SUPPLEMENTARY DATA

SUPPLEMENTARY TABLE 1. Female sex and liver disease correlated with CFRD onset by univariate (A) and multivariate (B) Cox regression analyses.

(A) Univariate analysis	**Study**	**HR**	**n**	**P-value**
Female sex*	CGS-D	1.78	1508	2.2×10^{-4}
	GMS-D	1.49	1267	1.5×10^{-4}
	TSS-D	1.21	379	0.28
	GMS-R	0.95	409	0.79
	CGS-R	2.15	285	0.11
	Combined	1.42	3848	9.3×10^{-7}
Liver disease*	CGS-D	2.86	1504	1.7×10^{-5}
	GMS-D	2.87	1267	2.8×10^{-11}
	TSS-D	1.87	361	0.017
	GMS-R	5.04	408	2.8×10^{-6}
	CGS-R	1.26	285	0.709
	Combined	2.75	3825	2.3×10^{-22}
F508del homozygosity	CGS-D	1.25	1505	0.16
	GMS-D	0.23†	1267	2.59×10^{-10}
	TSS-D	0.93	377	0.68
	GMS-R	1.25	409	0.03
	CGS-R	0.82	285	0.38
	Combined	0.99	3843	0.84
(B) Multivariate model	**Study**	**HR**	**n**	**P-value**
Female sex	CGS-D	1.82	1504	1.4×10^{-4}
	GMS-D	1.50	1267	9.0×10^{-5}
	TSS-D	1.17	361	0.41
	GMS-R	0.91	408	0.64
	CGS-R	2.15	285	0.11
	Combined	1.42	3825	1.0×10^{-6}
Liver disease	CGS-D	3.11		3.8×10^{-6}
	GMS-D	2.91		1.4×10^{-11}
	TSS-D	1.88		0.017
	GMS-R	5.09		2.5×10^{-6}
	CGS-R	1.27		0.705
	Combined	2.82		4.4×10^{-23}

SUPPLEMENTARY DATA

*HR for female sex and liver disease varied across the 5 study groups analyzed (each with I^2 >0.9). Higher HR for liver disease in the GMS is attributed to the stringent definition of liver disease used in the GMS. †The apparently protective effect of F508del homozygosity in the GMS study (HR=0.23) is due to study design; essentially all non-F508del-homozygous individuals have liver disease with cirrhosis which is a risk factor for CFRD. Abbreviations: "-D" and "-R" denote Discovery and Replication sample subsets, respectively.

SUPPLEMENTARY TABLE 2. Detailed analysis results for SNP rs4077468 in the Discovery sample.

	CGS	GMS	TSS	Combined
N (total)	1508	1263	288	3059
N (CFRD)	167	374	103	644
HR	1.36	1.38	1.40	1.38
P	6.16×10^{-3}	2.75×10^{-5}	2.31×10^{-2}	3.59×10^{-8}
Cochran Q	0.011	1.1×10^{-5}	0.017	0.028
P (Cochran)				0.99
I^2				0

Abbreviations: HR, hazard ratio; see Methods for definitions of Cochran Q, Cochran P, and I^2.

SUPPLEMENTARY TABLE 3. Reanalyses of SNPs in Table 2 when restricting to individuals homozygous for the F508del *CFTR* mutation (n=2303 total; n=516 with CFRD events) in the Discovery sample. Shown are the primary analysis with PCs as covariates (unadjusted) and the secondary analysis also including adjustment for female sex and liver disease ("adj").

SNP	Chr	NCBI36 position (Mb)	Risk/Other Allele	HR	P-value	HR adj	P adj	Annotation
rs7415921	1	204.2	G/T	1.36	$9.02\times10^{-7*}$	1.39	$4.24\times10^{-7*}$	SLC26A9
rs4077468†	1	204.2	T/C	1.36	1.90×10^{-6}	1.38	$1.07\times10^{-6*}$	SLC26A9
rs4077469†	1	204.2	G/A	1.36	1.90×10^{-6}	1.38	$1.07\times10^{-6*}$	SLC26A9
rs7419153	1	204.2	A/G	1.35	4.08×10^{-6}	1.37	2.74×10^{-6}	SLC26A9
rs7512462	1	204.2	T/C	1.35	4.21×10^{-6}	1.38	$1.61\times10^{-6*}$	SLC26A9
rs7555534	1	204.2	C/T	1.33	7.88×10^{-6}	1.38	$1.01\times10^{-6*}$	SLC26A9
rs1874361	1	204.2	A/C	1.29	4.06×10^{-5}	1.34	3.55×10^{-6}	SLC26A9
rs6981918	8	144.0	T/G	2.65	4.36×10^{-5}	2.79	1.69×10^{-5}	CYP11B2
rs995447	4	62.5	G/A	1.55	2.23×10^{-4}	1.74	4.66×10^{-6}	LPHN3
rs11902125	2	65.7	T/C	1.56	2.36×10^{-4}	1.62	8.23×10^{-5}	KRT18P33
rs4759088	12	53.2	T/C	1.26	6.39×10^{-4}	1.28	3.23×10^{-4}	NCKAP1L

*Study-wide suggestive

†These SNPs are in 100% linkage disequilibrium in the Discovery Sample

SUPPLEMENTARY DATA

SUPPLEMENTARY TABLE 4. Imputed SNPs within the *SLC26A9* locus associated with CFRD at a suggestive ($P<1.8\times10^{-6}$) or significant ($P<9.1\times10^{-8}$) level in the Discovery sample (n=3059)

SNP	Chr	NCBI36 Position (bp)	Risk/Other Allele	RAF	R^2	HR	P-value
rs4951271	1	204,180,471	A/G	0.42	0.98	1.39	$1.73\times10^{-8*}$
rs1342063	1	204,179,482	C/T	0.42	0.97	1.39	$2.96\times10^{-8*}$
rs1342064	1	204,179,696	T/C	0.41	0.97	1.38	$3.27\times10^{-8*}$
rs2036100	1	204,174,495	C/G	0.40	0.93	1.38	$7.40\times10^{-8*}$
rs61814953	1	204,176,703	T/C	0.40	0.94	1.38	$8.04\times10^{-8*}$
rs1342061	1	205,911,385	T/A	0.58	0.91	1.37	1.09×10^{-7}
rs6661355	1	205,910,631	C/T	0.57	0.90	1.36	1.48×10^{-7}
rs6673820	1	205,910,604	A/G	0.56	0.91	1.35	2.41×10^{-7}
rs7521316	1	205,907,796	T/C	0.67	0.87	1.36	5.34×10^{-7}
rs4951030	1	205,913,747	T/A	0.50	0.96	1.33	5.56×10^{-7}
rs6593976	1	205,909,285	A/C	0.53	0.92	1.34	6.16×10^{-7}
rs6593975	1	205,909,276	C/T	0.53	0.93	1.34	6.57×10^{-7}
rs11240598	1	205,908,630	C/G	0.65	0.88	1.34	1.28×10^{-6}
rs7418300	1	205,918,020	G/C	0.72	0.91	1.33	1.65×10^{-6}
rs1342062	1	205,912,786	G/T	0.63	0.76	1.40	1.70×10^{-6}

1567 SNPs located within 1Mb of SLC26A9 were imputed with reasonable quality (MACH quality score $R^2>0.3$) using phased haplotypes from the 1000 Genomes project (19). HR and P-values for association with CFRD onset are shown. *Study-wide significant; Abbreviations: RAF, risk allele frequency; HR, per-allele hazard ratio.

SUPPLEMENTARY DATA

SUPPLEMENTARY TABLE 5. Analysis of rs4077468 stratifying by and adjusting for meconium ileus (MI) in the Discovery sample.

	CGS	**GMS**	**TSS**	**Combined**
n (no MI)	1251	991	224	2466
n (MI)	238	198	48	484
n (total with MI information)	1489	1189	272	2950
CFRD analysis stratified by MI				
SNP HR (no MI)	1.36	1.36	1.40	1.37
SNP P-value (no MI)	1.13×10^{-2}	3.55×10^{-4}	3.14×10^{-2}	1.08×10^{-6}
SNP HR (MI)	1.75	1.53	1.58	1.58
SNP P-value (MI)	0.14	3.82×10^{-2}	0.378	7.48×10^{-3}
SNP HR (stratified)	1.39	1.39	1.41	1.39
SNP P-value (stratified)	4.20×10^{-3}	4.20×10^{-5}	2.06×10^{-2}	3.66×10^{-8}
CFRD analysis (covariate: MI)				
SNP HR	1.37	1.38	1.41	1.38
SNP P-value	6.00×10^{-3}	4.18×10^{-5}	2.10×10^{-2}	5.23×10^{-8}
MI HR	1.09	1.57	0.82	1.33
MI P-value	0.742	1.67×10^{-3}	0.516	1.37×10^{-2}
CFRD analysis (covariate: MI, SNP*MI)				
SNP HR	1.36	1.36	1.38	1.36
SNP P-value	1.12×10^{-2}	3.84×10^{-4}	3.85×10^{-2}	1.37×10^{-6}
MI HR	1.12	1.67	0.98	1.45
MI P-value	0.759	7.09×10^{-3}	0.969	1.79×10^{-2}
SNP*MI HR	0.97	0.90	0.77	0.90
SNP*MI P-value	0.922	0.634	0.618	0.552

Association of SNP with CFRD onset was analyzed using Cox proportional hazard analysis while adjusting for principal components. Analyses of separate MI strata were combined by fixed-effects meta-analysis (see Methods). Analysis for interaction included adjustment for MI and a linear interaction term.

SUPPLEMENTARY DATA

SUPPLEMENTARY TABLE 6. Analysis of rs4077468 stratifying by and adjusting for liver disease (LD) in the Discovery sample.

	CGS	GMS	TSS	Combined
n (no liver disease)	1432	1133	260	2825
n (with liver disease)	72	130	18	220
n (total with LD information)	1504	1263	278	3045
CFRD analysis stratified by liver disease				
SNP HR (no LD)	1.37	1.34	1.59	1.38
SNP P-value (no LD)	9.57×10^{-3}	2.45×10^{-4}	6.28×10^{-3}	2.25×10^{-7}
SNP HR (LD)	1.01	1.97	0.37	1.38
SNP P-value (LD)	0.977	4.81×10^{-3}	8.56×10^{-2}	9.06×10^{-2}
SNP HR (stratified)	1.33	1.40	1.42	1.38
SNP P-value (stratified)	1.34×10^{-2}	1.22×10^{-5}	3.21×10^{-2}	5.11×10^{-8}
CFRD analysis (covariate: liver disease)				
SNP HR	1.34	1.39	1.44	1.38
SNP P-value	9.13×10^{-3}	1.66×10^{-5}	2.19×10^{-2}	3.41×10^{-8}
Liver disease HR	2.58	3.02	1.69	2.66
Liver disease P-value	1.47×10^{-4}	5.70×10^{-12}	0.1	3.77×10^{-15}
CFRD analysis (covariate: liver disease, SNP*liver disease)				
SNP HR	1.36	1.33	1.57	1.37
SNP P-value	1.04×10^{-2}	3.58×10^{-4}	7.75×10^{-3}	4.19×10^{-7}
Liver disease HR	2.39	3.94	0.95	2.93
Liver disease P-value	1.20×10^{-2}	1.82×10^{-9}	0.931	1.98×10^{-9}
SNP*Liver HR	1.12	0.68	1.98	0.95
SNP*Liver P-value	0.747	0.13	0.125	0.763

Association of SNP with CFRD onset was analyzed using Cox proportional hazard analysis while adjusting for principal components. Analyses of separate liver disease strata were combined by fixed-effects meta-analysis (see Methods). Analysis for interaction included adjustment for liver disease and a linear interaction term.

SUPPLEMENTARY DATA

SUPPLEMENTARY TABLE 7. Analysis of rs4077468 stratifying by and adjusting for male/female sex.

	CGS	**GMS**	**TSS**	**Combined**
n (male)	825	675	148	1648
n (female)	683	588	140	1411
n (total)	1508	1263	288	3059
CFRD analysis stratified by female sex				
SNP HR (male)	1.64	1.45	1.04	1.43
SNP P-value (male)	3.55×10^{-3}	6.07×10^{-4}	0.876	2.35×10^{-5}
SNP HR (female)	1.20	1.28	1.48	1.28
SNP P-value (female)	0.231	2.34×10^{-2}	7.85×10^{-2}	2.63×10^{-3}
SNP HR (stratified)	1.37	1.36	1.26	1.35
SNP P-value (stratified)	4.64×10^{-3}	5.60×10^{-5}	0.161	3.42×10^{-7}
CFRD analysis (covariate: female sex)				
SNP HR	1.36	1.37	1.42	1.37
SNP P-value	6.47×10^{-3}	3.90×10^{-5}	1.80×10^{-2}	4.65×10^{-8}
female sex HR	1.84	1.46	1.30	1.52
female sex P-value	1.14×10^{-4}	2.99×10^{-4}	0.195	1.58×10^{-7}
CFRD analysis (covariate: female sex, SNP*female sex				
SNP HR	1.51	1.47	1.44	1.48
SNP P-value	1.45×10^{-2}	2.83×10^{-4}	0.105	3.10×10^{-6}
female sex HR	1.58	1.31	1.28	1.36
female sex P-value	5.46×10^{-2}	6.95×10^{-2}	0.416	7.49×10^{-3}
SNP*female HR	1.21	1.17	1.03	1.16
SNP*female P-value	0.396	0.309	0.929	0.208

Association of SNP with CFRD onset was analyzed using Cox proportional hazard analysis while adjusting for principal components. Analyses of separate male/female strata were combined by fixed-effects meta-analysis (see Methods). Analysis for interaction included adjustment for female sex and a linear interaction term.

SUPPLEMENTARY DATA

SUPPLEMENTARY TABLE 8. CFRD Risk (PAR, % of total variance) estimated for each genotyped SNP, all 5 SNPs considered together, and the 5-SNP risk score within the Discovery sample.

Annotation	SNP	Risk Allele Freq	HR	PAR	% of total variance
SLC26A9	rs4077468	0.58	1.39	0.32	3.2%
TCF7L2	rs7901695	0.32	1.31	0.18	2.0%
CDKAL1	rs7756992	0.28	1.25	0.13	1.2%
CDKN2A/B	rs1412829	0.59	1.26	0.25	1.4%
IGF2BP2	rs1470579	0.32	1.15	0.11	0.5%

Multi-SNP and non-SNP variables	PAR	% of total variance
5 SNPs (multivariate)	0.68	8.3%
5-SNP risk score	N/A	7.6%
Female sex	0.18	2.5%
Liver disease	0.15	5.7%
Female sex + liver disease	0.31	8.2%
5 SNPs + sex, liver disease	0.78	15.8%

Abbreviations: HR, hazard ratio, PAR, population attributable risk

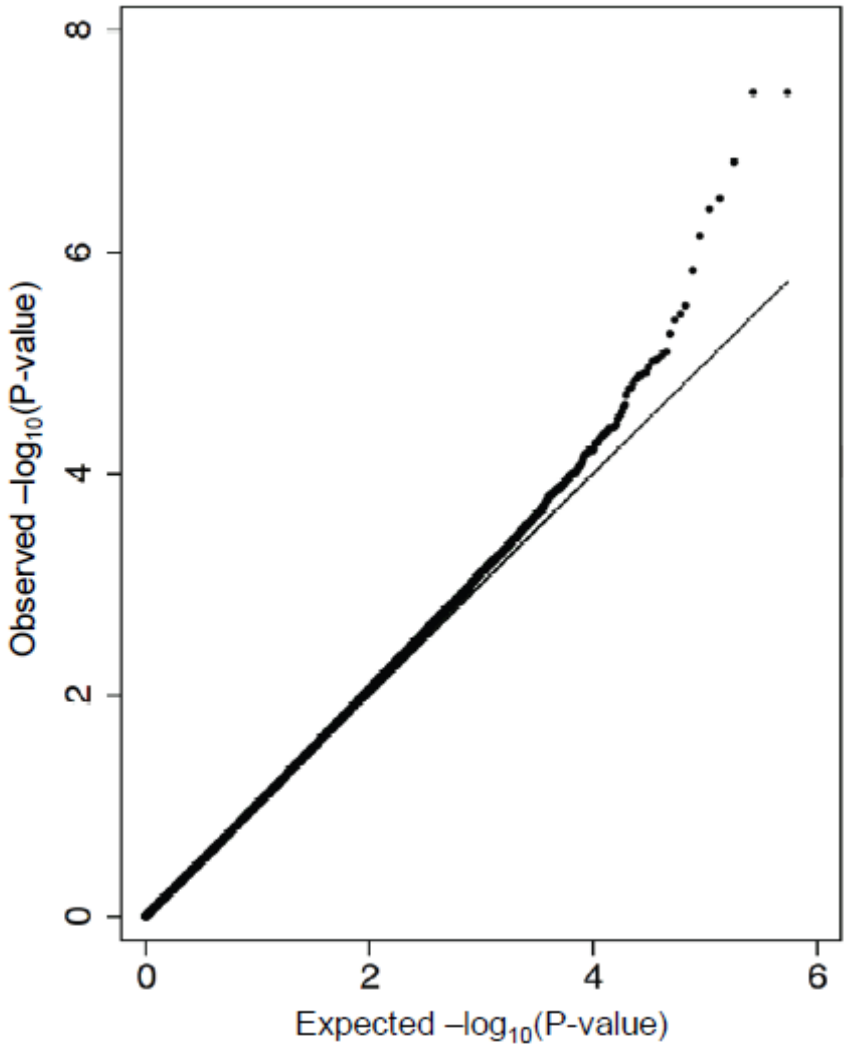
SUPPLEMENTARY TABLE 9. *SLC26A9* SNPs in DIAGRAMv3 study*

SNP	Risk Allele	Other Allele	P-value	OR	95% CI	Ncase	Ncontrol
rs4077468	G	A	3.1×10^{-3}	1.06	1.02 – 1.10	9,580	53,810
rs4077469	T	C	3.2×10^{-3}	1.06	1.02 – 1.10	9,580	53,810

*The DIAGRAM stage 1 analysis (10) included 12,171 type 2 diabetes cases and 56,862 controls, all of European descent. Genotypes for rs4077468 and rs4077469 were available for 78% of cases and 95% of controls (see above; genotyped on 5 and imputed on 7 of the 12 genotyping platforms). Results from the 12 study-specific analyses were combined in a fixed-effects model. Abbreviations: OR, odds ratio; N_{case}, number of cases, and $N_{control}$, number of controls.

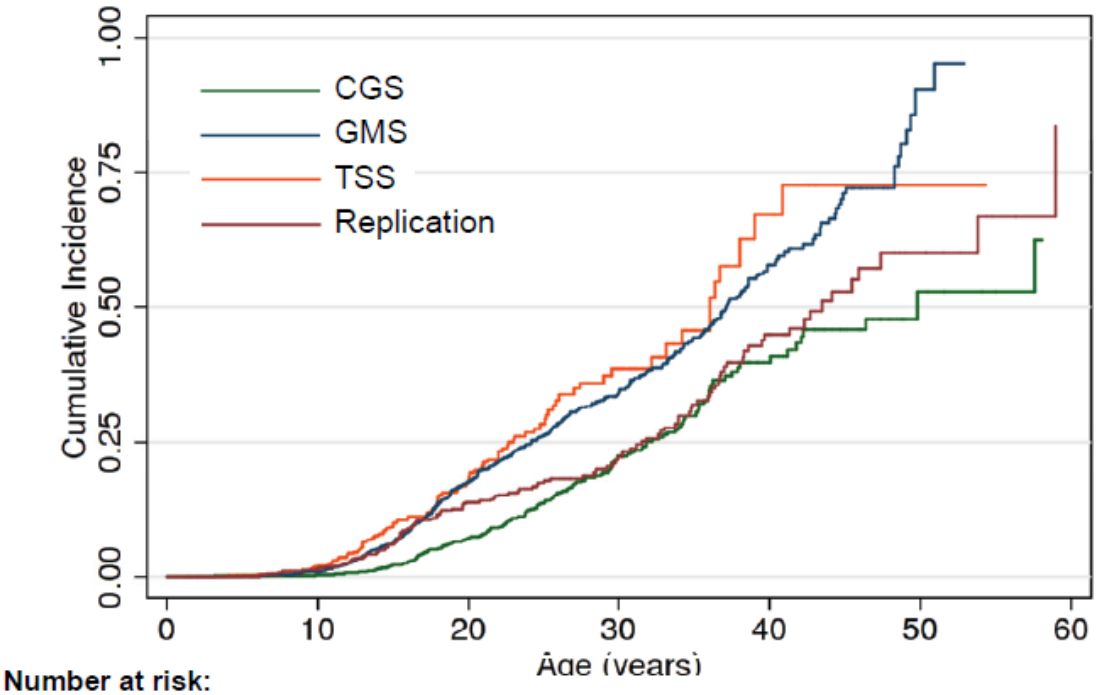
SUPPLEMENTARY DATA

Supplementary Figure 1. Q-Q plot (expected vs. observed –log10 P-values) for the Discovery sample meta-analysis. CFRD onset was analyzed for correlation with each SNP (-log10 P-value plotted as black dots) within each of the 3 component studies (CGS, TSS, and GMS) using Cox proportional hazards (covariates = PCs) and combined using inverse variance (fixed-effects) meta-analysis (see Methods).



Supplementary Figure 2. Cumulative incidence of CFRD for individuals in the Discovery sample (CGS, GMS, TSS shown separately), and Replication samples. The age-dependent incidence rates of CFRD in TSS (orange) and GMS (blue) subsets of the Discovery sample did not differ significantly (P=0.3, log rank). CFRD onset in the CGS (green) differed from TSS and GMS subsets of the Discovery sample ($P=3\times10^{-286}$, log rank). CFRD onset in the CGS and GMS subsets of the Replication sample did not differ from each other (P=0.93, log rank), but CFRD onset in the Replication sample overall (red) was differed from and was intermediate between CGS and TSS/GMS (P<0.05, log rank).

SUPPLEMENTARY DATA



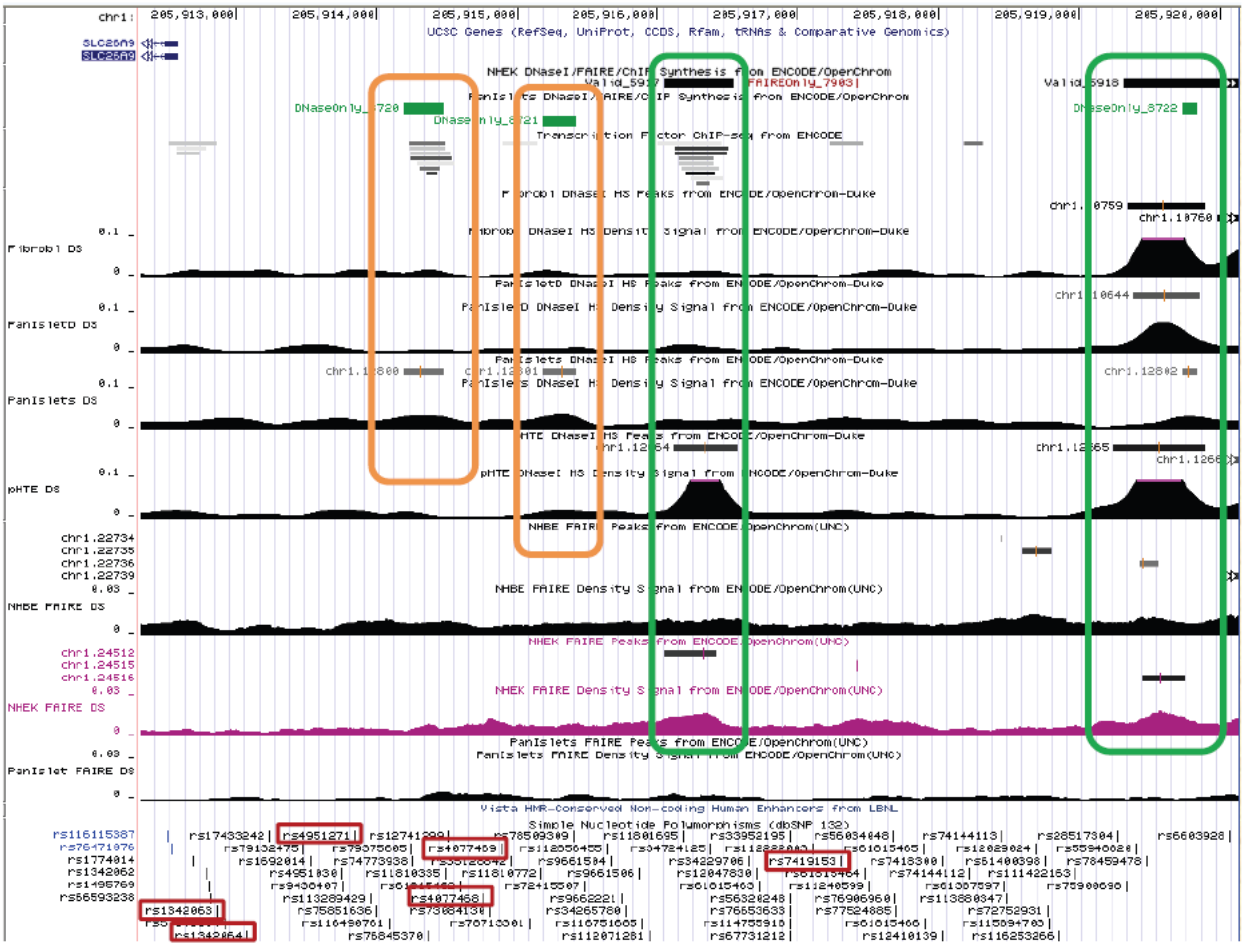
Number at risk:

CGS:	1508	1256	610	228	56	9	3
TSS:	786	629	210	42	7	1	0
GMS:	1263	1212	638	283	78	2	0
Repl.:	694	491	333	137	50	12	1

Supplementary Figure 3. Open chromatin in multiple cell lines around *SLC26A9* from ENCODE (http://genome.ucsc.edu/ENCODE/index.html). Green boxes represent regions of tissue non-specific open chromatin. Blue boxes represent regions of open chromatin specific to *SLC26A9* expressing tissues. Orange boxes show regions of open chromatin specific to pancreatic islet cells. The location of SNPs shown in Figure 4 that are associated with CFRD onset are outlined in red boxes. A. SNPs located in the *SLC26A9* 5′ region. B. SNPs located in the first intron. Fibrobl=fibroblasts, PanIsletD=dedifferentiated pancreatic islets, PanIslets=pancreatic islet cells, pHTE=Human tracheal epithelium, NHBE=normal human bronchial epithelium, NHEK=normal human epidermal keratinocyte.

SUPPLEMENTARY DATA

A. 5′ Region of *SLC26A9*



SUPPLEMENTARY DATA

B. *SLC26A9* Intron 1