Human Genetics (2026) 145:8
https://doi.org/10.1007/s00439-025-02800-7

ORIGINAL INVESTIGATION

Distinct regulatory elements of *SLC6A14* expression contribute to modification of cystic fibrosis phenotypes

Mohsen Esmaeili[1] · Cheng Wang[1] · Fan Lin[1] · Naim Panjwani[1] · Adele Chan[1] · Gengming He[1] · Katherine Keenan[1] · Julie Avolio[2] · Ann Harris[3] · Johanna M. Rommens[1] · Lisa J. Strug[1,4]

Received: 16 June 2025 / Accepted: 22 November 2025

Abstract

The *SLC6A14* gene on chromosome X modifies disease presentation in individuals with cystic fibrosis (CF). Population studies have revealed distinct proximal and distal SNP clusters associated with gastrointestinal and lung phenotypes, respectively. Given the co-localization of cis-eQTLs within the corresponding associated chromosomal regions, where less *SLC6A14* expression aligns with improved phenotypic outcomes, we sought to understand the SNP contributions to *SLC6A14* expression regulation. Using reporter gene assays, we show that the proximal cluster, aligning with pancreas eQTLs, provides variation in promoter activity that is allele-dependent. The more expansive distal cluster, aligning with primary nasal cell eQTLs, was surveyed and found to harbour an enhancer feature that augmented expression in conjunction with the promoter in cells of lung origin. The rs4446858 SNP present within the enhancer aligns with the binding motif of the IRF1 transcription factor, where the alternate C allele was found to respond to direct addition of IRF1 or its increase following IFN-γ stimulation, resulting in increased SLC6A14 expression. Our data support a conclusion that *SLC6A14* expression in airway tissue is regulated, in part, by immune-responsive genetic features, likely involving goblet cells. While our findings support previous model system studies indicating beneficial acute effects of *SLC6A14* expression, they also highlight a distinction between immune-related responses and cumulative effects of long term disease reflected as lung function in individuals with CF. Our studies emphasize the challenges of elucidating complex genetic loci where there are multiple levels of regulatory influence with competing contributions.

Introduction

Cystic Fibrosis (CF [MIM: 219700]) is a life-limiting recessive disease that affects the respiratory, digestive and reproductive systems, and is caused by loss of function variants in the Cystic Fibrosis Transmembrane Conductance Regulator (*CFTR*) (Kerem et al. 1989). Development of small molecules targeting the faulty CFTR protein, known as modulators, have resulted in improved lung function and quality of life for individuals with CF (Grasemann and Ratjen 2023; Keating et al. 2025; Middleton et al. 2019; Salvatore and Pepe 2025). Access to the current modulators is costly and there remain no effective modulators for individuals with null or minimally functioning *CFTR* alleles (Guo et al. 2022). Further, CF individuals with the same causal variants display variable lung, pancreatic and gastrointestinal disease severity (Kerem and Kerem 1996; Luisetti 1997; Rosenstein 1994) and exhibit variable response to the modulators (Alicandro et al. 2024; Gong et al. 2022; Heijerman et al. 2019; Middleton et al. 2019; Strug et al. 2016; Zemanick et al. 2024). Genome-wide association studies (GWAS) have identified multiple modifier genes including *SLC6A14* (Solute Carrier 6 family member A14), which contribute to CF disease morbidity across multiple organs (Corvol et al. 2015; Gong et al. 2019, 2022; Strug et al. 2016; Sun et al. 2012). *SLC6A14* encodes a Na^+/Cl^--coupled, neutral and cationic amino acid transporter that is

✉ Lisa J. Strug
lisa.strug@utoronto.ca

1 Program in Genetics and Genome Biology, The Hospital for Sick Children Research Institute, Toronto, Canada

2 Translational Medicine Program, The Hospital for Sick Children Research Institute, Toronto, Canada

3 Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, USA

4 Departments of Statistical Sciences and Computer Science and the Division of Biostatistics, University of Toronto, Toronto, Canada

Published online: 05 January 2026

expressed in epithelia of several organs including the lungs (Sikder et al. 2017; Sloan and Mager 1999). *SLC6A14* has been shown to contribute to risk of intestinal obstruction at birth (meconium ileus, MI) and lung disease severity in CF. The single nucleotide polymorphisms (SNPs) associated with MI reside near the *SLC6A14* transcription start site (TSS), while the SNPs associated with lung disease severity and *Pseudomonas aeruginosa* (PsA) infection localize to a far upstream intergenic region, more than 200 kb from the *SLC6A14* TSS (Corvol et al. 2015; Gong et al. 2019; Lin et al. 2024). Both clusters include cis-expression quantitative trait loci (eQTLs) for *SLC6A14*. Based on adult non-CF pancreatic tissue from the Genotype Tissue Expression project (GTEx; GTEx Consortium 2015) and naïve CF primary human nasal epithelia (HNEs; Gong et al. 2019), alleles associated with reduced MI risk or better lung function consistently implicated less *SLC6A14* transcript levels. The mechanisms by which the associated SNPs contribute to *SLC6A14* expression, and correspondingly the CF phenotypes, are poorly understood. Further, despite replication for the GWAS and eQTL evidence pointing to less *SLC6A14* expression, there appears to be some inconsistency as cell model investigations point to beneficial roles of SLC6A14 in the airway, at least with respect to pathogen engagement and capacity for wound healing in CF tissues (Di Paola et al. 2017; Mercier et al. 2022). In this study, we sought to understand the airway-distinct GWAS and eQTL signals, and gain insight into the underlying contributors to *SLC6A14* expression that influence CF phenotypes.

Materials and methods

Study population

The phenotypic dataset for MI occurrence was obtained from the Canadian CF Gene Modifier Study (CGMS) and has been described previously (Corvol et al. 2015; Gong et al. 2019; Sun et al. 2012).

Bulk RNA sequencing and analyses

RNA sequencing (RNA-Seq) was performed on HNEs collected from 107 individuals with CF from the CGMS. Samples were obtained through nasal brushing of the inferior turbinate tissues and RNA was extracted with the RNeasy Micro Kit (Qiagen, 74004) with sequencing on Illumina HiSeq 2000 or 2500 platforms (Gong et al. 2019; Polineni et al. 2018).

All sequenced reads were aligned to the human genome GRCh38 reference using STAR version 2.5.4b (Dobin et al. 2013). Aligned reads were converted into expression counts using RNA-SeQC version 2.0.0 (DeLuca et al. 2012) and analyzed as normalized measures of transcripts per million (TPM) (Wagner et al. 2012) or trimmed mean of M values (TMM) (Robinson and Oshlack 2010).

Cis-expression quantitative trait loci (eQTL) analysis

eQTLs for *SLC6A14* transcript levels were calculated from HNE bulk RNA-Seq data. Associations between TMM-normalized gene expression and SNP genotypes within 1 Mb up- and down-stream of the *SLC6A14* TSS were evaluated using FastQTL version 2.0 (Ongen et al. 2016). Covariates were included in the model to adjust for: sample quality based on RNA integrity number (RIN; RIN$\geq$7 for all samples), population structure (top three genotype principal components), individual and study background (sex, study source), immune cell contamination (*PTPRC*/*CD45* gene expression) and fifteen latent probabilistic estimation of expression residual (PEER) factors to account for background effects. Genotype principal components were calculated using the R package GENESIS (Conomos et al. 2015). PEER factors were generated by the R package PEER (Stegle et al. 2012).

Co-localization analyses

Analysis with the Simple Sum (SS) co-localization testing framework (Gong et al. 2019; Wang et al. 2022) was used to assess whether the same variants contribute to different traits. The chromosome X-associated locus from a previous lung function GWAS meta-analysis (Gong et al. 2019) was used to evaluate co-localization with *SLC6A14* eQTLs from primary HNEs and with GWAS summary statistics from the age of first PsA infection in individuals with CF. In brief, the lung GWAS included 6365 individuals with CF with a median age of 19.5 years from four separate studies in Canada, USA and France. Linear regression (for data from unrelated individuals) and linear mixed modeling (for data that included siblings) were used to test for association with lung function, measured in sex- and age-specific CF FEV_1 percentiles adjusted for mortality, revealing the extended associated locus between *AGTR2* (Angiotensin II receptor type 2) and *SLC6A14* genes. The age of first PsA infection GWAS included 1653 individuals with CF from the CGMS with mean age of 34 years (Lin et al. 2024) using linear mixed modeling. Linkage disequilibrium (LD) with the most significant lung function GWAS SNP was calculated based on the female European samples from the 1000 Genomes Project (1000 Genomes Project Consortium 2015). Both SS testing and co-localization visualizations displaying *p*-values plotted along the physical map in the

SLC6A14 region were generated using the web tool LocusFocus version 1.4.9 alpha (Panjwani et al. 2020).

Long range interaction and assessment of chromatin accessibility

Circular chromosome conformation capture with deep sequencing (4C-seq) analysis was performed with Calu-3 cells as described previously (NandyMazumdar et al. 2020) using *SLC6A14*-specific primers (Table S1). Assays for transposase-accessible chromatin with deep sequencing (ATAC-Seq) were performed with Calu-3 and 16HBE14o- cell lines, and for human bronchial epithelia grown at air liquid interface (HBE-ALI) as described previously (NandyMazumdar et al. 2020).

Co-expression analysis of SLC6A14 and IRF1/2

For the bulk HNE RNA-Seq samples, co-expression was tested using linear regression models with TMM-normalized expression data for *IRF1* or *IRF2* regressed on *SLC6A14* expression TMMs. The co-expression model was adjusted for the same set of covariates as the eQTL analysis, excluding *PTPRC/CD45* expression and 4 PEER factors associated with immune responses. *SLC6A14* is known to be subject to X-inactivation (Tukiainen et al. 2017).

For the analyses of single cell RNA-Seq data, raw read count data was obtained from the human lung cell atlas study by Travaglini and colleagues involving a set of 9394 cells and 44 cell types (Synapse: syn21041850/wiki/600865, accessed 2021-11-28; Travaglini et al. 2020).

SmartSeq2 (SS2) single cell data were used to assess *SLC6A14* co-expressions with *IRF1* and *IRF2* in different lung cell types. Zero-inflated negative binomial models were fitted by regressing *IRF1* or *IRF2* expression read counts on that of *SLC6A14* using a log link function. Proportion of genes with detected expression was estimated for each cell and included as a covariate in the zero-inflation portion of the model, using R function zeroinfl() from the package pscl in R version 3.5.1 (Zeileis et al. 2008). Bonferroni correction was applied for multiple hypothesis testing.

Cell culture

HEK-293, PANC-1 and IB3-1 cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS), 100 U/ml penicillin/streptomycin (pen/strep), 1mM sodium pyruvate, 2mM L-glutamine and 25mM HEPES (pH 7.8). GripTite 293 MSR cells (Thermo Fisher Scientific) were cultured in DMEM supplemented with 10% FBS, 25 U/ml pen/strep, 4mM L-glutamine, 0.1 mM MEM Non-Essential Amino Acids (NEAA) and 600 µg/ml Geneticin. CFPAC-1 and Capan-1 cells were cultured in IMDM with 100U/ml pen/strep plus 10 and 20% FBS, respectively. H6c7 cells (Kerafast, ECA001-FP), were cultured in keratinocyte serum free media supplemented with 25 mg bovine pituitary extract, 2.5 µg human recombinant epidermal growth factor and 100U/ml pen/strep. 16HBE14o- cells (MilliporeSigma, SCC150) were cultured in MEM supplemented with 10% FBS, 100U/ml pen/strep and 2mM L-glutamine. Calu-3 cells were grown as for 16HBE14o- with addition of 1mM sodium pyruvate and 0.1mM NEAA. MOR cells were grown in RPMI1640 supplemented with 10% FBS and 100U/ml pen/strep. All cell lines were cultivated in media without phenol red in a humidified atmosphere with 5% CO_2 at 37 °C. HNE samples were prepared and cultured using Collagen IV-coated transwell membranes as described previously (Gunawardena et al. 2023).

For the experiments involving interferon stimulation, human recombinant IFN-α (STEMCELL Technologies, 78076), IFN-ß (STEMCELL Technologies, 78113) and IFN-γ (Cell signaling, 80385) were reconstituted and used according to supplier protocols.

Plasmids and site directed mutagenesis

The restriction enzyme cut-free cloning strategy was carried out to generate expression vectors for human *SLC6A14* (GenBank: NM_007231), *IRF1* (GenBank: NM_002198) and *IRF2* (GenBank: NM_002199) genes, as described (Lou and Jin 2017). Briefly, coding sequences were amplified by PCR using two primer pairs (Table S2) to obtain products that were then mixed, denatured and re-annealed to generate gene fragments with desired cohesive ends. The *SLC6A14* insert was ligated to the EcoRI/BamHI-digested pTetOne vector (Clontech, 634310), and *IRF1*, *IRF2* inserts were ligated to the NotI-digested pCMV vector (Clontech, 631719), replacing *LacZ*. To generate the luciferase reporter plasmids with the *SLC6A14* proximal promoter, the 2442 bp (2.4 kb) fragment was first generated from human genomic DNA from individuals with CF with appropriate haplotypes by PCR, using SLC6A14-Prom 2442 primers (Table S2) for insertion into the luciferase reporter plasmid pGL4.10 (Promega, E6651) via Acc65I and HindIII sites. To generate truncated expression plasmid versions, specific primers (Table S2) and the Q5® Site-Directed Mutagenesis Kit (NEB, E0554S) was used as recommended.

The upstream segment (205 kb from the *SLC6A14* TSS) of 2102 base pairs (NC_000023.11:116229787–116231888; 2.1 kb) was added to selected promoter vectors using the restriction enzyme cut-free cloning strategy indicated above. The most common haplotype (D1) was amplified by PCR from human genomic DNA using primers 9 and 10 (Table S2) yielding Acc65I cohesive ends for direct insertion into

pGL4.10 reporter vectors in both orientations. *SLC6A14* reporter vectors with the SV40 enhancer were prepared similarly using primers indicated in Table S2.

Site directed mutagenesis was performed with appropriate primers (Table S2) to convert the common D1 haplotype to D1/D4 (rs4446858 (T>C) or D4 (rs4446858 (T>C) plus rs12009976 (G>A). Reporter plasmids containing only D1 fragments were generated by deletion of the *SLC6A14* proximal promoter fragments using an overlap extension PCR protocol (Lee et al. 2010). Briefly, PCR products representing the flanking regions of the sequence to be deleted were prepared separately using nonchimeric and chimeric primer sets (Fragments A and B, in Table S2). The products were merged in a second round of PCR and inserted into respective vectors at EcoRI and HindIII sites. The Q5® High-Fidelity DNA polymerase (NEB, M0491) was used for all vector constructions with verification by restriction enzyme mapping and Sanger sequencing.

Reporter gene assays

Cells in 6-well plates were cotransfected with 0.8 µg reporter vector and 0.2 µg pCMV-lacZ (Clontech, 631719) for internal normalization/transfection control, 24 h after seeding. As required, control vector or expression vectors for the human *IRF1* and *IRF2* genes were added to the transfection mixes. HEK-293 and PANC-1 cells were transfected with a modified $CaPO_4$ method (Esmaeili et al. 2016). After 16 h, the cells were washed with PBS and given fresh medium. Other cell lines were transfected using X-tremeGENE HP™ DNA Transfection Reagent (Roche, MilliporeSigma, 6366236001) according to manufacturer's protocol, with minor modification. Cells were harvested 72 h after transfection to measure both luciferase and beta-galactosidase activities (Esmaeili et al. 2016). Obtained luciferase units were normalized to those of beta-galactosidase.

Quantitative reverse transcription PCR (qRT-PCR)

RNA was isolated from the cells using TriPure™ isolation reagent (Roche, MilliporeSigma, 11667157001) according to the manufacturer's protocol and genomic DNA was removed with ezDNase™ enzyme (Invitrogen™, Thermo Fisher Scientific, 11766051). One-step qRT-PCR was conducted using SuperScript™ III Platinum™ SYBR™ Green One-Step qRT-PCR Kit (Invitrogen™, Thermo Fisher Scientific, 11736051), gene specific primers (Table S3) and a Bio-Rad CFX96™ Real Time PCR detection system. qRT-PCR results were analyzed using ΔΔCt method (Pfaffl 2001).

Antibodies and immunoblot analyses

Whole cell protein lysates were prepared by lysis in NETN buffer (100mM NaCl, 20mM Tris/HCl pH 8.0, 1mM EDTA, 0.5% NP-40) containing MS-SAFE Protease and Phosphatase Inhibitor Cocktail (MilliporeSigma, MSSAFE-1VL) with three freeze/thaw cycles (liquid nitrogen/37°C). 30 µg of protein extracts were resolved on SDS-PAGE. For deglycosylation experiments, the samples were treated with PNGase F (NEB, P0704) or Endo H (NEB, P0702) according to the supplier's protocol except that denaturation by heating was replaced with incubation at room temperature for 30 min. Primary antibodies used for immunodetection included SLC6A14 (Atlas Antibodies, HPA003193), IRF1 (Cell Signaling, 8478), IRF2 (Invitrogen, PA584335) and Calnexin (MilliporeSigma, C4731). Horseradish peroxidase-conjugated goat anti-rabbit IgG (MilliporeSigma, A6154) was used as the secondary antibody. Signal detection was captured with enhanced chemiluminescence (ECL) reagent (Bio-Rad, 1705060) using an Amersham Imager AI680 (GE Healthcare, 29270769).

Statistical and bioinformatic analyses

Two-tailed unpaired t-tests were performed for comparisons of groups in reporter gene and immunoblot analyses using the GraphPad QuickCalcs (https://www.graphpad.com/quickcalcs/ttest1/. Accessed February 2025). Chi-squared testing was used to compare MI frequency among CF individuals with various haplotypes. Bonferroni correction was applied for multiple comparisons. Bioinformatic analysis of the broad GWAS association region (chrX:116173577–116286195, GRCh38) was utilized to prioritize sub-regions for cis-regulatory potential using supporting tools from the UCSC Genome Browser (Perez et al. 2025); including evidence from open chromatin (ATAC-Seq) and chromatin conformation (4C) data, transcription factor (TF) ChIP-Seq data from ReMap (Hammal et al. 2022), TF binding motif predictions by JASPAR (Fornes et al. 2020) and annotations from the ENCODE project (ENCODE Project Consortium 2020).

Results

SLC6A14 transcript and protein expression

The MI and lung disease GWAS signals on the X chromosome showed narrow proximal and broad distal intergenic regions, respectively, indicating distinct determinants for *SLC6A14* in corresponding tissues (Gong et al. 2019). We first surveyed primary and respective tissue-derived cell

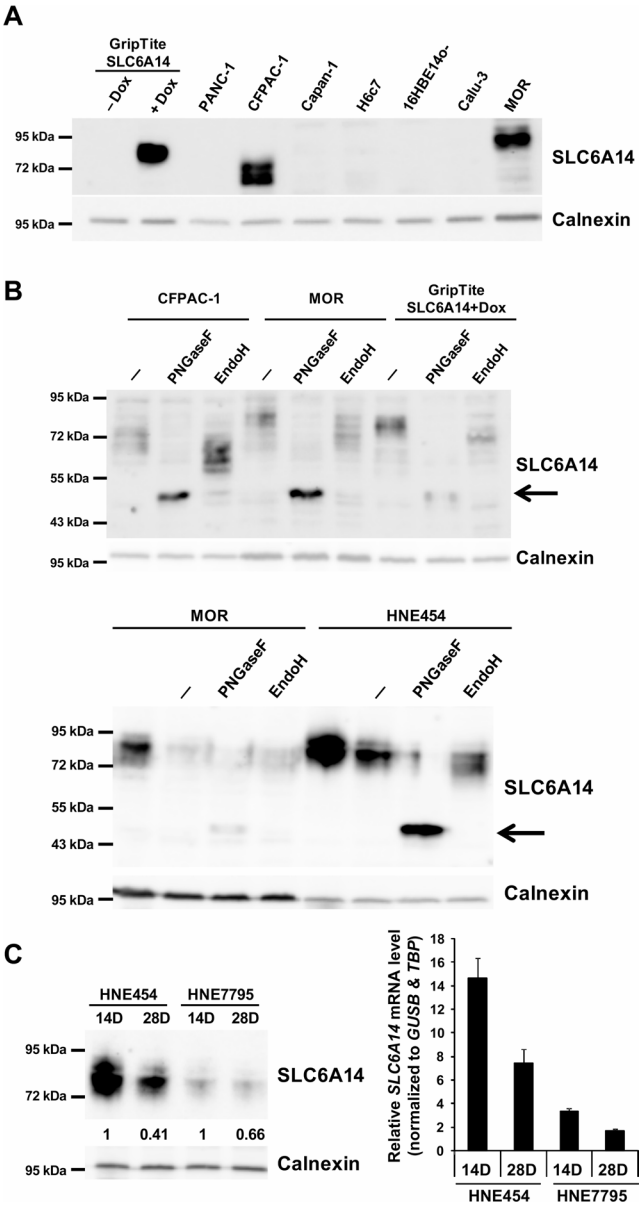


Fig. 1 Expression of SLC6A14 glycoisoforms in cell lines and primary respiratory epithelia. **A** Immunoblot analysis of whole cell extracts of the indicated cell lines were carried out with SLC6A14 and Calnexin (control) antibodies. The SLC6A14 antibody was validated with extracts of GripTite cells transiently transfected with an SLC6A14-inducible expression vector and treated with and without doxycycline (Dox; 100ng/ml) for 48 h. **B** Treatment of extracts of cell lines and HNEs with endoglycosidases or buffer only (–), revealed extensive glycosylation leaving residual SLC6A14 protein with PNGaseF treatment (indicated by the arrows). The lower panel includes untreated whole cell extracts in lanes 1 and 5. **C** Immunoblot analysis of protein extracts of HNE cells grown on transwell membrane as ALI cultures for 14 and 28 days (left panel). Quantification is shown normalized with Calnexin level to the respective 14 day culture. qRT-PCR analysis of RNA prepared from parallel cultures (right panel) with normalization to *GUSB* and *TBP* indicates alignment of RNA with protein levels. Error bars represent standard deviation of three technical replicates

lines, observing that expression of SLC6A14 protein could be detected by immunoblotting of cell extracts (Fig. 1A) with robust levels in HNE-ALI cultures, and in CFPAC-1 and MOR cell lines. SLC6A14 is an integral membrane protein with multiple predicted glycosylation sites. Treatments of protein extracts with EndoH and PNGaseF endoglycosidases supported complex and N-linked glycosylation with a consistent single core polypeptide (Sloan and Mager 1999) with either endogenous or heterologous expression (Fig. 1B). The higher and variable mobility bands reflect tissue glycosylation differences and multiple membrane spanning characteristics, consistent with previous reports (Kovalchuk et al. 2019; Rath and Deber 2013). *SLC6A14* mRNA could readily be detected in HNEs (Gong et al. 2019; He et al. 2022) and in multiple epithelial-derived cell lines of pancreas and lung origins by qRT-PCR aligning with the protein findings (Fig. S1). For the Calu-3 cell line, the absence of an immunoreactive signal was consistent with the comparatively low level of *SLC6A14* mRNA (Fig. S1). Further alignment of protein and RNA levels was evident in additional HNE-ALI samples (Fig. 1C), where expression was notably variable and readily influenced by culturing conditions. Consistency of protein and transcript expression has also been supported by direct functional investigation of SLC6A14 by others, including studies involving amino acid uptake assays (Ahmadi et al. 2019).

CF phenotype associations and proximal control of *SLC6A14* expression

The GWAS SNPs for MI risk and the most significant eQTLs for expression in the pancreas showed co-localization and alignment with the 5' end of *SLC6A14* (Gong et al. 2019). The absence of eQTLs directly within the early coding exons, pointed to expression regulatory capacity of the sequence near the TSS. This sub-region is conserved across mammals and has been annotated with two prominent candidate cis-Regulatory Elements (cCREs) by the ENCODE and Epigenomics Roadmap consortium projects (ENCODE Project Consortium 2020; Roadmap Epigenomics Consortium 2015; Satterlee et al. 2019) including one with a promoter-like signature (E2765491/prom), and one with a proximal enhancer-like signature (E2765490/enhP; Figs. 2A and S2A). A corresponding 2.4 kb segment containing the cCREs, the TSS and the transcript leader (haplotype PI, Fig. 2A, middle panel) was introduced into the pGL4.10 luciferase reporter expression vector system to evaluate promoter activity. Following transient transfection, robust reporter response was evident in both CFPAC-1 and MOR cell lines with less expression in Calu-3 cells (Fig. 2B), aligning with the SLC6A14 expression findings above.

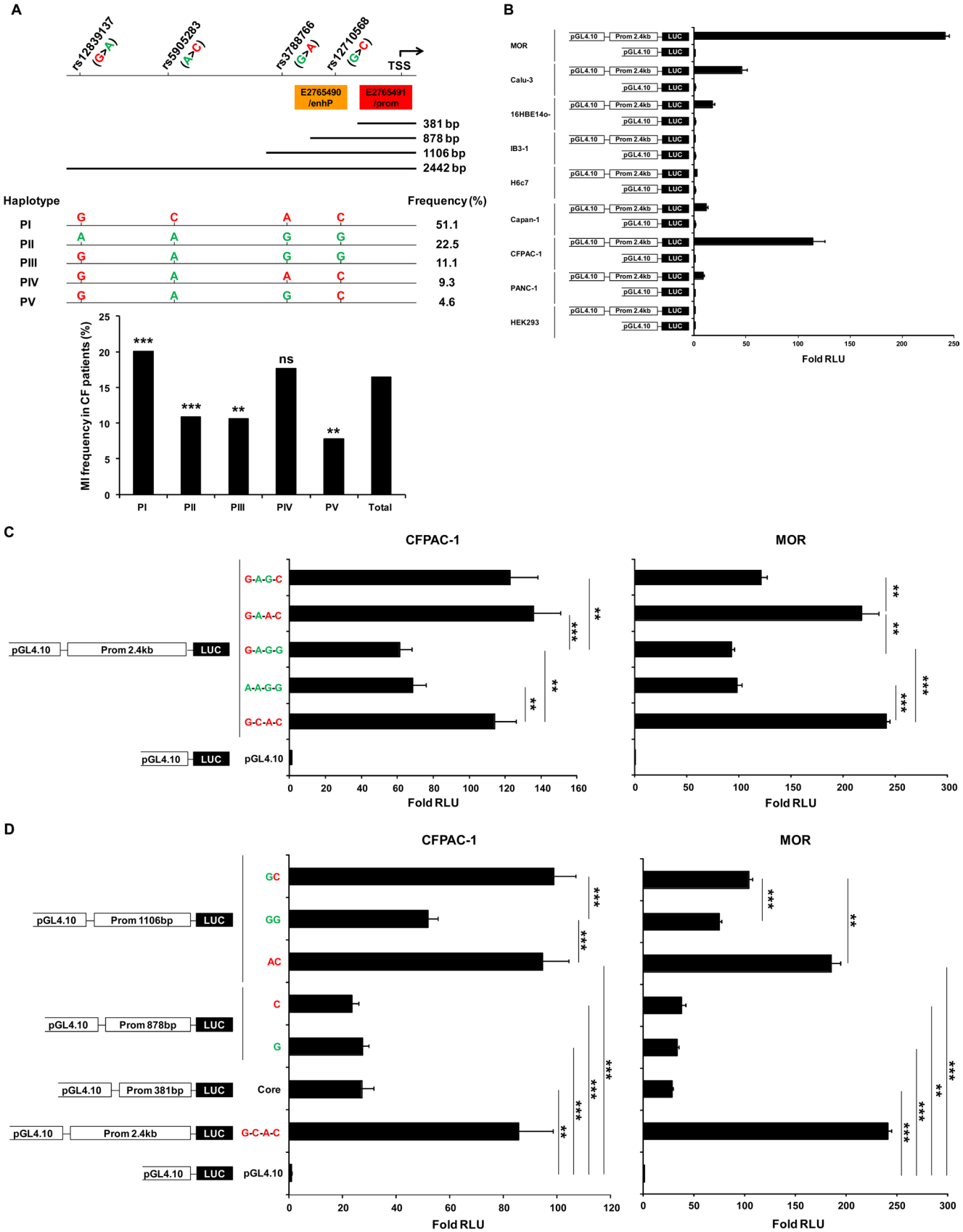

A
rs12839137 (G>A)
rs5905283 (A>C)
rs3788766 (G>A)
rs12710568 (G>C)
TSS
E2765490 /enhP
E2765491 /prom
381 bp
878 bp
1106 bp
2442 bp
Haplotype
Frequency (%)
PI G C A C 51.1
PII A A G G 22.5
PIII G A G G 11.1
PIV G A A C 9.3
PV G A G C 4.6
MI frequency in CF patients (%)
***
***
**
ns
**
PI PII PIII PIV PV Total
B
MOR
Calu-3
16HBE14o-
IB3-1
H6c7
Capan-1
CFPAC-1
PANC-1
HEK293
pGL4.10
Prom 2.4kb
LUC
Fold RLU
C
CFPAC-1
MOR
G-A-G-C
G-A-A-C
G-A-G-G
A-A-G-G
G-C-A-C
pGL4.10
Fold RLU
D
CFPAC-1
MOR
GC
GG
AC
C
G
Core
G-C-A-C
Prom 1106bp
Prom 878bp
Prom 381bp
Fold RLU

Fig. 2 MI risk haplotypes increase *SLC6A14* promoter activity in pancreatic and lung cell lines. **A** The genome segment with the MI-associated SNPs (risk and protective alleles in red and green, respectively), the TSS and 133 bp of the 5' untranslated region of the *SLC6A14* transcript is displayed with the annotated ENCODE cCREs (orange and red boxes) aligned above the complete 2.4 kb proximal promoter and trimmed segments tested for promoter activity in reporter assays. The middle panel describes the five common haplotypes along this proximal segment with indicated frequencies. The lower panel indicates the frequency of MI in individuals with CF grouped by haplotype (see Table S4 for haplotype interpretation). MI occurrence by haplotype was compared to the total occurrence by chi squared testing; ns: not significant, $^{**}P<0.01$, $^{***}P<0.001$ **B** Reporter assays with the common PI haplotype assessed promoter activity in pancreatic and lung cell lines. Beta-galactosidase was used as an internal control for transfection efficiency. The normalized luciferase values are shown as fold relative light units (RLU) after setting the values of pGL4.10-Luc arbitrarily at 1 for comparisons. Two to five independent assays were performed, each with two or three technical replicates. Error bars represent standard error of the mean (SEM). **C** Reporter gene assays were performed for all five common haplotypes in CFPAC-1 and MOR cell lines, with luciferase activity values determined as in (**B**). Mean ± SEM values from five independent assays for CFPAC-1 and two independent assays for MOR are shown, each with two or three technical replicates. Comparisons were performed using t tests; $^{**}P<0.01$, $^{***}P<0.001$. **D** Complete and trimmed segments with combinations of risk and protective alleles were then assessed as in (**C**). The mean ± SEM values from four independent assays for CFPAC-1 and two independent assays for MOR are shown, each with two technical replicates of each vector

Four *SLC6A14* eQTLs in the pancreas (Gong et al. 2019; GTEx V6) were present within the 2.4 kb segment including rs3788766 that exhibited the smallest *p*-value for genetic association with MI in consecutive GWAS studies by the International CF Gene Modifier Consortium (Gong et al. 2019; Sun et al. 2012; Fig. 2A, upper panel). Five allele combinations (PI to PV) explain > 98% of the haplotypes of the studied population with CF (6831 participants; Fig. 2A, middle panel). Groupings of individuals with these inferred haplotypes show MI frequencies that reflect the earlier GWAS (Fig. 2A, lower panel, Table S4). The PV and PIV haplotype groups are both limited in size, with haplotype PIV indicating no difference from overall MI occurrence of 16.5%. The larger PII and PIII groups exhibited lower MI frequency, and the PI group (with only risk alleles) exhibited higher frequency compared to the overall frequency of 16.5%.

To determine how the various haplotypes drive differential expression, reporter gene experiments were performed in CFPAC-1 and MOR cell lines. Both cell lines revealed expression contribution of the alleles, particularly for the rs3788766 and rs12710568 SNPs, where the respective A and C MI risk alleles were consistent with increased expression (compare PI or PIV versus PII or PIII; Fig. 2C).

Subsequent trimming of the 2.4 kb fragment to shorter lengths for additional reporter gene comparisons further supported the impact of the rs3788766 SNP in the MOR cell line, and in combination with the rs12710568 SNP in the CFPAC-1 cell line (Fig. 2D). The 1.1 kb fragment encompassing only the two SNPs near the TSS (AC haplotype for rs3788766 and rs12710568) reflected comparable levels of reporter activation as for the longer 2.4 kb segment (haplotype PI) in both CFPAC-1 and MOR cells (Fig. 2D). The modestly shorter 878 bp fragment containing only the rs12710568 SNP yielded response comparable to the shortest fragment tested with only 381 bp, implying that at least the segment containing the rs3788766 SNP is needed to contribute to functionality of the longer fragments, and that its individual alleles contribute to differential expression.

Moreover, the trimmed fragments align with the ENCODE-annotated cCREs for directing expression, as the inclusion of both E2765490/enhP and E2765491/prom yielded haplotype-responsive reporter gene expression comparable to the initial 2.4 kb segment, and the single E2765491/prom overlapping the TSS supported 'core' promoter activity in both CFPAC-1 and MOR cell lines.

Taken together, the reporter gene findings highlight the regulatory contribution of the near promoter region to expression, revealing the functional proximal and core promoter elements of *SLC6A14*. The alleles of the SNPs present also support the direction of effect as more expression was associated with increased MI risk in the initial GWAS and eQTL findings.

Lung phenotype associations and the distal *SLC6A14* locus

GWAS-associated variants with lung function in individuals with CF were annotated far upstream of *SLC6A14*. The distant locus encompasses genome-wide significant SNPs spanning at least 112.6 kb (from rs5194 to rs5952122; Gong et al. 2019). The more distal portion includes multiple associated SNPs in LD (Fig. 3A, red palette) that includes rs7879546, the SNP with the most significant GWAS *p*-value ($p=1.11\times10^{-9}$; Fig. 3A, purple dot), positioned 219.6 kb 5' to the TSS.

The associated region contains no annotated protein coding genes and is just downstream of the *AGTR2* gene with functional roles in the renin-angiotensin system, and negligible expression in airway tissues. Lung samples from the GTEx dataset report corresponding *AGTR2* transcript levels of only 1.1 TPM (GTEx V10) (GTEx Consortium 2015) and essentially none (0.00057 TPM) in a set of 107 primary CF HNE samples (increased by 44 samples from a previously reported analysis in Gong et al. 2019). This is in contrast to the notable levels of *SLC6A14* transcripts (average of 198.0 TPM) in HNEs and the readily detected SLC6A14 protein levels (Fig. 1B, C).

The many SNPs that are eQTLs for *SLC6A14* transcript levels in both GTEx tissues (non-CF) and the primary HNEs (from individuals with CF) broadened the locus to at least 161.3 kb (rs5950577 to rs5905248) with significant co-localization of the lung GWAS and HNE eQTL ($n=107$) statistics, with SS p-value $=6.02\times10^{-6}$ using 418 SNPs across a span of ±0.1 Mb (pale grey shading in Fig. 3A) relative to the rs7879546 SNP. The consistent direction of effects of the HNE eQTLs support that poorer lung function is associated with more SLC6A14 expression (Table 1). The eQTLs also apportioned the broad GWAS peak to at least two distinct segments (Fig. 3A, grey solid line).

An additional interrogation of the larger locus also considered association with age at first PsA infection, a secondary lung phenotype available in individuals with CF enrolled in the CGMS. The corresponding PsA association summary statistics colocalized with lung function GWAS association statistics with SS p-value $=6.02\times10^{-6}$, using 377 SNPs (Fig. 3A, yellow solid line), suggesting the same variants could influence the two phenotypes.

We then surveyed the entire broad distal interval using a layered approach to assess evidence for involvement in transcription regulation, with emphasis on information from nasal/airway tissue sources. Notable features included intermittent conservation and bioinformatic compilation of at least 26 cCREs, all annotated with distal enhancer signatures based on DNase hypersensitivity, histone modification or CTCF-binding as assigned by the ENCODE and Roadmap Epigenomics Consortium projects (ENCODE Project Consortium 2020; Roadmap Epigenomics Consortium 2015; Satterlee et al. 2019; Fig. S2A). These attributes together with the eQTLs and ATAC-Seq signals detected in differentiated HBE-ALI cultures and lung cell models (Fig. 3B), highlighted conserved sub-regions that could be prioritized with potential for expression modulation in the airway. These included at least a less distal 10 kb segment relative to the TSS of *SLC6A14* (116,278.7–116,288.7 kb; GRCh38; orange shading, Figs. 3A, S2A) with five cCREs (Fig. S2B) and a more distal 17 kb segment (116,226.7 to 116,243.7 kb; blue shading in Figs. 3A and S2A) with thirteen cCREs, encompassing several of the lung function GWAS-associated SNPs with the smallest p-values (Figs. 3C and S2C, upper panel).

The 10 kb sub-region was present in the eQTL peak segment where p-values were notable, but larger than those of the more distal segment (Figs. 3A; S2B). Of interest, the region contained a series of binding motifs for *NKX2* gene family members (Fig. S2B) with ChIP-Seq evidence for direct *NKX2-1* binding (Watanabe et al. 2013). The NKX2-1 homeodomain transcription factor is important early in lung development with expression levels that are maintained into adulthood in bronchial and alveolar epithelia. However, none of the eQTL alleles directly aligned with the implicated c/gTg/tGAGa/tGg/c consensus motif (Watanabe et al. 2013) present in the identified ChIP-Seq segments. We thus focused on the more distal 17 kb segment.

The 17 kb sub-region showed additional features relevant to expression regulation. Foremost, interrogations of the Calu-3 cell line by 4C analysis supported the linking of this region to the immediate *SLC6A14* promoter (Fig. 3B). Second, while the SNPs across this segment showed very strong LD, a further partitioning appeared evident by multiple eQTLs with the smallest p-values ($\leq5\times10^{-4}$) toward the more distal 6-7 kb sub-portion where the association with lung function was also strong (Fig. 3C). Third, multiple transcription factor motifs were present, consistent with the notable number of cCREs, as determined by the JASPAR database (Fig. S2C, upper panel). A compelling candidate feature included the alleles of the rs4446858 SNP with T reference and C alternate alleles. The T allele occurs within a potential Interferon Regulatory Factor (IRF) 1 motif, whereas the alternate allele C would provide stronger motif consensus (Fig. 3D). IRF1 and 2 motifs partially overlap and their corresponding transcription factor activities can counter each other depending on exposure context (Figs. 3D and S2C, lower panel). IRF2 is a general repressive factor (with some exceptions), whereas IRF1 is activating and responsive to interferon stimulation where consequent activities involve engagement in multiple cell pathways including immune responses (Tanaka et al. 1993; Wang et al. 2024). The rs4446858 SNP occurs within the annotated cCRE designated as E2765449/enhD and also displayed significant association with lung function ($p=1.69\times10^{-9}$, Table 1). Further, a ChIP-Seq study identified IRF1 binding at this position in pancreatic adenocarcinoma cells (Diaferia et al. 2016). A nearby cCRE designated as E2765448/enhD contained another SNP, rs12009976 with comparable association with lung function but no transcription factor motifs were assigned directly to this position (Fig. S2C, lower panel and Tables 1 and 2). Of direct relevance for proximal airway tissue, the ATAC-Seq findings in the HBE-ALI cultures suggested that the chromatin structure is open in this region. Taken together, a prioritized segment limited to 2.1 kb (highlighted in green within the 17 kb blue region, Fig. 3C) was then selected for more detailed assessment given the strong support from lung function GWAS variants and HNE eQTLs as well as the inclusion of the rs4446858 SNP directly implicated in transcription factor binding (Fig. S2C).

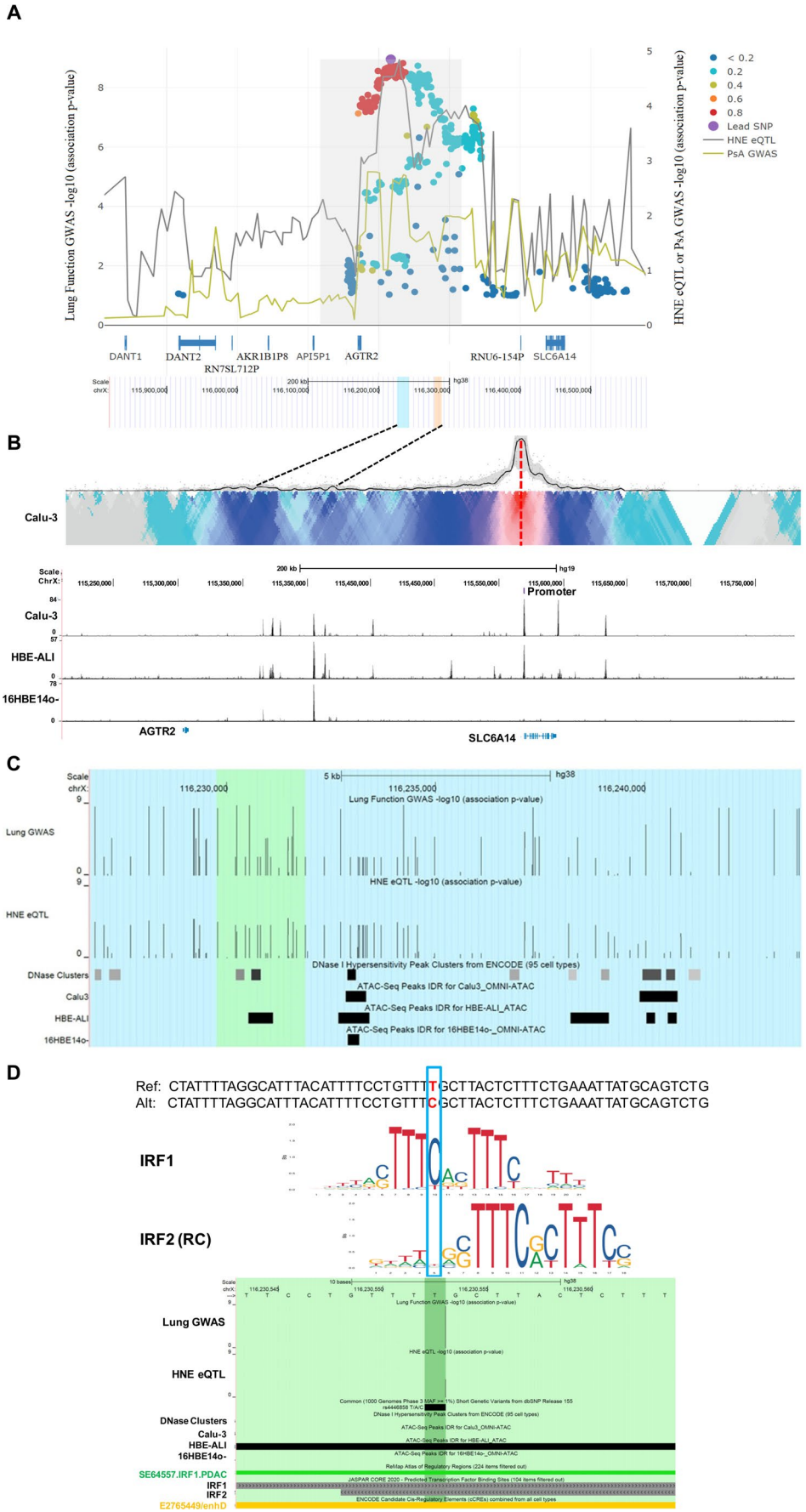


Fig. 3 The upstream locus associated with lung function in individuals with CF harbours regulatory features. **A** The schematic with annotated genes below highlight the *p*-values of the individual SNPs (coloured dots) from the lung function GWAS are shown overlying those of HNE eQTLs (continuous grey line), where the right and left y-axes indicate the respective *p*-value with -log10 scales. The dot colours indicate LD with the rs7879546 SNP (purple) with the smallest association *p*-value. The PsA association analysis is indicated by the solid yellow line. The pale grey-shaded portion of the plot indicates the region of SNPs (±0.1 Mb centred on the most significant Lung GWAS SNP) used to perform the SS colocalization tests. The 10 and 17 kb sub-regions are marked with orange and blue shadings. **B** Long range interaction of the GWAS association neighbourhood is shown in the upper panel with the 4C-seq domainogram from the *SLC6A14* promoter viewpoint, marked by the red vertical dashed line. The main trend of contact profile using a 5 kb window size is shown as a solid black line above the domainogram where relative interactions are normalized to the strongest point (which is set to 1). The domainogram uses color coded intensity values to show relative interactions with window sizes varying from 2 to 50 kb. Red denotes the strongest interactions and dark blue, through turquoise, to gray represent decreasing frequencies. Shown below are ATAC-seq tracks across the same genomic region in the indicated cells. **C** Features of the 17 kb sub-region are shown in refined tracks; including association *p*-values for GWAS SNPs for lung function and eQTLs for *SLC6A14* expression with –log10 scales, DNaseI hypersensitive clusters and ATAC-Seq peaks for cell lines and HBE-ALI as indicated. The 2.1 kb segment is shown with green shading. **D** The immediate sequence including the rs4446858 SNP within the prioritized 2.1 kb segment retaining annotations as in (**C**), is shown with the alternate C allele aligned directly with an IRF1 binding consensus motif. *RC* reverse complement.

Table 1 Top 15 significant SNPs from the CF lung function GWAS study and corresponding eQTLs in HNEs

SNP	Pos[a]	Ref[b]	Alt[c]	Lung Est[d]	Lung *P*[e]	eQTL Est[d]	eQTL *P*[f]	Region[g]
rs7879546	116,217,020	T	C	0.0726	1.11e–9	– 0.1807	1.65e–4	No
rs12559259	116,217,066	A	G	0.0723	1.20e–9	– 0.19532	2.93e–5	No
rs12556757	116,217,035	G	A	0.0724	1.43e–9	– 0.18264	1.16e–4	No
rs6520193	116,234,232	T	C	0.0708	1.50e–9	– 0.1807	1.65e–4	Blue
rs12009976	116,230,240	G	A	0.0706	1.65e–9	– 0.16848	5.08e–4	Blue
rs12559834	116,217,159	T	C	0.0711	1.66e–9	– 0.17972	1.88e–4	No
rs4446858	**116,230,553**	**T**	**C**	**0.0705**	**1.69e–9**	**– 0.1807**	**1.65e–4**	**Blue**
rs5952223	116,255,308	C	T	0.079	1.83e–9	– 0.07319	1.76e–1	No
rs7052638	116,217,348	A	C	0.0715	1.84e–9	– 0.19532	2.93e–5	No
rs5905350	116,225,152	T	C	0.0704	1.86e–9	– 0.1807	1.65e–4	No
rs5905340	116,222,048	T	C	0.0704	2.02e–9	– 0.1807	1.65e–4	No
rs10854906	116,228,156	C	A	0.0705	2.08e–9	– 0.20017	1.95e–5	Blue
rs4468049	116,211,666	C	G	0.0701	2.16e–9	– 0.1557	1.37e–3	No
rs5905214	116,228,498	C	T	0.0703	2.32e–9	– 0.18869	5.77e–5	Blue
rs5905282	116,209,043	C	T	0.0703	2.33e–9	– 0.19415	3.82e–5	No

[a]SNPs are located on chromosome X; positions are based on genome assembly GRCh38, [b]Reference allele, [c]Alternate allele, [d]Lung function GWAS and eQTL SNP effect estimates are given for the alternate alleles, [e]Lung GWAS *p*-value, [f]eQTL analysis *p*-value, [g]SNP location as blue indicates within prioritized 17 kb region, features of SNP in bold are shown in Fig. 3D

An enhancer from the distal GWAS locus contributes to the expression of *SLC6A14*

The prioritized 2.1 kb segment was first investigated for regulatory contribution to *SLC6A14* transcription via reporter gene assay. Insertion of the fragment (containing only the reference alleles of all sixteen eQTLs present, including the T allele of rs4446858, Table 2) into the pGL4.10 luciferase reporter expression vector alone did not drive response (Fig. S3), but its juxtaposition adjacent to the core promoter element (381 bp) or the trimmed promoter fragments containing only the proximal rs12710569 SNP (878 bp) or the rs12710569 and rs3788766 SNP combination (1106 bp; as described above, Fig. 2A) led to marked increase in reporter gene expression. Expression levels exceeded the core or trimmed promoter elements alone by 10- to 45-fold in the MOR cell line (Fig. 4A, right panel). Enhancement was achieved with either forward or reverse orientations, consistent with a classical enhancer definition. In direct contrast, no enhancement was observed in the CFPAC-1 cell line (Fig. 4A, left panel). The CFPAC-1 cell line is not refractory to reporter gene enhancement as activation of the core or trimmed promoter segments could be detected with a general alternate SV40 enhancer (Herr 1993) in a series of parallel vectors (Fig. S4). The distinction in cell line response with the distal upstream SLC6A14 segment would appear to reflect presence of at least one critical factor or binding complex in the lung-derived MOR cell line, but not in the pancreas-derived CFPAC-1 cell line, that could contribute gene expression regulation in airway-derived or lung tissues.

The distal enhancer element responds to IRF1 and IFN-γ

To assess underlying sequence differences in the putative enhancer that may explain genetic influence in lung and HNEs, the local haplotypes of the SNP alleles in the 2.1 kb segment of 6365 individuals with CF were considered. These SNPs exhibited notable eQTL effects for SLC6A14 (Table 2A), but due to extensive LD, only four sets of allele combinations explain > 99% of all haplotypes (named D1 to D4; Table 2B). Of these, D1 and D4 differ at only two of the sixteen analyzed SNPs, rs12009976 and rs4446858.

Activity of the 2.1 kb segment with corresponding D1 and D4 haplotypes, as well as a D1/D4 combination to specifically target the C allele of rs4446858, was then assessed. Consistent strong enhancement, with no obvious difference in capacities of the allele combinations in the MOR cell line was evident (Fig. 4B). Under the conditions used however, only basal transcription factor levels would be present, and response difference may require a stimulated situation. IRF genes encode transcription factors that are key effectors of interferons (reviewed in Wang et al. 2024). Before considering effects of interferons, we directly tested for effects of exogenous IRF1 and IRF2 factors via expression vectors added directly in the reporter gene assays. Addition of an IRF1 expression vector with the enhancer/core promoter reporter vector yielded heightened response compared to no addition (pCMV vehicle only), or to only IRF2 addition. The effect was dampened if IRF2 was added in addition to IRF1 (Fig. 4C). Notably, these effects were observed in the presence of the rs4446858 C allele and were irrespective of the rs12009976 allele. The findings implicate the rs4446858 allele in IRF1 binding with potential manifestation in

Table 2 *SLC6A14* eQTL associations and haplotypes of the prioritized 2.1 kb segment

A. Position and *SLC6A14* eQTL associations from HNE cells					
SNP	Pos[a]	Ref	Alt	eQTL Est	eQTL *P*
rs4824330	116,229,787	A	G	– 0.195045	2.69e–5
rs17095260	116,230,215	G	A	– 0.229871	8.05e–5
rs12009976	116,230,240	G	A	– 0.168484	5.07e–4
rs60405377	116,230,456	A	T	– 0.229871	8.05e–5
rs4446858	**116,230,553**	**T**	**C**	**– 0.180704**	**1.64e–4**
rs5952115	116,230,755	T	C	– 0.229871	8.05e–5
rs5952116	116,230,820	C	T	– 0.229871	8.05e–5
rs5905357	116,230,940	T	C	– 0.0621687	2.76e–1
rs5905215	116,230,979	G	A	– 0.0621687	2.76e–1
rs5952203	116,231,057	A	T	– 0.229871	8.05e–5
rs5905359	116,231,469	C	A	– 0.0649908	2.56e–1
rs4579644	116,231,477	G	A	– 0.233367	6.77e–5
rs5905360	116,231,568	T	G	– 0.0621687	2.76e–1
rs4427295	116,231,594	C	A	– 0.195045	2.69e–5
rs4623576	116,231,674	C	T	– 0.215003	3.32e–4
rs5952204	116,231,888	T	G	– 0.229871	8.05e–5

B. Haplotypes of contained SNP alleles and frequency			
Haplotype[b]	*N*	Percentage[c]	Haplotype code
0000000000000000	6511	51.13	D1
1111111001010111	2981	23.41	D2
1010100110101100	2685	21.09	D3
0010100000000000	495	3.89	D4
1010100000101100	49	0.38	D5
1010100000000100	4	0.03	D6
1010100000100100	3	0.02	D7
0010000000000000	2	0.02	D8
0000100110101100	1	0.01	D9
1010000110101100	1	0.01	D10
1010100110100100	1	0.01	D11
1111101001010111	1	0.01	D12

[a]Chromosome positions are from genome assembly version GRCh38, [b]Haplotypes of the 16 listed SNPs are shown left to right with reference (0) or alternate alleles (1), [c]Proportion (percentage) of sample with given haplotype identified from SNP typing of 6365 individuals of the CF lung function GWAS (Corvol et al. 2015; Gong et al. 2019)

immune (and presumably other) responses that could result in increases in *SLC6A14* expression in relevant stimulated states.

To induce IRF1 levels in the context of living cells, we then considered whether type I or II interferons (IFNs) could provide stimulus for the SNP response, with potential to then contribute to immune processes and perhaps impact lung function. The rapid interferon signaling of an innate immune response would readily boost IRF1 levels upon pathogen engagement. Indeed, addition of IFN-γ, but not IFN-α or IFN-β to the culturing media of MOR cells that did contain the alternate C allele at rs4446858 (data not shown) resulted in increased IRF1 levels most evident at 6 h, that waned by 24 h of treatment. IRF2 remained below detection with all treatments (Fig. 5A). Further, stimulation of mature SLC6A14 protein levels was then also evident by 24 h following onset of treatment with IFN-γ (Fig. 5B, C).

While the newly identified distal enhancer provides a key determinant to drive expression of SLC6A14, the rs4446858 SNP appears to direct an acute effect with prompted increase in SLC6A14 expression involving IRF1/IFN-γ stimulation.

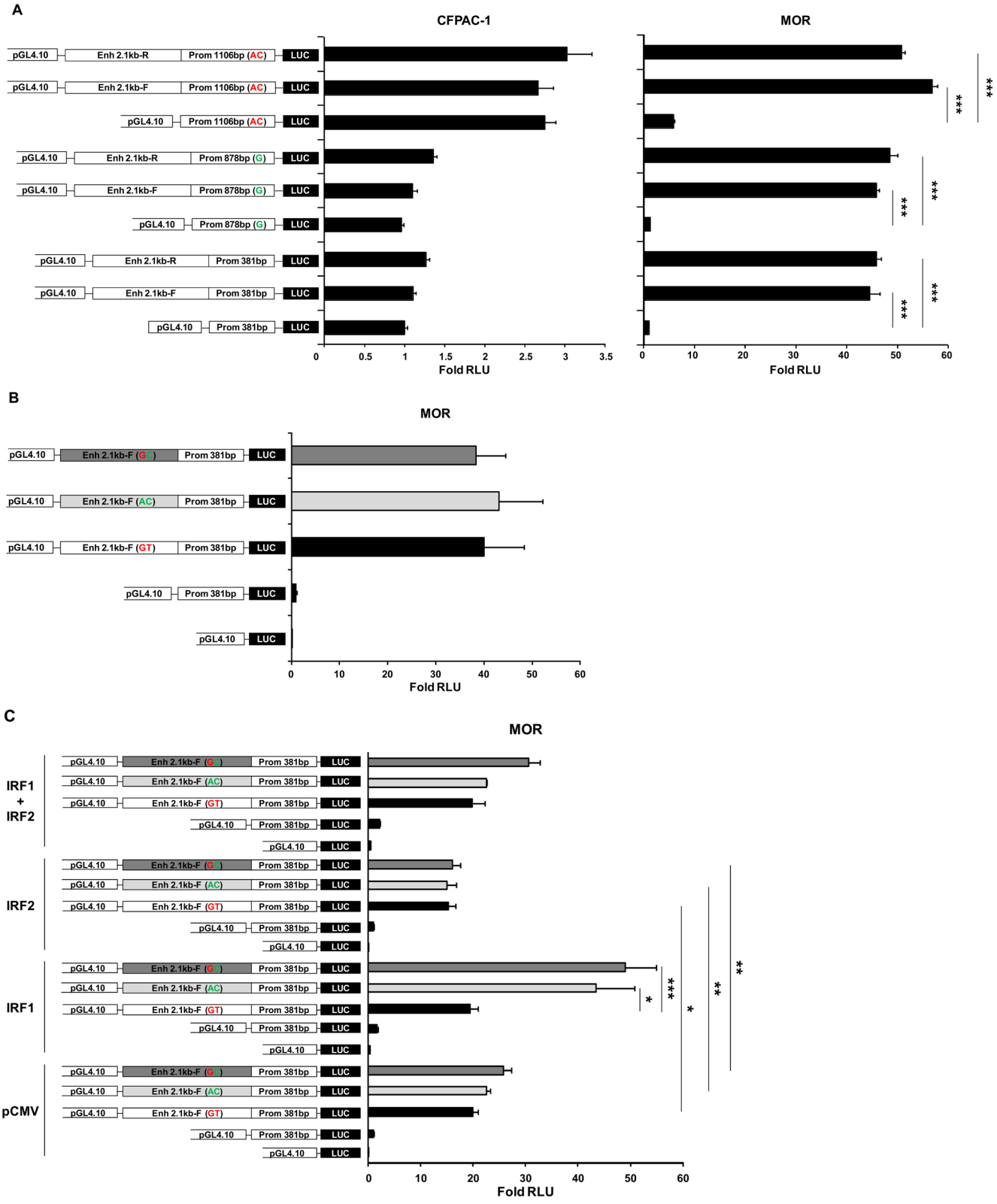
SLC6A14 co-expresses with IRF1 and IRF2

As SLC6A14 responds to stimulation by IRF1, driven by IFN-γ in the MOR cell line, we then sought to examine its expression together with *IRF1* and *IRF2* in vivo. *IRF1/2* genes are expressed in many cells types, whereas *SLC6A14* is most prominently expressed in airway and gastrointestinal epithelia. Using bulk and single cell RNA-Seq datasets, we evaluated *SLC6A14*, *IRF1* and *IRF2* expression specifically in primary nasal and proximal airway tissues, respectively. There was significant co-expression between SLC6A14 and IRF1 and SLC6A14 and IRF2 with effect estimates showing positive association of expression levels in naïve HNEs (Table 3). Recent single-cell lung expression studies include ones with sufficient depth to critically define multiple cell types in the adult upper airway. When assessing *SLC6A14* in the twelve defined epithelial cell types reported in the proximal airway by Travaglini and colleagues (Travaglini et al. 2020), expression was noted to be in 'differentiating basal' and 'club' cells at low levels with 7.8 and 3.2 TPM, respectively, and in 'goblet' cells at 28.6 TPM. This expression pattern is consistent with the derivation of goblet cells from club and basal cells (Davis and Wypych 2021). Co-expression was evident with both *IRF* genes in the goblet cells (Table 3), with the effect estimates showing positive association relationships as observed with the HNE bulk RNA-Seq results, indicating that this cell type of the proximal airway could contribute to IFN-γ and immune responses involving SLC6A14.

Discussion

Earlier GWAS studies in individuals with CF had implicated SNPs in the immediate and distant upstream regions of *SLC6A14* as contributing to disease phenotypes in diverse tissues (Gong et al. 2019). The eQTLs within these regions highlighted that less SLC6A14 expression aligns with better disease outcome.

The functional analyses of the immediate SNPs in their haplotype contexts defined the proximal promoter region in both pancreas- and lung-derived cell models, and supported that MI-associated SNPs, specifically the alternate alleles for rs3788766 and rs12710568 align with more expression and higher MI risk. The protective reference G allele for



rs3788766 alone has also pointed to less expression using a comparable reporter gene assay system in modified Calu-3 cells (Mercier et al. 2022). While MI occurs in the intestine (Gibson-Corley et al. 2016), an indirect role for SLC6A14 in the pancreas could reflect disturbed digestive enzyme release and activation due to abnormal pancreatic ductal

Fig. 4 Distal enhancer activity with augmented response to IRF1 depending on rs4446858 SNP allele. **A** Juxtaposition of the 2.1 kb segment of the 17 kb sub-region in forward (F) or reverse (R) orientation to the core promoter increased reporter gene activity in MOR cells, but not in CFPAC-1 cells. Luciferase activity values were determined as in Fig. 2, except that activity of core promoter only was set as 1 for comparisons. Two independent experiments were performed, each with two to four technical replicates for each vector. **B** The 2.1 kb segment with the most common SNP combination (haplotype D1; black), with alternate alleles A and C for rs12009976 and rs4446858, respectively (haplotype D4; pale) and with the alternate C allele for rs4446858 only (grey) yielded comparable reporter gene activities. Each bar represents the mean of five to ten replicates from two independent experiments. **C** Inclusion of exogenous IRF1 by co-transfection revealed further enhancement of the reporter gene activity with the alternate C allele of rs4446858. The exogenous inclusion of IRF2 modestly reduced enhancer activity, compared to the pCMV control vector only, and was independent of the C allele. A combination of IRF2 with IRF1 reduced the effect of IRF1 alone. The reporter gene assays were performed as in (**A**) except the core promoter fragment with pGL4.10 plus pCMV was set as 1 for comparisons. Each bar represents mean of six to twelve replicates from two independent experiments. For all experiments error bars represent SEM. $*P<0.05$, $**P<0.01$, $***P<0.001$ using unpaired t tests.

fluid (Venglovecz et al. 2021). This aligns with reported correlations between CFTR-causal variants and residual pancreatic function and occurrence of MI (Dupuis et al. 2016). The relatively early development and presentation of MI implicate influence *in utero*, at a time when there is already pathology in the CF pancreas. However, there remains limited information on endogenous SLC6A14 expression and its relationship to CFTR function. From single cell RNA sequencing studies of human fetal pancreas tissue, *CFTR* expression initiates with early exocrine differentiation well before digestive enzyme expression, but *SLC6A14* expression is more restricted and only seen after exocrine ductal cell emergence at ten to eleven weeks following conception (Ma et al. 2023).

The lung GWAS, involving cases from multiple centers with a wide range in ages did not show evidence of association between the SNPs and lung function in the proximal promoter region of *SLC6A14* (Corvol et al. 2015; Gong et al. 2019), but any modest expression alteration may have been influenced by a relatively stronger effect of the more distal region, with notable enhancer features. At least a couple of earlier studies of CF lung disease (Li et al. 2014; Pereira et al. 2017) had indicated some association for the proximal rs3788766 SNP, but these studies included fewer subjects and concentrated mostly on younger individuals with CF. Given the complexity and enduring disease progression in the CF lung, it remains a possibility that there is genetic influence at the proximal locus in the airways, but only at younger ages.

The large distal GWAS region implicated for lung phenotypes contains no protein coding genes but many *SLC6A14* eQTLs pointing to complexity in regulation of expression. We showed that at least one prioritized and conserved segment from this interval, positioned ~ 205 kb upstream of the *SLC6A14* TSS, exhibits strong enhancer activity for the proximal promoter in a lung-derived cell model. While signatures of chromatin accessibility for transcriptional machinery and long-range interactions were noted in HBE-ALI cultures, further assessment would be helpful to confirm tissue specificity of the enhancer. The context of regulatory function and the relationship of the identified enhancer with other local and neighboring sequences also remain unknown. Additional sub-regions, with evidence of open chromatin and the cCRE annotations that are derived from various cell lines and tissues do support further contributors for airway tissue and for regulatory complexity in other epithelial tissues. Indeed, alternate upstream contributors to regulation of *SLC6A14* have been suggested (Barilli et al. 2021; Di Paola et al. 2017). In addition, based on GTEx data (release V10), a set of at least six eQTLs occur within the distal GWAS interval for *SLC6A14* in minor salivary gland, where four overlap lung eQTLs and all are more distal of the prioritized 17 kb sub-region.

The initial *SLC6A14* eQTLs were calculated with RNA sampling from tissues of non-CF and overall healthy individuals while HNE analysis involved tissue directly from individuals with CF; both yielded consistent genetically-determined directions of effect for *SLC6A14* expression. While the enhancer itself would appear to be a prominent determinant of expression in the lung, the contained rs4446858 SNP challenges how SLC6A14 actually contributes to variation in lung-related phenotypes. Although many of the alternate *SLC6A14* eQTL alleles of the entire upstream locus including some within the enhancer segment contribute to stronger lung function via negative effects on expression (Tables 1 and 2A), the alternate C allele of rs4446858 with an allele frequency of 47% in European populations appears to exert effects on the IRF1/2 binding site while supporting further increased expression involving response to IRF1. It would thus appear that an immediate situation, determined by IFN-γ presence and stimulation in our experimental setting, adds a level of regulation to the role of the enhancer where the GWAS analysis then only reflected a net steady state without (or with limited) stimulation where less *SLC6A14* expression is predominant, and in line with observing better cumulative lung function.

A direct role for SLC6A14 has already been demonstrated when CF cells encounter pathogens, where arginine transport has been implicated in altering the epithelial surface liquid layer of the airway affecting early phases of bacterial infection (Barilli et al. 2021; Di Paola et al. 2017). Being a major contributor to the import of neutral and cationic amino acids, including arginine, SLC6A14 has also been implicated in other and multiple airway diseases. Arginine

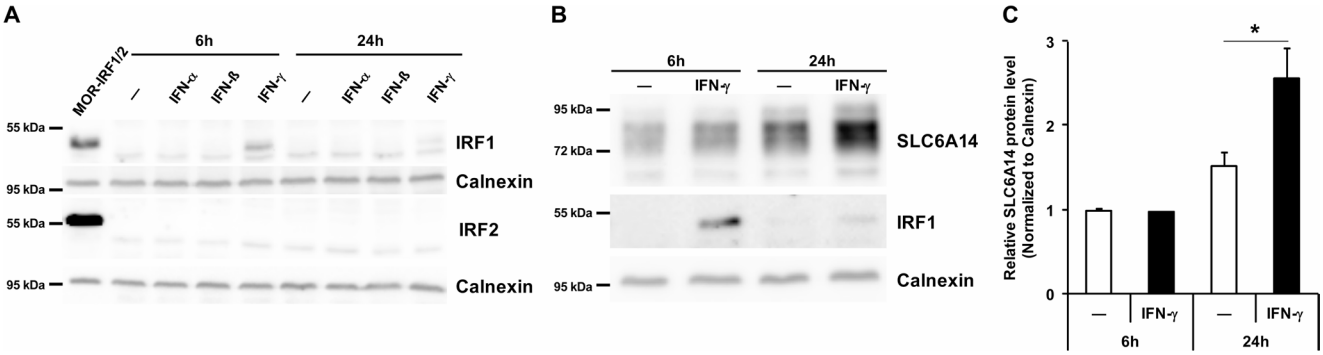


Fig. 5 IFN-γ induces IRF1 and SLC6A14. **A** MOR cells at ~50% confluence were treated with or without recombinant interferons (10ng/ml) for 6 or24 h durations. Immunoblot analysis of whole cell extracts were carried out with IRF1, IRF2 and Calnexin (control) antibodies. IRF1 and IRF2 antibodies were validated with extracts of MOR cells transfected with expression vectors without IFN treatment in far left lanes. **B** MOR cells were treated similarly with IFN-γ only to assess SLC6A14 expression; the immunoblot is a representative of three independent experiments, shown summarized in (**C**). **C** Bar graph indicates the mean (± SEM) of SLC6A14 abundance after normalization to Calnexin. The value of (-, 6 h) was arbitrarily set at 1, for comparisons. Data are from three experiments. *$P<0.05$ using an unpaired t test

Table 3 Co-expression of *SLC6A14* with *IRF1* and *IRF2*

Analysis	Tissue	Model	Effect estimate	*P*-value	Adjusted *P*[a]
Bulk RNA-Seq	HNE	IRF1	0.620020487	8.81e–10	1.76e–9
		IRF2	0.425270729	3.53e–5	7.06e–5
Single cell RNA-Seq	Goblet cells	IRF1	0.000340471	1.32e–11	5.25e–9
		IRF2	0.00017935	2.86e–5	2.01e–3
	Basal	IRF1	–0.030758985	5.09e–2	7.71e–1
		IRF2	–0.180431605	2.32e–1	1
	Ciliated	IRF1	0.000890285	2.45e–1	1
		IRF2	0.000408788	7.78e–1	1
	Club	IRF1	0.000283596	1.95e–1	1
		IRF2	–0.001583417	2.91e–1	1
	Neuroendocrine	IRF1	NA[b]	NA	NA
		IRF2	NA	NA	NA
	Ionocyte	IRF1	NA	NA	NA
		IRF2	NA	NA	NA
	Differentiating basal	IRF1	0.003644613	4.36e–1	1
		IRF2	0.00791775	1.17e–1	1

[a]Adjusted *p* : Bonferroni corrected *p*-values

[b]NA: data not available due to very low expression or insufficient cell numbers

metabolism is particularly relevant in airway inflammation involving nitric oxide and related arginase pathways (Asosingh et al. 2020; Barilli et al. 2021; Li et al. 2025).

To more fully understand the contribution of SLC6A14 to modification of airway phenotypes, interpretation will require understanding of the specific CF lung environment, including its progressive disease aspects. The CF airway involves immunodeficient mucosa (Cohen and Prince 2012) with serious and chronic infections in a hyperimmunoreactive milieu with impaired and disturbed innate and humoral responses (Bojanowski et al. 2021). It had been suggested that there may be increased expression of *SLC6A14* in severe CF lung disease versus healthy controls (Auslander et al. 2020) as also seen in the intestine of CF mice (Ikpa et al. 2020), however subsequent studies involving primary nasal samples from pediatric individuals with CF versus non-CF (Loske et al. 2024), and lung samples from older individuals with late stage CF versus non-CF (Carraro et al. 2021) show reduced expression of SLC6A14 in airway secretory epithelial cell types. These studies also emphasize the need to consider the possibility of altered cell component makeup as well as the effects of progressive disease in barrier airway epithelia function with CF (Carraro et al. 2021; Gustafsson and Hansson 2025). Of interest, the studies of nasal cells from pediatric patients highlighted the marked disturbance of interferon response pathways well before lung function would be markedly deteriorated (Loske et al. 2024) and thus a careful look at alteration in IFN-γ signaling reflecting genetic determination and variability would be warranted already at these early stages.

While treatments may alleviate symptoms, disease progression in the CF lung reflects a complex interplay of chronic infections, repeated cycles of infection and inflammation involving immune and tissue repair responses (Turcios 2020). Additional complications imposed by viral co-infections further aggravate progression (Kiedrowski and Bomberger 2018). Many studies of interferons and interferon responsive factors have highlighted their roles in immunity well beyond their long-established functions in combating viral infections. They facilitate immune responses but also contribute to supporting or hindering tissue damage responses depending on context (Boehmer and

Zanoni 2025). IFN-γ has substantial influence on multiple immune cell types and their respective cytokine production (Lee and Ashkar 2018). It has been observed that CFTR itself may respond to IRF1 stimulation based on cell model studies, but any direct consequent effects are presumably further hindered in the CF disease state (Zhang et al. 2013). We observed that SLC6A14 co-expresses with both IRF1 and 2 in goblet cells, and propose that at least short term enhanced expression is favorable, toward the dampening of inflammation. Better definitions of the states reflecting what are typically transitory responses of IFN-γ and IRF1 within the hyperinflammatory environment of the CF lung are needed to appreciate influence by genetic variation. This will enable better understanding of the relationship between SLC6A14 and lung function measures and cumulative disease. While acute and highly dynamic features are likely difficult to capture, the GWAS studies of large populations of individuals with CF do point to protective benefits of less expression for lung function, a direction that was also noted for the gastrointestinal phenotype involving reduced MI risk.

Conclusions

Our findings demonstrate that previously implicated associated SNP regions supported by colocalization analyses directly contribute to *SLC6A14* expression variation. The far upstream region includes cell/tissue-type differential regulation that explains the distinct locations for association with the different phenotypes studied. In the distal region, at least one potent regulatory element together with the proximal promoter enhances *SLC6A14* expression consistent with the strong expression in upper airway tissue including primary HNEs. While the population investigations point to benefits in lung function of less expression of SLC6A14, the investigations involving the IFN-γ cytokine and IRF1 stimulation support a complexity of regulation consistent with a favorable role of SLC6A14 in acute immunoregulatory responses involving goblet cells in individuals with CF harbouring the alternate C allele of the rs4446856 SNP. Lung function in CF is a complex measure reflecting a progressive multitude of effects including inflammation and its resolution, as well as damage and repair affecting residual tissue integrity and function. More precise delineation of when and how altered levels of SLC6A14 are beneficial or detrimental is needed to resolve and understand all of the effects of this modifier on CF disease. Our studies further emphasize the challenge of elucidating the contributions of GWAS loci where there are multiple levels of regulation with interacting and/or competing contributions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00439-025-02800-7.

Acknowledgements We thank H. Swahn, M. NandyMazumdar and S. Yin for generating some of the data shown in Fig. 3B. We also thank the CF participants, care providers, and clinic coordinators at CF Centers throughout Canada for their contributions to the CF Patient Registry and the CGMS.

Author contributions M.E. designed and conducted the experiments (with support from F.L. and A.C.), interpreted the data and drafted the manuscript. C.W. and G.H. analyzed RNA-seq data. CW performed eQTL analysis for HNE cells, performed the co-localization analysis and carried out the co-expression analyses. N.P. carried out genome bioinformatic analyses. A.H. provided the long range interaction data. K.K. and J.A. coordinated recruitment and collection for the CGMS. J.M.R. and L.J.S. conceptualized and supervised the work. L.J.S. secured funding for the project and oversees the CGMS. All authors contributed and read the manuscript.

Funding Funding was provided by a 2022 Basic and Clinical Research Grant (1009794) jointly funded by Cystic Fibrosis Canada (608828) and Canadian Institutes of Health Research-Institute of Circulatory and Respiratory Health (FRN: BCG 187014), Cystic Fibrosis Foundation (STRUG17PO), Canadian Institutes of Health Research Foundation Grant (FRN-167282), and by the Government of Canada through Genome Canada and Ontario Genomics Institute (OGI-148). This research was undertaken, in part, with funding from the Canada Research Chairs Program to L.J. Strug who is the Canada Research Chair in Genome Data Science.

Data availability Data and materials can be obtained as previously described (Gong et al. 2022) and by contacting the corresponding author. Code availability including the Simple Sum colocalization method and the web-accessible Locus Focus tool have been published previously, as referenced.

Declarations

Competing interests The authors declare no competing interests.

Ethics approval and consent to participate Current study approvals were obtained from the Research Ethics Board of The Hospital for Sick Children. Written informed consent was obtained from all participants or their guardians/parents. Related collection processes have been described previously (Gong et al. 2022). Direct consent for publication is not applicable.

References

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. Nature 526:68–74. https://doi.org/10.1038/nature15393

Ahmadi S, Wu YS, Li M, Ip W, Lloyd-Kuzik A, Di Paola M, Du K, Xia S, Lew A, Bozoky Z, Forman-Kay J, Bear CE, Gonska T (2019) Augmentation of cystic fibrosis transmembrane conductance regulator function in human bronchial epithelial cells via SLC6A14-dependent amino acid uptake. Implications for treatment of cystic fibrosis. Am J Respir Cell Mol Biol 61:755–764. https://doi.org/10.1165/rcmb.2019-0094OC

Alicandro G, Gramegna A, Bellino F, Sciarrabba SC, Lanfranchi C, Contarini M, Retucci M, Dacco V, Blasi F (2024) Heterogeneity in response to Elexacaftor/Tezacaftor/Ivacaftor in people with cystic fibrosis. J Cyst Fibros 23:1072–1079. https://doi.org/10.1016/j.jcf.2024.04.013

Asosingh K, Lauruschkat CD, Alemagno M, Frimel M, Wanner N, Weiss K, Kessler S, Meyers DA, Bennett C, Xu W, Erzurum S (2020) Arginine metabolic control of airway inflammation. JCI Insight 5:e127801. https://doi.org/10.1172/jci.insight.127801

Auslander N, Ramos DM, Zelaya I, Karathia H, Crawford TO, Schaffer AA, Sumner CJ, Ruppin E (2020) The GENDULF algorithm: mining transcriptomics to uncover modifier genes for Monogenic diseases. Mol Syst Biol 16:e9701. https://doi.org/10.15252/msb.20209701

Barilli A, Visigalli R, Ferrari F, Borsani G, Dall'Asta V, Rotoli BM (2021) Flagellin from *Pseudomonas aeruginosa* stimulates ATB(0,+) transporter for arginine and neutral amino acids in human airway epithelial cells. Front Immunol 12:641563. https://doi.org/10.3389/fimmu.2021.641563

Boehmer D, Zanoni I (2025) Interferons in health and disease. Cell 188:4480–4504. https://doi.org/10.1016/j.cell.2025.06.044

Bojanowski CM, Lu S, Kolls JK (2021) Mucosal immunity in cystic-fibrosis. J Immunol 207:2901–2912. https://doi.org/10.4049/jimmunol.2100424

Carraro G, Langerman J, Sabri S, Lorenzana Z, Purkayastha A, Zhang G, Konda B, Aros CJ, Calvert BA, Szymaniak A, Wilson E, Mulligan M, Bhatt P, Lu J, Vijayaraj P, Yao C, Shia DW, Lund AJ, Israely E, Rickabaugh TM, Ernst J, Mense M, Randell SH, Vladar EK, Ryan AL, Plath K, Mahoney JE, Stripp BR, Gomperts BN (2021) Transcriptional analysis of cystic fibrosis airways at single-cell resolution reveals altered epithelial cell States and composition. Nat Med 27:806–814. https://doi.org/10.1038/s41591-021-01332-7

Cohen TS, Prince A (2012) Cystic fibrosis: a mucosal immunodeficiency syndrome. Nat Med 18:509–519. https://doi.org/10.1038/nm.2715

Conomos MP, Miller MB, Thornton TA (2015) Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol 39:276–293. https://doi.org/10.1002/gepi.21896

Corvol H, Blackman SM, Boelle PY, Gallins PJ, Pace RG, Stonebraker JR, Accurso FJ, Clement A, Collaco JM, Dang H, Dang AT, Franca A, Gong J, Guillot L, Keenan K, Li W, Lin F, Patrone MV, Raraigh KS, Sun L, Zhou YH, O'Neal WK, Sontag MK, Levy H, Durie PR, Rommens JM, Drumm ML, Wright FA, Strug LJ, Cutting GR, Knowles MR (2015) Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. Nat Commun 6:8382. https://doi.org/10.1038/ncomms9382

Davis JD, Wypych TP (2021) Cellular and functional heterogeneity of the airway epithelium. Mucosal Immunol 14:978–990. https://doi.org/10.1038/s41385-020-00370-7

DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G (2012) RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 28:1530–1532. https://doi.org/10.1093/bioinformatics/bts196

Di Paola M, Park AJ, Ahmadi S, Roach EJ, Wu YS, Struder-Kypke M, Lam JS, Bear CE, Khursigara CM (2017) SLC6A14 is a genetic modifier of cystic fibrosis that regulates Pseudomonas aeruginosa attachment to human bronchial epithelial cells. mBio 8:e02073–e02017. https://doi.org/10.1128/mBio.02073-17

Diaferia GR, Balestrieri C, Prosperini E, Nicoli P, Spaggiari P, Zerbi A, Natoli G (2016) Dissection of transcriptional and cis-regulatory control of differentiation in human pancreatic cancer. EMBO J 35:595–617. https://doi.org/10.15252/embj.201592404

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Dupuis A, Keenan K, Ooi CY, Dorfman R, Sontag MK, Naehrlich L, Castellani C, Strug LJ, Rommens JM, Gonska T (2016) Prevalence of meconium ileus marks the severity of mutations of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Genet Med 18:333–340. https://doi.org/10.1038/gim.2015.79

ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A, Kaul R, Halow J, Van Nostrand EL, Freese P, Gorkin DU, Shen Y, He Y, Mackiewicz M, Pauli-Behn F, Williams BA, Mortazavi A, Keller CA, Zhang XO, Elhajjajy SI, Huey J, Dickel DE, Snetkova V, Wei X, Wang X, Rivera-Mulia JC, Rozowsky J, Zhang J, Chhetri SB, Victorsen A, White KP, Visel A, Yeo GW, Burge CB, Lecuyer E, Gilbert DM, Dekker J, Rinn J, Mendenhall EM, Ecker JR, Kellis M, Klein RJ, Noble WS, Kundaje A, Guigo R, Farnham PJ, Cherry JM, Myers RM, Ren B, Graveley BR, Gerstein MB, Pennacchio LA, Snyder MP, Bernstein BE, Wold B, Hardison RC, Gingeras TR, Stamatoyannopoulos JA, Weng Z (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583:699–710. https://doi.org/10.1038/s41586-020-2493-4

Esmaeili M, Jennek S, Ludwig S, Klitzsch A, Kraft F, Melle C, Baniahmad A (2016) The tumor suppressor ING1b is a novel corepressor for the androgen receptor and induces cellular senescence in prostate cancer cells. J Mol Cell Biol 8:207–220. https://doi.org/10.1093/jmcb/mjw007

Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranasic D, Santana-Garcia W, Tan G, Cheneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW, Mathelier A JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 48:D87-D92. https://doi.org/10.1093/nar/gkz1001

Gibson-Corley KN, Meyerholz DK, Engelhardt JF (2016) Pancreatic pathophysiology in cystic fibrosis. J Pathol 238:311–320. https://doi.org/10.1002/path.4634

Gong J, Wang F, Xiao B, Panjwani N, Lin F, Keenan K, Avolio J, Esmaeili M, Zhang L, He G, Soave D, Mastromatteo S, Baskurt Z, Kim S, O'Neal WK, Polineni D, Blackman SM, Corvol H, Cutting GR, Drumm M, Knowles MR, Rommens JM, Sun L, Strug LJ (2019) Genetic association and transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci. PLoS Genet 15:e1008007. https://doi.org/10.1371/journal.pgen.1008007

Gong J, He G, Wang C, Bartlett C, Panjwani N, Mastromatteo S, Lin F, Keenan K, Avolio J, Halevy A, Shaw M, Esmaeili M, Cote-Maurais G, Adam D, Begin S, Bjornson C, Chilvers M, Reisman J, Price A, Parkins M, van Wylick R, Berthiaume Y, Bilodeau L, Mateos-Corral D, Hughes D, Smith MJ, Morrison N, Brusky J, Tullis E, Stephenson AL, Quon BS, Wilcox P, Leung WM, Solomon M, Sun L, Brochiero E, Moraes TJ, Gonska T, Ratjen F, Rommens JM, Strug LJ (2022) Genetic evidence supports the development of SLC26A9 targeting therapies for the treatment of lung disease. NPJ Genom Med 7:28. https://doi.org/10.1038/s41525-022-00299-9

Grasemann H, Ratjen F (2023) Cystic fibrosis. N Engl J Med 389:1693–1707. https://doi.org/10.1056/NEJMra2216474

GTEx Consortium (2015) Human genomics. The Genotype-Tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348:648–660. https://doi.org/10.1126/science.1262110

Gunawardena TNA, Bozoky Z, Bartlett C, Ouyang H, Eckford PDW, Moraes TJ, Ratjen F, Gonska T, Bear CE (2023) Correlation of electrophysiological and fluorescence-based measurements of modulator efficacy in nasal epithelial cultures derived from people with cystic fibrosis. Cells 12:1174. https://doi.org/10.3390/cells12081174

Guo J, Garratt A, Hill A (2022) Worldwide rates of diagnosis and effective treatment for cystic fibrosis. J Cyst Fibros 21:456–462. https://doi.org/10.1016/j.jcf.2022.01.009

Gustafsson JK, Hansson GC (2025) Immune regulation of goblet cell and mucus functions in health and disease. Annu Rev Immunol 43:169–189. https://doi.org/10.1146/annurev-immunol-101721-065224

Hammal F, de Langen P, Bergon A, Lopez F, Ballester B (2022) ReMap 2022: a database of Human, Mouse, drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. Nucleic Acids Res 50:D316–D325. https://doi.org/10.1093/nar/gkab996

He G, Panjwani N, Avolio J, Ouyang H, Keshavjee S, Rommens JM, Gonska T, Moraes TJ, Strug LJ (2022) Expression of cystic fibrosis lung disease modifier genes in human airway models. J Cyst Fibros 21:616–622. https://doi.org/10.1016/j.jcf.2022.02.007

Heijerman HGM, McKone EF, Downey DG, Van Braeckel E, Rowe SM, Tullis E, Mall MA, Welter JJ, Ramsey BW, McKee CM, Marigowda G, Moskowitz SM, Waltz D, Sosnay PR, Simard C, Ahluwalia N, Xuan F, Zhang Y, Taylor-Cousar JL, McCoy KS (2019) Efficacy and safety of the Elexacaftor plus Tezacaftor plus Ivacaftor combination regimen in people with cystic fibrosis homozygous for the F508del mutation: a double-blind, randomised, phase 3 trial. Lancet 394:1940–1948. https://doi.org/10.1016/S0140-6736(19)32597-8

Herr W (1993) The SV40 enhancer: transcriptional regulation through a hierarchy of combinatorial interactions. Seminars Virol 4:3–13. https://doi.org/10.1016/1044-5773(93)80003-7

Ikpa PT, Meijsen KF, Nieuwenhuijze NDA, Dulla K, de Jonge HR, Bijvelds MJC (2020) Transcriptome analysis of the distal small intestine of Cftr null mice. Genomics 112:1139–1150. https://doi.org/10.1016/j.ygeno.2019.06.028

Keating C, Yonker LM, Vermeulen F, Prais D, Linnemann RW, Trimble A, Kotsimbos T, Mermis J, Braun AT, O'Carroll M, Sutharsan S, Ramsey B, Mall MA, Taylor-Cousar JL, McKone EF, Tullis E, Floreth T, Michelson P, Sosnay PR, Nair N, Zahigian R, Martin H, Ahluwalia N, Lam A, Horsley A (2025) Vanzacaftor-tezacaftor-deutivacaftor versus elexacaftor-tezacaftor-ivacaftor in individuals with cystic fibrosis aged 12 years and older (SKYLINE trials VX20-121-102 and VX20-121-103): results from two randomised, active-controlled, phase 3 trials. Lancet Respir Med 13:256–271. https://doi.org/10.1016/S2213-2600(24)00411-9

Kerem B, Kerem E (1996) The molecular basis for disease variability in cystic fibrosis. Eur J Hum Genet 4:65–73. https://doi.org/10.1159/000472174

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–1080. https://doi.org/10.1126/science.2570460

Kiedrowski MR, Bomberger JM (2018) Viral-bacterial co-infections in the cystic fibrosis respiratory tract. Front Immunol 9:3067. https://doi.org/10.3389/fimmu.2018.03067

Kovalchuk V, Samluk L, Juraszek B, Jurkiewicz-Trzaska D, Sucic S, Freissmuth M, Nalecz KA (2019) Trafficking of the amino acid transporter B(0,+) (SLC6A14) to the plasma membrane involves an exclusive interaction with SEC24C for its exit from the Endoplasmic reticulum. Biochim Biophys Acta Mol Cell Res 1866:252–263. https://doi.org/10.1016/j.bbamcr.2018.11.005

Lee AJ, Ashkar AA (2018) The dual nature of type I and type II interferons. Front Immunol 9:2061. https://doi.org/10.3389/fimmu.2018.02061

Lee J, Shin MK, Ryu DK, Kim S, Ryu WS (2010) Insertion and deletion mutagenesis by overlap extension PCR. Methods Mol Biol 634:137–146. https://doi.org/10.1007/978-1-60761-652-8_10

Li W, Soave D, Miller MR, Keenan K, Lin F, Gong J, Chiang T, Stephenson AL, Durie P, Rommens J, Sun L, Strug LJ (2014) Unraveling the complex genetic model for cystic fibrosis: pleiotropic effects of modifier genes on early cystic fibrosis-related morbidities. Hum Genet 133:151–161. https://doi.org/10.1007/s00439-013-1363-7

Li H, Liang Y, Deng J, Cheng Y, Chen S, Lian X, Ouyang S (2025) Targeting arginine metabolism in immune cells for the treatment of pulmonary inflammatory diseases. Curr Allergy Asthma Rep 25:35. https://doi.org/10.1007/s11882-025-01216-7

Lin B, Gong J, Keenan K, Lin F, Lin YC, Mesinele J, Calmel C, Mohand Oumoussa B, Boelle PY, Guillot L, Corvol H, Waters V, Sun L, Strug LJ (2024) Genome-wide association study of susceptibility to *Pseudomonas aeruginosa* infection in cystic fibrosis. Eur Respir J 64:2400062. https://doi.org/10.1183/13993003.00062-2024

Loske J, Voller M, Lukassen S, Stahl M, Thurmann L, Seegebarth A, Rohmel J, Wisniewski S, Messingschlager M, Lorenz S, Klages S, Eils R, Lehmann I, Mall MA, Graeber SY, Trump S (2024) Pharmacological improvement of cystic fibrosis transmembrane conductance regulator function rescues airway epithelial homeostasis and host defense in children with cystic fibrosis. Am J Respir Crit Care Med 209:1338–1350. https://doi.org/10.1164/rccm.202310-1836OC

Lou Y-j, Jin J (2017) Insert restriction enzyme cutting-free cloning strategy for expression plasmid construction. Biotechnol Biotechnol Equip 31:1033–1039. https://doi.org/10.1080/13102818.2017.1351310

Luisetti M (1997) Influence of a secondary genetic factor in cystic fibrosis genotype-phenotype correlations. Monaldi Arch Chest Dis 52:390–391

Ma Z, Zhang X, Zhong W, Yi H, Chen X, Zhao Y, Ma Y, Song E, Xu T (2023) Deciphering early human pancreas development at the single-cell level. Nat Commun 14:5354. https://doi.org/10.1038/s41467-023-40893-8

Mercier J, Calmel C, Mesinele J, Sutanto E, Merabtene F, Longchampt E, Sage E, Kicic A, Boelle PY, Corvol H, Ruffin M, Guillot L (2022) SLC6A14 impacts cystic fibrosis lung disease severity via mTOR and epithelial repair modulation. Front Mol Biosci 9:850261. https://doi.org/10.3389/fmolb.2022.850261

Middleton PG, Mall MA, Drevinek P, Lands LC, McKone EF, Polineni D, Ramsey BW, Taylor-Cousar JL, Tullis E, Vermeulen F, Marigowda G, McKee CM, Moskowitz SM, Nair N, Savage J, Simard C, Tian S, Waltz D, Xuan F, Rowe SM, Jain R (2019)

Elexacaftor-Tezacaftor-Ivacaftor for cystic fibrosis with a single Phe508del allele. N Engl J Med 381:1809–1819. https://doi.org/10.1056/NEJMoa1908639

NandyMazumdar M, Yin S, Paranjapye A, Kerschner JL, Swahn H, Ge A, Leir SH, Harris A (2020) Looping of upstream cis-regulatory elements is required for CFTR expression in human airway epithelial cells. Nucleic Acids Res 48:3513–3524. https://doi.org/10.1093/nar/gkaa089

Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O (2016) Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics 32:1479–1485. https://doi.org/10.1093/bioinformatics/btv722

Panjwani N, Wang F, Mastromatteo S, Bao A, Wang C, He G, Gong J, Rommens JM, Sun L, Strug LJ (2020) LocusFocus: Web-based colocalization for the annotation and functional follow-up of GWAS. PLoS Comput Biol 16:e1008336. https://doi.org/10.1371/journal.pcbi.1008336

Pereira SV, Ribeiro JD, Bertuzzo CS, Marson FAL (2017) Association of clinical severity of cystic fibrosis with variants in the SLC gene family (SLC6A14, SLC26A9, SLC11A1 and SLC9A3). Gene 629:117–126. https://doi.org/10.1016/j.gene.2017.07.068

Perez G, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee CM, Nassar LR, Raney BJ, Speir ML, van Baren MJ, Vaske CJ, Haussler D, Kent WJ, Haeussler M (2025) The UCSC genome browser database: 2025 update. Nucleic Acids Res 53:D1243–D1249. https://doi.org/10.1093/nar/gkae974

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45. https://doi.org/10.1093/nar/29.9.e45

Polineni D, Dang H, Gallins PJ, Jones LC, Pace RG, Stonebraker JR, Commander LA, Krenicky JE, Zhou YH, Corvol H, Cutting GR, Drumm ML, Strug LJ, Boyle MP, Durie PR, Chmiel JF, Zou F, Wright FA, O'Neal WK, Knowles MR (2018) Airway mucosal host defense is key to genomic regulation of cystic fibrosis lung disease severity. Am J Respir Crit Care Med 197:79–93. https://doi.org/10.1164/rccm.201701-0134OC

Rath A, Deber CM (2013) Correction factors for membrane protein molecular weight readouts on sodium Dodecyl sulfate-polyacrylamide gel electrophoresis. Anal Biochem 434:67–72. https://doi.org/10.1016/j.ab.2012.11.007

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M (2015) Integrative analysis of 111 reference human epigenomes. Nature 518:317–330. https://doi.org/10.1038/nature14248

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25. https://doi.org/10.1186/gb-2010-11-3-r25

Rosenstein BJ (1994) Genotype-phenotype correlations in cystic fibrosis. Lancet 343:746–747. https://doi.org/10.1016/s0140-6736(94)91832-5

Salvatore D, Pepe A (2025) Cystic fibrosis transmembrane conductance regulator modulators in cystic fibrosis: A review of Registry-Based evidence. J Clin Med 14:3978. https://doi.org/10.3390/jcm14113978

Satterlee JS, Chadwick LH, Tyson FL, McAllister K, Beaver J, Birnbaum L, Volkow ND, Wilder EL, Anderson JM, Roy AL (2019) The NIH common Fund/Roadmap epigenomics program: successes of a comprehensive consortium. Sci Adv 5:eaaw6507. https://doi.org/10.1126/sciadv.aaw6507

Sikder MOF, Yang S, Ganapathy V, Bhutia YD (2017) The Na^+/Cl–coupled, broad-specific, amino acid transporter SLC6A14 ($ATB^{0,+}$): emerging roles in multiple diseases and therapeutic potential for treatment and diagnosis. AAPS J 20:12. https://doi.org/10.1208/s12248-017-0164-7

Sloan JL, Mager S (1999) Cloning and functional expression of a human Na^+ and Cl–dependent neutral and cationic amino acid transporter B^{0+}. J Biol Chem 274:23740–23745. https://doi.org/10.1074/jbc.274.34.23740

Stegle O, Parts L, Piipari M, Winn J, Durbin R (2012) Using probabilistic Estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc 7:500–507. https://doi.org/10.1038/nprot.2011.457

Strug LJ, Gonska T, He G, Keenan K, Ip W, Boelle PY, Lin F, Panjwani N, Gong J, Li W, Soave D, Xiao B, Tullis E, Rabin H, Parkins MD, Price A, Zuberbuhler PC, Corvol H, Ratjen F, Sun L, Bear CE, Rommens JM (2016) Cystic fibrosis gene modifier SLC26A9 modulates airway response to CFTR-directed therapeutics. Hum Mol Genet 25:4590–4600. https://doi.org/10.1093/hmg/ddw290

Sun L, Rommens JM, Corvol H, Li W, Li X, Chiang TA, Lin F, Dorfman R, Busson PF, Parekh RV, Zelenika D, Blackman SM, Corey M, Doshi VK, Henderson L, Naughton KM, O'Neal WK, Pace RG, Stonebraker JR, Wood SD, Wright FA, Zielenski J, Clement A, Drumm ML, Boelle PY, Cutting GR, Knowles MR, Durie PR, Strug LJ (2012) Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. Nat Genet 44:562–569. https://doi.org/10.1038/ng.2221

Tanaka N, Kawakami T, Taniguchi T (1993) Recognition DNA sequences of interferon regulatory factor 1 (IRF-1) and IRF-2, regulators of cell growth and the interferon system. Mol Cell Biol 13:4531–4538. https://doi.org/10.1128/mcb.13.8.4531-4538.1993

Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, Chang S, Conley SD, Mori Y, Seita J, Berry GJ, Shrager JB, Metzger RJ, Kuo CS, Neff N, Weissman IL, Quake SR, Krasnow MA (2020) A molecular cell atlas of the human lung from single-cell RNA sequencing. Nature 587:619–625. https://doi.org/10.1038/s41586-020-2922-4

Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, Cummings BB, Castel SE, Karczewski KJ, Aguet F, Byrnes A, Lappalainen T, Regev A, Ardlie KG, Hacohen N, MacArthur DG (2017) Landscape of X chromosome inactivation across human tissues. Nature 550:244–248. https://doi.org/10.1038/nature24265

Turcios NL (2020) Cystic fibrosis lung disease: an overview. Respir Care 65:233–251. https://doi.org/10.4187/respcare.06697

Venglovecz V, Hegyi P, Gray M (2021) Ion channels in pancreatic duct epithelial cells in health and disease. Pancreapedia: Exocrine Pancreas Knowl Base. https://doi.org/10.3998/panc.2021.08

Wagner GP, Kin K, Lynch VJ (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci 131:281–285. https://doi.org/10.1007/s12064-012-0162-3

Wang F, Panjwani N, Wang C, Sun L, Strug LJ (2022) A flexible summary statistics-based colocalization method with application to the mucin cystic fibrosis lung disease modifier locus. Am J Hum Genet 109:253–269. https://doi.org/10.1016/j.ajhg.2021.12.012

Wang L, Zhu Y, Zhang N, Xian Y, Tang Y, Ye J, Reza F, He G, Wen X, Jiang X (2024) The multiple roles of interferon regulatory factor family in health and disease. Signal Transduct Target Ther 9:282. https://doi.org/10.1038/s41392-024-01980-4

Watanabe H, Francis JM, Woo MS, Etemad B, Lin W, Fries DF, Peng S, Snyder EL, Tata PR, Izzo F, Schinzel AC, Cho J, Hammerman PS, Verhaak RG, Hahn WC, Rajagopal J, Jacks T, Meyerson M (2013) Integrated cistromic and expression analysis of amplified NKX2-1 in lung adenocarcinoma identifies LMO3 as a functional transcriptional target. Genes Dev 27:197–210. https://doi.org/10.1101/gad.203208.112

Zeileis A, Kleiber C, Jackman S (2008) Regression models for count data in R. J Stat Softw 27:1–25. https://doi.org/10.18637/jss.v027.i08

Zemanick ET, Emerman I, McCreary M, Mayer-Hamblett N, Warden MN, Odem-Davis K, VanDevanter DR, Ren CL, Young J, Konstan MW (2024) Heterogeneity of CFTR modulator-induced sweat chloride concentrations in people with cystic fibrosis. J Cyst Fibros 23:676–684. https://doi.org/10.1016/j.jcf.2024.02.001

Zhang Z, Leir SH, Harris A (2013) Immune mediators regulate CFTR expression through a bifunctional airway-selective enhancer. Mol Cell Biol 33:2843–2853. https://doi.org/10.1128/MCB.00003-13