CORRELATION BETWEEN GENOTYPE AND PHENOTYPE IN PATIENTS WITH CYSTIC FIBROSIS

The Cystic Fibrosis Genotype–Phenotype Consortium*

Abstract  Background. Cystic fibrosis is the most common lethal autosomal recessive disorder among whites. Seventy-two percent of patients with this disease are homozygotes or compound heterozygotes for eight mutations of the cystic fibrosis transmembrane conductance regulator gene on chromosome 7: ∆F508, G542X, R553X, W1282X, N1303K, 621+1G→T, 1717−1G→A, and R117H. We studied the relation between genotype and phenotype in patients from 14 countries.

Methods. Each of 399 patients who were compound heterozygotes for ∆F508 and one other mutation was matched with the ∆F508 homozygote of the same sex who was the closest in age from the same center. A paired analysis was performed of the following outcome variables: age at diagnosis, sweat chloride concentration, growth percentiles, pulmonary-function values, chest-film score, pseudomonas colonization, nasal polyps, pancreatic sufficiency, pancreatitis, diabetes mellitus, meconium ileus, distal intestinal obstruction syndrome, rectal prolapse, cirrhosis, and gallbladder disease.

Cystic fibrosis is the most common lethal autosomal recessive childhood disorder in the white population, occurring in approximately 1 in 2500 live births. Patients with cystic fibrosis have abnormal chloride conduction across the apical membrane of epithelial cells, causing inspissated secretions in the airways, pancreas, intestines, and vas deferens. The disease is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene on chromosome 7, which encodes a protein of 1480 amino acid residues that functions as a chloride channel regulated by cyclic AMP.

The most frequent mutation, present on about 67 percent of cystic fibrosis chromosomes worldwide, results in the deletion of a phenylalanine residue at codon 508 (∆F508). The clinical manifestations in patients homozygous for this mutation have been extensively studied. They generally have pancreatic insufficiency of early onset with markedly elevated sweat chloride concentrations, but the pulmonary manifestations are widely variable.

Case reports of patients with various mutations have been reviewed, and several studies of patients with defined genotypes have recently been published. Homozygotes for W1282X, the most common cystic fibrosis mutation in the Ashkenazi Jewish population, were compared with W1282X/∆F508 compound heterozygotes. The groups were similar to each other and to ∆F508 homozygotes described elsewhere. A report of patients carrying the N1303K mutation revealed that this mutation is associated with pancreatic insufficiency of early onset and widely variable pulmonary disease. Compound heterozygotes for G551D and ∆F508 were indistinguishable from matched ∆F508 homozygotes except for a decreased risk of meconium ileus. Studies of patients carrying the R553X mutation found inconsistent results with regard to sweat chloride concentrations and growth.

This report describes the clinical features of compound heterozygotes for ∆F508 with one of seven of the next most frequent cystic fibrosis mutations: the nonsense mutations G542X, R553X, and W1282X; the missense mutations N1303K and R117H; and the splice-site mutations 621+1G→T and 1717−1G→A, as compared with age- and sex-matched ∆F508 homozygotes treated at the same center. In addition, homozygotes and compound heterozygotes for nonsense mutations known to cause severe reduction of CFTR mRNA (G542X, R553X, and W1282X), and W1282X were grouped and compared with matched ∆F508 homozygotes.

Results. The compound heterozygotes having the genotype R117H/∆F508 clearly differed from the age- and sex-matched ∆F508 homozygotes; they more often had pancreatic sufficiency (67 percent vs. 4 percent, P < 0.001), were older when the diagnosis was first made (mean ± SD) age, 10.2 ± 10.5 vs. 2.5 ± 4.3 years; P = 0.002), and had lower sweat chloride concentrations (80 ± 18 vs. 108 ± 14 mmol per liter; P < 0.001). There were no statistically significant differences between ∆F508 homozygotes and other compound heterozygotes with regard to any variable tested.

Conclusions. Prenatal and prognostic counseling for patients with the R117H/∆F508 genotype should include the likelihood that they will have long-term pancreatic sufficiency. Patients with the other genotypes should expect the early onset of pancreatic insufficiency. For none of the genotypes studied can predictions be made about the occurrence of common complications or the severity or course of pulmonary disease. (N Engl J Med 1993;329:1308-13.)

*The members of the Cystic Fibrosis Genotype-Phenotype Consortium are listed in the Appendix. The paper was written by Ada Hamosh, M.D., M.P.H., and Mary Corey, Ph.D., who assume responsibility for its overall content and integrity. Address reprint requests to Dr. Hamosh at CMSC 1004, Johns Hopkins Hospital, 600 N. Wolfe St., Baltimore, MD 21287-3914.

Supported in part by grants (DK02151, DK44003, and DK41890) from the National Institutes of Health, the Canadian Cystic Fibrosis Foundation; the Deutsche Gesellschaft zur Bekämpfung der Mukoviszidose and the Bundesministerium für Forschung und Technologie, Germany; the grant (Z-215) from the Czech Ministry of Health; and the Biotechnology and Bioinstrumentation, Genetic Engineering, and Progetto Finalizzato Prevenzione e Controllo dei Fattori di Malattia projects of the Italian Consiglio Nazionale delle Ricerche.
Genotyping was offered to all the patients in each center. Twenty-two centers screened 100 percent of their patients for the mutations studied, and two screened approximately 75 percent; eight centers did not specify the percentage of patients genotyped.

Each patient was matched with the ΔF508 homozygote closest in age and of the same sex who was currently under treatment at the same center. Patients were matched within centers to control for differences in medical management among centers.

**Demographic Data**

The demographic data collected included the patient's date of birth, the date of death (if applicable), the age at the most recent clinic visit, race, sex, and ethnic origin (if known).

**Clinical Variables**

General outcome variables included age at diagnosis; sweat chloride concentration (in millimoles per liter) measured by quantitative pilocarpine iontophoresis; current weight and height percentiles; the Shwachman–Kulczycki clinical score, which includes a subjective assessment of activity, the physical examination, growth, and nutrition; and the chest film.

Respiratory status was assessed by tests of forced vital capacity (FVC), expressed as a percentage of the normal predicted values; forced expiratory volume in one second (FEV1), expressed as a percent of the predicted value; pseudomonas colonization; and chest-film score. Pulmonary-function tests were performed with Sensormedics, Vitalograph, Jaeger, P.K. Morgan, and Biomedin equipment. The majority of centers reported the best of three reproducible efforts performed in one day, one reported the best of two, and one the best of four. The most recent representative study — i.e., one not performed during an acute pulmonary exacerbation — was reported.

Pseudomonas colonization was defined as the first positive culture after a series of negative cultures (or as the first positive culture on record). In all but six centers, throat cultures were performed for children too young to provide sputum samples. Cultures were routinely performed at every clinic visit (every three months) in all but two centers. The most recent chest film was scored on either the Brasfield scale (ranging from 0 to 25, with 25 representing the best possible score) or the Chrispin–Norman scale (ranging from 0 to 38, with 0 representing the best possible score). Centers were asked to specify the method they used routinely. One investigator at each center was requested to score all the chest films, without knowledge of the patients' identities or genotypes. Results for the two types of chest-film score were analyzed separately.

Categorical data about pancreatic status (i.e., whether function was sufficient or insufficient) were requested. The centers were asked which method they used to determine pancreatic function: 72-hour fecal-fat study, pancreatic-stimulation test, serum trypsinogen assay, or subjective assessment of the need for pancreatic enzyme supplementation. In addition, the patient's age at the onset of pancreatic insufficiency was requested, if applicable.

The occurrence of several common complications of cystic fibrosis was recorded: history of nasal polyps; meconium ileus; distal intestinal obstruction syndrome; death as the need for enemas, intestinal lavage, or surgery after the neonatal period; pancreatitis; diabetes mellitus, defined as insulin dependence; biliary cirrhosis; and gallbladder disease, defined as cholelithiasis or sludging on abdominal sonography.

**Statistical Analysis**

In the case of continuous variables, mean values were compared by a two-tailed paired t-test. For categorical variables, conditional logistic regression was used to compare proportions. Because pairs in which both patients have the same outcome for a binary variable do not contribute to a conditional analysis, many comparisons were based on small numbers. When there were fewer than 20 contributory pairs, exact binomial probabilities were computed to test the null hypothesis. In all analyses the nominal significance level was 0.05, but because multiple comparisons were performed, only very small probability values were interpreted as significant. All P values are two-tailed.

**Results**

Table 1 summarizes the data on patients in each genotype group. Each matched group of ΔF508 homozygotes had the same distribution of age and sex as the corresponding group of subjects with a specific genotype. Ages were matched within one year for 79 percent of pairs, and within two years for 92 percent. Those not matched within two years were all older than 20 years of age. For brevity, only the combined data for ΔF508 homozygotes are shown. However, all the analyses took account of the pairing of the study subjects. The variability in sex distribution in the different genotype groups was increased in the groups with smaller samples. The R117H compound heterozygotes were significantly older than the patients in all other genotype groups (P<0.001).

All the patients were white. Over half, primarily those from the United States, Canada, and New Zealand, were of unspecified or unknown ethnic origin. The remainder, in descending order of frequency, were Italian, German, Czech, French Canadian, French, Belgian, Scottish, Irish, Slovak, Aškenazi Jewish, English, Bulgarian, Austrian, Brazilian, Albanian, and Macedonian. Since only a few centers reported data on patients who had died, no attempt was made to analyze death rates.

There were no significant differences with regard to any variable tested between the ΔF508 homozygotes and the patients with the G542X/ΔF508, R553X/ΔF508, N1303K/ΔF508, W1282X/ΔF508, 1717-1G→A/ΔF508, and 621+1G→T/ΔF508 genotypes, or the G542X, R553X, W1282X homozygotes and compound heterozygotes. Pancreatic status and age at the onset of any pancreatic insufficiency were analyzed both in the entire data set and in those pairs for whom objective tests of pancreatic function were reported. There were no differences in the results. The mean age at the onset of pancreatic insufficiency was virtually identical to the mean age at diagnosis. The patients with the 621+1G→T/ΔF508 genotype were somewhat younger at diagnosis than the matched ΔF508 homozygotes (P = 0.03), and none of the patients with the G542X/ΔF508 genotype had pancreatic sufficiency, as compared with 5 of the 147 matched ΔF508 homozygotes (P = 0.03). These were quite likely to be chance findings, given the large number of comparisons and the lack of a general or interpretable trend in the data. Data on cirrhosis, diabetes, rectal prolapse, gallbladder disease, and pancreatitis are not shown in Table 1. The overall prevalence of each of these conditions was less than 8 percent, and there was no trend toward increased prevalence in any genotype.

The patients with the R117H/ΔF508 genotype clearly differed from the matched ΔF508 homozygotes (Ta-
Table 1. Characteristics of the Patients According to Genotype.

<table>
<thead>
<tr>
<th>VARIABLE*</th>
<th>ΔPSG/ΔPSO</th>
<th>G542X/ΔPSO</th>
<th>R553X/ΔPSO</th>
<th>N1303X/ΔPSO</th>
<th>GENOTYPE</th>
<th>1717-10→A/ΔPSO</th>
<th>621+1G→T/ΔPSO</th>
<th>R117H/ΔPSO</th>
<th>X/T</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>399</td>
<td>148</td>
<td>52</td>
<td>60</td>
<td>17</td>
<td>30</td>
<td>51</td>
<td>23</td>
<td>18</td>
</tr>
<tr>
<td>Male sex — no. (%)</td>
<td>216 (54)</td>
<td>85 (57)</td>
<td>25 (48)</td>
<td>33 (55)</td>
<td>7 (41)</td>
<td>19 (63)</td>
<td>28 (55)</td>
<td>12 (52)</td>
<td>7 (39)</td>
</tr>
<tr>
<td>Age (yr) — mean ± SD</td>
<td>13.0 ± 8.7</td>
<td>11.9 ± 8.7</td>
<td>12.5 ± 8.1</td>
<td>12.3 ± 8.0</td>
<td>11.0 ± 10.8</td>
<td>11.8 ± 8.0</td>
<td>14.6 ± 7.7</td>
<td>23.5 ± 9.6‡</td>
<td>12.2 ± 9.7</td>
</tr>
<tr>
<td>Age at diagnosis — yr</td>
<td>1.7 ± 3.0 (392)</td>
<td>1.6 ± 3.1 (147)</td>
<td>1.7 ± 2.7 (52)</td>
<td>1.5 ± 2.7 (58)</td>
<td>4.0 ± 9.9 (17)</td>
<td>2.0 ± 4.4 (28)</td>
<td>0.8 ± 1.1 (51)¶</td>
<td>10.2 ± 10.5 (23)‡</td>
<td>0.6 ± 0.6 (16)</td>
</tr>
<tr>
<td>Sweat chloride — mmol/liter</td>
<td>106 ± 22 (328)</td>
<td>109 ± 23 (128)</td>
<td>105 ± 18 (46)</td>
<td>104 ± 24 (56)</td>
<td>110 ± 18 (13)</td>
<td>107 ± 36 (26)</td>
<td>100 ± 20 (22)</td>
<td>82 ± 19 (20)¶</td>
<td>105 ± 19 (16)</td>
</tr>
<tr>
<td>FEV1 — % of predicted</td>
<td>70 ± 27 (269)</td>
<td>67 ± 27 (81)</td>
<td>64 ± 25 (36)</td>
<td>69 ± 24 (42)</td>
<td>75 ± 26 (12)</td>
<td>68 ± 26 (20)</td>
<td>73 ± 26 (41)</td>
<td>73 ± 22 (22)¶</td>
<td>70 ± 25 (11)</td>
</tr>
<tr>
<td>Shwachman clinical score†</td>
<td>75 ± 17 (245)</td>
<td>74 ± 19 (73)</td>
<td>79 ± 12 (30)</td>
<td>72 ± 19 (42)</td>
<td>79 ± 13 (12)</td>
<td>71 ± 20 (16)</td>
<td>75 ± 19 (45)</td>
<td>81 ± 14 (21)¶</td>
<td>76 ± 17 (13)</td>
</tr>
<tr>
<td>no. positive/no. studied (%)</td>
<td>10/936 (2.5)</td>
<td>0/147 (0)¶</td>
<td>1/51 (2.0)</td>
<td>0/99 (0)</td>
<td>0/16 (0)</td>
<td>1/30 (3.3)</td>
<td>1/51 (2.0)</td>
<td>20/23 (87.0)¶</td>
<td>1/17 (5.9)</td>
</tr>
<tr>
<td>Pancreatic sufficiency</td>
<td>210/377 (56)</td>
<td>58/138 (42.0)</td>
<td>33/50 (66.0)</td>
<td>29/55 (52.7)</td>
<td>14/17 (82.4)</td>
<td>14/29 (48.3)</td>
<td>32/51 (62.8)</td>
<td>7/23 (30.4)</td>
<td>7/17 (41.2)</td>
</tr>
<tr>
<td>Pseudomonas colonization</td>
<td>57/394 (14.5)</td>
<td>35/147 (23.8)</td>
<td>6/50 (12.0)</td>
<td>6/60 (10.0)</td>
<td>1/17 (5.9)</td>
<td>5/30 (16.7)</td>
<td>6/51 (11.8)</td>
<td>0/23 (0)</td>
<td>1/17 (5.9)</td>
</tr>
<tr>
<td>Meconium ileus</td>
<td>49/313 (15.6)</td>
<td>18/114 (15.8)</td>
<td>5/34 (14.7)</td>
<td>8/52 (15.4)</td>
<td>6/15 (40)</td>
<td>4/29 (13.8)</td>
<td>9/43 (20.9)</td>
<td>5/18 (27.8)</td>
<td>0/16 (0)</td>
</tr>
<tr>
<td>Nasal polyposis</td>
<td>40/309 (12.9)</td>
<td>23/111 (20.7)</td>
<td>3/68 (4.3)</td>
<td>7/51 (13.7)</td>
<td>2/15 (13.3)</td>
<td>1/29 (3.5)</td>
<td>9/47 (19.2)</td>
<td>3/17 (17.7)</td>
<td>1/16 (6.3)</td>
</tr>
</tbody>
</table>

*FEV1 denotes forced expiratory volume in one second, and DIOS distal intestinal obstruction syndrome.
†X/T indicates that the patient is a homozygote or compound heterozygote for G542X, R553X, or W1282X.
‡See Table 2 for a matched comparison with ΔPSO homozygotes with respect to this variable.
¶P = 0.03 for a paired comparison with matched ΔPSO homozygotes.

ble 2). Among 23 pairs of patients, 19 were discordant with respect to pancreatic status and contributed to the paired statistical analysis. All 19 patients with the R117H/ΔPSO genotype had pancreatic sufficiency, whereas all the matched ΔPSO homozygotes had pancreatic insufficiency. For only 10 of these pairs was an objective test cited for the determination of pancreatic status. The four pairs of patients in which pancreatic status was concordant included three with pancreatic insufficiency and one with pancreatic sufficiency. The methods of pancreatic-function testing were not cited for these four pairs.

Two of the three patients with the R117H/ΔPSO genotype who had pancreatic insufficiency had a history of pancreaticitis. There were only two cases of pancreatitis among the 308 patients homozygous for ΔPSO for whom this information was available, and two cases among 50 patients with the 621+1G→T/ΔPSO genotype; no cases were seen in any of the other genotype groups. Conversely, meconium ileus was not seen in the group with the R117H/ΔPSO genotype.

As expected with pancreatic sufficiency, the patients with the R117H/ΔPSO genotype were older at diagnosis than those homozygous for ΔPSO, and their mean sweat chloride concentrations were lower. There were no differences between the patients with the R117H/ΔPSO genotype and either the matched or the combined group homozygous for ΔPSO with respect to any other outcome measure.

An overall analysis comparing the ΔPSO homozygotes and heterozygotes was performed with a combined set of matched pairs, excluding the heterozygotes with the genotype R117H/ΔPSO, the patients with two nonsense mutations, and the ΔPSO homozygotes controls for these two groups. There were no significant findings and no trends suggestive of differences between the ΔPSO heterozygotes and homozygotes in the 358 matched pairs.

**DISCUSSION**

The results of this international collaborative study and the results reported elsewhere allow a phenotypic description of approximately 62 percent of patients with cystic fibrosis. A cystic fibrosis genotype is a predictor of pancreatic status is confirmed conclusively by this study. Kerem and colleagues initially suggested that one "mild" allele (causing pancreatic insufficiency) would dominate over a "severe" allele (causing pancreatic insufficiency) such that two alleles for pancreatic insufficiency would be necessary for pancreatic insufficiency to be manifest. The large proportion of patients with pancreatic sufficiency (20 of 23) who had the R117H/ΔPSO genotype supports...
this hypothesis and the earlier observation that the R117H mutation is associated with pancreatic sufficiency.

Our study was cross-sectional, and therefore a selection bias created by mortality may explain some of the lack of phenotypic differences. Canadian and U.S. registries of data on patients are now recording genotypes and will soon provide unselected population data on genotype-phenotype correlations. However, several years of follow-up will be required before the effects on mortality can be estimated. In addition, it is possible that some real phenotypic differences in a single mutation or group of mutations may have gone undetected in this study because of the small size of the genotype groups.

Studies performed before the identification of the cystic fibrosis gene suggested that patients with pancreatic sufficiency had milder lung disease than those with pancreatic insufficiency. Despite the strong association between the cystic fibrosis genotype and the pancreatic phenotype, this study demonstrates that the severity and course of pulmonary disease are not predicted by the genotype. This observation is in agreement with a study from a single center in which there was wide variability in the extent of pulmonary disease within each genotype group. Some patients with pancreatic sufficiency had severe pulmonary manifestations, and many patients with pancreatic insufficiency had normal pulmonary function. This suggests that factors other than the cystic fibrosis genotype affect the pulmonary phenotype. A recent study comparing the course of pulmonary disease in monozygotic and dizygotic twins suggested a strong genetic contribution to the pulmonary phenotype that was independent of the cystic fibrosis genotype.

Environmental factors, such as exposure to passive smoking, also influence the course of pulmonary disease in patients with cystic fibrosis.

The lower sweat chloride concentration in patients with the R117H/ΔF508 genotype demonstrates that the CFTR genotype correlates with the degree of sweat-gland dysfunction. This result was not unexpected, since lower sweat chloride concentrations had been reported in patients with pancreatic sufficiency before the identification of the CFTR gene. Among adult men with congenital bilateral absence of the vas deferens but normal pulmonary and pancreatic function, 41 percent carry one copy of the ΔF508 mutation. Interestingly, 14 percent of these patients carry the R117H mutation on the other allele. Transient transfection studies have shown that CFTR carrying the R117H mutation is partially functional when expressed in Xenopus oocytes and HeLa cells. However, the mechanism whereby the R117H mutation gives rise to two different clinical phenotypes remains to be studied.

There were no statistically significant differences in the incidence of common complications of cystic fibrosis among the genotypes studied, suggesting that their occurrence may be influenced by other genetic and environmental factors.

Prenatal and prognostic counseling should not differ for patients homozygous for ΔF508 and those with the G542X/ΔF508, R553X/ΔF508, W1282X/ΔF508, N1303K/ΔF508, 621+1G→T/ΔF508, or 1717−1G→A/ΔF508 genotype. All should be counseled to expect pancreatic insufficiency of early onset. The same applies to homozygotes or compound heterozygotes for the G542X, R553X, and W1282X nonsense mutations. Patients with the R117H/ΔF508 genotype and their parents should be counseled to expect long-term pancreatic sufficiency. No predictions can be made about the occurrence of common complications of cystic fibrosis or the severity or course of pulmonary disease, because of the wide variability in each group of patients carrying the cystic fibrosis genotypes studied.

The results of this study demonstrate that the only prognostic value of genotypic information is for the prediction of pancreatic status. In patients with pancreatic sufficiency, knowing the genotype may help predict future pancreatic status and thus affect expectant management. Since the patients' mean age at the onset of pancreatic insufficiency was less than two years for each genotype associated with pancreatic insufficiency, the evaluation of genotypes for prognostic purposes in patients with cystic fibrosis should be useful for patients with pancreatic sufficiency who are more than two years of age. Patients found to carry two alleles for pancreatic insufficiency should be followed closely for the development of this condition, whereas those who carry an allele associated with pancreatic sufficiency would not need such evaluation.

### Appendix

The following persons participated in the Cystic Fibrosis Genotype–Phenotype Consortium: Center for Medical Genetics, Johns Hopkins University School of Medicine, Baltimore, MD; University of Utah, Salt Lake City; Children's Hospital, Boston; University of Wisconsin, Madison; and the University of North Carolina at Chapel Hill.


Massachusetts Medical Society
Registry on Continuing Medical Education

To obtain information on continuing medical education courses in the New England area, call between 9:00 a.m. and 12:00 noon, Monday through Friday, (617) 893-4610 or in Massachusetts 1-800-322-2303, ext. 1342.