Narrative Review Article

Comparing Mutation Patterns Across Populations: *CFTR*, *MEFV*, and *HBB* Variants in Irish, Mediterranean-Associated, and Iranian Groups

Oliver Goulden

Bronxville High School, 177 Pondfield Road, Bronxville, NY 10708, United States

ABSTRACT

Population-specific variant data can improve genetic screening and interpretation, but reported frequencies are often drawn from evidence types that are not directly comparable. This narrative review examines the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene in cystic fibrosis (CF), the *MEFV* innate immunity regulator (*MEFV*) gene in familial Mediterranean fever (FMF), and the beta-globin (*HBB*) gene in beta thalassemia across Irish, Mediterranean-associated, and Iranian populations. All numerical values discussed were extracted from published affected-patient cohorts, screening studies, systematic reviews, or public reference databases; they are not newly generated population estimates. Published Northern Irish CF cohorts show a concentrated *CFTR* mutation spectrum dominated by c.1521_1523delCTT (p.Phe508del), whereas the Iranian literature reports a broader range of *CFTR* variants. *MEFV* studies repeatedly identify c.2080A>G (p.Met694Val) across regional cohorts, but frequencies vary with referral patterns, testing panels, and population definition. Interpretation is also complicated by the debated penetrance of c.442G>C (p.Glu148Gln). For *HBB*, malaria-related selection provides a broad evolutionary explanation for the persistence of beta thalassemia, while local demographic processes may account for the regional predominance of particular pathogenic variants. The main limitations are non-equivalent denominators, referral and testing-panel bias, uneven geographic coverage, and the inability of frequency patterns alone to prove founder effects or selection. Overall, local mutation data are more clinically useful than broad ancestry labels and should be validated within the populations in which screening is performed.

Keywords: Cystic fibrosis; Familial Mediterranean fever; Beta thalassemia; *CFTR*; *MEFV*; *HBB*; Founder effect; Natural selection

INTRODUCTION

Genetic disease variants are not distributed evenly among human populations. Demographic events, migration, population size, and local adaptation can all shape the frequency of disease-associated alleles, creating differences that matter for diagnosis, carrier testing, and screening (1). However, genetic studies have historically overrepresented populations of European ancestry, and the continuing lack of diversity limits the accuracy of

Corresponding author: Oliver Goulden, E-mail: olly_goulden@icloud.com.

Accepted August 6, 2026
https://doi.org/10.70251/HYJR2348.4476177I

variant interpretation in many groups (2, 3). Broad labels such as Irish, Mediterranean, or Iranian can help organize a question, but they can also combine populations with different regional histories and mutation spectra.

Ancient DNA has shown that populations across Europe, Central Asia, and South Asia were shaped by repeated migration and admixture rather than by a single ancestral line (4, 5). These findings provide useful background for understanding partial shared ancestry across Eurasia. They do not, however, demonstrate that a disease-associated variant in different modern populations has the same historical origin. That stronger conclusion would require evidence such as a shared haplotype, linkage disequilibrium, genealogical data, or demographic modeling.

This review compares three genes that represent different population-genetic questions. The *CFTR* gene was selected because affected-patient data from Northern Ireland show a mutation spectrum strongly concentrated around c.1521_1523delCTT (p.Phe508del), while published Iranian evidence appears more heterogeneous (6, 7). The *MEFV* gene was selected because recurring variants are reported across Mediterranean and Iranian populations, although penetrance and variant classification are not uniform (8, 9). The *HBB* gene was selected because beta thalassemia illustrates the broad evolutionary relationship between hemoglobin disorders and malaria, while also showing strong regional variation in the specific pathogenic variants reported (10, 11).

The guiding question is whether these mutation patterns are most consistent with shared migration history, founder effects and drift, natural selection, or a combination of mechanisms. The central argument is that broad ancestry may provide context, but it is not a sufficient explanation by itself. Each gene must be evaluated through the specific type of evidence available, the denominator used, the testing method, and the strength of the historical or functional evidence supporting a proposed mechanism.

EVIDENCE TYPES, DENOMINATORS, AND STUDY DESIGN

The evidence discussed in this review comes from several different source types. An affected-patient mutation spectrum reports the proportion of disease-associated alleles or chromosomes among people who already have a diagnosis. A carrier study estimates how often a variant is found among heterozygous individuals. Newborn screening data may report disease incidence, carrier frequency, or the performance of a particular screening algorithm. Public reference databases such as the Genome Aggregation Database (gnomAD) report allele frequencies in large sequenced populations, most of whom were not selected because they had the disease under study (12). Iranome provides a more focused reference for Iranian variation, beginning with 800 healthy individuals from eight ethnic groups and later expanding by 400 individuals from four additional groups (13). These sources answer different questions and should not be treated as estimates of the same outcome.

The reported mutation spectrum is also influenced by how participants entered a study. A national registry can capture a broad diagnosed population, whereas a referral clinic may overrepresent severe, unusual, or previously unresolved cases. A newborn screening program identifies individuals through a defined algorithm, while a retrospective systematic review combines studies that may differ in diagnostic criteria, geographic coverage, laboratory methods, and publication year. Targeted mutation panels can only detect variants included on the panel, whereas sequencing can identify a wider range of variants. Improvements in sequencing and variant classification over time can therefore make newer studies appear more heterogeneous even when the underlying population has not changed.

Variant classification creates a second layer of difficulty. A reported sequence change is not automatically a disease-causing mutation. Clinical classification considers population frequency, functional evidence, segregation, phenotype, and other criteria (14). This is especially important for *CFTR* c.350G>A (p.Arg117His) and *MEFV* c.442G>C (p.Glu148Gln), whose clinical effects depend on context or remain debated. For this reason, the discussion below separates the frequency of a reported variant from the strength of evidence that the variant is pathogenic.

CFTR AND POPULATION-SPECIFIC MUTATION SPECTRA

A 2025 publication describing the complete Northern Ireland CF population through 2021 reported 61 *CFTR* variants among 1,005 affected-patient alleles. *CFTR* c.1521_1523delCTT (p.Phe508del; historical name F508del) accounted for 626 of 1,005 alleles, or 62.3%. The next reported variants were c.350G>A (p.Arg117His) at 8.9%, c.1652G>A (p.Gly551Asp) at 5.0%, c.1624G>T (p.Gly542Ter) at 3.3%, c.1679G>C (p.Arg560Thr) at 2.8%, and c.200C>T (p.Pro67Leu) at 2.2%. Together,

these six variants represented 84.4% of the affected-patient alleles in that registry-based source (6).

An earlier Northern Irish family study used 412 CF chromosomes from 206 families and reported p.Phe508del in 68.0% of chromosomes. Thirty detected variants accounted for 93.9% of the CF chromosomes studied (15). The broad concentration around p.Phe508del was therefore present in both publications, although the studies were separated by approximately 25 years and used different laboratory methods. The later study identified 39 allelic variants not described in 1996, which may reflect improved sequencing, changing clinical ascertainment, or both (6, 15). Persistence of a common variant across time is consistent with population structure, but it does not by itself establish a founder event.

Republic of Ireland newborn screening data provide a different evidence type. During the first 6.5 years of the national program, the estimated CF incidence was 1 in 2,570 and the carrier frequency was approximately 1 in 25 (16). These values describe population screening, not the percentage of *CFTR* alleles in an affected cohort. They therefore add clinical context but should not be placed on the same quantitative scale as the Northern Irish mutation-spectrum percentages.

The Iranian evidence is less uniform. A systematic review combined 16 articles including 735 Iranian patients with CF and identified 101 reported *CFTR* variants. The review reported p.Phe508del at 21.22%, but this was a pooled result across separate studies rather than a single national registry denominator (7). Geographic coverage was uneven, laboratory methods differed, and some Iranian provinces had little or no published data. The broader reported spectrum may reflect genuine population heterogeneity, but it may also reflect the combination of referral cohorts, study overlap, panel differences, and sequencing methods. Figure 1 and Table 1 therefore present the Northern Irish and Iranian findings as descriptive affected-patient evidence, not as a direct test of screening sensitivity.

Clinical interpretation also depends on the variant. Large registry and functional analyses have classified many common *CFTR* variants as disease-causing, but others have variable clinical consequences (17). For example, the penetrance of c.350G>A (p.Arg117His) is modified by the intronic poly-thymidine tract. The variant is more likely to be disease-causing when it occurs with 5T and has much lower penetrance in other contexts

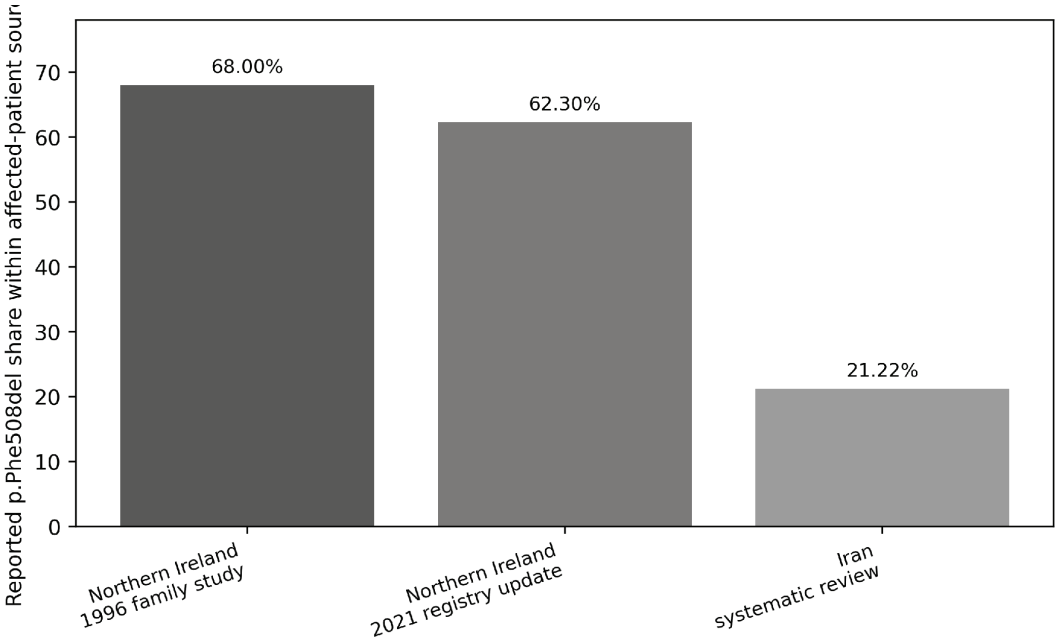

Figure 1. *Reported share of CFTR c.1521_1523delCTT (p.Phe508del) within selected affected-patient sources. The Northern Ireland values use affected-patient CFTR alleles or chromosomes; the Iranian value is pooled from 16 affected-patient studies. The bars are descriptive and do not represent a direct estimate of screening sensitivity.*

Table 1. *Published CFTR evidence summarized by source type and denominator. Percentages from affected-patient cohorts describe the share of CFTR alleles or chromosomes among people with cystic fibrosis. Newborn screening values describe incidence or carrier frequency and are not directly comparable.*

Source and population	Evidence type and denominator	Selected finding	Interpretive caution
Northern Ireland registry update (6)	Complete diagnosed CF population; 1,005 CFTR alleles; data through 2021	c.1521_1523delCTT (p.Phe508del) 626/1,005 alleles (62.3%); top six variants 84.4%	Registry-based affected-patient spectrum. Publication and testing methods differ from older studies.
Northern Ireland family study (15)	206 CF families; 412 CF chromosomes	p.Phe508del 68.0%; 30 detected variants 93.9%	Older laboratory methods and family-based ascertainment.
Republic of Ireland newborn screening (16)	National screening program; population-level incidence and carrier estimates	CF incidence 1 in 2,570; estimated carrier frequency 1 in 25	Screening context, not an affected-patient mutation-spectrum denominator.
Iranian systematic review (7)	16 affected-patient studies; 735 patients; pooled literature	101 reported variants; p.Phe508del reported at 21.22%	Not a single national registry. Geographic coverage, methods, and publication years varied.

(18). A raw frequency table that treats p.Arg117His as equivalent to p.Phe508del would therefore exaggerate the clinical meaning of the reported spectrum.

The high European frequency of p.Phe508del has led to several heterozygote-advantage hypotheses. Experimental studies proposed resistance to cholera toxin and reduced entry of Salmonella Typhi as possible mechanisms (19, 20). A later population-genetic model concluded that cholera and typhoid fever were unlikely to have provided enough historical selective pressure, while tuberculosis was presented as another possible candidate (21). Earlier syntheses also reviewed several proposed explanations (22). These studies show that selection has been proposed, but the affected-patient frequency data in this review cannot identify the responsible historical pressure. Stronger evidence would require haplotype age estimates, geographic and historical modeling, functional evidence tied to reproductive fitness, and comparison with neutral demographic expectations.

The clinical implication is narrower than estimating how many Iranian cases a p.Phe508del-centered panel would miss. Affected-cohort percentages do not equal panel sensitivity because panels contain multiple variants and are evaluated within specific screening algorithms. The evidence supports a more cautious conclusion: limited panels designed around variants common in one population may provide lower coverage in a more heterogeneous population and should be validated against local sequencing data (23).

MEFV FOUNDER MUTATIONS ACROSS REGIONAL POPULATIONS

Familial Mediterranean fever is clinically associated with populations around the Mediterranean basin and parts of the Middle East. Several *MEFV* variants recur across regional studies, particularly c.2080A>G (p.Met694Val; M694V), c.2040G>C (p.Met680Ile; M680I), c.2177T>C (p.Val726Ala; V726A), and c.2082G>A (p.Met694Ile; M694I). The repeated appearance of these variants has led to the term founder mutation, but frequency repetition alone does not prove that all copies descend from one ancestral chromosome.

In a Central Anatolian cohort, 10,033 referred patients were evaluated, 1,684 met Tel-Hashomer clinical criteria, and 1,223 had confirmed mutations. Across 5,753 reported *MEFV* alleles in the tested cohort, p.Met694Val accounted for 39.0%, p.Glu148Gln for 23.8%, p.Met680Ile for 14.8%, and p.Val726Ala for 11.2% (24). The study used reverse hybridization and pyrosequencing, and a high proportion of participants were referred because FMF was suspected. The percentages therefore describe alleles in a selected clinical population rather than population prevalence.

A Northwest Iranian study included 216 patients diagnosed by Tel-Hashomer criteria and screened them with a 12-variant reverse-hybridization panel. Among 432 possible alleles, p.Met694Val was reported in 23.61%, p.Val726Ala in 11.11%, p.Glu148Gln in 9.95%,

and p.Met680Ile in 9.02%. No panel variant was detected in 51 patients (25). The absence of a detected mutation in nearly one quarter of the clinical cohort illustrates how panel coverage can shape the apparent spectrum.

A separate study of 130 Azeri Turk patients in northwest Iran used targeted assays followed by sequencing of selected exons when initial testing was negative. It reported p.Met694Val at 40.19%, p.Glu148Gln at 17.64%, p.Val726Ala at 13.72%, and p.Met680Ile at 12.74% (26). An Iranian systematic review combined 16 articles and 4,256 patients and reported p.Met694Val at 20.27%, p.Glu148Gln at 10.27%, p.Val726Ala at 8.24%, and p.Met680Ile at 7.20% (9). Differences among these estimates may reflect local ancestry, but they also reflect referral patterns, clinical criteria, testing panels, sequencing depth, publication year, and possible overlap among regional reports. Table 2 and Figure 2 keep these denominators visible.

Table 2. *Selected MEFV evidence with study design and denominator. The reported percentages are specific to each clinical or reference population and should not be interpreted as directly comparable prevalence estimates.*

Source and population	Evidence type and denominator	Selected finding	Interpretive caution
Central Anatolia (24)	10,033 referred patients; 5,753 reported MEFV alleles in tested cohort	c.2080A>G (p.Met694Val) 39.0%; c.442G>C (p.Glu148Gln) 23.8%	Selected referral population; reverse hybridization and pyrosequencing.
Northwest Iran (25)	216 clinically diagnosed patients; 432 possible alleles; 12-variant panel	p.Met694Val 23.61%; p.Val726Ala 11.11%; p.Glu148Gln 9.95%	Panel could not detect untested variants; 51 patients had no panel variant.
Azeri Turk northwest Iran (26)	130 patients; targeted assays with sequencing of selected exons	p.Met694Val 40.19%; p.Glu148Gln 17.64%	Local clinical cohort; methods differ from the 216-patient panel study.
Iranian systematic review (9)	16 studies; 4,256 patients; pooled literature	p.Met694Val 20.27%; p.Glu148Gln 10.27%	Regional coverage and testing methods varied; possible study overlap.
Healthy northwest Iran (27)	224 healthy participants	p.Glu148Gln reported in 18.3% of participants	Supports caution in treating p.Glu148Gln as a fully penetrant disease-causing variant.

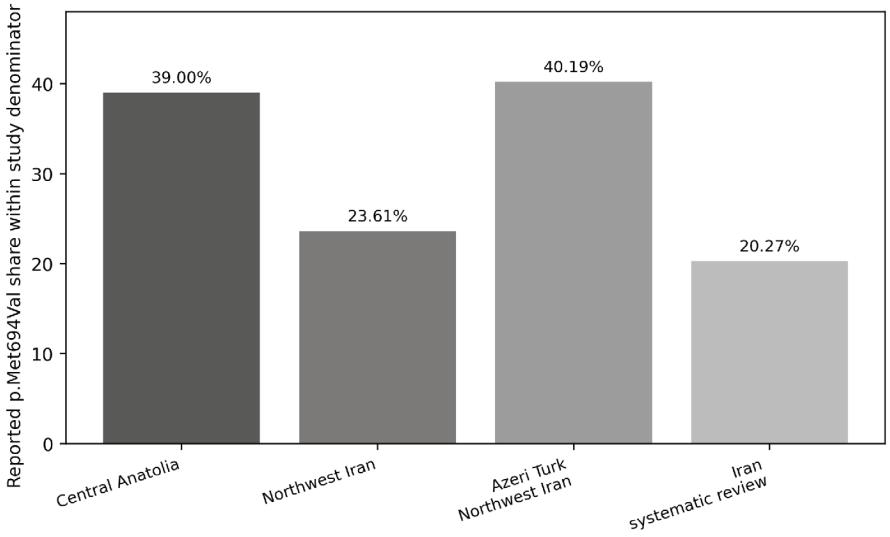

Figure 2. *Reported share of MEFV c.2080A>G (p.Met694Val) within selected clinical sources. Cohort selection, diagnostic criteria, laboratory methods, and denominators differ, so the bars should not be interpreted as controlled population comparisons.*

The interpretation of c.442G>C (p.Glu148Gln; E148Q) requires particular caution. A study of 224 healthy individuals in northwest Iran reported p.Glu148Gln in 18.3% of participants, a higher proportion than in the affected cohort from the same region (27). GeneReviews notes disagreement over whether p.Glu148Gln is pathogenic or benign and lists it as a variant of uncertain clinical significance (28). ClinVar has also contained conflicting submissions and emphasizes its high frequency among healthy individuals (29). The variant may act as a low-penetrance allele or modifier in some genetic contexts, but it should not be grouped automatically with well-established pathogenic exon 10 variants such as p.Met694Val.

Regional recurrence of p.Met694Val, p.Met680Ile, and p.Val726Ala is consistent with historically connected populations and possible founder effects. A classic review identified a recurring group of major *MEFV* variants across several Mediterranean populations (30). A separate population-genetic study found patterns in the *MEFV* region consistent with balancing selection and overdominance (31). However, the cohort percentages summarized here do not independently prove migration, a single founder origin, or heterozygote advantage. Those conclusions would require shared haplotypes, linkage disequilibrium, demographic modeling, and functional evidence that a heterozygous genotype altered survival or reproduction.

HBB, MALARIA SELECTION, AND REGIONAL FOUNDER EFFECTS

Beta thalassemia is caused by pathogenic variants in the *HBB* gene that reduce or eliminate beta-globin production. The broad geographic distribution of beta thalassemia overlaps regions where malaria was historically endemic, and heterozygote advantage is a well-supported explanation for the persistence of hemoglobin-disorder alleles at a regional level (10, 11). This broad evolutionary relationship does not mean that each local *HBB* variant was individually favored by malaria. The predominance of a particular mutation within one province may instead reflect migration, drift, consanguinity, or local founder events.

In Hamadan Province, 41 referred patients underwent *HBB* sequencing. The splice-site variant c.315+1G>A (historical name IVS-II-1 G>A) accounted for 25.61% of mutant alleles. The five most frequent pathogenic variants together accounted for 82.92% of mutant alleles (32). Because the study was small and hospital-based, these percentages describe the referred affected cohort and should not be treated as a province-wide carrier estimate.

A Tehran reference-clinic study included 2,315 patients who were homozygous or compound heterozygous for *HBB* variants and had been referred from multiple provinces between 2001 and 2016. The variants c.315+1G>A and c.92+5G>C (historical name IVS-I-5 G>C) together represented 62.40% of mutant alleles (33). This large referral cohort supports the importance of these variants in Iranian clinical practice, but it is enriched for diagnosed biallelic disease and is not a random sample of the Iranian population.

A study of 1,635 people with beta thalassemia trait in Mazandaran Province identified 24 *HBB* variants using reverse dot-blot testing and restriction-fragment methods. The c.315+1G>A variant represented 61.0% of characterized mutant alleles (34). Importantly, the study found this variant on different haplotypes and described it as the most heterogeneous of the common mutations. That finding weakens a simple single-founder explanation even though the regional concentration is striking.

Mediterranean and North African comparisons show the same principle: a limited group of pathogenic variants may dominate locally, but the leading variants differ by region. A multiplex assay designed for Mediterranean variants included c.92+1G>A, c.92+6T>C, c.93-21G>A, and c.118C>T among its targets (35). In an Albanian study, five variants accounted for nearly 90% of the 144 characterized beta thalassemia chromosomes, while combined Albanian studies found the same five variants in 92% of 379 mutant chromosomes (36). In eastern Morocco, 39 cases from 33 families yielded 52 mutant alleles, and c.118C>T (p.Gln40Ter; historical name codon 39 C>T) represented 72.54% (37). These studies support regional testing strategies, but their patient selection and laboratory methods differ.

The strongest causal claim that can be made is therefore at the level of hemoglobin disorders as a group: malaria-related selection contributed to the persistence of beta thalassemia alleles in historically endemic regions (10, 11). The predominance of c.315+1G>A in parts of Iran or c.118C>T in eastern Morocco requires a separate demographic explanation. Demonstrating a founder event for a specific *HBB* variant would require haplotype identity, genealogical links, or a demographic model showing descent from a small ancestral population.

Table 3. *Selected HBB evidence with denominator and regional context. Percentages describe mutant alleles within selected clinical or carrier cohorts, not general-population allele frequencies.*

Source and population	Evidence type and denominator	Selected finding	Interpretive caution
Hamadan Province, Iran (32)	41 referred patients; HBB sequencing; mutant-allele denominator	c.315+1G>A 25.61%; top five pathogenic variants 82.92%	Small hospital referral cohort.
Tehran reference clinic (33)	2,315 homozygous or compound heterozygous patients referred from multiple provinces	c.315+1G>A and c.92+5G>C together 62.40% of mutant alleles	Large clinical cohort enriched for diagnosed biallelic disease.
Mazandaran Province, Iran (34)	1,635 people with beta-thalassemia trait; 24 characterized variants	c.315+1G>A 61.0% of characterized mutant alleles	Variant occurred on different haplotypes, weakening a simple single-founder explanation.
Albanian cohort (36)	144 characterized beta-thalassemia chromosomes	Five variants accounted for nearly 90%	Regional clinical sample; combined Albanian literature reported 92% of 379 chromosomes.
Eastern Morocco (37)	39 cases from 33 families; 52 mutant alleles	c.118C>T (p.Gln40Ter) 72.54%	Small family-based regional cohort.

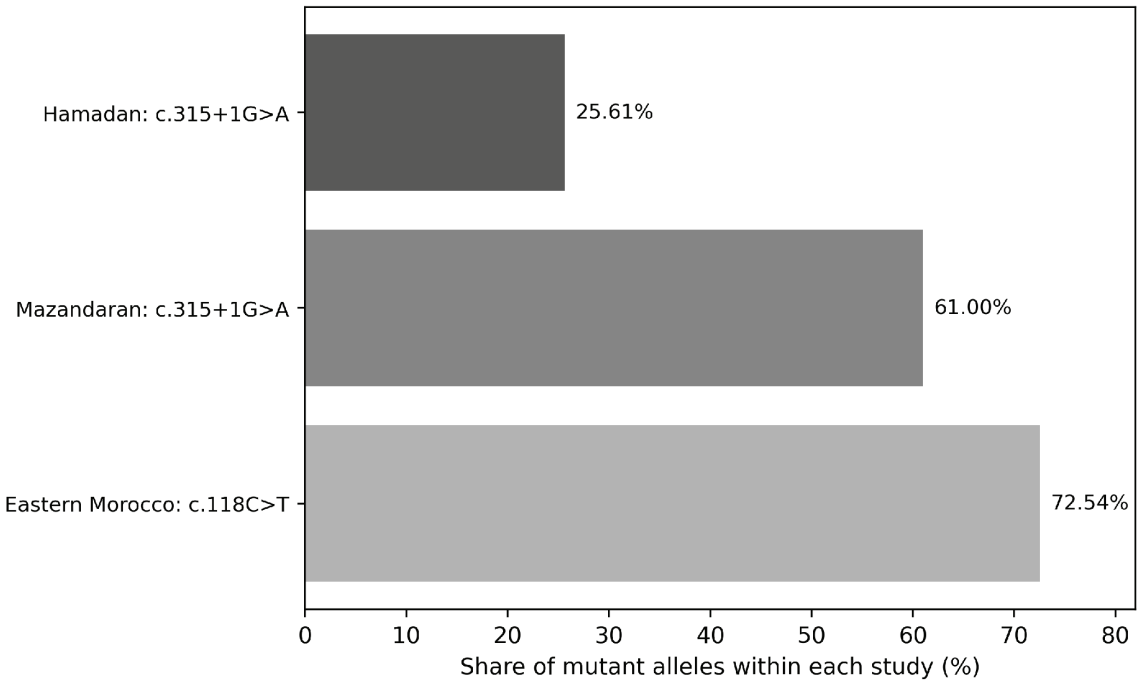

Figure 3. *Share of the leading HBB variant within three regional studies. The leading variant is c.315+1G>A in Hamadan and Mazandaran, but c.118C>T in eastern Morocco. The cohorts and methods differ, and the figure illustrates regional concentration rather than a direct comparison of prevalence.*

COMPARATIVE POPULATION-GENETIC INTERPRETATION

The three genes do not support one universal explanation. The Northern Irish *CFTR* literature shows a stable concentration around p.Phe508del, which is consistent with population structure and drift but does not prove a founder event. The *MEFV* literature shows recurring regional variants, yet the uncertain interpretation of p.Glu148Gln and the use of limited panels make a simple founder-mutation narrative unreliable. The *HBB* literature offers the strongest evidence for natural selection at the broader disease-class level, but local mutation spectra still require demographic explanations.

A founder effect is more than a common allele in one population. It implies that copies of the allele are identical by descent from a limited number of founders, usually supported by a shared haplotype, linkage disequilibrium, and a plausible demographic history (38). Similarly, distinguishing drift from selection requires comparing observed patterns with neutral demographic expectations. High disease-variant frequencies can result from founder effects, drift, selection, or combinations of these mechanisms (39). Frequency concentration should therefore be treated as a clue rather than proof.

Ancient DNA studies establish that major migrations connected parts of Europe and Asia (4, 5). They do not show that *CFTR* p.Phe508del, *MEFV* p.Met694Val, and *HBB* c.315+1G>A traveled together from one common ancestral population. Each variant has a different age, genomic context, and regional history. The evidence reviewed here is most consistent with a mixed model in which broad migration created background genetic connections, while local demographic events and environmental pressures shaped individual disease-associated variants.

IMPLICATIONS FOR GENETIC SCREENING

The clearest clinical implication is that screening panels should be evaluated against local mutation data. A panel based on common Northern European *CFTR* variants may have lower coverage in an Iranian clinical population with a broader reported spectrum. This does not mean that p.Phe508del is unimportant in Iran or that targeted panels are ineffective. It means that a panel should be validated using local affected-patient data and, where possible, sequencing-based follow-up (23).

The same principle applies to FMF and beta thalassemia. A limited *MEFV* panel may miss clinically diagnosed patients and may also detect low-penetrance variants whose significance is uncertain (25, 29, 30). Regional *HBB* panels can be efficient when a small number of pathogenic variants account for most mutant alleles, but sequencing remains important when the local spectrum is heterogeneous (32, 35-37). Broad ancestry labels may guide an initial hypothesis, but they should not replace precise population data, variant classification, and clinical context.

Table 4. *Comparative interpretation of the three gene examples. Frequency patterns suggest hypotheses but do not establish historical mechanisms without additional evidence.*

Gene and disease	Observed pattern	Cautious interpretation	Evidence needed for stronger causal inference
CFTR and cystic fibrosis	Northern Irish affected-patient spectra are concentrated around p.Phe508del; Iranian reports are more heterogeneous.	Consistent with population structure and drift. Heterozygote advantage has been proposed but is not established by these frequencies.	Shared haplotypes, variant-age estimates, demographic models, and functional evidence tied to historical fitness.
MEFV and familial Mediterranean fever	Several variants recur across Mediterranean and Iranian cohorts, with substantial local variation.	May reflect historically connected populations, founder effects, panel design, and penetrance differences.	Haplotype and linkage-disequilibrium evidence, broader sequencing, and functional clarification of low-penetrance variants.
HBB and beta thalassemia	Hemoglobin disorders overlap malaria-endemic regions, while leading HBB variants differ locally.	Malaria-related selection is supported broadly; local variant predominance may reflect founder and demographic processes.	Variant-specific haplotypes, genealogical evidence, and models separating selection from demography.

LIMITATIONS OF THE EVIDENCE

First, the main sources are not fully commensurable. The Northern Ireland *CFTR* evidence includes a complete regional registry-based affected population, whereas the Iranian *CFTR* estimate comes from a systematic review that combined 16 separate studies. The pooled Iranian result may include different diagnostic criteria, geographic regions, laboratory methods, and periods of data collection. It is therefore inappropriate to interpret the difference between 62.3% and 21.22% as a controlled population comparison.

Second, affected-patient mutation spectra, carrier frequencies, disease incidence, newborn screening outcomes, and reference-population allele frequencies use different denominators. Each provides useful information, but none can be substituted for another. The tables and figures in this review separate these evidence types, and the figures are intended as descriptive summaries only.

Third, the geographic labels remain broad. Northern Ireland and the Republic of Ireland are not identical populations, and Mediterranean-associated populations include groups with different histories. Iran contains substantial ethnic and regional diversity, which is one reason Iranome was developed (13). Several Iranian systematic reviews also report uneven geographic coverage (7, 9).

Fourth, variant detection is affected by study design. Referral clinics may overrepresent severe disease, targeted panels miss untested variants, and older studies had less complete sequencing than newer studies. Publication bias may also make known or panel-included variants more likely to appear repeatedly in the literature. Possible overlap among cohorts can further inflate the apparent amount of independent evidence.

Finally, frequency data alone cannot establish historical cause. Founder effects require haplotype and demographic evidence, while natural selection requires functional or population-genetic evidence that exceeds a simple geographic correlation. The interpretations offered here are therefore expressed as mechanisms that are consistent with the evidence, not as proven explanations.

CONCLUSION

The published evidence shows that disease-associated mutation patterns are strongly population-specific, but the mechanisms differ among *CFTR*, *MEFV*, and *HBB*. Northern Irish *CFTR* studies report a concentrated affected-patient spectrum dominated by p.Phe508del, while the Iranian literature reports a wider range of *CFTR* variants. *MEFV* studies repeatedly identify p.Met694Val and several other regional variants, but differences in panel design and the debated significance of p.Glu148Gln limit simple comparisons. For *HBB*, malaria-related selection provides a strong broad explanation for the persistence of beta thalassemia, while local founder and demographic processes may explain why different pathogenic variants dominate in particular regions. Shared migration history is therefore useful background, but it does not replace variant-specific evidence. The most defensible clinical conclusion is that limited panels may have lower coverage in genetically heterogeneous populations and require local validation. Future research should combine complete sequencing, consistent denominators, haplotype analysis, and better representation of under-studied populations.

ACKNOWLEDGEMENTS

The author thanks Sindhura Siddapureddy for mentorship throughout the Lumiere Research Fellowship. The author also thanks Grace for writing feedback and the Lumiere Research Fellowship team for support during the research and revision process.

FUNDING SOURCES

The author received no external research funding.

AI DISCLOSURE STATEMENT

During the preparation of this work, the author used ChatGPT for brainstorming, manuscript structure, spreadsheet organization, candidate reference identification, language editing, grammar review, and formatting assistance. Following the use of this tool, the author carefully reviewed, verified, and edited the content as necessary and takes full responsibility for the accuracy, originality, and integrity of the submitted work. AI tools were not used as scientific sources.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest related to this work.

REFERENCES

1. Prohaska A, Racimo F, Schork AJ, *et al*. Human disease variation in the light of population genomics. *Cell*. 2019; 177 (1): 115-131. doi:10.1016/j.cell.2019.01.052.
2. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016; 538 (7624): 161-164. doi:10.1038/538161a.
3. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. *Cell*. 2019; 177 (1): 26-31. doi:10.1016/j.cell.2019.02.048.
4. Haak W, Lazaridis I, Patterson N, *et al*. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 2015; 522 (7555): 207-211. doi:10.1038/nature14317.
5. Narasimhan VM, Patterson N, Moorjani P, *et al*. The formation of human populations in South and Central Asia. *Science*. 2019; 365 (6457): eaat7487. doi:10.1126/science.aat7487.
6. Blevings PJ, Moore JE, McCaughan J, Reid A, Rendall JC, Millar BC. Mutation characterisation of the cystic fibrosis transmembrane conductance regulator (CFTR) gene in people with cystic fibrosis in Northern Ireland. *Ulster Med J.* 2025; 94 (2): 64-76.
7. Alibakhshi R, Mohammadi A, Khamooshian S, Kazeminia M, Moradi K. CFTR gene mutation spectrum among 735 Iranian patients with cystic fibrosis: a comprehensive systematic review. *Pediatr Pulmonol.* 2021; 56 (12): 3644-3656. doi:10.1002/ppul.25647.
8. Gershoni-Baruch R, Shinawi M, Leah K, Badarnah K, Brik R. Familial Mediterranean fever: prevalence, penetrance and genetic drift. *Eur J Hum Genet*. 2001; 9 (8): 634-637. doi:10.1038/sj.ejhg.5200672.
9. Alibakhshi R, Mohammadi A, Ghadiri K, Khamooshian S, Kazeminia M, Moradi K. Spectrum of MEFV gene mutations in 4,256 familial Mediterranean fever patients from Iran: a comprehensive systematic review. *Egypt J Med Hum Genet.* 2022; 23: 5. doi:10.1186/s43042-022-00222-y.
10. Weatherall DJ. Recent insights into the population genetics and dynamics of the inherited disorders of hemoglobin. *Mediterr J Hematol Infect Dis*. 2009; 1 (1): e200922. doi:10.4084/MJHID.2009.022.
11. Hedrick PW. Population genetics of malaria resistance in humans. *Heredity*. 2011; 107 (4): 283-304. doi:10.1038/hdy.2011.16.
12. Karczewski KJ, Francioli LC, Tiao G, *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020; 581 (7809): 434-443. doi:10.1038/s41586-020-2308-7.
13. Iranome. About Iranome. Available from: https://www.iranome.com/about (accessed on 2026-08-01).
14. 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): 405-424. doi:10.1038/gim.2015.30.
15. Hughes DJ, Hill AJ, Macek M Jr, Redmond AO, Nevin NC, Graham CA. Mutation characterization of CFTR gene in 206 Northern Irish CF families: thirty mutations, including two novel, account for approximately 94% of CF chromosomes. *Hum Mutat*. 1996; 8 (4): 340-347.
16. Sasaki E, Kostocenko M, Lang N, *et al*. National newborn screening for cystic fibrosis in the Republic of Ireland: genetic data from the first 6.5 years. *Eur J Hum Genet*. 2020; 28 (12): 1669-1674. doi:10.1038/s41431-020-0661-5.
17. Sosnay PR, Siklosi KR, Van Goor F, *et al*. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet.* 2013; 45 (10): 1160-1167. doi:10.1038/ng.2745.
18. ClinVar. NM_000492.4(CFTR):c.350G>A (p.Arg117His). Available from: https://www.ncbi.nlm.nih.gov/clinvar/variation/7109/ (accessed on 2026-08-01).
19. Gabriel SE, Brigman KN, Koller BH, Boucher RC, Stutts MJ. Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science*. 1994; 266 (5182): 107-109. doi:10.1126/science.7524148.
20. Pier GB, Grout M, Zaidi T, *et al*. Salmonella typhi uses CFTR to enter intestinal epithelial cells. *Nature*. 1998; 393 (6680): 79-82. doi:10.1038/30006.
21. Poolman EM, Galvani AP. Evaluating candidate agents of selective pressure for cystic fibrosis. *J R Soc Interface*. 2007; 4 (12): 91-98. doi:10.1098/rsif.2006.0154.
22. Beckett KM. The cystic fibrosis heterozygote advantage: a synthesis of ideas. *Anthropologica*. 1997; 39 (1-2): 147-158.
23. Watson MS, Cutting GR, Desnick RJ, *et al*. Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. *Genet Med*. 2004; 6 (5): 387-391. doi:10.1097/01.GIM.0000139506.11694.7C.
24. Yasar Bilge NS, Sari I, Solmaz D, *et al*. Prevalence of MEFV gene mutations in a large cohort of patients with suspected familial Mediterranean fever in Central Anatolia. *Ann Saudi Med*. 2019; 39 (6): 382-387. doi:10.5144/0256-4947.2019.382.
25. Salehzadeh F, Emami D, Zolfegari AA, *et al*. MEFV gene profile in northwest of Iran: twelve common MEFV gene mutations in 216 patients with familial Mediterranean fever. *Iran J Med Sci*. 2015; 40 (1): 68-72.
26. Mohammadnejad L, Farajnia S. Mediterranean

fever gene analysis in the Azeri Turk population with familial Mediterranean fever: evidence for new mutations associated with disease. *Cell J*. 2013; 15 (2): 152-159.

27. Salehzadeh F, Sharghi A, Hoseini Asl S, *et al*. MEFV gene variant alleles in normal population of northwest of Iran, which is near to Mediterranean Sea. *Genet Res Int*. 2019; 2019: 6418759. doi:10.1155/2019/6418759.
28. Shohat M, Halpern GJ. Familial Mediterranean fever. In: Adam MP, Feldman J, Mirzaa GM, *et al*., editors. GeneReviews [Internet]. Seattle: University of Washington, Seattle; 2000-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1227/ (accessed on 2026-08-01).
29. ClinVar. NM_000243.3(MEFV):c.442G>C (p.Glu148Gln) and familial Mediterranean fever. Available from: https://www.ncbi.nlm.nih.gov/clinvar/RCV000002651/ (accessed on 2026-08-01).
30. Touitou I. The spectrum of familial Mediterranean fever mutations. *Eur J Hum Genet*. 2001; 9 (7): 473-483. doi:10.1038/sj.ejhg.5200658.
31. Fumagalli M, Cagliani R, Pozzoli U, *et al*. A population genetics study of the familial Mediterranean fever gene: evidence of balancing selection under an overdominance regime. *Genes Immun*. 2009; 10 (8): 678-686. doi:10.1038/gene.2009.59.
32. Jalilian M, Azizi Jalilian F, Ahmadi L, *et al*. The frequency of HBB mutations among beta-thalassemia patients in Hamadan Province, Iran. *Hemoglobin*. 2017; 41 (1): 61-64. doi:10.1080/03630269.2017.1302468.
33. Bazazzadegan N, Abedini SS, Azarkeivan A, *et al*. The spectrum of HBB mutations among 2315 beta thalassemia patients of a reference clinic in Tehran, Iran. *Hemoglobin*. 2023; 47 (4): 147-151. doi:10.1080/03630269.2023.2242787.
34. Akhavan-Niaki H, Derakhshandeh-Peykar P, Banihashemi A, *et al*. A comprehensive molecular characterization of beta thalassemia in a highly heterogeneous population. *Blood Cells Mol Dis*. 2011; 47 (1): 29-32. doi:10.1016/j.bcmd.2011.03.005.
35. Atanasovska B, Bozhinovski G, Chakalova L, *et al*. Efficient detection of Mediterranean beta-thalassemia mutations by multiplex single-nucleotide primer extension. *PLoS One*. 2012; 7 (10): e48167. doi:10.1371/journal.pone.0048167.
36. Babameto-Laku A, Mitre A, Berisha S, Mokini V, Roko D. Molecular genetic characterization of beta-thalassemia and sickle cell syndrome in the Albanian population. *Balkan J Med Genet*. 2011; 14 (1): 45-50. doi:10.2478/v10034-011-0017-0.
37. Belmokhtar I, Lhousni S, Errahhali ME, *et al*. Molecular heterogeneity of beta-thalassemia variants in the Eastern region of Morocco. *Mol Genet Genomic Med*. 2022; 10 (8): e1970. doi:10.1002/mgg3.1970.
38. Slatkin M. A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. *Am J Hum Genet*. 2004; 75 (2): 282-293. doi:10.1086/423146.
39. Zlotogora J. High frequencies of human genetic diseases: founder effect with genetic drift or selection? *Am J Med Genet*. 1994; 49 (1): 10-13. doi:10.1002/ajmg.1320490104.