Supplemental Material and Methods

Regulatory Profile Conservation Analysis of the region encompassing of *SLC26A9*

To explore regulatory profile of the region 5’ and within intron 1 of *SLC26A9*, we used the Open Regulatory Annotation database (ORegAnno) track on the UCSC genome browser (https://www.genome.ucsc.edu), which contains curated regulatory annotation including transcription factor general binding sites derived from experimental data (41). We also evaluated conservation in this region using the Vertebrate Multiz Alignment & Conservation track.

Single-cell RNA-sequencing of pancreatic cells

Preparation of single cells: Human pancreatic material not used for islet autotransplantation was immediately dissociated into single cells by enzymatic digestion by incubation with Accumax (Invitrogen). Cell clumps were removed with the MACS SmartStrainer 30μM. Cells were then prepared according to the 10X Genomics® Cell Preparation Guide for Single Cell Protocols and resuspended in PBS with 0.04% BSA. Cell viability (~80%) and concentration were determined using the Cellometer Auto 2000 Cell Viability Counter. The single cell cDNA library was prepared using droplet-based technology from 10X Genomics®. ~17,400 single cells were immediately loaded into the 10X Genomics® Chromium Controller to prepare gel bead-in-emulsions (GEMs). Single cell libraries were generated according to the 10X Genomics Chromium Single-Cell 3’ v2 protocol. The library was loaded onto an Illumina NextSeq500 with 2x75 cycle paired end sequencing.

Processing of RNA-Seq Reads: Processing of RNA-Seq reads was completed with the Cell Ranger Single Cell Software and pipeline v2.1.1 (http://software.10xgenomics.com/single-cell/overview/welcome). Raw base call files were demultiplexed into FASTQ files. Reads were aligned to GRCh38 supplied by 10X Genomics® using STAR. Cell barcodes and unique molecular identifiers (UMIs) were counted and filtered for barcodes corresponding to a known barcode sequence and for unique RNA molecules. Cells were filtered for those with UMI counts >10% of the 99th percentile, a cut-off identified by Cell Ranger. The Seurat R package (version 2.3.3) (43) was used for further quality control. Genes expressed in fewer than 3 cells and cells expressing fewer than 200 detected genes were filtered out. Cells with greater than 50% mitochondrial expression and >3000 unique gene counts (possible doublets) were also filtered out.

Plasmid construction

Reporter constructs were generated to contain regions of different lengths (5' 4.8kb, 2.3kb, 1.172kb and 1.173kb) corresponding to either high risk (HR) or low risk (LR) haplotypes (Supplemental Figure 3). Inserts were amplified from genomic DNA using specific primers with KOD Hot Start DNA polymerase. With overhangs added, inserts were fused upstream to the firefly luciferase reporter PGL4.10 vector (Promega) using the In-fusion Cloning Kit (Takara) according to manufacturer's instructions. After transformation in Stellar Competent Cells (provided by the In-fusion Cloning kit), plasmids resulting from both Spinsmart^{TM} Plasmid Miniprep DNA Purification Kit (Denville) and Plasmid Maxi Kit (Qiagen) were checked by sequence analyses. As needed, site-directed mutagenesis was used to modify key variants or unwanted changes as a result of subcloning to match the sequence corresponding to haplotypes-of-interest with the Site-Directed Mutagenesis Kit (NEB).

Variant association with gene expression

Pancreas and lung cis-eQTL association statistics of the CFRD-associated variants *(8)* were downloaded from GTEx(v7) (Supplemental Table 4). Directionality of beta value was modified from GTEx. A positive beta value indicates association of the high risk allele instead of the reference allele.

Statistics

The optimized sequence kernel association tests (SKAT-O) were used to check for association between sets of variants and CFRD. Statistical significance after correction for the number of windows used in the analysis was defined as a p-value <0.01/36=2.7E-4. Determining significance of co-expression of transcripts in scRNA-seq data: The hypergeometric test was used to measure the statistical significance of two genes being co-expressed in the same cell given the total number of cells they are expressed in. Significance of co-expression was only calculated if at least 1 cell expressed both genes. P-values <0.05 were considered significant. See methods for details. Dual luciferase-renilla assay: DLRA reading was performed three times for each of the transfection well. Each of the three readings were averaged. An α=0.05 using a Student’s t-test based on a difference between sample means was considered significant.

SUPPLEMENTAL MATERIAL APPENDIX:

Supplemental Figure 1. Haplotypes observed in 762 F508del homozygous samples (1,524 chromosomes) across the *SLC26A9* locus

Supplemental Figure 2. Violin plot of *CFTR* and *SLC26A9* expression in pancreatic cells

Supplemental Figure 3. Dual Luciferase-Renilla Experimental Design

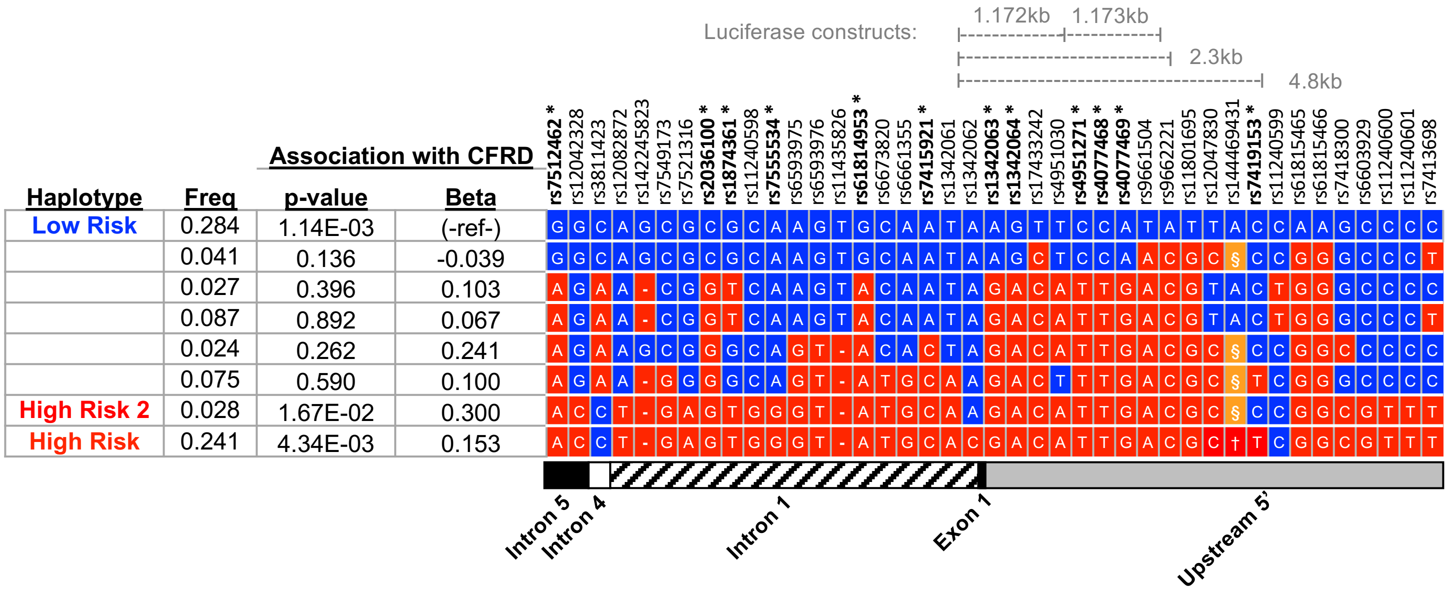
Supplemental Table 1. Summary statistics of associations for SNPs that reached genome-wide or suggestive significance in the genome-wide association study for CFRD onset (Blackman *et al.*, 2013)

Supplemental Table 2. The number of cells sequenced by predicted cell type

Supplemental Table 3. Average normalized gene expression values of selected genes in cells that express only *CFTR*, only *SLC26A9*, both or neither

Supplemental Table 4. Top CFRD-associated variants as eQTLs for *SLC26A9*

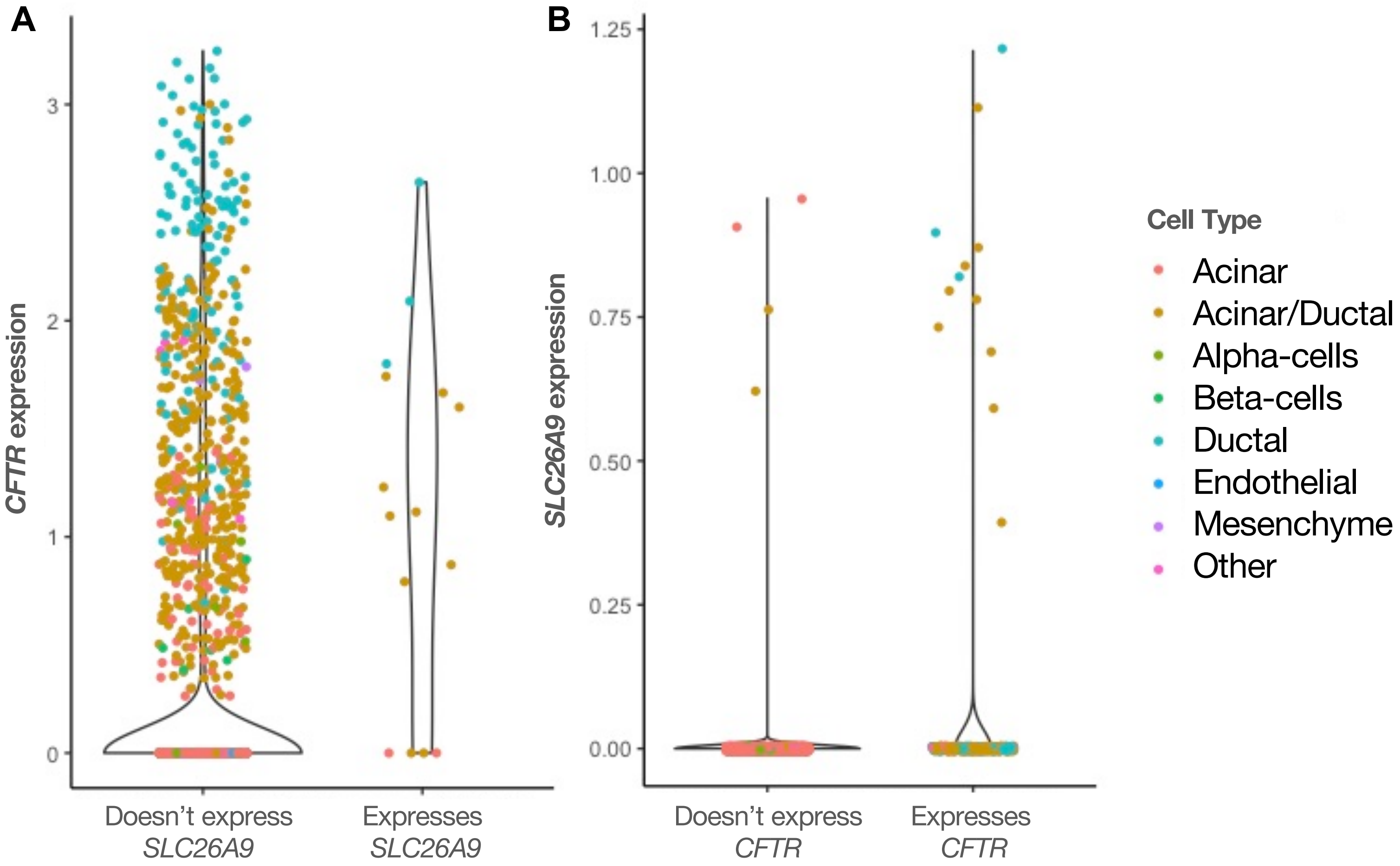
.



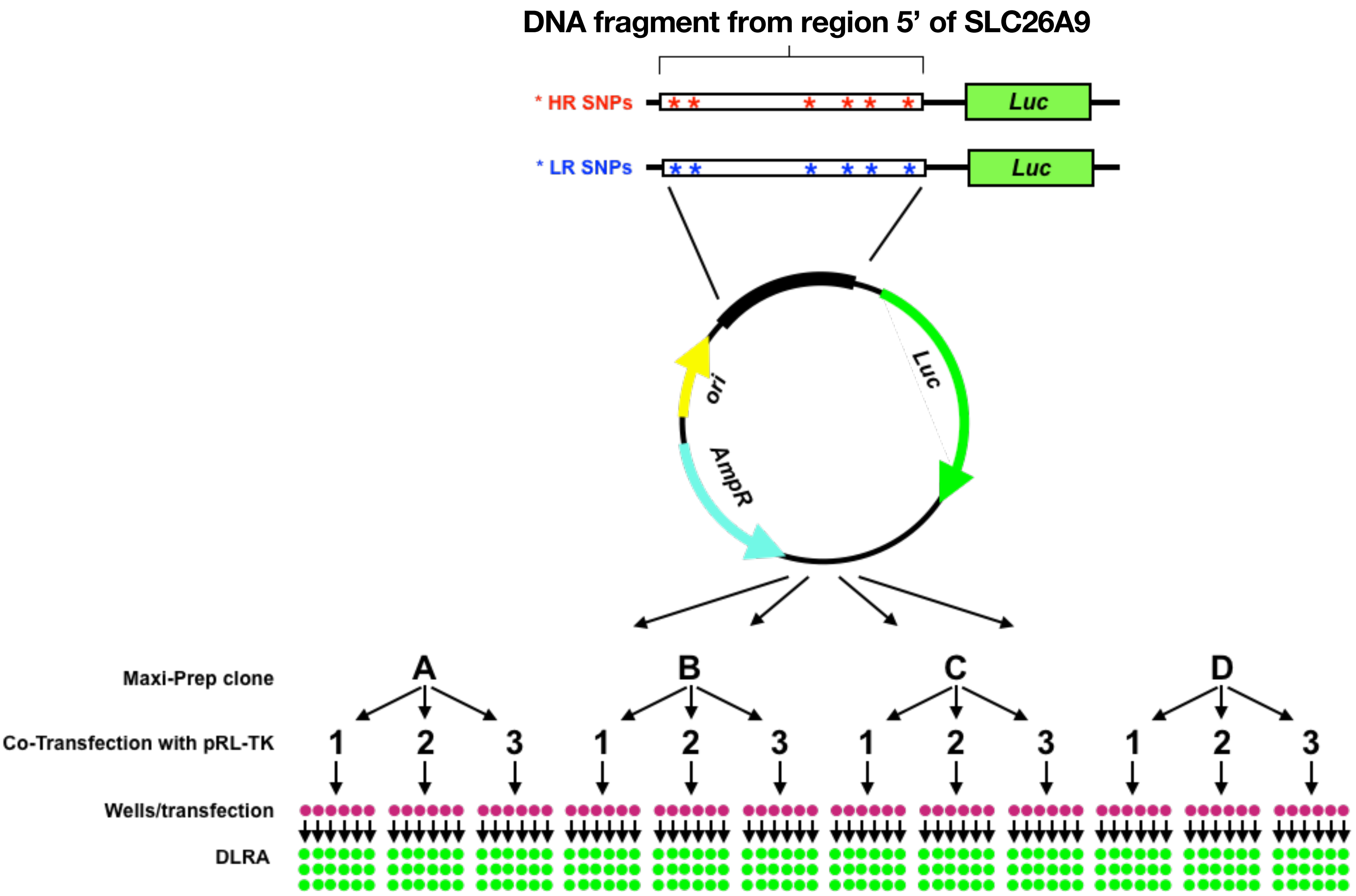
Haplotype	Freq	Association with CFRD		**rs7512462** *	rs12042328	rs3811423	rs12082872	rs142245823	rs7549173	rs7521316	**rs2036100** *	**rs1874361** *	rs11240598	**rs7555534** *	rs6593975	rs6593976	rs11435826	**rs61814953** *	rs6673820	rs6661355	**rs7415921** *	rs1342061	rs1342062
		p-value	Beta																				
Low Risk	0.284	1.14E-03	(-ref-)	G	G	C	A	G	C	G	C	G	C	A	A	G	T	G	C	A	A	T	A
	0.041	0.136	-0.039	G	G	C	A	G	C	G	C	G	C	A	A	G	T	G	C	A	A	T	A
	0.027	0.396	0.103	A	G	A	A	-	C	G	G	T	C	A	A	G	T	A	C	A	A	T	A
	0.087	0.892	0.067	A	G	A	A	-	C	G	G	T	C	A	A	G	T	A	C	A	A	T	A
	0.024	0.262	0.241	A	G	A	A	G	C	G	G	G	C	A	G	T	-	A	C	A	C	T	A
	0.075	0.590	0.100	A	G	A	A	-	G	G	G	G	C	A	G	T	-	A	T	G	C	A	A
High Risk 2	0.028	1.67E-02	0.300	A	C	C	T	-	G	A	G	T	G	G	G	T	-	A	T	G	C	A	A
High Risk	0.241	4.34E-03	0.153	A	C	C	T	-	G	A	G	T	G	G	G	T	-	A	T	G	C	A	C

Haplotype	**rs1342063** *	**rs1342064** *	rs17433242	rs4951030	**rs4951271** *	**rs4077468** *	**rs4077469** *	rs9661504	rs9662221	rs11801695	rs12047830	rs144469431	**rs7419153** *	rs11240599	rs61815465	rs61815466	rs7418300	rs6603929	rs11240600	rs11240601	rs7413698
Low Risk	A	G	T	T	C	C	A	T	A	T	T	A	C	C	A	A	G	C	C	C	C
	A	G	C	T	C	C	A	A	C	G	C	§	C	C	G	G	G	C	C	C	T
	G	A	C	A	T	T	G	A	C	G	T	A	C	T	G	G	G	C	C	C	C
	G	A	C	A	T	T	G	A	C	G	T	A	C	T	G	G	G	C	C	C	T
	G	A	C	A	T	T	G	A	C	G	C	§	C	C	G	G	C	C	C	C	C
	G	A	C	T	T	T	G	A	C	G	C	§	T	C	G	G	G	C	C	C	C
High Risk 2	G	A	C	A	T	T	G	A	C	G	C	§	C	C	G	G	C	G	T	T	T
High Risk	G	A	C	A	T	T	G	A	C	G	C	†	T	C	G	G	C	G	T	T	T

Supplemental Figure 1. Haplotypes observed in 762 F508del homozygous samples (1,524 chromosomes) across the *SLC26A9* locus. Representation of *SLC26A9* SNP haplotypes with MAF>15% and MHF>2%. Location of variants relative to *SLC26A9* are shown in box diagram below the haplotypes (Note: *SLC26A9* is on (-) DNA strand thus locations are shown 3' to 5' from left to right, not drawn to scale). Haplotype frequencies, p-values and beta values are shown to the left of the respective haplotype. rsIDs are shown above with the gray dotted lines denoting the SNPs that were included in the respective luciferase constructs. CFRD-associated variants reported by Blackman *et al*. 2013 are marked by an asterisk (*) and bolded. SNPs highlighted in blue indicate the most common ancestral haplotype. Variants highlighted in red indicate changes from the most common ancestral haplotype. § indicates TGGGGCCTCGGGTACCTCA, and † indicates TGGGGCCTCGGGTATCTCA. In addition to the Low Risk (LR) and High Risk (HR) haplotypes that we functionally test in this study, also labeled here is High Risk 2, which is identical at 11 of 12 CFRD-associated SNPs (exception is rs7419153).



Supplemental Figure 2. Violin plot of *CFTR* and *SLC26A9* expression in pancreatic cells. Expression is in log-normalized transcript counts. From our current study, each data point is a cell, colored by the predicted cell type. **(A)** Expression of *CFTR* in cells that do and do not express *SLC26A9*. **(B)** Expression of *SLC26A9* in cells that do and do not express *CFTR*.



Supplemental Figure 3. Dual Luciferase-Renilla Experimental Design. Constructs containing the a DNA fragment from the 5' region of SLC26A9 containing either high risk (HR) or low risk (LR) variants for risk of developing CFRD were cloned in a luciferase reporter plasmid. For each of the DNA fragments tested, clones obtained from two-four maxipreps were prepared and arbitrarily designated as clones A-D. These clones were co-transfected two-three independent times into PANC-1 or CFPAC-1 cells along with the same quantity per well of renilla luciferase encoding plasmid pRL-TK, a control reporter plasmid for the normalization of transfection efficiency. Each transfection consists of 6 wells per transfection. Dual Luciferase-Renilla reading were performed three times for each of the well. Each of the three readings were averaged. Note that the amount of experimental luciferase construct used was calculated based on the concentration of the plasmid adjusted for it size (molecular molar mass). Experimental plasmids and control plasmids are added in a 50:1 ratio.

Supplemental Table 1. Summary statistics of associations for SNPs that reached genome-wide or suggestive significance in the genome-wide association study for CFRD onset (Blackman *et al.*, 2013). Association conducted with martingale residuals of Cystic Fibrosis Related Diabetes in 762 508del homozygotes.

rsID	bp (hg19)	Location	Ref	Alt	Freq	p-value	beta
rs7512462	205899595	Intron 5	C	T	0.42	1.63E-06	-0.15
rs2036100	205907872	Intron 1	G	C	0.41	5.65E-06	-0.14
rs1874361	205908186	Intron 1	A	C	0.47	1.04E-04	0.12
rs7555534	205908867	Intron 1	C	T	0.34	1.00E-04	0.13
rs61814953	205910080	Intron 1	C	T	0.39	5.39E-05	-0.13
rs7415921	205910883	Intron 1	G	T	0.45	1.14E-05	0.14
rs1342063	205912859	Upstream	T	C	0.42	5.92E-05	-0.13
rs1342064	205913073	Upstream	C	T	0.42	6.04E-05	-0.13
rs4951271	205913848	Upstream	G	A	0.43	3.28E-05	-0.13
rs4077468	205914757	Upstream	G	A	0.42	3.83E-05	-0.13
rs4077469	205914885	Upstream	T	C	0.42	3.83E-05	-0.13
rs7419153	205917309	Upstream	A	G	0.38	6.17E-04	0.11

Supplemental Table 2. The number of cells sequenced by predicted cell type in our study ('Current Study') and two publicly available datasets (Baron *et al*., 2016 (GSE84133) ; Segerstolpe *et al*., 2016 (E-MTAB-5061)). Count and percentage of the cell types in each study is displayed.

Current Study			**Baron**			**Segerstolpe**		
Cell Type	**Count**	**Percent**	**Cell Type**	**Count**	**Percent**	**Cell Type**	**Count**	**Percent**
Acinar	2032	67.8	Beta	2525	29.5	Alpha cell	886	40.1
Acinar/Ductal	479	16.0	Alpha	2326	27.1	Ductal cell	386	17.5
Beta-cells	134	4.5	Ductal	1077	12.6	Beta cell	270	12.2
Ductal	133	4.4	Acinar	958	11.2	Gamma cell	197	8.9
Other	75	2.5	Delta	601	7.0	Acinar cell	185	8.4
Alpha-cells	65	2.2	Activated stellate	284	3.3	Delta cell	114	5.2
Mesenchyme	45	1.5	Gamma	255	3.0	PSC cell	54	2.4
Endothelial	36	1.2	Endothelial	252	2.9	Unclassified endocrine cell	41	1.9
Total	2999		Quiescent stellate	173	2.0	Co-expression cell	39	1.8
			Macrophage	55	0.6	Endothelial cell	16	0.7
			Mast	25	0.3	Epsilon cell	7	0.3
			Epsilon	18	0.2	Mast cell	7	0.3
			Schwann	13	0.2	MHC class II cell	5	0.2
			T_cell	7	0.1	Unclassified cell	2	0.1
			Total	8569		Total	2209	

Supplemental Table 3. **Average normalized gene expression values of selected genes in cells that express only *CFTR*, only *SLC26A9*, both or neither.** List encompasses selected ion channels, bicarbonate transporters, FOXI1 and WNK pathway genes. Gene expression values in our study ('Current Study'), and four previously published studies (Baron *et al.* (GSE84133), Wang *et al.* (GSE83139), Muraro *et al.* (GSE85241) and Segerstolpe *et al.* (E-MTAB-5061)) are shown. Each expression value has been colored according to its relative expression value within each study, where green indicates high expression and grey/white indicates lower expression. NA indicates that this gene was not detected in that study. *SLC26A9* average expression among the five studies are as follows: 'Current Study': 0.0043, 'Baron': 0.0066, 'Wang': 3.9926, ' Muraro': 0.0173 and 'Segerstolpe': 0.7564.

		Current Study				Baron				Wang				Muraro				Segerstolpe			
Ion Channel	Alternative Gene Name	Expresses CFTR only	Expresses SLC26A9 only	Both	Expresses Neither	Expresses CFTR only	Expresses SLC26A9 only	Both	Expresses Neither	Expresses CFTR only	Expresses SLC26A9 only	Both	Expresses Neither	Expresses CFTR only	Expresses SLC26A9 only	Both	Expresses Neither	Expresses CFTR only	Expresses SLC26A9 only	Both	Expresses Neither
ANO1	CACC, TMEM16A	0.00	0.00	0.01	0.01	0.08	0.50	0.11	0.06	25.32	0.24	12.36	15.04	0.30	0.00	0.21	0.20	1.61	0.00	4.64	1.17
AQP1		1.43	0.00	0.90	0.08	2.31	0.00	1.44	0.04	169.18	0.16	237.47	24.23	9.07	0.33	27.87	0.09	270.75	0.53	473.63	3.80
AQP5		0.00	0.00	0.00	0.00	0.04	0.17	0.11	0.00	0.18	0.16	0.17	0.18	0.00	0.00	0.00	0.00	0.73	0.00	2.91	0.21
ATP6V0D2		NA	NA	NA	NA	0.00	0.00	0.00	0.00	0.59	0.16	0.17	0.41	0.01	0.17	0.07	0.03	0.008	0.000	0.000	0.296
ATP6V1B1		0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.84	0.16	0.17	0.58	0.07	0.00	0.04	0.02	0.49	0.29	0.25	0.15
ATP6V1C2		NA	NA	NA	NA	0.00	0.00	0.00	0.00	422.48	670.54	309.71	337.51	0.00	0.00	0.00	0.01	0.30	0.28	0.31	0.18
FOXI1		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.16	0.17	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
OSR1		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.40	0.16	0.17	0.18	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.19
SCNN1A	ENaCA	0.26	0.00	0.17	0.01	0.11	0.00	0.11	0.01	44.84	63.75	126.91	12.15	2.26	0.00	3.47	0.07	49.16	15.97	61.13	1.60
SCNN1B	ENaCB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.63	0.16	1.13	0.18	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.03
SCNN1D	ENaCD	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.51	0.16	0.17	0.38	0.04	0.17	0.04	0.02	1.00	0.00	0.76	1.26
SCNN1G	ENaCG	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	2.39	0.16	0.17	0.19	0.06	0.00	0.25	0.00	3.06	0.00	6.39	0.00
SLC26A3		NA	NA	NA	NA	0.00	0.00	0.00	0.00	2.51	0.16	0.17	5.46	0.00	0.00	0.00	0.00	0.23	0.00	0.00	0.01
SLC26A4	PENDRIN	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.01	6.37	0.16	13.93	0.84	0.26	0.00	0.36	0.14	0.20	0.00	0.06	0.68
SLC26A6		0.00	0.00	0.01	0.01	0.05	0.17	0.22	0.03	19.69	0.16	4.22	17.03	0.15	0.00	0.14	0.09	5.86	12.08	7.19	4.56
SLC4A2	AE2	0.23	0.00	0.10	0.04	0.43	0.33	0.33	0.24	40.28	53.21	8.06	35.23	0.61	0.33	0.50	0.26	28.93	46.42	44.92	16.53
SLC4A4	NBCe1-B	2.51	0.56	1.94	0.13	2.89	0.83	2.33	0.12	760.69	24.95	2140.25	51.50	11.24	2.87	24.44	0.41	119.11	14.96	162.55	4.72
SLC4A7	NBCn1	0.08	0.00	0.01	0.02	0.10	0.50	0.44	0.11	155.67	116.36	34.66	118.66	1.93	0.67	1.66	0.82	6.03	1.14	8.51	4.22
SLC9A1	NHE1	0.00	0.00	0.05	0.01	0.24	0.50	0.11	0.11	11.08	26.94	3.13	11.12	0.86	0.67	1.47	0.29	7.50	16.72	14.22	3.51
SLC9A3	NHE3	NA	NA	NA	NA	0.00	0.00	0.00	0.00	0.35	0.16	0.25	0.40	NA	NA	NA	NA	0.41	0.28	0.34	0.42
STK39		0.00	0.00	0.05	0.03	0.19	0.33	0.33	0.12	65.22	557.72	10.78	69.82	1.83	2.87	2.52	0.65	6.60	8.20	6.76	3.28
STK39	SPAK	0.00	0.00	0.05	0.03	0.11	0.17	0.33	0.13	65.22	557.72	10.78	69.82	1.835	2.870	2.522	0.655	6.60	8.20	6.76	3.28
WNK1		0.61	0.00	0.11	0.05	0.50	0.00	0.89	0.42	429.15	597.75	254.59	340.66	5.29	6.85	5.69	2.86	18.81	32.62	28.16	19.93
WNK4		0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	3.23	0.16	0.21	8.63	0.110	0.167	0.000	0.098	1.23	0.73	0.00	3.50

Supplemental Table 4. Top CFRD-associated variants as eQTLs for *SLC26A9.* Top CFRD-associated SNPs (as determined by Blackman *et al.,* 2013*)* and their association with *SLC26A9* expression in the pancreas and lung, obtained from GTEx, v7. A positive beta value indicates the risk variants associated with higher gene expression.

		Pancreas		**Lung**	
rsID	**Risk/Alt Allele**	**β**	**p value**	**β**	**p-value**
rs7419153	T/C	-0.285	1.44E-04	0.041	0.175
rs4077469	G/A	-0.220	3.13E-03	0.003	0.910
rs4077468	T/C	-0.220	3.13E-03	0.003	0.910
rs4951271	T/C	-0.208	4.79E-03	-0.004	0.886
rs1342064	A/G	-0.223	2.88E-03	-0.001	0.973
rs1342063	G/A	-0.224	2.80E-03	-0.001	0.978
rs7415921	C/A	-0.166	2.10E-02	0.015	0.603
rs61814953	A/G	-0.247	1.06E-03	0.010	0.734
rs7555534	G/A	-0.093	2.26E-01	0.054	0.061
rs1874361	T/G	-0.156	2.47E-02	0.048	0.072
rs2036100	G/C	-0.263	3.65E-04	0.016	0.573
rs7512462	A/G	-0.246	6.26E-04	0.006	0.839
Overall effect of risk alleles on gene expression:		**Decrease**		**No effect**	