Genetic Modifiers of Cystic Fibrosis–Related Diabetes

Scott M. Blackman,1,2 Clayton W. Commander,3 Christopher Watson,2 Kristin M. Arcara,1 Lisa J. Strug,1,5 Jaclyn R. Stonebraker,3 Fred A. Wright,2 Johanna M. Rommens,6,7 Lei Sun,4,8 Rhonda G. Pace,3 Sarah A. Norris,3 Peter R. Durie,9,10 Mitchell L. Drumm,11 Michael R. Knowles,3 and Garry R. Cutting2

Diabetes is a common age-dependent complication of cystic fibrosis (CF) that is strongly influenced by modifier genes. We conducted a genome-wide association study in 3,059 individuals with CF (644 with CF-related diabetes [CFRD]) and identified single nucleotide polymorphisms (SNPs) within and 5’ to the SLC26A9 gene that associated with CFRD (hazard ratio [HR] 1.38; P = 3.6 × 10−15). Replication was demonstrated in 694 individuals (124 with CFRD; HR, 1.47; P = 0.007) (1), with combined analysis showing a P = 9.8 × 10−10; SLC26A9 is an epithelial chloride/bicarbonate channel that can interact with the CF transmembrane regulator (CFTR), the protein mutated in CF. We also hypothesized that common SNPs associated with type 2 diabetes also might affect risk for CFRD. A previous association of CFRD with SNPs in TCF7L2 was replicated in this study (P = 0.004; combined analysis P = 3.8 × 10−10), and type 2 diabetes SNPs at or near CDKAL1, CDKN2A/B, and IGF2BP2 were associated with CFRD (P < 0.004). These five loci accounted for 8.3% of the phenotypic variance in CFRD onset and had a combined population-attributable risk of 68%. Diabetes is a highly prevalent complication of CF, for which susceptibility is determined in part by variants at SLC26A9 (which mediates processes proximate to the CF disease-causing gene) and at four susceptibility loci for type 2 diabetes in the general population. Diabetes 62:3627–3635, 2013

Cystic fibrosis (CF) is a common life-limiting monogenic disease in Caucasians caused by defects in an epithelial chloride channel, CF transmembrane regulator (CFTR), which is expressed across tissues, including sweat glands, pancreas, and lung. Diabetes is an age-dependent complication of CF that affects 19% of adolescents and 40–50% of adults with CF (1). CF-related diabetes (CFRD) is associated with worse lung disease, malnutrition, and mortality (2), and treating CFRD substantially improves outcomes (1,3). Risk factors for CFRD include pancreatic exocrine insufficiency (4), female sex (5), and liver disease (6). Genetic modifiers (genes other than CFTR) contribute to the risk of CFRD (7), and identification of these modifiers could give insight into the pathophysiology of CFRD.

The clinical and histologic features of CFRD share some similarities with other forms of diabetes in the general population, but there are also distinct differences. For example, type 2 diabetes and CFRD are associated with a subacute decline in β-cell function and production of islet amyloid (8,9). In contrast, insulin sensitivity is reduced in type 2 diabetes but is generally normal in CFRD (except during exacerbations or glucocorticoid treatment) (1). These findings suggest that CFRD and type 2 diabetes have both common and distinct mechanisms; therefore, dissection of the contributing pathways could be informative for both conditions.

Identification of genes that confer risk for CFRD can address the degree of overlap with type 2 diabetes. There is a reasonable basis for a genetic approach as follows: >50 common gene variants associated with type 2 diabetes (10); a family history of type 2 diabetes (i.e., diabetes in non-CF family members) approximately triples the risk for CFRD (11); and in a total of 1,745 CF individuals, we previously demonstrated that TCF7L2, a susceptibility gene for type 2 diabetes, confers risk for CFRD (11). The formation of the CF Genetic Modifier Consortium and genome-wide single nucleotide polymorphism (SNP) typing of ~3,500 CF patients (12) afforded an opportunity to search for unique and shared risk factors using genome-wide association analysis. To increase the power for detecting variants affecting both CFRD and type 2 diabetes, we also tested candidate type 2 diabetes SNPs at 7 loci for association with CFRD.

RESEARCH DESIGN AND METHODS

Subjects. The study participants contributing to this analysis have been described previously. The Twin and Sibling Study (TSS) recruited sets of living twins and siblings with CF and their parents (13). The Genetic Modifier Study (GMS) recruited people with CF who were homozygous for F508del (the most common disease-causing CFTR mutation) and had either mild or severe lung function (14). Additional GMS participants were from an international CF liver disease study that included all CFTR genotypes (15). The Canadian CF Gene Modifier Study (CGS) recruited from the majority of CF centers in Canada and was representative of the national CF population (16). From those recruited, a sample was selected with severe exocrine pancreatic insufficiency or CFTR genotype (or both) expected to confer little or no residual CFTR function. A discovery sample was drawn from those individuals in CGS, TSS, and GMS lung and liver studies who were genotyped in 2007 (12,15). A separate replication sample was composed of individuals who were not genotyped initially for a variety of reasons, such as recruitment or acquisition of lung function data after the genotyping had been performed (16). CFRD status was ascertained from clinic records (7). Diabetes was defined by clinician diagnosis of CFRD plus insulin treatment for at least 1 year (in the TSS and GSS; 6 months in the GMS lung study). CFRD age of onset was
missing for 11 in the discovery sample and 10 in the replication sample. In the TSS, additional clinical information was used to exclude 390 individuals with intermediate glucose tolerance, as was performed previously (7). In the remaining 108 families with multiple children, one individual per family was included (with preference given to including individuals with diabetes, and then older individuals; this excluded 19 with diabetes and 89 without).

Demographics, CFTR genotype, and meconium ileus (MI) were defined by individual chart review, with adequate documentation of MI required as in previous studies (16,17). Liver disease was defined by clinician diagnosis in two studies (TSS and CGS), whereas the GMS study required documentation of portal hypertension attributable to cirrhosis (15). All study participants provided informed consent, and all the studies were approved by the Institutional Review Boards at participating institutions.

**Genotyping and quality control.** SNPs in the discovery sample were genotyped by Genome Quebec using the Illumina 610-Quad platform, and quality control was performed as described previously (12,16). Comparison with previous genotyping yielded low platform discordance as assessed by 542 Illumina GoldenGate SNPs typed in the GMS portion of the discovery sample (0.07%) and by the rs7903146 SNP typed in the TSS and GMS portions of the discovery sample (0.24%). SNPs that were monomorphic in any of the three discovery samples or that had overall minor allele frequency <1% were excluded, leaving 549,869 SNPs from chromosomes 1-22 and the X chromosome to be tested. SNPs in the replication sample were typed using TaqMan Assays-on-Demand (Applied Biosystems, Foster City, CA) (16).

SNPs within a 1-Mb region around SLC26A9 were imputed from discovery sample genotypes spanning 188.9-219.9 Mb (National Center for Biotechnology Information 36.3 coordinates), with MACH and Minimac (18) using reference haplotypes from the 1,000 Genomes Project (August 2010 release) (19). Genotypes for 1,567 SNPs were imputed with MACH quality score $R^2 > 0.3$ (19).

**Statistical methods.** SNPs were analyzed for association with CFRD age at onset using a proportional hazards model (event: diagnosis of diabetes; censoring: age at most recent diabetes testing). The “unadjusted” analysis used an additive genetic model along with three to eight genotype principal components (number selected by Scree plot) (20) as covariates, and the “adjusted” analysis also included covariates for female sex and liver disease. Study results were combined using a meta-analysis $Z$-statistic (21) calculated as $Z = W_{\text{discovery}} + W_{\text{replication}}$, where the weight ($W$) is inversely proportional to the SE. A common reference allele was used for each SNP to preserve direction of effect. The proportional hazards assumption was confirmed for all significantly associated SNPs by testing for time dependence of Schoenfeld residuals (22) (Stata `estat phtest`). Heterogeneity in meta-analysis was assessed using a $\chi^2$ test (23); for $\chi^2 > 0.25$, data also were analyzed using a Weibull model with shared frailty to allow for study-specific effect heterogeneity. The PLINK software package (24) was used for data handling, and R (http://www.r-project.org) and Stata 11 (StataCorp, College Station, TX) were used for analysis. Regional $P$ value plots were generated with LocusZoom (http://csg.sph.umich.edu/locuszoom/).

Observed versus expected plots of $P$ values on a log scale for each study (data not shown) and for the combined discovery meta-analysis (Supplementary Fig. 1) demonstrated no substantial deviation from the expected distribution of $P$ values, except among those with $P < 10^{-8}$.suggestive association was declared for $P$ values lower than the following conservative threshold: $1 / (\text{number of SNPs}) / 1 / 549,869 = 1.8 \times 10^{-6}$. Significant association was declared using a conservative Bonferroni-corrected threshold of $P < 0.05 / 549,869 = 9.1 \times 10^{-5}$. Genotypes among males of X-chromosome SNPs were encoded according to default PLINK settings (zero or one copy of the minor allele), but an alternative coding to zero or two copies (assuming X-inactivation) resulted in no qualitative changes in conclusions or in identification of the most significant SNPs on chromosome X (data not shown).

In testing type 2 diabetes SNPs for association with CFRD, seven loci were selected from the eight subsequently replicated loci reported in the earliest genome-wide association studies of type 2 diabetes. The eighth locus, FTO, was not tested as a CFRD candidate gene because the increased risk of type 2 diabetes appears to be mediated by increased BMI (not generally the case for people with CFRD). When the associated SNP in a given type 2 diabetes locus was not genotyped in this study, a proxy was chosen based on linkage disequilibrium. For PPARG, no good proxy was available and imputed genotypes were used. Statistical significance in the candidate study was defined as a two-sided $P < 0.004 (0.05 / 12$ SNPs tested). Because of linkage disequilibrium between SNPs at the same locus, the effective number of tests should be <12, so this threshold may be more conservative than necessary.

**RESULTS**

**Characteristics of the discovery and replication samples.** The discovery sample included 3,059 individuals (Table 1) from the CGS, GMS, and TSS samples. Because of differences in ascertainment criteria, the TSS-D (D denotes discovery) subjects had a mean age that was ~2 years younger, and almost all GMS-D subjects were homozygous for F508del. The replication sample included 694 individuals from the CGS and GMS. The CGS-R (R denotes replication) individuals had younger mean age (10.5 years) than the other study subsets but had a similar proportion of F508del homozygotes (64%). The GMS-R individuals were older than average (mean age, 33 years) and fewer (41%) were homozygous for F508del. Ethnicity was reported for 98% of participants, and 92% were Caucasian (GMS-D: 96; CGS-D: 89; TSS-D: 93; GMS-R: 94; and CGS-R: 81%). As expected for an age-dependent disorder, subjects with CFRD were older (discovery: $P = 7 \times 10^{-10}$; replication: $P = 3 \times 10^{-27}$) and CFRD prevalence increased with increasing age (odds ratio, 1.07 per year; 95% CI, 1.06–1.08). The cumulative incidence of CFRD, reflecting the CFRD rate by age, did not differ in the discovery sample between TSS and GMS (Supplementary Fig. 2). However, CGS had a somewhat lower rate of CFRD across all ages, which could reflect either patient health or CFRD screening practices. In the replication sample, CFRD onset did not differ between GMS-R and CGS-R ($P = 0.93$) and was between that of the discovery populations (log rank $P < 0.05$; Supplementary Fig. 2).

Female sex and liver disease were independent risk factors for CFRD onset (Supplementary Table 1), as seen in previous studies (5,7). Heterogeneity was evident across the three discovery and two replication subgroups for the

| TABLE 1 | Participant characteristics in the discovery and replication samples analyzed for CFRD onset |
|---------|------------------------------------------------------------------------------------------|
|         | PI subjects | DF/DF subjects, n (%) | Female (PI), n (%) | Diabetes (PI), n (%) | Diabetes (DF/DF), n (%) | Mean age (PI) (years) |
|         | (n)         | (n) %                   | (n) %                 | (n) %                          | (n) %                          | CFDR No CFRD Both |
| Discovery sample | | | | | | |
| GMS | 1,263 | 1,217 (96) | 588 (47) | 374 (30) | 354 (29) | 27.9 | 22.4 | 24.0 |
| CGS | 1,508 | 915 (51) | 683 (45) | 167 (11) | 100 (11) | 28.6 | 18.8 | 19.8 |
| TSS | 288 | 171 (59) | 140 (49) | 103 (36) | 62 (36) | 22.4 | 15.2 | 17.8 |
| Combined | 3,059 | 2,303 (75) | 1,411 (46) | 644 (21) | 516 (22) | 27.2 | 19.8 | 21.4 |
| Replication sample | | | | | | |
| GMS-R | 409 | 166 (41) | 188 (46) | 104 (25) | 42 (25) | 32.9 | 26.1 | 27.9 |
| CGS-R | 285 | 182 (64) | 135 (47) | 20 (7) | 12 (7) | 27.0 | 9.3 | 10.5 |
| Combined | 694 | 348 (50) | 323 (47) | 124 (18) | 54 (16) | 31.8 | 18.4 | 20.9 |

DF/DF, homozygous F508del; PI, pancreatic exocrine insufficiency; R, replication sample. |
degree of risk conferred by sex and liver disease (each with \( P < \) 10^{-6}, \( r^2 \geq 90\% \)), possibly reflecting cohort effects or differences in study design (e.g., subject ascertainment criteria or phenotype definitions). The greater hazard ratio (HR) for liver disease in the GMS than in CGS and TSS (whose HRs did not differ; \( P = 0.9 \)) is attributed to the stringent criteria for liver disease (i.e., presence of portal hypertension attributable to cirrhosis). When compared with other CFTR mutations conferring severe pancreatic exocrine insufficiency, F508del homozygosity was not a significant risk factor for CFRD onset.

**Genome-wide significant association between SNPs in SLC26A9 and CFRD.** Genome-wide association analysis of the discovery sample identified two SNPs on chromosome 1 associated with CFRD onset (Fig. 1 and Table 2). The region of significance was within and 5′ to SLC26A9 (Fig. 2A), and the two SNPs with the lowest \( P \) value (rs4077468 and rs4077469; \( P = 3.59 \times 10^{-8} \)) were in complete linkage disequilibrium in the discovery sample (\( r^2 = 1.0 \)). Therefore, results from the discovery sample apply equally to rs4077468 and rs4077469. Conditioning on the top-ranked SNP (rs4077468) reduced evidence for association with the remaining SNPs in the region to \( P > 0.05 \) (the lowest \( P \) value was for rs7555534; \( P = 0.06 \); Fig. 2B), indicating no significant evidence for locus heterogeneity at SLC26A9. Each “A” allele of rs4077468 conferred increased risk of CFRD in the discovery sample as a whole (HR, 1.38 per allele; 95% CI, 1.23–1.54; Fig. 3A) and in each of the three study subgroups (Supplementary Table 2). No locus other than the SLC26A9 locus contained SNPs with \( P \) values surpassing a genome-wide suggestive threshold in the unadjusted analysis. When adjusting for liver disease and female sex (Table 2, adjusted analysis), the same SNPs at the SLC26A9 locus were associated (rs1874361, \( P = 1.6 \times 10^{-8} \); rs4077468, \( P = 2.5 \times 10^{-8} \)). In addition, one SNP in each of four other loci (CYP11B2, KRT18P33, NCKAP1L, and LPHN3) had suggestive evidence for association in the adjusted analysis.

To determine if association could be reproduced, subjects in the replication sample (\( n = 124 \) with CFRD; Table 1) were genotyped for rs4077468. Again, the “A” allele of rs4077468 associated with CFRD onset (HR, 1.47; 95% CI, 1.11–1.94; \( P = 0.007 \); Fig. 3B). When analyzing the two subsets of the replication sample separately, association was seen in the GMS subset (\( n = 104 \) with CFRD; HR, 1.58; 95% CI, 1.16–2.15; \( P = 0.004 \)), but the CGS subset did not provide support for association (\( n = 285 \); HR; 1.05; 95% CI, 0.5–2.0; \( P = 0.9 \)), possibly because of the young age and low rate of CFRD in that subset. A meta-analysis of discovery and replication samples supports association of rs4077468 with CFRD onset (\( n = 3,753 \); HR, 1.39 per allele; 95% CI, 1.25–1.54; \( P = 9.8 \times 10^{-10} \)).

To test whether genetic variation at the CFTR locus might affect the association of SNPs at SLC26A9, a second analysis was performed with 2,303 individuals homozygous for F508del. Data from this reduced but more genetically homogeneous sample supported association of CFRD with SNPs at the SLC26A9 locus with the same magnitude of effect (e.g., rs4077468: HR, 1.36; 95% CI, 1.20–1.54; \( P = 1.9 \times 10^{-6} \); Supplementary Table 3). By regression and meta-analysis, there was no evidence for correlation or interaction between rs4077468 and CFTR genotype (homozygous F508del status or number of F508del alleles; \( P > 0.05 \)). The subset of 132 individuals with zero F508del mutations did not provide support for association (\( n = 132 \); HR, 1.13; 95% CI, 0.6–2.2).

Analysis of imputed SNPs in the SLC26A9 region revealed five additional SNPs, all with high-quality imputed
genotypes \((R^2 > 0.9)\), which were associated with CFRD onset (Supplementary Table 4). Conditional analysis (not shown) demonstrated that these five SNPs tag the same genetic association signal as rs4077468.

The CFRD modifier SNPs in SLC26A9 are located in the promoter region (<5 kb upstream of transcription start) and within the first intron (Fig. 4), suggesting a role in splicing or expression. Putative transcription factor binding regions were identified using published data collected within the UCSC Genome Browser, including FAIRE-seq (25), DNase I hypersensitivity, and ChIP-seq (Supplementary Fig. 3). The three SNPs 5′ of SLC26A9 with the most significant \(P\) values for CFRD onset (rs4077468, rs4077469, and rs4951271) flank a region of DNase I hypersensitivity present in pancreatic islets but absent in dedifferentiated islets, which also binds transcription factors according to ChIP-seq analysis (orange boxes, Supplementary Fig. 3A) (26). Associated SNPs in intron 1 are proximate to three regions that may represent transcription factor binding sites that are active in multiple tissues, including tracheal

### Table 2
Genotyped SNPs associated with CFRD at a suggestive or significant level in the discovery sample

| SNP       | Chr | NCBI36 position (bp) | Risk/other allele | RAF   | HR    | \(P\)    | HR adj  | \(P\) adj  | Annotation     |
|-----------|-----|----------------------|-------------------|-------|-------|---------|---------|-----------|----------------|
| rs4077468 | 1   | 204,181,380          | A/G               | 0.58  | 1.38  | \(3.6 \times 10^{-8}\) | 1.39    | \(2.5 \times 10^{-8}\) | SLC26A9       |
| rs4077469 | 1   | 204,181,508          | C/T               | 0.58  | 1.38  | \(3.6 \times 10^{-8}\) | 1.39    | \(2.5 \times 10^{-8}\) | SLC26A9       |
| rs7415921 | 1   | 204,177,506          | G/T               | 0.45  | 1.34  | \(1.6 \times 10^{-7}\) | 1.35    | \(1.8 \times 10^{-7}\) | SLC26A9       |
| rs1874361 | 1   | 204,174,809          | A/C               | 0.47  | 1.33  | \(3.3 \times 10^{-7}\) | 1.38    | \(1.6 \times 10^{-7}\) | SLC26A9       |
| rs7512462 | 1   | 204,166,218          | T/C               | 0.59  | 1.34  | \(4.0 \times 10^{-7}\) | 1.36    | \(2.9 \times 10^{-7}\) | SLC26A9       |
| rs7555534 | 1   | 204,175,490          | C/T               | 0.33  | 1.33  | \(7.2 \times 10^{-7}\) | 1.37    | \(8.8 \times 10^{-8}\) | SLC26A9       |
| rs7419153 | 1   | 204,183,932          | A/G               | 0.37  | 1.32  | \(1.5 \times 10^{-6}\) | 1.33    | \(2.0 \times 10^{-6}\) | SLC26A9       |
| rs6981918 | 8   | 144,004,941          | T/G               | 0.01  | 2.60  | \(2.9 \times 10^{-6}\) | 2.69    | \(1.6 \times 10^{-6}\) | CYP11B2       |
| rs11902125| 2   | 65,692,304           | T/C               | 0.26  | 1.62  | \(3.0 \times 10^{-6}\) | 1.67    | \(1.3 \times 10^{-6}\) | KRT18P33      |
| rs4759088 | 12  | 53,210,771           | T/C               | 0.30  | 1.31  | \(5.4 \times 10^{-6}\) | 1.33    | \(1.5 \times 10^{-6}\) | NCKAPIL       |
| rs995447  | 4   | 62,501,063           | G/A               | 0.06  | 1.54  | \(6.7 \times 10^{-5}\) | 1.72    | \(8.8 \times 10^{-5}\) | LPHN3         |

The 3,059 discovery samples were analyzed while adjusting for principal components (HR and \(P\) shown) and while adjusting also for female sex and liver disease (HR adj and \(P\) adj). Risk allele is defined as the allele associated with increased risk of CFRD. Suggestive: \(P < 1.8 \times 10^{-6}\). Significant: \(P < 9.1 \times 10^{-8}\). HR is shown per allele. adj, adjusted; Chr, chromosome; RAF, risk allele frequency. *Study-wide significant. †Study-wide suggestive. ‡These SNPs are in 100% linkage disequilibrium in the discovery sample.

---

**FIG. 2.** Regional plot of negative log \(P\) values for SNPs at or near SLC26A9 gene. **A:** Three-study meta-analysis illustrating maximum evidence for association at rs4077468 and rs4077469 (◆). **B:** Three-study meta-analysis performed while conditioning the number of minor alleles at rs4077468, demonstrating no evidence for locus heterogeneity (all \(P > 0.05\)).
epithelium and fibroblasts that show evidence for transcription factor binding by ChIP-seq (blue boxes, Supplementary Fig. 3).

Association between SLC26A9 and CFRD is independent of MI, CF-related liver disease, and sex. Several of the SLC26A9 SNPs associated with CFRD also were associated with MI, another important complication of CF (16). In the discovery sample, MI was weakly correlated with CFRD onset (HR, 1.3; P = 0.01), a relationship that was restricted to the GMS subset (HR, 1.7; P < 0.001; CGS: P = 0.8; TSS: P = 0.3), raising the possibility that MI could be a confounder. However, rs4077468 associated with CFRD in sample subsets both with MI and without MI (Supplementary Table 5; stratified analysis P = 3.66 × 10^{-5}) and when including an adjustment for MI (P = 5.23 × 10^{-5}, P value for MI covariate term = 0.01) or an interaction with MI (P value for SNP*MI interaction term = 0.6). In a similar set of analyses, no evidence was found for interaction between SNP effect and liver disease (Supplementary Table 6) or female sex (Supplementary Table 7). Thus, SLC26A9 SNPs modify risk of CFRD independently of MI, liver disease, and female sex.

The previous association study of CF-specific lung function (12) included the SLC26A9 SNPs and demonstrated no significant evidence of association with rs4077468 (P = 0.5; P = 0.8 when restricted to F508del homozygotes).

Type 2 diabetes risk alleles in CDKAL1, CDKN2A/B, and IGF2BP2 modify risk for CFRD. In the second part of our analysis, we tested whether specific type 2 diabetes candidate SNPs might associate with CFRD but, because of insufficient power, did not reach genome-wide significance in the genome-wide association studies. First, we found the previously identified association of TCF7L2 SNP rs7903146 to be replicated in 2,031 individuals unique to this study (288 TSS and 740 GMS subjects contributed to the earlier analysis) (11). Each “T” allele of rs7903146 increased risk of CFRD in the CGS subjects (n = 1,508; HR, 1.38; 95% CI, 1.1–1.7; P = 0.004) and in the GMS subjects unique to this study (n = 523; HR, 1.34; 95% CI, 1.08–1.66; P = 0.01). The combined evidence associates the TCF7L2 SNP with CFRD onset (n = 3,059; HR, 1.31; 95% CI, 1.2–1.5; P = 3.8 × 10^{-5}; Table 3).

More than 40 susceptibility genes harboring common type 2 diabetes SNPs have been identified in studies involving >100,000 cases and controls (10). The earliest genome-wide association studies, which were of similar sample size as this study, detected eight loci that were subsequently replicated. Hypothesizing that the effect size for CFRD might be similar to that of type 2 diabetes, we selected the 12 common SNPs in the following 7 loci other than FTO (see RESEARCH DESIGN AND METHODS): CDKAL1; HHEX-IDE; CDKN2A/B; IGF2BP2; SLC30A8; KCNJ11; and PPARG (Table 3). Study-wide significant association with CFRD was demonstrated for SNPs at CDKN2A/B (rs1412829: P = 5.1 × 10^{-5}) and CDKAL1 loci (rs7754840: P = 1.6 × 10^{-3}; rs7750892: P = 1.9 × 10^{-4}). For IGF2BP2, association reached study-wide significance for one of two SNPs tested (rs1470579: P = 4.2 × 10^{-5}). For every associated SNP, the same allele associated with increased risk of type 2 diabetes and CFRD onset, i.e., the direction of effect was the same.

Combined magnitude of effect for CFRD risk alleles. The five detected modifier loci each affected CFRD risk (HR) by 1.2–1.4-fold, and each accounted for 0.5–3% of the total variance in CFRD onset (Supplementary Table 8) and together accounted for 8.3% of the total variance or ~8–21% of the heritability (7). An alternative measure, the population-attributable risk (essentially, the fraction of CFRD cases that would not have occurred if no risk alleles were present), ranged from 11 to 32% for individual SNPs and was 68% for the 5 SNPs together (Supplementary Table 8). There was no detected interaction by Cox regression (P > 0.05). To illustrate the combined effect of these five loci, a risk score was calculated as the total number of high-risk alleles. When stratified by five-SNP risk score, the CFRD prevalence ranged from 11% in those with zero or one risk alleles to 40% in those with eight or nine risk alleles (Fig. 5). No individual had all 10 possible risk alleles.

SLC26A9 SNPs and type 2 diabetes risk in the general population. To test whether there was a detectable effect of the SLC26A9 SNPs on type 2 diabetes risk in the general population, published results from the DIAGRAM consortium meta-analysis were obtained (10). The DIAGRAMv3
stage 1 analysis included 9,580 type 2 diabetic cases and 53,810 controls, all of European descent, typed for the SLC26A9 SNPs. Both SNPs showed evidence for association with type 2 diabetes (odds ratio, 1.06; $P = 0.003$; Supplementary Table 9). Interestingly, in contrast to what was found for TCF7L2 and other type 2 diabetes SNPs, the alleles of rs4077468 and rs4077469 associated with CFRD were associated with decreased risk of type 2 diabetes.

**DISCUSSION**

Using genome-wide association and candidate-based approaches, four novel genetic modifiers of CFRD were identified. One modifier locus (SLC26A9) was identified without an a priori hypothesis but has previous evidence for a role in CFTR biology. The discovery of CFRD susceptibility SNPs at SLC26A9 suggests a heretofore unsuspected role for alternate ion conduction pathways that may be dependent or independent of CFTR. SNPs at the

---

**TABLE 3**

Genotyped SNPs in TCF7L2 and seven additional type 2 diabetes susceptibility loci tested for association with CFRD in the discovery sample

| Locus       | SNP               | T2D RA/non-RA | HR*   | $I^2$ | Two-sided $P$ | One-sided $P$ |
|-------------|------------------|---------------|-------|-------|---------------|---------------|
| TCF7L2      | rs7901695        | G/A           | 1.34  | 0     | $6.0 \times 10^{-7}$ | $3.0 \times 10^{-7}$ |
|             | rs7903146        | T/C           | 1.31  | 0     | $3.8 \times 10^{-6}$ | $1.9 \times 10^{-6}$ |
| KCNJ11      | rs5219           | T/C           | 0.95  | 0     | 0.35          | 0.82          |
|             | rs5215           | G/A           | 0.94  | 0     | 0.28          | 0.86          |
| CDKAL1      | rs7754840        | C/G           | 1.20  | 55    | $7.9 \times 10^{-4}$ | $9.7 \times 10^{-5}$ |
|             | rs7756992        | C/T           | 1.26  | 45    | $7.9 \times 10^{-4}$ | $9.7 \times 10^{-5}$ |
| HHEX / KIF-1 / IDE | rs1111875 | G/A         | 0.96  | 7     | 0.44          | 0.78          |
|             | rs5015480        | C/T           | 0.93  | 9     | 0.20          | 0.90          |
|             | rs7923387        | G/A           | 0.97  | 19    | 0.59          | 0.70          |
| SLC30A8     | rs13266634       | C/T           | 1.08  | 0     | 0.19          | 0.094         |
| CDKN2A/B    | rs1412829        | T/C           | 1.26  | 0     | $5.1 \times 10^{-3}$ | $2.5 \times 10^{-3}$ |
| IGF2BP2     | rs4402960        | A/C           | 1.14  | 0     | $2.2 \times 10^{-2}$ | 1.1 $\times 10^{-2}$ |
|             | rs1470579        | G/T           | 1.18  | 0     | $4.2 \times 10^{-3}$ | $2.1 \times 10^{-3}$ |
| PPARG       | rs1801282 (imputed§) | C/G | 1.12  | 0     | 0.20          | 0.10          |

---

$n = 3,059$. Association with CFRD onset was tested using a Cox proportional hazards model in each of the CGS, TSS, and GMS datasets and combined by fixed-effects meta-analysis. For SNPs with evidence of heterogeneity across studies ($I^2 \geq 25\%$), a shared frailty model was used with essentially identical results (not shown). HR is shown for CFRD onset for each T2D risk allele. RA, risk allele; T2D, type 2 diabetes. *HR >1 denotes the same allele associated with increased risk of T2D and CFRD. †Study-wide $P < 0.05$. ‡Study-wide $P < 0.01$ (Bonferroni correction for $n = 12$ SNPs at loci other than TCF7L2). §See [RESEARCH DESIGN AND METHODS](#). Imputed SNP rs1801282 had minor allele frequency of 0.125; MACH quality score ($R^2$) = 0.98.
SLC26A9 could modify the cellular phenotype of CFRD in several ways. In tissues expressing both SLC26A9 and CFTR, SLC26A9 could act as an alternative conduction pathway for chloride or bicarbonate, thereby modifying the loss of CFTR function. When expressed in the same cells, SLC26A9 may interact with mutant CFTR, leading to stabilization and moderation of the phenotype (30). Although F508del-CFTR has little residual function, a small increase in CFTR function may be sufficient to alter the ion conduction profile in affected epithelia. Either of these mechanisms also could account for the same SLC26A9 SNPs also modifying the risk of MI (16). Tissues expressing both proteins include lung, stomach, and small intestine, although cellular colocalization has not been demonstrated. Lack of association of SLC26A9 SNPs with lung phenotypes suggests that effects of these SNPs could be either tissue-specific or could be most important when there is little residual CFTR function, but this study does rule out the possibility that greater changes in SLC26A9 activity could affect other complications of CF.

Another possibility is that SLC26A9 activity itself could play a role in glucose metabolism (such as by modulating insulin secretion); in that case, SNPs that affect the expression of SLC26A9 could modify diabetes risk. SLC26A9 could be of greater importance in CFRD (compared with type 2 diabetes) if this pathway is already perturbed by interaction with F508del-CFTR (27,31). This scenario is supported by the association of SLC26A9 SNPs with type 2 diabetes (Supplementary Table 9), even though the opposite alleles conferred increased risk. Conceivably, the same alterations in SLC26A9 activity could normalize or exacerbate ion transport abnormalities depending on whether CFTR is functional. A similar paradoxical association was seen for opposite alleles of SNPs in TGFβ1 associating with lung disease in populations with CF and chronic obstructive pulmonary disease (14).

A pertinent related question is whether these genes are expressed in pancreatic β-cells, which could support the islet as a site of action (32). Reports of CFTR expression in β-cells are inconsistent (33,34), and SLC26A9 is expressed in multiple tissues, including bronchial epithelium (35,36), but expression in β-cells was not reported. Conductance of chloride (37) correlates with insulin release from β-cells, as illustrated by inhibition or ablation of SLC12A1 (Na⁺K⁺Cl⁻ cotransporter 1) (38–40). Thus, it is not unreasonable to speculate that alteration in the expression of CFTR or SLC26A9 (or both) might affect β-cell function directly by altering chloride or bicarbonate flow (or both), a hypothesis that is attractive because it accommodates all of the postulated mechanisms. Altered insulin secretory dynamics in a ferret CF model (41) suggest that CFTR may play a role in β-cell function.

Three SNPs in the 3′ UTR of SLC26A9 associate with decreased reporter transcription in A549 cells (42). Those SNPs did not show evidence of association with CFRD in this study (rs12031234: genotyped; minor allele frequency, 5.1%; HR, 1.05; 95% CI, 0.89–1.03; P = 0.6; rs2282429: imputed with R² = 0.6; and rs2282430: imputed with R² = 0.48, same results as rs2282429). These SNPs may have too little effect to impact CFRD risk or may not affect SLC26A9 transcription in cell types pertinent for CFRD.

Gene variants in four loci (TFC7L2, CDKAL1, CDKN2A/B, and IGF2BP2) associate with both type 2 diabetes and CFRD. Loci that contribute to both diseases support the concept that diabetes develops in individuals who may have underlying susceptibility to β-cell dysfunction (43).
in CDKAL1 are reported to impair proinsulin translation and to stimulate the endoplasmic reticulum stress response (44) that promotes apoptosis (45). Finding CDKAL1 as a CFRD modifier suggests these pathways are important in CFRD as well, such as if endoplasmic reticulum stress is induced by CFTR misfolding (46). Involvement of TCF7L2 (which may affect β-cell mass and proinsulin processing) (47,48) and CDKN2A/B (tumor suppressors with reported roles in both cellular senescence and insulin secretion) (49,50) suggest roles for growth/apoptosis and insulin processing in CFRD. Finally, although association with the other selected type 2 diabetes SNPs was not detected, effect sizes comparable with what are reported for type 2 diabetes could not be ruled out because of limited study power. Although CFRD is a disease distinct from type 2 diabetes, the shared genetic architecture highlights disease pathways that are similar.

The five CFRD modifier loci accounted for 8.3% of the variation in CFRD onset, with CFRD prevalence varying from 10 to 40% as a function of the five-SNP risk score (Fig. 5). Although not directly comparable because of differences in study design, it is notable that type 2 diabetes SNPs accounted for 10% of the variance in type 2 diabetes risk (51), with odds ratio varying by 1.5-fold (52) to 4-fold (51) as a function of a 39-SNP risk score. This may reflect greater disease heterogeneity in type 2 diabetes compared with CFRD and suggests that further study of CFRD is useful even with comparatively smaller sample sizes. However, CFRD currently cannot be predicted accurately from genes alone. That said, even an imperfect prediction may be of use in the care of people with CF, such as by prompting earlier CFRD screening in high-risk individuals.

A risk of genome-wide association studies is that effect sizes may be inflated because of the “winner’s curse” (arising from the large number of SNPs tested) and fail to replicate in subsequent studies. However, this was not the case for rs4077468 (SLC26A9), because the per-allele effect size was greater in the replication (HR, 1.47) than in the discovery (HR, 1.38) sample. The effect of winner’s curse in the type 2 diabetes candidate analysis should be much lower, because only 12 SNPs were considered. The five-SNP risk score assumes no interaction between modifier alleles, approximates each high-risk variant as having an equal effect size, and should be validated in a separate cohort. The loci with suggestive association in the discovery sample need further study before concluding there is association with CFRD. Many of the nondiabetic individuals are expected to develop CFRD in the future; therefore, future analyses of this study cohort, including updated CFRD information, will have increased statistical power to detect association.

In summary, genetic variation in SLC26A9, which encodes a bicarbonate and chloride transport protein that interacts with CFTR, is associated with CFRD, possibly through CFTR-dependent mechanisms. Greater understanding of the processes involved may lead to novel treatments for this highly prevalent complication of CF. Our understanding of the disease mechanism also is enhanced by the demonstration that three additional type 2 diabetes loci affect CFRD risk, highlighting disease pathways that are shared between these two diseases. People with CF are sensitized to development of diabetes at a young age and at a high rate. As such, CF could provide valuable insight into pathologic mechanisms operating before the onset of overt disease, during which time preventative intervention might be possible.

ACKNOWLEDGMENTS
This work was supported in part by the National Institutes of Health (K23 DK076446 to S.M.B., R01 HL068927 to G.R.C., R01 HL068890, R01 DK063638, and R01 HL065396 to M.R.K., and HG0004314 to L.J.S.), United States Cystic Fibrosis Foundation (CUTTIN06P0, R025-CR07, KNOWLE00A0, R026-CR07, KNOWLLE11P0, STONEB12, and DRUMMA00), and Pediatric Endocrine Society (S.M.B.). This work was supported in part by Genome Canada through the Ontario Genomics Institute (as per research agreement 2004-OGI-3-05 with the Ontario Research Fund), Research Excellence Program, the Ontario Ministry of Research and Innovation Early Researcher (L.J.S.), Cystic Fibrosis Canada grants (to P.R.D. and L.J.S.), Natural Sciences and Engineering Research Council of Canada (NSERC) (to L.J.S.), Canadian Institutes of Health Research (CIHR) 119556 (to L.J.S.), NSERC (250053-2008), and CHIR (MOP 84287) grants (to L.S.). Funding for genome-wide genotyping was provided by the United States Cystic Fibrosis Foundation.

No potential conflicts of interest relevant to this article were reported.

S.M.B. designed the study, researched data, performed analyses, wrote the manuscript, and performed the replication analysis. C.W.C., C.W., and K.M.A. researched data and performed analyses. L.J.S. designed the replication study, contributed to discussions, and reviewed and edited the manuscript. J.R.S. researched data, contributed to discussions, and reviewed and edited the manuscript. F.A.W. contributed to discussions and reviewed and edited the manuscript. P.R.D. designed the replication study, genotyped the replication samples, contributed to discussions, and reviewed and edited the manuscript. M.L.D. and M.R.K. contributed to discussions and reviewed and edited the manuscript. S.A.N. researched data and contributed to discussions. G.R.C. researched data, contributed to discussions, and reviewed and edited the manuscript. M.L.D. and M.R.K. contributed to discussions and reviewed and edited the manuscript. G.R.C. contributed to study design, discussions, and writing and editing of the manuscript. S.M.B. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Portions of this work were published as abstracts at the North American Cystic Fibrosis Conference, Anaheim, California, 3–5 November 2011; the North American Cystic Fibrosis Conference, Orlando, Florida, 11–13 October 2012; and ENDO 2012, The Endocrine Society’s 94th Annual Meeting and Expo, Houston, Texas, 13–16 June 2012.

The authors are grateful for the participation of the many CF patients, families, research coordinators, and clinicians in the Cystic Fibrosis Twin and Sibling Study, the Genetic Modifiers of Cystic Fibrosis Study, and the Canadian Consortium for Cystic Fibrosis Genetic Studies. The authors thank the DIAGRAM consortium (including Dr. Mark McCarthy and Dr. Andrew Morris, both at the Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K.) for sharing results for SNPs of interest.

REFERENCES
1. Moran A, Becker D, Casella SJ, et al; CFRD Consensus Conference Committee. Epidemiology, pathophysiology, and prognostic implications of cystic fibrosis-related diabetes: a technical review. Diabetes Care 2010; 33:2677–2683
17. Blackman SM, Deering-Brose R, McWilliams R, et al. Relative contribution of genetic and environmental factors to lung disease severity in cystic fibrosis. Cystic Fibrosis and Related Disorders Research Network. Pediatrics 2011;129:S63-69.

19. Abecasis GR, Altshuler D, Auton A, et al.; 1000 Genomes Project Consortium. Large-scale association analysis provides insights into the genetic architecture of type 2 diabetes. Nat Genet 2012;44:562-569.

22. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994;81:515-526.

24. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-575.

25. Gautlon KJ, Nammo T, Passquale L, et al. A map of open chromatin in human pancreatic islets. Nat Genet 2010;42:255-259.

26. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol 2011;9:e1001046.