Impact of Variants of Varying Clinical Consequence in Underrepresented Populations and Implications for a Minimum Variant Set for Pan-Ancestry Cystic Fibrosis Carrier Screening

G016

Gary J. Latham, Brian C. Haynes, Brad Hall, and John N. Milligan

Asuragen, a Bio-Techne Brand, Austin, TX

Summary

- Toward the goal of equitable cystic fibrosis (CF) testing and improved outcomes for minority patients, the ACMG recently updated the recommended number of *CFTR* variants for CF carrier screening (CS) to 100 variants ("ACMG-2023") by filtering candidates through the CFTR2 and gnomAD (v2.1.1 plus v3.1.2) databases and combining them to cover 95% of carriers across ancestries
- However, this approach omitted variants of varying clinical consequence (VVCC) and pathogenic structural variants (SVs) shown in recent US population-level NGS studies to have high prevalence in minority populations where delayed diagnosis and treatment lead to worse clinical outcomes
- Here, we leveraged pathogenicity scores calculated from ClinVar as a metric for scrutinizing VVCC to identify likely pathogenic variants, showing that VVCC above the threshold represent 25-45% of carriers in African, East Asian, and South Asian ancestries and remain underrepresented on most *CFTR* panels
- Inclusion of VVCC and SV with strong pathogenic evidence in targeted panels may further improve detection rates in minority populations toward the goal of equitable CF testing and screening

Introduction

Cystic Fibrosis (CF), a progressive autosomal recessive disease, is caused by pathogenic variants in the *CFTR* gene. CF carrier screening (CS) is the most common CS test performed, recommended by both the ACMG and ACOG for all women considering pregnancy.[1] Over 2,000 *CFTR* variants have been identified, but most are benign or have unknown significance, and variant frequencies differ significantly between ancestries. Historically, CS recommendations examined only 23 pathogenic variants in *CFTR*. In the last 4 years, several studies in the diverse US population showed that 31-44% of carriers were missed by this panel, with outsized impact on minority groups[2,3]. In response, the ACMG recently updated recommendations to include 100 variants ("ACMG-2023"), selected by filtering candidates through the CFTR2[4] and gnomAD (v2.1.1 plus v3.1.2) databases and combining them to cover 95% of the total US carrier frequency. By requiring variants to be CF-causing in CFTR2, however, all variants of varying clinical consequence (VVCC) were excluded. Further, the use of gnomAD to filter variants also excluded most structural variants (SV), as current population databases such as gnomAD lack sufficient sample size to prioritize SVs involved in CF and other recessive genetic disorders. Because recent studies have shown that a small subset of *CFTR* VVCC and SV are highly prevalent in minority populations,[2,3] this approach excludes key variants critical for equitable pan-ancestral coverage, resulting in reduced coverage in reduced populations.[5]

Here, we derived a method for calculating pathogenicity ratings (PR) from ClinVar, a frequently updated database that aggregates pathogenicity data from broad sources. We show that PR can be used to identify pathogenic VVCC in CFTR2, representing 24-45% of carriers in African, East Asian, and South Asian ancestries. We found a small set of 14 VVCC with both high pathogenicity ratings (PR ≥ 75%) and high prevalence (MAF ≥ 5e-5) in at least one of six ancestry groups. We demonstrate that most panels detected less than half of carriers in at least one ancestry when including these pathogenic VVCC in carrier frequency estimations from gnomAD. These data suggest that the inclusion of VVCC with strong pathogenic evidence in *CFTR* testing and screening panels would improve pan-ancestral coverage, consistent with recent UK biobank studies showing that patients with one CF-causing variant and one VVCC are 10-fold more prevalent than those with two CF-causing variants.[6]

Methods

Mirroring the ACMG 2023 guidelines, we generated a base set of pathogenic *CFTR* variants consisting of all variants listed as CF-causing variants in CFTR2 (April 7th, 2023 release)[4] and all variants previously in the recommended ACMG CF23 panel. Next, we filtered the pathogenic variant set based on their inclusion in gnomAD (v2.1.1/v3.1.2) in 6 ancestries: African/African American (AFR), Latino/Admixed American (AMR), Ashkenazi Jewish (ASJ), East Asian (EAS), non-Finnish European (NFE), and South Asian (SAS). Two variants (R117H, T338I) were manually included and two (S434X, 4209TGTT->AA) were manually excluded to mirror the ACMG 2023 guidelines.[1] However, because ACMG data referenced a previous version of CFTR2 (April 29th, 2022 release), our list contained one additional variant compared to their list resulting in a set of 198 pathogenic variants (CFTR2-198).

Next, we generated a set of variants that included both pathogenic/likely pathogenic and VVCC with pathogenic evidence of CF or *CFTR*-related disorders using ClinVar by calculating a pathogenicity rating:

$$PR = N_P / N_S * 100$$

where PR is pathogenicity rating calculated from ClinVar submissions reported as a percentage, N_P is the number of ClinVar submissions that are pathogenic, likely pathogenic, or drug response, and N_S is the total number of ClinVar submissions reporting clinical significance.

To establish an appropriate threshold for PR, we compared clinical significance determinations from CFTR2 and ClinVar to PR, determining that 75% sufficiently distinguished established benign and pathogenic variants in both databases (Figure 1). We also evaluated different thresholds for number of submissions in ClinVar and chose ≥ 5 submissions to prevent bias from a small number of submissions. Variants meeting both PR and submission thresholds were combined generating a total of 247 variants ("ClinVar-247"). The independent ClinVar-247 variant set included all pathogenic variants in the CFTR2-198 variant set, and a subset of VVCC from CFTR2. We compared percent coverage of nine commercially available panels and the ACMG 2023 panel of 100 variants with both ClinVar-247 and CFTR2-198 using variant frequencies in six major ancestries estimated from gnomAD (v2.1.1/v3.1.2).

Results

1.

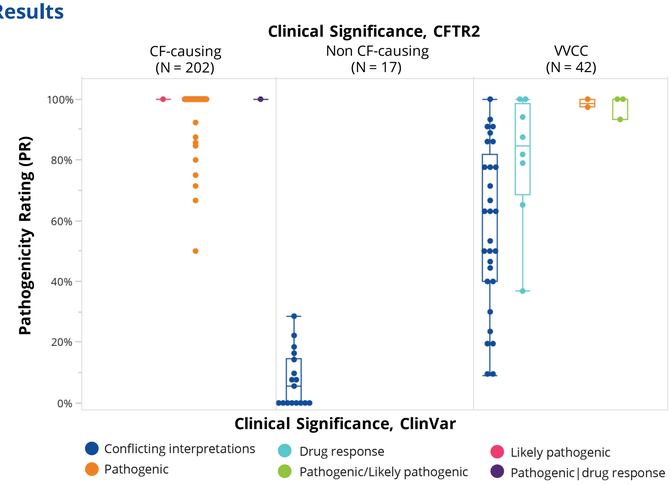

Figure 1. Pathogenicity Ratings are Strongly Correlated with Clinical Significance in CFTR2 and ClinVar. Data includes all variants that are present in CFTR2, ClinVar, and gnomAD. CFTR2 clinical significance indicated at top.[4] ClinVar clinical significance indicated by color, see legend. CF-causing variants had 100% agreement between CFTR2 and ClinVar, with PR = 50-100%. Benign variants had 100% agreement between CFTR2 and ClinVar, with PR = 0-27%. VVCC in CFTR2 included variants listed in ClinVar as conflicting interpretations, drug response, and pathogenic/likely pathogenic, with a broad range for PR = 9-100%. For variants listed as pathogenic/likely pathogenic or drug response in ClinVar, 97.7% (210/215) had PR ≥ 75%. For variants listed as CF-causing in CFTR2, 98.5% (199/202) had PR ≥ 75%. Based on these data, we set a PR threshold of 75% to establish pathogenicity of CFTR2 VVCC.

2.

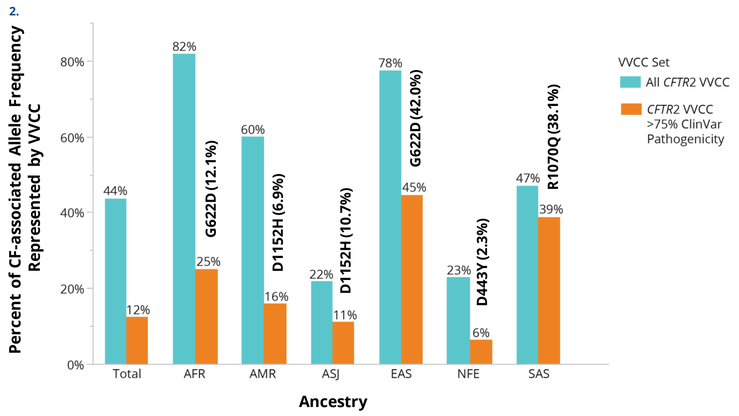

Figure 2. Percentage of Allele Frequencies Represented by CFTR2 VVCC's in gnomAD are Highest in Minority Ancestries. The percentage of overall allele frequencies represented by CFTR2 VVCC relative to the 240 variants considered CF-causing or VVCC by CFTR2 in each of six ancestries: African/African American (AFR), Latino/Admixed American (AMR), Ashkenazi Jewish (ASJ), East Asian (EAS), non-Finnish European (NFE), and South Asian (SAS). Data shown including all CFTR2 VVCC (teal)[4] and those with PR ≥ 75% (orange). The most prevalent VVCC and percent of total AF is indicated in parentheses for each ancestry. R117H is not considered as a VVCC in these calculations despite being included as such in CFTR2 because it is considered pathogenic by the ACMG.[1] While VVCC represent a high percentage of variants when ignoring pathogenicity in most ancestries, three ancestries considered minorities in the US (AFR, EAS, SAS) have high VVCC representation (25-45%) when applying pathogenicity thresholds. Surprisingly, only 14 VVCC had high pathogenicity ratings (PR ≥ 75%) and high prevalence (MAF ≥ 5e-5) in one or more ancestries: D443Y, D614G, D1152H, F312del, G622D, I618T, P5L, Q1476X, R117G, R117H, R1070Q, R1070W, S977F, and S1455X. Based on cell line studies, the average CFTR function as a percent of wildtype CFTR was 32.8% where data was available (8/14), including G622D (18.2%) and R1070Q (20.6%).[7] A single VVCC accounts for 7-42% of all carriers (CF-causing and VVCC with PR ≥ 75%) in each ancestry except non-Finnish Europeans. Critically, a single VVCC also explained 43-98% of carriers with VVCC in each of these minority groups.

3.

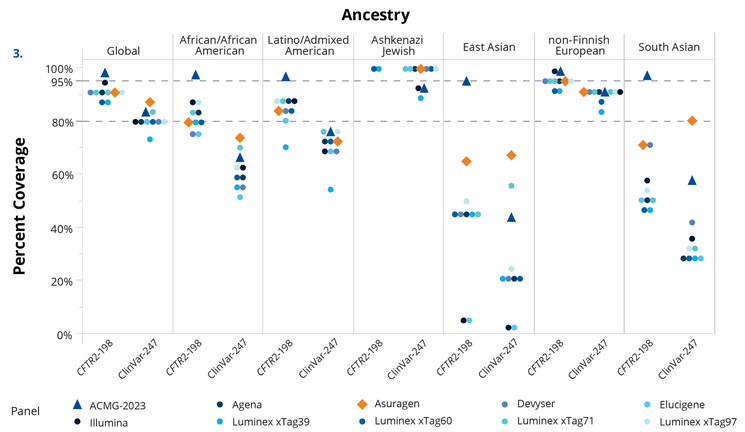

Figure 3. Incorporation of VVCC with High Pathogenicity Ratings in ClinVar Highlights Ancestry-specific Coverage Differences Between Panels. CFTR2-198 represents only pathogenic *CFTR* variants from CFTR2 (n = 198).[4] ClinVar-247 includes all variants in CFTR2-198 and includes VVCC in CFTR2 by assigning pathogenicity based on variants in ClinVar with ≥5 submissions and PR ≥ 75% (n = 247). Variant frequencies were estimated globally and across each ancestry using gnomAD (see Methods). Nine commercially available panels and the ACMG 2023 minimum 100 variant list are labeled as shown. Panels with the best overall coverage for each variant set indicated (ACMG-2023 for CFTR2-198, triangle; Asuragen for ClinVar-247, diamond). Of the ten panels evaluated, nine had at least one ancestry with <50% coverage when including VVCC with high pathogenicity scores (ClinVar-247). Only the Asuragen panel had >65% coverage in all ancestries when evaluated with ClinVar-247. All assays had 100% coverage for Ashkenazi Jewish ancestry with the CFTR-198 set.

Table 1. Pathogenic *CFTR* Structural Variants (SVs) with High Prevalence. Because gnomAD contains SV data for only a small subset of individuals, it is not sufficiently powered for calculating SV frequencies.[1] Only a single SV was considered by ACMG (c.1820_1903del), which did not meet their frequency threshold for inclusion.[1] However, two SVs listed as CF-causing in CFTR2 represent >0.2% of total alleles in CFTR2,[4] recent US population studies,[2] or both. Frequency is indicated as percent of total pathogenic alleles. SVs are listed as a sum of allele frequencies with a common deletion region; CFTRdele2,3 represents all exon 2-3 deletions, CFTRdele20 represents exon 18-20, 19-20, and 19-21 deletions. Other pathogenic large exon deletions were observed in both data sources, but not at a frequency >0.2%.

Variant	Clinical Significance, CFTR2	Frequency, CFTR2	Frequency, US Population[2]
CFTRdele2,3	CF-causing	0.01%	0.48%
CFTRdele20	CF-causing	0.29%	0.38%

Conclusions

- Pathogenicity ratings (PR) provide a useful means for scrutinizing pathogenicity of VVCC. PR is a ratio of ClinVar submissions listed as pathogenic, likely pathogenic, or drug response vs total submissions expressed as a percentage. CF-causing variants were correlated with PR in both variant databases: CFTR2 (98.5% CF-causing have PR ≥ 75%) and ClinVar (97.7% have PR ≥ 75%).
- VVCC with PR ≥ 75% represented a large proportion of carriers (25-45%) in ancestries considered minorities in the US (AFR, SAS, EAS) where delayed diagnosis leads to worse clinical outcomes;[6] a single VVCC explained 43-98% of carriers with VVCC in each of these minority groups
- Of nine commercial panels and the 100 variant ACMG 2023 list evaluated, only the 65-variant Asuragen panel had >65% coverage in all ancestries when VVCC with PR ≥ 75% were included in estimates (87% global coverage)
- These data suggest that test panels including just 14 VVCC with strong pathogenic evidence and high prevalence in one or more ancestries would improve equitable coverage; inclusion of pathogenic SVs with high prevalence (**Table 1**)[2] would further improve coverage

References:

1. Deignan, J.L. *et al*. Updated Recommendations for CFTR Carrier Screening: A Position Statement of the American College of Medical Genetics and Genomics (ACMG). Genet. Med. 2023, 100867.
2. Beauchamp, K.A. *et al*. Sequencing as a First-Line Methodology for Cystic Fibrosis Carrier Screening. Genet. Med. 2019, 21, 2569–2576.
3. Westemeyer, M. *et al*. Clinical Experience with Carrier Screening in a General Population: Support for a Comprehensive Pan-Ethnic Approach. Genet. Med. 2020, 22, 1320–1328.
4. The Clinical and Functional TRanslation of CFTR (CFTR2) Available online: http://cftr2.org (accessed on 1 August 2023).
5. Hall, B. *et al*. Latham GJ. Multi-site Verification of a Targeted CFTR Polymerase Chain Reaction/Capillary Electrophoresis Assay that Evaluates Pathogenic Variants across Diverse Ethnic and Ancestral Groups. Arch Pathol Lab Med. 2023. In Press.
6. McGarry, M.E. *et al*. Detection of Disease-causing CFTR Variants in State Newborn Screening Programs. Pediatr. Pulmonol. 2023, 58, 465–474.
7. Raraigh, K.S. *et al*. Functional Assays Are Essential for Interpretation of Missense Variants Associated with Variable Expressivity. Am. J. Hum. Genet. 2018, 102, 1062–1077.

Presented at AMP 2023

bio-techne®

Asuragen®