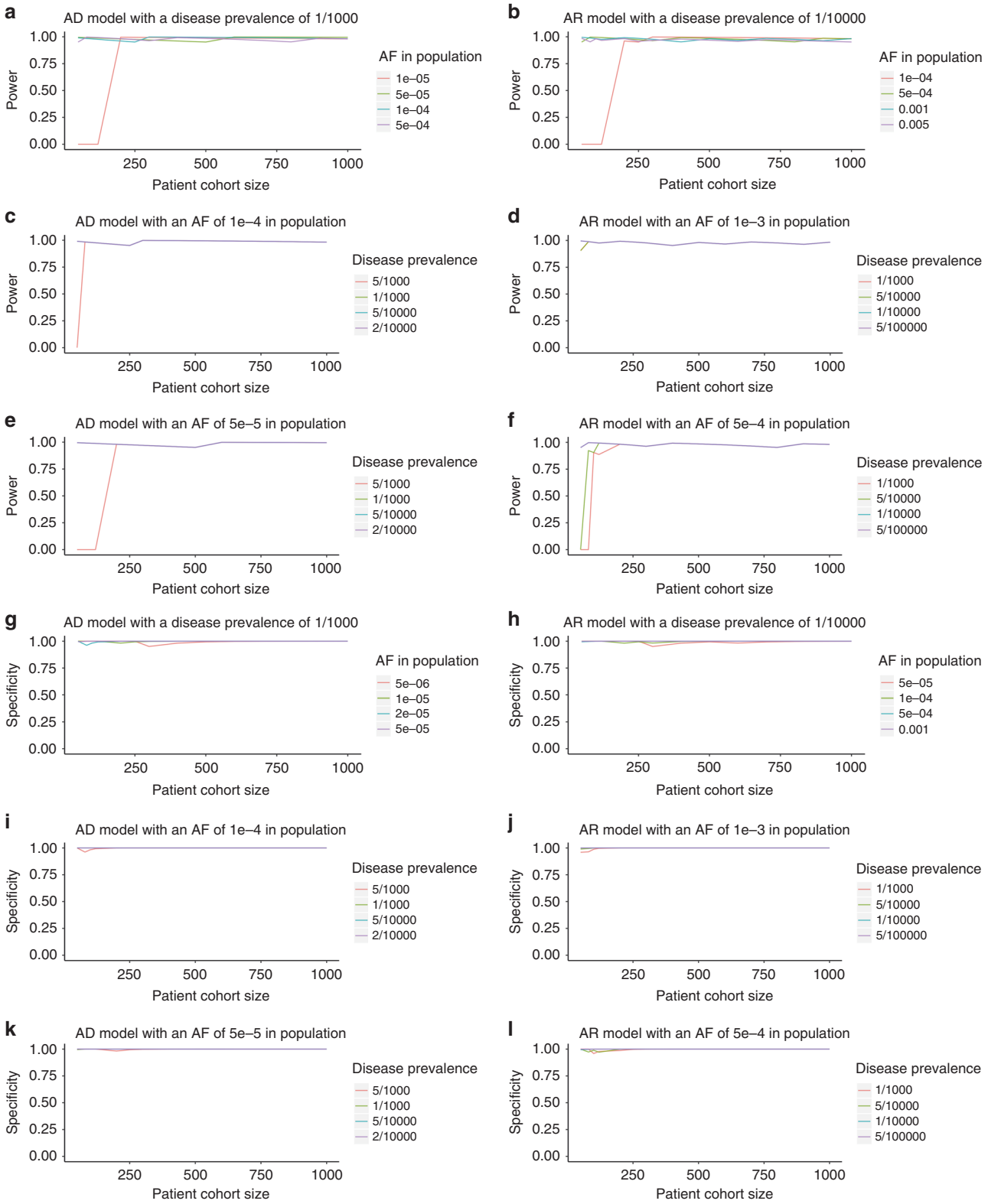
RESULTS

The simulation analysis of the test power and specificity

To assess the performance of the test, we simulated a variety of scenarios with different settings for the patient cohort size (n), the disease prevalence (Q), and the AF in the normal population (q_k).

The test power increases as the patient cohort size increases

The test power is positively correlated with the patient cohort size (Fig. **1a**). Larger sample size of patient cohort will allow the test to achieve high power and specificity in determining the pathogenicity of rarer variants. For example, under the autosomal dominant (AD) model, for a disease with a prevalence of 1/1000, when the patient cohort size is increased from 120 to 200, the test power will increase from 0 to about 100% in detecting rare benign variants with a population AF at 1×10^{-5} (Fig. **1a**, Supplementary Table S1). A similar trend was observed under the AR model (Fig. **1b**, Supplementary Table S1).



The power increases as the disease prevalence decreases

The test power is negatively correlated with the disease prevalence (Fig. **1c–f**, Supplementary Table S1). Namely, the test has higher power for rarer diseases. For example, under the AD model, to detect benign variants with a population AF of 5×10^{-5}, when the associated disease prevalence decreases from 1/200 to 1/1000, the patient cohort size that the test requires to achieve about 100% power will decrease from 200 to 50. A similar trend was observed under the AR model (Supplementary Table S1).

Fig. 1 The simulation analysis of the test power. As the patient cohort size increases, the test will have more power to distinguish the variants that are rare in the normal population under the autosomal dominant (AD) model (**a**) and autosomal recessive (AR) model (**b**). The test has more power to detect the benign variants for the disease with rarer disease prevalence than with relatively common prevalence, for variants with a population allele frequency (AF) of 1×10^{-4} under the AD model (**c**), a population AF of 1×10^{-3} under the AR model (**d**), a population AF of 5×10^{-5} under the AD model (**e**), and a population AF of 5×10^{-4} under the AR model (**f**). The test has a high specificity and is robust. The test specificity is not significantly affected by the AF in the normal population under the AD model (**g**) and AR model (**h**), and is also not significantly affected by the disease prevalence, for variants with a population AF of 1×10^{-4} under the AD model (**i**), a population AF of 1×10^{-3} under the AR model (**j**), a population AF of 5×10^{-5} under the AD model (**k**), and a population AF of 5×10^{-4} under the AR model (**l**).

The test has a high specificity

We found that the test specificity is robust, remaining close to 100%, and is largely unaffected by the AF in the normal population and disease prevalence. As long as the observed frequency of the variant in the patient cohort is greater than or similar to the expected frequency of a pathogenic variant based on the AF in the normal population and disease prevalence, the test will typically not consider the variant as a benign variant (Fig. 1g–l, Supplementary Table S1).

Sampling bias in patient cohort could affect the test power and specificity

When applying the test to detect the variants biasedly enriched in the sampled patient cohort, the test power to detect benign alleles will decrease (Fig. 2, Supplementary Table S2). For example, if the observed frequency of a variant in the sampled patient cohort is fivefold the true AF due to sampling bias, for an AR disease with a disease prevalence of 1/10,000, the test power of detecting variants with a population AF of 5×10^{-4} will decrease to 26% with a patient cohort size of 1000. Similarly, when applying the test to identify the variants artificially depleted in the sampled patient cohort due to sampling bias, the test specificity to detect pathogenic alleles will decrease when the patient sample size is small but will rapidly recover as the sample size increases (Fig. 2, Supplementary Table S2). For example, if the observed frequency of a pathogenic variant in the sampled patient cohort is at 10% of the true AF, for an AD disease with a disease prevalence of 1/1000, the test specificity of detecting a variant with a population AF of 5×10^{-5} will increase from 39% to 92% when the sample size increases from 50 to 250. A similar trend was observed under the AR model.

Misspecification of disease prevalence could affect test performance

As shown in Supplementary Fig. S1, in the case where the disease prevalence is underestimated, the test power to detect benign alleles and test specificity to detect pathogenic alleles are not significantly affected. On the other hand, when the disease prevalence is overestimated, the power of the test to detect benign alleles decreases but rapidly improves with increasing of the sample size, while the specificity of the test to detect pathogenic alleles is not significantly affected (Supplementary Material, Supplementary Fig. S1, Table S3).

The impact of allele penetrance on test performance

Based on simulation, if a disease is attributed to alleles with various penetrance values, the test power to detect benign alleles will not be significantly affected, whereas, test specificity to detect pathogenic alleles will decrease for alleles with low penetrance using small patient cohort sizes but will rapidly improve as the sample size increases (Supplementary Material, Supplementary Fig. S2, Table S4). Additionally, we performed the simulation for the scenario of a disease attributed to multiple pathogenic alleles with heterogeneous penetrance. Under this scenario, the test power to detect benign alleles is not significantly affected, and test specificity to detect pathogenic alleles with low penetrance decreases using small sample sizes but will rapidly improve as the sample size increases (Supplementary Material, Supplementary Fig. S3, Table S4). Overall, the test shows excellent power and specificity for alleles with penetrance ≥50%.

Estimation of the thresholds of population allele frequency for filtering variants

We performed simulation to determine the thresholds of population AF for filtering benign variants for diseases with a variety of disease prevalence under AR/AD models, respectively (Supplementary Materials). As shown in Supplementary Table S5, our test allows the filtering of alleles that are at least ten times less frequent in population compared with applying thresholds based on disease prevalence alone.

Real case studies

To evaluate the performance of our model, we applied our test to variants in three genes, *ABCA4*, *USH2A*, and *LRP5*, to represent three types of scenarios.

Screening of reported pathogenic variants in ABCA4 for Stargardt disease

The disease prevalence of Stargardt disease (STGD) is estimated as 1 in 10,000 individuals.[4] It has been estimated that about 70% of STGD patients carry variants in *ABCA4*.[5] Therefore, this represents the scenario of a recessive disease with a relatively homogeneous genetic cause. We screened 945 reported pathogenic variants in *ABCA4* genes collected in the Human Gene Mutation Database (HGMD). Among them, 11 variants are likely benign, as their population AF is higher than 0.7% ($\sqrt{1/20,000}$), the cutoff based on STGD disease

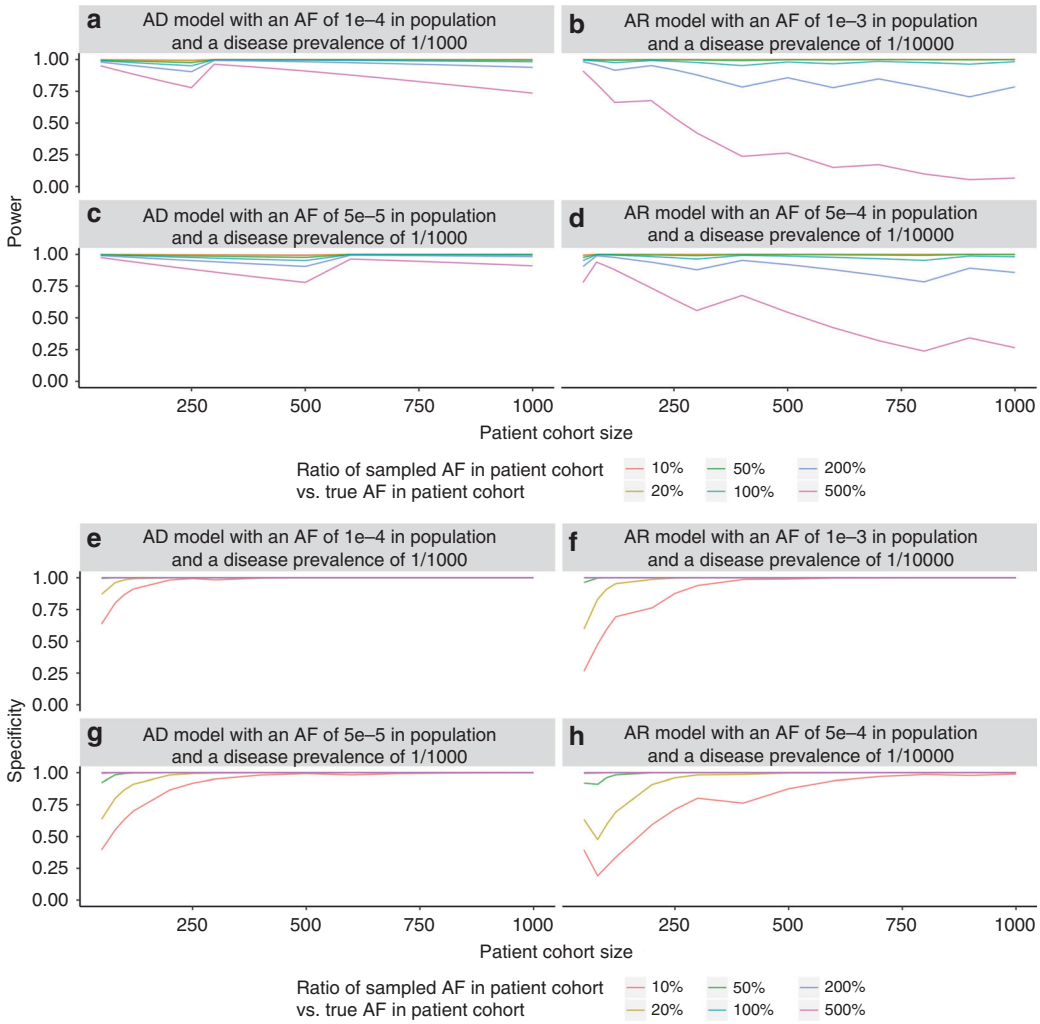


Fig. 2 The simulation analysis of sampling bias. Sampling bias can affect the test performance. The test power decreases for detecting the variants biasedly enriched in the patient sampling, and is not significantly affected for detecting the variants biasedly depleted in the sampling, for variants with a population allele frequency (AF) of 1×10^{-4} under the autosomal dominant (AD) model (**a**), a population AF of 1×10^{-3} under the autosomal recessive (AR) model (**b**), a population AF of 5×10^{-5} under the AD model (**c**), and a population AF of 5×10^{-4} under the AR model (**d**). The test specificity is not significantly affected for detecting the variants biasedly enriched in the patient sampling, and decreases for detecting the variants biasedly depleted in the sampling, which can be rapidly improved as the sample sizes increase, for variants with a population AF of 1×10^{-4} under the AD model (**e**), a population AF of 1×10^{-3} under the AR model (**f**), a population AF of 5×10^{-5} under the AD model (**g**), and a population AF of 5×10^{-4} under the AR model (**h**).

prevalence, and therefore they were excluded from further analysis. The remaining 934 variants were subjected to our test model. As a result, 26 variants with AF in the range of 0.46% to 0.03% were identified as likely benign (binomial test1, Bonferroni correction p value ≤ 0.05/934 and test2 Bonferroni correction p value > 0.05/934) (Fig. 3a).

To examine the test result, we further reviewed the original literatures and searched the genome aggregation database (gnomAD) for the 26 variants. Based on published literature, 8 of the 26 variants should be annotated as benign as they are reported as nonpathogenic polymorphisms in one or more papers (e.g., NM_000350.2:c.455G>A,[6,7] NM_000350.2:c.2701A>G[6]), behave similarly to wild type based on functional assays (e.g., NM_000350.2:c.5693G>A[8]), or do not segregate with the disease (e.g., NM_000350.2:c.4685T>C[7]) in the original report. Consistently, homozygous individuals were observed for five of the eight variants in gnomAD. Additionally, 13 of the 26 variants should be annotated as a variant of uncertain significance (VUS) based on data presented in the original papers due to the lack of definite association of the variant to the disease.[4,6,9–12] Interestingly, 6 of these 13 variants have homozygous individuals observed in gnomAD (Supplementary Tables S6, S7). Finally, 5 of the 26 variants were noted as likely pathogenic in the original reports, given the patients carry a second variant in *ABCA4*, or supported by functional evidence.[7,8,13–19] However, four of the five variants have homozygotes identified in gnomAD with the number ranging from 1 to 68 (Supplementary Tables S6, S7). Therefore, almost all the 26 variants predicted to be benign by our test are indeed unlikely to be pathogenic.

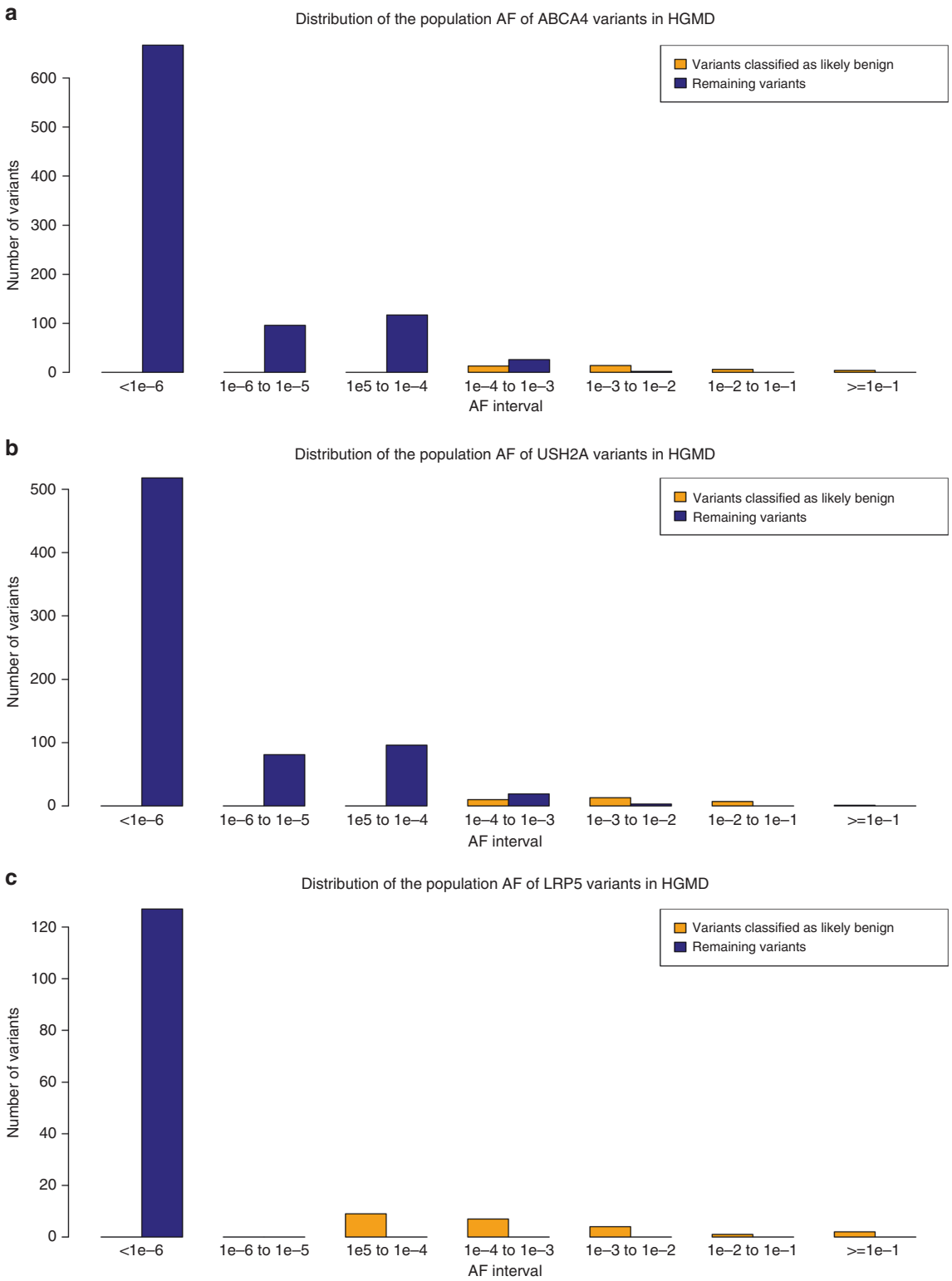


Fig. 3 The distribution of population allele frequency (AF) of the Human Gene Mutation Database (HGMD) variants in three genes. (**a**) The non-Finnish European (NFE) AF of *ABCA4* variants in HGMD, (**b**) the NFE AF of *USH2A* variants in HGMD, and (**c**) the East Asian AF of *LRP5* variants in HGMD. Orange indicates the HGMD variants identified as likely benign by our test. Blue indicates the other HGMD variants.

Screening of reported pathogenic variants in USH2A for retinitis pigmentosa

The disease prevalence of retinitis pigmentosa (RP) is estimated as 1 in 4000 (https://ghr.nlm.nih.gov/condition/retinitis-pigmentosa#statistics), and RP has been linked to more than 60 disease genes. About 10–15% of RP patients have been attributed to variants in *USH2A*.[20,21] Therefore, this represents the scenario where variants in a gene lead to a recessive disease with heterogeneous genetic basis. Seven hundred forty-eight *USH2A* variants collected in HGMD were evaluated. Among them, 13 variants with AF greater than 0.52% ($\sqrt{11/400,000}$) in gnomAD were considered benign and excluded from further analysis (Supplementary Tables S8 and S9). For the remaining 735 variants, 18 were identified as likely benign (binomial test1, Bonferroni correction p value ≤ 0.05/735 and test2 Bonferroni correction p value > 0.05/735), whose AF ranges from 0.04% to 0.5% (Fig. **3b**). Based on data in the original reports and the number of individuals carrying homozygous variants observed in gnomAD, at least 10 of the 18 variants are likely benign. Specifically, three variants were suggested to be benign by the original reference (e.g., NM_206933.2:c.15433G>A,[22] NM_206933.2:c.6587G>C[23,24]). All of them have homozygotes in gnomAD, with the number ranging from 4 to 120. Eight variants were annotated as VUS based on the original papers (e.g., NM_206933.2:c.10510C>G[16]) and three of them have homozygotes in gnomAD with the number ranging from one to nine (Supplementary Tables S8 and S9). Finally, for the seven variants that were suggested to be likely pathogenic by the original papers (e.g., NM_206933.2:c.11815G>A[25]), four of them have homozygotes in gnomAD with the number ranging from 1 to 95 (Supplementary Tables S8 and S9).

Identification of benign variants in LRP5 gene for familial exudative vitreoretinopathy

Familial exudative vitreoretinopathy (FEVR) is a rare retinal vascular disorder that is primarily dominantly inherited.[26] Variants in *LRP5* account for about 20–25% of FEVR patients. Therefore, this represents the scenario where variants in a gene lead to a dominant disease with heterogeneous genetic basis. One hundred fifty putative pathogenic variants in *LRP5* were analyzed. After excluding 7 variants based on the disease prevalence of FEVR (greater than 0.05%, $1-\sqrt{1-1/1000}$), 16 likely benign variants were identified (binomial test1, Bonferroni correction p value ≤ 0.05/143 and test2 Bonferroni correction p value > 0.05/143). The AF of the 16 variants is in the range 0.0058–0.029% (Fig. **3c**). Based on the data presented in the original report, nine variants were found in patients with diseases unrelated to the eye, including six variants related to bone diseases and three variants related to colorectal cancer, indicating these variants are not pathogenic variants for FEVR. Among the seven variants linked to eye diseases, two variants were identified in patients with a second *LRP5* variant in compound heterozygous form, implying each variant alone might be insufficient to lead to disease. For the remaining five variants, one variant was considered as benign, one as VUS, and three as likely pathogenic based on the original literature (Supplementary Tables S10 and S11). To further evaluate these five variants, functional assays were conducted as described below.

Comparison with in silico variant prediction scores and ClinVar annotation

We have compared our results with three commonly used in silico variant prediction tools, including CADD, REVEL, and phastcon_100way. The benign variants identified by our method have lower CADD, REVEL, and phastcon_100way scores than the other reported variants in the same gene for all the three genes, supporting our test result (Fig. **4a–c** and Supplementary Tables S6, S8, S10, S12, Supplementary Results).

To further assess our test results, we cross referenced to ClinVar annotation. As shown in Fig. **4d**, the ClinVar assessment largely supports our test results. Specifically, of the 74 benign variants identified by our test and with records in ClinVar (a total of 91 likely benign variants identied by our test), the vast majority are classified as benign (32 variants, 43.2%), have conflicting interpretations of pathogenicity (24 variants, 32.4%), or are VUS (15 variants, 20.3%) (Fig. **4d**, and Supplementary Tables S6, S8, and S10).

Functional assay of predicted benign variants in *LRP5*

To further evaluate our results, we performed a functional assay on five *LRP5* variants that were originally reported as pathogenic but are identified as likely benign by our test. As shown in Fig. **5**, three of the variants, *LRP5*.R1219H (NM_002335.3:c.3656G>A, two-tailed t test, p value < 0.005), *LRP5*.R1342Q (NM_002335.3:c.4025G>A, t test, p value > 0.05), and *LRP5*.H1383P (NM_002335.3:c.4148A>C, t test, p value < 0.002), have similar or higher WNT signaling activity than the wild-type control without or with WNT3A treatment, suggesting that these three variants are indeed likely benign. In contrast, *LRP5*.A422V (NM_002335.3:c.1265C>T) shows similar signaling activity to *LRP5*.WT without WNT3A treatment (t test, p value = 0.2733), but its activity is reduced by about 50% with WNT3A treatment (t test, p value = 4.95e-5) (Fig. **5**), suggesting that *LRP5*.A422V is likely to be a hypomorphic allele. One variant, *LRP5*.T1506M (NM_002335.3:c.4517C>T), shows lower signaling activity than *LRP5*.WT without (t test, p value = 1.01e-6) or with WNT3A treatment (t test, p value = 1.31e-7), suggesting it is likely to be a pathogenic allele (Fig. **5**, Supplementary Table S13, Supplementary Results).

DISCUSSION

To identify the variants likely to be benign in the context of Mendelian diseases, we designed a novel statistical test by integrating disease prevalence, AF in the patient cohort, and AF in the normal population into a robust statistical framework. Evaluation of our model with both the simulated and real data followed by experimental validation show that it is well powered to detect benign variants, and especially effective for variants with low AF in the normal population.

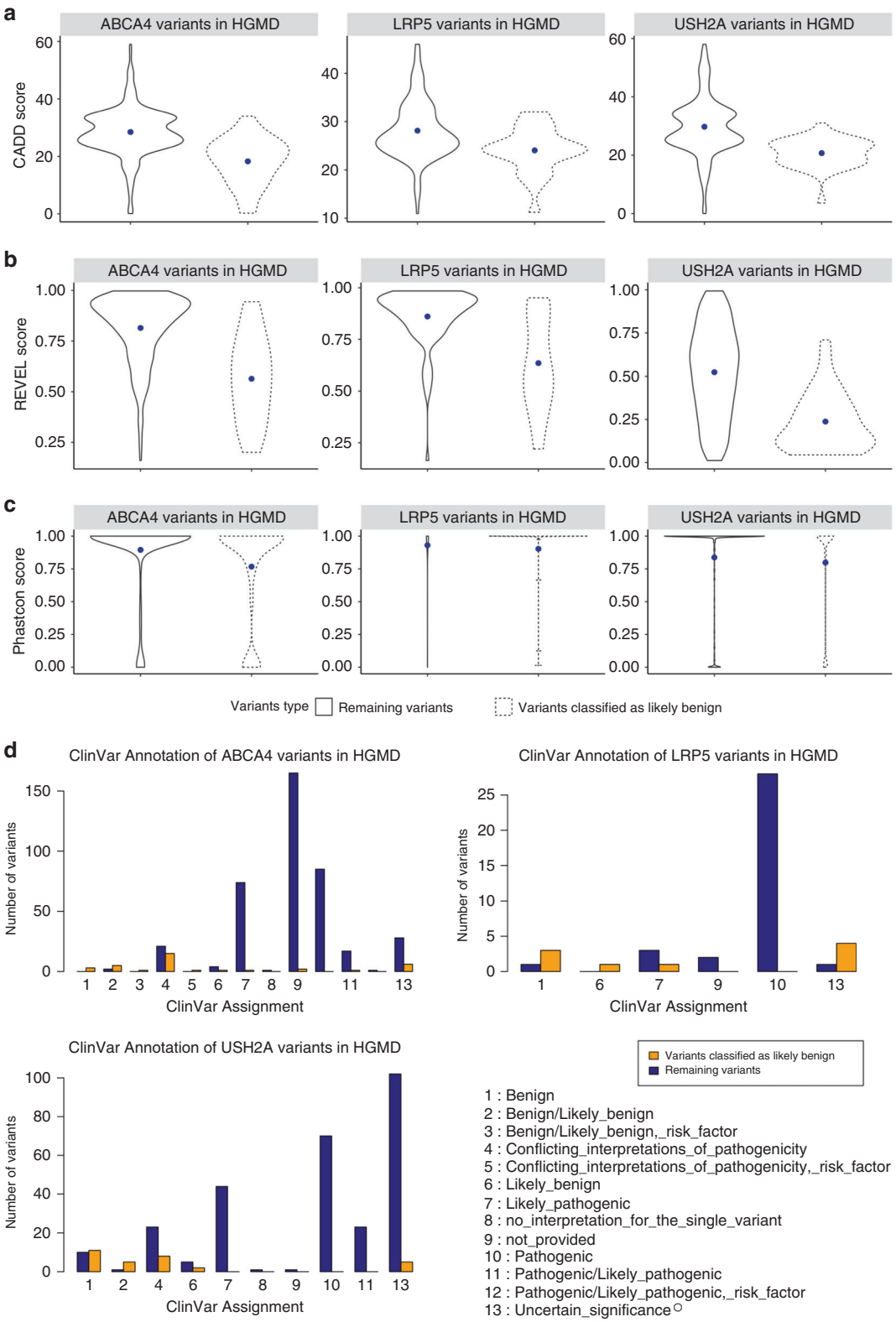


Fig. 4 The distribution of other variant prediction scores and ClinVar assignment for the Human Gene Mutation Database (HGMD) variants in *ABCA4*, *LRP5*, and *USH2A*. (**a**) Distribution of CADD phred scores, (**b**) distribution of REVEL scores, and (**c**) distribution of phastcon scores. Dotted line indicates the likely benign variants identified by our test. Solid line indicates the other HGMD variants. The blue dot indicates the mean of the distribution. (**d**) The ClinVar assignment. The bar plot shows the distribution of the variants among the ClinVar assignment categories. The ClinVar categories labeled with 1–13 were enumerated in the right bottom panel. Orange indicates the likely benign variants identified by our test. Blue indicates the other HGMD variants.

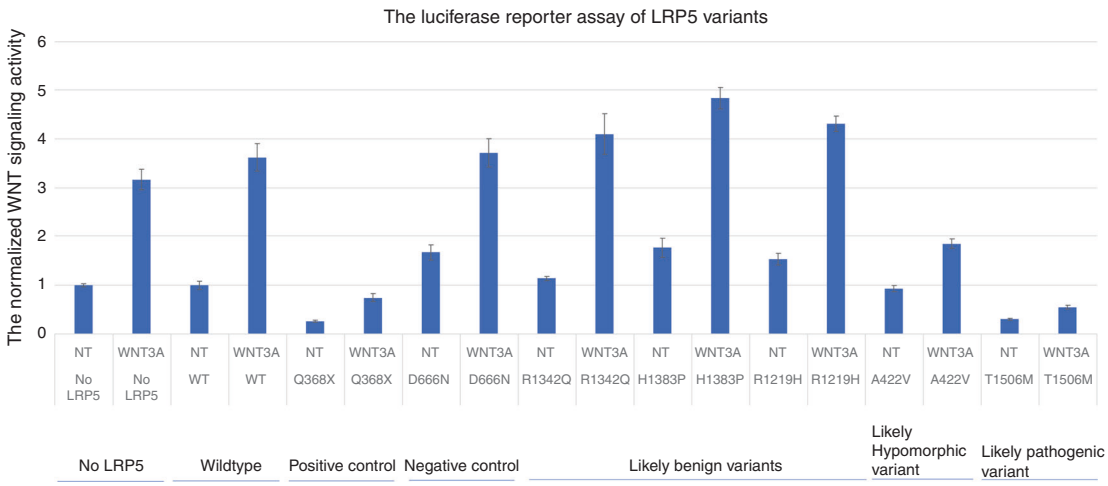


Fig. 5 The luciferase reporter assay of *LRP5* variants. We tested five variants that were identified as likely benign by our test along with a positive control and a negative control. The luciferase reporter assay of the positive control variant, p.Q368X, showed reduced WNT signaling activity, while that of the negative control variant, p.D666N, showed higher or similar WNT signaling activity to the wild-type allele. The variants, p.R1342Q, p.H1383P, and p.R1219H, showed higher or similar WNT activity to the wild-type allele, suggesting they are likely benign and consistent with test results. However, p.A422V showed similar WNT signaling to the wild-type allele without WNT3A treatment, but reduced WNT signaling with WNT3A treatment, suggesting it might be a hypomorphic allele. p.T1506M showed the reduced WNT signaling without or with WNT3A treatment, suggesting it might be a pathogenic allele. The *y*-axis shows the WNT signaling activity of the variants normalized by that of the wild-type allele without WNT3A treatment. *NT* no treatment. *WNT3A* with treatment.

Simulation analysis also suggests the test is more powerful when using a patient cohort with larger sample size, for dominant diseases versus recessive diseases, and for rarer diseases versus relatively common diseases. Furthermore, simulation indicates that the test has a high specificity that is less affected by the disease prevalence and AF in the normal population. Moreover, simulations show that the test performance is not significantly affected by the estimation of disease prevalence and allele penetrance, but could be affected by sampling bias. Additionally, in the case when prevalence of a disease approaches zero, several confounding factors likely co-occur: the variance of disease prevalence estimation likely increases, the available sample size will be smaller, and within these constraints, the penetrance of an allele possibly varies by its frequency. In those situations, the test power to detect extremely rare benign variants could be compromised and the test should be used with caution.

For real data sets, we applied our test to screen the variants in HGMD using real patient cohort data and AF in gnomAD, and identified the variants that are likely benign for three types of Mendelian diseases. The identified variants have the population AF in the range of 0.0058% to higher than 1%, consistent with the simulation analysis showing the high power of our test for variants with low AF in the population. Moreover, our test results are supported by multiple independent evidences. First, the original literature suggests that a large proportion of the predicted benign variants can indeed be classified as benign or VUS. Second, for recessive diseases, a large proportion of the predicted benign variants are found in homozygous state in multiple individuals in gnomAD, suggesting they are unlikely to cause disease in biallelic state. Third, other in silico variant prediction scores (i.e., CADD, REVEL, and phastcon) showed that the putatively benign variants are less deleterious than the other HGMD variants in the same genes. Fourth the majority of the putative benign variants are classified as benign, conflicting interpretations of pathogenicity, and uncertain significance by ClinVar.

Predictions from our model are further validated by a functional assay. The luciferase functional assay was performed on a subset of HGMD variants predicted to be benign in *LRP5* (with population AF in the range of 0.0058% to 0.029%). The assay shows that three of the five tested variants are likely benign. Another variant, *LRP5*.A422V, is likely a hypomorphic variant. This is also consistent with the original reference in which *LRP5*.A422V was only seen in *cis* with *LRP5*.R348W in the patient family,[27] therefore *LRP5*.A422V alone might not be severe enough to cause the phenotype. Interestingly, *LRP5*.R348W is predicted to be likely pathogenic by our test. Additionally, one tested variant, *LRP5*.T1507M, is likely pathogenic. The reason of the contradictory result for *LRP5*.T1507M might be due to sampling bias of this allele in our patient cohort. Specifically, undersampling of this variant in the patient cohort could lead to false positive interpretation. Indeed, our test suggested this allele is slightly enriched in the patient cohort compared with its AF in normal population with a p value of 0.01, though it did not pass the multiple testing correction of p value (p value $\leq 0.05/143$) to be assigned as a pathogenic allele.

One of the main limitations is that the test depends on the availability of patient sequencing data and the patient cohort size. With the dramatic reduction in sequencing cost, it has become common practice to perform panel exome or even genome sequencing as part of the diagnosis for patients with

rare genetic diseases. Hence, we expect that this bottleneck will soon be overcome. Additionally, although our model is for filtering the variants of a single gene, if the disease is caused by multiple genes, one can simply adjust the Q value to the frequency of the disease due to variants in the gene of interest and apply our test directly (e.g., the aforementioned cases of *USH2A* and *LRP5*). Moreover, it is straightforward to aggregate multiple genes from the same disease and run the test directly with the modified Q value accordingly. We envision that the performance of our test can be improved by increasing the patient cohort size and combining the test with other variant prediction scores, e.g., REVEL and CADD scores, which are based on types of information other than AF. In summary, as a general test that is mainly driven by the size of the data set and can be applied to any type of variant in the context of all Mendelian diseases, we believe our work provides a general framework for filtering benign variants with very low population AF, and its performance will continue to improve as more patient sequences become available.

SUPPLEMENTARY INFORMATION

The online version of this article (https://doi.org/10.1038/s41436-020-00948-3) contains supplementary material, which is available to authorized users.

CODE AVAILABILITY

Our code is available at https://github.com/fe4960/Binomial_test.

ACKNOWLEDGEMENTS

We are grateful to the lab of David Moore at Baylor College of Medicine for providing L cells for the functional assays of *LRP5* variants. We thank the computing cluster server in the Molecular and Human Genetics Department at Baylor College of Medicine for providing the computing resource. This work was supported by grants from the National Eye Institute (grant numbers R01EY022356, R01EY018571, EY002520 to R.C.); Retinal Research Foundation [to RC]; National Institutes of Health shared instrument grant (grant number S10OD023469 to R.C.); and the Competitive Renewal Grant of Knights Templar Eye Foundation (to J.W.).

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

1. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011;12:745–755.
2. Whiffin N, Minikel E, Walsh R, et al. Using high-resolution variant frequencies to empower clinical genome interpretation. Genet Med. 2017;19:1151–1158.
3. Clarke GM, Anderson CA, Pettersson FH, et al. Basic statistical analysis in genetic case-control studies. Nat Protoc. 2011;6:121–133.
4. Michaelides M, Hunt DM, Moore AT. The genetics of inherited macular dystrophies. J Med Genet. 2003;40:641–650.
5. Zernant J, Xie YA, Ayuso C, et al. Analysis of the ABCA4 genomic locus in Stargardt disease. Hum Mol Genet. 2014;23:6797–6806.
6. Thiadens AA, Phan TM, Zekveld-Vroon RC, et al. Clinical course, genetic etiology, and visual outcome in cone and cone-rod dystrophy. Ophthalmology. 2012;119:819–826.
7. Downes SM, Packham E, Cranston T, et al. Detection rate of pathogenic mutations in ABCA4 using direct sequencing: clinical and research implications. Arch Ophthalmol. 2012;130:1486–1490.
8. Sun H, Smallwood PM, Nathans J. Biochemical defects in ABCR protein variants associated with human retinopathies. Nat Genet. 2000;26:242–246.
9. Downs K, Zacks DN, Caruso R, et al. Molecular testing for hereditary retinal disease as part of clinical care. Arch Ophthalmol. 2007;125:252–258.
10. Simonelli F, Testa F, de Crecchio G, et al. New ABCR mutations and clinical phenotype in Italian patients with Stargardt disease. Invest Ophthalmol Vis Sci. 2000;41:892–897.
11. Eisenberger T, Neuhaus C, Khan AO, et al. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, noncoding exons and the overall variant load: the example of retinal dystrophies. PLoS One. 2013;8:e78496.
12. Rosenberg T, Klie F, Garred P, Schwartz M. N965S is a common ABCA4 variant in Stargardt-related retinopathies in the Danish population. Mol Vis. 2007;13:1962–1969.
13. Wiszniewski W, Zaremba CM, Yatsenko AN, et al. ABCA4 mutations causing mislocalization are found frequently in patients with severe retinal dystrophies. Hum Mol Genet. 2005;14:2769–2778.
14. Burke TR, Tsang SH, Zernant J, et al. Familial discordance in Stargardt disease. Mol Vis. 2012;18:227–233.
15. Testa F, Rossi S, Sodi A, et al. Correlation between photoreceptor layer integrity and visual function in patients with Stargardt disease: implications for gene therapy. Invest Ophthalmol Vis Sci. 2012;53:4409–4415.
16. van Huet RA, Pierrache LH, Meester-Smoor MA, et al. The efficacy of microarray screening for autosomal recessive retinitis pigmentosa in routine clinical practice. Mol Vis. 2015;21:461–476.
17. Zernant J, Schubert C, Im KM, et al. Analysis of the ABCA4 gene by next-generation sequencing. Invest Ophthalmol Vis Sci. 2011;52:8479–8487.
18. Suarez T, Biswas SB, Biswas EE. Biochemical defects in retina-specific human ATP binding cassette transporter nucleotide binding domain 1 mutants associated with macular degeneration. J Biol Chem. 2002;277:21759–21767.
19. Biswas-Fiss EE, Affet S, Ha M, Biswas SB. Retinoid binding properties of nucleotide binding domain 1 of the Stargardt disease-associated ATP binding cassette (ABC) transporter, ABCA4. J Biol Chem. 2012;287:44097–44107.
20. Huang L, Mao Y, Yang J, et al. Mutation screening of the USH2A gene in retinitis pigmentosa and USHER patients in a Han Chinese population. Eye (Lond). 2018;32:1608–1614.
21. Sandberg MA, Rosner B, Weigel-DiFranco C, et al. Disease course in patients with autosomal recessive retinitis pigmentosa due to the USH2A gene. Invest Ophthalmol Vis Sci. 2008;49:5532–5539.
22. Glockle N, Kohl S, Mohr J, et al. Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. Eur J Hum Genet. 2014;22:99–104.
23. Shearer AE, Eppsteiner RW, Booth KT, et al. Utilizing ethnic-specific differences in minor allele frequency to recategorize reported pathogenic deafness variants. Am J Hum Genet. 2014;95:445–453.
24. Garcia-Garcia G, Aparisi MJ, Jaijo T, et al. Mutational screening of the USH2A gene in Spanish USH patients reveals 23 novel pathogenic mutations. Orphanet J Rare Dis. 2011;6:65.
25. Tajiguli A, Xu M, Fu Q, et al. Next-generation sequencing-based molecular diagnosis of 12 inherited retinal disease probands of Uyghur ethnicity. Sci Rep. 2016;6:21384.
26. Li LH, Li N, Zhao JY, et al. Findings of perinatal ocular examination performed on 3573, healthy full-term newborns. Br J Ophthalmol. 2013;97:588–591.
27. Salvo J, Lyubasyuk V, Xu M, et al. Next-generation sequencing and novel variant determination in a cohort of 92 familial exudative vitreoretinopathy patients. Invest Ophthalmol Vis Sci. 2015;56:1937–1946.