Am. J. Hum. Genet. 74:176–179, 2004

Report

Variation in a Repeat Sequence Determines Whether a Common Variant of the Cystic Fibrosis Transmembrane Conductance Regulator Gene Is Pathogenic or Benign

Joshua D. Groman,[1] Timothy W. Hefferon,[1] Teresa Casals,[3] Lluís Bassas,[4] Xavier Estivill,[5] Marie Des Georges,[6] Caroline Guittard,[6] Monika Koudova,[7] M. Daniele Fallin,[2] Krisztina Nemeth,[8] Gyorgy Fekete,[8] Ludovit Kadasi,[9] Ken Friedman,[10] Martin Schwarz,[11] Cristina Bombieri,[12] Pier Franco Pignatti,[12] Emmanuel Kanavakis,[13] Maria Tzetis,[13] Marianne Schwartz,[14] Giuseppe Novelli,[15] Maria Rosaria D'Apice,[15] Agnieszka Sobczynska-Tomaszewska,[16] Jerzy Bal,[16] Manfred Stuhrmann,[17] Milan Macek, Jr.,[7] Mireille Claustres,[5] and Garry R. Cutting[1]

[1]Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, and [2]Bloomberg School of Public Health, Johns Hopkins University, Baltimore; [3]Medical and Molecular Genetics Center—Institut Recerca Oncologica, Hospital Duran i Reynals, Barcelona; [4]Institut d'Urologia, Nefrologia i Andrologia, Fundació Puigvert, Barcelona; [5]Genes and Diseases Program, Centre de Regulacio Genomica, Barcelona; [6]Laboratoire de Génétique Moléculaire, CHU de Montpellier, Institut Universitaire de Recherche Clinique (IURC), Montpellier; [7]Institute of Biology and Medical Genetics, Cystic Fibrosis Centre, Second School of Medicine and University Hospital Motol, Charles University, Prague; [8]Department of Pediatrics, Semmelweis University, Budapest; [9]Institute of Molecular Physiology and Genetics, Bratislava, Slovakia; [10]Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill; [11]Molecular Genetics Laboratory, Royal Manchester Children's Hospital, Manchester, United Kingdom; [12]Section of Biology and Genetics, Department of Mother and Child, Biology and Genetics, University of Verona, Verona; [13]Department of Medical Genetics, Athens University, St. Sophia's Children's Hospital, Athens; [14]Department of Clinical Genetics, Rigshospitalet, Copenhagen; [15]Dipartimento di Biopatologia e Diagnostica per Immagini Università di Roma "Tor Vergata," Rome; [16]Medical Genetics Department, Institute of Mother and Child, Warsaw; and [17]Institut für Humangenetik Medizinische Hochschule, Hannover

An abbreviated tract of five thymidines (5T) in intron 8 of the *cystic fibrosis transmembrane conductance regulator* (*CFTR*) gene is found in ~10% of individuals in the general population. When found in *trans* with a severe *CFTR* mutation, 5T can result in male infertility, nonclassic cystic fibrosis, or a normal phenotype. To test whether the number of TG repeats adjacent to 5T influences disease penetrance, we determined TG repeat number in 98 patients with male infertility due to congenital absence of the vas deferens, 9 patients with nonclassic CF, and 27 unaffected individuals (fertile men). Each of the individuals in this study had a severe *CFTR* mutation on one *CFTR* gene and 5T on the other. Of the unaffected individuals, 78% (21 of 27) had 5T adjacent to 11 TG repeats, compared with 9% (10 of 107) of affected individuals. Conversely, 91% (97 of 107) of affected individuals had 12 or 13 TG repeats, versus only 22% (6 of 27) of unaffected individuals ($P < .00001$). Those individuals with 5T adjacent to either 12 or 13 TG repeats were substantially more likely to exhibit an abnormal phenotype than those with 5T adjacent to 11 TG repeats (odds ratio 34.0, 95% CI 11.1–103.7, $P < .00001$). Thus, determination of TG repeat number will allow for more accurate prediction of benign versus pathogenic 5T alleles.

Genes that are responsible for inherited disorders can have mutations that cause disease phenotypes in some but not all individuals. These incompletely penetrant mutations become problematic if they are incorporated into clinical testing prior to an understanding of the factors affecting penetrance. One variant that presents a problem in the interpretation of genetic testing of the *cystic fibrosis transmembrane conductance regulator* (*CFTR*) gene is an abbreviated tract of five thymidines (5T) in intron 8 (Noone et al. 2000; Strom et al. 2002). Some individuals bearing 5T in *trans* with a severe cystic fibrosis–causing mutation may have nonclassic cystic fibrosis (CF [MIM 219700]); others may have male in-

Received September 22, 2003; accepted for publication October 23, 2003; electronically published December 18, 2003.

Address for correspondence and reprints: Dr. Garry R. Cutting, Johns Hopkins University School of Medicine, CMSC 9-125, 600 North Wolfe Street, Baltimore, MD 21287; E-mail: gcutting@jhmi.edu

0002-9297/2004/7401-0018$15.00

fertility due to congenital bilateral absence of the vas deferens (CBAVD [MIM 277180]); and ~40% may be healthy and fertile as a consequence of incomplete penetrance (Chillón et al. 1995; Zielenski et al. 1995). Thus, 5T frequently creates an interpretative challenge for clinicians due to the commonness of the variant (found in 1 in 10 individuals) and the large number of *CFTR* genetic tests performed each year.

The 5T variant is a stretch of five contiguous thymidines at the 3′ end of intron 8 that exacerbates skipping of exon 9, thereby resulting in reduced levels of functional CFTR protein (Chu et al. 1993). Prior studies have demonstrated that the number of TG repeats immediately adjacent to 5T correlates with exon 9 skipping (Cuppens et al. 1998; Niksic et al. 1999). Cuppens et al. (1998) have also suggested that increased exon 9 skipping from transcripts bearing more TG repeats increases the penetrance of 5T. However, an adequate calculation of the effect of TG repeat number on 5T penetrance has been hampered by small sample sizes. The aim of this collaborative study was to evaluate whether TG repeat number significantly aids in prediction of benign versus pathogenic 5T alleles, thereby warranting its use in clinical situations.

Identifying information was removed from all DNA samples prior to submission to this study. The relative frequencies of 5T and TG-T combinations were estimated by screening 1,269 DNA samples from the general populations of five countries (United States, Czech Republic, Poland, Italy, and Germany). To assess the effect of TG repeat variation on penetrance of 5T, we analyzed DNA samples from 27 unaffected fertile men, 98 patients with CBAVD, and 9 patients with nonclassic CF. Each of the 134 individuals in this study had a severe CF mutation on one *CFTR* gene and 5T on the other. Either the 5T variant was confirmed—by studying allele transmission in the family—to be in *trans* with a severe CF mutation, or it was assumed to be in *trans* if the severe CF mutation was a common mutation previously shown not to be in linkage disequilibrium with 5T. To identify unaffected individuals with this genotype, we screened for 5T in 1,685 men who fathered at least one child with classic CF (fertile, obligate carriers). Since >98% of men with CF are infertile, it is unlikely that fathers of patients with CF have an unrecognized form of the disease, making them an appropriate control population for this study. Each of the 27 fathers with 5T had one of the following mutations: ΔF508 (18), G542X (2), 1812-1G→A (2), ΔI507 (1), 936delTA (1), N1303K (1), 3600+2insT (1), or 1717-1 G→A (1). 5T was inferred to be in *trans* with ΔF508 and ΔI507, since these mutations have never been shown to be in linkage disequilibrium with 5T. The 5T variant was confirmed by use of pedigree analysis to be in *trans* with the other mutations. Samples from fathers of patients with CF were referred from CF care centers from the following countries (number of samples screened and number with the CF mutation/5T genotype in parentheses): Czech Republic (202, 5), Denmark (172, 2), France (109, 2), Germany (114, 1) , Greece (210, 1), Hungary (30, 2), Italy (230, 2), Poland (100, 5), Slovakia (46, 0), Spain (384, 5), and the United States (88, 2).

Patients with CBAVD were diagnosed by palpation, analysis of sperm, and/or rectal ultrasound. Each of the 98 patients with CBAVD had 5T with one of the following mutations: ΔF508 (78), G542X (6), N1303K (3), 711+1G→T (2), R1066C (2), R1162X (2), R764X (1), Y563X (1), H609R (1), L206W (1), or R334W (1). 5T was confirmed by use of pedigree analysis to be in *trans* with R764X, Y563X, and H609X and was inferred as in *trans* for the remaining mutations, since these mutations have never been found to be in linkage disequilibrium with 5T in previous population sampling. Patients with CBAVD were referred from the following countries (number with the CF mutation/5T genotype): Denmark (6), France (35), Germany (8), Greece (1), Italy (4), Poland (6), Spain (29), the United Kingdom (6), and the United States (3).

Nine patients with nonclassic CF were referred from CF care centers in the United States and were confirmed to have a 5T in *trans* with one of the following *CFTR* mutations: ΔF508 (6), 2814insA (1), G463D (1), or F693L (1). For each patient, we scanned the entire coding region of the *CFTR* gene to rule out the presence of rare mutations in addition to 5T (DNA sequencing $n = 6$, denaturating gradient gel electrophoresis [DGGE] $n = 3$). Each patient met current diagnostic criteria for CF by demonstrating both CFTR dysfunction, and at least one clinical feature of CF (Rosenstein and Cutting 1998). Specifically, six of the nine patients had respiratory disease consistent with CF, as well as CFTR dysfunction, as indicated by elevated mean sweat Cl^- concentration (58–70 mmol/L). Of the remaining three patients, each had evidence of CFTR dysfunction, as indicated by borderline sweat Cl^- concentration (43–49 mmol/L), and abnormally reduced low Cl^- buffer/isoproterenol response, as demonstrated by the nasal potential difference (NPD) test. Of these three patients, one had mild respiratory disease, one had elevated serum trypsinogen in the newborn period, and one had elevated quantitative fecal fat levels consistent with pancreatic insufficiency.

To detect samples positive for 5T, PCR products were fixed on Hybond N+ nylon membranes (Amersham Biosciences) and were probed with the ^{32}P end-labeled probe 5′-TGTGTGTGTTTTTAACAG. Samples that were positive for 5T were sequenced bidirectionally and analyzed by automated capillary electrophoresis (Applied Biosystems) to determine TG repeat number and phase of TG-T repeat combinations. Samples from

Table 1

Frequency of the 5T Allele and TG-5T Combinations in the General Population

Country of Origin	DNA Samples Screened	5T Allele Frequency	5T Combinations (Proportion of 5T Alleles)		
			TG11-5T	TG12-5T	TG13-5T
United States	517	0.048	36 (0.72)	13 (0.26)	1 (0.02)
Czech Republic	251	0.030	14 (0.93)	1 (0.07)	0 (0.00)
Poland	132	0.053	13 (0.93)	1 (0.07)	0 (0.00)
Italy	173	0.038	9 (0.70)	4 (0.30)	0 (0.00)
Germany	196	0.041	11 (0.69)	4 (0.25)	1 (0.06)
Total	1269	0.043	83 (0.77)	23 (0.21)	2 (0.02)

France, Italy, Greece, and Spain were screened using DGGE or restriction enzyme digestion.

Allele frequencies for 5T were similar among the five general population groups and were consistent with previous reports (table 1). 5T was found in *cis* with three different TG repeats (TG11-5T, TG12-5T, TG13-5T). Although variation in the frequencies of these combinations was observed among populations, TG11-5T was by far the most common in all five populations, followed by TG12-5T and TG13-5T (table 1).

TG12-5T and TG13-5T were more frequent in the affected patients than controls across each of the populations sampled (table 2), and the frequencies of the different TG-T combinations differed significantly between the unaffected and affected groups (table 3). The most frequent combination, TG11-5T, was found in 78% of unaffected individuals, versus 9% of affected individuals (table 3), suggesting that TG11-5T is generally benign. TG12-5T was the most common disease-associated combination, and was found in 76% of affected individuals versus 22% of unaffected individuals. It is interesting to note that TG13-5T was found exclusively in affected individuals (table 3). These findings demonstrate that, when 5T is found in *trans* with a severe CF mutation, the odds of pathogenicity are 28 and 34 times greater for TG12-5T and TG13-5T, respectively, than for TG11-5T (table 3).

Previous studies have suggested that another common variant in *CFTR*, a methionine or valine at codon 470, also influences the penetrance of 5T (Cuppens et al. 1998; de Meeus et al. 1998). To test whether this variant affects the penetrance of 5T, we established phase between TG-T combinations and this variant in 22 unaffected and 92 affected individuals. As reported elsewhere, TG11-5T was always found with M470, and, in all but five cases, TG12-5T was found with V470 (data not shown). The highly penetrant TG13-5T was found to occur only with M470 in all of the cases in which phase could be established ($n = 15$). The correlation between increased penetrance and TG repeat number, lack of correlation with variation at codon 470, and experimental evidence of increased exon 9 skipping with more TG repeats indicates that TG repeat number, rather than M470V status, is the major determinant of penetrance for 5T.

Our analysis demonstrates that knowledge of TG repeat number in individuals with 5T is of diagnostic value. When TG repeat number is determined, the sen-

Table 2

Distribution of the TG-5T Combinations in Unaffected and Affected Individuals, by Country of Origin

Country of Origin	No. of Occurrences in					
	Unaffected Individuals			Affected Individuals		
	TG11-5T	TG12-5T	TG13-5T	TG11-5T	TG12-5T	TG13-5T
France	2	…	…	2	28	5
Spain	4	1	…	4	25	…
United States	1	1	…	…	8	4
Poland	5	…	…	1	4	1
Germany	…	1	…	1	4	3
Denmark	2	…	…	2	2	2
Italy	2	…	…	…	4	…
United Kingdom	…	…	…	…	5	1
Czech Republic	3	2	…	…	…	…
Greece	1	…	…	…	1	…
Hungary	1	1	…	…	…	…

Table 3

Frequency of the TG-5T Combinations in Unaffected and Affected Individuals with a CF Mutation/5T Genotype

	No. of Individuals			
Status	All Samples	TG11-5T	TG12-5T[a]	TG13-5T[b]
Unaffected:				
Fertile men	27	21	6	0
Affected:				
CBAVD	98	10	76	12
Nonclassic CF	9	0	5	4
Total affected	107	10	81	16

Note.—Odds ratios were calculated by comparing the frequency of the TG12-T5 or TG13-T5 combination to the frequency of the putatively benign TG11-T5 combination in affected versus unaffected individuals.

[a] Odds ratio 28.4, 95% CI 9.3–86.9, $P < .00001$.

[b] Odds ratio 34.0, 95% CI 4.0–289.8, $P < .0001$.

sitivity for identifying affected individuals is 91%, whereas the specificity for identifying those who are unaffected is 78%. After adjusting for differences in sample sizes of the affected and unaffected groups, we used the proportion of each TG-T combination in affected individuals to estimate that the disease risk for males with TG11-5T is 0.10, TG12-5T is 0.78, and TG13-5T is 1.0. Since CBAVD accounts for a large component of the disease risk in this study, the risk of pathogenicity will likely be lower for females. Nonetheless, a female carrying TG12-5T or TG13-5T in *trans* with a CF mutation will be at higher risk of developing nonclassic CF than if she carried the TG11-5T allele.

The association between TG repeat number and disease penetrance demonstrated in this study is consistent with molecular studies correlating TG repeat number with aberrant *CFTR* splicing (Cuppens et al. 1998; Niksic et al. 1999; Hefferon et al., in press) and validates clinical use of TG repeat testing in counseling individuals with 5T. This information will have important clinical relevance, since ~25 million individuals in the United States are predicted to carry at least one 5T allele. Although TG repeat number proves to be a reliable predictor of penetrance for 5T, further studies are required to elucidate the factors that determine whether an individual will develop multisymptomatic (nonclassic CF) or monosymptomatic (CBAVD) disease.

Acknowledgments

We wish to thank M. Boyle, S. Hsu, and T. Holtzman, for review of the manuscript, and L. Ramos, J. Gimenez, C. Yurk, X. Wang, D. Arking, C. Castellani, M. De Gironcoli, F. Belpinati, M. Macek, Sr., A. Krebsova, and V. Vavrova, for technical and logistical assistance. This study was supported by grants from the Italian MIUR and Ministry of Health "Genetic Tests" project DL229/99 9 and MURST-Italy (to P.F.P. and C.B.); the Italian Ministry of Health and Ministry of Education (to G.N.); IGA MZCR-00000064203/6463–3, MSMT-111300003/LN00A079, EC-CFChip, and CRMGEN (to M.M.Jr.); F. La Marato TV3 (980410) and ISCiii (PI020099) and ICS (to T.C., L.C., and X.E.); and the Cystic Fibrosis Foundation (CUTTIN98AO) and the National Institutes of Health (DK44003 and HL68927) (to G.R.C.).

Electronic-Database Information

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for CF and CBAVD)

References

Chillón M, Casals T, Mercier B, Bassas L, Lissens W, Silber S, Romey MC, et al (1995) Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. N Engl J Med 332:1475–1480

Chu CS, Trapnell BC, Curristin S, Cutting GR, Crystal RG (1993) Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. Nat Genet 3:151–156

Cuppens H, Lin W, Jaspers M, Costes B, Teng H, Vankeerberghen A, Jorissen M, Droogmans G, Reynaert I, Goossens M, Nilius B, Cassiman JJ (1998) Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes: the polymorphic (TG)m locus explains the partial penetrance of the 5T polymorphism as a disease mutation. J Clin Invest 101:487–496

de Meeus A, Guittard C, Desgeorges M, Carles S, Demaille J, Claustres M (1998) Linkage disequilibrium between the M470V variant and the IVS8 polyT alleles of the CFTR gene in CBAVD. J Med Genet 35:594–596

Hefferon TW, Groman JD, Yurk CE, Cutting GR. A variable dinucleotide repeat in the CFTR gene contributes to phenotype diversity by forming RNA secondary structures that alter splicing. Proc Natl Acad Sci USA (in press)

Niksic M, Romano M, Buratti E, Pagani F, Baralle FE (1999) Functional analysis of *cis*-acting elements regulating the alternative splicing of human CFTR exon 9. Hum Mol Genet 8:2339–2349

Noone PG, Pue CA, Zhou Z, Friedman KJ, Wakeling EL, Ganeshananthan M, Simon RH, Silverman LM, Knowles MR (2000) Lung disease associated with the IVS8 5T allele of the CFTR gene. Am J Respir Crit Care Med 162:1919–1924

Rosenstein BJ and Cutting GR (1998) The diagnosis of cystic fibrosis: a consensus statement. Cystic Fibrosis Foundation Consensus Panel. J Pediatr 132:589–95

Strom CM, Huang D, Buller A, Redman J, Crossley B, Anderson B, Entwistle T, Sun W (2002) Cystic fibrosis screening using the College panel: platform comparison and lessons learned from the first 20,000 samples. Genet Med 4:289–296

Zielenski J, Patrizio P, Corey M, Handelin B, Markiewicz D, Asch R, Tsui LC (1995) CFTR gene variant for patients with congenital absence of vas deferens. Am J Hum Genet 57: 958–960