Exploring the role of epigenetic regulation in cystic fibrosis lung disease pathology using a new airway epithelial cell model

Leigh Henderson

Specific Aims

Lung disease is a leading cause of morbidity and mortality among the 70,000 people with CF worldwide. CF is caused by mutation of a single gene: cystic fibrosis transmembrane conductance regulator protein (CFTR). Bacterial infection precipitates CF lung disease via tissue-damaging inflammation and neutrophil-dominated immune response. Mechanisms linking CFTR mutation and the characteristic CF immune response are not well elucidated. Epigenetic changes (ie. alterations to chromatin that do not affect DNA sequence) associated with mutant CFTR may provide a mechanistic link. My laboratory previously reported altered epigenetic regulation of one gene, interleukin-8, leads to increased inflammatory signaling by CF airway epithelial cells. I propose to look genome-wide for altered regulation in CF using a better cell model than previously available.

I hypothesize that **epigenetic remodeling in CF cells prior to stress (aim 1) correlates with differential response to bacterial, inflammatory, and oxidative stress (aim 2)**. For this study, putative active enhancers, which are marked by specific histone modifications, will be compared between CF and non-CF airway epithelial cells. The goal is to identify enhancers and genes that influence CF-specific pulmonary stress response. Analysis will focus on enhancers of previously discovered gene modifiers of CF lung disease severity. Identified enhancers and genes are potential drug targets to mediate pulmonary inflammation in CF patients. Reducing inflammation is predicted to reduce the severity of CF lung disease.

Aim 1: Identify remodeled regions of the airway epithelial genome in CF airway epithelial cells prior to stress exposure.

This proposal focuses on airway epithelial cells because these cells play an important role in the pathogenesis of CF lung disease *and* express CFTR. Primary airway cells from the majority of 2-year old CF patients have already been exposed to bacterial infection, so it is difficult to obtain primary cells prior to stress exposure. Immortalized cell lines derived from non-CF cells are also not well suited to this study because the cells underwent differentiation with functional CFTR (many epigenetic marks are established during differentiation). Therefore I propose to **1.1)** create a panel of isogenically matched CF and non-CF airway epithelial cell lines, differentiated from human induced pluripotent stem cells (hiPSCs)**.**

Using the isogenically matched CF and non-CF airway epithelial cell lines, I will next: **1.2)** identify differentially activated putative enhancers in CF vs. non-CF cells and **1.3)** determine differential gene activity by measuring RNA transcript levels. Results from aims 1.2 thru 1.3 will be compared to results from two established CF airway epithelial cell models: **1.4)** human tracheal cells and **1.5)** airway epithelial cells from CF-mouse models.

Aim 2: Identify remodeled regions in CF airway epithelial cells exposed to stressors.

CF airway epithelial cells have a unique response to stress. Differential enhancer activation and gene expression contributes to this CF stress response at one known locus. I propose to find such loci genome-wide. To do so, differentially activated putative enhancers and transcripts will be measured in iPSC derived airway epithelial cells (developed in aim 1.1) under three stressors commonly found in CF lungs: **2.1)** bacterial, **2.2)** inflammatory, and **2.3)** oxidative stress.

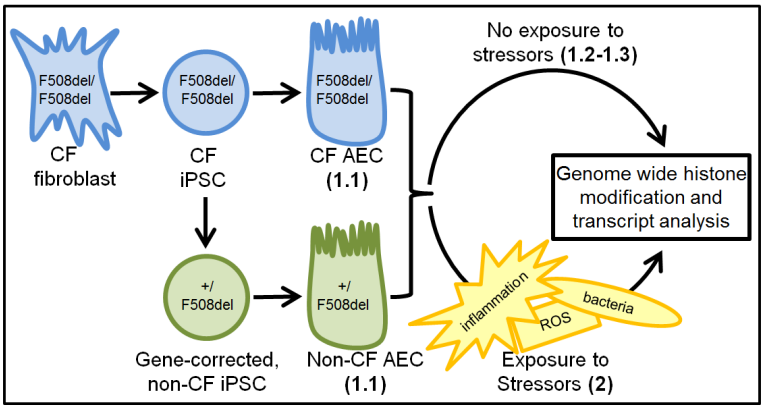

Figure 1: Experimental overview. Human iPSC-derived, isogenically matched airway epithelial cells (1.1) will be analyzed genome wide histone modification and transcript levels before (1.2-1.3) and after (2) exposure to stressors. Key: Airway epithelial cell (AEC).

Significance

Thirty thousand individuals in the United States have CF; 70,000 worldwide. Mutation of the gene CFTR causes CF. CFTR is a regulated channel through which chloride[1] and other anions such as bicarbonate[2,3] are conducted (see review [4]). The deletion of a phenylalanine at amino acid 508 of the protein (abbreviated F508del) is the most common CF-causing mutation, found in at least one CFTR allele of 88% of CF patients in the United States[5].

CFTR in airway epithelial cells directly maintains airway surface mucous viscosity by modulating the osmotic gradient. In CF, highly viscous airway surface mucous predisposes patients to pulmonary bacterial infections. Ensuing inflammation leads to lung tissue damage and, eventually, to lung disease. Inflammation begins when airway epithelial cells release pro-inflammatory chemoattractants such as interleukin 8 (Il-8). I hypothesize mutant CFTR indirectly influences airway epithelial cell stress response. Neutrophils recruited by pro-inflammatory chemoattractants release elastase and oxidants, which harm bacteria but also damage lung connective structure and cells. DNA released by dying bacteria and neutrophils contributes to mucous viscosity and plugging of the airways. Bronchiectasis and lung disease develop over time as a result of these recurrent processes (See review [6]). To treat pulmonary infections, airway clearance techniques to clear thick mucus from the lungs are prescribed to most patients along with frequent use of inhaled antibiotics. Despite treatment, 50% of CF patients in the United States aged 8-29 have moderate to severe lung disease[5]. New therapeutic strategies are needed.

Aside from CFTR, many other genetic loci modify CF disease severity. I am a member of Dr. X's laboratory, who is a leader in the identification and study of CF modifier genes[4,7–12]. Several CF-modifying loci are known to be affected by epigenetic modifications in the presence of mutant CFTR. For example, my laboratory reported hyperacetylation of the promoter of interleukin-8 (Il-8) in CF airway epithelial cells. Hyperacetylation leads to increased Il-8 expression during stress[13]. Increased histone dacetylase 2 activity in CF cells accounts for the increased acetylation at Il-8[14]. It is plausible that increased HDAC2 activity affects other loci as well. Therefore, I hypothesize that **genomic remodeling in CF cells prior to stress (aim 1) results in differential response to bacterial, inflammatory, and oxidative stress (aim 2)**. Technical limitations at the time of the previous study prevented genome-wide study of genomic remodeling. Now, chromatin immunoprecipitation techniques coupled with next-generation sequencing will allow us to explore how epigenetics affects CF-modifying genetic loci at a whole-genome scale. Discovered epigenetic modifications that predispose CF airway epithelial cells to stress will provide potential new drug targets for the treatment of CF lung disease.

Innovation

Novelty of whole-genome epigenetics studies in CF

Epigenetics is defined as "the sum of the alterations to the chromatin template that collectively establish and propagate different patterns of gene expression (transcription) and silencing from the same genome[15]." Multiple aspects of the chromatin template can be altered, including amino acids in histone tails, histone proteins, and cytosine bases in DNA; for the purposes of this study we are most interested in amino acid modifications on histone tails. Certain histone modifications mark genomic regulatory elements and are associated with repression or activation of gene transcription (see Table 2 in reference [16]). My laboratory previously reported hyperacetylation at the interleukin 8 (IL-8) promoter leads to increased expression of IL-8, which promotes inflammation[13]. At the time, there were no available antibodies to target specific histone acetylation sites and only a few loci could be studied because next generation sequencing was prohibitively expensive. The proposed study will target specific histone modifications with newly available antibodies and map histone modifications genome wide using next generation sequencing. To my knowledge, it will be the only genome-wide epigenetic study in CF-literature to date.

A better model of CF in airway epithelial cells is needed

For the purposes of the proposed epigenetic study, an ideal airway epithelial cell model would have the following characteristics:

1. Not an immortalized cell line. Immortalization often causes large-scale genomic rearrangements as well as epigenetic alterations.

2. Isogenic CF and control lines. CF lung disease severity can be accounted for by inheritance at loci other than CFTR[17]. Indeed, multiple SNPs have been found which significantly modify CF lung disease[7,11,12]. Epigenetic influences of these SNPs have not been ruled out as a mechanism of modification. Therefore, it is important to control for genetic diversity in the proposed study.
3. Have developed with mutant CFTR (for the CF line) or normal CFTR (for the non-CF line). Epigenetic modifications contribute to embryonic development[18]. CFTR may influence developmental epigenetic modification (this topic has not been studied). In order to capture the "true" epigenetic state of CF and control airway epithelial cells, then, we need to use cells that underwent differentiation as CF or control cells.
4. Form a ciliated epithelium. *In vivo* airway epithelia has cilia.
5. Have electrical resistance. Tight junctions between airway epithelial cells *in vivo* create electrical resistance across the epithelial membrane.

As shown in table 1, no existing CF human airway epithelial model has all 4 ideal characteristics. We propose to create a model that does. To do so, we will **1.1.1)** create a panel of non-CF human induced pluripotent stem cell (hiPSC) lines that are isogenically matched to CF lines by zinc finger nuclease (ZFN)-mediated recombination and then **1.1.2)** differentiate the hiPSCs to mature airway epithelium.

Table 1: Only the proposed CF airway epithelial cell model has all the ideal characteristics for an epigenetics study.

Cell Model (control/CF)	Description	Ideal characteristics of an airway epithelial cell line				
		Non-immortalized	Isogenic CF and control lines	Developed as CF/non-CF	Form a ciliated epithelium	Electrical resistance
9/HTEo$^-$ pCEP / pCEP-R[38]	CFTR function is inhibited by the overexpression of the protein's R domain		✓*			
S9/IB3[19,20]	CFTR function is rescued by expression from a virus in S9 but not the CF-patient derived IB3 cells		✓*			
Primary cell inhibited with $CFTR_{inh}$-172[30,36,37]	CFTR is inhibited with a drug	✓	✓		✓	✓
16HBE14o$^-$ sense/antisense[39]	CFTR function is inhibited by plasmid-based expression of part of human CFTR		✓*			✓**
Proposed cell model	Genomic CFTR is corrected by ZFN-mediated recombination in iPSCs, then iPSCs are differentiated to airway epithelial cells	✓	✓	✓	✓	✓

*Due to frequent genomic rearrangements in immortalized cell lines, isogenicity between CF and control lines is not certain. ** Electrical resistance in 16HBE14o$^-$ cells is significantly lower than in primary airway epithelium.

The proposed cell model is useful and novel

The proposed iPSC-derived airway epithelial cell model will have several advantages over existing cell lines. First, iPSCs by definition have unlimited self-renewal potential. Therefore, the gene corrected iPSC lines can and will be made available to other investigators. My collaborators at Case Western Reserve University have a successful history of developing and sharing resources (both cell and mouse lines) for the CF research community[21,22]. The unlimited self-renewal of iPSCs also obviates immortalization. Second, CFTR mutation or correction will be done on the endogenous genome, avoiding possible side effects due to the presence of plasmid or virus vectors. Finally, cells will undergo differentiation that mimics *in vivo* development (figure 2) with either a CFTR mutation or correction. CFTR-influenced epigenetic modifications made during development, if these exist, will be present at the final stage of differentiation.

The iPSC-derived airway epithelial cell model proposed is unique among basic science studies of CF. To my knowledge, current uses of this cell model are limited to: (1) therapeutic treatment development and (2) CFTR-correcting drug testing[23]. The novelty of the differentiation protocol and its technical difficulty may explain the lack of iPSC-derived airway epithelial cells in the basic-science literature. The proposed differentiation protocol mimics embryonic development of the tissue. This type of differentiation to lung cell fates became possible with the 2011 discovery of a method to differentiate to anterior foregut endoderm[24]. Since then, several protocols to differentiate to mature airway epithelium have been published[23,25]. Importantly for this proposal, CFTR is expressed in the mature airway epithelium to be generated[23]. Our collaborators at the Pluripotent Stem Cell

Facility at Case Western Reserve University, led by Dr. W, have the expertise to make this challenging differentiation protocol possible[26,27].

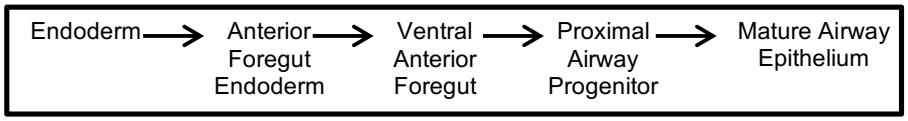

Figure 2: Stages of differentiation from stem cell to mature airway epithelium during embryonic development. The *in vitro* differentiation process to be used mimics this process.

Approach

Aim 1: Identify remodeled regions of the airway epithelial genome in CF airway epithelial cells

Rationale

In aim 1, remodeling of the unperturbed CF airway epithelial genome will be analyzed. Due to the immediate bacterial colonization of the CF airway after birth, it is debated whether excessive inflammation in the CF lung is innate to the system or a response to near continuous bacterial infection throughout life. We will use isogenically matched CF and non-CF airway epithelial cell lines differentiated from human induced pluripotent stem cells (hiPSCs, aim 1.1) to explore epigenetics in cells not previously exposed to the infection or inflammation. The effect of stressors such as bacterial infection will be examined in aim **2**. I hypothesize that **even before exposure to infection and inflammation characteristic of the CF lungs, epigenetic modifications resulting from reduced CFTR function predispose CF cells to perturbation.** Epigenetic modifications will also be studied in previously established cell models (human tracheal cells, aim 1.4, and airway epithelial cells from CF-mouse models, aim 1.5).

1.1) Create a panel of isogenically matched CF and non-CF airway epithelial lines

A better model of CF in airway epithelial cells is needed

I hypothesize that **epigenetic remodeling in CF cells prior to stress (aim 1) correlates with differential response to bacterial, inflammatory, and oxidative stress (aim 2).** This hypothesis will be addressed in the airway epithelial cell type because these cells play an important role in the pathogenesis of CF lung disease *and* express CFTR. Existing cell lines are not ideal for the study of epigenetic remodeling, as discussed in the innovation section. A new airway epithelial cell model is needed.

Study design and expected results (with preliminary data)

1.1a) Create a panel of isogenically matched CF and non-CF human induced pluripotent stem cell (hiPSC) lines by zinc finger nuclease (ZFN)-mediated recombination.

Rationale

Pluripotent cells, such as induced pluripotent stem cells (iPSCs) and embryonic stem cells (ESCs), can be differentiated to airway epithelial cells using published protocols[23–25]. For this study, iPSCs will be used for several reasons. Fibroblasts (which are obtainable through minimally-invasive skin biopsy) can be differentiated to iPSCs through a well-characterized protocol. CF cells are needed and it is more ethical to find CF fibroblast donors than CF embryonic stem cells. Furthermore, we are using CF fibroblasts from the Coriell Institute for Medical Research, which come with a medical history that will be informative to downstream analysis.

The fibroblasts will be differentiated to iPSCs, gene-corrected, and then differentiated to airway epithelial cells (figure 1). A zinc finger nuclease (ZFN) will be used to achieve gene correction, rather than a transcription-activator like effector nucleases (TALEN), because the ZFN is freely available to us through a collaborator. The ZFN will be used to correct CFTR (F508del/F508del) cells to the heterozygous state (+/F508del). Bi-allelic gene correction is prohibitively difficult to achieve, but one copy of the wild type CFTR allele is enough to prevent CF. Cells will be corrected at the iPSC stage, rather than the fibroblast stage, because iPSCs have unlimited self-renewal potential and will be more amenable to expansion after clonal selection. Correction at the iPSC stage will also allow cells to undergo differentiation to airway epithelial cells with functional CFTR in the corrected cells and reduced-function CFTR in the CF cells.

Derive hiPSCs from fibroblasts of individuals with F508del homozygous CF

Two F508del homozygous fibroblast cell lines were ordered from the Coriell Institute for Medical Research (ID GM01957 and GM04539). Introduction of a polycystronic lentivirus to the fibroblasts and doxycycline-induced expression of Oct4, Sox2, Klf4, and c-Myc resulted in reprogramming to iPSCs. Once reprogrammed, the lentiviral DNA was excised by cre-lox recombination. An assay for teratoma formation confirmed a stem-cell state in each of the two cell lines. My collaborators at the Case Western Reserve University Pluripotent Stem Cell Facility performed the reprogramming and teratoma assays.

Correct CFTR (F508del/F508del) iPSCs to CFTR (+/F508del)

Zinc finger nucleases (ZFNs) are engineered proteins containing DNA-binding domains (zinc fingers) and a DNA-cleaving domain (a Fok1 nuclease). When the zinc finger domains from two different ZFNs bind nearby DNA sequences, their accompanying nuclease domains come into contact and cleave a double-stranded break in the DNA. Cells can correct double-stranded breaks using homologous recombination, in which the cell degrades the DNA up and down-stream of the break and replaces it using an identical or near-identical template sequence. When a near-identical template sequence is introduced into the cell along with two ZFNs, homologous recombination results in genomic incorporation of the template sequence[28] (figure 3A).

Dr. Z has kindly provided me with ZFN-containing plasmids (pZFNs) that cleave a double stranded break in CFTR near the position corresponding to amino acid 508 in the CFTR protein (figure 3B). He has also provided a donor plasmid containing the normal sequence of CFTR in this region. When introduced together into a CFTR (F508del/F508del) cell, the mutant sequence is replaced by the normal sequence. When incorporated into the genome, several silent base pair changes in the donor plasmid as compared with CFTR's normal sequence disallow further ZFN binding & cleavage.

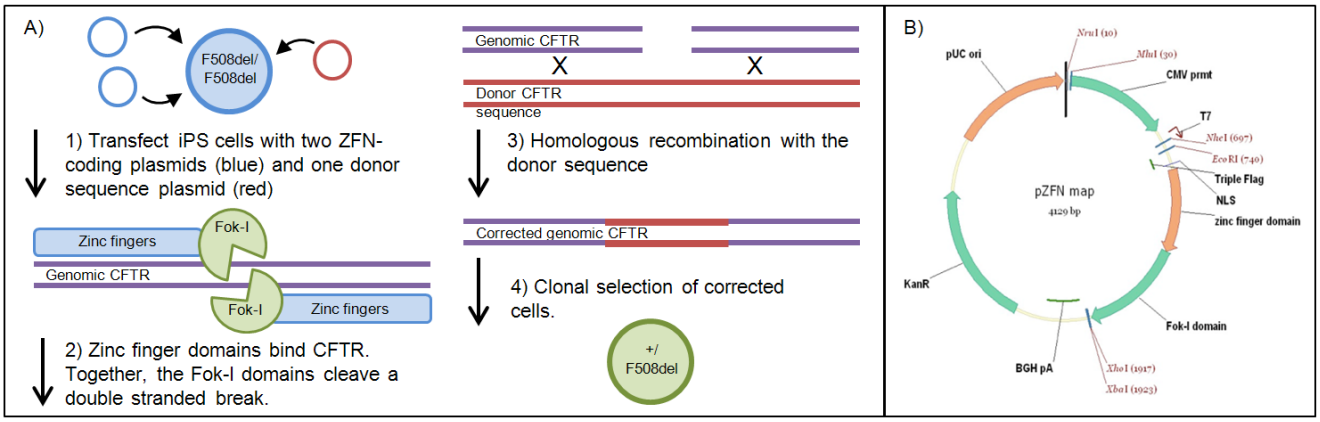

Figure 3: (A) Mechanism of gene correction by ZFNs. (B) Plasmid for the expression of CFTR-targeting ZFN.

Transfection of the pZFNs and donor plasmid to the iPSCs will be performed using the Neon Transfection System or the Lonza Nucleofector. Once transfected, clones in which CFTR is successfully corrected on one allele will need to be identified and selected for. The method of selection will depend on transfection and recombination efficiency.

To determine transfection efficiency, I have cloned the two ZFN-coding sequences into green and red fluorescent protein (GFP and RFP) encoding plasmids. Sequencing verified that the ZFN sequence in the cloned plasmids is not mutated from the original. Cells will be transfected with the two ZFN-FP plasmids plus the donor plasmid. The percent of cells positive for both GFP and RFP by either confocal fluorescence microscopy or fluorescence-activated cell sorting (FACS) will be the transfection efficiency.

To determine recombination efficiency, GFP and RFP positive cells will be isolated and cultured. Individual clones (~20 per cell line) will then be isolated from these plates, cultured separately, and CFTR genotype at position 508 will be determined by sequencing. The percent of GFP and RFP positive clones in which CFTR is corrected to (+/F508del) will be the recombination efficiency.

If transfection and recombination efficiency are high, then we can use the ZFN-only coding plasmids plus the donor plasmid to correct CFTR. This would eschew introduction of fluorescent proteins, which could be a confounding factor in later experiments. Otherwise, the ZFN + FP coding plasmids plus the donor plasmid will be used to correct CFTR. Sequencing will be used to verify CFTR correction.

1.1b) Differentiate the hiPSCs to mature airway epithelium

Rationale

The proposed *in vitro* differentiation procedure was selected because it mimics the *in vivo* embryonic development of airway epithelial cells (figure 2). The differentiated airway epithelial cells are similar to native airway epithelial cells in gene expression and morphology[23].

Differentiate human pluripotent stem cells to mature airway epithelium

Embryonic stem cells (ESCs) are the gold standard type of pluripotent stem cell, so the differentiation protocol will be optimized on human ESCs (hESCs) before moving to human iPSCs. As shown in figure 4, I have differentiated hESCs to Nkx2.1+ endoderm following a protocol adapted from Dr. Janet Rossant and colleagues[25,29] (figure 4). Cells will be differentiated to mature airway epithelia by 5 weeks of further culture at air-liquid interface[23]. After successfully differentiating hES cells to mature airway epithelia, the same protocol will be tested and optimized on non-CF hiPSCs and finally on CF hiPSCs. I am performing stem cell differentiation at the Pluripotent Stem Cell Facility at Case Western Reserve University under the advisement of Dr. W.

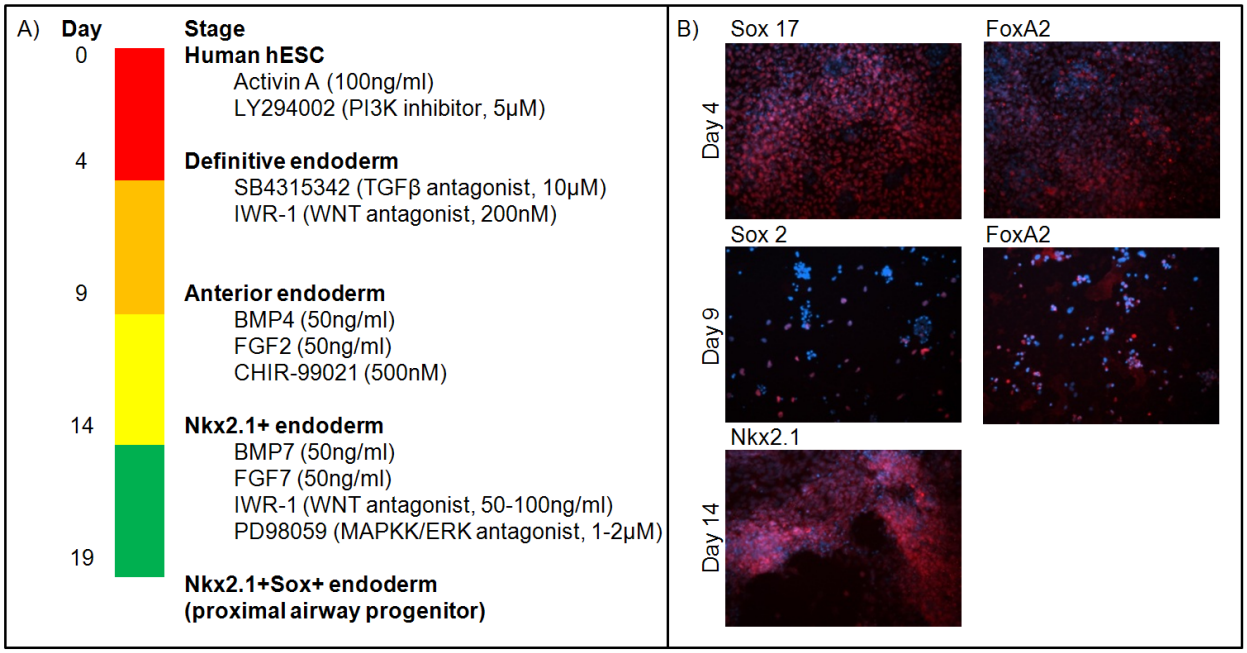

Figure 4: (A) Visual representation of differentiation protocol for the generation of Nkx2.1+Sox2+ endoderm, adapted from references [25,29]. Five weeks of proximal airway progenitor culture at air-liquid interface results in maturation to polarized, ciliated airway epithelium[23]. (B) Immunofluorescence of the indicated cell stage marker (red) and DAPI (blue) in hES cells differentiated following the protocol in part A.

Verify that differentiated human pluripotent stem cells are mature airway epithelium

The electrical properties, morphology, transcriptome, and protein expression of our differentiated cells will be characterized to verify that they are truly mature airway epithelial cells. Airway epithelial cell cultures should 1) form an electrical barrier, 2) have apical cilia, and 3) express mRNA and protein characteristic of airway epithelial cells and not of other endodermal-lineage cell types. Electrical measurements will be taken in an Ussing chamber. Cilia will be visualized by electron microscopy at the CWRU Electron Microscopy Core. Microarray and RT-PCR will be used to assay the transcriptome. Immunofluorescence will be used to detect protein expression. Primary airway epithelial cells (available through the CWRU Epithelial Cell Culture Core directed by Dr. Y) will serve as a control.

1.2) Identify differentially activated putative enhancers and associated genes in CF vs. non-CF cells

Rationale

I hypothesize that **epigenetic modifications resulting from reduced CFTR function predispose CF cells to perturbation, particularly near previously identified gene modifiers of the disease** (here defined as "candidate genes," see table 3). Specifically, I hypothesize that reduced CFTR function indirectly leads to differentially activated enhancers and thereby to changed gene transcription and a distinct CF response to stress. Discovery of candidate genes with differentially activated enhancers in CF vs. non-CF cells would support the hypothesis. Enhancer activation is indicated by enhancer-associated histone modifications; H3K4me1 and H3K27ac peaks mark active putative enhancers[30,31]. Therefore, we should look for genes with enhancers marked with H3K4me1 and H3K27ac peaks differentially in CF vs. non-CF cells.

1.2a) Sequence DNA bound to H3K4me1 and H3K27ac

Aim 1.2.1 will be accomplished by chromatin immunoprecipitation (ChIP) followed by next generation sequencing performed as outlined by Odom and colleagues[32] and in line with ENCODE version 2.0 guidelines for ChIP-seq eperiments[33]. Briefly, the first stage of ChIP involves formaldehyde-crosslinking DNA to proteins in cells, lysing the cells, fragmenting DNA by sonication, immunoprecipitation of protein-DNA complexes with antibodies bound to magnetic beads, and purification of the immunoprecipitated DNA. The second stage prepares isolated DNA for sequencing. A fraction of each sample undergoes ChIP without the immunoprecipitation step; this is an input control for downstream analysis. I am currently working out the ChIP protocol under the advisement from Dr. Q's laboratory. I have purified DNA of the expected 200-400bp fragment size from first stage of the ChIP protocol and am now optimizing the second stage.

Sequencing will be run at the CWRU Sequencing Core on the HiScanSQ. Reads will be single-end and 50bp in length. I expect to find 20-50k H3K4me1+ and H3K27ac+ peaks genome-wide in each sample[34]. Steps to be used in sequencing data analysis are listed in table 2. Computationally demanding scripts can be run at the High Performance Computing Cluster at Case Western Reserve University, available by subscription. The X laboratory server, which has one terabyte of space available for this project, will store data files. As suggested in the ENCODE 2.0 guidelines[33], ChIP-seq will be performed on at least two biological replicates of each cell type and condition and only replicates in which 80% of the top 40% of gene targets overlap will be considered for further analysis.

Table 2: Analysis pipeline for ChIP-sequencing data.

Step	Purpose	Software to be used
1. Assess read quality	Determine whether sequencing data is quality enough for further analysis and whether reads need to be trimmed	FastQC[35]
2. Trim low quality bases	Remove bases (and entire reads if necessary) that the sequencer called with low certainty from the dataset	BRAT[36]
3. Align to genome	Find the genomic location of each read	BWA[37], SAM Tools[38]
4. Call peaks	Determine genomic locations with significant enrichment for the histone modification of interest	MACS[39]
5. Annotate	Associate genomic sequences with gene, regulatory element, and other known functional information	MACS[39]
6. Analyze/visualize	Interpret the biological meaning of the data	Savant[40], IGV[41]

1.2b) Associate differentially activated putative enhancers with genes

Identification of the genes acted upon by identified putative active enhancers (sequences enriched for H3K4me1 and H3K27ac) presents a challenge. Existing methods to identify acted-upon genes all have limitations, so we will use several different methods. Methods to be used in this study are: (1) nearest gene, (2) CFCF-defined domain, and (3) PreSTIGE. The first identifies the gene nearest to a putative enhancer as its target. However, some enhancers are known to interact with distant genes, up to 1Mb away, and not interact with the nearest gene[42]. The nearest gene method also fails to recognize that genes and enhancers on opposite sides of an insulator rarely interact, even if in close proximity. Insulators are genetic regions often

identified by the binding of the insulator binding protein CTCF and which are variable between cell types (see review [43]). The second method identifies target genes as those in the same CTCF-defined domain as putative enhancers. The third method, PreSTIGE, is a computational tool being developed by Dr. Q's laboratory at CWRU (unpublished). PreSTIGE identifies gene-enhancer interactions by comparing genome-wide H3K4me1 peaks with gene expression data in many cell types. In cell types where a gene is expressed, its enhancers should be associated with H3K4me1; possible gene-enhancer interactions are eliminated when this does not hold true. The PreSTIGE database of predicted gene-enhancer pairs will be used to analyze data for this study. Gene-enhancer pairs identified by any of the three methods used will be considered for further analysis.

Table 3: Candidate gene list. Candidate genes are those genes or SNPs previously reported associate with modified lung disease severity.

Gene (or nearest gene)	
IL-8	EHF/APIP[7]
NOS2	AGTR2[7]
ENaC	HLA-DRA[7]
Il-6	EEA1[7]
Il-1	SLC8A3[7]
Il-17[44]	AHRR[7]
IFRD1[9]	CDH8[7]
MBL	CBLN4/MC3R[7]

1.3) Determine differential gene, primary transcript, and isoform expression by measuring RNA transcript levels

Rationale

Predicted interactions between differentially activated enhancers and genes need to be tested for effect on gene expression for two reasons. First, the predicted enhancer-gene relationship may not be real. Second, inactive enhancers may not prevent gene transcription. RNA-seq will be used to quantify gene, primary transcript, and isoform expression. Results will be compared between CF and non-CF cell lines to find differentially expressed transcripts from genes with differentially activated putative enhancers.

Approach

RNA-seq will be used to accomplish aim 1.3. Briefly, total RNA will be isolated from cultured cells (RNeasy Mini Kit with optional DNase step; Qiagen). Total RNA will be delivered to the CWRU Sequencing Core for preparation (TruSeq Stranded Total RNA Sample Preparation Kit[45]; Illumina) and sequencing. For sequencing, 4 bar-coded biological replicates (CF or non-CF) will be run per lane of the Illumina GA IIx sequencer with paired end, 100bp-long reads. Steps to be used in sequencing data analysis are listed in table 4.

Table 4: Analysis pipeline for RNA-sequencing data analysis.

Step	Purpose	Software to be used
1. Assess read quality	Determine whether sequencing data is quality enough for further analysis and whether reads need to be trimmed	FastQC[35]
2. Trim low quality bases	Remove bases (and entire reads if necessary) that the sequencer called with low certainty from the dataset	Dynamic trim
3. Align to genome	Find the genomic location of each read. Due to splicing, not all sequences from mRNA will be initially aligned. Find possible alignments for sequences that include splice sites	TopHat (uses Bowtie2)
4. Assemble, identify, and enumerate gene isoforms	Alternative splicing of some genes forms results in different mRNA transcripts (isoforms). Piece sequences together into transcripts and estimates the abundance of different isoforms. Graph results	Cufflinks, Cuffmerge, Cuffdiff, CummeRbund

Transcript levels of differentially activated candidate genes and non-candidate genes of interest (identified in aims 1.2 and 1.3, respectively) will be further verified through quantitative real time PCR. Because transcript levels do not necessarily reflect protein expression, we will also measure protein levels by western blotting for genes of interest with available antibodies.

1.4) Perform aims 1.2 to 1.3 in human tracheal cells

The proposed iPSC-derived airway epithelial lines (developed in aim 1.1) are novel. Therefore, epigenetic markings and RNA expression findings should be verified in existing airway epithelial cell models. Primary human tracheal cells will be the first model for verification.

The Epithelial Cell Culture Core at CWRU, directed by Dr. Y, isolates airway epithelial cells from autopsied or lung transplant-recipient tracheas (with IRB approval). I expect that more non-CF tracheas will be available for use. A CFTR inhibitor will be used to create CF-like cells isogenic to the non-CF cells. The core digests tracheas with trypsin and then isolates epithelial cells from fibroblasts by differential attachment. Isolated airway epithelial cells provided by the core will be cultured at an air-liquid interface as previously described[46]. After 3 weeks of culture, cells will be moved to liquid-liquid interface and treated with DMSO (1:1000 with vehicle control) or 20M CFTRinh-172 prepared in DMSO. Media will be replaced daily with drugs added to the basolateral and apical side of the culture. After 3 days of mock treatment or CFTR inhibition, differentially activated putative enhancers and differential RNA expression (as described in aims 1.2 and 1.3, respectively) will be measured.

I expect that epigenetic markings and RNA expression from non-CF iPSC-derived and primary human airway epithelial cell lines will cluster separately from CF cell lines.

1.5) Perform aims 1.2 to 1.3 in airway epithelial cells from CF-mouse models

Airway epithelial cells from mice will be the second model for verification of epigenetic marking and RNA expression in iPSC-derived airway epithelial cells. Cells will be obtained from the CFTR (F508del/F508del) mouse model. The CWRU CF Animal Core cares for the animals with approval from the CWRU IACUC. The Epithelial Cell Culture Core at CWRU will digest the tracheas with trypsin and then isolate epithelial cells from fibroblasts by differential attachment. My laboratory has developed a protocol for the culture of these cells adapted from Dorin and colleagues[47]. After 3 weeks of culture, differentially activated putative enhancers and differential RNA expression (as described in aims 1.2 and 1.3, respectively) will be measured.

I expect that a subset epigenetic markings and RNA expression from non-CF iPSC-derived and primary mouse airway epithelial cell lines will cluster separately from CF cell lines. I also expect a different subset of human-specific and mouse-specific genes to cluster together in cells from each species regardless of CFTR function.

Potential Problems and Alternate Strategies

Off-target DNA binding and cleavage by zinc fingers occur at a low frequency and can result in mutation[48]. Off target mutations can occur anywhere in the genome and require whole-genome sequencing to detect. Whole-genome sequencing is prohibitively expensive, so an alternate strategy is needed to assure that effects of ZFN-correction are not due to off-target mutation. First, multiple (3-5) clones of ZFN-corrected cell lines will be assessed and only epigenetic alterations observed in all clones will be considered unrelated to off-target effects. Second, potential sites of off-target mutation will be sequenced and only cell lines without mutation will be kept. Off target effects are non-random and most often occur at sequences that closely match the target sequences of the ZFN[48]. Potential off-target sites are identified by searching the genome for close match sequences[49]. Alternatively, multiple ZFN or transcription-like activator effector nuclease (TALEN) pairs that cleave near F508del but whose target sequences are different can be generated. It would be expected that the off-target cleavage sites of one ZFN/TALEN pair to be different than another. Epigenetic alterations observed in cell lines generated by different pairs can be assumed to be caused by the change to the true target region.

In vitro differentiation to definitive endoderm (the first step in the proposed differentiation protocol) modulates the epigenome[50], but whether the modulation is similar to that during developmental differentiation is unknown. Epigenetic differences between native airway epithelial cells and *in vitro* differentiated airway epithelial cells are also unknown. The proposed study therefore compares *in vitro* differentiated CF and non-CF cells, not *in vitro* differentiated cells to native cells. However, some of the epigenetic differences between CF and non-CF cells may reflect disparate responses to *in vitro* differentiation. To filter out such artifacts, epigenetic differences that do not also appear in other cell models (human tracheal cells in aim 1.4 and mouse airway cells in aim 1.5) will be removed from consideration.

Aim 2: Identify remodeled regions in CF airway epithelial cells exposed to stressors

Bacterial infection, inflammation, and oxidative stress often afflict the airways of people with CF. Epigenetic changes at regulatory elements of the genome contribute to the cellular response to stress. I hypothesize **epigenetic modifications resulting from reduced CFTR function dysregulate the response to stress in CF cells.** Discovery of differentially expressed transcripts with differentially activated enhancers between non-CF and CF airway epithelial cells under conditions of stress would support the hypothesis. In Aim 2, airway epithelial cells will be subjected to stressors common to the CF airways before assay for genome-wide active putative enhancers. We propose to search for differentially activated enhancers in cells exposed to **2.1)** bacterial, **2.2)** inflammatory, and **2.3)** oxidative stress using the same experimental design as proposed in aims 1.2-1.3:

1.2) identify differentially activated putative enhancers and associated genes in CF vs. non-CF cells
1.3) determine differential gene activity by measuring RNA transcript levels.

Results will be compared between CF and non-CF cells of the hiPSC-derived airway epithelial cell lines (developed in aim 1.1). Results will also be compared between cells under no stress (from aim 1) and under stress.

2.1) Identify differentially activated regulatory regions in airway epithelial cells upon bacterial stress

Rationale

Over 75% of CF patients are infected with *Pseudomonas aeruginosa* by age three[51]. By adulthood, *P. aeruginosa* and several other species of bacteria are usually the most numerous species in the lungs of healthy CF patients, although these species are rarely found in the lungs of healthy individuals without CF[52,53]. Airway epithelial cells respond to contact with *P. aeruginosa* by producing pro-inflammatory molecules, which propagates the inflammation and infection cycle that leads to CF lung disease. Among the products of *P. aeruginosa,* flagellin (the major structural protein of flagella) stimulates the greatest response in airway epithelial cells (see review [54]). In order to elicit a bacterial stress response, I therefore propose to expose airway epithelial cells to flagelin derived from *P. aeruginosa*.

Experimental approach

Flagellin will be purified from a clinical isolate of *P. aeruginosa,* available from Dr. R, as previously described[55]. A mock purification of flagellin from a strain of *P. aeruginosa* lacking flagella will serve as a negative control. After exposure to flagellin (50ng/ml) for 6 hours[56], cells will be assayed as described in aims 1.2 and 1.3.

2.2) Identify differentially activated regulatory regions in airway epithelial cells upon inflammatory stress

Rationale

The immune response to bacterial infection induces inflammation, a process that successfully targets bacteria in the short term but is harmful to lung tissue in the long run. To induce inflammation, airway epithelial cells release pro-inflammatory molecules such as IL-6 and IL-8. Airway epithelial cells also help to control inflammation through the release of anti-inflammatory molecules such as Il-10. There are reports of increased pro-inflammatory and decreased anti-inflammatory molecule release from CF airway epithelial cells as compared with controls under conditions of bacterial and/or inflammatory stress[57,58], although evidence to the contrary has also been published[59]. Altered cytokine release by CF cells may be linked to differential gene activation status, which I propose to study genome-wide while cells are exposed to pro-inflammatory cytokines.

Experimental approach

To induce inflammatory stress, airway epithelial cells will be treated with a pro-inflammatory cytokine mixture of Il-1β (0.167ng/μl) and tumor necrosis factor-α (TNF-α, 0.33ng/μl) in serum-free culture media (as described previously [12]). The cytokine concentrations were chosen to be physiologically relevant[60]. After exposure to inflammatory stress, cells will be assayed for putative active enhancers as described in aims 1.2 and 1.3.

2.3) Identify differentially activated regulatory regions in airway epithelial cells upon oxidative stress

Rationale

Chronic inflammation characteristic of CF lung disease is associated with an overabundance of neutrophils. Neutrophils kill bacteria with a variety of effector mechanisms, including the release of oxidants (see textbook [61]). Neutrophils cannot exit the airways alive and have a short life-span. Therefore, two mechanisms remove dead neutrophils from the lung: mucociliary clearance and phagocytosis by alveolar macrophages. Dead neutrophils accumulate in the CF airways because the high viscosity of airway mucous impairs mucociliary clearance and proteases released by neutrophils impair phagocytosis by alveolar macrophages. Oxidants such as hydrogen peroxide released by neutrophils during faulty apoptosis or during phagocytosis accumulate in the CF airways, damaging lung tissue and contributing to lung disease (See review [6]). Previous studies from my laboratory show oxidative stress causes increased acetylation of the Il-8 promoter in airway epithelial cells that correlated with higher Il-8 release[13]. I propose to discover genome-wide epigenetic changes in cells exposed to oxidative stress.

Experimental approach

To induce oxidative stress, airway epithelial cells will be exposed to 200 mM hydrogen peroxide in serum-free medium for 3 hours[13]. After exposure to oxidative stress, cells will be assayed as described in aims 1.2 and 1.3.

Potential Problems and Alternate Strategies

Due to the expense of whole-genome sequencing, the number of variables in stressor experiments is limited. I have attempted to choose the most biologically relevant time-points and stressors, as explained in the rationale for each sub aim.

Innate oxidative stress in CF cells presents a potential problem. CF airway epithelial cells have been shown to become more similar to non-CF cells upon antioxidant treatment, showing that oxidative stress may be inherent to CF cells[13,14]. The presence of inherent oxidative stress may mean that additional oxidative stress by exposure to hydrogen peroxide, as proposed, may not cause many alterations to the epigenome in CF cells. To test this hypothesis, we could assess the effect of antioxidants on the epigenome by exposure to 5 mM N-acetyl-cysteine (NAC) in serum-free medium for 2 hours[13]

Future Directions

From the proposed study, I expect to discover differentially activated putative enhancers in CF airway epithelial cells that correlate with differential response to stress. Among the genes affected, the most likely to affect *in vivo* response to stress will be selected. For example, I expect Il-8 to be among the genes selected because previous data from my lab shows increased Il-8 expression associated with epigenetic modification is present in CF cells[13]. Future work should test (1) whether the level of selected genes affects cellular stress response and (2) the mechanisms by which the enhancers of selected genes are modified. Genes and mechanisms to be identified by this approach are possible drug targets for the treatment of CF lung disease. Elucidating mechanisms behind enhancer modification will also expand the knowledge of CF lung disease pathogenesis.

Timetable

As described in figure 5, I expect experiments for this proposal to be completed in two years' time.

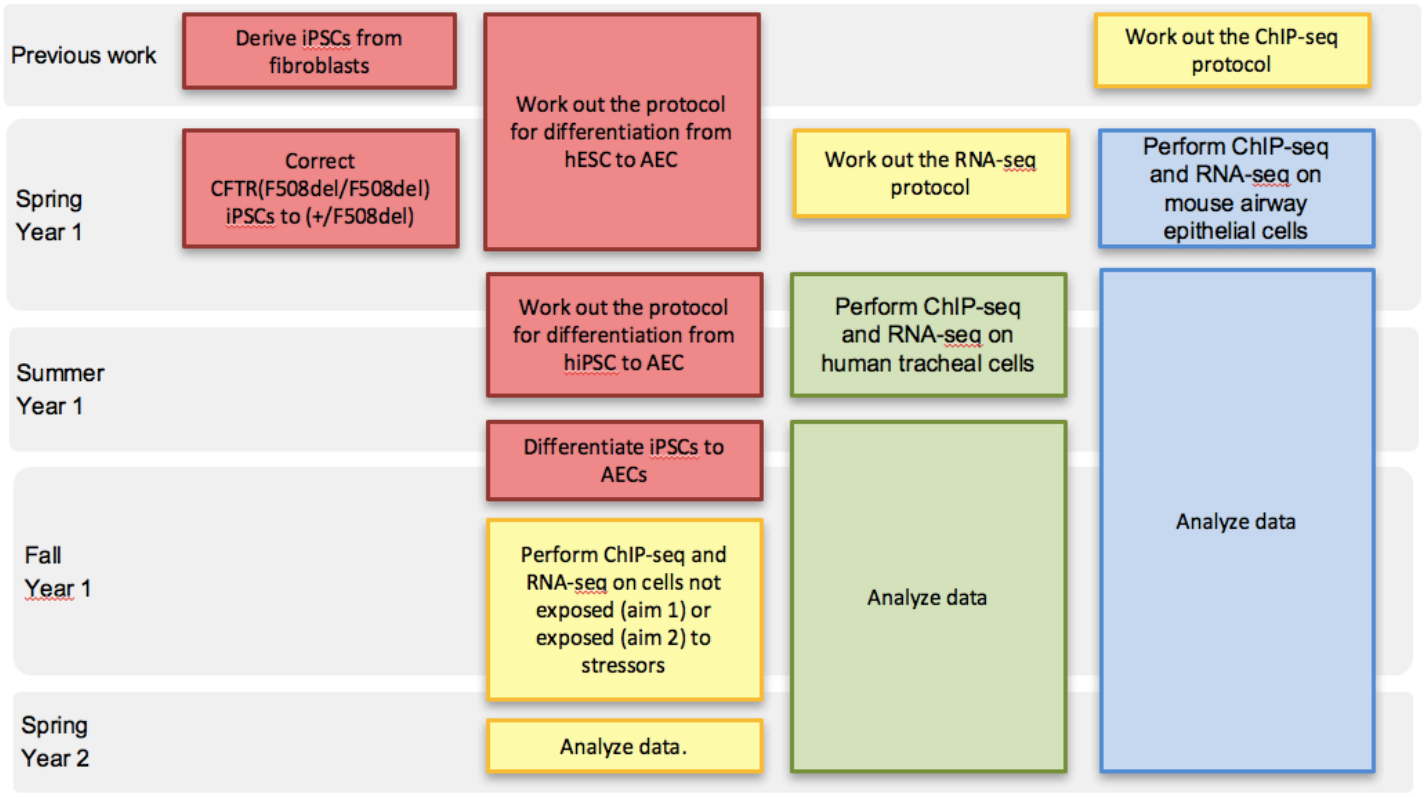

Figure 5: Planned timetable. Aims are color-coded: 1.1 (red), 1.2-1.3 (yellow), 1.4 (green), 1.5 (blue). Aims 2.1-2.3 are shown combined with aim 1.2-1.3 (yellow). Key: Airway epithelial cell (AEC).

Literature Cited

1. Bear, C. E. *et al.* Purification and functional reconstitution of the cystic fibrosis transmembrane conductance regulator (CFTR). *Cell* **68**, 809–18 (1992).

2. Choi, J. Y. *et al.* Aberrant CFTR-dependent HCO3- transport in mutations associated with cystic fibrosis. *Nature* **410**, 94–97 (2001).

3. Linsdell, P. *et al.* Permeability of wild-type and mutant cystic fibrosis transmembrane conductance regulator chloride channels to polyatomic anions. *The Journal of general physiology* **110**, 355–64 (1997).

4. Drumm, M. L., Ziady, A. G. & Davis, P. B. Genetic Variation and Clinical Heterogeneity in Cystic Fibrosis. *Annual review of pathology* 267–282 (2011).doi:10.1146/annurev-pathol-011811-120900

5. Cystic Fibrosis Foundation *Cystic Fibrosis Foundation Patient Registry 2010 Annual Data Report*. (Bethesda, MD, 2010).

6. Chmiel, J. F., Berger, M. & Konstan, M. W. The role of inflammation in the pathophysiology of CF lung disease. *Clinical reviews in allergy & immunology* **23**, 5–27 (2002).

7. Wright, F. a *et al.* Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nature genetics* **43**, 539–46 (2011).

8. Taylor, C. *et al.* A novel lung disease phenotype adjusted for mortality attrition for cystic fibrosis genetic modifier studies. *Pediatric pulmonology* **46**, 857–69 (2011).

9. Gu, Y. *et al.* Identification of IFRD1 as a modifier gene for cystic fibrosis lung disease. *Nature* **458**, 1039–42 (2009).

10. Pace, R. *et al.* Genetic Modifiers of Lung Disease in Cystic Fibrosis. *The New England journal of medicine* **353**, 1443–1453 (2005).

11. Henderson, L. B. *et al.* Variation in MSRA Modifies Risk of Neonatal Intestinal Obstruction in Cystic Fibrosis. *PLoS Genetics* **8**, e1002580 (2012).

12. Hillian, A. *et al.* Modulation of cystic fibrosis lung disease by variants in interleukin-8. *Genes and immunity* **9**, 501–8 (2008).

13. Bartling, T. R. & Drumm, M. L. Oxidative stress causes IL8 promoter hyperacetylation in cystic fibrosis airway cell models. *American journal of respiratory cell and molecular biology* **40**, 58–65 (2009).

14. Bartling, T. R. & Drumm, M. L. Loss of CFTR results in reduction of histone deacetylase 2 in airway epithelial cells. *American journal of physiology. Lung cellular and molecular physiology* **297**, L35–43 (2009).

15. Allis, D. C., Jenuwein, T. & Reinberg, D. Chapter 3: Overview and Concepts. *Epigenetics* 23–40 (2007).

16. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

17. Vanscoy, L. L. *et al.* Heritability of lung disease severity in cystic fibrosis. *American journal of respiratory and critical care medicine* **175**, 1036–43 (2007).

18. Leeb, M. & Wutz, A. Establishment of epigenetic patterns in development. *Chromosoma* **121**, 251–62 (2012).

19. Flotte, T. R. *et al.* Gene expression from adeno-associated virus vectors in airway epithelial cells. *American journal of respiratory cell and molecular biology* **7**, 349–356 (1992).

20. Eidelman, O. *et al.* Control of the proinflammatory state in cystic fibrosis lung epithelial cells by genes from the TNF-alphaR/NFkappaB pathway. *Molecular medicine* **7**, 523–34 (2001).

21. Perez, A., Risma, K. A., Eckman, E. A. & Davis, P. B. Overexpression of R domain eliminates cAMP-stimulated Cl- secretion in 9 / HTEo- cells in culture. *American journal of physiology. Lung cellular and molecular physiology* **271**, L–85–L92 (1996).

22. Hodges, C. A., Cotton, C. U., Palmert, M. R. & Drumm, M. L. Generation of a conditional null allele for Cftr in mice. *Genesis* **46**, 546–552 (2008).

23. Wong, A. P. *et al.* Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein. *Nature Biotechnology* (2012).doi:10.1038/nbt.2328

24. Green, M. D. *et al.* Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. *Nature biotechnology* **29**, 267–72 (2011).

25. Mou, H. *et al.* Generation of Multipotent Lung and Airway Progenitors from Mouse ESCs and Patient-Specific Cystic Fibrosis iPSCs. *Cell Stem Cell* **10**, 385–397 (2012).

26. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–9 (2007).

27. Najm, F. J. *et al.* Rapid and robust generation of functional oligodendrocyte progenitor cells from epiblast stem cells. *Nature methods* **8**, 957–62 (2011).

28. Carroll, D. Genome Engineering With Zinc-Finger Nucleases. *Genetics* **188**, 773–782 (2011).

29. Mou, H. *et al.* Protocol: Generation of Nkx2.1+ lung endoderm and Nkx2.1+Sox2+ proximal airway progenitors from human embryonic stem cells and iPS cells. (2012).

30. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–12 (2009).

31. Zentner, G. E., Tesar, P. J. & Scacheri, P. C. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome research* **21**, 1273–83 (2011).

32. Schmidt, D. *et al.* ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods (San Diego, Calif.)* **48**, 240–8 (2009).

33. ENCODE Consortium ENCODE and modENCODE Guidelines For Experiments Generating ChIP, DNase, FAIRE, and DNA Methylation Genome Wide Location Data, Version 2.0. 1–20 (2011).at <http://encodeproject.org/ENCODE/experiment_guidelines.html>

34. Zentner, G. E. & Scacheri, P. C. The chromatin fingerprint of gene enhancer elements. *The Journal of biological chemistry* **287**, 30888–96 (2012).

35. Bioinformatics, B. FastQC: A quality control tool for high throughput sequence data. at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

36. Harris, E. Y., Ponts, N., Levchuk, A., Roch, K. Le & Lonardi, S. BRAT: bisulfite-treated reads analysis tool. *Bioinformatics* **26**, 572–573 (2010).

37. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

38. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

39. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).

40. Fiume, M., Williams, V., Brook, A. & Brudno, M. Savant: genome browser for high-throughput sequencing data. *Bioinformatics (Oxford, England)* **26**, 1938–44 (2010).

41. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature biotechnology* **29**, 24–26 (2011).

42. Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics* **12**, 1725–1735 (2003).

43. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).

44. Brodlie, M. *et al.* Raised interleukin-17 is immunolocalised to neutrophils in cystic fibrosis lung disease. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* **37**, 1378–85 (2011).

45. Illumina Illumina TruSeq ® Stranded Total RNA Sample Preparation Guide. (2012).

46. Perez, A. *et al.* CFTR inhibition mimics the cystic fibrosis inflammatory profile. **4948**, 383–395 (2007).

47. Davidson, D. J., Kilanowski, F. M., Randell, S. H., Sheppard, D. N. & Dorin, J. R. A primary culture model of differentiated murine tracheal epithelium. *American journal of physiology. Lung cellular and molecular physiology* **279**, L766–78 (2000).

48. Gupta, A., Meng, X., Zhu, L. J., Lawson, N. D. & Wolfe, S. a Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic acids research* **39**, 381–92 (2011).

49. Sander, J. D. *et al.* ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic acids research* **38**, W462–8 (2010).

50. Jiang, W., Wang, J. & Zhang, Y. Histone H3K27me3 demethylases KDM6A and KDM6B modulate definitive endoderm differentiation from human ESCs by regulating WNT signaling pathway. *Cell research* 1–9 (2012).doi:10.1038/cr.2012.119

51. Burns, J. L. *et al.* Longitudinal assessment of Pseudomonas aeruginosa in young children with cystic fibrosis. *The Journal of infectious diseases* **183**, 444–452 (2001).

52. Bauernfeind, A. *et al.* Qualitative and quantitative microbiological analysis of sputa of 102 patients with cystic fibrosis. *Infection* **15**, 270–277 (1987).

53. Bittar, F. *et al.* Molecular detection of multiple emerging pathogens in sputa from cystic fibrosis patients. *PloS one* **3**, 1–7 (2008).

54. Prince, A. Flagellar activation of epithelial signaling. *American journal of respiratory cell and molecular biology* **34**, 548–51 (2006).

55. Zhang, J., Xu, K., Ambati, B. & Yu, F.-S. X. Toll-like Receptor 5-Mediated Corneal Epithelial Inflammatory Responses to Pseudomonas aeruginosa Flagellin. *Investigative Ophthalmology & Visual Science* **44**, 4247–4254 (2003).

56. Shanks, K. K., Guang, W., Kim, K. C. & Lillehoj, E. P. Interleukin-8 production by human airway epithelial cells in response to Pseudomonas aeruginosa clinical isolates expressing type a or type b flagellins. *Clinical and vaccine immunology* **17**, 1196–1202 (2010).

57. Bonfield, T. L., Konstan, M. W. & Berger, M. Altered respiratory epithelial cell cytokine production in cystic fibrosis. *The Journal of allergy and clinical immunology* **104**, 72–8 (1999).

58. Kube, D., Sontich, U., Fletcher, D. & Davis, P. B. Proinflammatory cytokine responses to P . aeruginosa infection in human airway epithelial cell lines. *American Journal of physiology. Lung cellular and molecular physiology* **280**, L493–L502 (2001).

59. Hybiske, K. *et al.* Effects of cystic fibrosis transmembrane conductance regulator and DeltaF508CFTR on inflammatory response, ER stress, and Ca2+ of airway epithelia. *American journal of physiology. Lung cellular and molecular physiology* **293**, L1250–60 (2007).

60. Osika, E. *et al.* Distinct sputum cytokine profiles in cystic fibrosis and other chronic inflammatory airway disease. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* **14**, 339–46 (1999).

61. Riches, D. W., Sawyer, R. T., Fenton, M. J. & Martin, T. R. Innate Immunity in the Lungs. *Murray and Nadel's Textbook of Respiratory Medicine* 255–284 (2010).