Received: 6 January 2021 | Revised: 22 June 2021 | Accepted: 24 June 2021

DOI: 10.1002/humu.24250

RESEARCH ARTICLE

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Elucidating clinical phenotypic variability associated with the polyT tract and TG repeats in CFTR

Keith Nykamp | **Rebecca Truty** | **Darlene Riethmaier** | **Julia Wilkinson** | **Sara L. Bristow** | **Sienna Aguilar** | **Dana Neitzel** | **Nicole Faulkner** | **Swaroop Aradhya**

Invitae, San Francisco, California, USA

Correspondence
Keith Nykamp, Invitae, 1400 16th St, San Francisco, CA, USA.
Email: keith.nykamp@invitae.com

Abstract

Biallelic pathogenic variants in *CFTR* manifest as cystic fibrosis (CF) or other CFTR-related disorders (CFTR-RDs). The 5T allele, causing alternative splicing and reduced protein activity, is modulated by the adjacent TG repeat element, though previous data have been limited to small, selective cohorts. Here, the risk and spectrum of phenotypes associated with the *CFTR* TG-T5 haplotype variants (TG11T5, TG12T5, and TG13T5) in the absence of the p.Arg117His variant are evaluated. Individuals who received physician-ordered next-generation sequencing of *CFTR* were included. TG[11-13]T5 variant frequencies (biallelic or with another CF-causing variant [CFvar]) were calculated. Clinical information reported by the ordering provider or the individual was examined. Among 548,300 individuals, the T5 minor allele frequency (MAF) was 4.2% (TG repeat distribution: TG11 = 68.1%, TG12 = 29.5%, TG13 = 2.4%). When present with a CFvar, each TG[11-13]T5 variant was significantly enriched in individuals with a high suspicion of CF or CFTR-RD (personal/family history of CF/CFTR-RD) compared to those with a low suspicion for CF or CFTR-RD (hereditary cancer screening, *CFTR* not requisitioned). Compared to CFvar/CFvar individuals, those with TG[11-13]T5/CFvar generally had single-organ involvement, milder symptoms, variable expressivity, and reduced penetrance. These data improve our understanding of disease risks associated with TG[11-13]T5 variants and have important implications for reproductive genetic counseling.

KEYWORDS

CFTR, CFTR-related disorder, cystic fibrosis, genetic counseling, penetrance, polyT tract, TG repeat

1 | INTRODUCTION

Cystic fibrosis (CF) and CFTR-related disorders (CFTR-RDs), caused by biallelic pathogenic variants in the *CFTR* gene, have a highly variable clinical phenotypic spectrum (Castellani & Assael, 2017; Farrell et al., 2017; Sosnay et al., 2016). Pathogenic variants in *CFTR* can alter the abundance, stability, or function of the encoded CFTR protein to varying degrees (Wang et al., 2014). The severity of CF or a CFTR-RD is largely dependent on the types of pathogenic variants and the resulting residual CFTR activity. Very low residual, cellular CFTR activity (<2% of normal), typically leads to a classic CF phenotype, characterized by multiorgan involvement, including severe lung and liver disease, pancreatic exocrine insufficiency, sweat gland dysfunction, and male infertility due to congenital absence of vas deferens (CAVD) (Knowles & Drumm, 2012). However, when residual, cellular CFTR activity ranges between 4% and 25%, nonclassic CF phenotypes are often observed, such as pancreatic sufficiency (i.e., CF-PS), isolated recurrent pancreatitis, bronchiectasis, chronic sinusitis, and/or CAVD (males only) (Castellani & Assael, 2017; Ramalho et al., 2002). In addition to residual activity, genetic modifiers, and/or environmental factors can also result in variable expressivity and penetrance of CFTR-related symptoms (Cutting, 2010).

As a class, *CFTR* variants that affect messenger RNA (mRNA) splicing are often associated with milder symptoms and reduced penetrance due to variable defects in splicing and protein activity. One example of this is illustrated by the polymorphic repeat region of intron 9, consisting of a thymidine-guanine (TG) repeat element (generally ranging from 10 to 13 successive TGs, NM_000492.4:c.1210-34TG[10], NM_000492.4:c.1210-34TG[11], NM_000492.4:c.1210-34TG[12], NM_000492.4:c.1210-34TG[13]) and a thymidine (polyT) tract (most often 9, 7, or 5 uninterrupted thymidines [T9, T7, T5]; NM_000492.4:c.1210-12T[9], NM_000492.4:c.1210-12T[7], NM_000492.4:c.1210-12T[5]). The length of the polyT tract is a major determinant of whether exon 10 is included or excluded in the *CFTR* mRNA during splicing (Chu et al., 1992, 1993; Cuppens et al., 1998; Mak et al., 1997; Rave-Harel et al., 1997; Teng et al., 1997). In individuals who are compound heterozygous for T9 and/or T7 alleles, greater than 75% of the *CFTR* mRNA includes exon 10 and is full length, while the remaining fraction of mRNA lacks exon 10 and encodes a nonfunctional CFTR protein (Delaney et al., 1993; Niksic et al., 1999; Strong et al., 1993). By contrast, individuals who have T5 on one or both chromosomes express <50% full-length CFTR, but this is further modulated by the number of TG repeats: the presence of a TG11T5 allele results in ~50% full-length CFTR (NM_000492.4:c.1210-34TG[11]T[5]), a TG12T5 allele results in ~25% (NM_000492.4:c.1210-34TG[12]T[5]), and a TG13T5 allele results in <25% (NM_000492.4:c.1210-34TG[13]T[5]) (Chu et al., 1991; Hefferon et al., 2004; Niksic et al., 1999).

Importantly, decreasing residual CFTR protein activity associated with the three TG[11-13]T5 haplotypes correlates with increasing severity and penetrance of CFTR-related symptoms. For example, CF-PS was reported in all children with a TG[11-13]T5 and a CF-causing variant (CFvar), and ~38% of these children with TG13T5/CFvar and in ~6% of children with TG12T5/CFvar ended up with a CF diagnosis (Salinas et al., 2016), though these rates may change with adolescence or adulthood. This is true for both males and females with a CFvar (i.e., <2% CFTR activity) on the opposite chromosome. However, in the same study, children with the TG11T5/CFvar genotype were not diagnosed with CF in follow-up evaluations, likely due to the elevated residual CFTR activity in these individuals (Salinas et al., 2016). CAVD, along with milder CF-related phenotypes, such as idiopathic pancreatitis, bronchiectasis, chronic rhinosinusitis, and intermediate sweat chloride levels, has been reported in >95% of males with TG13T5/CFvar, 75%–85% of males with TG12T5/CFvar, and 35%–55% of males with TG11T5/CFvar (Groman et al., 2004; Sun et al., 2006). Approximately 5% of females with a TG11T5/CFvar genotype have been reported to manifest milder CFTR-RDs (Sun et al., 2006). In addition, the common p.Arg117His variant (NM_000492.4:c.350G>A, p.Arg117His) has been found to occur on the same chromosome (*in cis*) as the TG12T5 variant at a low frequency (Kiesewetter et al., 1993; Thauvin-Robinet et al., 2009). While p.Arg117His may increase the penetrance and severity of the TG12T5/CFvar genotype, it is not required for the disease to manifest since overt CFTR-RDs have been associated with the TG12T5/CFvar genotype in the absence of p.Arg117His (Salinas et al., 2016).

Considering the relatively high prevalence of the 5T allele (~10% in the general population), and the wide disparity in clinical outcomes among individuals with TG11T5 compared to those with TG12T5 or TG13T5, it has been recommended that clinical interpretation and reporting of the 5T variant should be done in the context of the number of TG repeats, and irrespective of the presence of p.Arg117His (Deignan et al., 2020; Groman et al., 2004). Accordingly, we custom-developed a bioinformatics haplotype caller to accurately determine the T/TG genotype for each individual who had *CFTR* sequencing analysis performed. Here, we report our findings from more than 500,000 individuals tested through Invitae Corporation ("Invitae"), a commercial genetic testing laboratory, including the likelihood that individuals with TG[11-13]T5 genotypes are affected with CF and CFTR-RDs.

2 | MATERIALS AND METHODS

2.1 | Genetic testing panels and study population

Individuals tested through Invitae between 2014 and 2019 with *CFTR* sequencing results were included. *CFTR* sequencing is performed using targeted hybrid capture next-generation sequencing (NGS) assays designed for preconception carrier screening and diagnostic testing for pediatric disorders and for hereditary cancer syndromes. In cases where *CFTR* sequencing was included in an assay but not ordered by the referring clinician, the results were not reported to the clinician, even though the sequencing data was generated. Review and analysis of fully deidentified data were approved by the Western Institutional Review Board (WIRB 1167406).

2.2 | Genetic testing and variant interpretation

Genetic testing was conducted using targeted gene panels via short-read NGS as described previously (Lincoln et al., 2015; Truty et al., 2019). The homopolymer polyT and dinucleotide TG repeat region in *CFTR* can be difficult to accurately disambiguate using standard NGS bioinformatics tools. Therefore, an algorithm was developed and validated to specifically call the T/TG haplotypes. Interpretation of the T/TG haplotypes and all *CFTR* variants was performed as previously described (Hannah et al., 2019), using the proprietary Sherloc variant interpretation framework (Nykamp et al., 2017), which was based on guidelines from the American College of Medical Genetics and Genomics and the Association of Molecular Pathology (ACMG/AMP) (Richards et al., 2015). Notably, the T9 and T7 alleles and the TG10T5 variant have been classified as benign since they are not associated with the disease. The TG13T5 and TG12T5 variants are classified as pathogenic variants since they are associated with CF-PS or other CFTR-RDs in the majority of individuals with a CFvar on the opposite chromosome. The TG11T5 variant has been classified as pathogenic (low penetrance) due to much lower penetrance for CFTR-RDs compared to TG13T5 and TG12T5 alleles.

2.3 | CFTR cohorts and data analysis

All analyses were conducted on deidentified data sets. Three analyses were conducted within the study population. First, individuals were divided into those who had requested *CFTR* sequencing results in a clinical report (*CFTR* requested) and those who did not (sequencing of *CFTR* was performed but results were not requested or analyzed). The allele frequency for each of the TG[11-13]T5 variants was calculated for the total population and for each of the two subgroups individually. *T* tests were performed to determine if the MAF was different between these two groups ($p < .05$ considered significant).

Second, individuals were defined as having a "high suspicion" or "low suspicion" of CF or a CFTR-RD. The high suspicion cohort included individuals for whom a personal and/or family history of CF or a CFTR-RD was reported by the ordering physician on the test requisition form, regardless of whether *CFTR* testing was ordered by the clinician, using the following keywords: CF, sweat chloride, immunoreactive trypsinogen, CAVD, meconium ileus. The low suspicion group included individuals referred for hereditary cancer testing and did not request *CFTR* analysis as part of the test. All other individuals in the study population were excluded from this analysis. Within each of these groups, the proportion of individuals with each of the following genotype combinations was calculated: CFvar/CFvar, CFvar/TG[11-13]T5, TG13T5/TG13T5, TG12T5/TG12T5, TG11T5/TG11T5, TG13T5/TG12T5, TG13T5/TG11T5, TG12T5/TG11T5. The phase of the various allele combinations was not determined in all individuals in this study, though it is very likely that the vast majority of individuals in this study with CFvar/CFvar or CFvar/TG[11-13]T5 genotypes had one variant on each chromosome, given that none of the CFvar observed have been reported to be *in cis* with each other or with the TG[11-13]T5 alleles. The common p.Arg117His variant is classified as a variant of uncertain significance (VUS), and individuals with this variant were excluded from the analysis, due to uncertainty regarding the pathogenicity of this variant alone (de Nooijer et al., 2011; Thauvin-Robinet et al., 2009). Furthermore, since the p.Arg117His variant has been reported to modulate the severity of the TG12T5 variant but not the TG13T5 or TG11T5 variants (Kiesewetter et al., 1993; Waller & Simmonds, 2016), including individuals with the p.Arg117His variant would confound analyses comparing TG13T5, TG12T5, and TG11T5 alleles. Odds ratios (ORs) comparing the high and low suspicion groups were calculated for each genotype. Confidence intervals and *p* values were calculated for each OR with the MedCalc's OR online calculator (see Web Resources section). Pearson's χ^2 analysis compared OR effect across the different genotypes.

Third, phenotypic variability and severity associated with the CFvar/TG[11-13]T5 genotypes were assessed by reviewing CFTR-related symptoms (Table S1) in individuals with CFvar/CFvar or one of the CFvar/TG[11-13]T5 genotypes. Clinical information included in the analysis was provided either at the time of requisition by the referring clinicians or during a consultation with an Invitae genetic counselor by the tested individual. All clinical information was provided at the discretion of the ordering clinician or of the tested individual.

3 | RESULTS

3.1 | Summary of the study population

In total, 548,300 individuals had *CFTR* sequencing performed between 2014 and 2019 (~1.1 million alleles); *CFTR* sequencing results were specifically requisitioned for 34,195 individuals. The mean age at testing was 47.0 years and the majority of individuals were female (78.3%) and of self-reported White/Caucasian ethnicity (61.3%) (Table 1). The MAF of the T5 allele (4.2%) and the distribution of the TG repeats were consistent with previous reports (Groman et al., 2004; Sun et al., 2006) and similar regardless of whether *CFTR* results were requisitioned for the individual (Table 2).

3.2 | Distribution of T5 alleles in individuals with a high suspicion of a CFTR-related disorder

Individuals in the high suspicion group were more evenly split among males and females (48.0% vs. 52.0%, respectively) compared to the low suspicion group (15.8% vs. 84.2%). Age at time of testing was younger in the high suspicion group versus low suspicion group (mean 24.7 years vs. 52.6 years) (Table 1). When present with a CFvar (Table S2), each of the TG[11-13]T5 variants were significantly

enriched in individuals with a high suspicion of CF or a CFTR-RD compared to those with very low suspicion for CF or a CFTR-RD (Table 3). Enrichment was significantly lower ($p < .001$) for the CFvar/TG11T5 genotype (OR = 6.9) than for CFvar/TG12T5 (OR = 29.0) and CFvar/TG13T5 (OR = 78.8) genotypes. TG11T5 and TG12T5 homozygotes were absent from the high suspicion group. By contrast, TG13T5 homozygotes were significantly enriched (OR = 255.7) in the high suspicion group, similar ($p = .56$) to the CFvar/CFvar genotype (OR = 581.7), though it should be interpreted with caution as only three individuals in the cohort were homozygous for TG13T5.

TABLE 1 Demographic characteristics of study population

Characteristic	Total (n = 548,300)	CF/CFTR-RD "high" suspicion (n = 860)	CF/CFTR-RD "low" suspicion (n = 439,331)
Sex, n (%)[a]			
Female	429,414 (78.3)	447 (52.0)	369,699 (84.2)
Male	118,871 (21.7)	413 (48.0)	69,620 (15.8)
Age (years)			
Mean	47.0	24.7	52.6
Median	49	25	53
Race/ethnicity[b]			
White/Caucasian	335,955 (61.3)	523 (60.8)	284,579 (64.8)
Hispanic	46,640 (8.5)	138 (16.0)	30,776 (7.0)
Black/African-American	34,497 (6.3)	30 (3.5)	27,804 (6.3)
Asian	24,019 (4.4)	37 (4.3)	16,328 (3.7)
Askenazi Jewish	20,373 (3.7)	9 (1.0)	18,183 (4.1)
French Canadian	2626 (0.4)	2 (0.2)	2412 (0.5)
Native American	1878 (0.3)	2 (0.2)	1486 (0.3)
Mediterranean	1636 (0.3)	3 (0.3)	1020 (0.2)
Sephardic Jewish	1269 (0.2)	1 (0.1)	876 (0.2)
Pacific Islander	1233 (0.2)	0	837 (0.2)
Multiethnic	11,632 (2.1)	23 (2.7)	9566 (2.2)
Other/unknown	66,542 (12.1)	92 (10.7)	45,464 (10.3)

[a]Sex was unknown for 15 individuals in the overall population. Sex was unknown for 12 individuals included in the *CFTR* "low" suspicion group.
[b]Race and ethnicity was self-reported by the individuals receiving testing.

3.3 | Genotype-phenotype associations

Of the individuals with a CFvar/CFvar genotype, phenotype information was provided by the clinician in ~59% of cases. Phenotype information was available for 67% of individuals with a CFvar/TG13T5 genotype, for 28% of those with a CFvar/TG12T5 genotype, and for 39% of those with a CFvar/TG11T5 (n = 46 overall for CFvar/TG[11-13]T5). In this group, the chronic sinopulmonary disease was the most commonly observed category of symptoms, with 14 (30.4%) individuals listing this as a primary indication, which is similar to the percentage of CFvar/CFvar individuals (28.2%) reporting chronic sinopulmonary complications (Table 4). Recurrent pancreatitis was the next most common phenotypic category for CFvar/TG[11-13]T5 individuals (n = 10, 21.7%). Interestingly, the majority of these individuals (8/10) had the TG11T5 allele (Table 4), whereas recurrent pancreatitis was very rarely reported (1.3%) for CFvar/CFvar individuals. Eight of the individuals (17.4%) with the CFvar/TG[11-13]T5 genotype were reported to have intermediate sweat chloride levels (30–60 mEq/L), while none had diagnostically elevated levels (>60 mEq/L). This was in contrast to CFvar/CFvar individuals, among whom 28 (18.8%) had clinician-reported, diagnostically elevated sweat chloride levels and a much smaller fraction (6.0%) had intermediate levels. The remaining CFvar/CFvar individuals (75.2%) are expected to have elevated sweat chloride levels, although this information was not indicated by the ordering provider. Finally, 2 of 19 (10.5%) males with a CFvar/TG[11-13]T5 genotype reported obstructive azoospermia CAVD, which was slightly less common than in males with CFvar/CFvar genotypes (n = 6/42, 14.3%).

TABLE 2 Minor allele frequencies of *CFTR* TG[11-13]5T variants

Haplotype	*CFTR* analysis not requested (total alleles = 1,028,364)		*CFTR* analysis requested (total alleles = 69,956)		Total (total alleles = 1,098,320)	
	Alleles	MAF	Alleles	MAF	Alleles	MAF
TG[11-13]T5	42,448	0.0413	3011	0.0430	45,936	0.0418
TG11T5	29,189	0.0284	1753	0.0251	31,266	0.0285
TG12T5	12,280	0.0119	1126	0.0161	13,550	0.0123
TG13T5	979	0.000952	132	0.00189	1120	0.00102

Abbreviations: MAF, minor allele frequency; TG[11-13]T5, individual with one of the following genotypes: NM_000492.4:c.1210-34TG[11]T[5], NM_000492.4:c.1210-34TG[12]T[5], or NM_000492.4:c.1210-34TG[13]T[5]; TG11T5, NM_000492.4:c.1210-34TG[11]T[5]; TG12T5, NM_000492.4:c.1210-34TG[12]T[5]; TG13T5, NM_000492.4:c.1210-34TG[13]T[5].

TABLE 3 Frequency distribution of individuals with CFTR causative variants (CFvar) and TG[11-13]T5 variants for groups with a high suspicion and low suspicion of CF or CFTR-RD

Genotype	High suspicion (total $n = 860$)	Low suspicion (total $n = 439{,}331$)	OR (95% CI)	p value
2xCFvar	97	96	582 (435, 778)	<.0001
CFvar/TG13T5*	2	13	79 (18, 349)	<.0001
CFvar/TG12T5§	13	229	29 (17, 52)	<.0001
CFvar/TG11T5§,‡	7	523	7 (4, 15)	<.0001
Homozygous TG13T5	1	2	256 (23, 2823)	<.0001
Homozygous TG12T5§	0	108	2 (0.2, 38)	.55
Homozygous TG11T5§	0	461	0.5 (0.04, 9)	.68
TG13T5/TG12T5§	0	17	15 (0.9, 243)	.062
TG13T5/TG11T5§	0	11	22 (1, 377)	.032
TG12T5/TG11T5§	2	285	4 (0.9, 14)	.072

Abbreviations: CFvar, cystic fibrosis causing variant; CI, confidence interval; TG11T5, NM_000492.4:c.1210-34TG[11]T[5]; TG12T5, NM_000492.4:c.1210-34TG[11]T[5]; TG13T5, NM_000492.4:c.1210-34TG[13]T[5].

*$p < .05$, vs. 2xCFvar.

§$p < .01$, vs. 2xCFvar.

‡$p < .01$, vs. CFvar/TG13T5 and CFvar/TG12T5.

TABLE 4 Distribution of CFTR-related symptoms for TG[13-11]T5 variants

	CFvar/ CFvar	CFvar/ TG13T5	CFvar/ TG12T5	CFvar/ TG11T5	Total
Symptoms					
Chronic sinopulmonary	42	1	6	7	56
Pancreatic insufficiency	9	0	0	0	9
Recurrent pancreatitis	2	1	1	8	12
Gastrointestinal	22	0	5	0	27
Elevated sweat Cl	28	0	0	0	28
Intermediate sweat Cl	9	2	3	3	17
Stated CF Dx	75	0	3	0	78
CAVD	6	0	2	0	8
Not affected	0	0	3	6	9
Patient responses					
Provided indications	149 (59%)	4 (67%)	19 (28%)	23 (39%)	195 (51%)
No information provided	102 (41%)	2 (33%)	50 (72%)	36 (61%)	190 (49%)
Total individuals	251	6	69	59	385

Abbreviations: CAVD, congenital absence of the vas deferens; Cl, Chloride; CFvar, cystic fibrosis causing variant; Dx, diagnosis; TG11T5, NM_000492.4:c.1210-34TG[11]T[5]; TG12T5, NM_000492.4:c.1210-34TG[11]T[5]; TG13T5, NM_000492.4:c.1210-34TG[13]T[5].

4 | DISCUSSION

In the largest cohort to date, we describe the significant complexities underlying the clinical interpretation of genotypes involving the *CFTR* T/TG tract in the absence of the p.Arg117His variant based on evidence from a diverse population of unaffected and affected individuals. In addition, this study included a large population of individuals who presented with a range of clinical symptoms within the full spectrum of CFTR-RDs, providing further insight into the clinical outcomes for T/TG alleles. The data reported here demonstrate that

while each of the TG[11-13]T5 genotypes are associated with a milder spectrum of symptoms compared to the CFvar/CFvar genotype, the risk of disease increases with the number of TG repeats. In a recent update to the technical genetic testing standards for *CFTR*, the ACMG suggested that clinical reporting of the 5T allele should include the number of adjacent TG repeats if possible, given the relatively high frequency of 5T in the population (about 1 in 12 are carriers) and the wide disparity in disease risk associated with TG11T5 carriers compared to TG12T5 or TG13T5 carriers (Deignan et al., 2020). Studies of the clinical implications of TG[11-13]T5 variants illuminate the significant nuances in correlations between these variants and associated clinical symptoms (Groman et al., 2004), and our study further provides important clarity for interpreting the clinical significance of the T/TG haplotypes in the absence of the p.Arg117His variant. In addition to providing useful information for diagnostic testing, the data here may help to inform counseling to individuals and couples pursuing carrier screening, as it is clear that the number of TG repeats may influence the risk of developing CF or CFTR-RDs.

In comparing individuals with a high or low suspicion of CF or a CFTR-RD, enrichment in the high suspicion group was statistically lower for the CFvar/TG11T5 genotype compared to CFvar/TG12T5 and TG13T5 genotypes, consistent with reduced penetrance for TG11T5. Interestingly, TG11T5 and TG12T5 homozygotes were absent from the high suspicion group, suggesting that these genotypes rarely cause CF or CFTR-RDs. By contrast, TG13T5 homozygotes were statistically enriched in the high suspicion group and their clinical phenotype was not significantly different from that of individuals with a CFvar/CFvar genotype. In examining the relationship between TG[11-13]T5 variants and clinical indications for testing, we observed an increased but variable risk for disease associated with all three TG[11-13]T5 variants when they were present with a CFvar. Importantly, individuals with a CFvar and TG13T5 or TG12T5 were more likely to have *CFTR*-related symptoms than those with CFvar in combination with TG11T5. Moreover, the phenotypes in TG[11-13]T5 individuals were much less severe and often affected a single organ compared to individuals with CFvar/CFvar variants, consistent with previous reports. Interestingly, we found that pancreatitis and chronic sinopulmonary disease were most commonly reported among CFvar/TG11T5 individuals while chronic sinopulmonary disease, intermediate sweat chloride levels, and a suspected diagnosis of CF were most commonly reported among CFvar/TG12T5 individuals. Unfortunately, the total number of individuals with a CFvar/TG13T5 genotype who reported a phenotype was very low ($n = 4$), which limits the spectrum of findings for these individuals. Nevertheless, these results are also consistent with previous reports. Among individuals with clinical information, the observed phenotypic spectrum for those with TG[11-13]T5 alleles illustrates the wide variability and milder severity associated with these alleles compared to CFvar/CFvar individuals.

While phenotypic analyses in individuals with CFvar/TG[11-13]T5 genotypes have been reported in the published literature, there has been limited information available on individuals who are homozygous or compound heterozygous for the TG[11-13]T5 variants. One case report of a male individual homozygous for TG12T5 demonstrated that this genotype may be associated with CFTR-RD (Montagnani et al., 2013), specifically exhibiting recurrent episodes of pancreatitis and elevated sweat chloride levels, but the likelihood with which TG12T5 homozygous individuals suffer from disease was not clear from this study. With the large population of sequenced individuals in our study, we were able to conclude that TG11T5 and TG12T5 individuals are quite unlikely to experience symptoms. Unfortunately, the small number of individuals in the population with a TG13T5 variant limited the number of homozygous individuals available for assessment, although these individuals showed a trend in enrichment similar to patients with CFvar/CFvar genotypes. Interestingly, although only 2 of 287 TG12T5/TG11T5 individuals reported *CFTR*-related symptoms at the time of testing (high suspicion group), one individual who was not initially included in the high suspicion cohort reported chronic sinopulmonary disease upon follow-up with our genetic counseling team. Therefore, the number of patients with *CFTR*-related symptoms, and corresponding ORs, particularly with milder, single-organ symptoms, is expected to be higher than observed in this study.

Additional studies utilizing more complete medical histories among individuals with CFvar and TG[11-13]T5 variants in larger sample sizes will provide additional insights into the variable expressivity and severity of symptoms associated with CF and CFTR-RDs. Further, population-based studies will be necessary to obtain a more precise understanding of the penetrance and likely course of disease for T/TG carriers.

This study is limited by the amount of clinical information available for individuals who have received *CFTR* testing. For example, only 59% of ordering clinicians provided any clinical information about the patient's disease when a CFvar/CFvar genotype was observed. Phenotypic information for individuals with a CFvar and a TG[11-13]T5 allele was reported less frequently (34%). It is unclear if this means individuals with CFvar/TG[11-13]T5 genotypes are more likely asymptomatic, or more likely to have milder symptoms that were overlooked by the referring clinician. Importantly, this limitation highlights both the challenge with interpreting variants with partial functional activity and the importance of having complete information on a patient's phenotype when trying to understand the expressivity and penetrance of mild variants. These limitations should be considered in the context of counseling individuals undergoing either diagnostic testing or carrier screening. Future studies will help to clarify the findings here.

The results from this study complement the recent technical standards from the ACMG (Deignan et al., 2020) by describing a very large cohort of individuals with *CFTR* sequencing results and describing the spectrum of clinical symptoms observed in individuals with various TG[11-13]T5 genotypes. These data support improved counseling of individuals who have one or more T/TG alleles and highlight the need to include polyT tract and TG repeat analysis in routine *CFTR* testing, regardless of p.Arg117His status.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

WEB RESOURCES

MedCalc's Online OR Calculator—https://www.medcalc.org/calc/odds_ratio.php

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request. Additionally, all variants are shared with ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/).

ORCID

Sara L. Bristow http://orcid.org/0000-0002-1173-4718

REFERENCES

Castellani, C., & Assael, B. M. (2017). Cystic fibrosis: A clinical view. *Cellular and Molecular Life Sciences*, *74*(1), 129–140.

Chu, C. S., Trapnell, B. C., Curristin, S., Cutting, G. R., & Crystal, R. G. (1993). Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. *Nature Genetics*, *3*(2), 151–156.

Chu, C. S., Trapnell, B. C., Curristin, S. M., Cutting, G. R., & Crystal, R. G. (1992). Extensive posttranscriptional deletion of the coding sequences for part of nucleotide-binding fold 1 in respiratory epithelial mRNA transcripts of the cystic fibrosis transmembrane conductance regulator gene is not associated with the clinical manifestations of cystic fibrosis. *The Journal of Clinical Investigation*, *90*(3), 785–790.

Chu, C. S., Trapnell, B. C., Murtagh, J. J., Jr., Moss, J., Dalemans, W., Jallat, S., Mercenier, A., Pavirani, A., Lecocq, J. P., & Cutting, G. R. (1991). Variable deletion of exon 9 coding sequences in cystic fibrosis transmembrane conductance regulator gene mRNA transcripts in normal bronchial epithelium. *The EMBO Journal*, *10*(6), 1355–1363.

Cuppens, H., Lin, W., Jaspers, M., Costes, B., Teng, H., Vankeerberghen, A., Jorissen, M., Droogmans, G., Reynaert, I., Goossens, M., Nilius, B., & Cassiman, J. J. (1998). Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes. The polymorphic (Tg)m locus explains the partial penetrance of the T5 polymorphism as a disease mutation. *The Journal of Clinical Investigation*, *101*(2), 487–496.

Cutting, G. R. (2010). Modifier genes in Mendelian disorders: The example of cystic fibrosis. *Annals of the New York Academy of Sciences*, *1214*, 57–69.

Deignan, J. L., Astbury, C., Cutting, G. R., Del Gaudio, D., Gregg, A. R., Grody, W. W., & Richards, S., ACMG Laboratory Quality Assurance Committee. (2020). CFTR variant testing: A technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, *22*, 1288–1295. https://doi.org/10.1038/s41436-020-0822-5

Delaney, S. J., Rich, D. P., Thomson, S. A., Hargrave, M. R., Lovelock, P. K., Welsh, M. J., & Wainwright, B. J. (1993). Cystic fibrosis transmembrane conductance regulator splice variants are not conserved and fail to produce chloride channels. *Nature Genetics*, *4*(4), 426–431.

de Nooijer, R. A., Nobel, J. M., Arets, H. G., Bot, A. G., van Berkhout, F. T., de Rijke, Y. B., de Jonge, H. R., & Bronsveld, I. (2011). Assessment of CFTR function in homozygous R117H-7T subjects. *Journal of Cystic Fibrosis*, *10*(5), 326–332.

Farrell, P. M., White, T. B., Ren, C. L., Hempstead, S. E., Accurso, F., Derichs, N., Howenstine, M., McColley, S. A., Rock, M., Rosenfeld, M., Sermet-Gaudelus, I., Southern, K. W., Marshall, B. C., & Sosnay, P. R. (2017). Diagnosis of cystic fibrosis: Consensus guidelines from the Cystic Fibrosis Foundation. *The Journal of Pediatrics*, *181S*, S4–S15.e1.

Groman, J. D., Hefferon, T. W., Casals, T., Bassas, L., Estivill, X., Des Georges, M., Guittard, C., Koudova, M., Fallin, M. D., Nemeth, K., Fekete, G., Kadasi, L., Friedman, K., Schwarz, M., Bombieri, C., Pignatti, P. F., Kanavakis, E., Tzetis, M., Schwartz, M., … Cutting, G. R. (2004). Variation in a repeat sequence determines whether a common variant of the cystic fibrosis transmembrane conductance regulator gene is pathogenic or benign. *American Journal of Human Genetics*, *74*(1), 176–179.

Hannah, W. B., Truty, R., Gonzales, V., Kithcart, G. P., Ouyang, K., Zeman, M. K., Li, C., Drumm, M., Nykamp, K., & Gaston, B. M. (2019). Frequency of cystic fibrosis transmembrane conductance regulator variants in individuals evaluated for primary ciliary dyskinesia. *The Journal of Pediatrics*, *215*, 172–177.

Hefferon, T. W., Groman, J. D., Yurk, C. E., & Cutting, G. R. (2004). A variable dinucleotide repeat in the CFTR gene contributes to phenotype diversity by forming RNA secondary structures that alter splicing. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(10), 3504–3509.

Kiesewetter, S., Macek M, J.r, Jr., Davis, C., Curristin, S. M., Chu, C. S., Graham, C., Shrimpton, A. E., Cashman, S. M., Tsui, L. C., & Mickle, J. (1993). A mutation in CFTR produces different phenotypes depending on chromosomal background. *Nature Genetics*, *5*(3), 274–278.

Knowles, M. R., & Drumm, M. (2012). The influence of genetics on cystic fibrosis phenotypes. *Cold Spring Harbor Perspectives in Medicine*, *2*(12), a009548.

Lincoln, S. E., Kobayashi, Y., Anderson, M. J., Yang, S., Desmond, A. J., Mills, M. A., Nilsen, G. B., Jacobs, K. B., Monzon, F. A., Kurian, A. W., Ford, J. M., & Ellisen, L. W. (2015). A systematic comparison of traditional and multigene panel testing for hereditary breast and ovarian cancer genes in more than 1000 patients. *The Journal of Molecular Diagnostics*, *17*(5), 533–544.

Mak, V., Jarvi, K. A., Zielenski, J., Durie, P., & Tsui, L. C. (1997). Higher proportion of intact exon 9 CFTR mRNA in nasal epithelium compared with vas deferens. *Human Molecular Genetics*, *6*(12), 2099–2107.

Montagnani, M., Cazzato, S., Mutignani, M., Cevenini, M., Guidetti, E., Zvi, I. B., Aldini, R., Saraceni, G., Cavoli, C., Garagnani, P., Ferrari, S., & Mantovani, V. (2013). A patient with pancreas divisum, recurrent acute pancreatitis, and homozygosity for the cystic fibrosis transmembrane regulator-associated protein 5T allele. *Clinical Gastroenterology and Hepatology*, *11*(5), 579–581.

Niksic, M., Romano, M., Buratti, E., Pagani, F., & Baralle, F. E. (1999). Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. *Human Molecular Genetics*, *8*(13), 2339–2349.

Nykamp, K., Anderson, M., Powers, M., Garcia, J., Herrera, B., Ho, Y.-Y., Kobayashi, Y., Patil, N., Thusberg, J., Westbrook, M., Invitae Clinical Genomics, G., & Topper, S. (2017). Sherloc: A comprehensive refinement of the ACMG-AMP variant classification criteria. *Genetics in Medicine*, *19*(10), 1105–1117.

Ramalho, A. S., Beck, S., Meyer, M., Penque, D., Cutting, G. R., & Amaral, M. D. (2002). Five percent of normal cystic fibrosis transmembrane conductance regulator mRNA ameliorates the severity of pulmonary disease in cystic fibrosis. *American Journal of Respiratory Cell and Molecular Biology*, *27*(5), 619–627.

Rave-Harel, N., Kerem, E., Nissim-Rafinia, M., Madjar, I., Goshen, R., Augarten, A., Rahat, A., Hurwitz, A., Darvasi, A., & Kerem, B. (1997). The molecular basis of partial penetrance of splicing mutations in cystic fibrosis. *American Journal of Human Genetics*, *60*(1), 87–94.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., & ACMG Laboratory Quality Assurance, C., ACMG Laboratory Quality Assurance Committee. (2015). 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. *Genetics in Medicine*, *17*(5), 405–424.

Salinas, D. B., Azen, C., Young, S., Keens, T. G., Kharrazi, M., & Parad, R. B. (2016). Phenotypes of California CF newborn screen-positive children with CFTR 5T allele by TG repeat length. *Genetic Testing and Molecular Biomarkers*, *20*(9), 496–503.

Sosnay, P. R., Raraigh, K. S., & Gibson, R. L. (2016). Molecular genetics of cystic fibrosis transmembrane conductance regulator: Genotype and phenotype. *Pediatric Clinics of North America*, *63*(4), 585–598.

Strong, T. V., Wilkinson, D. J., Mansoura, M. K., Devor, D. C., Henze, K., Yang, Y., Wilson, J. M., Cohn, J. A., Dawson, D. C., & Frizzell, R. A. (1993). Expression of an abundant alternatively spliced form of the cystic fibrosis transmembrane conductance regulator (CFTR) gene is not associated with a cAMP-activated chloride conductance. *Human Molecular Genetics*, *2*(3), 225–230.

Sun, W., Anderson, B., Redman, J., Milunsky, A., Buller, A., McGinniss, M. J., Quan, F., Anguiano, A., Huang, S., Hantash, F., & Strom, C. (2006). CFTR 5T variant has a low penetrance in females that is partially attributable to its haplotype. *Genetics in Medicine*, *8*(6), 339–345.

Teng, H., Jorissen, M., Van Poppel, H., Legius, E., Cassiman, J. J., & Cuppens, H. (1997). Increased proportion of exon 9 alternatively spliced CFTR transcripts in vas deferens compared with nasal epithelial cells. *Human Molecular Genetics*, *6*(1), 85–90.

Thauvin-Robinet, C., Munck, A., Huet, F., Génin, E., Bellis, G., Gautier, E., Audrézet, M. P., Férec, C., Lalau, G., Georges, M. D., Claustres, M., Bienvenu, T., Gérard, B., Boisseau, P., Cabet-Bey, F., Feldmann, D., Clavel, C., Bieth, E., Iron, A., … Girodon, E. (2009). The very low penetrance of cystic fibrosis for the R117H mutation: A reappraisal for genetic counselling and newborn screening. *Journal of Medical Genetics*, *46*(11), 752–758.

Truty, R., Paul, J., Kennemer, M., Lincoln, S. E., Olivares, E., Nussbaum, R. L., & Aradhya, S. (2019). Prevalence and properties of intragenic copy-number variation in Mendelian disease genes. *Genetics in Medicine*, *21*(1), 114–123.

Waller, M. D., & Simmonds, N. J. (2016). Phenotypic variability of R117H-CFTR expression within monozygotic twins. *Paediatric Respiratory Reviews*, *20*(Suppl), 21–23.

Wang, Y., Wrennall, J. A., Cai, Z., Li, H., & Sheppard, D. N. (2014). Understanding how cystic fibrosis mutations disrupt CFTR function: From single molecules to animal models. *The International Journal of Biochemistry & Cell Biology*, *52*, 47–57.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Nykamp, K., Truty, R., Riethmaier, D., Wilkinson, J., Bristow, S. L., Aguilar, S., Neitzel, D., Faulkner, N., & Aradhya, S. (2021). Elucidating clinical phenotypic variability associated with the polyT tract and TG repeats in CFTR. *Human Mutation*, 42, 1165–1172. https://doi.org/10.1002/humu.24250