NORTHWESTERN UNIVERSITY

Ets Homologous Factor Regulates Genes Involved in

Airway Epithelial Development, Function and Response to Injury

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Biological Sciences: Driskill Graduate Training Program in the Life Sciences (DGP)

By

Sara Fossum

EVANSTON, ILLINOIS

December 2016

ABSTRACT

Ets Homologous Factor Regulates Genes Involved in

Airway Epithelial Development, Function and Response to Injury

Sara Fossum

The airway epithelium forms an active barrier between the internal and external environments and has multiple roles critical to normal function of the lung. Its development and function are controlled by a network of transcription factors that regulate gene expression in response to varying stimuli. Dysfunction of this tissue plays an important role in pathogenesis of multiple pulmonary diseases, including cystic fibrosis (CF). Ets homologous factor (EHF) is a transcription factor that is expressed in the airway epithelium, and its levels are altered in inflammatory states. *EHF* maps adjacent to an intergenic region that has been implicated as a modifier of CF airway disease through a replicated genome-wide association study. The role of EHF in regulation of gene expression in the bronchial epithelium was investigated. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) identified a binding signature for EHF in Calu-3 lung adenocarcinoma and primary human bronchial epithelial (HBE) cells. EHF binding sites were enriched at promoters and for corresponding histone modifications. In Calu-3 cells, regions of EHF occupancy overlap those of activator protein-1 (AP-1) genome-wide, and these sites are found near loci important for epithelial differentiation and wound repair.

EHF modulation followed by RNA-sequencing (RNA-seq) in Calu-3 cells and A549 lung carcinoma cells altered expression of genes involved in epithelial differentiation, barrier function and response to wounding. RNA-seq after EHF depletion in primary HBE cells identified putative direct targets that are enriched for genes involved in degradation of response to wounding, the immune response, and gene regulation. EHF expression is increased by the Th2 interleukin IL-13. EHF increases levels of the SAM pointed domain-containing ETS transcription factor (SPDEF), a target of IL-13-induced signaling that contributes to goblet cell hyperplasia in the airway epithelium. Furthermore, EHF represses the cystic fibrosis transmembrane conductance regulator (*CFTR*) locus. Changes in gene expression correspond to

alterations in cell phenotype; EHF depletion slows wound repair and alters secretion of a neutrophil chemokine. In summary, these results support a strong role for EHF in regulating epithelial function in CF and other airway diseases.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Ann Harris, for your guidance over the last 5 years. You taught me how to ask interesting questions and design well-controlled experiments to answer them, how to think critically about research and results, and how to write grants and manuscripts. I appreciate the time and energy you have invested into my development as a scientist.

I also thank the other members of the Harris lab, both past and present, for your support and friendship.   I would especially like to acknowledge Dr. Shih-Hsing Leir, who was an invaluable mentor throughout my years in the lab, and Mike Mutolo, who was central to getting this project running.

I thank my thesis committee, Dr. Jason Brickner, Dr. Jindan Yu, and Dr. Debabrata Chakravarti, for your helpful discussions and commitment to overseeing my progress throughout my PhD.

Finally, I would like to thank my family and friends, especially my parents, Brad and Gloria, for supporting me throughout my schooling, understanding my desire to pursue this path, and raising me to recognize the importance of hard work and dedication. To my husband, Dr. Joe Maslak, thank you for following me to Chicago and believing in me always. To my sisters, Andrea and Kari, thanks for being my best friends and keeping me grounded. To my new family, the Maslaks, thank you for your care and encouragement. And finally, to my Chicago friends, thank you for being my local support system.

LIST OF ABBREVIATIONS

ADMA	Asymmetric dimethylarginine
AGAP1	ArfGAP with GTPase domain, ankyrin repeat and PH domain 1
AHR	Airway hyperreactivity
AJ	Adherens junction
ALI	Air-liquid interface
AP-1	Activator Protein-1
ASL	Air-surface liquid
ATF	Activating transcription factor
ATP	Adenosine triphosphate
B2M	β-2-microglobulin
BALF	Bronchoalveolar lavage fluid
BEGM	Bronchial epithelial growth medium
BETA	Binding and Expression Target Analysis
Bp	Base pair
bZIP	Basic region-leucine zipper
CBP	CREB binding protein
CCL5	Chemokine (C-C Motif) ligand 5
CCSP	Clara cell secretory protein
CEAS	Cis-regulatory Element Annotation System
CEBP	CCAAT/enhancer binding protein
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
ChIP	Chromatin immunoprecipitation
ChIP-qPCR	ChIP followed by quantitative PCR
ChIP-seq	ChIP followed by deep sequencing
CM	Conditioned media
co-IP	Co-immunoprecipitation
COL	Collagen
COPD	Chronic obstructive pulmonary disease
CRE	cAMP response element
CRISPR	Clustered regularly interspaced short palindromic repeats
CXCL	Chemokine (C-X-C motif) ligand
DAVID	Database for Annotation, Visualization and Integrated Discovery
DC	Dendritic cell
DDAH1	Dimethylarginine dimethylaminohydrolase 1
DDIT3	DNA Damage Inducible Transcript 3
DEG	Differentially expressed gene
DHS	DNase hypersensitive site
DMEM	Dulbecco's Modified Eagle's Medium, low glucose
DNase-seq	DNase I digestion followed by sequencing
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylenediamine tetraacetic acid
EGF	Epidermal growth factor
EGS	Ethylene glycolbis[succinimidyl succinate]
EHF	Ets homologous factor
ELF	E74 like ETS transcription factor
ELISA	Enzyme-linked immunosorbent assay
ENCODE	Encyclopedia of DNA Elements
EMSA	Electrophoretic mobility shift assay

EMT	Epithelial to mesenchymal transition
ENaC	Epithelial sodium channel
eQTLS	Expression quantitative trait loci
ERBB2	Erb-b2 receptor tyrosine kinase 2
ERK	Extracellular signal-regulated kinase
ESE	Epithelial-specific Ets subfamily
FBS	Fetal bovine serum
FDR	False discovery rate
FEV_1	Forced expiratory volume in 1 second
FOX	Forkhead box
FPKM	Fragments per kilobase of transcript per million fragments mapped
FZD1	Frizzled class receptor 1
GATA6	GATA binding protein 6
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GO	Gene ontology
GWAS	Genome wide association study
H3K27ac	Histone 3, lysine 27 acetylation
H3K27me3	Histone 3, lysine 27 trimethylation
H3K4me1	Histone 3, lysine 4 monomethylation
H3K4me3	Histone 3, lysine 4 trimethylation
HBE	Human bronchial epithelial cells
HEK	Human embryonic kidney 293 cells
HLA-DMA	Major histocompatibility complex, class II, DM alpha
HOPX	HOP homeobox
HRP	Horseradish peroxidase
HTE	Human tracheal epithelial cells
IB	Inclusion body
ID2	Inhibitor of DNA binding 2
IDR	Irreproducible discovery rate
IFN	Interferon
IL	Interleukin
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRAK	Interleukin 1 receptor associated kinase
IRF1	Interferon regulatory factor 1
ITGA2	Integrin, α2
JDP2	Jun dimerization protein 2
JNK	c-Jun NH2-terminal kinase
Kb	Kilobase
KD	Knockdown
KLF	Krüppel-like factor
KRT	Keratin
LB	Lysogeny broth
LIC	Ligation independent cloning
LPS	Lipopolysaccharide
LT	Life Technologies
MAPK	Mitogen-activated protein kinase
MD-2	Myeloid differentiation protein-2
MDS	Multi-dimensional scaling
MMP	Matrix metallopeptidase
MS	Mass spectrometry

MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, inner salt
MUC	Mucin
MYC	V-myc avian myelocytomatosis viral oncogene homolog
MyD88	Myeloid differentiation primary response 88
NC	Negative control
NF-κB	Nuclear factor Kappa B
NF-Y	Nuclear factor Y
Ni-NTA	Nickel charged Tris-NTA
NTA	Dip and Read™ Ni-NTA biosensor
Nm	Nanomoles
NOS	Nitric oxide synthase
OAS	2'-5'-oligoadenylate synthetase
PAMP	Pathogen-associated molecular patterns
PBS	Phosphate buffered saline
PBST	PBS with 0.1% Tween 20
PC	Principle component
PCA	Principle component analysis
PCR	Polymerase chain reaction
PIC	Protease inhibitor cocktail
PLAU	Plasminogen activator, urokinase
PLAUR	PLAU receptor
PLCH1	Phospholipase C, eta 1
PMS	Phenazine methosulfate
PNEC	Pulmonary neuroendocrine cell
PNT	Pointed domain
Pol II	RNA polymerase II
qPCR	Quantitative polymerase chain reaction
PRR	Pattern-recognition receptors
RARB	Retinoic acid receptor beta
RBM	Reticular basement membrane
RTK	Receptor tyrosine kinase
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
S100	S100 calcium binding protein
SC	Santa Cruz
SD	Standard deviation
SDC1	Syndecan 1
SEM	Standard error of the mean
SLC	Solute carrier
SOX	SRY-box
SP1	Specificity protein 1
SPDEF	SAM pointed domain-containing ETS transcription factor
SUMO	Small ubiquitin-like modifier
TAE	Tris/Acetic Acid/EDTA
TAD	Topologically associating domain
TBP	TATA-binding protein
TBS	Tris-buffered saline
TEMED	Tetramethylethylenediamine
TER	Transepithelial resistance
TF	Transcription factor
Th2	Type 2 T helper cell
Th17	Type 17 T helper cell

TICAM1	Toll like receptor adaptor molecule 1
TJ	Tight junction
TJP	Tight junction protein
TLR	Toll-like receptor
TLSP	Thymic stromal lymphopoietin
TIMP2	TIMP metallopeptidase inhibitor 2
TNF-α	Tumor necrosis factor-α
TNS	Trypsin Neutralizing Solution
TPA	12-O-tetradecanoylphorbol-13-acetate
TRAF6	TNF receptor associated factor 6
HDAC1	Histone deacetylase 1
TRIM5	Tripartite motif-containing protein 5
TSS	Transcription start site
UBB	Ubiquitin B
uPA	Plasminogen Activator, Urokinase
uPAR	uPA receptor
UTR	Untranslated region
WT	Wild type
ZO	Zonula occludens

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1. INTRODUCTION

1A The lung epithelium in health and disease

1A.1 Lung epithelial biology

The lungs are responsible for respiration, the exchange of oxygen (O_2) and carbon dioxide (CO_2). O_2 diffuses from the air to the pulmonary capillary blood, and CO_2 moves in the opposite direction. In eukaryotic cells, O_2 is consumed during oxidative phosphorylation, and CO_2 is generated as a waste product of the citric acid cycle. Both of these processes are necessary to generate adenosine triphosphate (ATP), a major energy source for cells (1). Thus, respiration is necessary for the body to generate energy.

The lungs have evolved into a progressively bifurcating structure that expands the surface over which O_2 and CO_2 gases can be exchanged; at each generation, the diameter of the airway decreases while the total surface area increases (1). The airway begins with one trachea, which branches into the left and right mainstem bronchi. Then come multiple generations of bronchi, which contain cartilage, followed by cartilage-free bronchioles. The trachea, bronchi, and bronchioles are all conducting airways; they do not participate in gas exchange. Following the terminal bronchioles, alveoli begin to bud off, and eventually the airway is completely lined with alveoli (1). The alveolus is the site of gas exchange.

The airways are lined with a layer of epithelial cells. Although it is continuous from the trachea through the alveoli, the types of differentiated cell types that make up this layer change. The major cell types found in the conducting airways include ciliated, mucous, basal, Clara and pulmonary neuroendocrine cells (PNECs) (2,3). There are also submucosal glands, which have highest density in the upper airways (4). Ciliated columnar cells are the most common cell type in the airways, and function to transport mucus out of the lung. Goblet cells release acid mucins from granules. Basal cells are involved in columnar cell attachment to the basement membrane, and are thought to be the primary stem cell of the lung. Clara cells secrete bronchiolar surfactant and anti-proteases. The trachea and large

bronchi are lined with pseudostratified cells, which become cuboidal in the smaller airways. Furthermore, as the larger airways branch to smaller ones, goblet and basal cells become less frequent, and secretory Clara cells become more common (5). The alveoli are lined with 2 types of differentiated epithelial cells; type I alveolar pneumocytes are the major route of gas diffusion, and type II pneumocytes secrete pulmonary surfactant, which eases expansion of the lungs during inspiration (1). The lung epithelium forms an active barrier between the external and internal environments, and serves many critical roles in airway function (Figure 1A.1).

B. Mucociliary escalator

C. Pathogen recognition

D. Wound repair

A. Barrier function
(cell-cell adhesions)

i.
ii.
iii.

IL-4
IL-13
IL-17a

IL-8
CXCL1
CXCL6

C. Inflammation
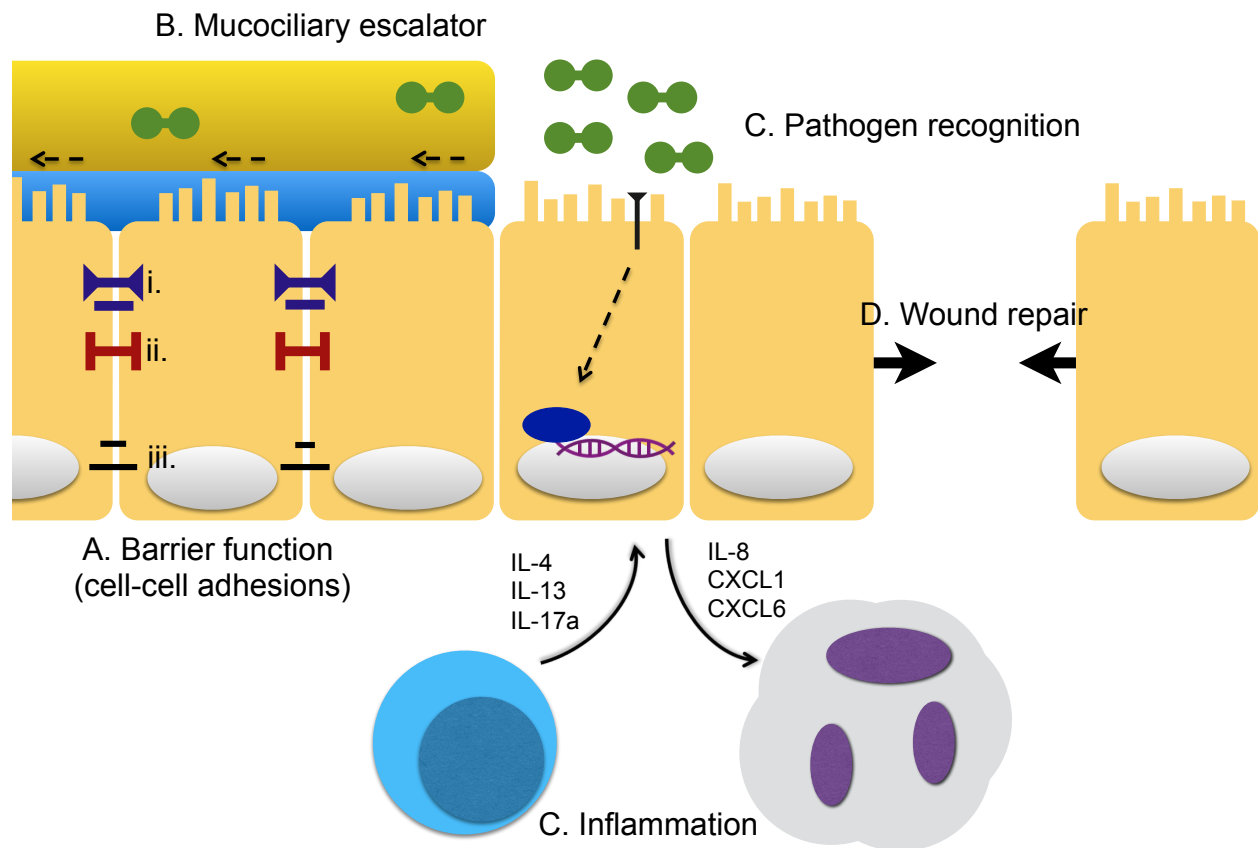
Figure 1A.1 Function of the airway epithelium. The epithelial cells perform several important roles in the lung. A) They form an active barrier between the internal and external environments through junctional complexes with neighboring cells, including i) tight junctions ii) adherens junctions, and iii) desmosomes. B) The airway epithelium contributes to the air-surface liquid (ASL) and drives the mucociliary escalator to remove foreign debris and pathogens from the lung. C) The epithelial cells respond to pathogens and cytokines, including interleukins released by T cells, leading to intracellular signaling and transcription factor regulation of expression of inflammatory genes. Bronchial epithelial cells secrete chemokines that recruit neutrophils to the lung tissue, including IL-8, CXCL1, and CXCL6. D) Following damage, the lung epithelium initiates multiple pathways to repair the wound.

1A.1.1 Epithelial barrier function

The airway epithelium provides a physical barrier between the outside environment and other tissues within the lung (Figure 1A.1 A). The structural integrity of this tissue is maintained by intercellular tight junctions, and by additional mechanisms that adhere the epithelial cells to each other and to the underlying basement membrane. Epithelial cells form junctional complexes with neighboring cells composed of the apical tight junctions (TJs) and the underlying adherens junctions (AJs) (6). TJs, which are the main determinants of paracellular permeability, contain a multitude of proteins, including the transmembrane claudins and occludin. The AJs are formed by the transmembrane proteins E-cadherin and nectin family members. The cytosolic zonula occludens (ZO) link the transmembrane junction proteins to the cytoskeleton (6). The most basal adhesions are the desmosomes, which consist of non-classical cadherins and connect to the filamentous cytoskeleton (7). The heterodimeric integrin proteins interact with extracellular matrix (ECM). A subset of these proteins serves to adhere to the ECM, while others bind ECM proteins for the purpose of cell signaling. The epithelial cells also adhere to the basement membrane, which forms a barrier between this cell layer and the underlying mesenchyme and serves to anchor the epithelium (2). Epithelial barrier function can be studied *in vitro* by polarizing epithelial cells on semi-permeable membranes (5). Permeability can then be determined by measuring transepithelial resistance (TER). Airway epithelial barrier function is altered in airway diseases including CF and asthma (6,8), and is regulated by cytokines (9,10).

1A.1.2 Air-surface liquid and the mucociliary escalator

Because of its constant exposure to foreign pathogens and particles, the lung has developed a mechanism to clear external debris, termed the mucociliary escalator (Figure 1A.1 B). The airway epithelium contributes to the production and homeostasis of airway surface liquid (ASL), which consists of a mucus gel overlaying a watery periciliary layer (11). The mucins forming the gel, including MUC5AC and MUC5B, are secreted in the large airways by surface goblet cells and submucosal glands, and more distally by other secretory cells (4,12). The mucus-secreting capacity is highest in the glands, followed by

goblet then Clara cells (4). The pericilliary layer contains tethered mucins that form a densely packed

glycocalyx, which serves as a selective barrier. MUC1 localizes to microvilli, while MUC4 and MUC20 are

found at the cilia shaft; MUC16 is tethered to the apical membrane of goblet cells (12). The volume of this

fluid is determined by the NaCl concentration in the airway lumen, which is in turn partially controlled by

sodium absorption through the epithelial sodium channel (ENaC) and chloride secretion via the cystic

fibrosis transmembrane conductance regulator (CFTR) and calcium-activated chloride channels (13). The

thickness of the periciliary layer is critical for proper propulsion of the overlaying mucus layer by the cilia.

The mucus layer forms a barrier for foreign pathogens and particles; debris becomes trapped in the gel,

which is then propelled from the distal to the proximal airways, where it is swallowed or coughed out of

the airway (13). The ASL also contains anti-microbial substances, secreted by secretory epithelial cells,

that kill and inhibit growth of bacteria during the time period that it takes the mucociliary escalator to clear

them (14).

The ASL is greatly altered in airway diseases including CF asthma, and chronic obstructive

pulmonary disease (COPD). Dysfunction in CF is characterized by an abnormal mucus layer and reduced

periciliary layer, impeding ciliary clearance of pathogens (14,15). There is increased secretion of

MUC5AC and MUC5B in the CF airway, particularly during disease exacerbations, and MUC5AC and the

tethered MUC2 are upregulated by infectious agents in the lung (16). In asthma, increased mucin

secretion by the epithelia leads to alteration in the consistency of the mucus gel layer and mucus plugging

of the airways. In COPD, many factors contribute to increased mucin production and decreased ASL

hydration, leading to airway obstruction (13).

1A.1.3 The immune response and inflammation

The pulmonary epithelium is partially responsible for detecting pathogens in air inhaled into the

lungs and for initiating an appropriate inflammatory response (Figure 1A.1 C). Pattern-recognition

receptors (PRRs) detect highly conserved portions of pathogens called pathogen-associated molecular

patterns (PAMPs), and then initiate a signaling cascade that leads to expression of inflammatory genes

(17). Examples of PRRs are the Toll-like receptors (TLRs). Surface TLRs mainly recognize components of microbial membranes, while TLRs expressed on intracellular vesicles identify microbial nucleic acids (17). This TLR response is part of the innate immune system.

To initiate infection, inhaled respiratory bacteria and viruses bind to specific receptors on the surface of airway epithelial cells, allowing them to colonize the epithelial surface (bacteria) or enter the cells and replicate intracellularly (bacteria and viruses). This stimulates the innate immune system and the interferon (IFN) response (7). The airway epithelium is also a link between innate and adaptive immunity. In response to bacteria, viruses, and inflammatory mediators, the airway epithelial cells release multiple inflammatory cytokines and chemokines, which contribute to both branches of the immune system (7). The role of the airway epithelium in immunity is complex; three specific responses will be discussed in subsequent sections.

1A.1.3.i Type 2 immune response

The type 2 immune response evolved to protect against parasites, but also contributes to allergic inflammatory diseases. In the airway, this response has mainly been studied in the context of asthma. Following lung damage or exposure to allergens or pathogens, the lung epithelium releases cytokines including interleukin 25 (IL-25), IL-33 and thymic stromal lymphopoietin (TLSP) (18). These mediators activate a subset of innate lymphoid cells, ILC2s. ILC2 cells release IL-13, recruiting a subset of dendritic cells (DCs) (19). The DCs activate T cells, generating type 2 T helper (Th2) cells. The Th2 and ILC2 cells release classic type 2 cytokines, including IL-4 and IL-13 (18) (Figure 1A.1.3.i).
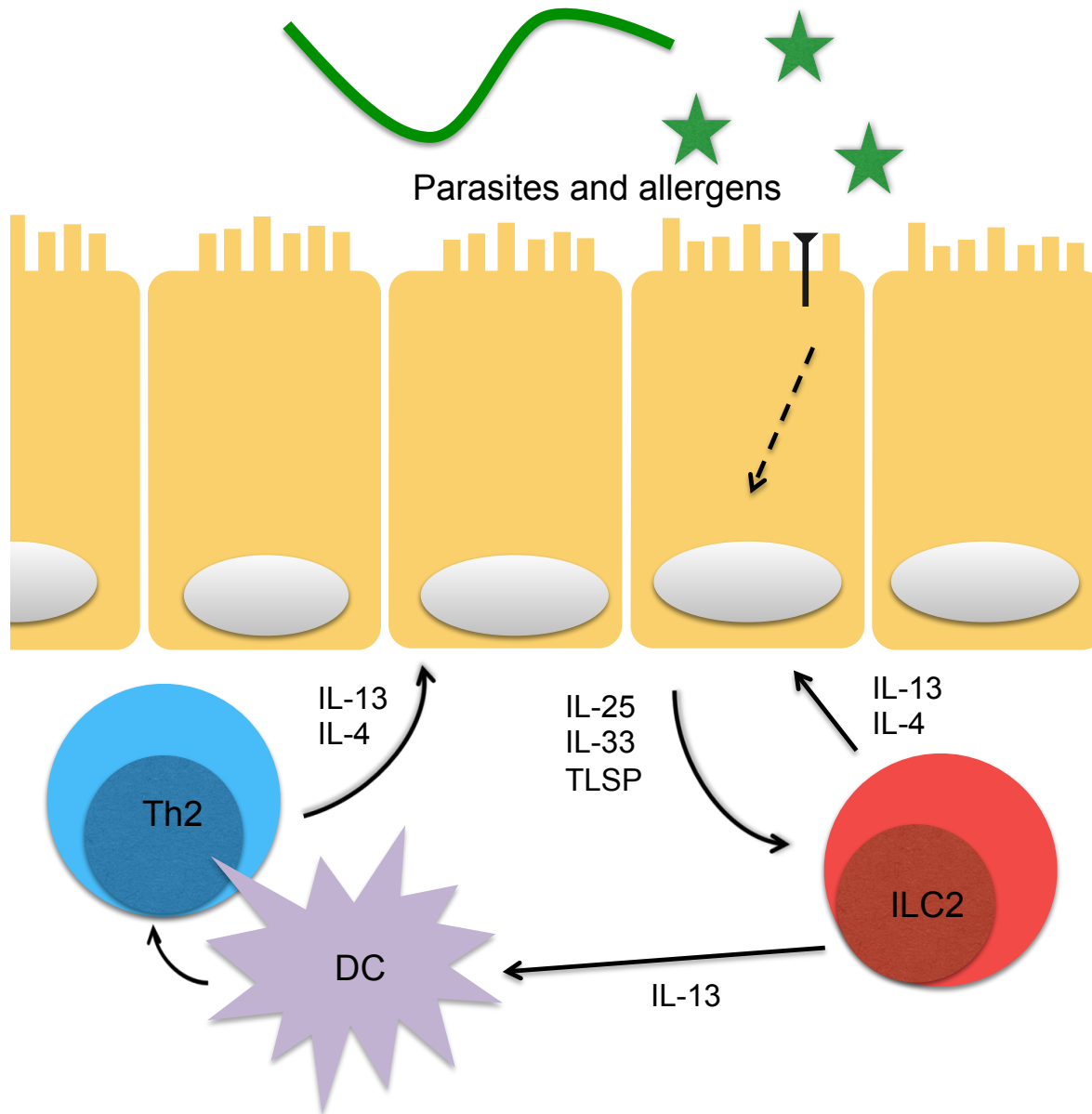
Figure 1A.1.3.i The type 2 immune response in the airway epithelium. The airway epithelium responds to pathogens and allergens by releasing cytokines that stimulate innate lymphoid cells (ILC2s), which in turn release IL-13 and recruit a subset of dendritic cells (DCs). These DCs cause activation of type 2 helper (Th2) cells and release of more type 2 immune mediators.

IL-4 and IL-13 feed back to affect lung epithelial function. When airway epithelial cells are exposed to either cytokine, intercellular permeability is increased with a corresponding reduction in expression of the tight junction proteins zonula occludens 1 (ZO-1) and occludin (9). Treatment of lung epithelial cells with IL-4 and IL-13 also slows wound closure (9). IL-13 is thought to be especially critical in the pathogenesis of asthma. In mice, IL-13 action on epithelial cells is necessary and sufficient to induce airway hyperreactivity and mucus overproduction (20). In human bronchial epithelial cells, IL-13 treatment induces gel-forming and tethered mucin gene expression and differentiation to goblet cells (21). At a molecular level, IL-13 signaling is mediated through STAT6, and acts upon the SAM pointed domain-containing ETS transcription factor (SPDEF) to cause changes in gene expression (22).

IL-13 has been associated with multiple airway diseases in humans. Following allergen challenge, the IL-13 concentration in the bronchoalveolar lavage fluid (BALF) is significantly higher in asthmatic patients than in healthy controls (23). This implies a role for IL-13 in the allergic response in asthma, consistent with findings in mice and *in vitro* studies. IL-13 in sputum samples is increased in patients with COPD as compared with healthy controls (24). In patients with CF, IL-13 concentration in airway BALF is higher than in healthy controls, and is highest in CF patients with pulmonary exacerbation and *Pseudomonas aeruginosa* infection. In these patients, there is a correlation between IL-13 concentration and structural changes in the airway and lung disease severity (24-26). This suggests that IL-13 may be involved in the pathogenesis of CF airway disease.

1A.1.3.ii Th17 response

IL-17-secreting T helper (Th17) cells are generated by exposure to IL-23 in the absence of IL-4 and IFN-γ. In the lung, chronic IL-17a overexpression results in leukocyte infiltration and changes to the tissue including altered ECM and increased mucus production (27). IL-17a acts on the bronchial epithelium to increase release of multiple cytokines, including the neutrophil chemoattractants IL-8 and chemokine (C-X-C motif) ligands 1 and 6 (CXCL1 and CXCL6) (28,29), which has implicated IL-17a as a

mediator of neutrophil recruitment. Indeed, loss of IL-17 in the mouse leads to impaired neutrophil recruitment to the alveolar space (30).

IL-17a has been associated with CF lung disease in humans. IL-17a-secreting T cells are found in the airway submucosa of patients with CF, both those that are newly diagnosed and those with later-stage disease. Furthermore, IL-17a concentration is higher in the BALF of patients with CF compared with healthy controls, and higher in CF patients with pulmonary exacerbation than without. IL-17a concentration is correlated with increased airway neutrophil recruitment (26,31). These findings suggest it could be an early mediator of CF pathogenesis through recruitment of neutrophils to the lung.

1A.1.3.iii Lipopolysaccharide (LPS)-induced immune response

Lipopolysaccharide (LPS) is a major component of the outer membrane of Gram-negative bacteria. There are multiple common airway pathogens that express LPS, including *P. aeruginosa, Haemophilus influenza,* and *Klebsiella pneumonia* (32). Thus, the airway epithelial cells have evolved an immune response to LPS exposure. The major receptor for LPS is TLR4 in complex with myeloid differentiation protein-2 (MD-2), which is expressed in the intracellular compartment of airway epithelial cells. When this complex interacts with internalized LPS, it initiates a signaling pathway involving the intermediates myeloid differentiation primary response 88 (MyD88), interleukin 1 receptor associated kinase (IRAK), and TNF receptor associated factor 6 (TRAF6), culminating in activation of the transcription factor nuclear factor Kappa B (NF-κB) and increased expression of pro-inflammatory genes, including *IL8* (33-35).

Impaired TLR4 signaling results in a decreased ability to clear bacteria and diminished survival in response to increased bacterial load (32), illustrating how important the response to LPS is to proper immune function in the lung. Children with CF have repeated respiratory infections with *H. influenza*, and *P. aeruginosa* is the predominant infection found in CF patients (36). Thus the response to LPS is likely critical in this population.

1A.1.4 Wound repair

Following damage due to infection, inflammation, trauma, or other stressors, the lung epithelium initiates multiple pathways to repair the wound (Figure 1A.1 D). Nearby epithelial cells migrate to spread over the denuded area, then proliferate and differentiate to restore functioning tissue (37,38). The underlying ECM is remodeled, and integrins on the epithelial cells interact with it to regulate cell migration (39,40). Matrix metallopeptidases (MMPs) are upregulated by the migrating epithelial cells and participate in ECM remodeling and migration (41,42). Wound closure is altered in response to signaling molecules including epidermal growth factor (EGF); the coagulation cascade proteins tissue factor, fibrinogen, and FXIII subunit A (FXIIIA); and the cytokines IL-1β, IL-4, IL-13, and IFN-γ (9,38,43).

Following injury, the lung epithelial cells can activate pathways leading to normal repair and regeneration, or they may activate aberrant pathways leading to remodeling and altered differentiation of the epithelium. The pathways responsible for normal and abnormal wound repair are thought to overlap with those involved in lung development and epithelial to mesenchymal transition (EMT) (44). In EMT, epithelial cells lose contact with adjacent cells and become more motile (45). EMT is a cancer-associated process that contributes to tumor metastasis and drug resistance (46). Similar pathways may be involved in fibrosis, which is the result of aberrant wound repair (47). Altered wound repair is a hallmark of multiple airway diseases, including COPD, chronic asthma, and CF (44,48,49).

1A.2 Diseases of the airway epithelium

Dysfunction of the airway epithelium is central to the pathogenesis of multiple airway diseases. Abnormal epithelial function plays a role in cystic fibrosis, asthma, and COPD.

1A.2.1 Cystic fibrosis

1A.2.1.i Epidemiology

CF is an autosomal recessive disease caused by mutations in the CFTR gene (50,51). It is the second most common lethal inherited disorder in the United States (52). There are nearly 29,000 people living with CF in the US (53). An estimated 1 in 29 Americans are carriers of the mutations; they have one mutated allele but do not have clinical symptoms of CF (54). This disease significantly decreases life expectancy; the median predicted age of survival was 39.3 years in 2014 (53). The most common cause of mortality in these patients is respiratory failure, which causes about 70% of deaths (53). The decreased life expectancy combined with the frequency of the mutation amongst the US population and lack of a cure for the most common mutation make CF a critical disease to understand.

1A.2.1.ii Molecular defect

CFTR is a chloride channel that is regulated in a cAMP-dependent manner (55-57). It also conducts bicarbonate and decreases sodium transport through the epithelial sodium channel (ENaC) (36,57). CFTR mutations fall into 6 different classes based on the effect they have on the protein or transcript: no functional protein (Class I), CFTR trafficking defect (II), defective channel regulation (III), decreased conductance through the channel (IV), reduced synthesis of the *CFTR* transcript (V), and reduced stability of the transcript (VI) (36). The more severe class I, II, and III mutations generally lack residual functioning CFTR at the cell surface, whereas the others have some remaining CFTR and a milder phenotype (36). The most common mutation, F508del, is a class II defect in trafficking; however, a small proportion is properly localized to the membrane, where it shows a defect in channel activity. Eighty-six percent of CF patients carry at least one copy of F508del, and 46.4% are homozygous for the allele (36).

1A.2.1.iii Lung disease in CF

Although CF affects epithelial function in many different organ systems, the major cause of mortality is lung dysfunction (52). This impaired function of the airway is complex; altered ion transport across the epithelia leads to a decreased ability to clear microorganisms through the mucociliary escalator, chronic infection, and inflammation. Although there is no increase in the number of submucosal glands and goblet cells, both are hypertrophied in CF lungs (4,58). Due to altered ion transport, their secretions are abnormally viscous and tether to the epithelium (15). The ASL is decreased in depth, and mucociliary transport is severely delayed (59). Furthermore, altered acidity of the secretions impairs killing of microbes (60). Bacteria colonize the abnormal secretions, leading to inflammation and structural damage (4).

Airway inflammation, resulting from chronic infection and an innate immune defect from CFTR loss, contributes significantly to CF pulmonary disease pathogenesis (61). Lung pathology in these patients is characterized by early neutrophil-dominated inflammation (62,63). Neutrophils release proteases, which damage the lung epithelium (64,65). Immune cells and pathogens such as *P. aeruginosa* cause an increase in the release of inflammatory mediators, which activates the transcription factor NF-κB. This leads to the production of more cytokines and an amplified cycle of inflammation (25,26,31,66,67). Inflammation and chronic infection lead to destruction of lung tissue. This damage causes remodeling and fibrosis, which are correlated with pulmonary dysfunction (68-70). Progressive bronchiectasis, mucus plugging and air trapping are prominent structural features of the disease (4).

1A.2.2 Asthma

Asthma is a chronic respiratory illness that affects approximately 300 million people worldwide (71). Unlike CF, the molecular defects causing it pathogenesis are not well understood. It is a heterogeneous disease characterized by recurrent episodes of reversible airway obstruction and hyperreactivity (AHR) (72). Inflammation is also a prominent feature of the disease; there is increased infiltration of eosinophils, lymphocytes, and macrophages in both the epithelium and lamina propria in the

lung of asthmatic patients as compared with healthy controls (73). Asthma is mediated by Th2 cytokines, including IL-4, IL-5, IL-9 and IL-13 (20).

The airway epithelium is central to asthma pathogenesis. It responds to the inflammatory milieu, particularly IL-13, resulting in AHR, mucus production, and goblet cell hyperplasia (20,74). Hypersecretion of mucus is highly involved in airway obstruction, which can lead to asphyxiation in severe episodes (74). There is likely also a defect in airway epithelial barrier function, including decreased expression of junctional proteins and increased intercellular permeability (6).

1A.2.3 Chronic obstructive pulmonary disease (COPD)

COPD is a chronic airway disease that is a major cause of morbidity and mortality worldwide, affecting about 64 million people (71,75). The pathogenesis of COPD is not well understood, but it is characterized by impaired airflow that is usually progressive and associated with an increased chronic inflammatory response to inhaled noxious particles, most notably cigarette smoke. The airflow limitation is due to a combination of obstructive bronchiolitis and destruction of the lung parenchyma (75).

There are increased numbers of goblet cells and inflammatory leukocytes in the epithelium of peripheral airways of patients with COPD as compared with healthy controls (76). Goblet cell hyperplasia leads to mucus hypersecretion, which in turn causes a decline in lung function. The combination of increased mucus secretion and epithelial thickening leads to airflow obstruction (77). Impaired mucus clearance is also thought to play a role in disease pathogenesis by leading to increased infection rate, which is correlated to disease severity (13).

1B Gene regulation in the lung epithelium

1B.1 Gene regulation in eukaryotes

Transcription is the process by which genomic DNA is copied into RNA. Regulation of gene transcription is tightly regulated in eukaryotes. This process begins at the transcription start site (TSS),

which is generally located at the 5' end of genes. Proteins comprising the general transcription machinery, including TATA-binding protein (TBP), mediator, TFIIA, TFIIB, TFIID, TFIIE, and TFIIH, bind to the TSS. They stabilize the unwinding of the DNA strands and recruit RNA polymerase II (Pol II) (78). Pol II is phosphorylated and then moves to the pause site. Finally, more transcription factors (TFs), co-factors, and elongation factors are recruited to the site of Pol II pausing, causing the release of Pol II and transcription of the gene (or of enhancer RNAs at *cis*-regulatory elements). Recruitment of TFs and other proteins at every step is often downstream of cell-signaling events. (79).

The accessibility of DNA for transcription is regulated by chromatin structure. Nucleosomes, the basic units of chromatin, are comprised of an octamer core containing 2 of each canonical histone protein: H2A, H2B, H3 and H4. Approximately 147 bp of DNA is wrapped around the histone octamer (80). Chromatin dynamics are partially controlled by post-translational modifications of histones; these reversible modifications are added and removed by enzymes known as writers and erasers, respectively (81). TSSs are occupied by nucleosomes when in an inactive state. To alter the chromatin to a more active conformation, specific TFs and nucleosome remodelers move the nucleosomes to the side. This makes the TSS accessible for binding of the transcription machinery (81).

The development of methods utilizing next generation sequencing, including chromatin immunoprecipitation-sequencing, DNase-sequencing, and RNA-sequencing, has allowed scientists to begin to determine the functions of non-coding regions of the genome. The Encyclopedia of DNA Elements (ENCODE) project has mined these data and integrated it across sources to map functional elements throughout the human genome (82,83). From these genome-wide studies, chromatin patterns specific to regulatory elements have emerged. Promoters, TSSs, and distal regulatory elements tend to be found in regions of DNA that are accessible to DNase digestion, termed DNase hypersensitive sites (DHS) (84). These regions represent open chromatin, which is depleted of nucleosomes. Regulatory elements, which are partially defined by histone modifications and TF binding, tend to be cell-type specific (85-87). Promoters and activating distal *cis*-regulatory elements, called enhancers, have specific characteristics distinguishing them from other genomic elements.

1B.1.1 Promoters

Promoters are sequences of DNA that determine where gene transcription begins, the direction of transcription and which strand of genomic DNA is copied into RNA. Promoters are located at the 5' end of genes immediately surrounding the TSS. The core promoter consists of the DNA sequence about 35 bp upstream of the gene and is the location of recruitment of the basal machinery for Pol II transcription. There are two major classes: broad, CpG-rich, TATA-less and narrow, TATA-box-containing promoters (83). The TATA sequence is recognized by TBP and is necessary for transcription initiation in the TATA-box-containing subtype (88). Histone modifications are found in a broad distribution around TSSs, especially in the CpG-rich promoters. The distribution is bimodal, consistent with a nucleosome-depleted central region of TF binding that is associated with active regulatory elements (83,89). Active promoters tend to be enriched for the histone 3, lysine 4 trimethylation (H3K4me3) and histone 3, lysine 27 acetylation (H3K27ac) modifications (83,90). Poised, GpC-rich promoters typically have the H3K4me3 and histone 3, lysine 27 trimethylation (H3K27me3) marks, while those that are inactive only have H3K27me3 (80). In contrast to histone modifications, transcription factors bind to a narrow region around the TSS (83).

Promoters are generally evolutionarily conserved. Some genes utilize multiple promoters, which correspond to transcripts with different 5' UTRs or first exons, differentially spliced transcripts, or the production of distinct proteins form the same locus (91). Promoters can be bi-directional, initiating transcription of two genes in opposite directions (92).

1B.1.2 Enhancers

Enhancers differ from promoters in their characteristic histone modifications and cell-type specificity. They are *cis*-acting regulatory elements, generally located distally from the TSS of genes. They are active irrespective of sequence orientation (93,94). Enhancers are nucleosome-depleted when active. Unlike promoters, poised and active enhancers tend to show enrichment of histone 3, lysine 4 monomethylation (H3K4me1), not H3K4me3 (90). H3K27ac distinguishes active from inactive and poised

enhancers (95). The chromatin state at enhancers is more variable between cell types than at promoters and correlates with cell-type-specific gene expression (93).

Because they are often located distally from TSSs, enhancers frequently work via chromatin looping, in which the enhancer communicates with the promoter through a direct interaction, forming a large loop of spacer DNA between the two genetic elements (96,97). In order for this conformation to form, stable protein-DNA complexes are produced at both sites (98). DNA-binding proteins, including CTCF and cohesin, stabilize the looped structure (99,100). The looping conformation is specifically seen in cell types expressing the gene controlled by the promoter and enhancer (97), and disruption of the long-distance interaction decreases expression of the regulated gene (99,101). These findings suggest that chromatin looping is important for cell-type-specific regulation of gene expression. Many genes are regulated by multiple enhancers, which work together to regulate expression of the loci (102). Like at promoters, TF binding is important for enhancer function (102,103).

1B.1.3 Transcription factors

Transcription factors are proteins that recognize and bind to short degenerate DNA motifs (102). They occupy genomic sites including promoters, enhancers, and other regulatory elements to control gene expression. Once bound, TFs serve multiple functions, including recruitment of other transcription factors and transcriptional co-regulators, recruitment of the Pol II machinery, stabilization of enhancer-promoter interactions, and remodeling of chromatin (79,104-106). Unlike most TFs, which can only access open chromatin, pioneer TFs are able to bind DNA that is wound around the nucleosome. They then prime the site for more rapid activation or remodel the chromatin, opening it for other factors to bind (106). TFs tend to bind genomic regions that have clusters of binding sites; the combined activity of all TFs binding at the region defines the activity of the promoter or regulatory element. The resulting transcriptional output can be a function of concentration of one TF, or TFs can interact through cooperative binding for a synergistic effect (102). Many factors influence whether a TF binds to its motif, including presence of cofactors and accessibility of the chromatin (85,107,108). The development of

genome-wide methods has allowed for a greater understanding of the characteristics of TF binding to DNA.

1B.2 Genome-wide methods to study gene regulation

Methods have been developed to identify promoters, *cis*-regulatory elements, and open chromatin sites genome-wide. They have allowed scientists to determine the patterns of TF binding and histone modifications that mark these regions (80). These data have been integrated with information about the transcriptome to determine how regulatory regions and the TFs that control them alter gene expression. Three important genome-wide methodologies are DNase I digestion followed by sequencing (DNase-seq), chromatin immunoprecipitation (ChIP) followed by sequencing (ChIP-seq), and RNA sequencing (RNA-seq).

1B.2.1 DNase-seq

Mapping of DHS is useful for the identification of regulatory elements. DNase-seq was developed to map DHS across the genome in a non-biased manner (84). In this technique, nuclei are isolated from cells and exposed to a small quantity of DNase I, which preferentially digests nucleosome-depleted DNA (109). Digested DNA ends are ligated to linkers, purified and PCR amplified to generate sequencing libraries. The sequencing reads are mapped to the genome, and DHS are identified as regions with high densities of reads (84,87).

Mapping regions of open chromatin provides useful information about the location of important regulatory elements. These areas of the genome are enriched for histone modifications, TSSs, and TF binding sites (82,84). TF-bound sites are relatively protected from DNase I cleavage as compared to the surrounding DHS, a phenomenon which can be used to predict locations of TF binding within an open chromatin site (108). Comparing DHS across diverse cell types can identify cell type-specific or –selective DHS, which are often found near genes important for differentiated function of that tissue (87).

Furthermore, intergenic SNPs that associate with disease severity in genome-wide association studies are often found in regions of open chromatin, suggesting that they may be influencing regulatory elements controlling genes important for disease pathogenesis (83,87).

1B.2.2 ChIP-seq

ChIP-seq was developed to probe for histone modifications and TF binding sites genome-wide (110,111). ChIP exploits antibodies specific to the protein or post-translational modification of interest, which are used to isolate the protein-DNA complexes. ChIP-seq combines ChIP with high-throughput DNA sequencing to look for protein-DNA interactions across the genome (110,111). The assay begins with crosslinking of the TF or histone to DNA using an agent such as formaldehyde or ethylene glycolbis[succinimidyl succinate] (EGS) (112). Alternatively, native ChIP can be performed without the use of a crosslinker (113). Then chromatin is isolated from cells and sheared to generate genome fragments between 200 and 400 bp in size. Chromatin containing the TF or histone modification of interest is enriched using immunoprecipitation with a specific antibody. Next, the crosslinks are reversed, the DNA is purified and libraries are prepared for high throughput sequencing (114,115). The short sequencing reads are aligned to the genome (116), and areas of enrichment of mapped reads are determined compared to input DNA, sheared chromatin that is not enriched for a DNA-binding protein. The significant areas of enrichment as determined by a combination of absolute signal threshold and minimum enrichment are considered to be significant binding sites for TFs or regions of histone modification (117).

Biological replicates of ChIP-seq experiments must be performed in order to determine whether experiments are reproducible and to quantify variation between samples. A study of Pol II ChIP determined that after two biological replicates of ChIP-seq, there is little gain in sensitivity (118). The ENCODE consortium has set 2 biological replicates as its standard for ChIP-seq experiments (119). In order to assess reproducibility between biological replicates and set thresholds for genuine ChIP signals, the irreproducible discovery rate (IDR) statistic has been developed (120). This method ranks the ChIP-

seq peaks within each replicate based on significance, then compares these ranks between replicates; the most significant peaks, which likely represent real ChIP signals, will be more consistent between replicates while the least significant peaks, which correspond to background noise, will be less consistent. When the consistency between ranked pairs is plotted, there is a transition from signal to noise, which serves as an internal threshold for real peaks (120). The IDR method has been endorsed by ENCODE to assess the quality of datasets (119).

Several factors are important for obtaining high-quality ChIP-seq data. The most critical is antibody quality. In order for an experiment to be successful, an antibody should be both sensitive and specific in order to enrich chromatin for the protein of interest while minimizing background (115). It is also necessary to properly lyse the cells to isolate the chromatin, a process that we have optimized in primary epithelial cells (121). Finally, it is important to properly shear chromatin to obtain DNA fragments of about 200 to 400 bp to increase precision when characterizing binding sites (122).

High-confidence TF ChIP-seq binding sites identified by the IDR method are used for downstream analysis. A *de novo* motif analysis can be used to determine which degenerate motifs are enriched among binding sites (123). The motif analysis can confirm that the correct protein was immunoprecipitated and identify novel motifs recognized by the TF or secondary motifs of putative co-regulators/co-factors of specific TF function (124). Data from ChIP-seq experiments using antibodies to different factors can be intersected to identify TFs that co-regulate gene expression through overlapping binding sites and determine whether binding sites are enriched for histone modifications (125,126). Genome-wide binding data can also be intersected with RNA-sequencing results to identify putative direct targets of TFs (127).

1B.2.3 RNA-seq

RNA-seq was developed to characterize the transcriptome of an organism (128). In this method, total RNA from cells or tissue is isolated and then enriched for poly-A containing mRNAs. cDNA is transcribed from the enriched mRNA, and then sequencing libraries are generated by end repair, tailing,

and ligation of Illumina adaptors, short nucleotide sequences that allow for PCR amplification of DNA and binding to a flow cell for deep sequencing. Libraries are sequenced and subject to downstream analysis (128).

Sequenced reads are aligned to the genome with software that maps across exon-exon junctions (116,129). The resulting fragments are assembled, annotated, and normalized to gene length and the number of million mapped reads. Finally, significantly differentially expressed transcripts between different samples are identified (129,130). RNA-seq can be applied to compare transcript expression in samples across stages of development, following treatment with pharmacologic agents, or after depletion or overexpression of a transcription factor or other regulator of transcription (131-133). Genome-wide expression data can be used to identify expression quantitative trait loci (eQTLs), genes that show changes in expression that correlate with alterations in genotypes, such as single nucleotide polymorphisms (SNPs) that associate with disease severity in genome-wide association studies (GWAS) (134). RNA-seq results can also be integrated with ChIP-seq and DNase-seq data to identify regulatory elements involved in gene expression and compare them across cellular environments (87,127,135).

1B.3 Gene regulation in the lung epithelium

It is critical to understand how gene expression is regulated in the lung epithelium because an appropriate response to the external environment is necessary to maintain a healthy airway. Abnormal or exaggerated reactions to stimuli can lead to a wide range of diseases (44). Networks of transcription factors control gene expression, thereby regulating development, the inflammatory response, and regeneration. The application of molecular and genome-wide approaches to study the critical TFs for lung epithelial differentiation and function is beginning to elucidate the molecular basis of these pathways (133).

1B.3.1 Transcription factors in development

The role of TFs in lung function begins at the earliest stages of development. Lung specification

begins at about 28 days gestation in humans. NKX2-1 expression, which is downstream of Wnt and BMP

signaling, is the first marker of anterior foregut cells that will become respiratory epithelial progenitors

(136). These cells form the trachea. After budding of the large airways, the buds expand and develop

rapidly in a process called branching morphogenesis. During this phase, proximal endoderm progenitor

cells express SRY-box 2 (SOX2), while distal progenitors express SOX9 and inhibitor of DNA binding 2

(ID2). The former will give rise to ciliated and secretory cells, while the latter will become alveolar

epithelial cells (136). Retinoic acid receptor signaling is also necessary for proper branching

morphogenesis (137). As the last generations of the peripheral airway are formed, the pulmonary

epithelium begins to differentiate. GATA binding protein 6 (GATA6) is involved in epithelial differentiation

throughout development and works synergistically with NKX2-1 (138-140). HOP homeobox (HOPX) acts

downstream of NKX2-1 to repress expression of pulmonary genes and is critical for alveolar development

(141). Krüppel-like factor 5 (KLF5) is necessary for maturation of the respiratory epithelium, especially in

the alveoli (142).

Forkhead box A1 and A2 (FOXA1 and FOXA2) are pioneer transcription factors that are

important for branching morphogenesis early in development and epithelial differentiation at later stages.

There is some redundancy in their roles. Combined depletion of both factors leads to loss of expression

of alveolar epithelial type II, ciliated and secretory cell proteins (143,144). As pioneer factors, they are

able to bind and open compacted chromatin in the absence of ATP-dependent chromatin remodeling

complexes (145). FOXA1 and FOXA2 regulate *CFTR* in the intestinal epithelium through action at *cis-*

regulatory elements. They influence chromatin accessibility, alter histone modifications, recruit other

transcription factors to these sites, and stabilize the chromatin looping at the locus (105,146).

Ets transcription factors also contribute to epithelial development. ELF5 is expressed distally in

the lung during branching morphogenesis, but its expression becomes limited to the proximal lung as

development progresses. Aberrant expression of ELF5 disrupts branching morphogenesis and

suppresses proliferation and differentiation of epithelial cells (147). SPDEF interacts with NKX2-1 to activate gene transcription in proximal airways (22). It promotes airway goblet cell differentiation and mucus production and antagonizes the action of FOXA2 (22,148). There is still much to be learned about the TF network that controls lung development, but it is clear that multiple TFs work together to specify both lung organogenesis and differentiation of epithelial cells towards different fates.

1B.3.2 Transcription factors in the immune response

Some of the transcription factors involved in lung development also participate in the immune response. SPDEF in particular is a critical downstream mediator of Th2 cytokines. Its expression is induced by allergen exposure in an IL-13-dependent manner, leading to goblet cell hyperplasia and increased mucin production (22,149). SPDEF levels are also increased by rhinovirus infection (148). Furthermore, its expression augments the Th2 response while impairing the TLR response. It increases expression of the Th2 cytokines IL-33, IL-25 and IL-13 and eosinophil chemoattractants. The result is eosinophilic pulmonary inflammation, elaboration of Th2 cells, and increased AHR (150). Conversely, it binds to and inhibits the TLR signaling adapters Myd88 and toll like receptor adaptor molecule 1 (TICAM1), dampens IL-8 secretion in response to LPS, and impairs the NF-κB-mediated changes in gene expression downstream of TLR activation (148).

NF-κB is considered a master regulator of both the innate and adaptive immune system. It is activated downstream of multiple signaling pathways, including those initiated by tumor necrosis factor-α (TNF-α), IL-1 and LPS (151). In the airway epithelium, NF-κB activation leads to release of inflammatory mediators, including IL-6, IL-8, and granulocyte-macrophage colony-stimulating factor (GM-CSF), and lung inflammation (152,153). NF-κB contributes to airway lung disease; it shows increased activation by TNF-α in CF airway epithelial cells as compared to their wild type counterparts (154).

1B.4 Regulation of *CFTR*

Because multiple classes of *CFTR* mutations are characterized by decreased transcript (Class V and VI) or residual CFTR function (F508del, class III), increasing the amount of transcribed *CFTR* has therapeutic potential; it could decrease disease severity by elevating the amount of functional protein. The *CFTR* promoter drives basal transcription of the gene, but cannot account for the varied tissue-specific expression of the locus. *Cis*-regulatory elements that control gene expression have been identified upstream and downstream of *CFTR,* and also within its introns (146,155-159). In cell types where *CFTR* is expressed, the locus forms a complex looped structure, where cell-type-specific *cis*-regulatory elements interact with each other and the promoter. In cells where it is not expressed, these conformations are not established, and only the CTCF/cohesion bound topologically associating domain (TAD) is present (155,158). Regulatory elements evident in the airway epithelium include those found at -80.1 kilobases (kb), -44 kb and -35 kb upstream of the gene, and +48.9 kb downstream of the gene, in an intron of *CTTBP2.* The -80.1 and +48.9 kb sites are thought to mark the 5' and 3' boundaries of the TAD, respectively, and are present in multiple epithelial cell types (158). In contrast, the -35 kb site is a cell-type-selective regulatory element, present in the epithelium of the airway but not that of the intestine. It encompasses an enhancer for *CFTR* that binds interferon regulatory factor 1 (IRF1) and nuclear factor Y (NF-Y) (156).

1B.5 Ets homologous factor (EHF)

Ets homologous factor (EHF) is another transcription factor expressed in the lung epithelium (160,161), although its role in controlling gene expression in this cell type has not been previously characterized. It is a member of the Ets family of transcription factors.

1B.5.1 Ets family of transcription factors

The Ets family of transcription factors is characterized by the presence of the ETS DNA-binding domain. This domain is a variant of the winged helix-turn-helix motif that recognizes the conserved core motif 5'-GGA(A/T)-3'. The major point of protein-DNA contact occurs between the GGA sequence in the major groove of the DNA and the third α-helix of the ETS domain (162,163). The ETS domain is also a site of protein-protein interactions (164), and proteins in this TF family frequently interact with co-regulatory proteins (165). Approximately 1/3 of Ets transcription factors have a second conserved domain, called the pointed (PNT) domain. This domain has diverse properties, including docking of kinases, interactions with co-factors, and self-association of proteins (163).

Ets family members are regulated by post-translational modifications. Some members are phosphorylated downstream of receptor tyrosine kinase (RTK)/RAS/mitogen-activated protein kinase (MAPK) signaling, leading to relief of repression through breaking up of transcriptional complexes, activation through recruitment of co-factors, or changes in protein conformation that alter transcriptional activity (166). Fox example, the Ets factor ELK1 is under control of signaling pathways leading to phosphorylation and sumoylation of the protein; phosphorylation increases activation of transcription, while sumoylation represses its transcriptional activity (167).

The Ets transcription factors are divided into four classes based on preference for DNA-binding sequence. Class II Ets factors primarily recognize the sequence 5'-CCCGGAAGT-3', with variation seen at the first nucleotide (163). Three transcription factors that show high homology with one another make up the epithelial-specific (ESE) subfamily of class II Ets factors: E74 like ETS transcription factor 3 (ELF3), ELF5, and EHF. This subfamily is named for the initial observation that expression of these proteins is restricted to a subset of epithelial tissues in humans (168).

1B.5.2 EHF

EHF (also called ESE-3) is an Ets transcription factor that contains a carboxy-terminal ETS DNA-binding domain and a PNT domain near the amino terminus. The ETS domain is 84% and 65% identical

to the ELF3 and ELF5 domains, respectively. Its PTN domain has lesser homology with similar domains found in other Ets factors (168). It recognizes the consensus Ets GGAA core motif (168). There are three isoforms of EHF (Figure 1B.5.2). The j form has an alternative exon 1 that differentiates it from the b and a forms. The a variant lacks exon 6 and is thereby the shortest form. All three isoforms have both the ETS and PTN domains (169).
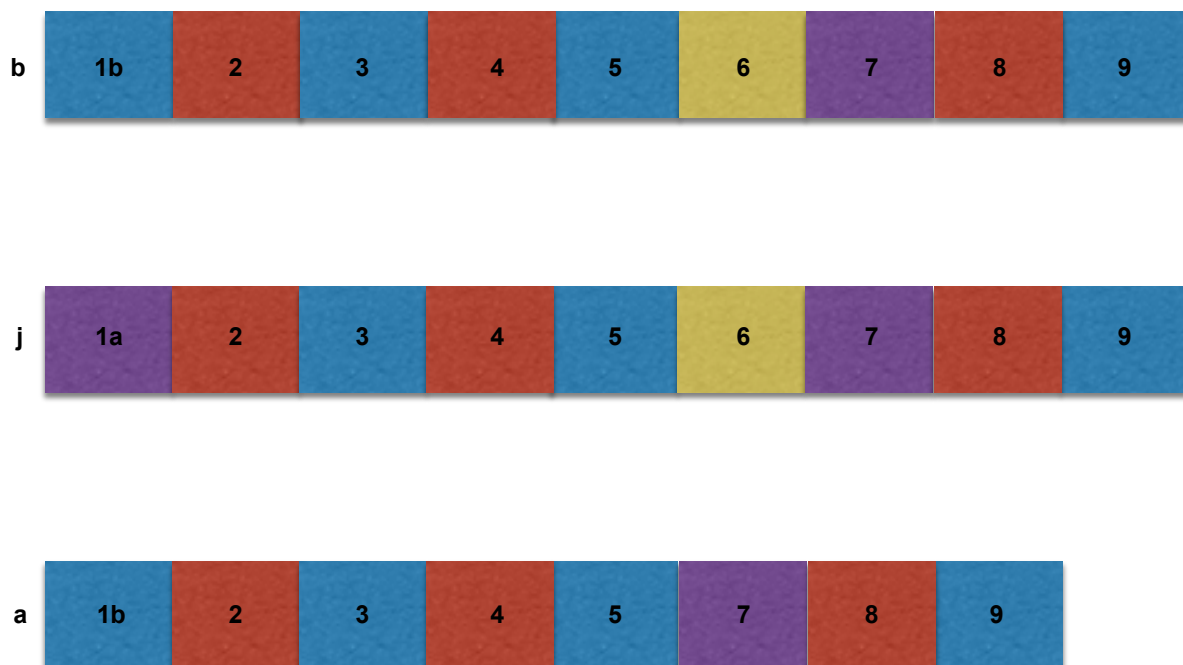
Figure 1B.5.2 Exon sequence of the EHF transcript variants. Each exon is represented by a colored box and numbered. The j form has an alternative exon 1(1a) that is marked in purple. The a form lacks exon 6 (yellow). Diagram modified from (169).

EHF has high expression in differentiated epithelial tissue, especially in cells with secretory function in the prostate, pancreas, salivary gland, and trachea (160,170). It is expressed constitutively in the bronchial epithelium, and its levels are highest in the serous and mucus glands and secretory ducts of the airways (160,161). The b variant is the main isoform found in the airway epithelium (160).

EHF can activate or repress transcription at promoters depending on the cellular context (160,171,172). Its regulation of gene expression contributes to carcinogenesis. EHF maps to chr11p13, which is deleted in some breast, prostate, and lung carcinomas (170). As expected, EHF expression is lost in a subset of bladder, prostate, oral epithelium, and breast ductal cancers (160,172,173). It is thought to be a tumor suppressor in the prostate; depletion of EHF in prostate cancer cells increases their migration and survival, with elevated expression of markers of EMT (172,174). Overall, low EHF expression in prostate tumors is correlated with a worse prognosis for patients (174). Reconstitution of EHF in these cancer cells decreases clonogenic survival and induces apoptosis by increasing expression of caspase-3 (173). Conversely, EHF overexpression in colon tumor cells is required for their survival; in these cells, it inhibits p53-mediate apoptosis (171).

Previous studies of EHF suggest a role for this protein in development. It binds to the promoters of genes involved in epithelial development *in vitro* and increases their activity (168). In the mouse corneal epithelium, *EHF* is significantly upregulated during development, and siRNA depletion of EHF leads to down-regulation of multiple genes involved in epithelial development and response to wounding (175). EHF ChIP-seq in the mouse cornea identified a genome-wide binding signature for the protein. About 1/3 of sites were located in promoters, suggesting EHF has a major role in gene regulation through binding to these regions. EHF bound to promoters of genes involved in epithelial differentiation, cell migration, and regulation of transcription (175).

Of interest to lung epithelial biology, *EHF* is regulated by cytokines in the bronchial epithelium. IL-1β and TNF-α increase *EHF* expression in this cell type through NF-κB binding at the *EHF* promoter (176). Induction of *EHF* by these cytokines also occurs in bronchial smooth muscle cells (161). This suggests that EHF has a role in inflammation in the airway.

1B.6 Activator Protein-1 (AP-1) Complex

Ets factors work with other transcription factors to regulate gene expression. EHF represses transcription of loci that have adjacent Ets and activator protein-1 (AP-1) motifs in their promoter; the repression is lost with mutation of either motif (160). This suggests that EHF works with AP-1 to repress gene expression. There is evidence that other Ets proteins interact with AP-1. The ETS domain of multiple TFs binds to the basic domain of the AP-1 protein c-Jun, and the Ets TF ELF1 co-immunoprecipitates with multiple Jun proteins (164). Adjacent Ets and AP-1 motifs are found near RAS/extracellular signal-regulated kinase (ERK) responsive genes genome-wide (177). Furthermore, the Ets factor ETV4 and the AP-1 protein JunD co-localize at binding sites throughout the genome (178).

AP-1 is a transcription factor complex made up of homo- and heterodimers of basic region-leucine zipper (bZIP) proteins of the Jun, Fos, Maf and ATF families. The dimer recognizes the 12-*O*-tetradecanoylphorbol-13-acetate (TPA) response element (5'-TGA[G/C]TCA-3') or cAMP response element (CRE, 5'-TGACGTCA-3') (179). Jun proteins can form homodimers or more stable heterodimers with members of the Fos family. They also form heterodimers with proteins in other AP-1 subfamilies (180).

In the lung there is high expression of Jun family members during development and into adulthood; the Fos protein expression is much lower (181). In response to mitogenic or toxic stimuli, expression of these proteins is rapidly induced. In the lung, AP-1 is activated in response to exposure to exogenous stimuli including cigarette smoke, airborne particulate matter, asbestos, and oxidants. Activation occurs downstream of MAPK signaling through phosphorylation by c-Jun NH2-terminal kinases (JNKs) and ERKs (181,182). ERK1 may be more active in CF cells than their wild type counterparts, leading to increased AP-1 activation and expression of inflammatory genes (183). This suggests a role for AP-1 in airway disease pathogenesis.

Once activated, AP-1 has many downstream effects in the lung. It regulates expression of differentiation markers, specifically surfactant proteins and Clara cell secretory protein (CCSP). AP-1 also

controls levels of ECM proteins, MMPS, and growth factors in the context of injury and repair (181).

Regulation of these genes makes AP-1 particularly interesting with respect to lung pathology.

1B.7 The 11p13 region and CF lung disease

Over 2,000 mutations within *CFTR* have been described, many of which are known to cause

disease (184). Specific mutations in *CFTR* do not correlate with lung disease progression (185).

However, twin and sibling studies found that airway disease is under strong genetic control (186),

implicating a role for loci other than *CFTR* in modulating phenotypic variation in CF patients. A GWAS

was performed to identify gene modifiers of lung disease severity in a cohort of 2,535 homozygous

F508del CF patients from North America. To quantify lung disease severity in subjects, a phenotype

based on multiple measures of forced expiratory volume in 1 second (FEV_1) over three years was used

(187). FEV_1 is the best predictor of mortality due to CF lung pathology (188). SNPs in non-coding regions

of chr11p13 and chr20q13.2 reached genome-wide significance as associating with lung disease severity

in the study (187). A meta-analysis of 4,139 F508del patients, which included the initial cohort plus

additional patients from North American and Europe, confirmed the association at chr11p13, and also

identified SNPs at 4 other loci that associate with the same phenotype (at chr3q29, chr5p15.3, chr6p21.3,

and chrXq22-q23) (189). Multiple SNPs are in linkage disequilibrium with the one at chr11p13 that shows

the most significant disease association, making it difficult to determine the causative SNP(s) (187).

However, it does implicate the region in modifying lung disease severity in CF patients. Since the SNPs

were found in an intergenic region, the critical variants probably impact *cis*-acting regulatory elements for

nearby genes that are required for normal lung function. These sequences are often found in regions of

open chromatin. Multiple DHS are present in primary airway epithelial cells in chr11p13, representing

putative regulatory elements (Figure 1B.7) (87). *EHF* maps to the 5' end of the critical intergenic region of

chr11p13 (Figure 1B.7). This suggests that if the disease-causing SNP impacts expression of EHF, the

TF could be a modifier of lung pathology.
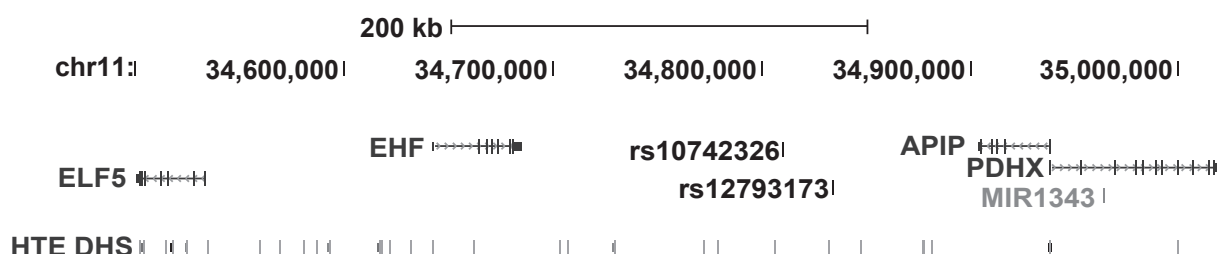
Figure 1B.7 DNase hypersensitive sites and disease-associated SNPs at chr11p13. A UCSC genome browser representation of chr11p13 displaying the most significant disease-associated SNPs from the initial GWAS (rs12793173) and meta-analysis (rs10742326) and DHS from primary human tracheal epithelial (HTE) cells (87). The nearest genes to the significant SNPs are also shown.

1C Hypothesis

Although regulation of gene expression in the lung epithelium is critical to its response to the environment, the TF regulatory network in this tissue is still only partially characterized. EHF is a transcription factor expressed in the airway epithelium, with highest levels in mucus glands, and its expression is increased in response to inflammatory mediators. Furthermore, it maps adjacent to an intergenic region of the genome implicated in lung disease severity in CF patients. We hypothesize that EHF controls lung epithelial function as part of a transcription factor network that regulates genes involved in critical biological pathways, particularly those that impact the severity of airway pathology.

2. MATERIALS AND METHODS

2.1 Utilization of cultured cells

2.1.1 Cell culture

Calu-3 lung adenocarcinoma (190), A549 lung carcinoma (191), and HEK-293 embryonic kidney (HEK) (192) cell lines were obtained from ATCC® and grown by standard methods in Dulbecco's Modified Eagle's Medium, low glucose (DMEM) with 10% fetal bovine serum (FBS).

Primary human bronchial epithelial (HBE) cells were obtained and cultured as follows. Passage 1 (P1) HBE cells were obtained from Dr. Scott Randell at the Tissue Procurement and Cell Culture Core, Marsico Lung Institute, University of North Carolina-Chapel Hill. The cells were harvested using protease XIV dissociation, and grown on collagen-coated plastic (193). HBE cells were passaged when 70-90% confluent. The cells were washed twice with phosphate buffered saline (PBS), and then 3.5 mL of 0.25% Trypsin-ethylenediamine tetraacetic acid (EDTA) (Lonza) was added to each 10 cm petri dish (2 mL for a 25 cm^2 flask). The dish was incubated at room temperature until the cell monolayer was detached (about 5 minutes). The Trypsin-EDTA was then neutralized with the same volume of Trypsin Neutralizing Solution (TNS, Lonza). The cells were centrifuged at 200 × g for 5 minutes at 4°C, and the cell pellet was resuspended in bronchial epithelial growth medium (BEGM, Lonza). The cells were seeded at 1.5 × 10^4 cells/cm^2 in BEGM. P3 or P4 cells were used for all experiments.

For air-liquid interface (ALI) experiments, HBE cells were seeded onto transwell permeable support inserts (24 well, 0.4 µm pore size, Corning). BEGM was added to the apical (0.5 mL) and basolateral (1 mL) side of the insert. Cells were grown for about 1 week until a confluent monolayer had formed, at which point ALI culture was initiated. ALI media (BEGM [with 2x supplements]: DMEM-high glucose 50:50, Pen/Strep [100 U/mL/100 µg/mL], 0.50 ng/mL EGF, and 1 mM $CaCl_2$) was added to the basolateral side of the insert (0.65 mL), and media was removed from the apical side. Cells were

polarized at ALI for 2 weeks, with media changed every other day. Transepithelial resistance (TER) was monitored using an epithelial voltohmmeter (EVOM2, World Precision Instruments).

2.1.2 Generation of EHF-overexpressing A549 clones

A549 cells were seeded into 6-well plates. Five µg of plasmid DNA (pcDNA 3.1(-) empty vector or pcDNA3.1(-)-EHF) was linearized in a 20 µL reaction with 1x NEBuffer 3.1 and 1 µL NruI (all reagents from NEB) at 37°C overnight. When A549 cells were 40% confluent, they were transfected using Lipofectin® transfection reagent (Life Technologies, LT). Briefly, 10 µL Lipofectin® was diluted into 100 µL of Opti-MEM® reduced serum media (LT) and 5 µg of linearized plasmid DNA was diluted into a separate aliquot of 100 µL Opti-MEM®. The mixtures were incubated at room temperature for 30 minutes. Lipofectin®-Opti-MEM® and DNA-Opti-MEM® were combined and incubated for 15 minutes at room temperature. Media on the A549 cells was removed and replaced with 1.8 mL of Opti-MEM®, and then 200 µL of Lipofectin®-DNA-Opti-MEM® mixture was added to each well. Cells were incubated in transfection mixture for 8 hours at 37°C, then media was removed and replaced with DMEM. Transfected cells were incubated for 72 hours at 37°C, then replaced with DMEM containing 800 µg/mL G418 (Thermo Fisher) to select for cells containing the pcDNA3.1(-) vector, which encodes for neomycin resistance. Media was replaced with fresh DMEM containing 800µg/mL G418 every 48-72 hours. Resistant clones were selected and moved to 24-well plates, then expanded into 6-well plates. EHF expression was determined by western blot.

2.1.3 siRNA depletion of transcription factors

For EHF knockdown, 30nM Silencer® Select negative control siRNA #2 (NC) or EHF siRNA (s25399, both from Ambion) were transfected into cells using RNAiMAX™ transfection reagent (LT). Calu-3 cells were transfected at 30-40% confluency, and incubated at 37°C for 48 hours (for RNA-seq and quantitative reverse transcription polymerase chain reaction [RT-qPCR]), 72 hours (for wound scratch

assay) or as otherwise described. HBE cells were transfected at 70-80% confluency and incubated at 37°C for 48 hours (for RNA-seq) or 72 hours (for RT-qPCR and wound scratch assay). HBE cultures from different donors were used for each transfection. For FOXA1 knockdown, HBE cells were treated with 20nM of control siRNA-A (sc-37007) or HNF-3α siRNA (sc-37930, both from Santa Cruz [SC]) using RNAiMax^{TM}. Transfections were incubated for 48 hours at 37°C.

For RNAiMAX^{TM} forward transfections in 24-well plates, 1µL RNAiMAX^{TM} was added to a tube containing 50µL Opti-MEM®, and the siRNA was added to a second tube containing 50µL Opti-MEM®. The RNAiMAX^{TM}-Opti-MEM® and siRNA-Opti-MEM® were combined and incubated at room temperature for 15 minutes. The media on the well was removed and replaced with 0.5mL fresh DMEM. The transfection mixtures were then added to the wells and incubated at 37°C. Transfections were scaled up for 6-well plates.

2.1.4. Lipopolysaccharide (LPS) treatment

To test the optimal length of LPS treatment, 150,000 Calu-3 cells were seeded in a 24-well plate and grown for 72 hours in DMEM with 10% FBS. The media was then removed and replaced with serum-free DMEM. Cells were serum starved for 24 hours, then treated with PBS (carrier) or 1 µg/mL *Pseudomonas aeruginosa* LPS (L9134, Sigma) for 4, 8, or 24 hours. One µg/mL LPS for 4 hours yielded the most consistent changes in gene expression and was used for subsequent studies.

Calu-3 cells were forward transfected with NC or EHF siRNA as above and grown in DMEM with 10% FBS at 37°C. Forty-eight hours after transfection, media was removed and replaced with serum-free DMEM. Cells were serum starved for 24 hours, and then treated with PBS (carrier) or 1 µg/mL *P. aeruginosa* LPS for 4 hours; RNA was then isolated for RT-qPCR.

2.1.5 IL-17a treatment

To determine the optimal concentration of IL-17a treatment, 150,000 Calu-3 cells were seeded in a 24-well plate and then grown for 72 hours in DMEM with 10% FBS. The media was removed and replaced with serum-free DMEM. Cells were serum starved for 24 hours, then treated with BSA (carrier) or 1 ng/mL, 10 ng/mL, or 100 ng/mL IL-17a (R&D Systems) for 24 hours. Treatment with 10 ng/mL of IL-17a caused an increase in target genes and was used for subsequent studies.

Calu-3 cells were forward transfected with NC or EHF siRNA as above and grown in DMEM with 10% FBS. Forty-eight hours after transfection, media was removed and replaced with serum-free DMEM. Cells were serum starved for 24 hours and then treated with BSA or 10ng/mL of IL-17a for 24 hours; RNA was then isolated for RT-qPCR.

2.1.6 IL-13 treatment

HBE cells were polarized on transwell inserts for 2 weeks in ALI media, and then BSA or 20 ng/mL IL-13 (PeproTech) in ALI media was added to the basolateral side of the membrane for 48 hours. Cells were lysed in NET buffer for western blot.

2.2 Manipulating nucleic acids

2.2.1 Ethanol precipitation

To precipitate DNA or RNA, the volume of the sample was measured, and then 1/10 the volume of 3 M sodium acetate (pH 5.2) and 2-3 volumes of 100% ethanol were added. The sample was incubated at -20°C for at least 1 hour, then centrifuged at 16,100 x g at 4°C for 30 minutes. The supernatant was removed and the pellet washed with 70% ethanol, centrifuged at 16,100 x g at 4°C for 5-10 minutes, and air-dried. The pellet was resuspended in an appropriate amount of nuclease–free water.

2.2.2 Agarose gel electrophoresis

Agarose gels were made by dissolving 1-2% w/v electrophoresis grade agarose (LT) in 1x

Tris/Acetic Acid/EDTA (TAE) buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA). Ethidium bromide (0.2

µg/mL) was added to all gels- except chromatin gels- to visualize the DNA. 6x loading dye (40% sucrose

in H_2O, 0.25% Orange G [Sigma]) was diluted to 1x in DNA samples, and the samples were loaded on

the gel with a 100 base pair (bp) or 1 kb ladder (NEB). Gels were run at 4.5 volts/cm for 5 minutes, then

at 6-8 volts/cm until desired resolution was achieved. Chromatin gels were post-stained in ethidium

bromide (0.2 µg/mL) for 30 minutes, and then destained in dH_2O for 10 minutes. DNA was visualized

using a UV transilluminator.

2.3. Bacterial cloning of DNA constructs

2.3.1 Polymerase chain reaction (PCR)

Phusion® High-Fidelity DNA Polymerase (NEB) was used to amplify DNA fragments for cloning

into plasmid constructs. The PCR reaction was set up in a total volume of 50 µL, with 5-10 ng of plasmid

template DNA, 1x Phusion® HF Buffer (NEB), 200 µM dNTPs mix (NEB), 0.5 µM each of forward and

reverse primer, 3% DMSO, and 1 unit/PCR reaction of Phusion DNA Polymerase. A list of plasmids used

as template DNA is found in table 2.3.1.i, and a list of primers with annealing temperatures and restriction

sites is found in Table 2.3.2.ii. Cycling conditions are as follows:

> 30 seconds at 98°C
> 35 cycles of:
>> 10 seconds at 98°C
>> 30 seconds at annealing temperature
>> 60 seconds per kb amplified DNA at 72°C
> 10 minutes at 72°C

Reactions were separated on a 1% gel to determine that the correct-sized PCR product was generated.

Table 2.3.1.i Plasmid templates for Phusion® PCR

cDNA gene name	Plasmid	Plasmid source
EHF	pCMV6-XL4-EHF (SC115508)	OriGene
c-Jun	pMIEG3-c-Jun (#40348)	AddGene

Table 2.3.1.ii Primers for Phusion® PCR

Primer Name	Primer Sequence	Restriction sites	Annealing temperature
EHF F	CACCTCGAGATGATTCTGGAAGGAGG	XhoI	-
EHF R	CACGGTACCTGTCCAAAGTATTGGCA	KpnI	-
Flag-EHF F	CGACTCGAGATGGACTACAAAGACGATGACGACAAG ATTCTGGAAGGAGG	XhoI	40°C
Flag-EHF R	CGAGGTACCGTATTGGCAGCTTCAG	KpnI	40°C
LIC EHF F	TACTTCCAATCCAATGCAATTCTGGAAGGAGG	LIC*	40°C
LIC EHF R	TTATCCACTTCCAATGTTATTAGTATTGGCAGCTTCAG	LIC*	40°C
c-Jun LIC F	TACTTCCAATCCAATGCAACTGCAAAGATGGAAAC	LIC*	53°C
c-Jun LIC R	TTATCCACTTCCAATGTTATTAAAATGTTTGCAACTGCT	LIC*	53°C

*Ligation independent cloning (LIC) ends are described in section 2.14.1

2.3.2 Plasmid construction

The Phusion® PCR product and the pcDNA3.1(-) vector were digested using the enzymes corresponding to the restriction sites on the PCR products. Digestions were performed in a total volume of 20 µL with 1x Buffer (NEB), 0.5 µL each enzyme, and 500 ng PCR product or vector at 37°C for 4 hours. The digested PCR product was purified using the QIAquick® PCR Purification Kit (Qiagen) according to the manufacturer's protocol. Five volumes of Buffer PB were added to 1 volume PCR product. Samples were added to the PCR cleanup column and centrifuged at 16,100 x *g*, then washed with 750µL Qiagen Buffer PE. The column was emptied and centrifuged at 16,100 x *g* for 1 minute to dry. The DNA was eluted with 50 µL Qiagen Buffer EB. The vector was de-phosphorylated; 2 µL of 10x Antarctic Phosphatase Reaction Buffer and 1 µL Antarctic Phosphatase (both from NEB) were added to the entire 20µL digestion reaction. The reaction was incubated at 37°C for 1 hour, and then heat

inactivated for 5 minutes at 65°C. The ligation was performed in a 20 µL reaction with 20 ng vector, a 3:1

molar ratio of insert (PCR product) to vector, 1x T4 DNA Ligase Buffer, and 1 µL NEB T4 Ligase (both

from NEB). The ligation reaction was incubated overnight at 16°C.

2.3.3 Bacterial transformation of plasmid constructs

Plasmids were transformed into NEB® 5-alpha (NEB) competent *Escherichia coli* cells. The

entire 10-20 µL ligation sample was added to 25 µL cells in a chilled 14 mL polypropylene round-bottom

tube and incubated on ice for 30 minutes. The cells were heat shocked at 42°C for 30 seconds and then

incubated on ice for 2 minutes. Pre-warmed lysogeny broth (LB) (10 g/L peptone, 5 g/L yeast extract, 10

g/L NaCl) media was added to the cells to bring the volume up to 1 mL, and the transformed cells were

incubated at 60-90 minutes with shaking at 250 rpm at 37°C. The entire reaction was spread on an LB

agar plate with appropriate antibiotic (30 µg/mL kanamycin or 50 µg/mL ampicillin), and the plate was

incubated overnight at 37°C.

2.3.4 Plasmid preparations

Minipreps. One colony of transformed bacteria was selected from an LB agar plate and used to

inoculate 5mL LB media with appropriate antibiotic (30 µg/mL kanamycin or 50 µg/mL ampicillin). The

culture was incubated overnight at 37°C with shaking at 250 rpm. The next day, bacteria were harvested

by centrifugation of 1.5 mL of overnight culture at 16,100 x *g* for 1 minute at room temperature. The DNA

was extracted from bacteria using the Fast Plasmid Miniprep kit (5 PRIME) according to the

manufacturer's protocol. The bacteria pellets were lysed by resuspension in 400 µL of 4°C Complete

Lysis solution. Lysates were incubated at room temperature for 3 minutes, transferred to a spin column in

a 2 mL microcentrifuge tube, and centrifuged at 16,100 x *g* for 1 minute at room temperature. Samples

were washed with 400 µL of Wash buffer and centrifuged at 16,100 x *g* for 1 minute at room temperature.

The flow-through was discarded and the column dried by centrifugation at 16,100 x *g* for 1 minute at room

temperature. The column was transferred to a new 1.5 mL microcentrifuge tube, and the DNA was eluted with 30 µL of elution buffer.

Midipreps and maxipreps. The glycerol stock of the plasmid was streaked onto an LB-agar plate supplemented with the appropriate antibiotic (30 µg/mL kanamycin or 50 µg/mL ampicillin). The plate was incubated overnight at 37°C. The next day, a 5 mL starter culture of LB containing the antibiotic was inoculated with a single colony from the streaked plate. The culture was incubated for 8 hours at 37°C with shaking at 250 rpm and diluted 1:500 into 50 mL (100 mL) of LB supplemented with antibiotic for a midiprep (or maxiprep). The culture was incubated overnight at 37°C with shaking at 250 rpm, and then the bacteria were harvested by centrifugation at 6000 x g for 15 minutes at 4°C. The DNA was isolated from lysed bacteria using the QIAGEN® Plasmid Midi (or Maxi) Kit. The bacteria pellet was resuspended in 4 mL (10 mL) 4°C Buffer P1. Four mL (10 mL) Buffer P2 was added and incubated at room temperature for 5 minutes. Four mL (10 mL) ice cold Buffer P3 was added and incubated on ice for 15 minutes (20 minutes). To collect cellular debris, lysates were centrifuged at 20,000 x g for 30 minutes at 4°C. Meanwhile, a QIAGEN-tip 100 (or 500) was equilibrated by applying 4 mL (10 mL) buffer QBT and allowing the column to empty by gravity flow. The supernatant from the centrifuged lysate was then added to the equilibrated column and allowed to enter the resin by gravity flow. The column was washed 2 times with 10 mL (30 mL) Buffer QC, and the DNA was eluted into a clean tube with 5 mL (15 mL) Buffer QF. DNA was precipitated by adding 3.5 mL (10.5 mL) room temperature isopropyl alcohol followed by centrifugation at 15,000 x g for 30 minutes at 4°C. The supernatant was removed and the DNA pellet was washed with 1.5 mL 70% ethanol and centrifuged at 16,100 x g for 45 minutes at 4°C. The pellet was air-dried and resuspended in nuclease-free H_2O. If the DNA was to be used for mammalian transfections, an additional ethanol precipitation step was performed.

Plasmids were sequenced using traditional methods on an ABI 3730 High-Throughput DNA Sequencer at the Northwestern University Feinberg School of Medicine Center for Genetic Medicine NUSeq Core Facility to confirm that the correct plasmid had been generated.

2.4. Protein analysis

2.4.1 Cell lysis in NET buffer.

Cells were lysed in NET buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA) with 1x protease inhibitor cocktail (Sigma) and 1% v/v Triton X-100. Cells were washed with PBS, and then NET buffer was added and incubated at room temperature for 5 minutes with rocking. Lysates were collected into 1.5 mL microcentrifuge tubes, incubated on ice for 15 minutes, and then centrifuged at 16,100 x g at 4°C for 5 minutes. Supernatants were saved and stored at -20°C or -80°C.

2.4.2 Western blot analysis

Lysates were separated on 10-12% SDS-polyacrylamide gels (37.5:1 acrylamide:bisacrylamide solution, 375 mM Tris-HCl pH 8.8, 0.1% SDS, 0.05% ammonium persulfate, 0.1% tetramethylethylenediamine [TEMED]) with a 3% stacking gel (37.5:1 acrylamide:bisacrylamide solution, 125 mM Tris-HCl pH 6.8, 0.1% SDS, 0.05% ammonium persulfate, 0.1% TEMED). 5x SDS loading dye (1.5 M Tris-HCl pH 6.8, 5.0 mL glycerol, 1.0 g SDS, 1.2 mg bromophenol blue, dH$_2$O to 10 mL) was combined with 2-mercaptoethanol (4:1 ratio), and then diluted 1:4 in protein lysate. Samples were boiled for 5 minutes at 95°C and loaded onto the gel, which was run at 65V until proteins entered the resolving gel and then 90V until desired separation was achieved. The proteins were transferred onto Immobilon®-P PVDF (Millipore) membranes at 300 mA for 1-1.5hr. The membrane was blocked in 5% w/v milk in PBS with 0.1% Tween 20 (PBST) for at least 1 hour at room temperature, and then incubated in primary antibody overnight at 4°C. A list of antibodies and dilutions is found in table 2.4.2. Rat anti-ESE3, clone 5A.5 was initially available commercially from Thermo Scientific. Some experiments were performed using rat hybridoma supernatant supplemented with 20% FBS (anti-ESE3, clone 5A.5), kindly provided by A. Tugores from the Complejo Hospitalario Universitario Insular Materno Infantil. Membranes were washed 3 times in PBST for 15 minutes, and then incubated in secondary antibody diluted 1:5000 in the

same diluent as the primary antibody. The secondary antibodies used were polyclonal goat anti-mouse immunoglobulins/Horseradish peroxidase (HRP) (P0447, Dako), rabbit anti-goat immunoglobulins/HRP (P0449, Dako), Anti-Rabbit IgG (whole molecule)–Peroxidase antibody produced in goat (A0545, Sigma), and goat anti-rat IgG-HRP (sc-2006, SC). Membranes were washed 3 times in PBST, incubated in Pierce® enhanced chemiluminescence western blotting substrate (Thermo Fisher) for 1 minute and exposed to film. Films were scanned using the Perfection V800 Photo Scanner (Epson) with SilverFast8 software (LaserSoft Imaging) and protein was quantified by gel densitometry using Adobe Photoshop.

Table 2.4.2 Antibodies used for western blots.

Antibody	Dilution	Diluent
Mouse anti-β-tubulin, Sigma T4026	1:5000	1% Milk-PBST
Mouse anti-Flag, Sigma F3165	1:3000	5% Milk-PBST
Rabbit anti-GAPDH, Cell Signaling	1:5000	5% Milk-PBST
Rabbit anti-c-Jun, Ab31419	1:1000	1% Milk-PBST
Rabbit anti-JunD, sc-74	1:2000	1% Milk-PBST
Rat anti-ESE3, clone 5A.5, Thermo Scientific	1:300	5% Milk-PBST
Rat anti-ESE3 rat hybridoma supernatant with 20% FBS, A. Tugores	1:300	PBST (primary) 5% Milk-PBST (secondary)

2.5 RNA isolation and cDNA preparation

2.5.1 RNA isolation using TRIzol®

Total RNA was extracted from cells using TRIzol® reagent (Invitrogen) following the manufacturer's protocol. Cells were lysed in 0.5mL TRIzol® for a 24-well plate or 1mL TRIzol® for a 6-well plate, incubated for 5 minutes at room temperature with rocking, homogenized by repeat pipetting and transferred to a 1.5 mL microcentrifuge tube. 0.2 mL of chloroform per 1 mL of initial TRIzol® reagent used was added to the sample, which was shaken for 15 seconds and then incubated at room temperature for 2 minutes. The sample was centrifuged at 12,000 x g at 4°C for 15 minutes to separate

the phases. The top, aqueous phase was collected into a clean 1.5 mL microcentrifuge tube. 0.5 mL

100% isopropyl alcohol per 1 mL of TRIzol® and 10 µg glycogen (Thermo Scientific) (24-well plates only)

were added and incubated for 10 minutes at room temperature to precipitate the RNA. The sample was

centrifuged at 12,000 x *g* at 4°C for 10 minutes, and the pellet was washed in 1 mL 70% ethanol per 1 mL

of TRIzol® and centrifuged at 7,600 x *g* at 4°C for 5 minutes. The pellet was air-dried, resuspended in

nuclease-free H_2O, and incubated at 55°C for 10-15 minutes. The RNA concentration was determined

using a Nanodrop 1000 (Thermo Scientific).

2.5.2 Reverse transcription

cDNA was synthesized from RNA using the TaqMan® Reverse Transcription kit (LT). One µg of

RNA was added to a 50µL reaction with 5.5 mM $MgCl_2$, 500 µM each dNTP, 1x RT buffer, 2.5 µM random

hexamers, 0.4 units/µL RNase inhibitor, and 1 unit/µL MultiScribe reverse transcriptase. There was also a

control reaction with no reverse transcriptase added (RT- reaction). Reverse transcription conditions were

as follows: 10 minutes at 25°C, 1 hour at 48°C, then 5 minutes at 95°C. To remove RNA from the sample,

9.8 µg/mL RNase (DNase-free, Roche) was added and incubated for 30 minutes at 37°C.

2.6 Quantitative PCR (qPCR)

2.6.1 SYBR® Green qPCR

qPCR reactions were assembled in 96-well plates with 1x Power SYBR® Green PCR Master Mix

(LT), 0.25 µM each forward and reverse primer, 3-5 µL of DNA (cDNA generated by reverse transcription

or purified ChIP DNA), and nuclease-free H_2O to 20 µL. Primers are listed in table 2.6.1. The reactions

were run on the Applied Biosystems QuantStudio 6 Flex (LT) machine with the following settings: Fast

96-well (0.1 mL block), Standard Curve, SYBR® Green Reagents, Standard run. The PCR conditions are

as follows:

2 minutes at 50°C
10 minutes at 95°C
40 cycles of:
      15 seconds at 95°C
      60 seconds at 60°C

Table 2.6.1 SYBR® qPCR Primers

A) ChIP-qPCR Primers	Sequences
AGAP1_Intron F	CTGTTTGCTGAGGGTGGACT
AGAP1_Intron R	GAGGGAATGACTGTGGCTCC
CEPBG-intron1 F	CCTCACCTACCACACACGTT
CEPBG-intron1 R	CTACAGGGTCCACGTTCAGC
CXCL1_promoter F	TTTTTGGCGTGGTAGTCACA
CXCL1_promoter R	ACCCCATGGTAACATTGCTC
FOXA1_5'intergenic F	CTGTGTTTAAGACGGGTCTGC
FOXA1_5'intergenic R	CGCACCCTACAGTCCTCACT
HOPX_3'-intergenic F	AGGGAGGCCAATGATGCAAG
HOPX_3'-intergenic R	GTCCCACCCCAGGTCCTAAT
IL8_5'intergenic F	TGAGGAAATCAGGCAAGAGG
IL8_5'intergenic R	CCCCCAAGAAAAGAAGGAAG
ITGA2_Promoter F	GAAGTGTCCTGCCTCCCAAA
ITGA2_Promoter R	AAGTCAGCCAGGTTTCAGGG
JDP2_5'intergenic F	CCAGCTTCCTTGTTGTTGGC
JDP2_5'intergenic R	CCACCCTAGCATGTTCCCAG
OAS3_promoter F	GGCTGGAAGCAAGGAGATGA
OAS3_promoter R	GACCTGACACCCACTTCCTG
PLAU_promoter F	GGTGTCACGCTTCATAACGGT
PLAU_promoter R	CCCTAGCAGCTCTCATGACTC
PLCH1_5'Intergenic F	AGTAACTCACTGACGCTGGC
PLCH1_5'Intergenic R	GCCTCTTCTCTCTGTGGCTG
RARB_5'intergenic F	TGTGCCAGTCTGTTTCCTGT
RARB_5'intergenic R	TGTTCAGGGAGCAAAAGGAC
SPDEF_promoter F	GGCTCCTGATTAACCCTTGC
SPDEF_promoter R	ACGAGTGAATGAGCGAGTGA
TJP2_Intron F	CCTGGTTTCTGCGTTACCCT
TJP2_Intron R	AGCATGTGCATCCTCACCAA

NC F	TCCTTCCAGGTTTTGGCTCC
NC R	GCCCCAGATCAGGAGAGAGA
B) RT-qPCR Primers	**Sequences**
β-2 Microglobulin F	CTCTCTCTTTCTGGCCTGGAG
β-2 Microglobulin R	TCTGCTGGATGACGTGAGTA
CEBPG F	GCAACACCTACGGGTCGAATA
CEBPG R	CTTCACATGTACCTCCGCTG
CXCL1 F	CTTGCCTCAATCCTGCATC
CXCL1 R	CCTTCTGGTCAGTTGGATTTG
CXCL6 F	GGAAGCAAGTTTGTCTGGACC
CXCL6 R	AGAAAACTGCTCCGCTGAAG
DDAH1 F	TGCAACTTTAGATGGCGGAGA
DDAH1 R	GCCAAGATTTCAGCACCTCG
FOXA1 F	TCCAGGATGTTAGGAACTGTG
FOXA1 R	AGGCCTGAGTTCATGTTGCT
FOXA2 F	CAGTATGCTGGGAGCGGTGAAGA
FOXA2 R	AGTAGCCCTCGGGCTCTGCAT
HOPX F	TTTCCGAGGAGGAGACCCA
HOPX R	AAGCTGTCAATGCCTGCCA
IL8 F	AGCTGGCCGTGGCTCTCT
IL8 R	CTGACATCTAAGTTCTTTAGCACTCCTT
ITGA2 F	TTGGAACGGGACTTTCGCAT
ITGA2 R	TCGGCTTTCTCATCAGGTTTCA
JDP2 F	TGGAGGTGAAACTGGGCAAG
JDP2 R	ACTTTGTTCTTCTCCCGGCG
KLF5 F	ACCTCCATCCTATGCTGCTAC
KLF5 R	TTTGTGCAACCAGGGTAATC
RARB F	ACACTGCGAGTCCGTCTTC
RARB R	GCTGGTGCTCTGTGTTTCAA
SPDEF F	TGCTCAAGGACATCGAGACG
SPDEF R	CACTTCTGCACATTGCTGGG
TJP1 F	AGAGACAAGATGTCCGCCAG
TJP1 R	CAAATCCAAATCCAGGAGCCC
TJP2 F	GTGTCCGGAGGCAGAGACAA
TJP2 R	TGCCATTGACCATGACCACTC

2.6.2 Quantitation of *CFTR* expression

Calu-3 cells were reverse transfected with 40 nM of siGenome Human EHF siRNA SMARTpool (M-012421-00) or Non-targeting siRNA Pool #2 (D-001206-14-05, both from Dharmacon) using RNAiMax^{TM} in a 96-well plate. siRNA and RNAiMax^{TM} were combined with 30 µL Opti-MEM®, incubated for 15 minutes at room temperature, and then added to the 96-well plate. 17,500 Calu-3 cells in 120 µL DMEM with 10% FBS were added to the well. Non-targeting controls were transfected in quadruplicate and EHF siRNA in duplicate in three replicate experiments. Seventy-two hours after transfection, the cells were lysed and cDNA was synthesized using the Cells-to-CT™ Kit (Ambion). *CFTR* and β-2-microglobulin (*B2M*) levels were measured using a multiplexed TaqMan® assay (194). Reactions were performed in 96-well plates with 10 µL TaqMan® Fast Advanced master mix (LT), 0.9 µM each CFTR forward and reverse primer, 0.9 µM each B2M forward and reverse primer, 0.25 µM FAM-conjugated CFTR probe, 0.25 µM Joe-conjugated B2M probe, 4µL cDNA, and nuclease-free H$_2$O to 20 µL. Primers are listed in table 2.6.2. The reactions were run on the Applied Biosystems QuantStudio 6 Flex (LT) machine with the following settings: Fast 96-well (0.1mL block), Standard Curve, TaqMan® Reagents, Fast run. The PCR conditions are as follows:

> 20 seconds at 95°C
> 40 cycles of:
> > 1 second at 95°C
> > 20 seconds at 60°C

Table 2.6.2 TaqMan® qPCR Primers and Probes

Primer/Probe Name	Sequence
A) CFTR	
Forward	AGCTGTCAAGCCGTGTTCTAGATA
Reverse	ATGAGGAGTGCCACTTGCAAA
Probe	56-FAM/CACACGAAATGTGCCAATGCAAGTCCCT/36-TAMSp
B) B2M	
Forward	AAGTGGGATCGAGACATGTAAG
Reverse	GCAAGCAAGCAGAATTTGGA
Probe	56-JOEN/TCATGGAGG/ZEN/TTTGAAGATGCCGCA/3IABkFQ

2.7 RNA-sequencing

2.7.1 Library preparation and sequencing

RNA was isolated from 6-well plates using TRIzol® and then ethanol precipitated. Total RNA purity and quality were assessed by Nanodrop and Bioanalyzer (Agilent). For EHF knockdown RNA-seq, 5 samples each of control and EHF siRNA transfected cells were sequenced. For the EHF overexpression experiment, 3-4 samples each of 3 pcDNA and 3 EHF-overexpression clones were sequenced. Libraries were generated from 2 µg of total RNA using the TruSeq^{TM} sample preparation kit (Illumina) as per the manufacturer's Low-Throughput protocol. Poly-A containing mRNAs were purified using poly-T-oligo-attached magnetic beads in 2 rounds of purification. cDNA was synthesized from the mRNA, then the ends were repaired and 3' adenylated. Adapters were ligated and cDNA was amplified using PCR to generate the final library. The libraries were sequenced on Illumina HiSeq2500 machines at the University of North Carolina High-Throughput Sequencing Facility using standard protocols for paired-end 2 x 100 base reads, yielding 2.6–3.5 x10^7 reads per library

2.7.2 RNA-seq data analysis

For the EHF depletion in Calu-3 cells and EHF overexpression in A549 cells RNA-seq experiments,

data in fastq format were mapped to the reference human genome (hg19) with ENSEMBL Release 74 of

human genome annotation as transcriptome mapping using TopHat (v2.0.10) (116) with the following

command:

tophat -r 150 -p 12 --no-coverage-search -o <output-file> --library-type fr -unstranded --transcriptome-index <R1-fastq.qz R2-fastq.qz>

Quantification and differential expression of transcripts from the aligned reads were performed using

Cufflinks (v2.1.1) (129) with a false discovery rate (FDR) ≤ 0.1 as differential expression threshold using

the following command:

./cuffdiff -o <output-file> -L <Sample1,Sample2> -p 12 -b hg19.fa -u -FDR 0.1 <Sample1-1/acceptedhits.bam,…,Sample1-n/acceptedhits.bam> <Sample2-1/acceptedhits.bam,…,Sample2-n/acceptedhits.bam>

Samples were normalized and expressed as fragments per kilobase of transcript per million fragments

mapped (FPKM). Genes with a fold change ≥ 1.5 were considered differentially expressed. Gene

ontology (GO) processes enriched among differentially expressed genes were identified using Database

for Annotation, Visualization and Integrated Discovery (DAVID) (195).

For the EHF depletion in HBE cells, data in fastq format were mapped to the reference human

genome (hg19) and the UCSC hg19 human gene annotations with TopHat (v2.0.9) (116) using the

following command:

tophat -p 12 -G Homo_sapiens/UCSC/hg19/Annotation/Genes/genes.gtf -o <Sample_thout> <R1-fastq.gz,R2-fastq.gz>

Transcripts were assembled and merged using Cufflinks and Cuffmerge (129) with the following

commands:

cufflinks -p 8 -o <Sample_clout> Sample_thout/accepted_hits.bam

cuffmerge –g Homo_sapiens/UCSC/hg19/Annotation/Genes/genes.gtf –s UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa –p 12 assemblies.txt

The assemblies.txt file contained a list of all files corresponding to *Sample_clout/transcripts.gtf.*

Quantification and differential expression of transcripts from the aligned reads were performed using

Cufflinks (v2.1.1) (129) with the following command:

cuffdiff -o diff_out -b /Homo_sapiens/UCSC/hg19/Sequence/WholeGenomeFasta/genome.fa -p 12 -L
<Sample1,Sample2> -u merged_asm/merged.gtf <Sample2-1/acceptedhits.bam,…,
Sample2-n/acceptedhits.bam>

Samples were normalized and expressed as FPKM values. Putative direct targets of EHF were identified

using Binding and Expression Target Analysis (BETA) (127). This software assigns transcription factor

binding peaks to genes using a model that decreases the regulatory potential of the TF binding site as the

distance to the TSS increases, then sums the contribution of multiple sites near a gene (in this analysis,

within 10 kb or a CTCF site, whichever is closer). It also takes into account the significance of gene

expression changes following modulation of the transcription factor (with an FDR <0.1). Enrichment

analysis was performed with MetaCore from Thomson Reuters version 6.26 using the Enrichment

Analysis workflow.

2.8 Chromatin immunoprecipitation (ChIP)

2.8.1 Chromatin preparation.

2.8.1.i One-step lysis.

Cells were seeded on 10 cm plates and grown until 100% confluent. Media was replaced with 10

mL fresh media, and formaldehyde was added to a final concentration of 1% to crosslink cells for 10

minutes at room temperature. The reaction was quenched with 0.125 M glycine for 5 minutes at room

temperature. Cells were washed 2 times with 5 mL ice-cold PBS, then scraped into 5 mL PBS in a 15 mL

conical tube. The sample was centrifuged at 2,000 RPM at 4°C for 4 minutes in a bucket centrifuge, and

the pellet was resuspended in 1 mL fresh lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1,

and 1x protease inhibitor cocktail [PIC, Roche]). The lysate was divided into 2 500 µL aliquots and

incubated on ice for at least 10 minutes. Chromatin was then sheared to the appropriate size using an

Ultrasonic Processor (Cole Parmer). 500 µL of chromatin in a microcentrifuge tube was placed in an ice water bath with the probe 2/3 into the solution and sonicated with an amplitude of 40% and pulses of 25 seconds on and 59 seconds off. The total sonication time was determined for each cell type and is listed in table 2.8.1.i.

Table 2.8.1.i Chromatin sonication conditions for one-step lysis

Cell type	Total sonication time
A549	3 minutes 20 seconds
Calu-3	4 minutes 10 seconds
primary human bronchial epithelial cells	3 minutes 45 seconds
primary human tracheal epithelial cells	4 minutes 10 seconds

2.8.1.ii Two-step lysis

Cells were seeded on 10 cm plates and grown until 100% confluent. Media was replaced with 10 mL fresh media, and formaldehyde was added to a final concentration of 1% to crosslink cells for 10 minutes at room temperature. The reaction was quenched with 0.125 M glycine for 5 minutes at room temperature. Cells were washed 2 times with 5 mL ice cold PBS and then incubated in 5 mL of 0.05% Trypsin-EDTA for 10 minutes at 37°C. The trypsin was quenched with 1 mL FBS warmed to 37°C for 5 minutes at room temperature. Cells from 2 10cm plates were scraped into 1 15mL conical tube, and then the sample was centrifuged at 2,000 RPM at 4°C for 5 minutes in a bucket centrifuge. The supernatant was removed and replaced with 5 mL PBS with 1x PIC. The sample was centrifuged at 2,000 RPM at 4°C for 5 minutes in a bucket centrifuge, and then the pellet was resuspended in 2 mL ice cold Farnham lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40, 1x PIC). Single cell suspension was achieved by passing the lysate through a 25-gauge needle 5 times, and then cells were sheared in a manual homogenizer with 10 strokes. The lysate was incubated on ice for 10 minutes and then centrifuged at

2,000 RPM at 4°C for 5 minutes in a bucket centrifuge. The pellet was resuspended in 1 mL 4°C RIPA

buffer (1x PBS, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 1x PIC), divided into 2 500µL aliquots

and incubated on ice for at least 10 minutes. Chromatin was sonicated with an amplitude of 40%, and

pulses of 25 seconds on and 59 seconds off. The total sonication time was determined for each cell type

and is listed in 2.8.1.ii.

Table 2.8.1.ii Chromatin sonication conditions for two-step lysis

Cell type	Total sonication time
Calu-3	7 minutes 05 seconds
Primary human bronchial epithelial cells	5 minutes 50 seconds

2.8.1.iii Testing chromatin for proper sonication.

A 15 µL aliquot of sonicated chromatin was removed and combined with 15 µL H_2O. 0.5 µg of

RNase (DNase-free, Roche) was added to the sample and incubated for 1 hour at 37°C. Next, 20µg of

Proteinase K (Sigma) was added and incubated for 1 hour at 65°C. The sample was separated on a 2%

agarose gel and then post-stained with ethidium bromide. Chromatin sheared to 100-500 bp, with most

material around 250 bp, was used for chromatin immunoprecipitation (ChIP).

2.8.2 Chromatin immunoprecipitation.

2.8.2.i Agarose bead method.

For FOXA1 ChIP, 1 mL of chromatin prepared using 1-step lysis was diluted in 8.5 mL chromatin

dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl)

and 400µL 25x PIC. Next, 160µL Protein A agarose beads with salmon sperm DNA (Millipore) were

added to the diluted sample, and the solution was mixed on a rotator at 4°C for 1 hour to pre-clear the chromatin. The sample was centrifuged at 1,000 RPM at 4°C for 4 minutes in a bucket centrifuge to pellet the beads. One hundred µL of supernatant was saved as input, and the remaining supernatant was divided into 2 15mL conical tubes. Ten µg of normal rabbit IgG (Millipore) was added to one tube, and 10µg Rabbit-anti-FOXA1 (Ab23738, Abcam) was added to the second. The samples were rotated overnight at 4°C. The next day, 60µL protein A agarose beads were added to each tube and rotated at 4°C for 1 hour. Samples were centrifuged at 1,000 RPM at 4°C for 4 minutes in a bucket centrifuge to collect the beads, and the supernatant was discarded. The beads were re-suspended in 1 mL low salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl, 1x PIC) and transferred to a 1.5mL microcentrifuge tube. The samples were rotated at 4°C for 4 min, then centrifuged at 1,000 RPM for 1 minute in a benchtop centrifuge at 4°C. The resuspension, rotation and centrifugation steps were repeated for washes with high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 500 mM NaCl, 1x PIC), lithium chloride immune complex wash buffer (250 mM LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.1, 1x PIC), and two washes with TE (10 mM Tris-HCl pH 7.5, 1 mM EDTA). Following the final TE wash, the beads were resuspended in 250 µL fresh elution buffer (1% SDS, 0.1M NaHCO$_3$), vortexed, and rotated at room temperature for 15 minutes. Samples were centrifuged at 1,000 RPM for 1 minute in a benchtop centrifuge at room temperature, and the supernatant was transferred to a 1.5mL microcentrifuge tube. The elution step was repeated and the supernatants were combined. The input sample volume was brought up to 500µL with elution buffer. Twenty µL of 5 M NaCl was added to the input and IP samples, which were then incubated overnight at 65°C to reverse the crosslinks. The next day, 0.5 µg of RNase (DNase-free, Roche) was added to the samples and incubated for 1 hour at 37°C. Next, 20 µg of Proteinase K (Sigma), 10µL of 0.5M EDTA, and 20µL of 1M Tris-HCl pH 6.5, were added and incubated for 1 hour at 65°C. The DNA was purified using phenol/chloroform extraction. An equal amount of UltraPure^{TM} phenol:chloroform:isoamyl alcohol (25:24:1 v/v) (Invitrogen) was added to the DNA sample, and the phases were separated by centrifugation at 16,100 x *g* for 5 minutes in MaXtract^{TM} High Density

tubes (Qiagen). The aqueous phase was removed and ethanol precipitated with 20 μg of tRNA from baker's yeast (Sigma) as a carrier. Pellets were re-suspended in 100-200 μL of nuclease-free H_2O.

2.8.2.ii Dynabead method

For ChIP using magnetic beads, 200μL Dynabead Protein A or G slurry was added to 1 mL PBS/BSA (5 mg/mL BSA in PBS with 1x PIC) in a 1.5 mL microcentrifuge tube. The tube was placed on a magnetic rack to collect the beads. The beads were resuspended in 1 mL PBS/BSA, and then collected by placement on the magnet. This step was repeated 2 additional times to wash the beads, which were then resuspended in 1 mL PBS/BSA. An appropriate amount of antibody was added to each tube. For ChIP-seq, only one sample with the antibody against the protein of interest was generated. For ChIP-qPCR there were 2 samples: one with the ChIP antibody and the second with an isotype-matched IgG. Antibodies and amounts used for ChIP, dynabead type, and corresponding IgG are listed in Table 2.8.2. Calu-3 EHF ChIP-seq was performed with rat anti-ESE3 (Thermo Scientific); the first replicate of the HBE EHF ChIP-seq was done with the commercial antibody and the second with purified anti-ESE3 from rat hybridoma (A. Tugores). The samples were rotated at 4°C for at least 2 hours. Following incubation, the beads were washed 3 times in 1mL PBS/BSA and then resuspended in 100 μL PBS/BSA. For ChIP-seq, 1mL of chromatin generated using the two-step lysis method was added to 100μL of resuspended beads. For ChIP-qPCR, 500μL of chromatin generated using the two-step lysis method was added to each of 2 1.5 mL microcentrifuge tubes, and then resuspended IgG-conjugated beads were added to one tube, and ChIP antibody-conjugated beads were added to the second. Twenty-five μL/mL of input was removed from chromatin before combining with beads. Samples were incubated on the rotator overnight at 4°C. The next day, the beads were collected on the magnetic rack, resuspended in LiCl IP wash buffer (100mM Tris pH 7.5, 500mM LiCl, 1% NP-40, 1% sodium deoxycholate) and incubated on the rotator for 3 minutes at 4°C. The sample was placed on the magnetic rack to gather the beads. These steps were repeated 4 times for a total of 5 LiCl washes. After the final wash, the beads were resuspended in 1x TE (pH 7.5) and rotated at 4°C for 1 min. The beads were collected on the magnetic rack and resuspended in

200 µL IP elution buffer (1% SDS, 0.1 M NaHCO3). Beads were incubated in elution buffer for 1 hour in a

65°C heat bath, with vortexing every 15 minutes. Samples were then centrifuged at 16,100 x *g* at room

temperature for 3 minutes, and the supernatants were removed to a clean 1.5 mL microcentrifuge tube.

Input samples were brought to 200 µL with IP elution buffer. Samples were incubated overnight in a 65°C

heat bath to reverse the crosslinks. The next day, DNA was purified with the QIAquick® PCR Purification

Kit (Qiagen) according to the manufacturer's protocol. The DNA was eluted with 30µL Qiagen Buffer EB

warmed to 55°C for ChIP-seq and 2x50µL warmed buffer for ChIP-qPCR.

Table 2.8.2 Antibodies used for ChIP

Antibody	Corresponding IgG	Dynabead Type	Amount for ChIP	Amount for ChIP-seq
Rabbit anti-c-Jun, sc-1694 (SC)	Normal rabbit IgG (Millipore)	Protein A	20µg	20µg
Rat anti-ESE3 (EHF), clone 5A.5	Rat IgG (SC)	Protein G	5µg	20µg
Rabbit anti-JunD, sc-74 (SC)	Normal rabbit IgG (Millipore)	Protein A	10µg	10µg
Rabbit anti-H3K27ac Ab4729 (Abcam)	NA	Protein A	NA	5µg
Rabbit anti-H3K4Me1, Ab8895 (Abcam)	NA	Protein A	NA	5µg
Rabbit anti-H3K4me3, 07-0473 (Millipore)	NA	Protein A	NA	5µg

2.8.2.iii Quantitative PCR

qPCR was performed on ChIP samples using SYBR® Green and primers listed in table 2.6.1. Percent of input was calculated using the following formulas:

Normalized input (nIn)= C_tIn – $\log_2$(input dilution factor)

% of Input= 100 x 2^(nIn-C_tIP)

Where C_tIn is the C_t value obtained for the input for a given primer set, and C_tIP is the C_t value for the IgG or ChIP sample for the same primer set.

2.8.3 ChIP-sequencing (ChIP-seq) library preparation and sequencing

Sequencing libraries were generated from 5-10 ng of immunoprecipitated DNA. DNA ends were blunted in a 50 µL reaction with 1x T4 DNA ligase buffer, 0.4 mM dNTP mix, 1.4 units T4 DNA polymerase, 0.5 units DNA Polymerase I Large (Klenow) Fragment, and 4.5 units T4 Polynucleotide Kinase (all from NEB). The reaction was incubated at 20°C for 30 min. End-repaired DNA was purified using Agencourt® AMPure® XPbeads. The end-repair reaction was combined in a 1:1:1 ratio with AMPure XP beads and clouding buffer (30% PEG, 1.25 M NaCl) and incubated for 15 minutes at room temperature. Beads were collected on a magnetic rack, washed 2 times with 80% ethanol, and air-dried. DNA was eluted in 11.5µL of elution buffer (Qiagen). Adenine bases were added to the end-repaired DNA in the following reaction: 1x NEBuffer 2, 0.67 mM dATP, 2.5 units Klenow Fragment (3' → 5' exo-) (all from NEB), and eluted DNA in a total volume of 15µL. The reaction was incubated for 30 minutes at 37°C. The sample was added to 1x DNA Quick Ligase Buffer (NEB), 2 µL of a 1:160 dilution (in NEBuffer 2) of Library Quantification DNA Standard-Illumina adaptor (1-6) (Kapa Biosystems), and 1.5 µL Quick Ligase (NEB) in a total volume of 37 µL and then incubated for 25 minutes at room temperature to add Illumina adapters. The DNA was purified using AMPure XP beads, eluted into 12 µL elution buffer, and then converted to double stranded DNA in a PCR reaction with 1 µM each TruSeq Primer (F:AATGATACGGCGACCACCGAGATCTACA, R:CAAGCAGAAGACGGCATACGAGAT) and 1x

NEBNext® High-Fidelity PCR master mix (NEB). The DNA was amplified with 5 cycles of PCR:

> 30 seconds at 98°C
> 5 cycles of:
> > 10 seconds at 98°C
> > 30 seconds at 65°C
> > 30 seconds at 72°C
> 5 minutes at 72°C

To size select DNA, the PCR reaction was separated on a 2% Electrophoresis grade agarose:SeaKem® LE agarose (Lonza) gel (3:1 ratio) and post-stained in 1x SYBR® Gold nucleic acid gel stain (ThermoFisher Scientific) in TAE. The gel was cut between 200 and 350bp, and the DNA was purified using the MinElute® Gel Extraction Kit (Qiagen). The cut gel was weighed and dissolved in 5 volumes buffer QG with rotation at room temperature. One gel volume of isopropanol was added, and the sample was loaded onto the spin column and centrifuged at 16,1000 x g for 1 minute. The column was washed with 500 µL buffer QG, then 750µL buffer PE; both washes were centrifuged at 16,1000 x g for 1 minute. The flow-through was discarded and the column centrifuged at 16,1000 x g for 1 minute to dry. The DNA was eluted with 24 µL buffer EB into a clean 1.5 mL microcentrifuge tube, and then amplified with 6 additional cycles of PCR. DNA concentration was quantified by Qubit® using the dsDNA HS assay kit (LT) and library quality was checked by Bioanalyzer. Sequencing was performed on an Illumina HiSeq 2000 or HiSeq 4000. The numbers of reads generated for each experiment are listed in table 2.8.3.

Table 2.8.3 List of ChIP-seq samples and number of reads generated.

Sample ID	Cell type	Immunoprecipitated Protein	Number of reads
AH20	Calu-3	Input	24,147,887
AH21	Calu-3	EHF	14,000,252
AH23	Calu-3	H3K27ac	25,999,778
AH29	Calu-3	H3K4me1	15,953,914
AH52	Calu-3	EHF	12,203,275
AH71	HBE	H3K4me1	18,441,084
AH72	HBE	H3K27ac	19,162,051
AH74	HBE	Input	18,517,840
AH137	HBE	EHF	16,013,245
AH198	Calu-3	c-Jun	25,281,846
AH201	Calu-3	JunD	15,566,239
AH214	Calu-3	H3K4me3	35,103,559
AH221	HBE	H3K4me3	46,471,558
AH249	HBE	EHF	56,569,934
AH281	Calu-3	c-Jun	53,091,280
AH282	Calu-3	JunD	70,380,384

2.8.4 ChIP-seq data analysis

Fastq files were aligned to the hg19 version of the genome using Bowtie (v 1.0.0) (116) with a

seed length of 40, a maximum of 3 mismatches allowed in the seed, and only unique reads mapped with

the following command:

bowtie -q -l 40 -n 3 -m 1 hg19 --best --chunkmbs 200 <input fastq file> <Output file>.

For EHF ChIP-seq in HBE cells, replicates were performed on cells from 2 donors. For EHF, c-Jun and

JunD ChIP-seq, two biological replicates were performed on Calu-3 cells. Peaks were called using the

HOMER-IDR method (IDR<0.05) (196). For histone modification ChIP-seq, one replicate was performed

in Calu-3 and primary HBE cells. Histone tag directories were generated using HOMER (123) with the

following command:

makeTagDirectory <output directory name> <Bowtie output file>.

Downstream analysis was performed using HOMER and Cis-regulatory Element Annotation System

(CEAS) (197). A *de novo* motif analysis was performed on the significant IDR peaks within 200 bp of the

center of the peak with the command:

  findMotifsGenome.pl <peak file> hg19 <output> -size 200

Calu-3 ChIP-seq peaks were annotated using the HOMER command:

  annotatePeaks.pl <peak file> hg19 > <output>.

Histograms for histone modifications mapped from the center of EHF ChIP-seq peaks were generated

using the HOMER command:

  annotatePeaks.pl <peak file> hg19 -size 4000 -hist 10 –d <input tag directory> <H3K4me1 tag
  directory> <H3K4me3 tag directory> <H3K27ac tag directory> > <output>

Heatmaps for ChIP-seq tag counts mapped from the center of EHF ChIP-seq peaks were generated

using the HOMER command:

  annotatePeaks.pl <peak file> hg19 -size 100 -hist 10 -ghist -d <Tag Directory 1>…<Tag Directory n> >
  <output>

Samples were then grouped using hierarchical clustering with Cluster 3.0 software and the histogram was

visualized with Java TreeView (198).

        Annotation of genomic locations of ChIP-seq peaks and determination of enrichment at these

sites was performed using CEAS with the following command:

  ceas -g hg19.refGene -b <peak file>

2.9 DNase-seq

        DNase-seq was performed on two biological replicates of Calu-3 cells and 3 donor cultures of

HBE cells as described previously (87). Briefly, nuclei isolated from Calu-3 cells were digested with

DNase I. Sequencing libraries were generated using the Crawford lab method, which involves ligating a

biotinylated linker to the ends of DNase-digested DNA, enrichment on streptavidin beads, ligation to

adapters, and PCR amplification (84). Sequencing reads were aligned to the genome using Bowtie (116),

tag directories were generated and significant DHS were identified using HOMER (123).

2.10 Wound scratch assay

Calu-3 cells from 3 different passages, wild type HBE cells from 3 donor cultures, and CF mutant HBE cells from 3 donors were used in a wound scratch assay. The CF genotypes were ΔF508/ΔF508 (2 donors) and ΔF508/3849+10kb C>T (1 donor). Cells were seeded into a 35mm dish, forward transfected with NC or EHF siRNA as above, and grown to a confluent layer. Seventy-two hours post-transfection, the confluent lawn was damaged with a comb; 3 linear scratches were made per dish. Images were taken at 3,6, 9, and 12 hours post-scratch for Calu-3 cells, and 0, 3, and 6 hours post-scratch for HBE cells. For each scratch, 3 2 mm boxes were analyzed (9 boxes/plate). The scratch area was measured using ImageJ software. For Calu-3 cells, the rate of wound closure was determined by the following formulas:

Average distance across scratch (Avg Dist) = scratch area/scratch length

Change in Avg Dist (ΔDist) = Avg $Dist_N$ – Avg $Dist_{N-1}$

Rate of wound closure = ΔDist/ΔTime

Where N is the current time point, N-1 is the previous time point, and ΔTime is the time difference between N and N-1.

For HBE cells, the percent of initial injury was determined by the following formula:

Percent of initial injury = 100 x (Wound $size_{TN}$/Wound $size_{T0}$)

Where TN is the current time point and T0 is the 0 hour time point.

2.11 Cell proliferation assay

To measure cell proliferation, the CellTiter 96® AQ$_{ueous}$ Non-Radioactive Cell Proliferation Assay (Promega), which uses the MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3- carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, inner salt) substrate, was performed on 3 biological replicates of Calu-3 cells. Cells were trypsinized and a standard curve was generated using a 5-fold serial dilution starting at 75,000 cells/well with a 1:2 dilution at each step. The known dilutions were seeded in triplicate in a 96-well plate and allowed to adhere to the plate for 1 hour at 37°C. Phenazine methosulfate (PMS) was added to the MTS reagent at a ratio of 50 µL PMS per 1 mL MTS. The media on the cells was replaced with 100 µL fresh media and 20 µL of the MTS-PMS solution. The cells were incubated for 80 minutes at 37°C, and then the absorbance was read at 490 nm using a VersaMax tunable microplate reader (Molecular Devices). To determine the effect of EHF depletion on cell proliferation, 8,000 Calu-3 cells per well were seeded in a 96-well plate. Twenty-four hours after seeding, cells were forward transfected with 30 nM of NC or EHF siRNA as above. At this time, the MTS-PMS solution was added to 3 non-transfected wells to determine the cell density (0 hour time point). The cells were incubated for 80 minutes at 37°C, then the absorbance was read at 490 nm. This was repeated for transfected cells 24, 48 and 72 hours after transfection. There were 3 replicates per treatment at each time point. At 72 hours, one well per treatment was lysed and run on a western blot to confirm EHF knockdown.

2.12 Enzyme-linked immunosorbent assay (ELISA)

Calu-3 cells seeded in a 24-well plate were forward transfected with NC or EHF siRNA as above. Forty-eight hours post-transfection, media was replaced with serum-free DMEM and conditioned for 24 hours. Then, the media was removed and centrifuged at 1,500 x *g* for 10 minutes at 4°C to remove cell debris. The top portion of the supernatant was saved for analysis by enzyme-linked immunosorbent assay (ELISA). Transfected cells used to condition the media were lysed and run on a western blot to confirm EHF depletion. To determine the concentration of neutrophil chemokines secreted into the media following EHF knockdown, a quantikine colorimetric sandwich ELISA for IL-8 (D800C) and CXCL6

(DGC00, both from R&D) was performed according to the manufacturer's protocol. For the IL-8 ELISA, conditioned media (CM) was diluted 1:50 in 1x Diluent RD5P (R&D). For the CXCL6 ELISA, the CM was diluted 1:2 in RD5-24 (R&D). Standard curves were set up using serial dilutions of 1:2 starting with 2000 pg/mL. Each standard, blank, and CM sample was read in duplicate; 3 biological replicates were investigated. The standard, blank or CM sample was added to a microplate coated with the capture antibody and incubated at room temperature for 2 hours. The well was washed, and then a conjugated antibody specific to the cytokine was added. The plate was incubated for 1-2 hours at room temperature, and then the wells were washed. The colorimetric substrate solution was added, incubated for 30 minutes protected from light, and then quenched with stop solution. The plate was read at 450 and 540 nm in a VersaMax tunable microplate reader. The 540nm absorbance was subtracted from the 450 nm reading to correct for optical imperfections in the plate. The duplicated normalized absorbances from each standard and sample were averaged, and the blank was subtracted. A standard curve was generated by Prism software (GraphPad) using a four parameter logistic (4-PL) curve fit. The standard curve was used to determine the concentration of the sample, which was then multiplied by the dilution factor.

2.13 Co-immunoprecipitation

HEK cells grown on a 10 cm plate were transfected with either pcDNA 3.1(-) empty vector or pcDNA3.1(-)-Flag-EHF. Ten µg of DNA plasmid was diluted in 500 µL Opti-MEM®, and 10 µL of Lipofectin® transfection reagent (LT) was diluted in a separate aliquot of 500µL of Opti-MEM®. The mixtures were incubated at room temperature for 30 minutes, then combined and incubated for 15 minutes at room temperature. Media was removed from the HEK cells and replaced with 4 mL Opti-MEM®. One mL of the DNA-Lipofectin® complexes was added to the cells and incubated overnight at 37°C. The next morning, the Opti-MEM® was removed and replaced with 10 mL DMEM with 10% fetal bovine serum and incubated at 37°C. Seventy-two hours after transfection, cells were lysed in 1 mL NET buffer. 20 µL of lysate was removed as input material.

Two immunoprecipitations (IPs) were performed in parallel, one with negative control lysate and the second with Flag-EHF-expressing cell lysate. Eighty μL of Anti-FLAG M2 affinity gel was added to 1mL Tris-buffered saline (TBS) (50 mM Tris-HCl pH 7.5, 150 mM NaCl) in a 1.5 mL microcentrifuge tube. The resin was centrifuged at 8,200 x g for 30s, and then washed 2 additional times in 1 mL TBS. During the final wash, the resuspended resin was evenly distributed to 2 1.5mL microcentrifuge tubes. To one tube, 1mL lysate from cells transfected with pcDNA 3.1(-) empty vector was added; to the second tube, 1 mL lysate from pcDNA3.1(-)-Flag-EHF-transfected cells was added. The IPs were incubated at 4°C for 2 hours on the rotator, and then the resin was collected by centrifugation at 8,200 x g for 30s and washed 3 times in TBS. Protein was eluted from the resin by adding 100 μL of 150 ng/μL 3X FLAG peptide in TBS and incubating on the rotator at 4°C for 30 minutes. The resin was collected by centrifugation at 8,200 x g for 30 seconds; the supernatant was saved and probed by western blot. The transfection and IP were performed 3 times.

2.14 Epitope-tagged protein expression and purification

Purified epitope tagged proteins were produced for bio-layer interferometry (Section 2.15). Bacterial constructs encoding N-terminal His6-EHF and Flag-c-Jun were generated. Expression of the constructs in *E. coli,* lysis of bacteria and purification of the proteins were optimized following the schematic shown in Figure 2.14.

**Generation of bacterial
expression plasmids**

Protein expression in *E. Coli*
Optimize:
Strain
Temperature
Time
IPTG concentration
Growth media

Bacterial lysis
Optimize:
Lysis buffer
Lysis method

**Insoluble protein in
inclusion bodies**

Soluble protein

Protein purification
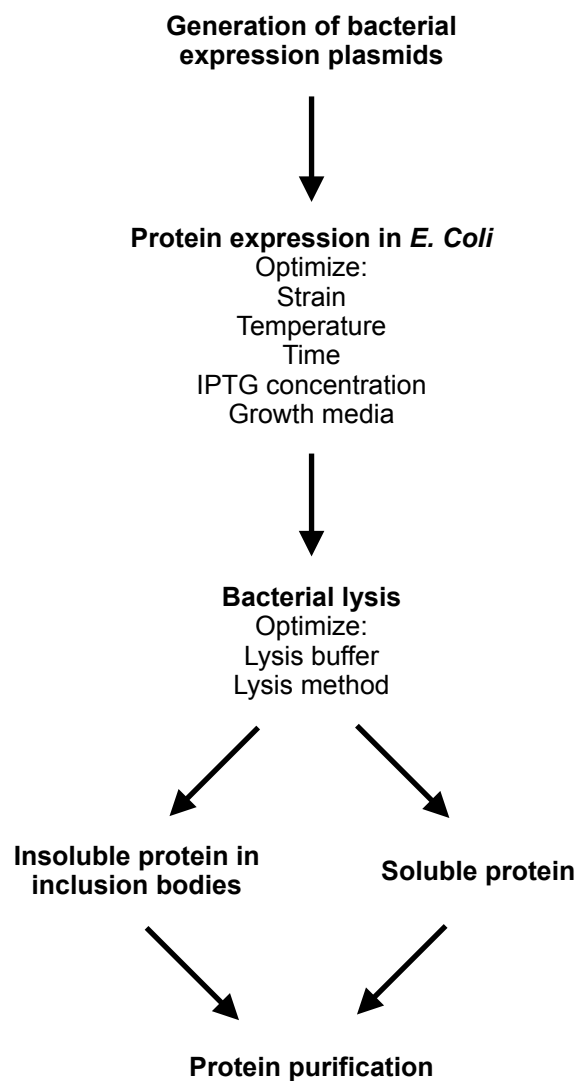
Figure 2.14 Schematic of His6-EHF and Flag-c-Jun protein expression and purification. Bacterial expression plasmids for His6-EHF and Flag-c-Jun were generated and separately cloned into the appropriate strain of E. Coli. To determine optimal conditions for protein expression, multiple temperatures, incubation times, isopropyl β-D-1-thiogalactopyranoside (IPTG) concentrations for induction, and growth media were tested. The bacteria were then lysed, which yielded either soluble protein or inclusion bodies (insoluble aggregates). Lysis was optimized to generate the highest yield of soluble protein or isolate the inclusion bodies. Finally, the epitope-tagged proteins were purified.

2.14.1 Construction of expression plasmids

N-terminal His6-tagged EHF and N-terminal Flag-tagged c-Jun expression plasmids were generated using ligation independent cloning (LIC). pET His6 TEV LIC cloning vector (1B) and pET Flag TEV LIC cloning vector (1L) were gifts from Scott Gradia (Addgene plasmids #29653 and #29662, respectively). pMIEG3-c-Jun was a gift from Alexander Dent (Addgene plasmid #40348) (Table 2.3.1.i). EHF (NM_012153) Human cDNA Clone (SC115508) was obtained from OriGene (Table 2.3.1.i). Midipreps of all constructs were generated. EHF and c-Jun were amplified using plasmid templates and primers listed in Table 2.3.1.ii, which were designed to amplify the open reading frames with specified LIC ends (F: TACTTCCAATCCAATGCA, R: TTATCCACTTCCAATGTTATTA). The plasmid vector (pET His6 TEV LIC cloning vector [1B] or pET Flag TEV LIC cloning vector [1L]) was digested in a 20 μL reaction with 500 ng plasmid DNA, 1x NEBuffer SspI (NEB), and 2.5 units SspI (NEB) for 1 hour at 37°C. The reactions were separated on a 1% agarose gel, and the digested vector was gel purified with the QIAquick® Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. Briefly, the DNA fragment was cut from the agarose gel with a clean scalpel and then weighed. Three volumes of Buffer QG were added to the gel, then incubated at 50°C until the slice had dissolved. The pH was adjusted with 10 μL 3 M sodium acetate (pH 5.2). One gel volume of isopropanol was added to the sample, which was then applied to a spin column by centrifugation at 16,100 x g for 1 minute at room temperature. The DNA was then washed with 750 μL Buffer PE and centrifuged for 1 minute. The flow through was discarded and the column dried by centrifugation for an additional minute. The DNA was eluted into a clean 1.5 mL microcentrifuge tube with 50 μL Buffer EB. The PCR-amplified EHF and c-Jun were ligated into the digested, gel-purified pET His6 TEV LIC cloning vector (1B) and pET Flag TEV LIC cloning vector (1L), respectively. The LIC reaction for the vector was performed in 20 μL with 100 ng of digested vector, 2.5 mM dGTP, 1x NEBuffer 2.1 (NEB), 5 mM dithiothreitol (DTT), and 1.2 units of T4 DNA polymerase (NEB). The LIC reaction for the PCR product was performed in a total volume of 20 μL with 100 ng of PCR product, 2.5 mM dCTP, 1x NEBuffer 2.1 (NEB), 5 mM DTT, and 1.2 units of T4 DNA polymerase (NEB). Both reactions were incubated at 22°C for 30 minutes, then 75°C for 20 min. For the final LIC step, 2 μL

vector LIC product was combined with 2 µL PCR LIC product and 6 µL H_2O and incubated at room temperature for 10 minutes to anneal. The LIC product was transformed into competent *E. Coli*, and the plasmids were prepared as described in section 2.3.

2.14.2 Analytical expression of proteins in *E. Coli*

Plasmids were transformed into competent *E. coli.* Two strains were tested: BL21 (DE3) (NEB) and BL21-CodonPlus (DE3)-RP (#230255, Agilent). Twenty-five µL of the competent cells were added to a pre-chilled 14 mL polypropylene round-bottom tube. XL10-Gold β-mercaptoethanol was diluted 1:10 with H_2O, and then 0.5 µL of the dilution was added to the competent cells and incubated on ice for 10 minutes. Twenty-five ng of plasmid was added to the cells and incubated on ice for 30 minutes. The reactions were heat-shocked for 20 seconds at 42°C, and then incubated on ice for 2 minutes. Pre-heated LB medium was added to the cells, which were then allowed to recover at 37°C for 1.5 hours with shaking at 250 rpm. Using a sterile spreader, 100 µL of the transformation was spread on an LB-agar plate containing 30 µg/mL kanamycin. The plate was incubated overnight at 37°C.

The optimal conditions for induction of protein expression were determined in small cultures. A single transformed colony was used to inoculate 1mL of LB or NZCYM (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 1 g/L Casamino acids and 1 g/L $MgSO_4$) media supplemented with 30 µg/mL kanamycin, and the culture was incubated overnight at 37°C with shaking at 250 rpm. The next day, the starter culture was diluted 1:100 in multiple aliquots of 5 mL of LB or NZCYM media containing 30 µg/mL kanamycin. The cultures were incubated at 37°C with shaking at 250 rpm for about 2 hours, until the OD_{600} was ≥ 0.6. One mL of each culture was removed and centrifuged at 16,100 x *g* at room temperature for 1 minute (0 hour sample). The remainder of the culture was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a concentration ranging from 10 µM to 1 mM to determine the optimal concentration. As a negative control, 1 culture per growth condition was not treated with IPTG (- IPTG sample). The cultures were incubated at 24°C or 37°C for 4 or 24 hours with shaking at 250 rpm to

determine optimal growth conditions. One mL of culture was then transferred to a 1.5 mL microcentrifuge tube and centrifuged at 16,100 x g at room temperature for 1 minute to collect the bacteria.

The samples were run on a silver stain analytical gel to determine protein expression under each condition. Bacterial pellets were resuspended in 5x SDS loading dye with 20% 2-mercaptoethanol, boiled at 95°C for 15 minutes, and then centrifuged at 16,100 x g at room temperature for 10 minutes. Five µL of supernatant was loaded onto a 12% SDS-polyacrylamide gel with a 3% stacker. Gels were run at 65 V until proteins entered the resolving gel, then at 90 V until adequate resolution was achieved. The gel was washed in 50% v/v MeOH, 10% v/v acetic acid for 20 minutes, followed by a 10-minute wash in 20% v/v ethanol. It was then rinsed with dH_2O for 1 hour, during which the dH_2O was replaced every 10 minutes. Next, the gel was sensitized with 0.03% w/v sodium thiosulfate for 1 minute, and then washed twice with dH_2O. The gel was incubated in 0.1% w/v silver nitrate for 20 minutes in the cold room, and then rinsed 3 times with dH_2O. Finally, it was developed in cold developer solution (0.04% v/v formaldehyde, 2% w/v sodium carbonate) for 10-20 minutes, until the protein was visible. The reaction was quenched with 5% v/v acetic acid, and the gel was scanned using the Perfection V800 Photo Scanner with SilverFast8 software.

2.14.3 Large-scale expression in *E. Coli*

2.14.3.i His6-EHF

To optimize lysis of His6-EHF-expressing bacteria, a glycerol stock of BL21-CodonPlus (DE3)-RP cells transformed with pET-His6-TEV-EHF was used to inoculate 10 mL of LB media containing 30 µg/mL kanamycin and 50 µg/mL chloramphenicol. The starter culture was incubated overnight at 37°C with shaking at 250 rpm and then diluted 1:100 in LB media supplemented with 30 µg/mL kanamycin and 50 µg/mL chloramphenicol. The culture was incubated for about 2 hours at 37°C with shaking at 250 rpm, until the OD_{600} was ≥ 0.6, induced with 100 µM of IPTG and then incubated for 24 hours at 37°C.

For expression of His6-EHF at the Northwestern University Recombinant Protein Production Core, a glycerol stock of BL21-CodonPlus (DE3)-RP cells transformed with pET-His6-TEV-EHF was used to inoculate an LB starter culture supplemented with 34 µg/mL chloramphenicol and 35 µg/ml kanamycin. The starter culture was incubated overnight at 37°C with shaking at 200 rpm. The next day, the entire starter culture was diluted into 1 L Terrific Broth media (Fisher Scientific) supplemented with 35 µg/mL kanamycin and incubated in a BIOSTAT® Bplus Bioreactor (Sartorius Stedim) with the following fermentation conditions: 300 rpm stirring, 7.2 pH, 3 L/minute gas flow, 20% O_2, 37°C. The cell culture was induced by 0.6 mM IPTG when the OD_{600} reached 0.8 and incubated overnight at 20°C.

2.14.3.ii Flag-c-Jun

A glycerol stock of BL21-CodonPlus (DE3)-RP cells transformed with pET-Flag-TEV-c-Jun was used to inoculate 10 mL of LB media containing 30 µg/mL kanamycin and 50 µg/mL chloramphenicol. The starter culture was incubated overnight at 37°C with shaking at 250 rpm and then diluted 1:100 in LB media supplemented with 30 µg/mL kanamycin and 50 µg/mL chloramphenicol. The culture was incubated for about 2 hours at 37°C with shaking at 250 rpm, until the OD_{600} was ≥ 0.6, then induced with 10 µM IPTG and incubated for 24 hours at 24°C.

2.14.4 Lysis

2.14.4.i Lysis of E. coli expressing His6-EHF

To optimize lysis of *E. coli* expressing His6-EHF, multiple methods were investigated. Two base buffers, 50mM Tris-HCl and 50mM sodium phosphate, at 2 pHs, 7.5 and 8, were tested. The buffers were supplemented with additional agents, including 5% glycerol, 1% Triton-X, 25% sucrose, 1mM DTT, and 0.5% NP-40. Three mechanisms to achieve lysis were tested: enzymatic (lysozyme) only, enzymatic with rapid freeze-thaw cycles, and enzymatic with sonication using the Ultrasonic Processor (Cole Parmer).

With all methods and buffers tested, the His6-EHF protein aggregated in inclusion bodies (IBs). Therefore, the IBs were extracted from cells as follows: bacteria were collected by centrifugation at 8,000 x g for 15 minutes at 4°C, and the pellet was resuspended in a 1:10 ratio (w/v) of pellet to lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 5mM imidazole, 8M urea, lysozyme and DNase I) by stirring for 1 hour at 4°C. The lysate was clarified by sonication using the Ultrasonic Processor (Cole Parmer), with the amplitude set to 65% in pulses of 9 seconds on and 6 seconds off for a total of 4 minutes. The IBs were collected by centrifugation at 12,000 x g for 20 minutes at 4°C, and then resuspended in 1:10 ratio (w/v) of 1x PBS with 8 M urea, pH 7.5.

For lysis at the Northwestern University Recombinant Protein Production Core, the bacteria were harvested by centrifugation at 8,000 x g for 15 minutes at 4°C. The cells were resuspended in lysis buffer (1.5 mM magnesium acetate, 1 mM $CaCl_2$, 250 mM NaCl, 100 mM ammonium sulfate, 40 mM Na_2HPO_4, 3.25 mM citric acid, 5% Glycerol, 5 mM imidazole, 5 mM β-mercaptoethanol, 0.08% n-dodecyl-D-maltoside, 1 mM PMSF, 20 μM leupeptin), homogenized in an EmulsiFlex-C5 High Pressure Homogenizer (Avestin) 2 times, and clarified in the centrifuge at 18,000 x g at 4°C for 40 min.

2.14.4.ii Lysis of E. coli expressing Flag-c-Jun

Bacteria were harvested by centrifugation at 8,000 x g for 15 minutes at 4°C. Pellets were resuspended in 15 mL of lysis buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 1mM DTT, 1% Triton-X, 10 mM Imidazole, 5% glycerol, 1 mg/mL lysozyme, 100 mM PMSF, DNase I) per gram of pellet and stirred at 4°C for about 2 hours, until the pellet was resuspended. The sample was then sonicated using the Ultrasonic Processor (Cole Parmer), with the amplitude set to 65% in pulses of 9 seconds on and 6 seconds off for a total of 15 minutes. The sonicated lysate was centrifuged at 17,000 x g at 4°C for 30 minutes to collect cell debris, and the supernatant was saved.

2.14.5 Protein purification

2.14.5.i His6-EHF purification

Refolding and purification of His6-EHF from IBs was performed using the BioLogic DuoFlow™ system and Ni-charged Profinity™ IMAC Resin (both from Bio-Rad). For column equilibration, lysate loading, and wash steps, a flow rate of 1 mL/min and was used. The column was equilibrated with 5 column volumes of binding buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 8 M Urea, pH 8.0). The lysate was loaded onto the resin with 3 column volumes of binding buffer. The column was washed with 10 volumes of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 8 M Urea, pH 8.0), then 10 volumes of urea binding buffer (20 mM NaH_2PO_4, 20 mM imidazole, 300 mM NaCl, 1 mM β-mercaptoethanol, 8 M urea, pH 8.0). For refolding and elution, a flow rate of 0.5 mL/min was used. The His6-EHF was refolded by applying a linear gradient from 100% urea binding buffer to 100% refolding buffer (20 mM NaH_2PO_4, 20 mM imidazole, 300 mM NaCl, 1 mM β-mercaptoethanol, pH 8.0) for 60 minutes. The protein was eluted by employing a linear gradient from 100% refolding buffer to 100% phosphate elution buffer (20 mM NaH_2PO_4, 500 mM imidazole, 300 mM NaCl, 1 mM β-mercaptoethanol, pH 8.0) for 60 minutes.

Purification of His6-EHF at the Northwestern University Recombinant Protein Production Core was performed in three steps on the AKTAxpress (GE Healthcare Life Science) FPLC purification system. First, affinity purification was performed on a HisTrap FF 5 mL column (GE Healthcare Life Science) with a sample loading and elution flow rate of 3 mL/minute. The binding buffer was 10 mM Tris-HCl, 500 mM NaCl, 5 mM β-mercaptoethanol, pH 8.3. Before the elution, a wash step with 25 mM imidazole was applied to remove unspecific binding. The protein was eluted in 10 mM Tris-HCl, 500 mM NaCl, 500 mM Imidazole, pH 8.3. The second step was size exclusion purification, performed on a HiLoad 26/600 Superdex 200 prep grade column (GE Healthcare Life Science) at a flow rate of 2 mL/minute. The protein was eluted in 10 mM Tris-HCl, 500 mM NaCl, 5 mM β-mercaptoethanol, pH 8.3. The third step was ion

exchange purification on a HiTrap Q HP column (GE Healthcare Life Science) with a sample loading and elution flow rate of 4 mL/minute. The binding buffer was 20 mM Tris-HCl, 50 mM NaCl, pH 8.3. The elution buffer was 20 mM Tris-HCl, 800 mM NaCl, pH 8.3. The fractions were separated by shallow gradient elution in 40 column volumes. The purified fractions were collected and concentrated in Vivaspin Protein Concentrator Spin Columns with a molecular weight cutoff of 10,000 Daltons (GE Healthcare Life Science). The protein was stored in His-purification Tris buffer (10 mM Tris-HCl pH 8.3, 500 mM NaCl).

2.14.5.ii Flag-c-Jun purification

Flag-c-Jun was purified using Anti-FLAG M2 affinity gel (Sigma). Resin was resuspended, and then 120 µL was added to a clean 15 mL conical tube. Four mL of 1x TBS was added to the resin. The mixture was centrifuged at 4,200 x g for 1 minute at 4°C, and the supernatant was removed. The resin was washed 2 additional times with 4 mL TBS. After the final wash, 160 µL of PIC and 4 mL of lysate containing Flag-c-Jun protein were added to the resin. The sample was rotated overnight at 4°C. The next day, the supernatant was removed and the resin was washed 3 times with 4 mL TBS. During the third wash, the beads were transferred to a 1.5 mL microcentrifuge tube. To elute the protein, 150 µL of 150 ng/µL 3x FLAG peptide (Sigma) in TBS was added to the resin and incubated on the rotator overnight at 4°C. The next day, the resin was pelleted by centrifugation at 5,000 x g for 1 min. The supernatant containing the purified protein was aliquoted and stored at -20°C.

Flag-c-Jun in TBS was dialyzed to replace the buffer with NETN buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% NP-40). BioDesignDialysis Tubing™ with a molecular weight cutoff of 8,000 Daltons (BioDesign) was hydrated in dH$_2$O, and then filled with 100 µL of Flag-c-Jun in TBS. The tubing was submerged in 20 mL of dialysis buffer (19.85 mM Tris pH 8.0, 99.75 mM NaCl, 1 mM EDTA, 0.5% NP-40). The sample was dialyzed at 4°C for 4 hours; the buffer was changed 3 times.

2.15 Bio-layer interferometry

Purified His6-EHF and Flag-c-Jun were investigated for direct interaction using the BLItz[TM] Bio-Layer Interferometer (199). This technology utilizes a nickel charged Tris-NTA (Ni-NTA) biosensor to capture His6-tagged proteins. The sensor is then submerged in a solution of purified interacting protein. White light is sent down the biosensor, and any reflected light is collected. Reflected wavelengths of light are affected by the thickness of the proteins bound to the biosensor. Any shift in wavelengths captured by a spectrometer corresponds to changes in thickness of this biological layer.

To perform the assay, Dip and Read™ Ni-NTA (NTA) Biosensors (FortéBio) were hydrated in 200 µL of His-purification Tris buffer overnight at room temperature in a 96-well plate. The next day, the biosensor was mounted onto the BLItz[TM] system and hydrated in 250 µL of His-purification Tris buffer for 5 minutes. An advanced kinetics experiment was run using the BLItz Pro[TM] software for data collection and analysis; the settings used are found in table 2.15. For all steps in which the sensor was incubated in the tube position, 250µL of buffer was added to a black 0.5 mL microcentrifuge tube and placed in the tube holder. For steps in the drop position, 4 µL of sample was added to the drop holder, which was then slid into place. The initial baseline step was always performed in His-purification Tris buffer, the storage buffer for His6-EHF. For the loading step, the sensor was incubated in either His6-EHF protein in His-purification Tris buffer (119 µg/mL) or His-purification Tris buffer only. For the second baseline and dissociation steps, the sensor was incubated in the Flag-c-Jun storage buffer (either TBS or NETN buffer). During the association step, the sensor was submerged in Flag-c-Jun protein stored in either TBS (15 µM) or NETN buffer (10µM).

Table 2.15 BLItz^{TM} Advanced Kinetics Settings

Step number	Step Type	Duration (seconds)	Position
1	Initial baseline	30	Tube
2	Loading	200	Drop
3	Baseline	30	Tube
4	Association	120	Drop
5	Dissociation	120	Tube

3. RESULTS

Most of the data presented here are part of 2 publications from our laboratory:

Fossum, S.L., Mutolo, M.J., Tugores, A., Gosh, S., Randell, S., Jones, L. C., Leir, S.H., Harris, A. Ets homologous factor has critical roles in epithelial dysfunction in airway disease. *In submission.*

Fossum, S.L., Mutolo, M.J., Yang, R., Dang, H., O'Neal, W.K., Knowles, M.R., Leir, S.H., Harris, A. 2014. Ets homologous factor regulates pathways controlling response to injury in airway epithelial cells. *Nucleic Acids Res*. 42, 13588-91358.

Michael Mutolo performed the initial EHF ChIP-seq replicates (shown in Figure 3.1.3 and 3.3.1) and the EHF depletion experiments where CFTR mRNA and protein expression were measured (Figure 3.9). He also generated the A549 EHF-overexpressing clones (Figure 3.5.2) and produced FOXA1 ChIP-seq data in HBE cells that were used to inform the ChIP-qPCR experiments shown in Figure 3.7.2. Dr. Shih-Hsing Leir grew the primary human bronchial epithelial cells and made some of the HBE chromatin. Rui Yang established and optimized protocols required for ChIP-seq and the HOMER pipeline used to analyze the data in the Harris lab. He made the initial ChIP-seq libraries, and generated the Calu-3 and HBE DNase-seq data shown in Figure 3.1.3 and Figure 3.6.2, respectively. Dr. Sujana Ghosh produced the H3K4me3 ChIP-seq results presented in Figures 3.1.3, 3.3.2, and 3.6.2. Dr. Sergii Pshenychnyi from the Northwestern Recombinant Protein Production Core purified His6-EHF protein from bacterial culture. Dr. Hong Dang from the University of North Carolina (UNC) did the CuffDiff analysis for the Calu3 EHF depletion and A549 overexpression experiments. Lisa Jones from UNC made the RNA-seq libraries. Dr. Antonio Tugores from Complejo Hospitalario Universitario Insular Materno Infantil provided the EHF antibody necessary to complete these studies.

3.1 EHF ChIP-seq in Calu-3 lung adenocarcinoma cells characterizes a genome-wide binding profile for the protein

The role of EHF in regulating gene expression in the bronchial epithelium has not been characterized previously. As a transcription factor (TF), EHF alters gene expression by binding to the genome at regulatory regions. We therefore sought to identify its sites of occupancy in a non-biased, genome-wide manner using ChIP-seq. Initially, Calu-3 lung adenocarcinoma cells were used as a model system for the airway epithelium (190). Calu-3 cells are highly differentiated and have attributes of normal human airway epithelial function. They form polarized monolayers in culture on transwell inserts with tight junctions that lead to high transepithelial resistance (TER), and they respond to cAMP agonists to drive chloride secretion across the epithelial layer (190). Calu-3 cells grown on plastic were utilized with the intent of learning about EHF function in these cells and determining whether the necessary reagents were suitable for use in the more valuable primary human bronchial epithelial (HBE) cells.

The antibody used in the ChIP experiments was first tested for specificity. EHF was depleted in Calu-3 cells using a specific siRNA, and the lysates were probed by western blot (Figure 3.1.1). The antibody detected protein at the predicted size for EHF that was not present in the siRNA-treated cell lysates. Thus, the antibody was deemed adequately specific and used for ChIP.
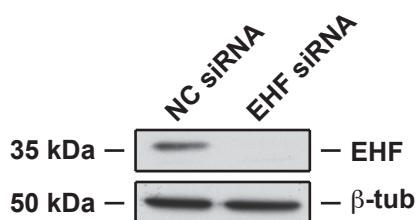
Figure 3.1.1 Validation of EHF antibody for ChIP-seq. An antibody against EHF (Clone 5A.5, Thermo Scientific) was tested for specificity. Calu-3 cells were treated with either negative control (NC) or EHF targeted siRNA. Cell lysates were run on a western blot and probed for EHF with β-tubulin as a loading control. Protein was detected at the predicted size of EHF, 35 kDa, and it decreased in intensity in the knockdown sample.

Two biological replicates of EHF ChIP-seq were performed in Calu-3 cells. The HOMER

irreproducible discovery rate (IDR) pipeline (196) called reproducible peaks from the two samples, with

ChIP input DNA used as a background control. A total of 1,643 peaks were identified using an IDR <

0.05. There was a significant correlation of tag counts at the significant peaks between the 2 samples

(Figure 3.1.2), which confirms that these sites showed similar enrichment of EHF binding in the replicates.

These high-confidence EHF bindings sites were used for further analysis.
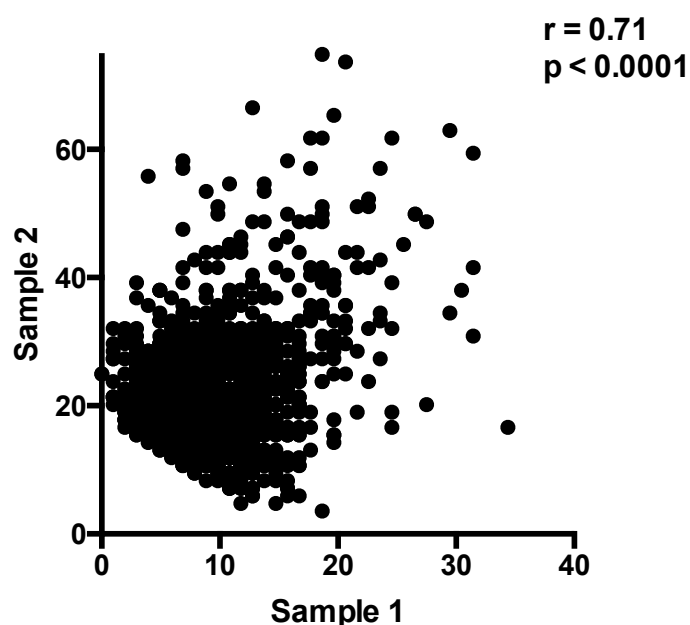
Figure 3.1.2 Correlation of replicates of EHF ChIP-seq in Calu-3 cells. Two biological replicates of
EHF ChIP-seq were generated in Calu-3 cells. The HOMER irreproducible discovery rate (IDR) pipeline
identified reproducible peaks from both data sets. The normalized tag counts at every significant peak
were counted for each biological replicate and plotted. The Pearson correlation coefficient (r) between the
pairs was computed (r = 0.71, p < 0.0001).

Peaks were annotated based on their genomic location using the Cis-regulatory Element Annotation System (CEAS) software (197) (Figure 3.1.3 A). Although the majority of EHF peaks were found at either intergenic or intronic sites, as predicted for cell-type specific regulatory elements (87,200), EHF binding was enriched at promoters (defined as −1 kb relative to the transcription start site) and 5' untranslated regions (UTRs) (Figure 3.1.3 B). The EHF ChIP-seq output was intersected with open chromatin data for the Calu-3 cells, which were generated by DNase I digestion followed by deep sequencing (DNase-seq). EHF occupancy sites showed increased DNase hypersensitivity, suggesting that this factor binds to open chromatin (Figure 3.1.3 C). As a background control, tag counts for Calu-3 DNase-seq were plotted within 2 kb of genomic regions with the epithelial-specific Ets motif recognized by EHF, but with no EHF binding site (Figure 3.1.3 D). These regions did not show the same increase in DNase hypersensitivity, suggesting this characteristic is specific to regions where EHF binds.

Regulatory sites are marked by specific histone modifications. Histone 3, lysine 4 trimethylation (H3K4me3) is enriched at active and poised promoters, whereas enhancers are marked by histone 3, lysine 4 monomethylation (H3K4me1) (90). Active enhancers and promoters are also characterized by histone 3, lysine 27 acetylation (H3K27ac) (83,95). To determine whether the EHF ChIP-seq peaks were enriched for these histone modifications, as predicted based on their genomic locations, one biological replicate of ChIP-seq for H3K4me1, H3K4me3 and H3K27ac in Calu-3 cells was performed. The generated normalized tag counts were plotted from the center of EHF peaks. The results showed a bimodal distribution of enrichment for the H3K4me3 and H3K27ac modifications surrounding the center of EHF ChIP-seq peaks (Figure 3.1.3 E), which was not seen in the input control. This corresponds to EHF occupancy at the promoter. The H3K4me1 modification was not enriched. As a control, tag counts for histone modifications were plotted within 2 kb of genomic regions that lacked EHF occupancy but contained the epithelial-specific Ets motif (Figure 3.1.3 F). There was no enrichment of any histone modification, suggesting EHF specifically binds to regions marked with H3K4me3 and H3K27ac.
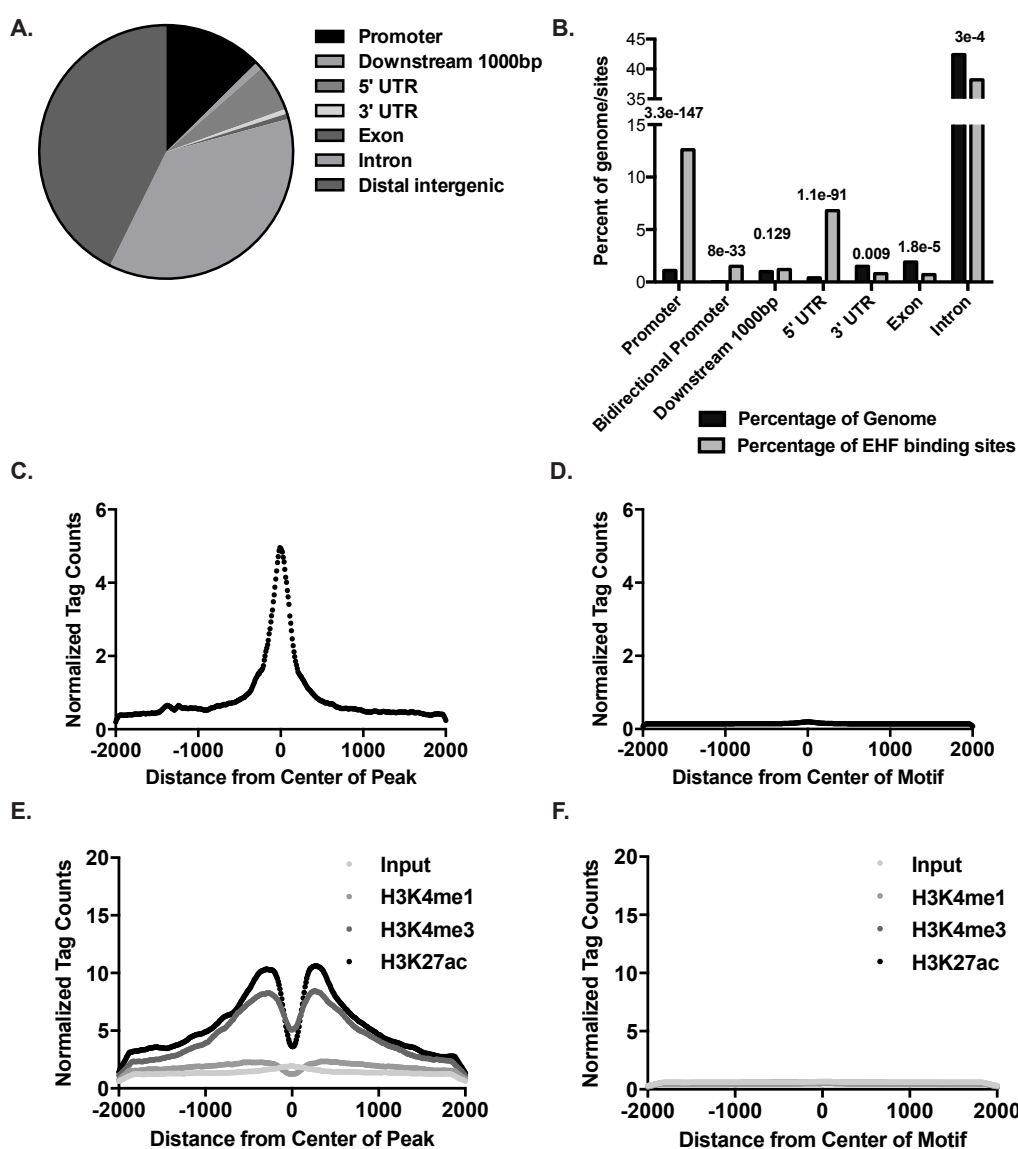
Figure 3.1.3 Characteristics of EHF binding sites in Calu-3 cells. A) Percentage of EHF binding sites found at promoters (1,000 base pairs [bp] upstream of the transcription start site [TSS]), downstream of genes (1000 bp), 5' untranslated regions (UTRs), 3' UTRs, exons, introns, and intergenic sites. B) Percentage of EHF ChIP-seq peaks found at specific sites as compared with genomic background. Bidirectional promoters are within 2500 bp of the TSS. P values, derived using CEAS (one-sided binomial test), are written above bars. C) Normalized tag counts from Calu-3 DNase-seq as measured from the center of EHF peaks. D) As C, but measured from the center of epithelial-specific Ets motif-containing genomic regions with no EHF binding peaks. E) Normalized tag counts from H3K4me1, H3K4me3, and H3K27ac ChIP-seq in Calu-3 cells, as measured from the center of all EHF binding sites. Normalized tag counts from input sample are included as a control. F) As E but measured from the center of epithelial-specific Ets motif-containing genomic regions with no EHF binding peaks.

To validate the ChIP-seq findings, ChIP followed by quantitative PCR (ChIP-qPCR) was performed on two Calu-3 chromatin samples (Figure 3.1.4). The sites tested included 2 promoter (for integrin, α2 [*ITGA2*] and 2'-5'-oligoadenylate synthetase 3 [*OAS3*]), 2 intronic (in tight junction protein 2 [*TJP2*] and ArfGAP with GTPase domain, ankyrin repeat and PH domain 1 [*AGAP1*]), and 2 intergenic sites (near Jun dimerization protein 2 [*JDP2*] and phospholipase C, eta 1 [*PLCH1*]). These locations were chosen due to their close proximity to genes with a significant change in expression following EHF depletion (*ITGA2, OAS3, JDP2, TJP2*) and because they all contain an Ets motif. For comparison, one negative control (NC) region near EHF that lacked a TF ChIP-seq peak, a predicted EHF binding motif and histone modification enrichment was selected. Peaks identified in ChIP-seq were enriched over an IgG control in both replicates, at all six sites tested, while no difference was seen in the negative control (Figure 3.1.4).
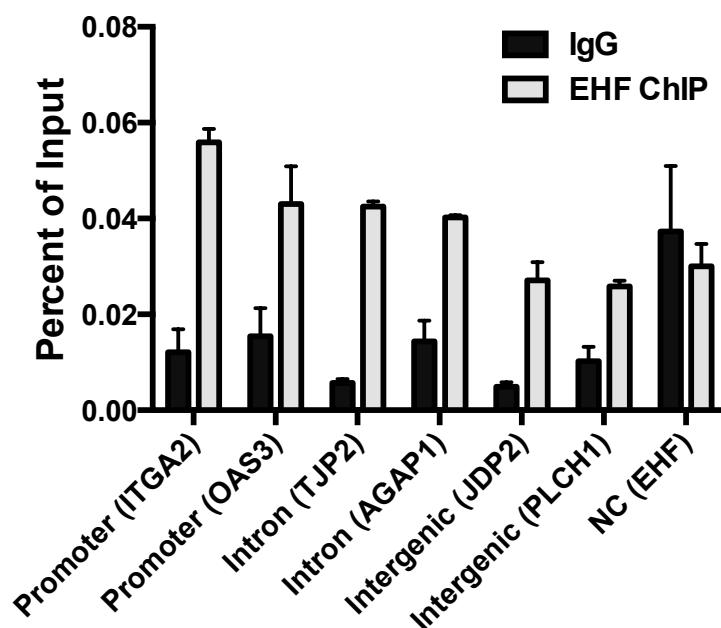
Figure 3.1.4 Validation of EHF ChIP-seq peaks in Calu-3 cells. EHF ChIP followed by quantitative PCR (ChIP-qPCR) was performed in Calu-3 cells. EHF binding was validated at 2 promoter sites, near integrin, α2 (*ITGA2*) and 2'-5'-oligoadenylate synthetase 3 (*OAS3*); 2 intronic locations, in tight junction protein 2 (*TJP2*) and ArfGAP with GTPase domain, ankyrin repeat and PH domain 1 (*AGAP1*); and 2 intergenic sites, near Jun dimerization protein 2 (*JDP2*) and phospholipase C, eta 1 (*PLCH1*). A negative control (NC) site lacked EHF occupancy. Peaks are labeled by their genomic location. Average of both replicates with standard deviation (SD) (n = 2).

3.2 Annotation of EHF binding sites in Calu-3 cells

The nearest gene annotation method was used to predict which genes were likely regulated by occupancy of EHF at *cis*-regulatory elements. This method has inherent limitations, since many regulatory elements lie very far from the genes they control and can be located within introns of genes they do not regulate. A gene ontology (GO) process enrichment analysis was performed on these loci using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (195) to identify processes that were enriched among them (Table 3.2). Genes located nearest to EHF binding sites were involved in cell signaling, cell adhesion, and cell motion. The cell motion process included multiple integrins (*ITGA2, ITGA6*), *CD44,* erb-b2 receptor tyrosine kinase 2 (*ERBB2*), and mitogen-activated kinase 8 (*MAPK8*).

Table 3.2 EHF binds near genes involved in cell signaling, adhesion, and motion in Calu-3 cells. A gene ontology (GO) analysis by DAVID was performed on the nearest genes to each EHF peak.

GO Term	P Value
enzyme linked receptor protein signaling pathway	9.23E-07
regulation of cell adhesion	4.62E-06
cell projection	4.89E-06
cell motion	9.01E-06
basolateral plasma membrane	9.66E-06
phosphate metabolic process	1.10E-05
regulation of cell motion	1.15E-05
cell leading edge	2.04E-05

3.3 EHF binding sites are over-represented at active promoters in HBE cells

Because TF binding to regulatory elements is often cell-type specific (85), it is critical to study this process in cells that best model the *in vivo* cell type. HBE cells were isolated from lung tissue specimens using protease digestion, then grown on collagen-coated plastic (193). Growth of undifferentiated cells on plastic enables collection of sufficient material for ChIP-seq, although they do not recapitulate the *in vivo* tissue as closely as differentiated primary cultures grown on transwell inserts. We performed 2 replicates of EHF ChIP-seq in HBE cells derived from different donors. The HOMER IDR pipeline (196) was used to call a total of 11,326 peaks with an IDR < 0.05. Sonicated input DNA was used as background. The normalized tag counts at each called peak were significantly correlated between the two biological replicates (r = 0.29, p < 0.0001).

CEAS software (197) was used to annotate peaks based on genomic location (Figure 3.3.1 A). As seen in the Calu-3 cells, EHF sites were over-represented at promoters and 5' UTRs of genes (Figure 3.3.1 B), although to a greater extent. EHF also occupied intergenic and intronic regions, which may represent *cis*-regulatory elements. To determine whether EHF peaks are marked by histone modifications corresponding to their genomic location, ChIP-seq was performed for H3K4me1, H3K4me3 and H3K27ac in HBE cells. EHF sites were enriched for all modifications with a bimodal distribution around the center of the peak, a pattern suggestive of a nucleosome-depleted center region where TFs bind (Figure 3.3.2 A). Consistent with binding of EHF at promoters, H3K4me3 was the most enriched mark. The distribution of histone modifications differed based on genomic annotation (Figure 3.3.2 B-D). The promoter sites were enriched for H3K4me3 but depleted for H3K4me1 sites (Figure 3.3.2 B), with the opposite pattern seen in intergenic peaks (Figure 3.3.2 C). The intronic binding regions had higher normalized tag counts for H3K4me3 than the intergenic sites (Figure 3.3.2 D). There was no enrichment of histone modification tag counts at genomic regions with the epithelial-specific Ets motif but no EHF occupancy, suggesting that EHF specifically binds to sites containing these modifications (Figure 3.3.2 E). As an additional control, the histone modifications were plotted from the center of HBE DHS where EHF does not bind (Figure 3.3.2 F). The pattern of H3K4me3 and H3K27ac enrichment at these sites was similar to that

seen in the EHF binding sites, which suggests that the modifications are likely present due to EHF

binding to open chromatin.
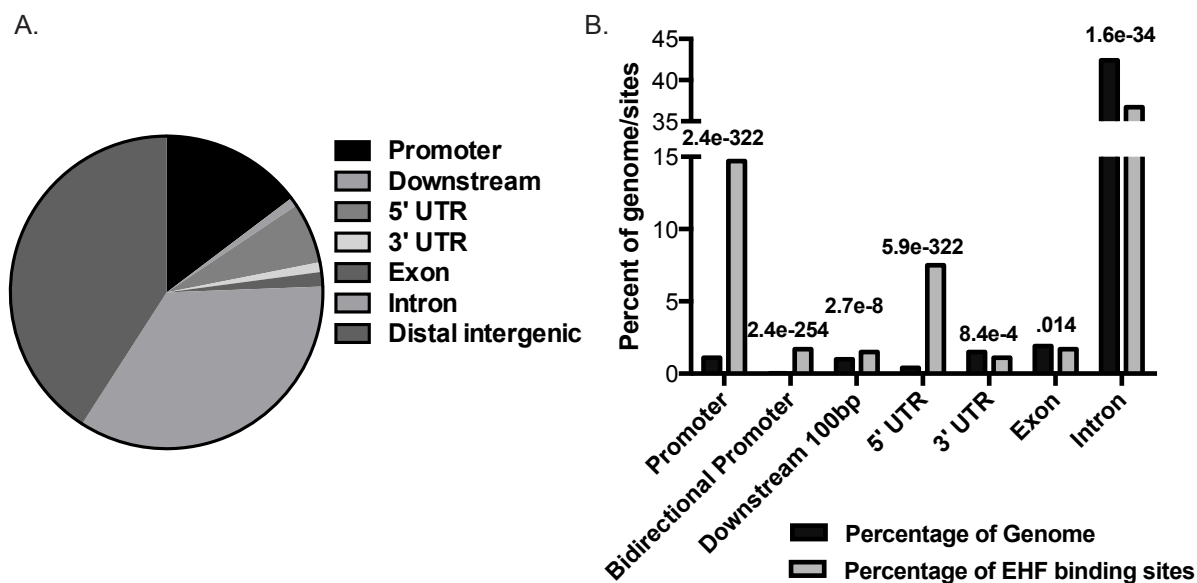
Figure 3.3.1 EHF binding sites are over-represented at promoters of genes in HBE cells. A) Percentage of EHF binding sites found at promoters, downstream of genes, 5' UTRs, 3' UTRs, exons, introns, and intergenic sites. B) Percentage of EHF ChIP-seq peaks found at specific sites as compared with genomic background. P values, derived using CEAS (one-sided binomial test), are written above bars.
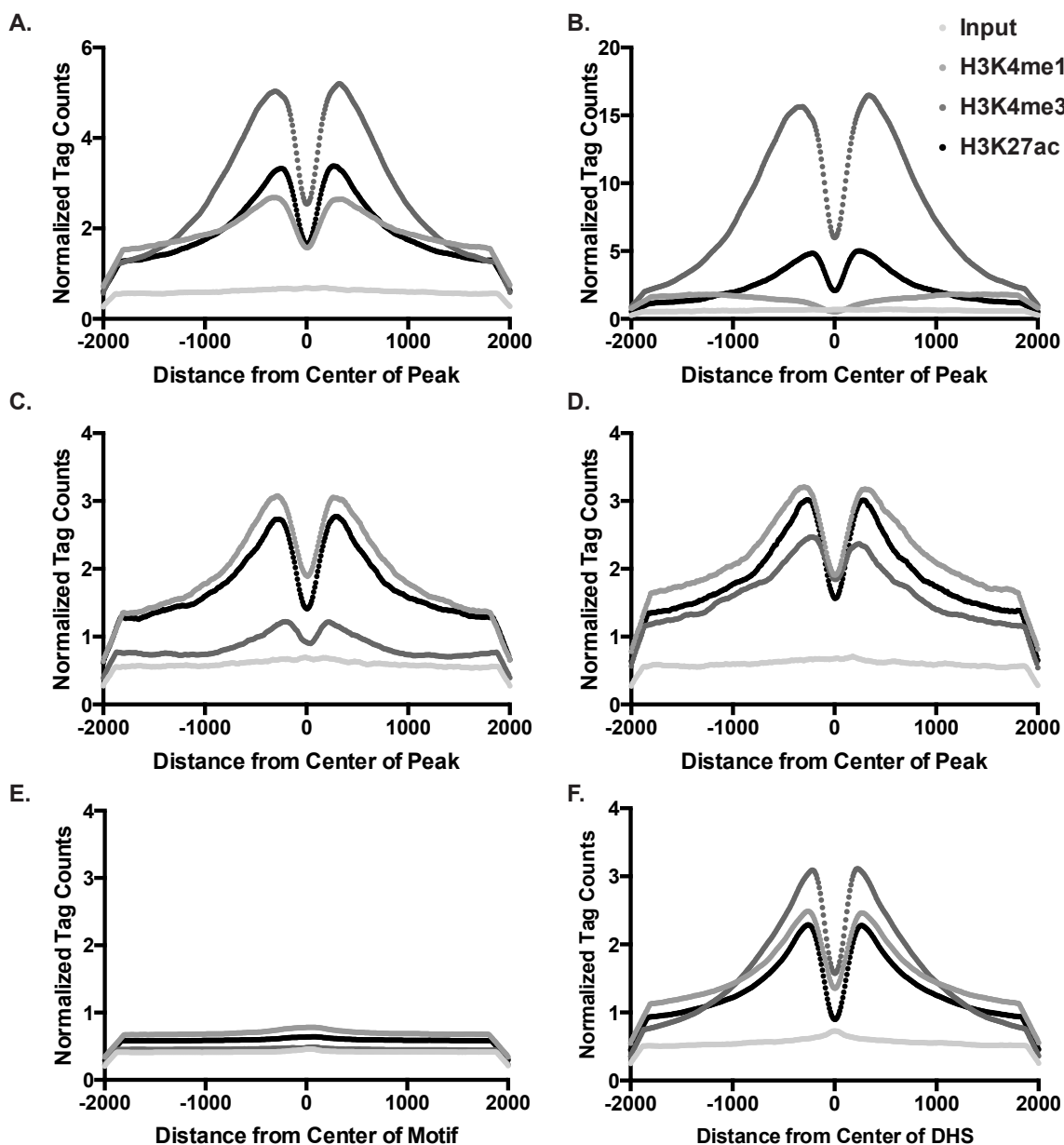
Figure 3.3.2 Peaks of EHF occupancy are enriched for active histone modifications. A) Normalized tag counts for H3K4me1, H3K4me3, and H3K27ac ChIP-seq in HBE cells, plotted from the center of all EHF binding sites. Normalized tag counts from an input sample are included as a control. B) As A, but for promoters occupied by EHF. C) As A, but for EHF-bound intergenic sites (outside of the promoter and gene body). D) As A, but for EHF sites in introns. E) As A, but plotted from the center of epithelial-specific Ets motif-containing genomic regions with no EHF occupancy. F) As A, but plotted from the center of HBE DHS where EHF does not bind. All panels use the legend found in the upper right.

Histone ChIP-seq tag counts were mapped at individual EHF sites, which were then grouped by hierarchical clustering (198). A pattern emerged for subsets of EHF occupancy peaks as defined by their histone modifications (Figure 3.3.3). Binding sites were marked by either H3K4me1 or H3K4me3, with little overlap of these modifications. All EHF occupancy sites containing H3K4me3 also had H3K27ac. However, the H3K4me1-marked regions were further subdivided into those with high or low mapped H3K27ac tag counts. There was a minority of EHF peaks that lacked all histone modifications studied.
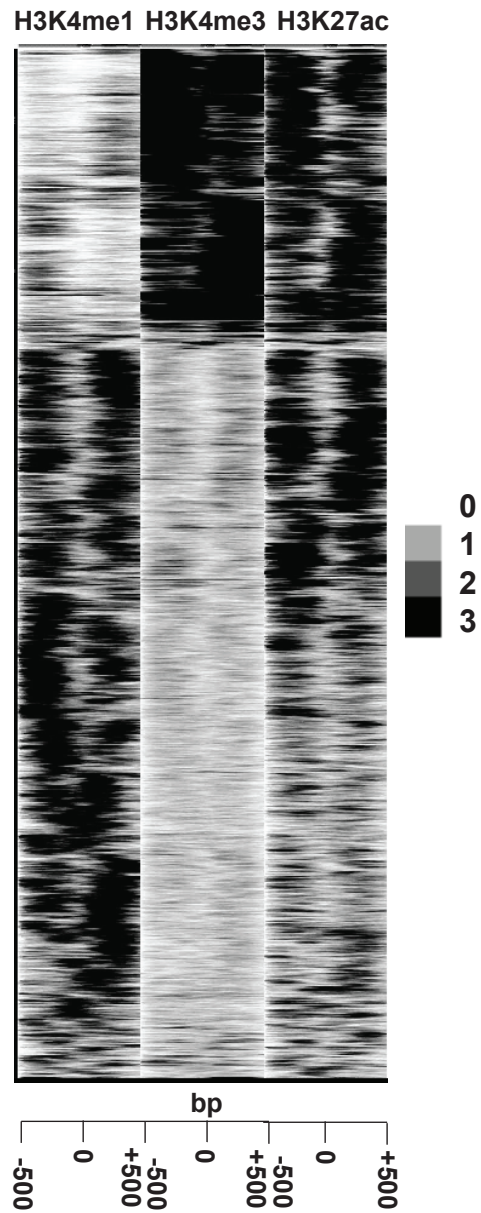
Figure 3.3.3 EHF occupancy sites in HBE cells show distinct patterns of histone modifications.
Histogram of H3K4me1 (left), H3K4me3 (center) and H3K27ac (right) tag counts in HBE cells mapped from the center of significant HBE EHF binding sites ± 500 bp. The rows represent individual EHF peaks, which were grouped by hierarchical clustering. The x-axis is measured in bp.

The EHF ChIP-seq IDR peaks from Calu-3 and HBE cells were intersected to determine the overlap. A total of 1,117 sites were common to both data sets. This represents 68% of the total peaks revealed by the Calu-3 ChIP-seq, which identified fewer binding sites than the HBE experiment. The promoter sites were also overlapped. Of the 304 total promoter peaks identified in the Calu-3 cells, 256 (84%) were found in the HBE cells.

3.4 EHF and activator protein 1 (AP-1) bind at overlapping sites genome-wide

De novo motif analysis was performed on the significant EHF ChIP-seq peaks in Calu-3 (Figure 3.4.1 A) and HBE (Figure 3.4.1 B) cells. In both cell types, the most enriched motif was an Ets motif similar to the consensus binding sites for the epithelial-specific Ets transcription factor subfamily members (EHF, ELF3 and ELF5). This motif was found in 76.32% (p = 1e-1248) and 84.25% (p = 1e-5785) of EHF binding sites within 200 bp of the center of the peak in Calu-3 and HBE cells, respectively, and confirms that the antibody specifically immunoprecipitated an Ets protein. This analysis also identified secondary enriched motifs, which can identify co-regulators of EHF gene regulation. In both cell types, the activator protein 1 (AP-1) motif was the second most enriched, found in 22.76% (p=1e-159) and 20.94% (p=1e-944) percent of sites in Calu-3 and HBE cells, respectively. This is consistent with previous results showing that EHF binds near AP-1 motifs in repressive contexts (160). In Calu-3 cells, the distribution of the Ets motif and AP-1 motif differ slightly. The Ets motif is most over-represented at the center of EHF peaks, with the AP-1 motif showing a plateau of enrichment within 100 bp of the center of the peak (Figure 3.4.1 C). In HBE cells, the binding motifs for Forkhead domain and Krüppel-like factor (KLF) transcription factor families were also among the most enriched secondary motifs, found in 15%, and 13.54% of peaks, respectively (Figure 3.4.1 B).

A.

Ets motif

P-Value	% of Targets	% of Background
1e-1248	76.32	5.03

AP-1 motif

P-Value	% of Targets	% of Background
1e-159	22.76	4.13

B.

Ets motif

P-Value	% of Targets	% of Background
1e-5785	84.25	15.23

AP-1 motif

P-Value	% of Targets	% of Background
1e-944	20.94	4.04

Forkhead motif

P-Value	% of Targets	% of Background
1e-172	15.00	7.27

Krüppel-like factor motif

P-Value	% of Targets	% of Background
1e-156	13.54	6.52

C.

Figure 3.4.1 EHF binding sites are enriched for the epithelial-specific Ets, Activator Protein 1 (AP-1), and other binding motifs. A) A *de novo* motif analysis within 200bp of the center of EHF ChIP-seq binding sites in Calu-3 cells identified the epithelial-specific Ets motif as the most enriched. The AP-1 motif is also enriched. Motif logos constructed in the analysis are shown with P values derived by HOMER. B) As A but for EHF ChIP-seq binding sites in HBE cells. The Forkhead and Krüppel-like factor motifs were also enriched. C) Incidence of the epithelial specific Ets and AP-1 motifs plotted from the center of Calu-3 EHF ChIP-seq peaks in bp.

Since EHF occupies the relatively common Ets motif, and it can increase or decrease gene expression, its interactions with other TFs are likely critical to specify binding location and mechanisms of action. The AP-1 binding matrix is the most enriched secondary motif in EHF occupancy sites, making it a candidate co-regulator of EHF function. AP-1 is a transcription factor complex composed of homo- and hetero-dimers of Jun, Fos, and other families of proteins (179). To confirm that EHF and AP-1 co-localize in the genome, ChIP-seq for c-Jun and JunD, members of the AP-1 protein complex expressed in the lung epithelium (195), was performed in Calu-3 lung adenocarcinoma cells. ENCODE consortium-validated antibodies were used to perform 2 biological replicates for each protein, and the HOMER IDR method was used to call significant peaks of transcription factor binding. A total of 11,517 and 7,524 sites were identified for c-Jun and JunD, respectively. There was a significant correlation between tag counts in the two samples at the significant peaks for c-Jun ($r = 0.76$, $p < 0.0001$) and JunD ($r = 0.69$, $p < 0.0001$) replicate experiments. These data sets were intersected with that of the EHF ChIP-seq in Calu-3 cells (Figure 3.4.2 A). Twenty-five percent of the EHF peaks overlapped at least one c-Jun or JunD peak, with 59% of these containing both c-Jun and JunD sites, 32% having only c-Jun and 9% containing only JunD. EHF, c-Jun and JunD tag counts were mapped with respect to the center of each EHF site, and these were grouped using hierarchical clustering (198). A heatmap was generated from the results (Figure 3.4.2 B). The pattern of c-Jun and JunD occupancy with respect to EHF was very similar. A subset of EHF sites had Jun binding at the center of the peak, while others showed displacement in either the 5' or 3' direction. As expected from the overlap, some EHF peaks did not have Jun occupancy. As a control, EHF, c-Jun and JunD tag counts were plotted from the center of Calu-3 DHS that lacked significant EHF binding sites (Figure 3.4.2 C). A fraction of these sites show EHF enrichment below the threshold required to call a significant binding region. A subset of these DHS have co-occupancy of c-Jun and JunD, as expected from the overlap with EHF that showed the majority of their binding sites do not overlap EHF. However, the total proportion of these DHS that contain c-Jun and JunD is lower than that of the EHF sites, suggesting the overlap is significant. Furthermore,
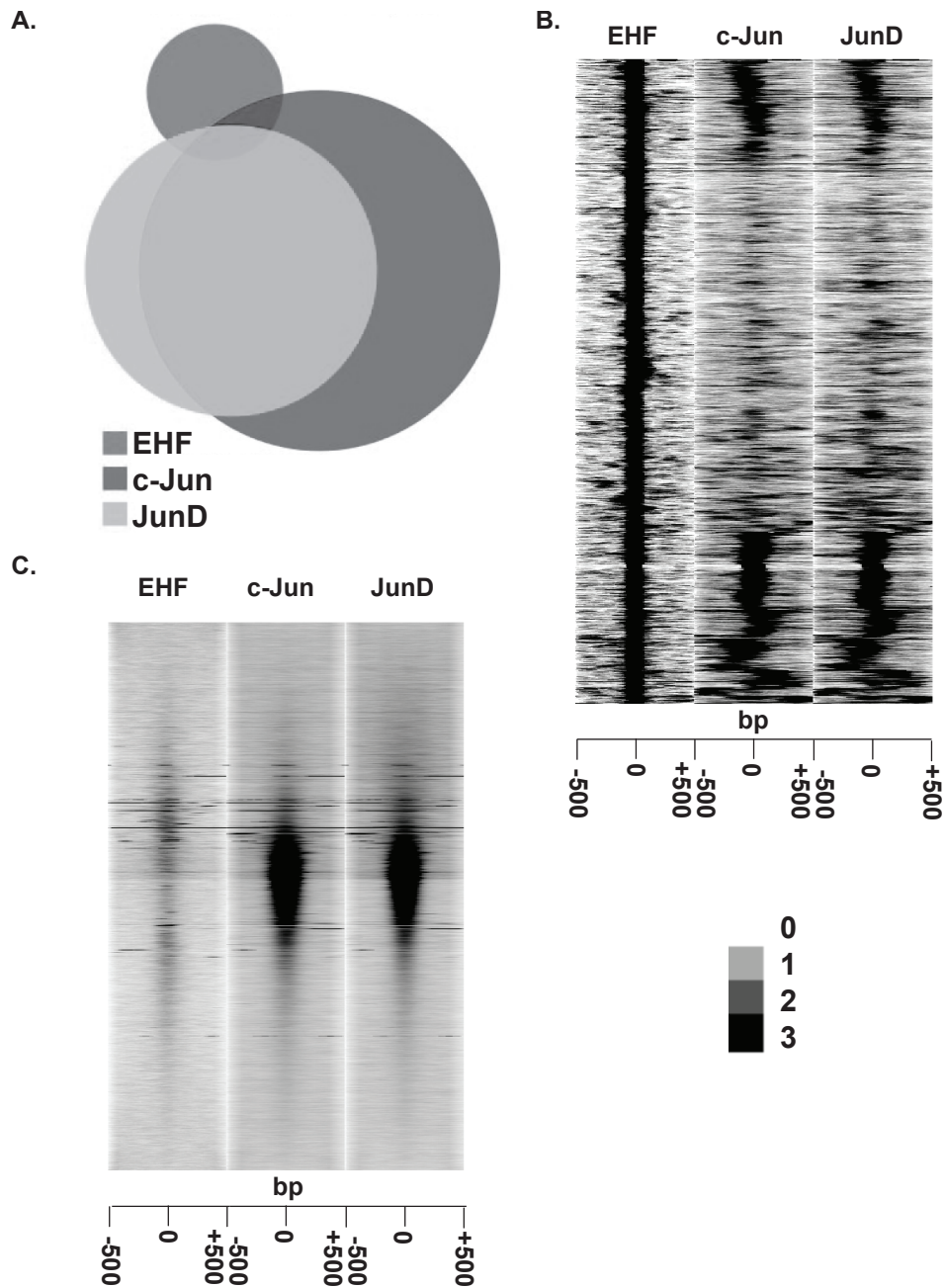
Figure 3.4.2 EHF and AP-1 bind at overlapping sites genome-wide. A) Overlap of EHF, c-Jun and JunD ChIP-seq binding sites in Calu-3 lung adenocarcinoma cells. B) Histogram of EHF (left), c-Jun (center) and JunD (right) tag counts mapped from the center of significant EHF binding sites ± 500 bp. Hierarchical clustering was performed on the EHF peaks, each of which is represented by a row. The x-axis is measured in bp. C) As B, but plotted from the center of Calu-3 DHS that do not contain an EHF binding site.

CEAS was used to determine genomic locations of EHF/AP-1 co-occupancy sites (Figure 3.4.3 A). All sites were over-represented at promoters and 5' UTRs, but EHF sites that coincide with only a JunD binding peak were more enriched at these genomic locations than those intersecting with a c-Jun peak. Tag counts for the histone modifications H3K4me1, H3K4me3, and H3K27ac from Calu-3 cells were plotted from the center of the EHF/c-Jun/JunD (Figure 3.4.3 B), EHF-c-Jun (Figure 3.4.3 C), and EHF/JunD (Figure 3.4.3 D) overlapping sites. The EHF sites coinciding with AP-1 binding are very enriched for H3K27ac, higher than at all EHF binding sites, suggesting they are found in active elements. The EHF/JunD sites show higher H3K4me3 normalized tag counts than EHF/c-Jun sites, corresponding to higher over-representation at the promoter. The nearest gene annotation method was used to determine potential genes regulated through TF binding to EHF/AP-1 sites, and these were subjected to a GO process enrichment analysis using MetaCore from Thomson Reuters version 6.26 (Figure 3.4.3 E). The overlapping sites are found near loci involved in differentiation, development and wound repair, processes that correspond to known functions of AP-1 in the lung (181). These data suggest that EHF may work with AP-1 to regulate these phenotypes important for airway pathology.
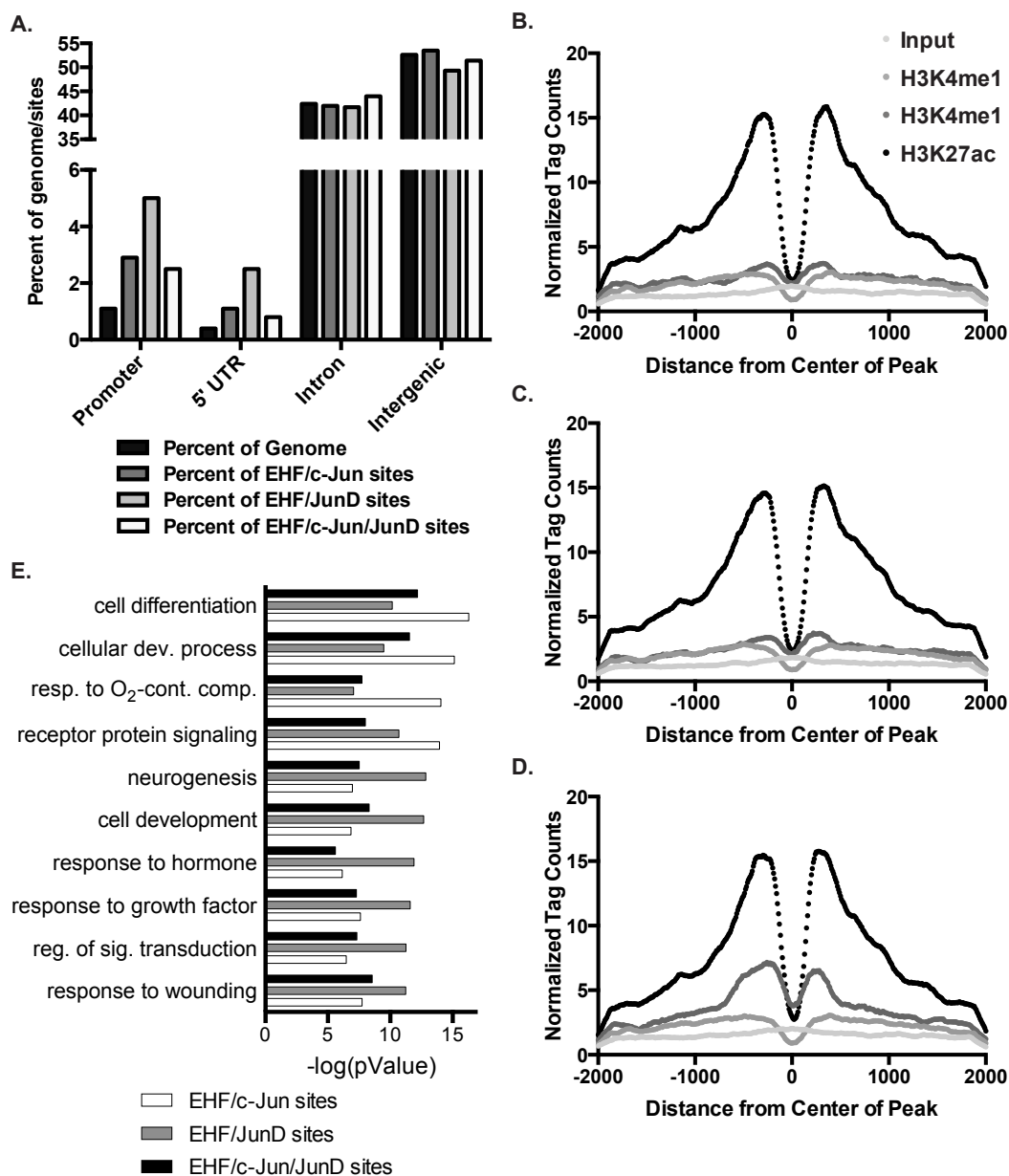
Figure 3.4.3 Sites of EHF and AP-1 binding are enriched for active histone marks and found near genes involved in cell differentiation and wound repair. A) Percentage of EHF/AP-1 co-occupancy sites found at promoters, 5' UTRs, introns, and intergenic sites as compared with genomic background. Enrichment of overlapping sites is seen at the promoter and 5' UTR. B-D) Normalized tag counts for H3K4me1, H3K4me3, and H3K27ac ChIP-seq in Calu-3 cells, mapped from the center of B) EHF/c-Jun/JunD, C) EHF/c-Jun, and D) EHF/JunD overlapping peaks. Normalized tag counts from input sample are included as a control. E) A gene ontology enrichment analysis was performed on the nearest gene to each co-occupancy site. Significance of enrichment is displayed as the –log (P Value).

The co-occupancy of EHF and AP-1 across the genome suggests that the two proteins may directly interact. The ETS domain of multiple TFs has been shown to bind to Jun family members through co-immunoprecipitation experiments (co-IPs) (164). To determine whether EHF directly interacts with c-Jun or JunD, a co-IP was performed in human embryonic kidney 293 (HEK) cells. Flag-EHF was expressed in the cells, which were then lysed in NET buffer. EHF was immunoprecipitated using Anti-FLAG M2 affinity gel, and enriched protein complexes were probed on a western blot for presence of EHF, c-Jun and JunD (Figure 3.4.4). In the input, (non-immunoprecipitated lysate), Flag-EHF, c-Jun and JunD were present. In the IP, Flag-EHF was detected, suggesting that the resin efficiently pulled down the protein. However, in the IP samples, neither c-Jun nor JunD was detected. This suggests that EHF does not interact with these proteins in this *in vitro* system. However, if the interaction between EHF and c-Jun or JunD is weak or transient, it may not be easily isolated in a co-IP experiment. Thus, a different method that can detect the kinetics of interactions was used.
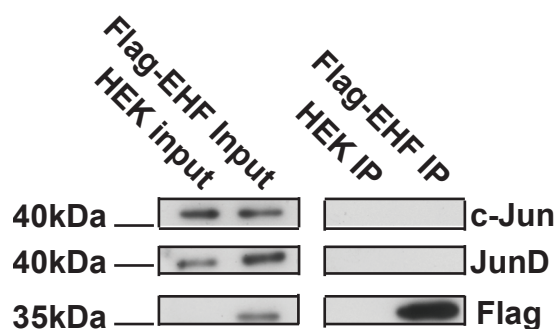
Figure 3.4.4 Co-immunoprecipitation (Co-IP) does not support a direct interaction between EHF and Jun proteins. Input was removed from lysates of HEK cells that were transfected with an empty vector (HEK input) or vector expressing N-terminal Flag-tagged EHF (Flag-EHF input). The remaining lysate was immunoprecipitated using Anti-FLAG M2 affinity gel, and enriched protein was eluted from the resin (HEK IP or Flag-EHF IP). All samples were probed for presence of c-Jun (top), JunD (middle) or Flag (bottom) by western blot. The Flag-EHF was specifically pulled down, but c-Jun and JunD were not co-immunoprecipitated. Shown is one representative experiment (n=3).

As a second measure of the interaction between EHF and c-Jun, BLItz^{TM} bio-layer interferometry was utilized (199). This assay requires purified proteins, one of which must have a His epitope tag for binding to the biosensors on the platform. EHF was cloned into a bacterial expression vector with an N-terminal His6 epitope tag. Its expression in *E. Coli* and lysis of the bacteria were optimized and monitored using silver stain gels. The growth condition parameters tested and their respective effects on His6-EHF protein levels are summarized in Table 3.4.1. Two *E. coli* strains, growth temperatures, lengths of incubation and types of growth media, and 4 concentrations of IPTG were tested. Induction of His6-EHF with 100 μM IPTG for 24 hours at 37°C in LB medium led to highest expression of His6-EHF (Figure 3.4.5 A-C). The lysis buffers and supplements tested to solubilize His6-EHF in bacterial lysate are listed in Table 3.4.2. With all lysis methods tested, the expressed protein aggregated in insoluble inclusion bodies (IBs) (Figure 3.4.6). Therefore, the IBs were collected in a urea buffer and then refolded and purified on a Ni-charged resin (Figure 3.4.7 A-B). A small amount of refolded protein was obtained, but not a sufficient quantity for use in the bio-layer interferometry. The Northwestern Recombinant Protein Production Core expressed and purified sufficient His6-EHF for use in the assay; the correct protein was detected by western blot (Figure 3.4.7 C).

Table 3.4.1 Growth conditions tested to optimize His6-EHF expression in *E. Coli.* Listed are the category of the condition (growth strain, temperature, time, concentration of IPTG for induction, and growth media), the different conditions tested, and the effect on His6-EHF expression.

Category	Condition	Result
Growth strain	BL21 (DE3)	No expression
	BL21-CodonPlus (DE3)-RP	Expression of His6-EHF
Growth temperature	37°C	Expression of His6-EHF
	24°C	No expression of His6-EHF
Growth time	4 hours	Expression of His6-EHF
	24 hours	Expression of His6-EHF (similar to 4 hours)
Concentration of IPTG	10µM	Low expression- does not improve solubility
	100µM	Highest expression
	500µM	Moderate expression
	1mM	Moderate expression
Growth media	NZCYM	Medium expression of His6-EHF
	LB	Highest expression of His6-EHF
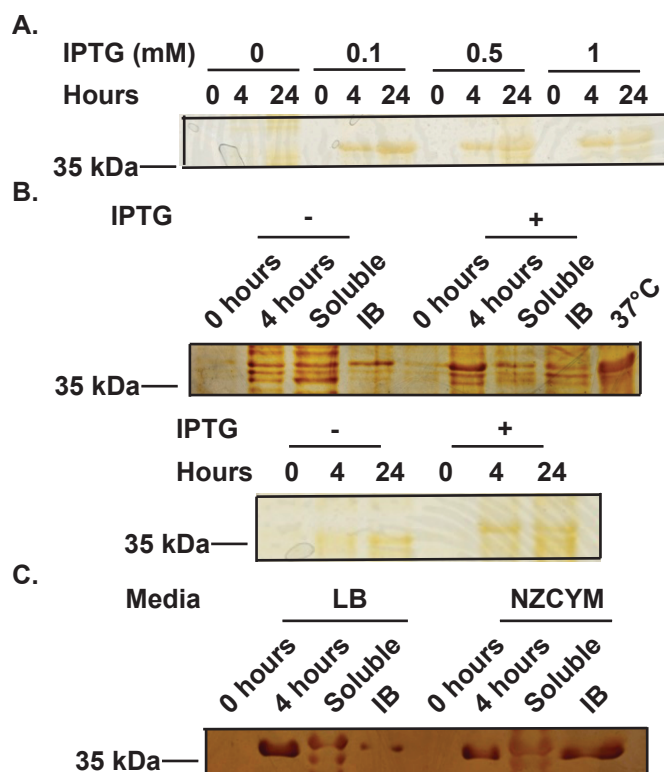
Figure 3.4.5 Optimization of His6-EHF expression in *E. Coli.* BL21-CodonPlus (DE3)-RP bacteria were transformed with pET-His6-TEV-EHF and then induced with IPTG under different conditions to optimize His6-EHF expression. Bacterial pellets were collected and probed for protein levels by silver stain gel. A) Four concentrations of IPTG were tested for induction of protein expression for 0, 4, or 24 hours at 37°C. B) Cells were induced with 100 μM IPTG and incubated for 0, 4, or 24 hours at 24°C (top) or 37°C (bottom). The cells incubated at 25°C were lysed, and the soluble and insoluble inclusion body (IB) fractions were also probed for protein expression. C) Bacteria were grown in LB or NZCYM media and induced with 100 μM IPTG for 0 or 4 hours at 37°C. The bacteria were lysed, and the soluble and IB fractions were also probed for protein expression.

Table 3.4.2 Lysis conditions tested to solubilize His6-EHF. Listed are the base buffers, additives, and pHs tested to solubilize His6-EHF.

Category	Condition
Base buffer	50 mM sodium phosphate
	50 mM Tris-HCl
Additive	5% Glycerol
	1% Triton-X
	0.5% NP-40
	25% sucrose
	1 mM DTT
pH	8
	7.5
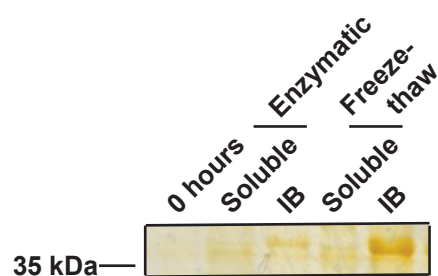
Figure 3.4.6 Lysis of bacteria expressing His6-EHF yields protein aggregation in insoluble inclusion bodies. pET-His6-TEV-EHF transformed BL21-CodonPlus (DE3)-RP *E. Coli* were induced with 100 µM IPTG for 0 or 24 hours at 37°C. After 24 hours, bacteria were lysed in 50 mM sodium phosphate buffer supplemented with 300 mM NaCl, lysozyme and DNase I using an enzymatic or freeze-thaw mechanism. Both the soluble and insoluble (IB) fractions were probed for protein by silver stain.
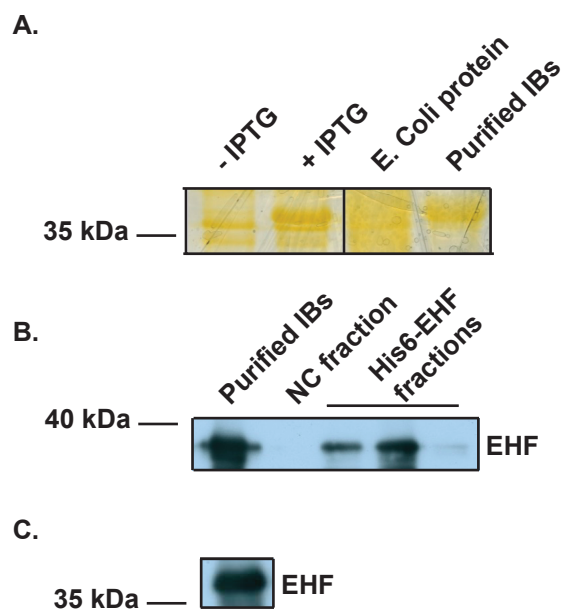
Figure 3.4.7 Purification of His6-EHF. A) pET-His6-TEV-EHF transformed BL21-CodonPlus (DE3)-RP *E. Coli* were induced with 100 µM IPTG for 24 hours at 37°C (+IPTG). A non-induced culture (- IPTG) was included as a control. The IBs were isolated from the soluble protein (E. Coli protein) using a urea-based buffer. The bacterial pellets, soluble *E. coli* protein, and purified IBs were probed for protein expression by silver stain. B) The isolated IBs were refolded and further purified on a Ni-charged column. The first eluted fraction was saved as a negative control (NC fraction). The purified IBs, NC fraction, and eluted fractions containing His6-EHF were probed for the presence of EHF by western blot. C) Detection of purified His6-EHF generated by the Northwestern University Recombinant Protein Production Core by western blot. This purified protein was used for the bio-layer interferometry assay.

An N-terminal Flag tagged c-Jun construct was generated and transformed into BL21-CodonPlus (DE3)-RP *E. Coli.* Induction of Flag-c-Jun with 2 concentrations of IPTG (10 µM or 1mM), incubation times (4 or 24 hours) and temperatures (24 or 37°C) was investigated. Treatment with 10 µM of IPTG for 24 hours at 24°C yielded soluble protein (Figure 3.4.8 A-B), which was purified with Anti-FLAG M2 affinity gel and tested for purity by silver stain and presence of the correct protein by western blot (Figure 3.4.8 C).
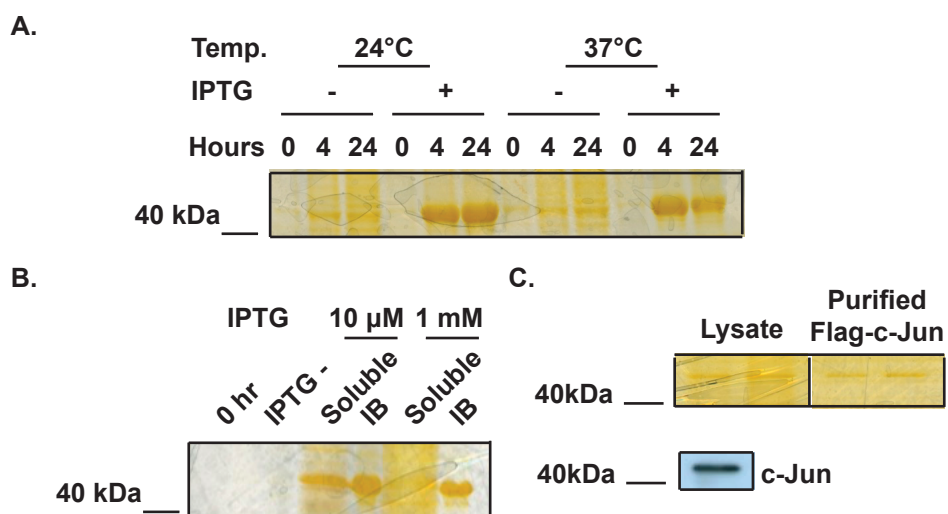
Figure 3.4.8 Expression and purification of Flag-c-Jun. BL21-CodonPlus (DE3)-RP bacteria were transformed with pET-Flag-TEV-c-Jun. A) Bacteria were induced with 10 µM IPTG for 0, 4, or 24 hours at 24°C or 37°C to optimize expression of Flag-c-Jun. Un-induced cultures (- IPTG) were included as a control. Bacterial pellets were probed for presence of protein by silver stain gel. B) Cells were induced with 10 µM or 1mM IPTG and incubated for 24 hours at 24°C. Bacteria were lysed, and the soluble and IB fractions were probed for protein expression by silver stain gel. Un-induced (- IPTG) and 0 hour samples were include as controls. C) c-Jun was purified using Anti-FLAG M2 affinity gel. The input lysate and purified sample were probed for the presence of protein by silver stain (top). Purified Flag-c-Jun used for bio-layer interferometry was detected by western blot (bottom).

For the bio-layer interferometry, His6-EHF was immobilized on a Ni-NTA biosensor (Figure 3.4.9 A). The sensor was then incubated in TBS buffer containing Flag-c-Jun and binding was measured for 120 seconds (Figure 3.4.9 B). There was no binding of Flag-c-Jun to the His6-EHF coated sensor above that seen in the negative control, Flag-c-Jun binding to an uncoated sensor. In a previous study, the ETS domain of multiple Ets transcription factors was co-immunoprecipitated with c-Jun in NETN buffer (164). To test whether EHF and c-Jun interact in this buffer, the purified Flag-c-Jun was dialyzed with NETN buffer. Then, the bio-layer interferometry was repeated with His6-EHF bound to the Ni-NTA biosensor and dipped into Flag-c-Jun in NETN. Again, no binding of Flag-c-Jun to His6-EHF was seen (Figure 3.4.9 C). These data do not support a direct interaction between EHF and c-Jun.
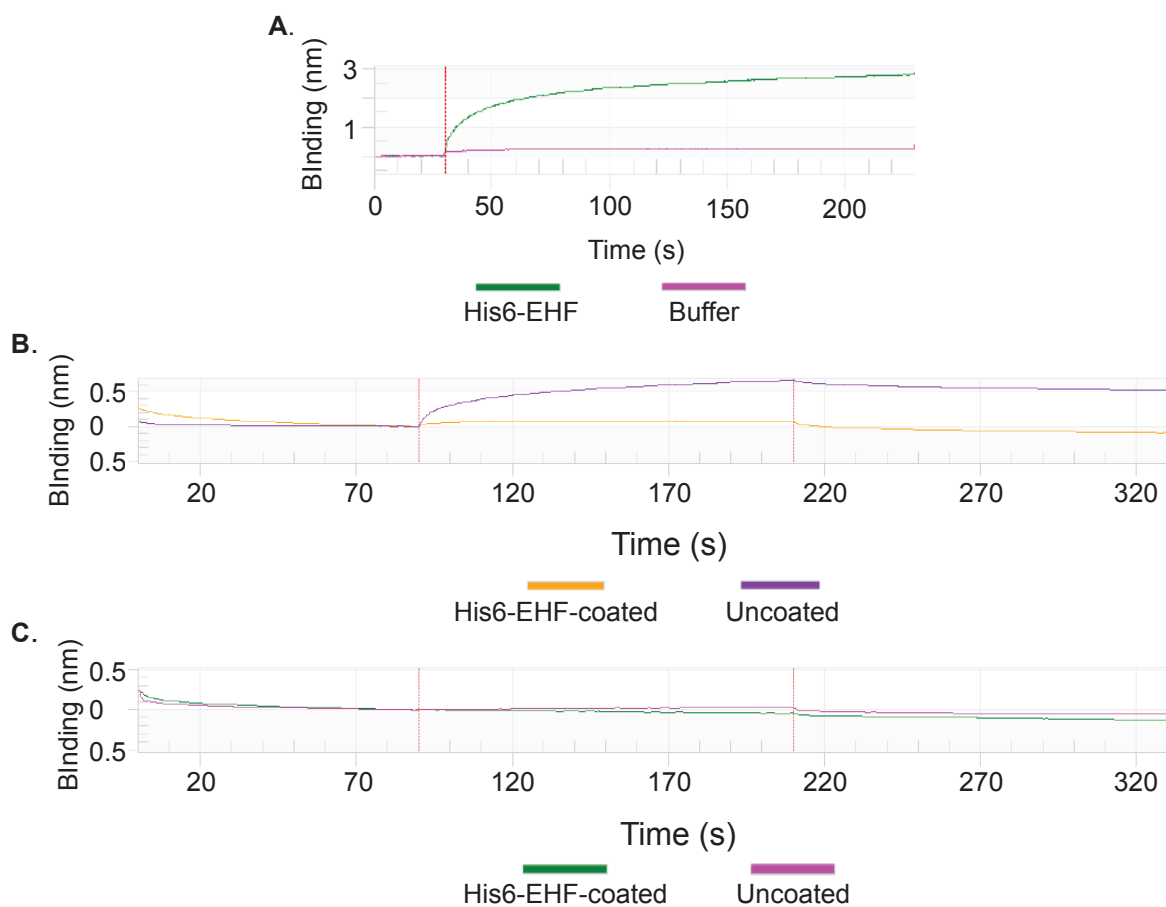
Figure 3.4.9 Bio-layer interferometry does not support a direct interaction between EHF and c-Jun.
A) BLItz[TM] tracing of His6-EHF binding to the Ni-NTA sensor. The tracing shows 30 seconds (s) of initial baseline in buffer followed by 200 s of incubation with His6-EHF (green) or buffer only (purple) to coat the sensor. B-C) BLItz[TM] tracing of incubation of sensor in Flag-c-Jun in buffer. Tracings show 90 s of baseline in buffer, 120 s of incubation in Flag-c-Jun, then 120 s of dissociation in buffer. B) Tracing of Flag-c-Jun in TBS binding to His6-EHF-coated (yellow) or -uncoated (purple) sensor. There was some non-specific binding of Flag-c-Jun to the uncoated sensor, but no binding of Flag-c-Jun to the His6-EHF-coated sensor. C) Tracing of Flag-c-Jun in NETN buffer binding to His6-EHF-coated (green) or -uncoated (purple) sensor. There was no binding of Flag-c-Jun to the His6-EHF-coated or –uncoated sensor. For all tracings, binding was measured in nanomoles (nm).

3.5 Modulation of EHF in airway epithelial cell lines alters genes involved in epithelial differentiation, barrier function and response to wounding.

3.5.1 EHF depletion in Calu-3 cells alters genes important for epithelial differentiation, cell-cell adhesion, and response to wounding

To identify target genes regulated by EHF, either directly or indirectly, a specific siRNA was used to deplete EHF in Calu-3 cells (Figure 3.5.1.i A). Total RNA was isolated from five NC siRNA and five EHF siRNA-treated samples. RNA quality was confirmed by Nanodrop measurement of OD 260/280 and 260/230 ratios. RNA-seq was performed on libraries generated from the 10 samples, and the resulting sequence reads were mapped to the genome using TopHat. CuffLinks (129) was used to generate gene expression values in the format of fragments per kilobase per million mapped fragments (FPKM), which was then compiled into an expression value matrix. Principle component analysis on the FPKM values showed that the five samples within each treatment group clustered together, and the different treatment groups segregated along PC #2 (Figure 3.5.1.i B). CuffDiff (129) was used to determine differentially regulated genes following EHF depletion. A total of 256 genes showed ≥ 1.5-fold difference in expression following EHF knockdown (Figure 3.5.1.i C). One hundred and twenty one and 135 genes showed decreased or increased levels, respectively, suggesting that EHF can positively or negatively regulate transcription. The average distance between EHF-bound sites and EHF-activated or -repressed genes (2930 kb and 2524 kb, respectively) was not significantly different (P = 0.435).
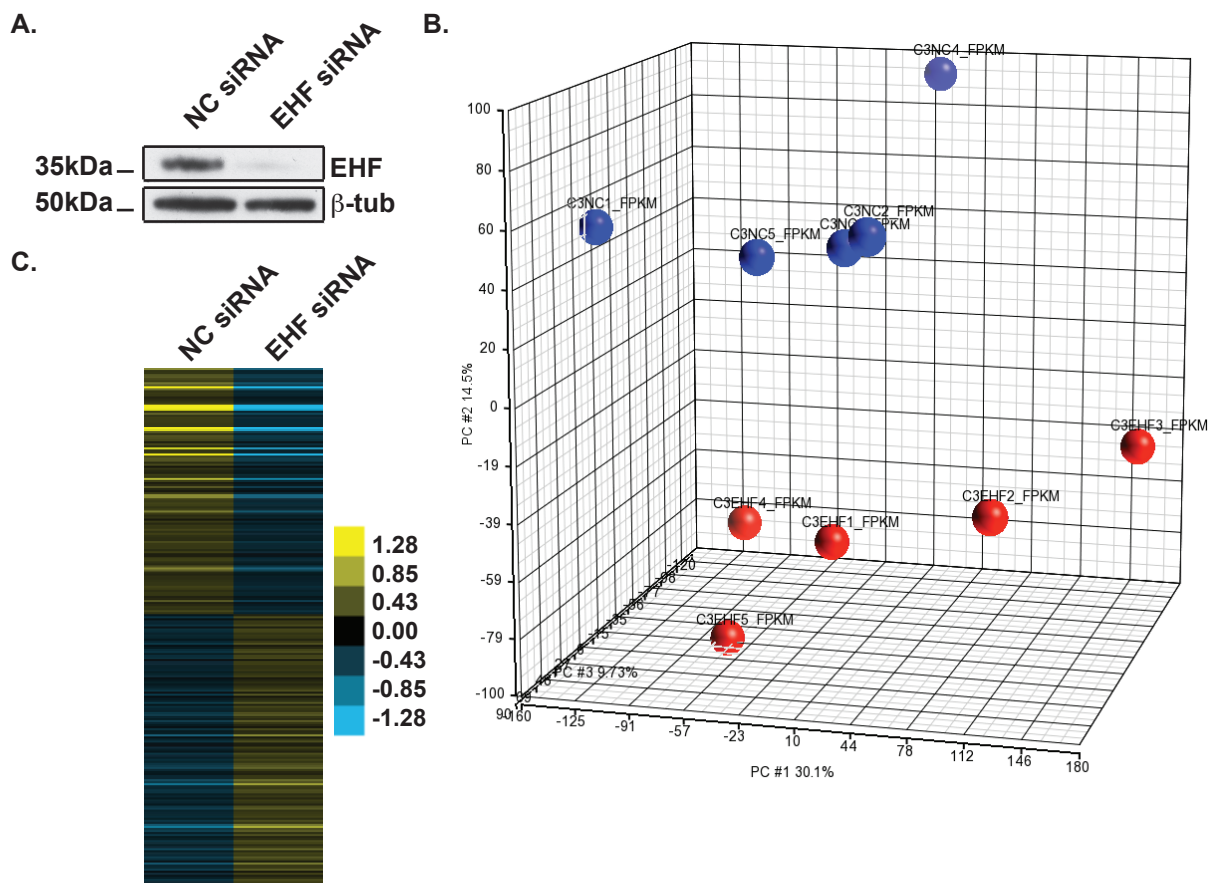
Figure 3.5.1.i EHF depletion followed by RNA-seq alters gene expression in Calu-3 cells. A) siRNA depletion of EHF in Calu-3 cells measured by western blot compared to β-tubulin. B) Whole transcriptome expression values in FPKM at gene level were used to compare similarity among the samples through dimension reduction analysis of principle component analysis (PCA). The top 3 Principle Components (PCs) that explain the largest variability of the data were plotted on a 3D plot, with NC samples in blue and EHF siRNA-treated samples in red). Samples from different treatment groups are segregated along PC #2, which explains 14.5% of all variability in the data. C) Relative expression of 256 genes that showed ≥ 1.5-fold change in expression following siRNA depletion of EHF. Each row represents a different gene. The scale is $\log_2$ fold change relative to the average gene expression value.

A GO process enrichment analysis was performed on the differentially expressed genes (DEGs) following EHF depletion using DAVID (Table 3.5.1). The DEGs were enriched for GO pathways crucial for lung development and repair. Genes involved in epithelial cell differentiation and development included frizzled class receptor 1 (*FZD1*), components of the cornified envelope including small proline-rich proteins *SPRR1A* and *SPRR1B*, and keratins 4, 5 and 14 (*KRT4, KRT5, KRT14*). Genes involved in locomotory behavior and response to wounding were also enriched, with DEGs including S100 calcium binding proteins A8 and A9 (*S100A8, S100A9*), chemokine (C-X-C motif) ligands 1 and 6 (*CXCL1, CXCL6*), chemokine (C-C Motif) ligand 5 (*CCL5*), and plasminogen activator, urokinase (*PLAU*).

Table 3.5.1 EHF depletion in Calu-3 cells alters genes important for epithelial differentiation, cell-cell adhesion, and response to wounding. GO analysis by DAVID of differentially expressed genes (DEGs) with ≥ 1.5-fold change in expression following EHF knockdown.

GO Term	P Value
epithelial cell differentiation	6.85E-05
regulation of epidermis development	1.61E-04
extracellular region	6.17E-04
epithelium development	7.13E-04
ectoderm development	0.001089007
epidermis development	0.002608863
locomotory behavior	0.002899408
response to wounding	0.003509121
regulation of response to external stimulus	0.004469944

Gene expression changes following EHF reduction were validated using siRNA depletion followed by quantitative reverse transcription polymerase chain reaction (RT-qPCR). DEGs involved in cell–cell junctions, immune function and cell motility were tested. Expression changes seen in the RNA-seq data were replicated in the qPCR validation (Figure 3.5.1.ii A). *JDP2, CXCL1*, interleukin-8 (*IL-8*), and dimethylarginine dimethylaminohydrolase 1 (*DDAH1*) increased in expression in EHF-depleted cells, while *TJP2* and *ITGA2* decreased and tight junction protein 1 (*TJP1*) showed no change. These alterations corresponded to the RNA-seq results, with the exception of the lack of increase in *TJP1*. The RT-qPCR and RNA-seq signals are significantly correlated (r = 0.81, P = 0.027). *ITGA2* showed decreased expression following EHF KD, and contains an EHF binding site at its promoter. To confirm that transcriptional changes identified in the RNA-seq data were also seen at the protein level, whole cell lysates from negative control and EHF siRNA-treated samples were analyzed by western blot to look for changes in integrin, α2 protein levels. A specific antibody showed that integrin, α2 protein decreased in the EHF-depleted cells as compared with the NC cells, consistent with the mRNA data (Figure 3.5.1.ii B). This was repeated for three different transfections and quantified relative to β-tubulin using gel densitometry, showing a significant decrease (p<0.05). Since EHF binds to its promoter, this suggests *ITGA2* is a direct target of EHF activation.
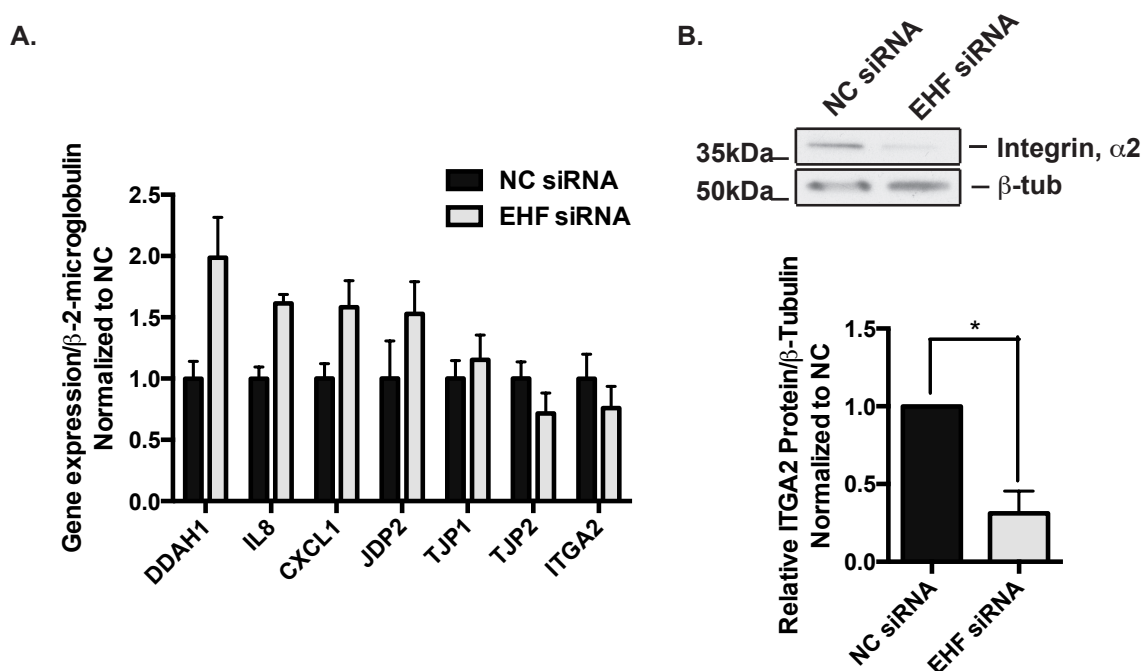
Figure 3.5.1.ii Validation of RNA-seq results in Calu-3 cells confirms that EHF regulates targets involved in cell-cell adhesion, cell motility, and inflammation. A) EHF siRNA depletion followed by RT-qPCR confirmed changes in gene expression identified by RNA-seq (n=2). Dimethylarginine dimethylaminohydrolase 1 (*DDAH1*), interleukin-8 (*IL-8*), *CXCL1*, and *JDP2* indreased, while *TJP2* and *ITGA2* decreased; these alterations were consistent with RNA-seq results. Tight junction protein 1 (*TJP1*) did not change. B) Integrin, α2 protein was measured by western blot (top) and quantified by gel densitometry (bottom). Integrin, α2 protein significantly decreased following EHF depletion (n=3), corresponding to the observed decrease in RNA abundance for this gene. *p<0.05 by an unpaired two-tailed Student's t test.

3.5.2 Overexpression of EHF in A549 cells changes expression of genes involved in cell adhesion, cell proliferation, and response to wounding

A549 cell clones that stably overexpress EHF from a transfected plasmid (pcDNA 3.1(-)) were generated. This alveolar-like lung carcinoma cell line (191) was chosen due to its low levels of endogenous EHF, in contrast to the high levels seen in Calu-3 cells. RNA-seq was performed on 3 A549 clones carrying the empty vector and 3 overexpressing EHF. Western blots showed increased EHF protein in the latter clones as compared with the former (Figure 3.5.2 A). Total RNA from 3–4 replicates of each clone was isolated, tested for quality, and sequenced. At the transcript level, there was an 8.6-fold increase in EHF in the over-expressing clones. A total of 1,153 genes showed a significant change in expression of ≥ 1.5-fold in the EHF-overexpressing clones as compared to the empty vector clones (Figure 3.5.2 B), with 761 increasing and 392 decreasing. These data confirm that EHF can positively or negatively regulate gene expression, but suggest a slightly larger role for EHF in activating gene transcription.
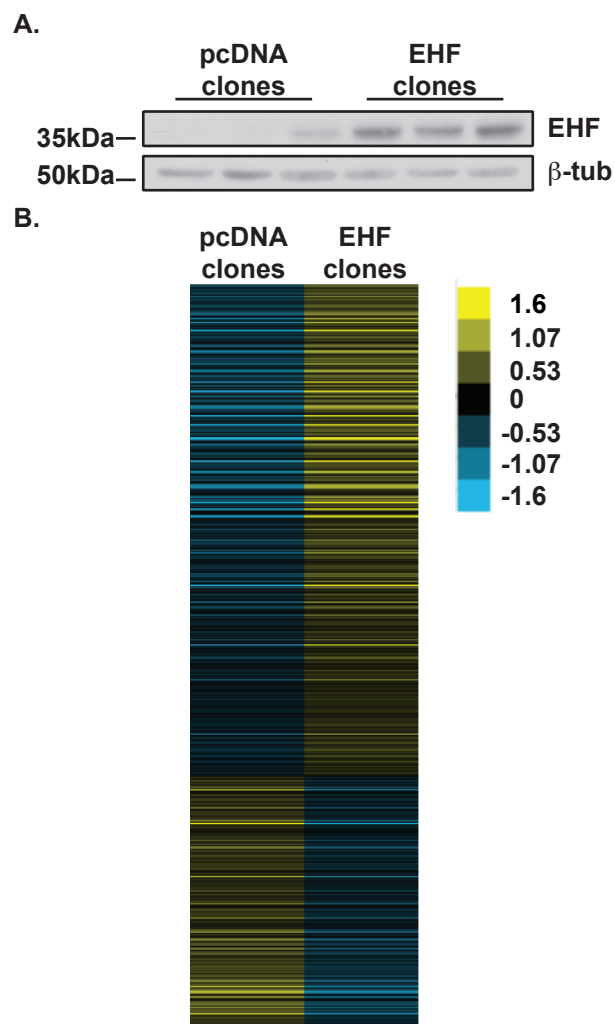
Figure 3.5.2 Overexpression of EHF followed by RNA-seq identifies differentially expressed genes in A549 cells. A). Western blot probed with an antibody to EHF shows low levels of EHF in A549 cell clones containing the pcDNA 3.1(-) vector alone (pcDNA clones) and higher levels in clones expressing pcDNA 3.1(-)-EHF (EHF clones). β-tubulin provided the loading control. B) Relative expression of 1,153 genes that showed ≥ 1.5-fold change in expression in EHF overexpressing clones versus control clones. Each row represents a different gene. The scale is $\log_2$ fold change relative to the average gene expression value.

A GO process enrichment analysis was performed on the DEGs using DAVID (Table 3.5.2). These genes were enriched for processes including the extracellular matrix, plasma membrane structure and function, cell proliferation and response to wounding. Genes encoding proteins important for the extracellular region include multiple collagens (*COL4A3, COL5A1, COL5A2, COL7A1* and *COL12A1*) and mucins (*MUC1, MUC3A, MUC5AC* and *MUC13*). Transcripts related to the plasma membrane include matrix metallopeptidase 14 (*MMP14*), *CD44* and multiple solute carrier family members relevant to lung epithelial function (*SLC6A14, SLC7A2, SLC7A7, SLC12A2, SLC16A7* and *SLC23A2*). *SLC6A14* is a gene modifier of lung disease severity and age of first *Pseudomonas aeruginosa* infection in CF patients (201). DEGs involved in response to wounding include interleukins (*IL1A, IL6, IL11* and *IL15*) and the PLAU receptor (*PLAUR*).

Table 3.5.2 Overexpression of EHF in A549 cells changes expression of genes involved in cell adhesion, cell proliferation, and response to wounding. A GO analysis was performed on the DEGs with ≥ 1.5-fold change in expression in EHF overexpressing clones using DAVID.

GO Term	P Value
extracellular region part	2.37E-13
extracellular region	6.97E-11
plasma membrane part	4.74E-10
extracellular space	2.01E-09
plasma membrane	1.65E-08
intrinsic to plasma membrane	2.60E-08
integral to plasma membrane	2.66E-08
regulation of cell proliferation	1.38E-07
response to hormone stimulus	1.58E-07
response to wounding	1.97E-07
calcium ion binding	7.31E-07
cell proliferation	1.68E-06
urogenital system development	2.18E-06
response to endogenous stimulus	2.60E-06
response to peptide hormone stimulus	3.81E-06

Fifty-eight genes were positively regulated in both Calu-3 and A549 cells (expression decreased in EHF depleted Calu-3 cells and increased in EHF-overexpressing A549 clones). A DAVID gene ontology enrichment analysis showed that these 58 genes were associated with processes of epithelial differentiation, anchoring junction and regulation of cell proliferation. Twenty-eight genes were negatively regulated in both cell types, many of which encode cell surface proteins.

3.6 EHF depletion in HBE cells identifies targets involved in transcriptional regulation, the immune response, and response to wounding

To detect genes that are directly and indirectly regulated by EHF, RNA-seq was performed on EHF-depleted HBE cells. A targeted siRNA was used to knock down EHF (Figure 3.6.1 A), then RNA was extracted from NC and EHF siRNA treated cells and sequenced. Cufflinks (129) calculated the expression of all genes in FPKM for each sample. A multi-dimensional scaling (MDS) analysis was performed to determine the similarity between samples within the same treatment group and between groups (Figure 3.6.1 B). Within each treatment group, 4 out of 5 samples clustered together and away from the other treatment group. The groups differed along M1 and M2. Cufflinks was used to identify a total of 1,145 genes that were differentially regulated following EHF depletion (FDR <0.05), with 625 increasing and 520 decreasing in expression (Figure 3.6.1 C).

A.

	NC siRNA	EHF siRNA	
35 kDa —			— EHF
50 kDa —			— β-tub

C.

B.
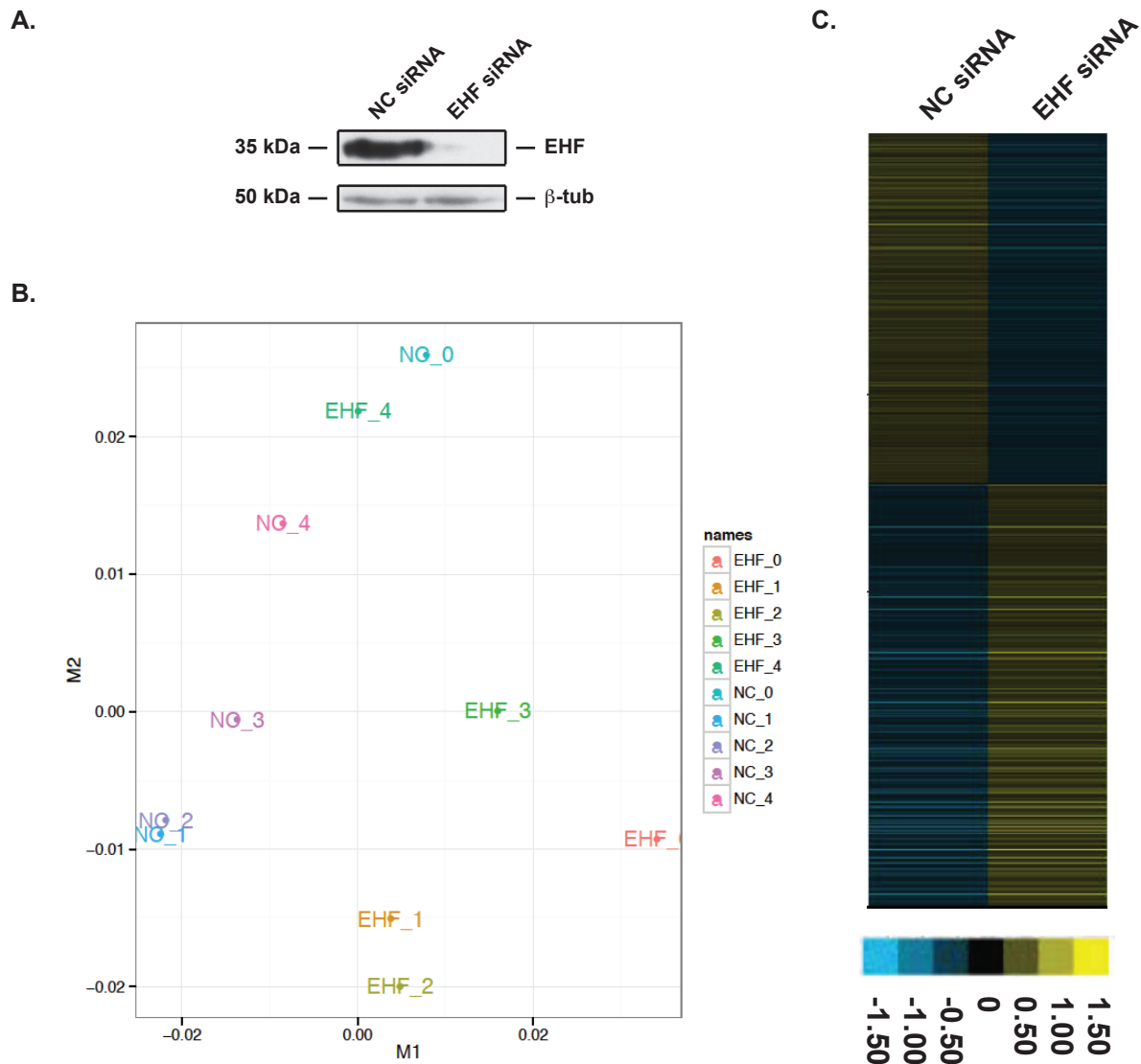
Figure 3.6.1 EHF depletion followed by RNA-seq identifies differentially expressed genes in HBE cells. A) siRNA depletion of EHF in primary HBE cells was measured by western blot with β-tubulin as a loading control. B) Multi-dimensional scaling (MDS) plot of individual samples based on FPKM values. Whole transcriptome expression values in FPKM at gene level were used to compare similarity among the samples through dimension reduction analysis. C) Relative expression of 1,145 genes that show a significant change in expression following EHF depletion in HBE cells (FDR < 0.05). Each row signifies a different gene. The scale is $\log_2$ fold change relative to the average gene expression value.

EHF binding sites are enriched at the promoters of genes, which implies that EHF exerts many of its effects through occupation of these regions. The loci that have EHF bound to their promoters and are differentially expressed following EHF depletion are the most likely to be direct targets of the protein. To identify these putative direct targets, the EHF ChIP-seq and RNA-seq following depletion data in HBE cells were intersected. In total, there were 71 EHF-activated and 78 EHF-repressed genes with promoter binding sites in HBE cells; EHF also occupied 2,290 promoters of genes that were not differentially expressed following EHF depletion in HBE cells. The median distances from EHF binding sites to activated and repressed TSSs were not different (84 and 82 bp, respectively). The activated targets included the transcription factor *SPDEF*, which is involved in airway goblet cell development, *S100A9,* and the integrin *ITGB6.* The repressed genes include the transcription factors *KLF5,* Activating Transcription Factor 3 (*ATF3*), and DNA Damage Inducible Transcript 3 (*DDIT3*)*, and JDP2.*

To determine a potential mechanism for the difference in EHF function at these sites, DNase-seq and histone modification tag counts in HBE cells were mapped to the promoter-bound EHF sites. The EHF binding sites were all hypersensitive to DNase I digestion, although the distribution of DNase-seq tag counts differed between the EHF activated/non-targeted genes and the repressed loci (Figure 3.6.2 A). The former 2 showed a narrow, bimodal peak of hypersensitivity, consistent with transcription factor binding to the center. At the binding sites near EHF-repressed genes, the DNase-seq tag counts were skewed downstream of the center of the site, and a bimodal pattern was not seen. This suggests that EHF may bind to the 5' end of the DHS. The EHF peaks were enriched for H3K4me3 and H3K27ac, regardless of whether they were near activated, repressed, or non-targeted genes (Figure 3.6.2 B).

Another mechanism that could explain differences in EHF function is interaction with other transcription factors. To identify motifs of potential co-factors, a *de novo* motif analysis was performed on the EHF binding sites near activated, repressed, and non-targeted genes (Figure 3.6.2 C). As expected, the epithelial specific Ets motif recognized by EHF was the most enriched motif at all sites. However, the secondary motifs differed between sites. At the sites near activated and non-targeted genes, motifs

similar to those recognized by KLF4 and Specificity Protein 1 (SP1) were the second most enriched. In

the EHF peaks near repressed genes, the motif recognized by ATF4 was over-represented.
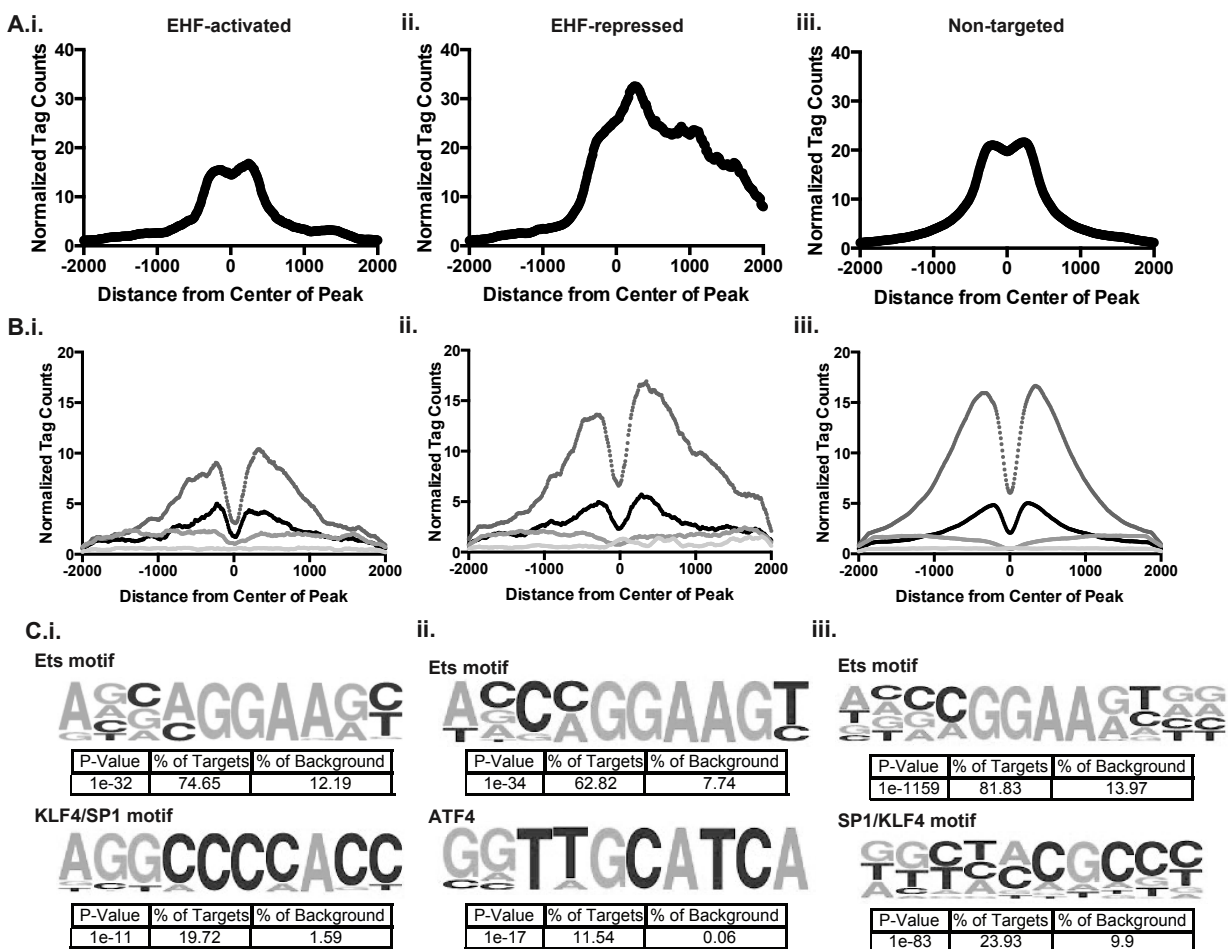
Figure 3.6.2 Promoters bound by EHF near EHF-activated and -repressed genes show different characteristics. A) Normalized tag counts from HBE DNase-seq plotted from the center of EHF binding sites found at the promoters of EHF- i) activated ii) repressed and iii) non-targeted genes. B) Normalized tag counts for H3K4me1, H3K4me3, and H3K27ac ChIP-seq in HBE cells, plotted from the center of EHF- i) activated ii) repressed and iii) non-targeted genes. Normalized tag counts from an input sample are included as a control. C) A *de novo* motif analysis within 200bp of the center of EHF ChIP-seq binding sites in HBE cells at the promoters of EHF- i) activated ii) repressed and iii) non-targeted genes. Motif logos constructed in the analysis are shown with P values derived by HOMER.

Although DEGs with EHF bound to the promoter represent the highest-confidence list of putative direct targets, considering only these loci would overlook those targets regulated through EHF binding to nearby *cis*-regulatory elements. To identify these targets of EHF in HBE cells, the binding sites identified in EHF ChIP-seq and DEGs from RNA-seq were intersected using Binding and Expression Target Analysis (BETA) (127), which assigns binding sites to genes based on a combination of distance to TSS and differential gene expression following TF modulation. EHF sites within 10kb of the TSS and DEGs identified with an FDR ≤ 0.1 were considered for this analysis. EHF was found to have significant activating and repressive function (Figure 3.6.3), consistent with gene expression changes in the positive and negative direction following EHF depletion. This analysis identified 190 putative direct targets that are repressed by EHF, and 188 that are activated. The BETA motif analysis found that the HOXB5 motif is differentially enriched in sites near EHF-repressed DEGs as compared with those that are activated (p=3.68e-5). The HNF1 homeobox A (HNF1A) and FOXA1 motifs were enriched in EHF sites near repressed genes as compared with non-regulated genes (p= 7.21e-7 and 4.63e-06, respectively by a T test). The nuclear factor I C (NFIC) and FOXK1 motifs were over-represented in EHF sites near activated DEGs are compared with those that were not differentially expressed (p= 1.43e-5 and 3.35e-5, respectively). The EHF-repressed genes included the following that encode transcription factors: CCAAT/enhancer binding protein (C/EBP) gamma and delta (*CEBPG and CEBPD*)*,* and *KLF5.* Genes activated by EHF include those involved in the immune response, including *SPDEF,* transforming growth factor beta 1 (*TGFB1*)*,* and *S100A9.*
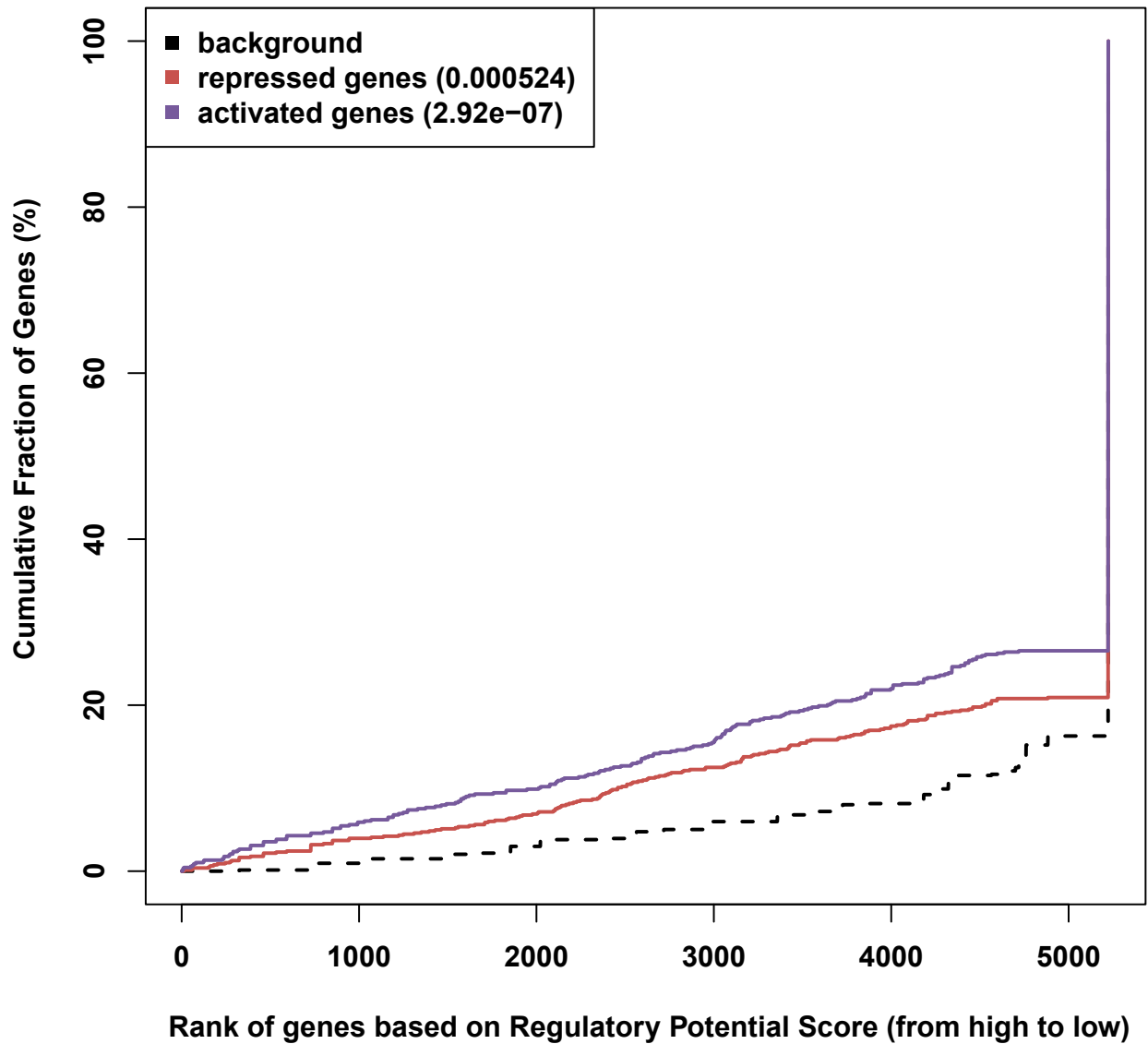
Figure 3.6.3 EHF has significant activating and repressive function. Activating and repressive function prediction of EHF in primary HBE cells as determine by BETA software. The red, purple, and dashed lines indicate the down-regulated, up-regulated, and non-differentially expressed genes, respectively. P values of activated and repressed genes versus non-regulated genes were determined by a Kolmogorov-Smirnov test.

An enrichment analysis was performed on up-regulated and down-regulated putative direct targets using MetaCore (Figure 3.6.4). EHF-activated genes were involved in GO processes including response to stress, positive regulation of NF-κB transcription factor activity, and response to wounding and diseases including viral and bacterial infections. EHF activates expression of *MMP2*, *PLAU,* transforming growth factor beta 1 (*TGFB1*) and TIMP metallopeptidase inhibitor 2 (*TIMP2*), which contribute to response to wounding. The infection processes include the loci tripartite motif-containing protein 5 (TRIM5), *TGFB1*, and major histocompatibility complex, class II, DM alpha (*HLA-DMA*). . EHF-repressed genes were enriched in gene ontologies involving regulation of metabolic and cellular processes. These gene ontologies include multiple transcription factors (*ATF4*, *KLF5, KLF4, CEBPG, CEBPD*), and histone deacetylase 1 (*HDAC1*)..
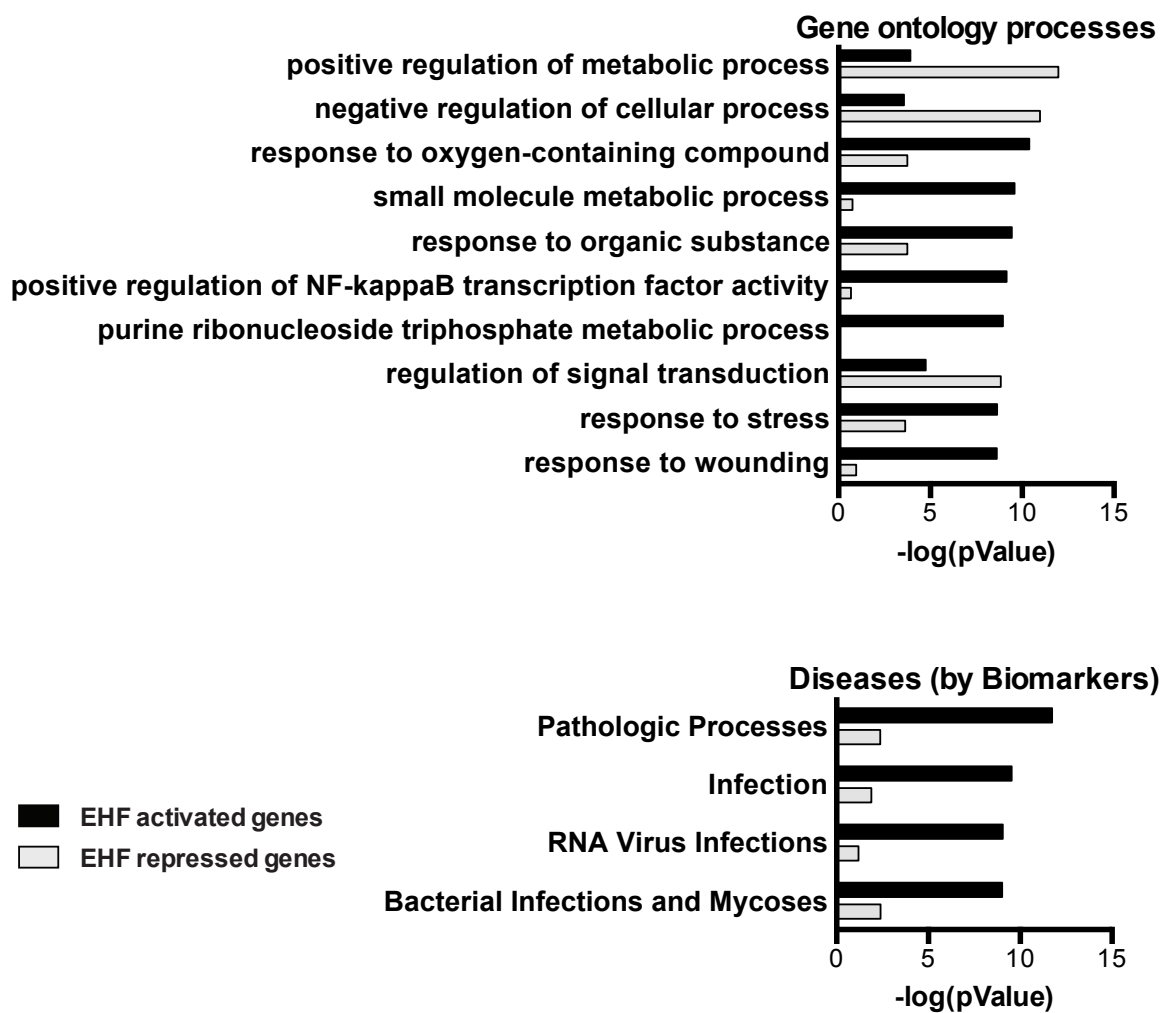
Figure 3.6.4 Predicted direct targets of EHF in HBE cells are involved in transcription regulation, the immune response, and response to wounding. A MetaCore enrichment analysis was performed on putative direct targets of EHF. The analysis included EHF-activated (black) and –repressed (grey) genes, and was divided into gene ontology processes (top) and diseases (bottom). Significance of enrichment is displayed as the –log (P Value). The legend for all panels is at the bottom left.

3.7 EHF and FOXA1 may co-regulate other transcription factors involved in lung organogenesis and epithelial differentiation

Lung organogenesis and epithelial differentiation are complex processes coordinated by a network of transcription factors. Multiple TFs within this network were identified as putative EHF targets in our HBE genome-wide data. These include genes with nearby peaks of EHF occupancy (Figure 3.7.1) and/or DEGs after EHF depletion, among which are *GATA6, HOPX, KLF5*, *SPDEF*, retinoic acid receptor beta (*RARB*)*, FOXA1* and *FOXA2.* Furthermore, EHF activated transcription of *SPDEF and FOXA1* in Calu-3 and A549 cells. To validate direct or indirect targets, EHF was depleted in primary HBE cell cultures derived from 4 donors. Expression of key TFs was determined using RT-qPCR (Figure 3.7.2 A). *CEBPG, HOPX, KLF5,* and *RARB* significantly increased in expression following EHF depletion, while *FOXA2* showed a non-significant increase. Levels of *FOXA1* and *SPDEF* decreased significantly. *GATA6* showed no change*.* Nearby sites of occupancy identified by ChIP-seq suggests that most of these genes are direct targets (Figure 3.7.1). To confirm this prediction, ≥ 4 biological (donor) replicates of EHF ChIP-qPCR were performed in primary HBE cultures (Figure 3.7.2 B). EHF binding was validated at intron 1 of *CEBPG,* intergenic sites 5' of *FOXA1, HOPX,* and *RARB,* and the promoter of *SPDEF.*

The sites of EHF occupancy near *CEBPG, HOPX, RARB*, and *SPDEF* also contain sequences with the Forkhead domain motif. Since EHF represses all but one of these genes (*SPDEF*), and the FOXA1 motif is enriched amongst EHF binding sites near repressed genes genome-wide, we next asked whether FOXA1 co-regulates gene expression with EHF at these loci. FOXA1 was depleted in HBE primary cultures from 3 donors using a specific siRNA and a non-targeting control, and gene expression was measured by RT-qPCR (Figure 3.7.2 C). *EHF, FOXA2, HOPX* and *KLF5* significantly increased in expression, while *SPDEF* decreased and *RARB* did not change. With the exception of *RARB* and *EHF,* all alterations paralleled those seen following EHF depletion. Furthermore, the EHF sites near *HOPX* and *SPDEF* were bound by FOXA1 as shown by 4 biological replicates of ChIP-qPCR in primary HBE cells from 2 donors (Figure 3.7.2 D). The FOXA1 enrichment near *RARB* was lower than the other sites, which may explain why FOXA1 knockdown does not alter *RARB* expression.
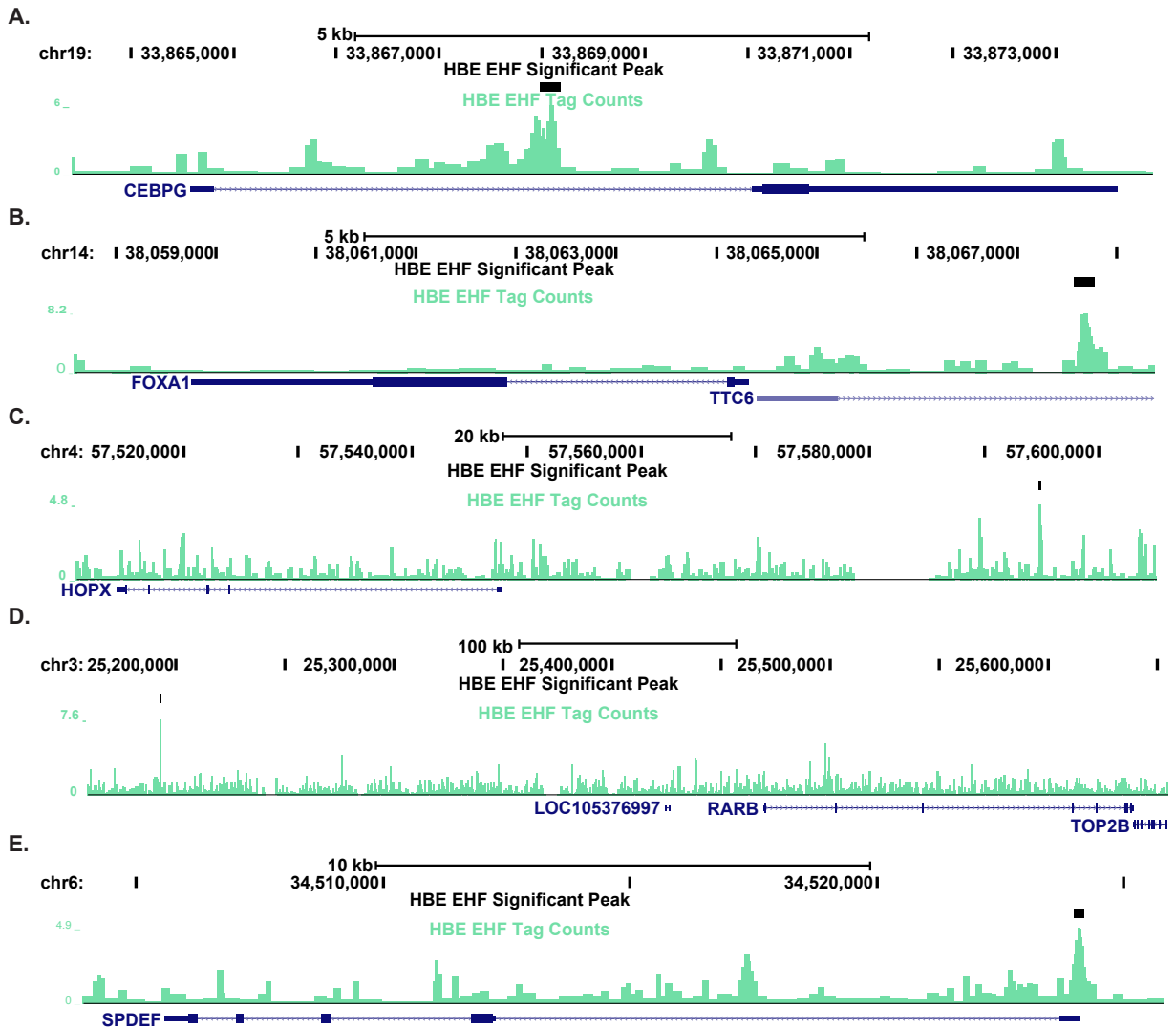
Figure 3.7.1 EHF ChIP-seq peaks in HBE cells are found near transcription factors involved in lung development. A-E) UCSC genome browser tracks displaying the significant EHF peaks identified by IDR in HBE cells (top, black), the EHF ChIP-seq tag counts (middle, green), and the gene (bottom, blue) for A) CEPBG, B) FOXA1, C) HOPX, D) RARB, and E) SPDEF.
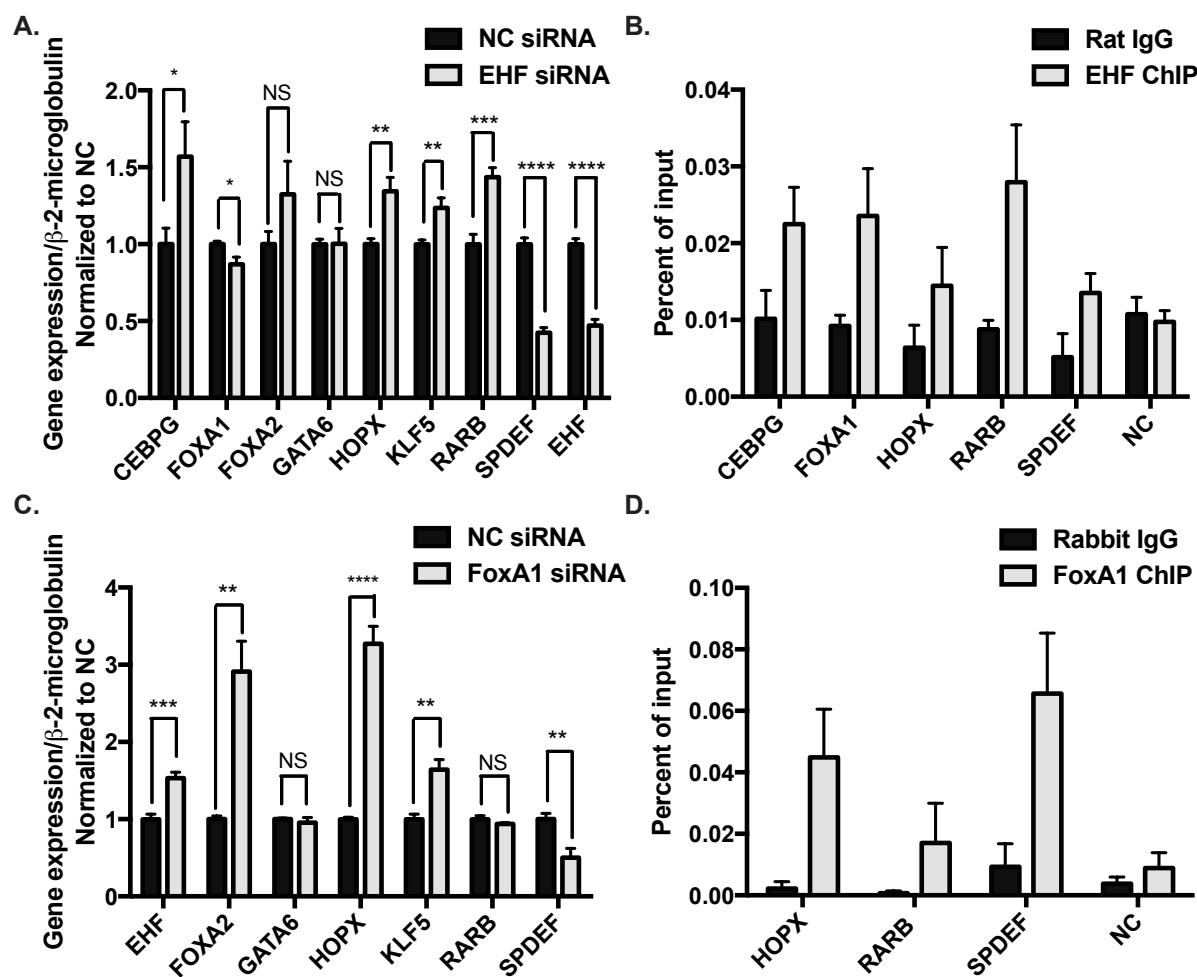
Figure 3.7.2 EHF and FOXA1 regulate other transcription factors involved in lung organogenesis and epithelial differentiation. A) EHF depletion by siRNA followed by RT-qPCR identified changes in transcription factor gene expression. All genes except *FOXA1* and *GATA6* showed changes consistent with RNA-seq. Shown is average of all experiments, with error bars representing standard error of the mean (SEM) (n=4). B) EHF ChIP-qPCR confirmed binding at select sites identified in ChIP-seq, including intron 1 of *CEBPG,* intergenic sites 5' of *FOXA1, HOPX,* and *RARB,* and the promoter of *SPDEF.* Shown is average of all experiments with SEM (n=4-5). C) FOXA1 siRNA depletion followed by RT-qPCR showed alterations in expression of transcription factors regulated by EHF, and *EHF* itself. All genes changed in the same direction as in EHF knockdown, except *RARB* and *EHF.* Average of all experiments with SEM (n=3). D) FOXA1 ChIP-qPCR confirmed binding to a subset of EHF sites. Average of all experiments with SEM (n=4). *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001 by an unpaired two-tailed Student's t test.

3.8 Treatment of primary HBE cells with IL-13 increases EHF expression

EHF represses many transcription factors involved in airway development, while activating

SPDEF. SPDEF contributes to goblet cell differentiation and mucus production in the airway, and is

induced by IL-13 (22,149). *MUC5AC,* a target of SPDEF, was increased in EHF-overexpressing A549

cells; Calu-3 and HBE cells grown on plastic do not express this gene. These data suggest that EHF may

participate in an IL-13-mediated cell differentiation pathway by activating SPDEF and repressing other

transcription factors. To determine whether IL-13 increases EHF expression, primary HBE cells were

polarized at air liquid interface (ALI) for 2 weeks, and then treated with IL-13 for 48 hours. Cell lysates

were prepared and EHF protein expression was measured by western blot and quantified by gel

densitometry (Figure 3.8). EHF expression was increased in IL-13 treated HBE cells as compared to

controls. There was significant variation in the induction of EHF by IL-13 between replicates, likely due to

inherent differences between the donor cells.
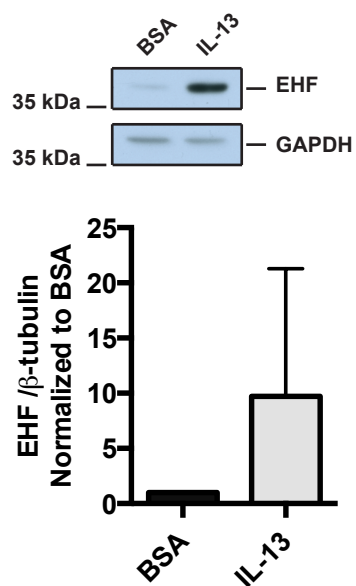
Figure 3.8 Treatment of polarized HBE cells with IL-13 increases EHF expression. EHF protein expression in BSA (carrier) or IL-13-treated polarized HBE cells as measured by western blot (top) and quantified using gel densitometry (bottom, average with SEM) (n=3).

3.9 EHF directly regulates *CFTR* expression

Our genome-wide EHF ChIP-seq dataset in HBE cells suggests that EHF may directly regulate *CFTR*. Peaks of EHF occupancy are seen at previously-characterized regulatory elements for *CFTR*, in DNase I hypersensitive sites -35 kb from the transcription start site (156), within intron 23 (legacy nomenclature) (155,202), and at 48.9kb 3' to the locus, in an intron for *CTTBP*2 (158) (Figure 3.7 A). CFTR protein expression in HBE cells grown on plastic is below the level detectable by western blot, so the effect of EHF on its expression was investigated in Calu-3 cells. EHF was depleted by siRNA in Calu-3 cells, and CFTR mRNA and protein expression was measured by RT-qPCR and western blot, respectively (Figure 3.7 B-C). EHF depletion significantly increased *CFTR* mRNA expression, which corresponded to an increase in protein levels, identifying EHF as a repressor of CFTR.
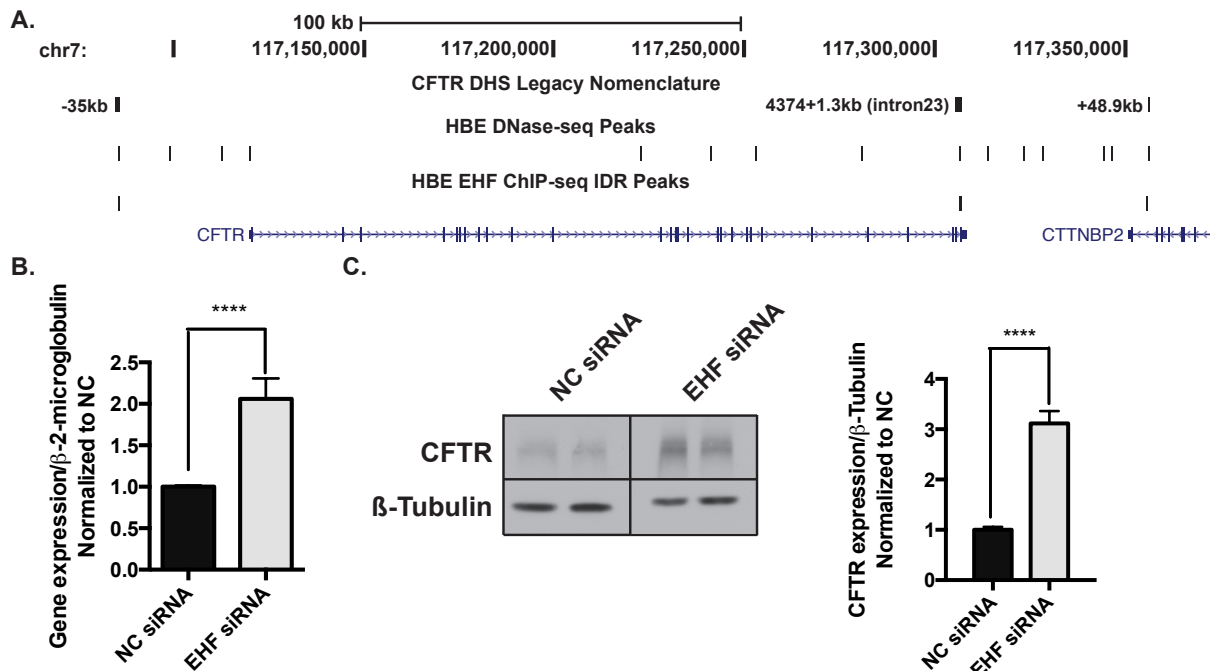
Figure 3.9 EHF represses CFTR expression. A) EHF ChIP-seq peaks at the *CFTR* locus. Legacy nomenclature for CFTR DHS is displayed. The HBE DHS track was downloaded from GEO (GSE74709) (158). B) EHF depletion in Calu-3 cells significantly increased *CFTR* mRNA expression, as measured by RT-qPCR. Shown is average of all experiments with SEM (n=3). C) CFTR protein expression significantly increased in EHF-depleted Calu-3 cells as measured by western blot (left) and quantified by gel densitometry (right, average with SEM) (n=4). ****$p<0.0001$ by an unpaired two-tailed Student's t test.

3.10 EHF depletion slows wound closure in lung epithelial cells

Our genome-wide data suggest that EHF regulates genes important for cell motility, which is essential for wound repair, including *CD44, ITGA2, MMP14, VIM, S100A8,* and *S100A9* (39,203-206). In HBE cells, genes activated by EHF are enriched in the gene ontology "response to wounding". Following epithelial injury, nearby epithelial cells migrate to repopulate the damaged area (37). This mechanism may be impaired in airway epithelial cells that carry a *CFTR* mutation, suggesting that it could play a role in CF lung pathology (48,49).

Wound closure can be measured *in vitro* using a scratch assay (203). To determine the effect of EHF depletion on wound repair in airway epithelial cells, this test was performed on 3 biological replicates of Calu-3 cells, 3 donor cultures of WT HBE cells, and 2 donor cultures of CF HBE cells. Cells were transfected with NC or EHF siRNA (Calu-3 cells when at 30-40% confluence, HBE cells at 70-80%) and grown to confluence. Seventy-two hours after transfection, the cell monolayer was scratched in a uniform manner using a comb. Scratches were then imaged at 3, 6, 9 and 12 hours after transfection in the Calu-3 cells (Figure 3.10.1 A), and at 0, 3, and 6 hours in the HBE cells (Figure 3.10.2 A). The Calu-3 cells had cellular debris at the scratch at 0 hours, impeding measurement of the scratch area. The HBE cells lacked this obstruction, and closed more rapidly than the Calu-3 cells, necessitating earlier measurements. The wound area was measured at each time point using ImageJ software. In Calu-3 cells at 3 hours, the average area was 0.60 mm^2 for NC cells and 0.53 mm^2 for EHF siRNA-treated cells; this difference was not significant (p = 0.25). The rate of wound closure in mm/hour was determined; the EHF siRNA-treated cells migrated at 67% the rate of NC treated cells from 3-6 hours (p < 0.01) and at 71% from 6-9 hours (p < 0.001) (Figure 3.10.1 B). This difference was no longer apparent at the 9-12 hour interval. The cells were grown for 27 hours post-scratch to ensure wound closure and then lysed to measure EHF depletion. Western blots of cell lysates probed for EHF with β-tubulin as a loading control showed that EHF depletion was maintained throughout the wound repair assay; EHF protein levels were 15% of those in negative control treated cells at 27 hours post-scratch (Figure 3.10.1 C).
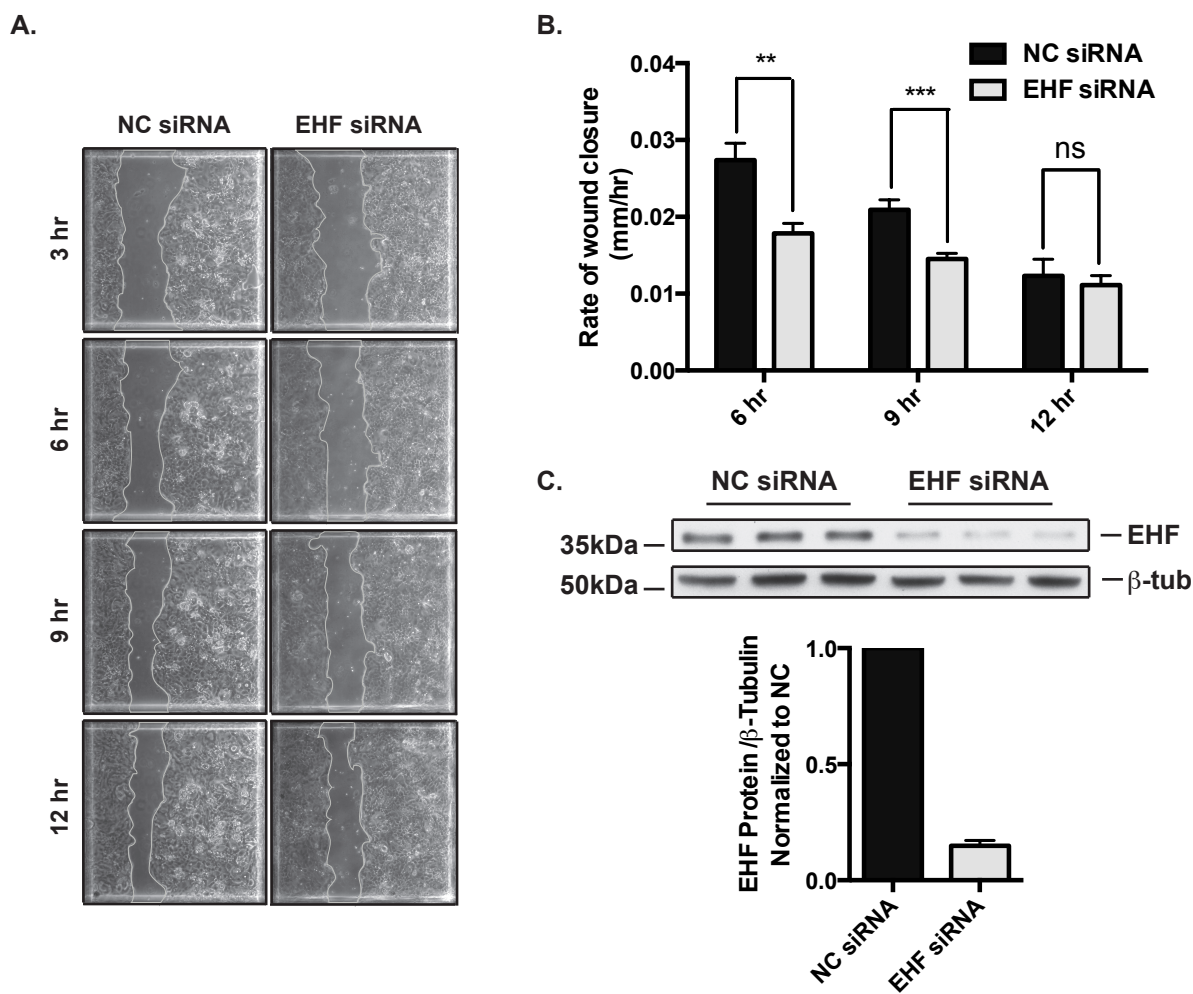
Figure 3.10.1 Depletion of EHF slows wound closure in Calu-3 cells. A) Images of NC and EHF siRNA treated cells at 3, 6, 9 and 12 hours after wounding. The outline of the wound is traced with a gray line. B) Depletion of EHF significantly reduced the rate of wound closure at 6 and 9 hours after wounding (n = 3). This difference was no longer evident 12 hours after wounding. **p<0.01, ***p<0.001 by an unpaired two-tailed Student's t test. C) siRNA knockdown of EHF was measured by western blot in lysates taken 27 hours after wounding (top) and the signal was quantified using gel densitometry (bottom). EHF depletion was maintained throughout the experiment.

In HBE cells, wound closure was measured as percentage of initial injury size. The initial wound size in the WT cells was 0.284 mm^2 for NC and 0.279 mm^2 for EHF cells; the difference was not significant (p = 0.79). The difference between the starting areas for NC CF cells (0.295 mm^2) and EHF siRNA treated CF cells (0.280 mm^2) was also not significant (p = 0.59). In the WT HBE cells at the 3 hour time point, the scratches in the EHF depleted cells remained significantly larger than those in the NC cells (p < 0.01) (Figure 3.10.2 B). There was a similar trend at 6 hours. In the CF cells, EHF knockdown significantly delayed the reduction in wound size at both 3 and 6 hours (Figure 3.10.2 C). This corresponds to a slowing of wound closure in the EHF-depleted cells compared to NC in both the WT and CF cells. Twenty-four hours after the scratch, cells were lysed to confirm EHF knockdown. EHF expression relative to β-tubulin was measured by western blot and quantitated using gel densitometry (Figure 3.10.2 D-E). Depletion of EHF was maintained through the scratch assay.
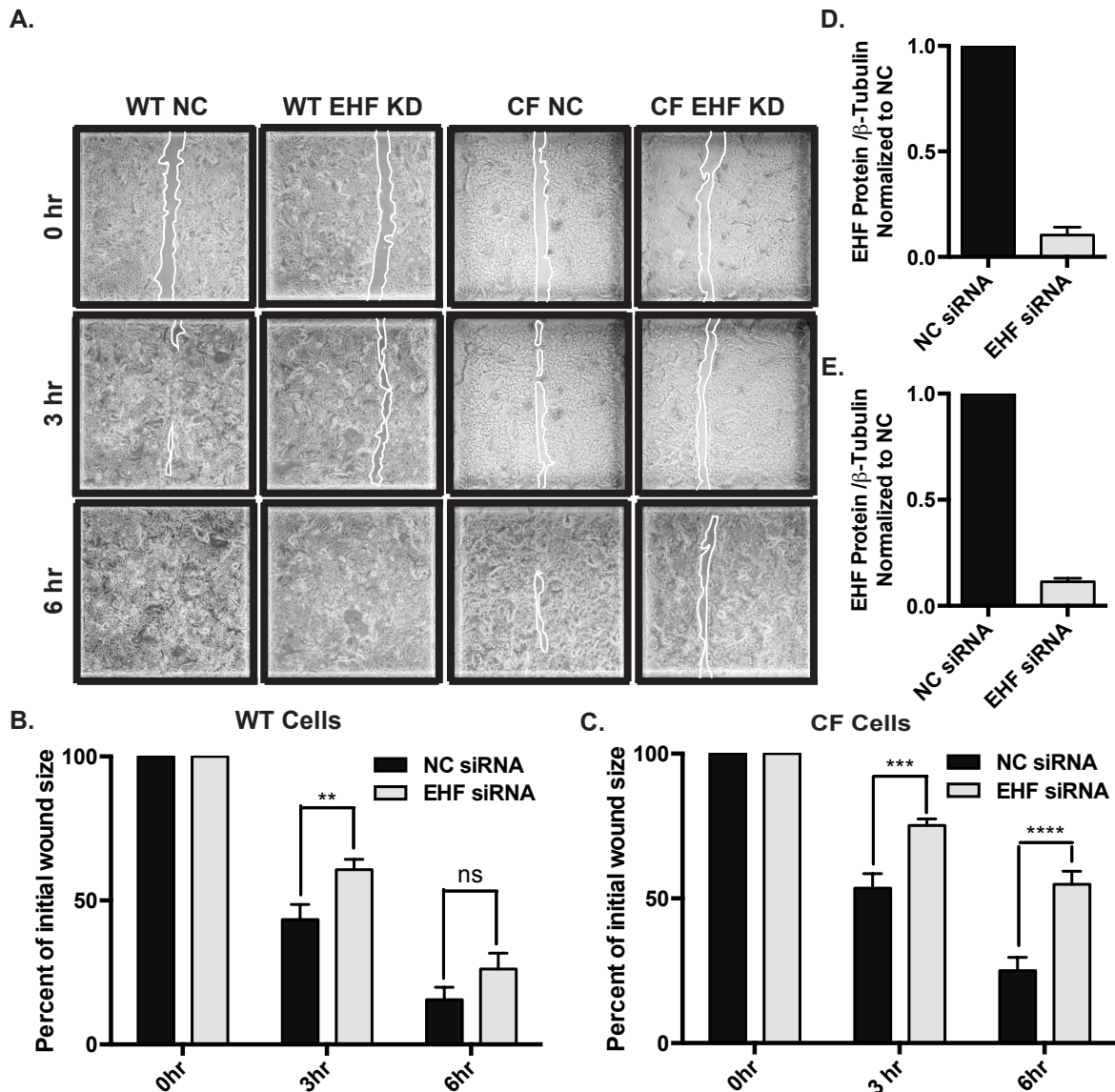
Figure 3.10.2 EHF depletion slows wound closure in wild type (WT) and cystic fibrosis (CF) primary HBE cells. A) Images of NC and EHF siRNA-treated (KD) WT and CF HBE cells at 0, 3, and 6 hours after initial wounding. Outlines of wounds are traced with a white line. B-C) Percent of initial wound size was measured at all time points in B) WT (n=3) and C) CF (n=3) cells. In both cell types, wound area remained larger in the EHF-depleted cells compared to the controls. **p<0.01, ***p<0.001, ****p<0.0001 by an unpaired two-tailed Student's t test. D-E) siRNA depletion of EHF was measured by western blot in lysates taken 24 hours after wounding, and the signal was quantified using gel densitometry. EHF depletion was maintained throughout the experiment in D) WT and E) CF cells. All graphs show the average with SEM.

3.11 Depletion of EHF does not alter cell proliferation in Calu-3 cells

Scratch assays measure wound closure, a process that involves epithelial cell migration and cell proliferation. Cell proliferation was identified as an enriched GO term among DEGs following EHF overexpression in A549 cells, suggesting EHF may regulate this process. To test whether differences in cell proliferation contribute to slowing of wound closure, the effect of EHF depletion on proliferation was measured in a growth assay. The assay utilizes MTS (3-[4,5-dimethylthiazol-2-yl]-5-[3-carboxymethoxyphenyl]-2-[4-sulfophenyl]-2H-tetrazolium, inner salt) to quantify live cells. Cells metabolize MTS to a formazan product, the absorbance of which can be measured in a microplate reader. The absorbance is proportional to the number of live cells, from which cell proliferation can be calculated. Cells were seeded in a 96-well plate, then transfected with NC or EHF siRNAs. Live cell numbers were measured at 0, 24, 48 and 72 hours post-transfection (Figure 3.11). There was no significant difference in live cell density between NC and EHF siRNA-treated cells at 24 ($p = 0.27$), 48 ($p = 0.49$), or 72 hours ($p = 0.87$) in 3 biological replicates, suggesting no variance in cell proliferation.
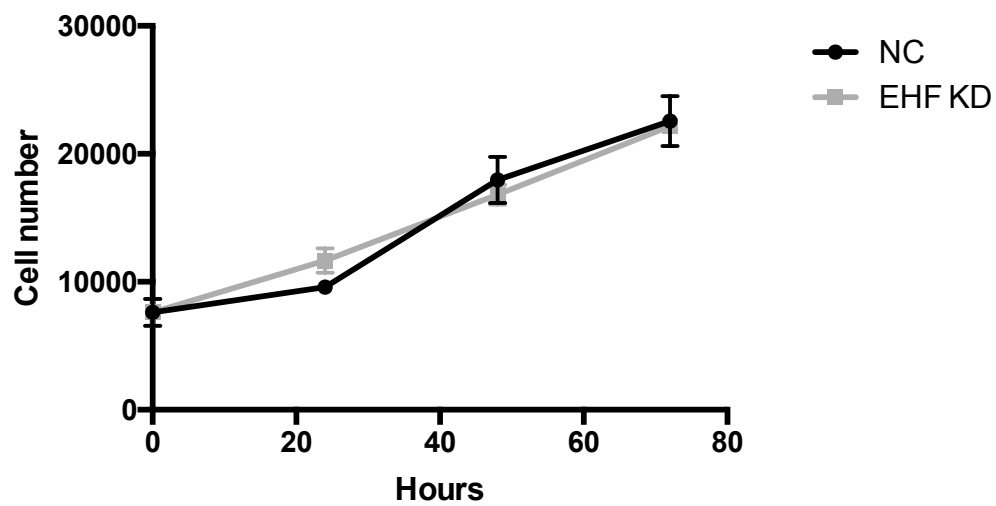
Figure 3.11 Loss of EHF expression does not alter cell proliferation. Calu-3 cells were treated with NC or EHF siRNA. Cell density was measured using an MTS assay from 0 to 72 hours post-transfection. There was no difference in cell density at 24 (p=0.27), 48 (p=0.49), or 72 hours (p=0.87) (n=3).

3.12 EHF regulates expression of a neutrophil chemokines

3.12.1 EHF binds to sites at chr4q13 near genes encoding neutrophil chemokines in HBE cells

Targets of EHF are enriched for genes involved in infection in HBE cells and for the GO process "response to wounding", which encompasses multiple loci involved in inflammation, in Calu-3, HBE and A549 cells. In HBE cells, 2 EHF binding sites were observed at the chr4q13.3 region encompassing the genes for the neutrophil chemokines *CXCL1, CXCL6,* and *IL8*, one 5' to *IL8* and the other 3' to *CXCL1* (Figure 3.12.1). EHF depletion followed by RNA-seq and RT-qPCR in Calu-3 cells identified increases in *CXCL1* and *IL8* expression, suggesting EHF regulates both transcripts in this cell type. Airway epithelial cells release these pro-inflammatory mediators, which are important for neutrophil survival and chemotaxis (28,207-209). Neutrophils mediate the acute inflammatory response and are found early in the lungs of cystic fibrosis patients (62,63), which makes these chemokines important for lung pathology. The sites identified by EHF ChIP-seq were confirmed by ChIP-qPCR in HBE cell chromatin from 4 donors (Figure 3.12.1).
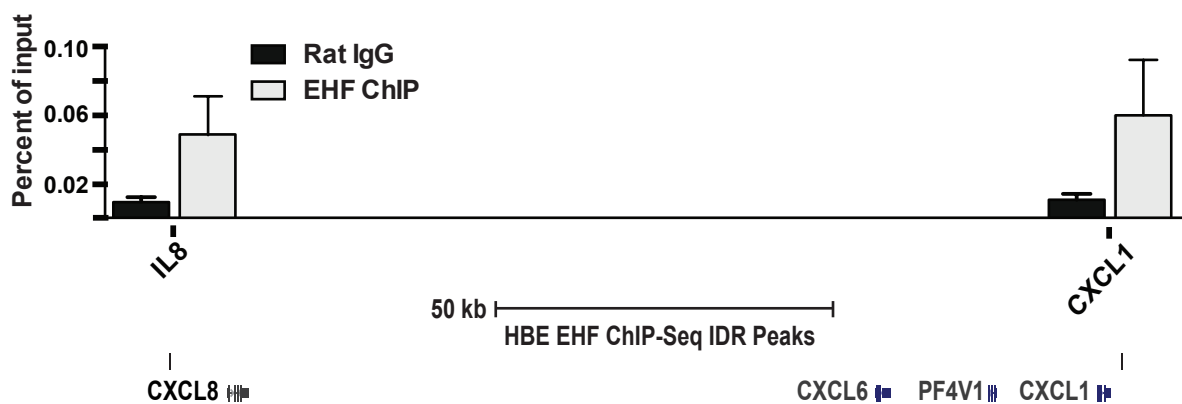
Figure 3.12.1 EHF binds to sites at chr4q13.3 near genes encoding neutrophil chemokines. EHF ChIP-seq in HBE cells identified binding sites on chr4q13.3, one 5' of *IL8* and the other 3' to *CXCL1*. EHF ChIP-qPCR in HBE cells confirmed binding to these sites. Shown is the average of all ChIP-qPCR experiments with SEM (n=4) (top) and a UCSC Genome Browser representation of the locus showing the significant EHF ChIP-seq peaks in HBE cells (bottom).

3.12.2 Calu-3 cells are responsive to interleukin-17 (IL-17a) and lipopolysaccharide (LPS)

The bronchial epithelium releases *CXCL1, CXCL6,* and *IL8* in response to immune stimuli including IL-17a (28,31) and lipopolysaccharide (LPS) (34,35). Th17 cells release IL-17a, which induces neutrophil infiltration in the lung (30). LPS is a component of the outer membrane of Gram-negative bacteria, including *P. aeruginosa* (32)*.* The role of EHF in regulating these immune mediators was studied in Calu-3 cells, which express *CXCL1, CXCL6* and *IL8*.

Initially, we sought to investigate whether Calu-3 cells grown on plastic were responsive to IL-17a and determine the optimal dose for treatment. Cells were serum starved for 24 hours, then incubated with BSA (carrier), or 1 ng/mL, 10 ng/mL, or 100 ng/mL IL-17a for 24 hours. RNA was isolated from treated cells, and expression of *CXCL1*, *CXCL6* and *IL8* was measured using RT-qPCR (Figure 3.12.2 A). The levels of *CXCL1* and *CXCL6* increased in a dose-dependent manner in response to IL-17a. At 10ng/mL of IL-17a, *CXCL1* and *CXCL6* were induced 2.5-fold, while *IL8* increased 1.5-fold. At 100ng/mL IL-17a, *IL8* expression was not elevated to the same extent, so 10ng/mL was selected for further experiments. The dose-dependent changes in gene expression also confirmed that the Calu-3 cells were responsive to IL-17a.

The response of Calu-3 cells to LPS and ideal incubation time for treatment was also determined. Calu-3 cells were serum starved for 24 hours, and then treated with PBS (carrier) or 1µg/mL *P. aeruginosa* LPS for 4, 8, or 24 hours. Expression of *CXCL1, IL8* and *EHF* was measured by RT-qPCR (Figure 3.12.2 B). Treatment for 4 hours increased expression of *CXCL1* and *IL8* 1.5-fold, whereas longer incubation times did not induce the same increase; therefore LPS treatment for 4 hours was used for subsequent experiments. These changes in mRNA expression suggest that the Calu-3 cells are responsive to LPS from *P. aeruginosa.*
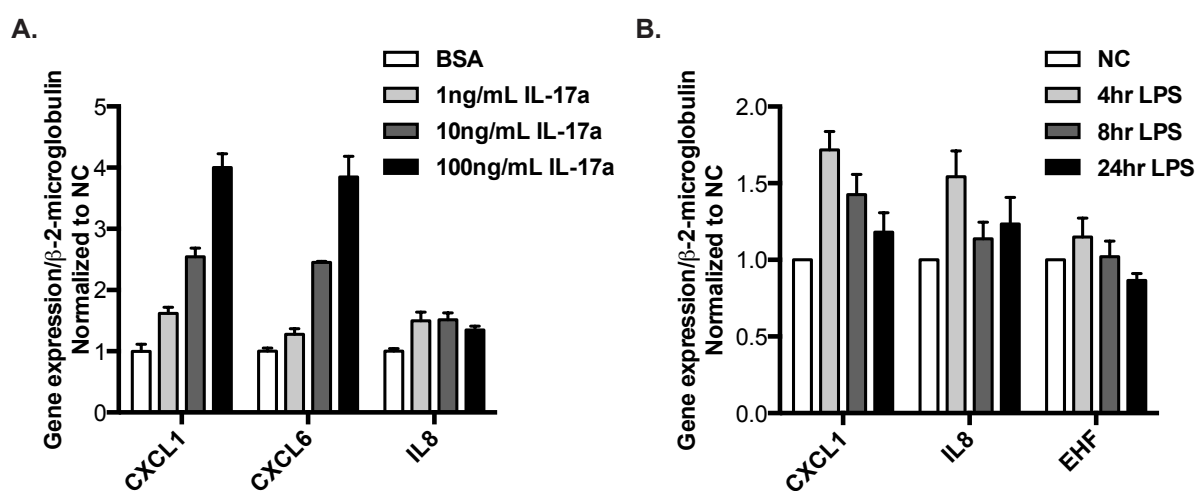
Figure 3.12.2 Calu-3 cells are responsive to interleukin-17 (IL-17a) and lipopolysaccharide (LPS).
A) Serum-starved Calu-3 cells were treated with BSA (carrier), or 1 ng/mL, 10 ng/mL, or 100 ng/mL IL-17a for 24 hours. Gene expression was measured using RT-qPCR. Shown is average of technical replicates from one experiment with SD. B) Serum-starved Calu-3 cells were treated with PBS (carrier) or 1µg/mL *P. aeruginosa* LPS for 4, 8, or 24 hours and gene expression was measured by RT-qPCR. Average with SEM is shown (n=3).

3.12.3 EHF depletion alters expression of neutrophil chemokines at basal levels and in response to inflammatory mediators

To confirm whether EHF regulates secretion of *CXCL1, CXCL6,* and *IL8*, both at a basal level and in an inflammatory milieu, EHF was depleted in Calu-3 cells using a specific siRNA, with a non-targeting siRNA as control. Cells were then subject to treatment with carrier, IL-17a or LPS from *P. aeruginosa* and RNA was extracted. Chemokine expression was quantified using RT-qPCR (Figure 3.12.3 A-B). *CXCL6* expression was reduced upon EHF depletion in normal conditions and following LPS and IL-17a treatment. EHF knockdown led to an increase in *IL8* expression, both at basal levels and following LPS treatment. *CXCL1* was not significantly regulated by EHF in either condition. This may be due to the time point at which the RNA was isolated in this experiment, which was longer after the transfection than in the studies that identified *CXCL1* as being repressed by EHF.

To determine if the changes in *CXCL6* and *IL8* transcript abundance altered cytokine secretion, colorimetric sandwich ELISAs were utilized to quantify their release into conditioned media. EHF-depleted or NC cells were allowed to condition serum-free media for 24 hours. EHF depletion decreased CXCL6 secretion (Figure 3.12.3 C), but did not significantly alter IL8 secretion (Figure 3.12.3 D).
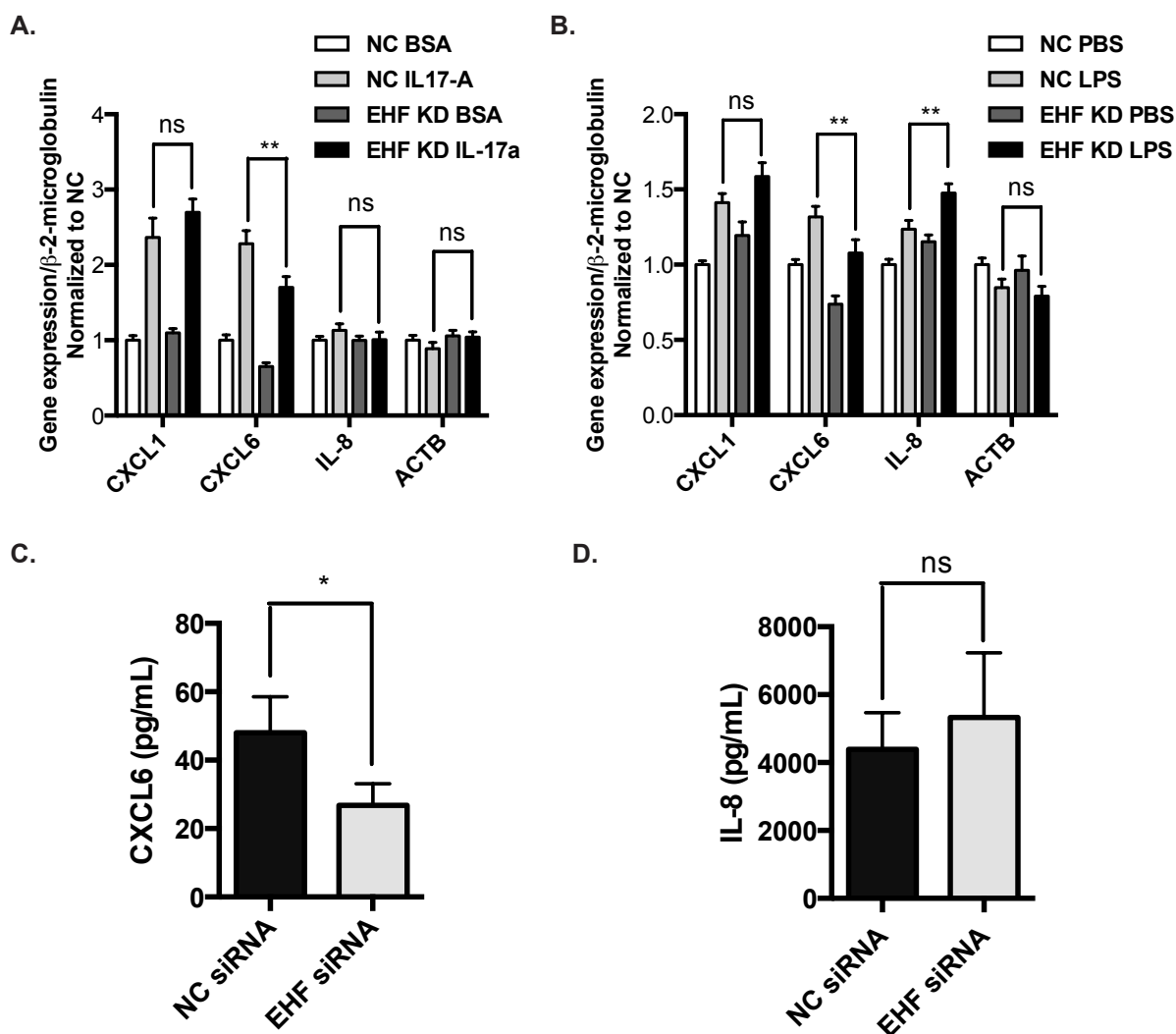
Figure 3.12.3 EHF depletion alters expression of neutrophil chemokines at basal levels and in response to inflammatory mediators. A-B) Calu-3 cells were subject to EHF siRNA depletion, then treated with A) carrier (BSA) or IL-17a or B) carrier (PBS) or LPS. Gene expression was measured by RT-qPCR. β-actin (ACTB) was included as a negative control. **p<0.01 by a 2-way ANOVA plus multiple comparisons test (n=3). C-D) Secretion of CXCL6 (C) and IL-8 (D) into media conditioned by NC- and EHF-depleted Calu-3 cells, quantified by colorimetric sandwich ELISA (n=3). *P<0.05 by a paired two-tailed Student's t test.

4. DISCUSSION

Understanding regulation of gene expression in lung epithelial cells is critical to determining how they develop, differentiate and respond to the environment. Because of the location of this tissue at the interface between the external and internal atmospheres, its response to a wide variety of signals and stressors is especially important to function of the airway, and abnormal reactions are often central to lung disease. Transcription factors (TFs) are responsible for integrating signals and coordinating cellular processes for normal airway function. Several transcription factors are known to have an important role in lung epithelial development and function, including FOXA1/A2 (143), SOX family members (210) and NKX2-1 (210), among others. Here, we characterize the role of EHF in coordinating gene expression profiles in lung epithelial cells.

4.1 EHF binding sites are over-represented at promoters and enriched for characteristic chromatin modifications

TFs function by binding to DNA at regulatory elements to influence expression of genes through multiple mechanisms. Genome-wide binding profiles generated by ChIP-seq are critical to characterizing the role of a factor in a specific tissue; however, no such study had been performed for EHF in the airway epithelium. We completed EHF ChIP-seq in Calu-3 lung adenocarcinoma cells and primary human bronchial epithelial (HBE) cells to identify its binding sites across the genome. EHF occupancy is over-represented at the promoters of genes, defined as 1,000 bp upstream of the transcription start site (TSS), bidirectional promoters (2,500 bp upstream of the TSS), and 5' UTRs in both cell types. The percentage of EHF binding peaks found at these genomic sites is slightly increased in the HBE cells as compared with the Calu-3 cells, but the overall pattern is similar. This suggests that it exerts a major effect on gene regulation through binding at promoters, and this influence may be greater in the HBE cells. There are also EHF binding sites found at putative *cis*-regulatory elements in introns and intergenic regions.

The intersection of EHF binding data with DNase-seq and ChIP-seq for histone modifications gives an overall indication of the function of EHF with respect to gene regulation. In Calu-3 cells, EHF binding sites show increased DNase hypersensitivity, which is not seen in genomic sites that contain the Ets motif but where EHF does not bind. This suggests that EHF is specifically bound at open chromatin sites, consistent with binding to regulatory elements. A previous study found that chromatin accessibility best predicts cell-type-specific binding of transcription factors (85), and it is likely a factor that limits the ability of EHF to occupy sites containing its motif.

Another important characteristic of EHF binding sites is the presence of active histone modifications, which reflects EHF occupancy at regions of open chromatin. In both Calu-3 cells and HBE cells, EHF sites are enriched for the H3K4me3 histone modification, which is found at poised and active promoters (90). In HBE cells, the EHF occupancy regions containing H3K4me3 also have H3K27ac, which marks active regulatory elements (83). Promoter-binding EHF sites are especially enriched for H3K4me3 and H3K27ac as compared with intronic or intergenic sites. These patterns suggest that EHF binds to active promoters, which is consistent with a role for increasing expression of genes. There is less enrichment of H3K4me1 at EHF sites, which marks distal enhancers (90), even though there are more sites marked by this modification. Intergenic EHF binding sites are enriched for H3K4me1 and depleted for H3K4me3. The H3K4me1 sites are sub-divided into those with high or low H3K27ac, which suggests that only a subset is active. The regions with less H3K27ac may represent poised enhancers (95). Determining the differences between these poised and active regions may shed light on the mechanism of EHF activity at *cis*-regulatory elements.

Of interest, H3K27ac enrichment is much higher at EHF sites in Calu-3 cells than in primary HBE cells. EHF has a known role in carcinogenesis (171-174), which is associated with alterations of the chromatin landscape (211). The Calu-3 cell line is derived from a lung adenocarcinoma, so the differences in H3K27ac may be reflective of a different epigenetic profile in this cell type as compared with normal lung epithelial cells.

Multiple studies have found that transcription factor binding at regulatory elements is cell-type specific (85-87). We have generated two high-confidence ChIP-seq data sets in Calu-3 cells and primary HBE cells. Although Calu-3 cells are derived from a lung adenocarcinoma, they show similarities to serous gland epithelial cells (190). The primary HBE cells are comprised of a mixture of epithelial cell types (212). There is a significant difference in the quantity of peaks identified in the two cell types, suggesting that the immunoprecipitation was more successful in the HBE cells. Only 68% of the peaks in the smaller Calu-3 set are present in the HBE cells. However, when the subsets of binding sites found at promoters are intersected, 84% of the Calu-3 sites are found in the HBE cells. This supports a previous study that found promoters are more similar across cell types than enhancers (93). Even in cell cultures as similar as Calu-3 cells and primary HBE, there are significant differences in TF binding, which stresses how critical the appropriate selection of cell culture models is to obtain results that are applicable to the tissue or disease of interest.

4.2 Co-regulators of EHF may control activating versus repressive transcriptional activity of the factor

EHF bound to the promoters of loci for which expression was increased and decreased following its depletion in HBE cells. A role for EHF in both activating and repressing gene expression was also confirmed by the BETA output, which showed significant activity in both directions through binding at promoters and nearby *cis*-acting enhancers. This is consistent with previous studies at specific promoters that show that EHF has both active and repressive action, depending on the cellular context (160,171,172,174). The mechanism whereby EHF coordinates its transcriptional targets to achieve this is of interest; our data show that EHF occupies the promoter of some DEGs identified after EHF modulation, but many lack a promoter binding site. The DEGs with EHF binding sites at the promoter are the most likely to be direct targets. Therefore, the characteristics of these sites were investigated with the aim of identifying potential mechanisms of EHF-mediated activation and repression. The distance between the EHF site and the TSS was not significantly different at repressed or activated promoters, which suggests

that the location of its occupancy within the promoter does not determine its transcriptional activity. The histone modification distribution at EHF-bound promoters near activated, repressed, and non-targeted genes was also similar, which suggests that they do not contribute to activating versus repressive action. However, the chromatin structure may still be involved, as the pattern of DNase-seq tag counts at promoters near repressed genes differed from the bimodal distribution around the center of EHF sites near activated and non-targeted genes. The peak of DNase hypersensitivity at the former regions skewed 3' of the center of the EHF peak, suggesting that EHF is bound at the 5' end of these regions, and other TFs bound further downstream could be contributing to the local chromatin landscape.

The most striking contrast between the EHF binding sites near activated- and repressed DEGs was the difference in enriched TF binding motifs. Although the epithelial-specific Ets motif was the most over-represented in both classes of peaks, the most significant secondary motif differed between the two. The motif for ATF4 was significantly enriched at EHF sites found near promoters of genes showing increased expression following EHF depletion, but not those that decreased. Because it is only present in promoters of EHF-repressed genes, it is a prime candidate for a co-regulator that impacts the transcriptional mechanism of EHF. ATF4 is a transcriptional enhancer induced by endoplasmic reticulum (ER) stress; it increases expression of DDIT3, leading to apoptosis. The expression of ATF4 is also induced by LPS in the lung (213), and it increases the levels of genes elevated in patients with COPD (214). EHF depletion elevates expression of DDIT3; this suggests that EHF may bind to ATF4-induced promoters and inhibit its ability to activate transcription. By decreasing another factor's ability to enhance transcription, EHF would appear to repress gene expression. The mechanism of this repression is of interest. Because the histone modifications are similar at EHF sites near activated and repressed genes, this makes EHF recruitment of histone deacetylases or methyltransferases unlikely. Co-occupancy of a promoter by EHF and ATF4 could lead to recruitment of co-repressors. These proteins may be involved in chromatin remodeling, have DNA methylase activity or directly interfere with co-activators of transcription (215). The state of DNA methylation and binding of co-activators such as p300 can be determined in NC and EHF depleted cells to identify a mechanism by which EHF inhibits ATF4-mediated induction of

transcription. Furthermore, occupancy of ATF4 at these sites can be compared in control and EHF knockdown cells to determine whether binding of EHF directly affects recruitment of ATF4 to these regions.

In EHF-activated promoters, the motif recognized by KLF4 and SP1 was significantly enriched and found in 19.72% of sites; in EHF sites at repressed promoters, it was only found in 12.82% of sites and less significantly over-represented. SP1 and Ets transcription factors, including EHF, cooperatively increase gene expression through binding to proximal promoters (216-218), suggesting a synergistic role for these two families of transcription factors through recruitment of the coactivators CREB binding protein (CBP) and p300 (216). KLF4 has a described role in the lung epithelium; its expression is induced following exposure to *Streptococcus pneumonia*, and it inhibits downstream expression of *IL8* through binding to the promoter (219). SP1 and KLF4 belong to the same superfamily of transcription factors characterized by a C-terminal zinc-finger DNA binding, however SP1 binds more strongly to GC-boxes than CACCC-boxes, with KLF4 showing the reverse preference (220). The two TFs have opposite effects on regulating expression of *SLC4A7* (221), which implicates them in antagonistic activity. The motif in the EHF peaks found near activated promoters was most similar to that recognized by KLF4; in the sites near non-targeted loci, the motif coincided more closely with that of SP1. This may suggest that differential occupancy of KLF4 and SP1 (or other factors in the transcription factor superfamily with similar binding preferences) may influence whether EHF is able to activate gene expression at a promoter. ChIP-seq for SP1 and KLF4, both of which are expressed in HBE cells, could confirm whether preferential binding of KLF4 to EHF occupancy sites increases transcriptional activation.

The presence of EHF ChIP-seq peaks in intronic and intergenic regions suggests that some DEGs may be controlled through distal EHF-regulated *cis*-elements. It is also likely that many genes are indirect targets, since EHF regulates other transcription factors. Together, these data suggest that EHF binds to active promoters and *cis*-regulatory elements in open chromatin sites to orchestrate its transcriptional program.

4.3 AP-1 is a potential co-regulator of EHF function

4.3.1 AP-1 and EHF co-occupy many sites genome-wide

EHF binds to the relatively common Ets motif, and it can increase or decrease gene expression. Therefore its interactions with other TFs are likely critical to specify binding location and mechanism of action. The possibility of ATF4 and SP1/KLF4 influencing EHF mechanism of action specifically at promoters was discussed in the previous section. A motif analysis was performed on all significant EHF binding sites in both Calu-3 and primary HBE cells showed that the most enriched secondary motif is for activator protein 1 (AP-1). The AP-1 transcription factor complex, which is made up of hetero- and homo-dimers of Jun, Fos and other family members, regulates genes involved in differentiation, cell injury, inflammation and apoptosis in the lung (181,182). EHF represses genes that contain both EHF and AP-1 motifs at the promoter (160), which suggests that the two co-regulate gene expression.

To confirm that EHF and AP-1 co-localize genome-wide, as is suggested by the presence of the AP-1 motif in EHF binding sites, ChIP-seq for AP-1 members was performed in Calu-3 cells. In the lung, there is high expression of Jun family members (181), so c-Jun and JunD were investigated. These data were intersected with the Calu-3 EHF ChIP-seq, yielding significant overlap between binding sites for EHF and both AP-1 proteins. The majority of co-occupied sites were bound by all 3 proteins (EHF, c-Jun and JunD). Furthermore, when c-Jun and JunD ChIP-seq tag counts (representing relative proportions of binding at base pair resolution) were mapped relative to the center of EHF peaks, there was a similar binding pattern for both proteins, which suggests that EHF co-localizes with a hetero-dimer of c-Jun and JunD. To confirm the co-occupancy of EHF, c-Jun and JunD, ChIP-reChIP (222) should be performed at specific sites of interest.

The EHF/AP-1 overlapping binding sites were more enriched for the H3K27ac histone modification than total EHF binding sites in Calu-3 cells, implying that these regions represent more active regulatory elements. They were found near genes involved in multiple pathways important for lung

epithelial function in disease, including differentiation and wound repair. These pathways are consistent with known roles for AP-1 in the lung, where the complex is activated in response to toxic stimuli to control expression of differentiation markers, extracellular matrix (ECM) proteins and growth factors (181). Of interest to wound repair, there were EHF/AP-1-bound sites at the *ITGA2* promoter and a putative enhancer 5' of the gene encoding syndecan 1 (*SDC1*). Syndecan 1 alters conformation of the $\alpha_2\beta_1$ integrin in injured airway epithelial cells, thereby coordinating cell migration and spreading during wound repair (41). Through co-regulating multiple genes involved in the re-epithelialization process, EHF and AP-1 could control the ability of the lung to repair damaged areas. Further investigation of these loci through double knockdown of EHF and AP-1 proteins could elucidate potential targets to modulate inflammation and wound repair in the airway.

4.3.2 *In vitro* studies do not support a direct interaction between EHF and Jun proteins

The significant overlap of EHF binding sites with those of c-Jun and JunD indicates that the transcription factors may interact to regulate gene expression. The ETS domain of multiple transcription factors was shown to directly interact with Jun proteins (164); the same domain in EHF may interact with c-Jun or JunD. Two *in vitro* methods were used to probe for direct interactions. In a Co-IP experiment, Flag-EHF was immunoprecipitated, and the enriched complexes were probed for c-Jun or JunD. Although the IP successfully pulled down Flag-EHF, no associated c-Jun or JunD was detected. Bio-layer interferometry was also utilized; it has the added ability to measure the kinetics of an interaction, and thus could detect transient or weak binding (199). Like in the Co-IP, although His6-EHF was successfully immobilized on the biosensor, no interaction was detected when the coated sensor was incubated in a solution containing Flag-c-Jun.

Several possible conclusions can be drawn from these results. The first is that EHF and the Jun proteins do not directly interact, and their co-occupancy of sites is a result of indirect cooperation. This may involve mechanisms including joint recruitment of a transcriptional complex, chromatin remodeling by one factor to allow binding of the second, or induction of local DNA bending to promote TF recruitment

(102). Alternatively, EHF may interact with one or both proteins, but this interaction was not detectable in the methods used. The interaction may be weak or transient, requiring cross-linking to be detectable by co-IP (223). Alternatively, it may require another co-factor or a post-translational modification downstream of a signaling pathway. Both AP-1 and Ets TFs are known to be downstream of MAPK signaling, for example, and for some Ets factors, phosphorylation is required for protein-protein interaction (166,182). Co-IPs can be performed using a cross-linking agent, in an airway epithelial cell type that expresses putative co-factors, or following treatment known to activate pathways leading to post-translational modifications of one or both TFs.

The proteins purified for the bio-layer interferometry could be misfolded, even though the denatured proteins are recognized by a specific antibody via western blot. The epitopes, although located on the N terminus distal to the putative interacting domain, may be impeding the interaction. Since the DNA-binding Ets and basic domains are the predicted interacting regions for EHF and c-Jun/JunD, respectively, an electrophoretic mobility shift assay (EMSA) can be used to confirm that the protein is in the correct conformation for binding to DNA (224).

A different method could be used to probe for protein-protein interactions; affinity purification of EHF coupled with mass spectrometry (MS) can determine interacting partners of EHF in a non-biased fashion (225). In this method, endogenous or Flag-tagged EHF is immunoprecipitated, and the components of the targeted complex identified by MS. This would allow for confirmation of an interaction (or lack thereof) between EHF and Jun proteins, but could also identify other proteins that interact with EHF.

4.4 EHF regulates processes including epithelial repair and inflammation in the lung epithelium

EHF modulation in several airway epithelial cell types showed that differentially expressed genes (DEGs) are enriched for common processes. Depletion of EHF in Calu-3 cells alters loci important for epithelial and ectoderm development and response to wounding, and its overexpression in A549 cells differentially regulates genes involved in response to wounding and proliferation. In primary HBE cells,

EHF-activated genes contribute to the immune response and response to wounding. Our data correlate well with the reported effects of EHF knockdown in corneal epithelium in the mouse; in this tissue, EHF-regulated genes are enriched for processes including response to wounding, epithelial and ectoderm development, and proliferation (226). Although the specialized biological properties of the airway and corneal epithelium are different, they share common features of all epithelial layers in the body.

4.4.1 EHF regulates genes involved in epithelial wound closure

The airway epithelium is constantly under stress from infection, inflammation, trauma, and noxious particles, and by necessity has developed a complex response to damage. Cells adjacent to the wound spread over the denuded area, then proliferate and differentiate to reconstitute a functioning epithelial layer (37,38). EHF regulates genes involved in wound closure in Calu-3, A549 and HBE cells. It is also implicated in wound repair in the mouse corneal epithelium (226) and epithelial cell migration in prostate cells during carcinogenesis (172,174). This is likely a common role for EHF across epithelial cells types.

There are multiple processes involved in epithelial migration; cell-cell adhesions must be modified (227) and the ECM remodeled through degradation and secretion of proteins. Epithelial adhesion to the ECM is tightly regulated, a balance between adhering enough to gain traction, but not so much as to impede movement (41,42). EHF regulates multiple genes involved in these processes across the airway epithelial cells studied. In HBE and Calu-3 cells, EHF depletion decreases *S100A8* and *S100A9*, which encode proteins that are pro-migratory for epithelial cells (228,229). Expression of the tight junction protein *TJP2* is increased along with EHF in Calu-3 cells. EHF also regulates genes involved in ECM remodeling; it activates expression of the pro-migratory matrix metallopeptidases *MMP2* in HBE cells and *MMP14* in A549 cells (42). Expression of type I and type IV collagens (*COL1A1, COL4A3, COL4A4, COL4A5, COL4A6*), which are migratory substrates for the airway epithelium, are differentially regulated by EHF in A549 cells; the integrins that bind them are activated by EHF (*ITGA2* in Calu-3, *ITGA6* in A549)

(39,40). Although the gene targets differ between the cell types, the general pathway is common and indicates a pro-migratory role for EHF.

One common target of EHF across all cell types studied is Plasminogen Activator, Urokinase (*PLAU,* uPA) and its receptor (*PLAUR,* uPAR). *PLAU* expression decreases with EHF in Calu-3 cells and HBE cells, and *PLAUR* increases with EHF in A549 cells, implicating them as loci activated by the factor. Blocking the interaction between uPA and its receptor attenuates wound repair in primary human bronchial epithelial cells (230). Furthermore, uPA induces migration and expression of markers of EMT in primary small airway epithelial cells (231). This receptor-ligand interaction has been implicated in human pulmonary diseases; *PLAU* expression may be increased in tracheal epithelial cultures with F508del mutated *CFTR* as compared to wild type cells (232). uPA levels are elevated in response to exposure to cigarette smoke and in patients with COPD (231); uPAR is increased in subjects with asthma (230).

To determine whether alterations in gene expression correlated with changes in wound repair *in vitro,* we performed a scratch assay on Calu-3 and wild type HBE cells. Wound closure is impaired upon EHF depletion, consistent with down-regulation of genes that promote cell motility. Of interest to CF lung biology, cells that have mutant *CFTR* may show slowed wound repair (48,49), which is problematic in a disease characterized by inflammation and resulting damage to the lung epithelium. Depletion of EHF slowed scratch closure in CF HBE cells. By further impairing a phenotype that is already compromised in *CFTR* mutant epithelial cells, EHF could modify lung disease severity in CF patients.

4.4.2 EHF controls expression of loci important for inflammation and immunity

Airway inflammation works in tandem with re-epithelialization in wound repair, and is also critical to the general health of the lung. It is a necessary response to trauma and pathogens, but when exaggerated or impaired can contribute to pulmonary pathology. Our genome-wide studies strongly implicate EHF in regulating inflammatory genes in airway epithelial cells. Its targets are enriched for processes including infection and response to wounding, which encompasses genes involved in cell migration, as discussed in the previous section, and those that encode immune mediators.

Inflammation is an integral part of the response to injury and determines, in part, whether damaged epithelium undergoes proper repair or remodeling (38). Secondary inflammation may worsen acute epithelial injury, chronic inflammation contributes to structural changes in the airway, and repeated cycles of damage, inflammation, and repair can lead to remodeling and dysfunction of the lung tissue (44). In the cell types studied, EHF regulated inflammatory loci involved in wound repair, remodeling, and fibrosis. Interleukin 1 beta (*IL1B*), which augments airway epithelial repair in multiple models, decreases in expression following EHF depletion in HBE cells (38). Two genes with altered expression in EHF-depleted Calu-3 cells, *CXCL1* and *CXCL6*, have potential pro-fibrotic roles; CXCL1 is over-expressed in fibrotic versus non-fibrotic lungs (233) and CXCL6 contributes to inflammation and fibrosis in the bleomycin mouse lung fibrosis model (234). *IL6* and *IL11* show increased expression with elevated EHF in A549 cells; overexpression of these cytokines in the mouse lung leads to structural changes and altered responsiveness to methacholine in the airway (235). Overall, EHF activates most of these genes, suggesting that it enhances the remodeling response.

Of the genes tested to validate the Calu-3 RNA-seq results, *DDAH1* shows the greatest change in expression following EHF depletion. DDAH1 metabolizes asymmetric dimethylarginine (ADMA), a competitive inhibitor of nitric oxide synthase (NOS). ADMA potentiates lung inflammation in a mouse model of allergic asthma (236). Furthermore, DDAH1 overexpression attenuates airway inflammation in a different mouse model of asthma by decreasing IgE expression and eosinophil migration in the lung. It also decreases levels of genes encoding cytokines, chemokines and matrix metallopeptidases (237). Since EHF represses *DDAH1* in our studies, this suggests a role for the transcription factor in potentiating allergic lung inflammation.

Il-17a and LPS induce bronchial epithelial cells to release neutrophil chemokines including *CXCL1, CXCL6* and *IL8* (28,31,34,35). These cytokines promote migration and survival of neutrophils. This is critical for CF lung disease, as a neutrophil-dominant inflammation is present early in the illness, and their release of proteases contributes to lung damage (62-65,238). These loci map to chr4q13.3, a region that encompasses 2 EHF binding sites in the primary HBE cells. Furthermore, EHF depletion in

lung adenocarcinoma cells increases *IL8* expression and decreases *CXCL6*, both at basal levels and in response to immune mediators. This correlates with changes in secretion of CXCL6, but not IL8. By increasing CXCL6 secretion, EHF could increase neutrophil recruitment and survival, thereby exacerbating inflammation in the CF lung. Gram-negative bacteria stimulate the innate immune system through LPS binding to TLRs (33-35), and IL-17a is secreted by Th17 cells in a branch of the adaptive immune system (27). The ability of EHF to control expression of chemokines downstream of both signaling pathways suggests a role for it in modulating innate and adaptive immunity.

EHF regulates many loci involved in the immune response, activating some and repressing others. The net effect of EHF on the immune system is likely complex. However, in general EHF activates expression of pro-inflammatory mediators while repressing those that dampen inflammation. Combined with the observation that EHF expression increases in response to inflammatory mediators in the lung epithelium (176), we predict a pro-inflammatory role for EHF in this tissue type.

4.5 EHF participates in a network of transcription factors responsible for epithelial development and differentiation

It is critical to understand how gene expression is regulated in the lung epithelium, because an appropriate response to the environment is necessary to maintain a healthy airway, and abnormal or exaggerated reactions can lead to disease. Many TFs have been investigated in the pulmonary epithelium, particularly through loss of function studies in rodents. Here, we have shown a role for EHF in the transcription factor network controlling pulmonary gene expression (Figure 4.4).
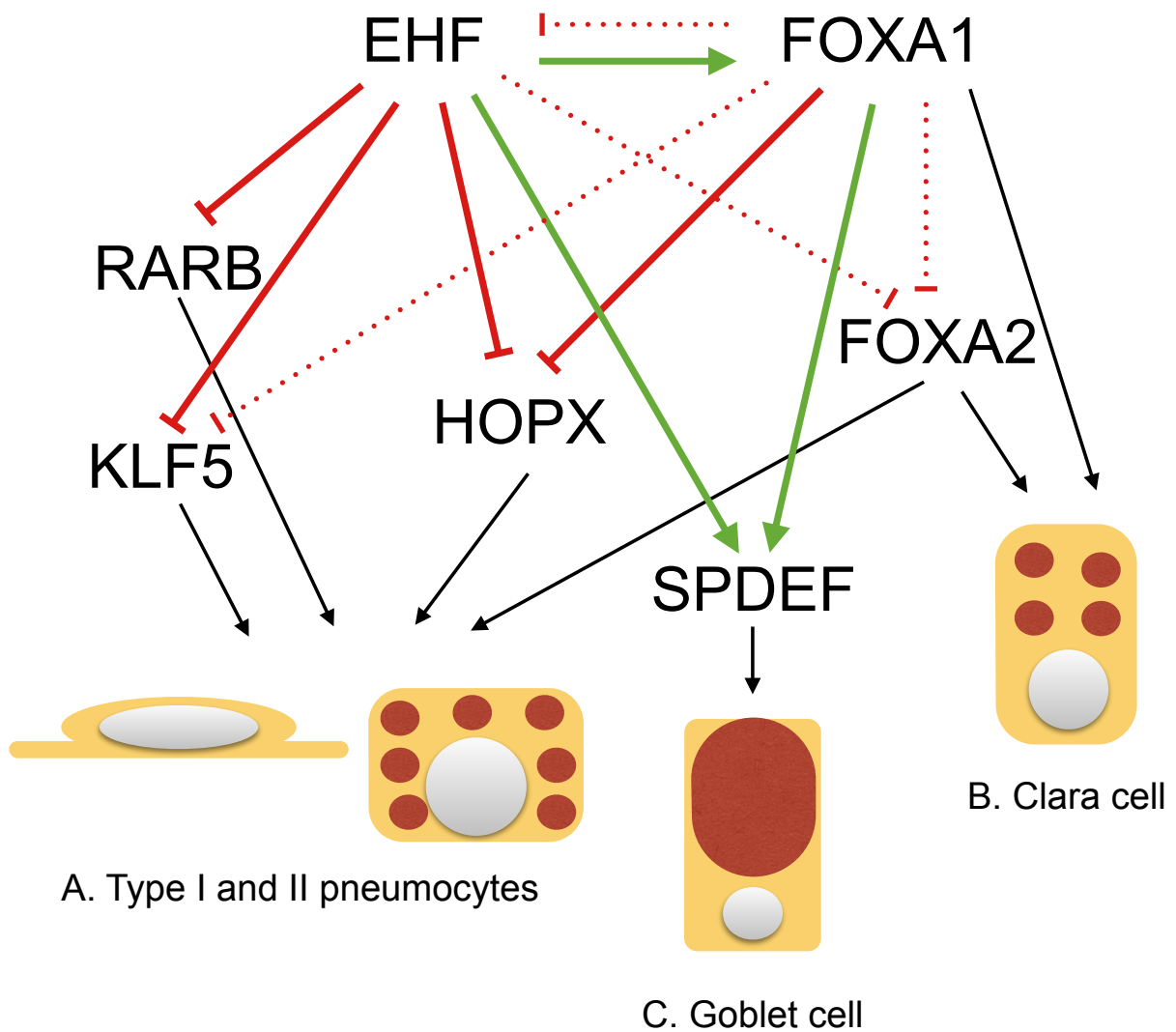
Figure 4.5 EHF is part of a transcription factor network that regulates airway epithelial differentiation. A) EHF and FOXA1 directly (putative direct targets, solid red line) and indirectly (dashed red line) repress TFs that activate differentiation of type I (left) and type II (right) pneumocytes. B) EHF regulates transcription factors necessary for Clara cell maturation. C) EHF and FOXA2 activate (solid green line) SPDEF expression, which increases goblet cell differentiation.

4.5.1 EHF represses expression of transcription factors involved in epithelial differentiation

EHF-regulated genes are enriched for gene ontology processes including epithelial differentiation and development in Calu-3 cells. In the HBE cells, EHF represses multiple transcription factors involved in lung epithelial development, including *HOPX, KLF5, RARB*, and *FOXA2*. In rodents, Hopx acts downstream of Nkx2-1 to repress GATA6/Nkx2-1-induced expression of type II pneumocyte genes (141). Although it is initially expressed in a bipotent alveolar progenitor, Hopx is eventually limited to type I pneumocytes, the alveolar cells involved in gas exchange (239). Loss of Hopx in the developing lung leads to impaired type II pneumocyte differentiation, abnormal alveolar development and the presence of increased surfactants (141). In the mouse, loss of Klf5 or FoxA2 in the developing lung epithelium results in a lack of differentiated alveolar epithelial cells (142,144,240). RARβ is also expressed in type II pneumocytes, and its loss impairs alveolarization (241). By repressing *HOPX, KLF5*, *RARB,* and *FOXA2* in the human airway*,* EHF may limit differentiation down the alveolar epithelial cell lineage. EHF repression of *RARB, HOPX,* and *KLF5* likely occurs through binding to the promoter or a nearby *cis*-regulatory element; they are putative direct targets. Inhibition of *FOXA2* expression is likely indirect, as there is no nearby EHF binding site, but may involve *FOXA1.*

EHF regulation of *FOXA1* and *FOXA2* may impact differentiation of conducting airway epithelial cells. In mice where both proteins are deleted, Foxj1*,* a differentiation marker of ciliated bronchial epithelia, and Clara cell secretory protein (CCSP) are absent (143). *FOXJ1* is not expressed in the HBE cells grown on plastic, but *CCSP* expression decreases following EHF depletion. This indicates that EHF may increase differentiation towards the Clara cell lineage despite a decrease in *FOXA2* expression. EHF may also contribute to branching morphogenesis, the process by which the lung buds divide and expand into the respiratory tree. Retinoic acid receptor signaling through RARβ and the combined deletion of FoxA1 and FoxA2 inhibit branching morphogenesis (137,143). The effect of EHF depletion on branching morphogenesis cannot be predicted from changes in gene expression, as it represses both *RARB* and *FOXA2,* while activating *FOXA1.*

A potential role for EHF in lung organogenesis and differentiation of specific epithelial cell types can only be hypothesized from our siRNA depletion studies. This developmental process is extremely complex, so to confirm the effect of the TF on development, an animal model with knockout of EHF specifically in the embryonic airway epithelium should be developed. This would allow for observations of the net effect of EHF deletion on maturation of the lung.

4.5.2 EHF and FOXA1 have overlapping roles in regulation of other transcription factors in the lung epithelium

FOXA1 is a pioneer transcription factor; it binds to and opens compacted chromatin for other proteins to bind (145). Loss of FoxA1 expression in mice leads to delayed maturation of the epithelium and alveolarization, however, these perturbations are compensated after birth (242). The compensation may be due to redundancy in roles between FoxA1 and FoxA2 (143). FOXA1 expression is increased in the airway epithelium of smokers with airway obstruction (as seen in COPD) compared with those with no impairment (243). In HBE cells, EHF binding peaks are enriched for the Forkhead factor motif that is bound by both FOXA1 and FOXA2. The motif is enriched in sites found near EHF-repressed genes compared with those near non-regulated genes, specifically in peaks within 10kb of the TSS but not those found at promoters. These data suggest that FOXA1 may work with EHF to mediate expression by action at *cis*-regulatory elements. This is confirmed by FOXA1 depletion in primary HBE cells, which causes alterations in expression of *FOXA2, HOPX, KLF5* and *SPDEF* that parallel those seen in EHF-depleted cells, suggesting that EHF and FOXA1 may cooperate in the complex transcription factor network that regulates airway epithelial differentiation. EHF and FOXA1 also regulate each other, implying that they perhaps participate together in a feedback loop.

Furthermore, FOXA1 was bound to the EHF sites near *HOPX* and *SPDEF.* FOXA1 may be binding to and remodeling closed chromatin to open it for EHF binding. To confirm that EHF and FOXA1 are bound to these sites at the same time, ChIP-reChIP can be used. The effect of FOXA1 depletion on

EHF binding to DNA should also be investigated to determine whether binding of FOXA1 is necessary for EHF to occupy these regions.

4.5.3 EHF activates a transcriptional program responsible for goblet cell hyperplasia

Although it repressed many transcription factors involved in lung development, EHF activated expression of *SPDEF* in all airway epithelial cell types studied, indicating an important role for EHF in controlling its transcription. SPDEF is primarily responsible for inducing goblet cell differentiation and mucus production in the airway (149). It is a mediator of the Th2 response, and its expression is induced by allergen and IL-13 exposure. In the conducting airways, an increase in SPDEF expression leads to goblet cell hyperplasia, increased mucus secretion, airway hyperresponsiveness, and eosinophilic inflammation (22,148-150). It inhibits TLR signaling, driving the inflammatory response towards a Th2 signature (150). SPDEF also represses transcription of FOXA2, which in turn inhibits SPDEF expression, goblet cell hyperplasia, mucin production, and eosinophilic inflammation in the airway (22,144,149,244,245). FOXA2 expression is reduced in patients with asthma, implicating perturbations in the balance between FOXA2 and SPDEF in human disease (244). By activating expression of *SPDEF,* and repressing *FOXA2,* EHF adds another layer of regulation to these opposing transcription factors. These data implicate EHF in activating goblet cell differentiation, which is consistent with previous studies in which immunohistochemical analysis showed the most intense staining for EHF in mucus gland epithelial cells, with a smaller fraction of other bronchial cells being positive for the factor (160,161). Further support for this model comes from the effect of IL-13 to increase EHF expression in polarized HBE cells.

4.6 EHF as a modifier of lung disease

These data show that EHF is a critical regulator of multiple lung epithelial functions that are integral to airway diseases, including cystic fibrosis. *EHF* maps adjacent (5') to an intergenic region of chr11p13 that contains SNPs that associate with lung disease severity in CF patients (187,189). Since the SNPs were found in an intergenic region, the critical variants probably impact *cis*-acting regulatory elements for nearby genes that are required for normal lung function. The disease-causing SNP(s) has (have) not been identified due to strong linkage disequilibrium between the most significant SNP identified by the GWAS and multiple variants in the region. Furthermore, to date no correlation has been established between the disease-associated SNPs and EHF activity. A method to investigate a potential link involves obtaining airway epithelial samples from patients with either the risk or the non-risk haplotype, and then determining relative expression of nearby genes, including EHF. The changes in expression could be subtle, so a large sample number may be required. Another option is to utilize clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 technology (246,247) to alter the endogenous genetic landscape in cell lines or primary cells treated with Rho kinase inhibitors to extend survival in culture (248,249). As an initial screen, whole regulatory elements can be removed to identify those that impact EHF expression. Then, single nucleotide changes can be introduced to determine the effect of SNPs on gene transcription. Through a multi-faceted approach, the function of the 11p13 region may be elucidated; however, regardless of whether its role is characterized, EHF is a strong candidate for being a modifier of lung disease severity due to its impact on *CFTR* expression, inflammation, and wound repair.

4.6.1 EHF regulation of CFTR

EHF binding sites were identified at DHS across the *CFTR* locus, including at a site -35 kb 5' to the gene, which encompasses an airway-selective *CFTR* enhancer that binds IRF1 and NF-Y (156). EHF represses CFTR mRNA and protein expression in Calu-3 cells. Besides a direct mechanism, EHF may

indirectly regulate CFTR through FOXA1 and FOXA2, which have a combined contribution to activating

transcription of CFTR in the intestinal epithelium (105). They regulate *CFTR* through action at multiple

sites, including an enhancer in intron 11 of the gene (146). A DHS may be evident at this site in HBE cells

(155), suggesting FOXA1 and FOXA2 may have a similar role in this cell type. By repressing *FOXA2*

expression, EHF may indirectly decrease expression of *CFTR.* Some disease-causing *CFTR* mutations

result in a small fraction of the protein being properly localized and functional (36). In patients with these

genotypes, repression of the locus could have deleterious effects on ion transport in the lung, and an

increase in EHF could further exacerbate CF lung disease.

4.6.2 Altered EHF expression modulates epithelial phenotypes that are abnormal in CF lung disease

In many of our studies, the effect of decreased EHF expression on epithelial function is opposite

to the defects seen in the cystic fibrosis lung (Figure 4.5.2). EHF depletion represses secretion of the

neutrophil cytokine CXCL6, whereas neutrophilic inflammation is an early finding in CF lung disease that

contributes to tissue destruction (62-65). Loss of EHF also depresses SPDEF expression; abnormal

submucosal glands play a pivotal role in CF lung disease (4). SPDEF-mediated changes in the airway

epithelium are induced by exposure to IL-13, which is increased in the CF lung (22,24-26,149). IL-13

treatment of polarized HBE cells also increases EHF expression. EHF represses *CFTR* expression,

possibly through direct binding to the *CFTR* locus or through regulation of an intermediate transcription

factor. Together, these data suggest that an increase in EHF expression will exacerbate CF pulmonary

pathology through increased neutrophil recruitment, goblet cell hyperplasia, and decreased *CFTR*

expression. If the 11p13 region is modifying lung disease through EHF, we predict that causative SNPs

increase its transcription.

Unlike the other phenotypes studied, the effect of decreased EHF on wound repair seemingly

parallels that seen in CF epithelial cells. Cells with mutated CFTR and depleted EHF may have delayed

wound closure *in vitro* (48,49,250), which implicates loss of EHF in diminishing a process already

impaired in CF. However, wound repair and remodeling in the CF lung is more complex than the *in vitro* scratch assay. Both the inherent CFTR defect and external stress from repeated cycles of inflammation and infection are thought to contribute to remodeling and tissue dysfunction. Thickening of the reticular basement membrane (RBM) due to wound repair processes of the overlying epithelium is present early in the disease, although its impact on pulmonary function is unclear (68,70,251). Proteases released primarily from neutrophils overwhelm anti-protease activity, damaging the epithelial layer and inducing goblet cell metaplasia (64). These proteases also degrade the ECM, leading to release of growth factors important for epithelial repair and likely RBM fibrosis (65). The complexities of tissue remodeling make it difficult to accurately predict how changes in an *in vitro* assay translate to alterations in wound repair and remodeling *in vivo.* Perhaps an increase in EHF activates wound repair pathways that lead to increased fibrosis or abnormal reconstitution of the epithelial layer, or as suggested by the scratch assay, an increase in EHF is protective by offsetting the inherent defect in wound repair caused by loss of CFTR function.

EHF clearly impacts lung epithelial function, regulating pathways that are particularly important for pathogenesis of airway disease. However, because of the complexities of these pathways, the impact of increased or decreased EHF expression on net pulmonary function is not certain. An animal model with EHF deleted in the lung epithelia would greatly improve our understanding of the net effect of EHF loss on airway function.
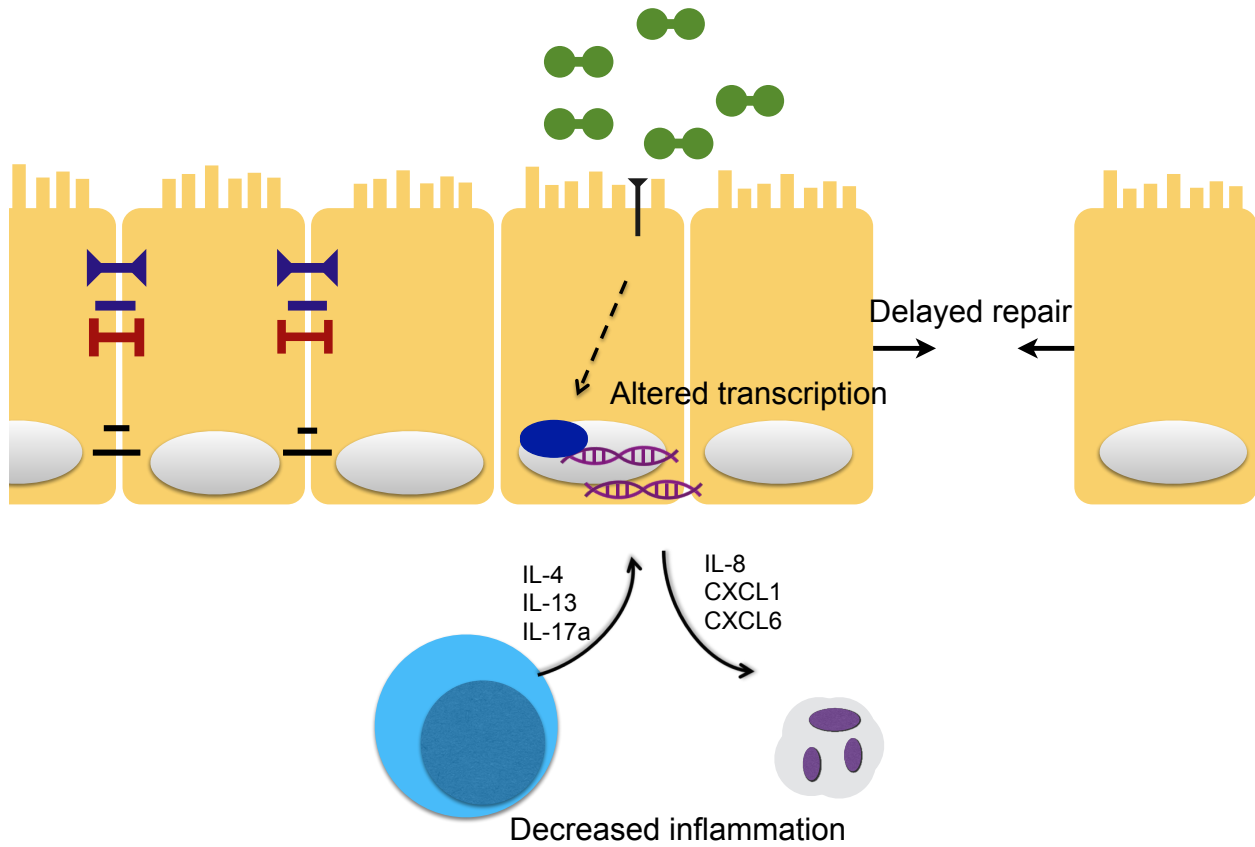
Figure 4.6.2 Model of the effect of EHF depletion on airway epithelial function. When EHF is depleted from airway epithelial cells, transcription is altered, causing delayed wound repair and decreased inflammation.

4.7 Conclusions

Overall, this study makes a strong case for EHF as a potential modifier of lung pathology in cystic fibrosis and other diseases in which the immune system and epithelial repair mechanisms play a critical role, including COPD and asthma. It activates a transcriptional program that increases goblet cell differentiation, inflammation, and wound repair. Alterations in EHF levels likely lead to multiple downstream effects on epithelial processes, thereby altering lung function. In the future, EHF should be studied as a potential target to alleviate airway dysfunction in the context of existing disease.

REFERENCES

1.   Boron, W.F. (2005), *Medical Physiology*. Updated Edition ed. Elsevier, pp. 593-605.

2.   Knight, D.A. and Holgate, S.T. (2003) The airway epithelium: structural and functional properties in health and disease. *Respirology*, **8**, 432-446.

3.   Branchfield, K., Nantie, L., Verheyden, J.M., Sui, P., Wienhold, M.D. and Sun, X. (2016) Pulmonary neuroendocrine cells function as airway sensors to control lung immune response. *Science*, **351**, 707-710.

4.   Widdicombe, J.H. and Wine, J.J. (2015) Airway Gland Structure and Function. *Physiol Rev*, **95**, 1241-1319.

5.   Crystal, R.G., Randell, S.H., Engelhardt, J.F., Voynow, J. and Sunday, M.E. (2008) Airway epithelial cells: current concepts and challenges. *Proc Am Thorac Soc*, **5**, 772-777.

6.   Rezaee, F. and Georas, S.N. (2014) Breaking barriers. New insights into airway epithelial barrier function in health and disease. *Am J Respir Cell Mol Biol*, **50**, 857-869.

7.   Hallstrand, T.S., Hackett, T.L., Altemeier, W.A., Matute-Bello, G., Hansbro, P.M. and Knight, D.A. (2014) Airway epithelial regulation of pulmonary immune homeostasis and inflammation. *Clin Immunol*, **151**, 1-15.

8.   LeSimple, P., Liao, J., Robert, R., Gruenert, D.C. and Hanrahan, J.W. (2010) Cystic fibrosis transmembrane conductance regulator trafficking modulates the barrier function of airway epithelial cell monolayers. *J Physiol*, **588**, 1195-1209.

9.   Ahdieh, M., Vandenbos, T. and Youakim, A. (2001) Lung epithelial barrier function and wound healing are decreased by IL-4 and IL-13 and enhanced by IFN-γ. *Am J Physiol Cell Physiol*, **281**, C2029-C2038.

10.   Petecchia, L., Sabatini, F., Usai, C., Caci, E., Varesio, L. and Rossi, G.A. (2012) Cytokines induce tight junction disassembly in airway cells via an EGFR-dependent MAPK/ERK1/2-pathway. *Lab Invest*, **92**, 1140-1148.

11.   Widdicombe, J.H. (2002) Regulation of the depth and composition of airway surface liquid. *J Anat*, **201**, 313-318.

12.   Kesimer, M., Ehre, C., Burns, K.A., Davis, C.W., Sheehan, J.K. and Pickles, R.J. (2013) Molecular organization of the mucins and glycocalyx underlying mucus transport over mucosal surfaces of the airways. *Mucosal Immunol*, **6**, 379-392.

13.   Fahy, J.V. and Dickey, B.F. (2010) Airway mucus funciton and dysfunction. *N Engl J Med*, **363**, 2233-2247.

14.   Knowles, M.R. and Boucher, R.C. (2002) Mucus clearance as a primary innate defense mechanism for mammalian airways. *J Clin Invest*, **109**, 571-577.

15.   Gustafsson, J.K., Ermund, A., Ambort, D., Johansson, M.E., Nilsson, H.E., Thorell, K., Hebert, H., Sjovall, H. and Hansson, G.C. (2012) Bicarbonate and functional CFTR channel are required for proper mucin secretion and link cystic fibrosis with its mucus phenotype. *J Exp Med*, **209**, 1263-1272.

16.   Kreda, S.M., Davis, C.W. and Rose, M.C. (2012) CFTR, mucins, and mucus obstruction in cystic fibrosis. *Cold Spring Harb Perspect Med*, **2**, a009589.

17.   Kawai, T. and Akira, S. (2010) The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol*, **11**, 373-384.

18.   Licona-Limon, P., Kim, L.K., Palm, N.W. and Flavell, R.A. (2013) TH2, allergy and group 2 innate lymphoid cells. *Nat Immunol*, **14**, 536-542.

19.   Halim, T.Y., Steer, C.A., Matha, L., Gold, M.J., Martinez-Gonzalez, I., McNagny, K.M., McKenzie, A.N. and Takei, F. (2014) Group 2 innate lymphoid cells are critical for the initiation of adaptive T helper 2 cell-mediated allergic lung inflammation. *Immunity*, **40**, 425-435.

20.   Kuperman, D.A., Huang, X., Koth, L.L., Chang, G.H., Dolganov, G.M., Zhu, Z., Elias, J.A., Sheppard, D. and Erle, D.J. (2002) Direct effects of interleukin-13 on epithelial cells cause airway hyperreactivity and mucus overproduction in asthma. *Nat Med*, **8**, 885-889.

21.   Zhen, G., Park, S.W., Nguyenvu, L.T., Rodriguez, M.W., Barbeau, R., Paquet, A.C. and Erle, D.J. (2007) IL-13 and epidermal growth factor receptor have critical but distinct roles in epithelial cell mucin production. *Am J Respir Cell Mol Biol*, **36**, 244-253.

22.   Park, K.S., Korfhagen, T.R., Bruno, M.D., Kitzmiller, J., Wan, H., Wert, S., Hershey, G.K.K., Chen, G. and Whitsett, J.A. (2007) SPDEF regulates goblet cell hyperplasia in the airway epithelium. *J Clin Invest*, **117**, 978-988.

23. Batra, V., Musani, A.I., Hastie, A.T., Khurana, S., Carpenter, K.A., Zangrilli, J.G. and Peters, S.P. (2004) Bronchoalveolar lavage fluid concentrations of transforming growth factor (TGF)-β1, TGF-β2, interleukin (IL)-4 and IL-13 after segmental allergen challenge and their effects on a-smooth muscle actin and collagen III synthesis by primary human lung fibroblasts. *Clin Exp Allergy*, **34**, 437-444.

24. Eickmeier, O., Huebner, M., Herrmann, E., Zissler, U., Rosewich, M., Baer, P.C., Buhl, R., Schmitt-Grohe, S., Zielen, S. and Schubert, R. (2010) Sputum biomarker profiles in cystic fibrosis (CF) and chronic obstructive pulmonary disease (COPD) and association between pulmonary function. *Cytokine*, **50**, 152-157.

25. Hartl, D., Griese, M., Kappler, M., Zissel, G., Reinhardt, D., Rebhan, C., Schendel, D.J. and Krauss-Etschmann, S. (2006) Pulmonary T(H)2 response in Pseudomonas aeruginosa-infected patients with cystic fibrosis. *J Allergy Clin Immunol*, **117**, 204-211.

26. Tiringer, K., Treis, A., Fucik, P., Gona, M., Gruber, S., Renner, S., Dehlink, E., Nachbaur, E., Horak, F., Jaksch, P., Doring, G., Crameri, R., Jung, A., Rochat, M.K., Hormann, M., Spittler, A., Klepetko, W., Akdis, C.A., Szepfalusi, Z., Frischer, T. and Eiwegger, T. (2013) A Th17- and Th2-skewed cytokine profile in cystic fibrosis lungs represents a potential risk factor for Pseudomonas aeruginosa infection. *Am J Respir Crit Care Med*, **187**, 621-629.

27. Park, H., Li, Z., Yang, X.O., Chang, S.H., Nurieva, R., Wang, Y.H., Wang, Y., Hood, L., Zhu, Z., Tian, Q. and Dong, C. (2005) A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nat Immunol*, **6**, 1133-1141.

28. Prause, O., Laan, M., Lötvall, J. and Lindén, A. (2003) Pharmacological modulation of interleukin-17-induced GCP-2-, GRO-α- and interleukin-8 release in human bronchial epithelial cells. *Eur J Pharmacol*, **462**, 193-198.

29. McAllister, F., Henry, A., Kreindler, J.L., Dubin, P.J., Ulrich, L., Steele, C., Finder, J.D., Pilewski, J.M., Carreno, B.M., Goldman, S.J., Pirhonen, J. and Kolls, J.K. (2005) Role of IL-17A, IL-17F, and the IL-17 Receptor in Regulating Growth-Related Oncogene- and Granulocyte Colony-Stimulating Factor in Bronchial Epithelium: Implications for Airway Inflammation in Cystic Fibrosis. *J Immunol*, **175**, 404-412.

30. Ye, P., Rodriguez, F.H., Kanaly, S., Stocking, K.L., Schurr, J., Schwarzenberger, P., Oliver, P., Huang, W., Zhang, P., Zhang, J., E., S.J., Bagby, G.J., Nelson, S., Charrier, K., Peschon, J.J. and Kolls, J.K. (2001) Requirement of interleukin 17 receptor signaling for lung CXC chemokine and granulocyte colony-stimulating factor expression, neutrophil recruitment, and host defense. *J Exp Med*, **194**, 519-527.

31. Tan, H.L., Regamey, N., Brown, S., Bush, A., Lloyd, C.M. and Davies, J.C. (2011) The Th17 pathway in cystic fibrosis lung disease. *Am J Respir Crit Care Med*, **184**, 252-258.

32. Sender, V. and Stamme, C. (2014) Lung cell-specific modulation of LPS-induced TLR4 receptor and adaptor localization. *Commun Integr Biol*, **7**, e29053.

33. Guillot, L., Medjane, S., Le-Barillec, K., Balloy, V., Danel, C., Chignard, M. and Si-Tahar, M. (2004) Response of human pulmonary epithelial cells to lipopolysaccharide involves Toll-like receptor 4 (TLR4)-dependent signaling pathways: evidence for an intracellular compartmentalization of TLR4. *J Biol Chem*, **279**, 2712-2718.

34. Pace, E., Ferraro, M., Siena, L., Melis, M., Montalbano, A.M., Johnson, M., Bonsignore, M.R., Bonsignore, G. and Gjomarkaj, M. (2008) Cigarette smoke increases Toll-like receptor 4 and modifies lipopolysaccharide-mediated responses in airway epithelial cells. *Immunology*, **124**, 401-411.

35. Wu, Y.L., Kou, Y.R., Ou, H.L., Chien, H.Y., Chuang, K.H., Liu, H.H., Lee, T.S., Tsai, C.Y. and Lu, M.L. (2010) Glucosamine regulation of LPS-mediated inflammation in human bronchial epithelial cells. *Eur J Pharmacol*, **635**, 219-226.

36. Elborn, J.S. (2016) Cystic fibrosis. *The Lancet*.

37. Zahm, J.-M., Chevillard, M. and Edith Puchelle, E. (1991) Wound repair of human surface respiratory epithelium. *Am J Respir Cell Mol Biol*, **5**, 242-248.

38. Crosby, L. and Waters, C.M. (2010) Epithelial repair mechanisms in the lung. *Am J Physiol Lung Cell Mol Physiol*, **298**, L715-L731.

39. Kim, H.J., Henke, C.A., Savik, S.K. and Ingbar, D.H. (1997) Integrin mediation of alveolar epithelial cell migration on fibronectin and type I collagen. *Am J Physiol* **273**, L134-141.

40. White, S.R., Dorscheid, D.R., Rabe, K.F., Wojcik, K.R. and Hamann, K.J. (1999) Role of very late adhesion integrins in mediating repair of human airway epithelial cell monolayers after mechanical injury. *Am J Respir Cell Mol Biol*, **20**, 787-796.

41. Chen, P., Abacherli, L.E., Nadler, S.T., Wang, Y., Li, Q. and Parks, W.C. (2009) MMP7 shedding of syndecan-1 facilitates re-epithelialization by affecting $\alpha_2\beta_1$ integrin activation. *PLoS One*, **4**, e6565.

42. Chen, P. and Parks, W.C. (2009) Role of matrix metalloproteinases in epithelial migration. *J Cell Biochem*, **108**, 1233-1243.

43.   Perrio, M.J., Ewen, D., Trevethick, M.A., Salmon, G.P. and Shute, J.K. (2007) Fibrin formation by wounded bronchial epithelial cell layers in vitro is essential for normal epithelial repair and independent of plasma proteins. *Clin Exp Allergy*, **37**, 1688-1700.

44.   Beers, M.F. and Morrisey, E.E. (2011) The three R's of lung health and disease: repair, remodeling, and regeneration. *J Clin Invest*, **121**, 2065-2073.

45.   Thiery, J.P. and Sleeman, J.P. (2006) Complex networks orchestrate epithelial-mesenchymal transitions. *Nat Rev Mol Cell Biol*, **7**, 131-142.

46.   Singh, A. and Settleman, J. (2010) EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene*, **29**, 4741-4751.

47.   Kim, K.K., Kugler, M.C., Wolters, P.J., Robillard, L., Galvez, M.G., Brumwell, A.N., Sheppard, D. and Chapman, H.A. (2006) Alveolar epithelial cell mesenchymal transition develops in vivo during pulmonary fibrosis and is regulated by the extracellular matrix. *Proc Natl Acad Sci U S A*, **103**, 13180-13185.

48.   Schiller, K.R., Maniak, P.J. and O'Grady, S.M. (2010) Cystic fibrosis transmembrane conductance regulator is involved in airway epithelial wound repair. *Am J Physiol Cell Physiol*, **299**, C912-C921.

49.   Itokazu, Y., Pagano, R.E., Schroeder, A.S., O'Grady, S.M., Limper, A.H. and Marks, D.L. (2014) Reduced GM1 ganglioside in CFTR-deficient human airway cells results in decreased beta1-integrin signaling and delayed wound repair. *Am J Physiol Cell Physiol*, **306**, C819-830.

50.   Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., Drumm, M.L., Iannuzzi, C., Collins, F.S. and Tsui, L.C. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, **245**, 1066-1073.

51.   Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M. and Tsui, L.C. (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science*, **245**, 1073-1080.

52.   Grosse, S.D., Boyle, C.A., Botkin, J.R., Comeau, A.M., Kharrazi, M., Rosenfeld, M. and Wilfond, B.S. (2004). National Center on Birth Defects and Developmental Disabilities, Vol. 53, pp. 1-36.

53.   Cystic Fibrosis Foundation Patient Registry. (2015) 2014 Annual Data Report.

54.   Strom, C.M., Crossley, B., Buller-Buerkle, A., Jarvis, M., Quan, F., Peng, M., Muralidharan, K., Pratt, V., Redman, J.B. and Sun, W. (2011) Cystic fibrosis testing 8 years on: lessons learned from carrier screening and sequencing analysis. *Genet Med*, **13**, 166-172.

55.   Cheng, S.H., Gregory, R.J., Marshall, J., Paul, S., Souza, D.W., White, G.A., O'Riorden, C.R. and Smith, A.E. (1990) Defective Intracellular Transport and Processing of CFTR is the Molecular Basis of Most Cystic Fibrosis. *Cell*, **63**, 827-834.

56.   Bear, C.E., Li, C., Kartner, N., Bridges, R.J., Jensen, T.J., Ramjeesingh, M. and Riordan, J.R. (1992) Purification and Functional Reconstitution of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR). *Cell*, **68**, 809-818.

57.   Rowe, S.M., Miller, S.M. and Sorscher, E.J. (2005) Cystic Fibrosis. *N Engl J Med*, **352**, 1992-2001.

58.   Bedrossian, C.W., Greenberg, S.D., Singer, D.B., Hansen, J.J. and Rosenberg, H.S. (1976) The lung in cystic fibrosis. A quantitative study including prevalence of pathologic findings among different age groups. *Hum Pathol.*, **7**, 195-204.

59.   Birket, S.E., Chu, K.K., Liu, L., Houser, G.H., Diephuis, B.J., Wilsterman, E.J., Dierksen, G., Mazur, M., Shastry, S., Li, Y., Watson, J.D., Smith, A.T., Schuster, B.S., Hanes, J., Grizzle, W.E., Sorscher, E.J., Tearney, G.J. and Rowe, S.M. (2014) A functional anatomic defect of the cystic fibrosis airway. *Am J Respir Crit Care Med*, **190**, 421-432.

60.   Shah, V.S., Ernst, S., Tang, X.X., Karp, P.H., Parker, C.P., Ostedgaard, L.S. and Welsh, M.J. (2016) Relationships among CFTR expression, HCO_3^- secretion, and host defense may inform gene- and cell-based cystic fibrosis therapies. *Proc Natl Acad Sci U S A*, **113**, 5382-5387.

61.   Cohen, T.S. and Prince, A. (2012) Cystic fibrosis: a mucosal immunodeficiency syndrome. *Nat Med*, **18**, 509-519.

62.   Khan, T.Z., Wagener, J.S., Martinez, J., Accurso, F.J. and Riches, D. (1995) Early pulmonary inflammation in infants with cystic fibrosis. *Am J Respir Crit Care Med*, **151**, 1075-1082.

63.   Konstan, M.W., Hilliard, K.A., Norvell, T.M. and Berger, M. (1994) Bronchoalveolar lavage findings in cystic fibrosis patients with stable, clinically mild lung disease suggest ongiong infection and inflammation. *Am J Respir Crit Care Med*, **150**, 448-454.

64.   Birrer, P., McElvaney, N.G., Ruderberg, C., Sommer, W., Liechti-Gallati, S., Kraemer, R., Hubbard, R. and Crystal, R.G. (1994) Protease-antiprotease imbalance in the lungs of children with cystic fibrosis. *Am J Respir Crit Care Med*, **150**, 207-213.

65.   Voynow, J.A., Fischer, B.M. and Zheng, S. (2008) Proteases and cystic fibrosis. *Int J Biochem Cell Biol*, **40**, 1238-1245.

66.   Muir, A., Soong, G., Sokol, S., Reddy, B., Gomez, M.I., Van Heeckeren, A. and Prince, A. (2004) Toll-like receptors in normal and cystic fibrosis airway epithelial cells. *Am J Respir Cell Mol Biol*, **30**, 777-783.

67.   Rottner, M., Kunzelmann, C., Mergey, M., Freyssinet, J.M. and Martinez, M.C. (2007) Exaggerated apoptosis and NF-kappaB activation in pancreatic and tracheal cystic fibrosis cells. *FASEB J*, **21**, 2939-2948.

68.   Durieu, I., Peyrol, S., Dominique, G., Bellon, G., Durand, D.V. and Pacheco, Y. (1998) Subepithelial fibrosis and degradation of the bronchial extracellular matrix in cystic fibrosis. *Am J Respir Crit Care Med*, **158**, 580-588.

69.   Hamutcu, R., Rowland, J.M., Horn, M.V., Kaminsky, C., MacLaughlin, E.F., Starnes, V.A. and Woo, M.S. (2002) Clinical Findings and Lung Pathology in Children with Cystic Fibrosis. *Am J Respir Crit Care Med*, **165**, 1172-1175.

70.   Hilliard, T.N., Regamey, N., Shute, J.K., Nicholson, A.G., Alton, E.W., Bush, A. and Davies, J.C. (2007) Airway remodelling in children with cystic fibrosis. *Thorax*, **62**, 1074-1080.

71.   Gohy, S.T., Hupin, C., Pilette, C. and Ladjemi, M.Z. (2016) Chronic inflammatory airway diseases: the central role of the epithelium revisited. *Clin Exp Allergy*, **46**, 529-542.

72.   Haley, K.J. and Drazen, J.M. (1998) Inflammation and airway function in asthma: what you see is not necessarily what you get. *Am J Respir Crit Care Med*, **157**, 1-3.

73.   Laitinen, L.A., A., L. and Haahtela, T. (1993) Airway mucosal inflammation even in patients with newly diagnosed asthma. *Am Rev Respir Dis*, **147**, 697-704.

74.   Ordoñez, C.L., Khashayar, R., Wong, H.H., Ferrando, R., Wu, R., Hyde, D.M., Hotchkiss, J.A., Zhang, Y., Novikov, A., Dolganov, G. and Fahy, J.V. (2001) Mild and moderate asthma is associated with airway goblet cell hyperplasia and abnormalities in mucin gene expression. *Am J Respir Crit Care Med*, **163**, 517-523.

75.   Vestbo, J., Hurd, S.S., Agusti, A.G., Jones, P.W., Vogelmeier, C., Anzueto, A., Barnes, P.J., Fabbri, L.M., Martinez, F.J., Nishimura, M., Stockley, R.A., Sin, D.D. and Rodriguez-Roisin, R. (2013) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*, **187**, 347-365.

76. Saetta, M., Turato, G., Baraldo, S., Zanin, A., Braccioni, F., Mapp, C.E., Maestrelli, P., Cavallesco, G., Papi, A. and Fabbri, L.M. (2000) Goblet cell hyperplasia and epithelial inflammation in peripheral airways of smokers with both symptoms of chronic bronchitis and chronic airflow limitation. *Am J Respir Crit Care Med*, **161**, 1016-1021.

77. Kim, V., Rogers, T.J. and Criner, G.J. (2008) New concepts in the pathobiology of chronic obstructive pulmonary disease. *Proc Am Thorac Soc*, **5**, 478-485.

78. Hahn, S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol*, **11**, 394-402.

79. Jonkers, I. and Lis, J.T. (2015) Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol*, **16**, 167-177.

80. Zhou, V.W., Goren, A. and Bernstein, B.E. (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*, **12**, 7-18.

81. Venkatesh, S. and Workman, J.L. (2015) Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol*, **16**, 178-189.

82. Encode Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., Kuehn, M.S., Taylor, C.M., Neph, S., Koch, C.M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J.A., Andrews, R.M., Flicek, P., Boyle, P.J., Cao, H., Carter, N.P., Clelland, G.K., Davis, S., Day, N., Dhami, P., Dillon, S.C., Dorschner, M.O., Fiegler, H., Giresi, P.G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K.D., Johnson, B.E., Johnson, E.M., Frum, T.T., Rosenzweig, E.R., Karnani, N., Lee, K., Lefebvre, G.C., Navas, P.A., Neri, F., Parker, S.C., Sabo, P.J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F.S., Dekker, J., Lieb, J.D., Tullius, T.D., Crawford, G.E., Sunyaev, S., Noble, W.S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I.L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H.A., Sekinger, E.A., Lagarde, J., Abril, J.F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J.S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M.C., Thomas, D.J., Weirauch, M.T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K.G., Sung, W.K., Ooi, H.S., Chiu, K.P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M.L., Valencia, A., Choo, S.W., Choo, C.Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T.G., Brown, J.B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C.N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J.S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R.M., Rogers, J., Stadler, P.F., Lowe, T.M., Wei, C.L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S.E., Fu, Y., Green, E.D., Karaoz, U., Siepel, A., Taylor, J., Liefer, L.A., Wetterstrand, K.A., Good, P.J., Feingold, E.A., Guyer, M.S., Cooper, G.M., Asimenos, G., Dewey, C.N., Hou, M., Nikolaev, S., Montoya-Burgos, J.I., Loytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N.R., Holmes, I., Mullikin, J.C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W.J.,

Stone, E.A., Program, N.C.S., Baylor College of Medicine Human Genome Sequencing, C., Washington University Genome Sequencing, C., Broad, I., Children's Hospital Oakland Research, I., Batzoglou, S., Goldman, N., Hardison, R.C., Haussler, D., Miller, W., Sidow, A., Trinklein, N.D., Zhang, Z.D., Barrera, L., Stuart, R., King, D.C., Ameur, A., Enroth, S., Bieda, M.C., Kim, J., Bhinge, A.A., Jiang, N., Liu, J., Yao, F., Vega, V.B., Lee, C.W., Ng, P., Shahab, A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M.J., Inman, D., Singer, M.A., Richmond, T.A., Munn, K.J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J.C., Couttet, P., Bruce, A.W., Dovey, O.M., Ellis, P.D., Langford, C.F., Nix, D.A., Euskirchen, G., Hartman, S., Urban, A.E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T.H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C.K., Rosenfeld, M.G., Aldred, S.F., Cooper, S.J., Halees, A., Lin, J.M., Shulha, H.P., Zhang, X., Xu, M., Haidar, J.N., Yu, Y., Ruan, Y., Iyer, V.R., Green, R.D., Wadelius, C., Farnham, P.J., Ren, B., Harte, R.A., Hinrichs, A.S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A.S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R.M., Karolchik, D., Armengol, L., Bird, C.P., de Bakker, P.I., Kern, A.D., Lopez-Bigas, N., Martin, J.D., Stranger, B.E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrimsdottir, I.B., Huppert, J., Zody, M.C., Abecasis, G.R., Estivill, X., Bouffard, G.G., Guan, X., Hansen, N.F., Idol, J.R., Maduro, V.V., Maskeri, B., McDowell, J.C., Park, M., Thomas, P.J., Young, A.C., Blakesley, R.W., Muzny, D.M., Sodergren, E., Wheeler, D.A., Worley, K.C., Jiang, H., Weinstock, G.M., Gibbs, R.A., Graves, T., Fulton, R., Mardis, E.R., Wilson, R.K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D.B., Chang, J.L., Lindblad-Toh, K., Lander, E.S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B. and de Jong, P.J. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799-816.

83.   Encode Project Consortium, Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B.R., Landt, S.G., Lee, B.K., Pauli, F., Rosenbloom, K.R., Sabo, P., Safi, A., Sanyal, A., Shoresh, N., Simon, J.M., Song, L., Trinklein, N.D., Altshuler, R.C., Birney, E., Brown, J.B., Cheng, C., Djebali, S., Dong, X., Dunham, I., Ernst, J., Furey, T.S., Gerstein, M., Giardine, B., Greven, M., Hardison, R.C., Harris, R.S., Herrero, J., Hoffman, M.M., Iyer, S., Kelllis, M., Khatun, J., Kheradpour, P., Kundaje, A., Lassman, T., Li, Q., Lin, X., Marinov, G.K., Merkel, A., Mortazavi, A., Parker, S.C., Reddy, T.E., Rozowsky, J., Schlesinger, F., Thurman, R.E., Wang, J., Ward, L.D., Whitfield, T.W., Wilder, S.P., Wu, W., Xi, H.S., Yip, K.Y., Zhuang, J., Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., Snyder, M., Pazin, M.J., Lowdon, R.F., Dillon, L.A., Adams, L.B., Kelly, C.J., Zhang, J., Wexler, J.R., Green, E.D., Good, P.J., Feingold, E.A., Bernstein, B.E., Birney, E., Crawford, G.E., Dekker, J., Elinitski, L., Farnham, P.J., Gerstein, M., Giddings, M.C., Gingeras, T.R., Green, E.D., Guigo, R., Hardison, R.C., Hubbard, T.J., Kellis, M., Kent, W.J., Lieb, J.D., Margulies, E.H., Myers, R.M., Snyder, M., Starnatoyannopoulos, J.A., Tennebaum, S.A., Weng, Z., White, K.P., Wold, B., Khatun, J., Yu, Y., Wrobel, J., Risk, B.A., Gunawardena, H.P., Kuiper, H.C., Maier, C.W., Xie, L., Chen, X., Giddings, M.C., Bernstein, B.E., Epstein, C.B., Shoresh, N., Ernst, J., Kheradpour, P., Mikkelsen, T.S., Gillespie, S., Goren, A., Ram, O., Zhang, X., Wang, L., Issner, R., Coyne, M.J., Durham, T., Ku, M., Truong, T., Ward, L.D., Altshuler, R.C., Eaton, M.L., Kellis, M., Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Roder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Batut, P., Bell, I., Bell, K., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H.P., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Li, G., Luo, O.J., Park, E., Preall, J.B., Presaud, K., Ribeca, P., Risk, B.A., Robyr, D., Ruan, X., Sammeth, M., Sandu, K.S., Schaeffer, L., See, L.H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard,

T.J., Reymond, A., Antonarakis, S.E., Hannon, G.J., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigo, R., Gingeras, T.R., Rosenbloom, K.R., Sloan, C.A., Learned, K., Malladi, V.S., Wong, M.C., Barber, G.P., Cline, M.S., Dreszer, T.R., Heitner, S.G., Karolchik, D., Kent, W.J., Kirkup, V.M., Meyer, L.R., Long, J.C., Maddren, M., Raney, B.J., Furey, T.S., Song, L., Grasfeder, L.L., Giresi, P.G., Lee, B.K., Battenhouse, A., Sheffield, N.C., Simon, J.M., Showers, K.A., Safi, A., London, D., Bhinge, A.A., Shestak, C., Schaner, M.R., Kim, S.K., Zhang, Z.Z., Mieczkowski, P.A., Mieczkowska, J.O., Liu, Z., McDaniell, R.M., Ni, Y., Rashid, N.U., Kim, M.J., Adar, S., Zhang, Z., Wang, T., Winter, D., Keefe, D., Birney, E., Iyer, V.R., Lieb, J.D., Crawford, G.E., Li, G., Sandhu, K.S., Zheng, M., Wang, P., Luo, O.J., Shahab, A., Fullwood, M.J., Ruan, X., Ruan, Y., Myers, R.M., Pauli, F., Williams, B.A., Gertz, J., Marinov, G.K., Reddy, T.E., Vielmetter, J., Partridge, E.C., Trout, D., Varley, K.E., Gasper, C., Bansal, A., Pepke, S., Jain, P., Amrhein, H., Bowling, K.M., Anaya, M., Cross, M.K., King, B., Muratet, M.A., Antoshechkin, I., Newberry, K.M., McCue, K., Nesmith, A.S., Fisher-Aylor, K.I., Pusey, B., DeSalvo, G., Parker, S.L., Balasubramanian, S., Davis, N.S., Meadows, S.K., Eggleston, T., Gunter, C., Newberry, J.S., Levy, S.E., Absher, D.M., Mortazavi, A., Wong, W.H., Wold, B., Blow, M.J., Visel, A., Pennachio, L.A., Elnitski, L., Margulies, E.H., Parker, S.C., Petrykowska, H.M., Abyzov, A., Aken, B., Barrell, D., Barson, G., Berry, A., Bignell, A., Boychenko, V., Bussotti, G., Chrast, J., Davidson, C., Derrien, T., Despacio-Reyes, G., Diekhans, M., Ezkurdia, I., Frankish, A., Gilbert, J., Gonzalez, J.M., Griffiths, E., Harte, R., Hendrix, D.A., Howald, C., Hunt, T., Jungreis, I., Kay, M., Khurana, E., Kokocinski, F., Leng, J., Lin, M.F., Loveland, J., Lu, Z., Manthravadi, D., Mariotti, M., Mudge, J., Mukherjee, G., Notredame, C., Pei, B., Rodriguez, J.M., Saunders, G., Sboner, A., Searle, S., Sisu, C., Snow, C., Steward, C., Tanzer, A., Tapanari, E., Tress, M.L., van Baren, M.J., Walters, N., Washieti, S., Wilming, L., Zadissa, A., Zhengdong, Z., Brent, M., Haussler, D., Kellis, M., Valencia, A., Gerstein, M., Raymond, A., Guigo, R., Harrow, J., Hubbard, T.J., Landt, S.G., Frietze, S., Abyzov, A., Addleman, N., Alexander, R.P., Auerbach, R.K., Balasubramanian, S., Bettinger, K., Bhardwaj, N., Boyle, A.P., Cao, A.R., Cayting, P., Charos, A., Cheng, Y., Cheng, C., Eastman, C., Euskirchen, G., Fleming, J.D., Grubert, F., Habegger, L., Hariharan, M., Harmanci, A., Iyenger, S., Jin, V.X., Karczewski, K.J., Kasowski, M., Lacroute, P., Lam, H., Larnarre-Vincent, N., Leng, J., Lian, J., Lindahl-Allen, M., Min, R., Miotto, B., Monahan, H., Moqtaderi, Z., Mu, X.J., O'Geen, H., Ouyang, Z., Patacsil, D., Pei, B., Raha, D., Ramirez, L., Reed, B., Rozowsky, J., Sboner, A., Shi, M., Sisu, C., Slifer, T., Witt, H., Wu, L., Xu, X., Yan, K.K., Yang, X., Yip, K.Y., Zhang, Z., Struhl, K., Weissman, S.M., Gerstein, M., Farnham, P.J., Snyder, M., Tenebaum, S.A., Penalva, L.O., Doyle, F., Karmakar, S., Landt, S.G., Bhanvadia, R.R., Choudhury, A., Domanus, M., Ma, L., Moran, J., Patacsil, D., Slifer, T., Victorsen, A., Yang, X., Snyder, M., White, K.P., Auer, T., Centarin, L., Eichenlaub, M., Gruhl, F., Heerman, S., Hoeckendorf, B., Inoue, D., Kellner, T., Kirchmaier, S., Mueller, C., Reinhardt, R., Schertel, L., Schneider, S., Sinn, R., Wittbrodt, B., Wittbrodt, J., Weng, Z., Whitfield, T.W., Wang, J., Collins, P.J., Aldred, S.F., Trinklein, N.D., Partridge, E.C., Myers, R.M., Dekker, J., Jain, G., Lajoie, B.R., Sanyal, A., Balasundaram, G., Bates, D.L., Byron, R., Canfield, T.K., Diegel, M.J., Dunn, D., Ebersol, A.K., Ebersol, A.K., Frum, T., Garg, K., Gist, E., Hansen, R.S., Boatman, L., Haugen, E., Humbert, R., Jain, G., Johnson, A.K., Johnson, E.M., Kutyavin, T.M., Lajoie, B.R., Lee, K., Lotakis, D., Maurano, M.T., Neph, S.J., Neri, F.V., Nguyen, E.D., Qu, H., Reynolds, A.P., Roach, V., Rynes, E., Sabo, P., Sanchez, M.E., Sandstrom, R.S., Sanyal, A., Shafer, A.O., Stergachis, A.B., Thomas, S., Thurman, R.E., Vernot, B., Vierstra, J., Vong, S., Wang, H., Weaver, M.A., Yan, Y., Zhang, M., Akey, J.A., Bender, M., Dorschner, M.O., Groudine, M., MacCoss, M.J., Navas, P., Stamatoyannopoulos, G., Kaul, R., Dekker, J., Stamatoyannopoulos, J.A., Dunham, I., Beal, K., Brazma, A., Flicek, P., Herrero, J., Johnson, N., Keefe, D., Lukk, M., Luscombe, N.M., Sobral, D., Vaquerizas, J.M., Wilder, S.P., Batzoglou, S., Sidow, A., Hussami, N., Kyriazopoulou-Panagiotopoulou, S., Libbrecht, M.W., Schaub, M.A., Kundaje, A., Hardison, R.C., Miller, W., Giardine, B., Harris, R.S., Wu, W., Bickel, P.J., Banfai, B., Boley, N.P., Brown, J.B., Huang, H., Li, Q., Li, J.J., Noble, W.S., Bilmes, J.A., Buske, O.J., Hoffman, M.M., Sahu, A.O., Kharchenko, P.V., Park, P.J., Baker, D., Taylor, J., Weng, Z., Iyer, S., Dong, X., Greven, M., Lin, X., Wang, J., Xi,

H.S., Zhuang, J., Gerstein, M., Alexander, R.P., Balasubramanian, S., Cheng, C., Harmanci, A., Lochovsky, L., Min, R., Mu, X.J., Rozowsky, J., Yan, K.K., Yip, K.Y. and Birney, E. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.

84.   Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S. and Crawford, G.E. (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell*, **132**, 311-322.

85.   Arvey, A., Agius, P., Noble, W.S. and Leslie, C. (2012) Sequence and chromatin determinants of cell-type-specific transcription factor binding. *Genome Res*, **22**, 1723-1734.

86.   Kang, K., Robinson, G.W. and Hennighausen, L. (2013) Comprehensive meta-analysis of Signal Transducers and Activators of Transcription (STAT) genomic binding patterns discerns cell-specific cis-regulatory modules. *BMC Genomics*, **14**, 4.

87.   Bischof, J.M., Ott, C.J., Leir, S.H., Gosalia, N., Song, L., London, D., Furey, T.S., Cotton, C.U., Crawford, G.E. and Harris, A. (2012) A genome-wide analysis of open chromatin in human tracheal epithelial cells reveals novel candidate regulatory elements for lung function. *Thorax*, **67**, 385-391.

88.   Smale, S.T. and Kadonaga, J.T. (2003) The RNA polymerase II core promoter. *Annu Rev Biochem*, **72**, 449-479.

89.   Pundhir, S., Bagger, F.O., Lauridsen, F.B., Rapin, N. and Porse, B.T. (2016) Peak-valley-peak pattern of histone modifications delineates active regulatory elements and their directionality. *Nucleic Acids Res*, **44**, 4037-4051.

90.   Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E. and Ren, B. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*, **39**, 311-318.

91.   Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D. and Ren, B. (2005) A high-resolution map of active promoters in the human genome. *Nature*, **436**, 876-880.

92.   Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., Forrest, A.R., Alkema, W.B., Tan, S.L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S.M., Wells, C.A., Orlando, V., Wahlestedt, C., Liu, E.T., Harbers, M., Kawai, J., Bajic, V.B., Hume, D.A. and Hayashizaki, Y.

(2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*, **38**, 626-635.

93. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., Ching, K.A., Antosiewicz-Bourget, J.E., Liu, H., Zhang, X., Green, R.D., Lobanenkov, V.V., Stewart, R., Thomson, J.A., Crawford, G.E., Kellis, M. and Ren, B. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, **459**, 108-112.

94. Shlyueva, D., Stampfel, G. and Stark, A. (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*, **15**, 272-286.

95. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A. and Jaenisch, R. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*, **107**, 21931-21936.

96. Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F. and Fraser, P. (2002) Long-range chromatin regulatory interactions in vivo. *Nat Genet*, **32**, 623-626.

97. Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. and Laat, W. (2002) Looping and Interaction between Hypersensitive Sites in the Active beta-globin Locus. *Molecular Cell*, **10**, 1453-1465.

98. Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J. and Blobel, G.A. (2005) Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell*, **17**, 453-462.

99. Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G. and Merkenschlager, M. (2009) Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, **460**, 410-413.

100. Zuin, J., Dixon, J.R., van der Reijdan, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., Grosveld, F.G., Ren, B. and Wendt, K.S. (2014) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A*, **111**, 996-1001.

101. Hou, C., Zhao, H., Tanimoto, K. and Dean, A. (2008) CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A*, **105**, 20398-20403.

102. Spitz, F. and Furlong, E.E. (2012) Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*, **13**, 613-626.

103. Xu, J., Watts, J.A., Pope, S.D., Gadue, P., Kamps, M., Plath, K., Zaret, K.S. and Smale, S.T. (2009) Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev*, **23**, 2824-2838.

104. Heinz, S., Romanoski, C.E., Benner, C. and Glass, C.K. (2015) The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*, **16**, 144-154.

105. Kerschner, J.L., Gosalia, N., Leir, S.H. and Harris, A. (2014) Chromatin remodeling mediated by the FOXA1/A2 transcription factors activates CFTR expression in intestinal epithelial cells. *Epigenetics*, **9**, 557-565.

106. Zaret, K.S. and Carroll, J.S. (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*, **25**, 2227-2241.

107. Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J. and Mann, R.S. (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell*, **147**, 1270-1282.

108. Boyle, A.P., Song, L., Lee, B.K., London, D., Keefe, D., Birney, E., Iyer, V.R., Crawford, G.E. and Furey, T.S. (2011) High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res*, **21**, 456-464.

109. Crawford, G.E., Davis, S., Scacheri, P.C., Renaud, G., Halawi, M.J., Erdos, M.R., Green, R., Meltzer, P.S., Wolfsberg, T.G. and Collins, F.S. (2006) DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat Methods*, **3**, 503-509.

110. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823-837.

111. Johnson, D.S., Mortazavi, A., Myers, R.M. and Wold, B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science*, **316**, 1497-1502.

112. Zeng, P.-Y., Vakoc, C., Chen, Z.-C., Blobel, G. and Berger, S. (2006) In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *BioTechniques*, **41**, 694-698.

113. O'Neill, L. and Turner, B.M. (2003) Immunoprecipitation of native chromatin: NChIP. *Methods*, **31**, 76-82.

114. Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J. and Odom, D.T. (2009) ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods*, **48**, 240-248.

115. Park, P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, **10**, 669-680.

116. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**, R25.

117. Pepke, S., Wold, B. and Mortazavi, A. (2009) Computation for ChIP-seq and RNA-seq studies. *Nature Methods Supllement*, **6** 522-532.

118. Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M. and Gerstein, M.B. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol*, **27**, 66-75.

119. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K.I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A.J., Hoffman, M.M., Iyer, V.R., Jung, Y.L., Karmakar, S., Kellis, M., Kharchenko, P.V., Li, Q., Liu, T., Liu, X.S., Ma, L., Milosavljevic, A., Myers, R.M., Park, P.J., Pazin, M.J., Perry, M.D., Raha, D., Reddy, T.E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J.A., Tolstorukov, M.Y., White, K.P., Xi, S., Farnham, P.J., Lieb, J.D., Wold, B.J. and Snyder, M. (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*, **22**, 1813-1831.

120. Li, Q., Brown, J.B., Huang, H. and Bickel, P.J. (2011) Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics*, **5**, 1752-1779.

121. Browne, J.A., Harris, A. and Leir, S.H. (2014) An optimized protocol for isolating primary epithelial cell chromatin for ChIP. *PLoS One*, **9**, e100099.

122. Furey, T.S. (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet*, **13**, 840-852.

123. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, **38**, 576-589.

124. Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M. and Wong, W.H. (2008) An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol*, **26**, 1293-1300.

125. Zanconato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., Rosato, A., Bicciato, S., Cordenonsi, M. and Piccolo, S. (2015) Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nat Cell Biol*, **17**, 1218-1227.

126. Zhao, J., Li, X., Guo, M., Yu, J. and Yan, C. (2016) The common stress responsive transcription factor ATF3 binds genomic sites enriched with p300 and H3K27ac for transcriptional regulation. *BMC Genomics*, **17**, 335.

127. Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., Tang, Q., Meyer, C.A., Zhang, Y. and Liu, X.S. (2013) Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nat Protoc*, **8**, 2502-2515.

128. Wilhelm, B.T., Marguerat, S., Goodhead, I. and Bahler, J. (2010) Defining transcribed regions using RNA-seq. *Nat Protoc*, **5**, 255-266.

129. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, **7**, 562-578.

130. Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L. and Pachter, L. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*, **31**, 46-53.

131. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*, **28**, 511-515.

132. Shu, S., Lin, C.Y., He, H.H., Witwicki, R.M., Tabassum, D.P., Roberts, J.M., Janiszewska, M., Huh, S.J., Liang, Y., Ryan, J., Doherty, E., Mohammed, H., Guo, H., Stover, D.G., Ekram, M.B., Peluffo, G., Brown, J., D'Santos, C., Krop, I.E., Dillon, D., McKeown, M., Ott, C., Qi, J., Ni, M., Rao, P.K., Duarte, M., Wu, S.Y., Chiang, C.M., Anders, L., Young, R.A., Winer, E.P., Letai, A., Barry, W.T., Carroll, J.S., Long, H.W., Brown, M., Liu, X.S., Meyer, C.A., Bradner, J.E. and Polyak, K. (2016) Response and resistance to BET bromodomain inhibitors in triple-negative breast cancer. *Nature*, **529**, 413-417.

133. Gao, X.G., Vockley, C.M., Pauli, F., Newberry, K.M., Xue, Y., Randell, S.H., Reddy, T.E. and Hogan, B.L. (2013) Evidence for multiple roles for grainyheadlike 2 in the establishment and maintenance of human mucociliary airway epithelium. *Proc Natl Acad Sci U S A*, **110**, 9356-9361.

134. Majewski, J. and Pastinen, T. (2011) The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet*, **27**, 72-79.

135.   Hawkins, R.D., Hon, G.C. and Ren, B. (2010) Next-generation genomics: an integrative approach. *Nat Rev Genet*, **11**, 476-486.

136.   Herriges, M. and Morrisey, E.E. (2014) Lung development: orchestrating the generation and regeneration of a complex organ. *Development*, **141**, 502-513.

137.   Mollard, R., Ghyselinck, N.B., Wendling, O., Chambon, P. and Mark, M. (2000) Stage-dependent responses of the developing lung to retinoic acid signaling. *Int J Dev Biol.*, **44**, 457-462.

138.   Yang, H., Lu, M.M., Zhang, L., Whitsett, J.A. and Morrisey, E.E. (2002) GATA6 regulates differentiation of the distal lung epithelium. *Development*, **129**, 2233-2246.

139.   Zhang, Y., Rath, N., Hannenhalli, S., Wang, Z., Cappola, T., Kimura, S., Atochina-Vasserman, E., Lu, M.M., Beers, M.F. and Morrisey, E.E. (2007) GATA and Nkx factors synergistically regulate tissue-specific gene expression and development in vivo. *Development*, **134**, 189-198.

140.   Zhang, Y., Goss, A.M., Cohen, E.D., Kadzik, R., Lepore, J.J., Muthukumaraswamy, K., Yang, J., DeMayo, F.J., Whitsett, J.A., Parmacek, M.S. and Morrisey, E.E. (2008) A Gata6-Wnt pathway required for epithelial stem cell development and airway regeneration. *Nat Genet*, **40**, 862-870.

141.   Yin, Z., Gonzalez, D., Kolla, V., Rath, N., Zhang, Y., Lu, M.M., Kimura, S., Ballard, P.L., Beers, M.F., Epstein, J.A. and Morrisey, E.E. (2006) Hop functions downstream of Nkx2.1 and GATA6 to mediate HDAC- dependent negative regulation of pulmonary gene expression. *Am J Physiol Lung Cell Mol Physiol*, **291**, L191-L199.

142.   Wan, H., Luo, F., Wert, S.E., Zhang, L., Xu, Y., Ikegami, M., Maeda, Y., Bell, S.M. and Whitsett, J.A. (2008) Kruppel-like factor 5 is required for perinatal lung morphogenesis and function. *Development*, **135**, 2563-2572.

143.   Wan, H., Dingle, S., Xu, Y., Besnard, V., Kaestner, K.H., Ang, S.L., Wert, S., Stahlman, M.T. and Whitsett, J.A. (2005) Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem*, **280**, 13809-13816.

144.   Wan, H., Kaestner, K.H., Ang, S.L., Ikegami, M., Finkelman, F.D., Stahlman, M.T., Fulkerson, P.C., Rothenberg, M.E. and Whitsett, J.A. (2004) Foxa2 regulates alveolarization and goblet cell hyperplasia. *Development*, **131**, 953-964.

145.   Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M. and Zaret, K.S. (2002) Opening of Compacted Chromatin by Early Developmental Transcrpton Factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, **9**, 279-289.

146. Kerschner, J.L. and Harris, A. (2012) Transcriptional networks driving enhancer function in the CFTR gene. *Biochem J*, **446**, 203-212.

147. Metzger, D.E., Stahlman, M.T. and Shannon, J.M. (2008) Misexpression of ELF5 disrupts lung branching and inhibits epithelial differentiation. *Dev Biol*, **320**, 149-160.

148. Korfhagen, T.R., Kitzmiller, J., Chen, G., Sridharan, A., Haitchi, H.-M., Hegde, R.S., Divanovic, S., Karp, C.L. and Whitsett, J.A. (2012) SAM-pointed domain ETS factor mediates epithelial cell-intrinsic innate immune signaling during airway mucous metaplasia. *Proc Natl Acad Sci U S A*, **109**, 16630-16635.

149. Chen, G., Korfhagen, T.R., Xu, Y., Kitzmiller, J., Wert, S., Maeda, Y., Gregorieff, A., Clevers, H. and Whitsett, J.A. (2009) SPDEF is required for mouse pulmonary goblet cell differentiation and regulates a network of genes associated with mucus production. *J Clin Invest*, **119**, 2914-2924.

150. Rajavelu, P., Chen, G., Xu, Y., Kitzmiller, J.A., Korfhagen, T.R. and Whitsett, J.A. (2015) Airway epithelial SPDEF integrates goblet cell differentiation and pulmonary Th2 inflammation. *J Clin Invest*, **125**, 2021-2031.

151. Hayden, M.S. and Ghosh, S. (2004) Signaling to NF-κB. *Genes Dev*, **18**, 2195-3334.

152. Tang, H., Sun, Y., Shi, Z., Huang, H., Fang, Z., Chen, J., Xiu, Q. and Li, B. (2013) YKL-40 induces IL-8 expression from bronchial epithelium via MAPK (JNK and ERK) and NF-kappaB pathways, causing bronchial smooth muscle proliferation and migration. *J Immunol*, **190**, 438-446.

153. Cheng, D.s., Han, W., Chen, S.M., Sherrill, T.P., Chont, M., Park, G.Y., Sheller, J.R., Polosukhin, V.V., Christman, J.W., Yull, F.E. and Blackwell, T.S. (2007) Airway Epithelium Controls Lung Inflammation and Injury through the NF- B Pathway. *J Immunol*, **178**, 6504-6513.

154. Venkatakrishnan, A., Stecenko, A.A., King, G., Blackwell, T.R., Brigham, K.L., Christman, J.W. and Blackwell, T.S. (2000) Exaggerated activation of nuclear factor-κB and altered IκB-β processing in cystic fibrosis bronchial epithelila cells. *Am J Respir Cell Mol Biol*, **23**, 396-403.

155. Ott, C.J., Blackledge, N.P., Kerschner, J.L., Leir, S.H., Crawford, G.E., Cotton, C.U. and Harris, A. (2009) Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus. *Proc Natl Acad Sci U S A*, **106**, 19934-19939.

156. Zhang, Z., Leir, S.H. and Harris, A. (2013) Immune mediators regulate *CFTR* expression through a bifunctional airway-selective enhancer. *Mol Cell Biol*, **33**, 2843-2853.

157. Zhang, Z., Leir, S.H. and Harris, A. (2015) Oxidative stress regulates CFTR gene expression in human airway epithelial cells through a distal antioxidant response element. *Am J Respir Cell Mol Biol*, **52**, 387-396.

158. Yang, R., Kerschner, J.L., Gosalia, N., Neems, D., Gorsic, L.K., Safi, A., Crawford, G.E., Kosak, S.T., Leir, S.H. and Harris, A. (2016) Differential contribution of cis-regulatory elements to higher order chromatin structure and expression of the CFTR locus. *Nucleic Acids Res*, **44**, 3082-3094.

159. Gosalia, N. and Harris, A. (2015) Chromatin Dynamics in the Regulation of CFTR Expression. *Genes (Basel)*, **6**, 543-558.

160. Tugores, A., Le, J., Sorokina, I., Snijders, A.J., Duyao, M., Reddy, P.S., Carlee, L., Ronshaugen, M., Mushegian, A., Watanaskul, T., Chu, S., Buckler, A., Emtage, S. and McCormick, M.K. (2001) The epithelium-specific ETS protein EHF/ESE-3 is a context-dependent transcriptional repressor downstream of MAPK signaling cascades. *J Biol Chem*, **276**, 20397-20406.

161. Silverman, E.S., Baron, R.M., Palmer, L.J., Le, L., Hallock, A., Subramaniam, V., Riese, R.J., McKenna, M.D., Gu, X., Libermann, T.A., Tugores, A., Haley, K.J., Shore, S., Drazen, J.M. and Weiss, S.T. (2002) Constitutive and Cytokine-Induced Expression of the ETS Transcription Factor ESE-3 in the Lung. *Am J Respir Cell Mol Biol*, **27**, 697-704.

162. Donaldson, L.W., Petersen, J.M., Graves, B.J. and McIntosh, L.P. (1996) Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J*, **15**, 125-134.

163. Hollenhorst, P.C., McIntosh, L.P. and Graves, B.J. (2011) Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem*, **80**, 437-471.

164. Bassuk, A.G. and Leiden, J.M. (1995) A direct physical association between ETS and AP-1 transcription factors in normal human T cells. *Immunity*, **3**, 223-237.

165. Sharrocks, A.D. (2001) The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol*, **2**, 827-837.

166. Tootle, T.L. and Rebay, I. (2005) Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. *Bioessays*, **27**, 285-298.

167. Yang, S.-H., Jaffray, E., Hay, R.T. and Sharrocks, A.D. (2003) Dynamic Interplay of the SUMO and ERK Pathways in Regulating Elk-1 Transcriptional Activity. *Molecular Cell*, **12**, 63-74.

168. Kas, K., Finger, E., Grall, F., Gu, X., Akbarali, Y., Boltax, J., Weiss, A., Oettgen, P., Kapeller, R. and Libermann, T.A. (2000) ESE-3, a Novel Member of an Epithelium-specific Ets Transcription Factor Subfamily, Demonstrates Different Target Gene Specificity from ESE-1. *J Biol Chem*, **275**, 2986-2998.

169. Sprater, F., Hovden, A.O. and Appel, S. (2012) Expression of ESE-3 isoforms in immunogenic and tolerogenic human monocyte-derived dendritic cells. *PLoS One*, **7**, e49577.

170. Kleinbaum, L.A., Duggan, C., Ferreira, E., Coffey, G.P., Butticé, G. and Burton, F.H. (1999) Human chromosomal localization, tissue/tumor expression, and regulatory function of the ets family gene EHF. *Biochem Biophys Res Commun*, **264**, 119-126.

171. Taniue, K., Oda, T., Hayashi, T., Okuno, M. and Akiyama, T. (2011) A member of the ETS family, EHF, and the ATPase RUVBL1 inhibit p53-mediated apoptosis. *EMBO Rep*, **12**, 682-689.

172. Kunderfranco, P., Mello-Grand, M., Cangemi, R., Pellini, S., Mensah, A., Albertini, V., Malek, A., Chiorino, G., Catapano, C.V. and Carbone, G.M. (2010) ETS transcription factors control transcription of EZH2 and epigenetic silencing of the tumor suppressor gene Nkx3.1 in prostate cancer. *PLoS One*, **5**, e10547.

173. Cangemi, R., Mensah, A., Albertini, V., Jain, A., Mello-Grand, M., Chiorino, G., Catapano, C.V. and Carbone, G.M. (2008) Reduced expression and tumor suppressor function of the ETS transcription factor ESE-3 in prostate cancer. *Oncogene*, **27**, 2877-2885.

174. Albino, D., Longoni, N., Curti, L., Mello-Grand, M., Pinton, S., Civenni, G., Thalmann, G., D'Ambrosio, G., Sarti, M., Sessa, F., Chiorino, G., Catapano, C.V. and Carbone, G.M. (2012) ESE3/EHF controls epithelial cell differentiation and its loss leads to prostate tumors with mesenchymal and stem-like features. *Cancer Res*, **72**, 2889-2900.

175. Stephens, D.N., Herndon Klein, R., Salmans, M.L., Gordon, W., Ho, H. and Andersen, B. (2013) The Ets transcription factor EHF as a regulator of cornea epithelial cell identity. *J Biol Chem*.

176. Wu, J., Duan, R., Cao, H., Field, D., Newnham, C.M., Koehler, D.R., Zamel, N., Pritchard, M.A., Hertzog, P., Post, M., Tanswell, A.K. and Hu, J. (2008) Regulation of epithelium-specific Ets-like factors ESE-1 and ESE-3 in airway epithelial cells: potential roles in airway inflammation. *Cell Res*, **18**, 649-663.

177. Plotnik, J.P., Budka, J.A., Ferris, M.W. and Hollenhorst, P.C. (2014) ETS1 is a genome-wide effector of RAS/ERK signaling in epithelial cells. *Nucleic Acids Res*, **42**, 11928-11940.

178. Hollenhorst, P.C., Ferris, M.W., Hull, M.A., Chae, H., Kim, S. and Graves, B.J. (2011) Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev*, **25**, 2147-2157.

179. Shaulian, E. and Karin, M. (2002) AP-1 as a regulator of cell life and death. *Nat Cell Biol*, **4**, E131-136.

180. Shaulian, E. and Karin, M. (2001) AP-1 in cell proliferation and survival. *Oncogene*, **20**, 2390-2400.

181. Reddy, S.P. and Mossman, B.T. (2002) Role and regulation of activator protein-1 in toxicant-induced responses of the lung. *Am J Physiol Lung Cell Mol Physiol*, **283**, L1161-L1178.

182. Mossman, B.T., Lounsbury, K.M. and Reddy, S.P. (2006) Oxidants and signaling by mitogen-activated protein kinases in lung epithelium. *Am J Respir Cell Mol Biol*, **34**, 666-669.

183. Verhaeghe, C., Remouchamps, C., Hennuy, B., Vanderplasschen, A., Chariot, A., Tabruyn, S.P., Oury, C. and Bours, V. (2007) Role of IKK and ERK pathways in intrinsic inflammation of cystic fibrosis airways. *Biochem Pharmacol*, **73**, 1982-1994.

184. Cystic Fibrosis Mutation Database. *Updated: April 25,2011*, http://www.genet.sickkids.on.ca/StatisticsPage.html.

185. The Cystic Fibrosis Genotype-Phenotype Consortium. (1993) Correlation between genotype and phenotype in patients with cystic fibrosis. *N Engl J Med*, **329**, 1308-1313.

186. Vanscoy, L.L., Blackman, S.M., Collaco, J.M., Bowers, A., Lai, T., Naughton, K., Algire, M., McWilliams, R., Beck, S., Hoover-Fong, J., Hamosh, A., Cutler, D. and Cutting, G.R. (2007) Heritability of lung disease severity in cystic fibrosis. *Am J Respir Crit Care Med*, **175**, 1036-1043.

187. Wright, F.A., Strug, L.J., Doshi, V.K., Commander, C.W., Blackman, S.M., Sun, L., Berthiaume, Y., Cutler, D., Cojocaru, A., Collaco, J.M., Corey, M., Dorfman, R., Goddard, K., Green, D., Kent, J.W., Jr., Lange, E.M., Lee, S., Li, W., Luo, J., Mayhew, G.M., Naughton, K.M., Pace, R.G., Pare, P., Rommens, J.M., Sandford, A., Stonebraker, J.R., Sun, W., Taylor, C., Vanscoy, L.L., Zou, F., Blangero, J., Zielenski, J., O'Neal, W.K., Drumm, M.L., Durie, P.R., Knowles, M.R. and Cutting, G.R. (2011) Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nat Genet*, **43**, 539-546.

188. Kerem, E., Reisman, J., Corey, M., Canny, G.J. and Levison, H. (1992) Prediction of mortality in patients with cystic fibrosis. *N Engl J Med*, **326**, 1187-1191.

189.   Corvol, H., Blackman, S.M., Boelle, P.Y., Gallins, P.J., Pace, R.G., Stonebraker, J.R., Accurso, F.J., Clement, A., Collaco, J.M., Dang, H., Dang, A.T., Franca, A., Gong, J., Guillot, L., Keenan, K., Li, W., Lin, F., Patrone, M.V., Raraigh, K.S., Sun, L., Zhou, Y.H., O'Neal, W.K., Sontag, M.K., Levy, H., Durie, P.R., Rommens, J.M., Drumm, M.L., Wright, F.A., Strug, L.J., Cutting, G.R. and Knowles, M.R. (2015) Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nat Commun*, **6**, 8382.

190.   Shen, B.-Q., Finkbeiner, W.E., Wine, J.J., Mrsny, R.J. and Widdicombe, J.H. (1994) Calu-3: a human airway epithelial cell line that shows cAMP-dependent Cl⁻ secretion. *Am J Physiol Lung Cell Mol Physiol*, **10**, L493-L501.

191.   Smith, B.T. (1977) Cell line A549: a model system for the study of alveolar type II cell function. *Am Rev Respir Dis*, **115**, 285-293.

192.   Aiello, L., Guilfoyle, R., Huebner, K. and Weinmann, R. (1979) Adenovirus 5 DNA sequences present and RNA sequences transcribed in transformed human embryo kidney cells (HEK-Ad-5 or 293). *Virology*, **94**, 460-469.

193.   Fulcher, M.L., Gabriel, S., Burns, K.A., Yankaskas, J.R. and Randell, S.H. (2005) Well differentiated human airway epithelial cell cultures. *Methods Mol Med*, **107**, 183-206.

194.   Mouchel, N., Henstra, S.A., McCarthy, V.A., Williams, S.H., Phylactides, M. and Harris, A. (2004) HNFα is involved in tissue-specific regulation of CFTR gene expression. *Biochem J*, **378**, 909-918.

195.   Dennis Jr., G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, **4**, P3.

196.   Karmel, A. (2014) Homer-idr: Second pass updated.

197.   Ji, X., Li, W., Song, J., Wei, L. and Liu, X.S. (2006) CEAS: cis-regulatory element annotation system. *Nucleic Acids Res*, **34**, W551-554.

198.   de Hoon, M.J., Imoto, S., Nolan, J. and Miyano, S. (2004) Open source clustering software. *Bioinformatics*, **20**, 1453-1454.

199.   Do, T., Ho, F., Heidecker, B., Witte, K., Chang, L. and Lerner, L. (2008) A rapid method for determining dynamic binding capacity of resins for the purification of proteins. *Protein Expr Purif*, **60**, 147-150.

200. Gaulton, K.J., Nammo, T., Pasquali, L., Simon, J.M., Giresi, P.G., Fogarty, M.P., Panhuis, T.M., Mieczkowski, P., Secchi, A., Bosco, D., Berney, T., Montanya, E., Mohlke, K.L., Lieb, J.D. and Ferrer, J. (2010) A map of open chromatin in human pancreatic islets. *Nat. Genet.*, **42**, 255-259.

201. Li, W., Soave, D., Miller, M.R., Keenan, K., Lin, F., Gong, J., Chiang, T., Stephenson, A.L., Durie, P., Rommens, J., Sun, L. and Strug, L.J. (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.

202. Zhang, Z., Ott, C.J., Lewandowska, M.A., Leir, S.H. and Harris, A. (2012) Molecular mechanisms controlling CFTR gene expression in the airway. *J Cell Mol Med*, **16**, 1321-1330.

203. Leir, S.H., Holgate, S.T. and Lackie, P.M. (2003) Inflammatory cytokines can enhance CD44-mediated airway epithelial cell adhesion independently of CD44 expression. *Am J Physiol Lung Cell Mol Physiol*, **285**, L1305-L1311.

204. McDonald, T.M., Sumner, A.J., Reyes, J.F., Pascual, A.S., Uppalapati, C.K., Cooper, K.E., Leyva, K.J. and Hull, E.E. (2013) Matrix metalloproteinases and collective cell migration in 24 h primary zebrafish explant cultures: MMP13 plays an inhibitory role and MMP14 may respond to stretch during reepithelialisation. *Cell Biol Int Rep*, **20**, 24-36.

205. Mendez, M.G., Kojima, S.I. and Goldman, R.D. (2010) Vimentin induces changes in cell shape, motility, and adhesion during the epithelial to mesenchymal transition. *FASEB J*, **24**, 1838-1851.

206. Yan, L.L., Huang, Y.J., Yi, X., Yan, X.M., Cai, Y., He, Q. and Han, Z.J. (2015) Effects of silencing S100A8 and S100A9 with small interfering RNA on the migration of CNE1 nasopharyngeal carcinoma cells. *Oncol Lett*, **9**, 2534-2540.

207. Smart, S.J. and Casale, T.B. (1994) Pulmonary epithelial cells facilitate TNF-α-induced neutrophil chemotaxis. *J Immunol*, **152**, 4087-4094.

208. Moser, B., Clark-Lewis, I., Zwahlen, R. and Baggiolini, M. (1990) Neutrophil-activating properties of the melanoma growth-stimulatory activity. *J Exp Med*, **171**, 1797-1802.

209. Cowburn, A.S., Deighton, J., Walmsley, S.R. and Chilvers, E.R. (2004) The survival effect of TNF-alpha in human neutrophils is mediated via NF-kappa B-dependent IL-8 release. *Eur J Immunol*, **34**, 1733-1743.

210. Park, K.S., Wells, J.M., Zorn, A.M., Wert, S.E., Laubach, V.E., Fernandez, L.G. and Whitsett, J.A. (2006) Transdifferentiation of ciliated cells during repair of the respiratory epithelium. *Am J Respir Cell Mol Biol*, **34**, 151-157.

211. Akhtar-Zaidi, B., Cowper-Sal-lari, R., Corradin, O., Saiakhova, A., Bartels, C.F., Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M.F., Willis, J., Moore, J.H., Tesar, P.J., Laframboise, T., Markowitz, S., Lupien, M. and Scacheri, P.C. (2012) Epigenomic enhancer profiling defines a signature of colon cancer. *Science*, **336**, 736-739.

212. Gruenert, D.C., Finkbeiner, W.E. and Widdicombe, J.H. (1995) Culture and transformation of human airway epithelial cells. *Am J Physiol* **238**, L347-L360.

213. Endo, M., Oyadomari, S., Suga, M., Mori, M. and Gotoh, T. (2005) The ER stress pathway involving CHOP is activated in the lungs of LPS-treated mice. *J Biochem*, **138**, 501-507.

214. Steiling, K., van den Berge, M., Hijazi, K., Florido, R., Campbell, J., Liu, G., Xiao, J., Zhang, X., Duclos, G., Drizik, E., Si, H., Perdomo, C., Dumont, C., Coxson, H.O., Alekseyev, Y.O., Sin, D., Pare, P., Hogg, J.C., McWilliams, A., Hiemstra, P.S., Sterk, P.J., Timens, W., Chang, J.T., Sebastiani, P., O'Connor, G.T., Bild, A.H., Postma, D.S., Lam, S., Spira, A. and Lenburg, M.E. (2013) A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. *Am J Respir Crit Care Med*, **187**, 933-942.

215. Payankaulam, S., Li, L.M. and Arnosti, D.N. (2010) Transcriptional repression: conserved and evolved features. *Curr Biol*, **20**, R764-771.

216. Tsai, E.Y., Falvo, J.V., Tsytsykova, A.V., Barczak, A.K., Reimold, A.M., Glimcher, L.H., Fenton, M.J., Gordon, D.C., Dunn, I.F. and Goldfeld, A.E. (2000) A lipopolysaccharide-specific enhancer complex involving Ets, Elk-1, Sp1, and CREB Binding Protein and p300 is recruited to the tumor necrosis factor alpha promoter In Vivo. *Mol Cell Biol* **20**, 6084-6094.

217. Reddy, S.P., Vuong, H. and Adiseshaiah, P. (2003) Interplay between proximal and distal promoter elements is required for squamous differentiation marker induction in the bronchial epithelium: role for ESE-1, Sp1, and AP-1 proteins. *J Biol Chem*, **278**, 21378-21387.

218. Hao, H., Qi, H. and Ratnam, M. (2003) Modulation of the folate receptor type beta gene by coordinate actions of retinoic acid receptors at activator Sp1/ets and repressor AP-1 sites. *Blood*, **101**, 4551-4560.

219. Zahlten, J., Herta, T., Kabus, C., Steinfeldt, M., Kershaw, O., Garcia, P., Hocke, A.C., Gruber, A.D., Hubner, R.H., Steinicke, R., Doehn, J.M., Suttorp, N. and Hippenstiel, S. (2015) Role of Pneumococcal Autolysin for KLF4 Expression and Chemokine Secretion in Lung Epithelium. *Am J Respir Cell Mol Biol*, **53**, 544-554.

220. Black, A.R., Black, J.D. and Azizkhan-Clifford, J. (2001) Sp1 and Krüppel-Like Factor family of transcription factors in cell growth regulation and cancer. *J Cell Physiol*, **188**, 143-160.

221.	Gorbatenko, A., Olesen, C.W., Morup, N., Thiel, G., Kallunki, T., Valen, E. and Pedersen, S.F. (2014) ErbB2 upregulates the Na+,HCO3(-)-cotransporter NBCn1/SLC4A7 in human breast cancer cells via Akt, ERK, Src, and Kruppel-like factor 4. *FASEB J*, **28**, 350-363.

222.	Truax, A.D. and Greer, S.F. (2012) ChIP and Re-ChIP assays: investigating interactions between regulatory proteins, histone modifications, and the DNA sequences to which they bind. *Methods Mol Biol*, **809**, 175-188.

223.	Agou, F., Ye, F. and Véron, M. (2004) In vivo protein cross-linking. *Methods Mol Biol*, **261**, 427-442.

224.	Carey, M.F., Peterson, C.L. and Smale, S.T. (2012) Experimental strategies for the identification of DNA-binding proteins. *Cold Spring Harb Protoc*, **2012**, 18-33.

225.	Dunham, W.H., Mullin, M. and Gingras, A.C. (2012) Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics*, **12**, 1576-1590.

226.	Stephens, D.N., Klein, R.H., Salmans, M.L., Gordon, W., Ho, H. and Andersen, B. (2013) The Ets transcription factor EHF as a regulator of cornea epithelial cell identity. *J Biol Chem*, **288**, 34304-34324.

227.	Friedl, P. and Gilmour, D. (2009) Collective cell migration in morphogenesis, regeneration and cancer. *Nat Rev Mol Cell Biol*, **10**, 445-457.

228.	Moon, A., Yong, H.Y., Song, J.I., Cukovic, D., Salagrama, S., Kaplan, D., Putt, D., Kim, H., Dombkowski, A. and Kim, H.R. (2008) Global gene expression profiling unveils S100A8/A9 as candidate markers in H-ras-mediated human breast epithelial cell invasion. *Mol Cancer Res*, **6**, 1544-1553.

229.	Duan, L., Wu, R., Ye, L., Wang, H., Yang, X., Zhang, Y., Chen, X., Zuo, G., Zhang, Y., Weng, Y., Luo, J., Tang, M., Shi, Q., He, T. and Zhou, L. (2013) S100A8 and S100A9 are associated with colorectal carcinoma progression and contribute to colorectal carcinoma cell survival and migration via Wnt/β-catenin pathway. *PLoS One*, **8**, e62092.

230.	Stewart, C.E., Nijmeh, H.S., Brightling, C.E. and Sayers, I. (2012) uPAR regulates bronchial epithelial repair in vitro and is elevated in asthmatic epithelium. *Thorax*, **67**, 477-487.

231.	Wang, Q., Wang, Y., Zhang, Y., Zhang, Y. and Xiao, W. (2015) Involvement of urokinase in cigarette smoke extract-induced epithelial-mesenchymal transition in human small airway epithelial cells. *Lab Invest*, **95**, 469-479.

232. Bastonero, S., Priol, Y.L., Armand, M., Bernard, C.S., Reynaud-Gaubert, M., Olive, D., Parzy, D., de Bentzmann, S., Capo, C. and Mege, J.L. (2009) New microbicidal functions of tracheal glands: defective anti-infectious response to *Pseudomonas aeruginosa* in cystic fibrosis. *PLoS One*, **4**, e5357.

233. Zuo, F., Kaminski, N., Eugui, E., Allard, J., Yakhini, Z., Ben-Dor, A., Lollini, L., Morris, D., Kim, Y., DeLustro, B., Sheppard, D., Pardo, A., Selman, M. and Heller, R.A. (2002) Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc Natl Acad Sci U S A*, **99**, 6292-6297.

234. Besnard, A.G., Struyf, S., Guabiraba, R., Fauconnier, L., Rouxel, N., Proost, P., Uyttenhove, C., Van Snick, J., Couillin, I. and Ryffel, B. (2013) CXCL6 antibody neutralization prevents lung inflammation and fibrosis in mice in the bleomycin model. *J Leukoc Biol*, **94**, 1317-1323.

235. Kuhn III, C., Homer, R.J., Zhu, Z., Ward, N., Flavell, R.A., Geba, G.P. and Elias, J.A. (2000) Airway hyperresponsiveness and airway obstruction in transgenic mice: morphologic correlates in mice overexpressing interleukin (IL)-11 and IL-6 in the lung. *Am J Respir Cell Mol Biol*, **22**, 289-295.

236. Klein, E., Weigel, J., Buford, M.C., Holian, A. and Wells, S.M. (2010) Asymmetric dimethylarginine potentiates lung inflammation in a mouse model of allergic asthma. *Am J Physiol Lung Cell Mol Physiol*, **299**, L816-825.

237. Kinker, K.G., Gibson, A.M., Bass, S.A., Day, B.P., Deng, J., Medvedovic, M., Figueroa, J.A.L., Hershey, G.K.K. and Chen, W. (2014) Overexpression of dimethylarginine dimethylaminohydrolase 1 attenuates airway inflammation in a mouse model of asthma. *PLoS One*, **9**, e85148.

238. Rosenfeld, M., Gibson, R.L., McNamara, S., Emerson, J., Burns, J.L., Castile, R., Hiatt, P., McCoy, K., Wilson, C.B., Inglis, A., Smith, A., Martin, T.R. and Ramsey, B.W. (2001) Early Pulmonary Infection, Inflammation, and Clinical Outcomes in Infants With Cystic Fibrosis. *Pediatr Pulmonol.*, **32**, 356-366.

239. Jain, R., Barkauskas, C.E., Takeda, N., Bowie, E.J., Aghajanian, H., Wang, Q., Padmanabhan, A., Manderfield, L.J., Gupta, M., Li, D., Li, L., Trivedi, C.M., Hogan, B.L. and Epstein, J.A. (2015) Plasticity of Hopx(+) type I alveolar cells to regenerate type II cells in the lung. *Nat Commun*, **6**, 6727.

240. Wan, H., Xu, Y., Ikegami, M., Stahlman, M.T., Kaestner, K.H., Ang, S.L. and Whitsett, J.A. (2004) Foxa2 is required for transition to air breathing at birth. *Proc Natl Acad Sci U S A*, **101**, 14449-14454.

241. Snyder, J.M., Jenkins-Moore, M., Jackson, S.K., Goss, K.L., Dai, H.H., Bangsund, P.J., Giguere, V. and McGowan, S.E. (2005) Alveolarization in retinoic acid receptor-beta-deficient mice. *Pediatr Res*, **57**, 384-391.

242. Besnard, V., Wert, S.E., Kaestner, K.H. and Whitsett, J.A. Stage-specific regulation of respiratoy epithelial cell differentiation by FoxA1. *Am J Physiol Lung Cell Mol Physiol*, **289**, L750-L759.

243. Hansel, N.N., Ruczinski, I., Rafaels, N., Sin, D.D., Daley, D., Malinina, A., Huang, L., Sandford, A., Murray, T., Kim, Y., Vergara, C., Heckbert, S.R., Psaty, B.M., Li, G., Elliott, W.M., Aminuddin, F., Dupuis, J., O'Connor, G.T., Doheny, K., Scott, A.F., Boezen, H.M., Postma, D.S., Smolonska, J., Zanen, P., Mohamed Hoesein, F.A., de Koning, H.J., Crystal, R.G., Tanaka, T., Ferrucci, L., Silverman, E., Wan, E., Vestbo, J., Lomas, D.A., Connett, J., Wise, R.A., Neptune, E.R., Mathias, R.A., Pare, P.D., Beaty, T.H. and Barnes, K.C. (2013) Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. *Hum Genet*, **132**, 79-90.

244. Park, S.W., Verhaeghe, C., Nguyenvu, L.T., Barbeau, R., Eisley, C.J., Nakagami, Y., Huang, X., Woodruff, P.G., Fahy, J.V. and Erle, D.J. (2009) Distinct roles of FoxA2 and FoxA3 in allergic airway disease and asthma. *Am J Respir Crit Care Med*, **180**, 603-610.

245. Chen, G., Wan, H., Luo, F., Zhang, L., Xu, Y., Lewkowich, I., Wills-Karp, M. and Whitsett, J.A. (2010) Foxa2 programs Th2 cell-mediated innate immunity in the developing lung. *J Immunol*, **184**, 6133-6141.

246. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.

247. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62-67.

248. Liu, X., Ory, V., Chapman, S., Yuan, H., Albanese, C., Kallakury, B., Timofeeva, O.A., Nealon, C., Dakic, A., Simic, V., Haddad, B.R., Rhim, J.S., Dritschilo, A., Riegel, A., McBride, A. and Schlegel, R. (2012) ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. *Am J Pathol*, **180**, 599-607.

249. Horani, A., Nath, A., Wasserman, M.G., Huang, T. and Brody, S.L. (2013) Rho-associated protein kinase inhibition enhances airway epithelial Basal-cell proliferation and lentivirus transduction. *Am J Respir Cell Mol Biol*, **49**, 341-347.

250. Trinh, N.T., Bardou, O., Prive, A., Maille, E., Adam, D., Lingee, S., Ferraro, P., Desrosiers, M.Y., Coraux, C. and Brochiero, E. (2012) Improvement of defective cystic fibrosis airway epithelial wound repair after CFTR rescue. *Eur Respir J*, **40**, 1390-1400.

251.   Regamey, N., Jeffery, P.K., Alton, E.W., Bush, A. and Davies, J.C. (2011) Airway remodelling and its relationship to inflammation in cystic fibrosis. *Thorax*, **66**, 624-629.

CURRICULUM VITAE

Sara L. Fossum

Current address: Email:
441 E Erie Apt 2611   sara-fossum@fsm.northwestern.edu
Chicago, Il 60611   Phone:
                                              651-226-5251

EDUCATION

NORTHWESTERN UNIVERSITY   Chicago, IL
MD   Expected 2018
PhD   Expected 2016
GPA: 3.9/4.0

UNIVERSITY OF NOTRE DAME   Notre Dame, IN
BS, Major: Biology   May 2010
Summa cum Laude
GPA: 3.93/4.0

UNIVERSITY OF NOTRE DAME LONDON PROGRAM London, UK
Semester Abroad   Fall 2008

HONORS

•Junior Investigator Basic Science Semi-Finalist, North American Cystic Fibrosis Conference (October 2015).
•Stanley Manne Children's Research Institute Travel Award (March 2015).
•Scientific Poster Award-Graduate Student, Lurie Children's Research Center Biomedical Research Symposium (March 2014)
• Cell and Molecular Basis of Disease Training Grant, supported by grant 2T32GM008061-31, Northwestern University (March 2013-March 2015)
•Heller Award for Excellence in the Pre-Clinical Curriculum, Northwestern Feinberg School of Medicine (2012)
•Outstanding Biological Scientist Award, University of Notre Dame (2010)
•Member, Phi Beta Kappa Society (2010)
•Dean's List, University of Notre Dame (Fall 2006-Spring 2010)
•College of Science Summer Undergraduate Research Fellowship, University of Notre Dame (2009)
•Claire Boothe Luce Fellowship, University of Notre Dame (2009)
•National Merit Scholar (2006)

SCIENTIFIC EXPERIENCE

•Graduate Student with Dr. Ann Harris, Human Molecular Genetics Program, Stanley Manne Children's Research Institute and Department of Pediatrics, Northwestern University Feinberg School of Medicine. *Characterization of the role of the EHF transcription factor in regulating lung epithelial gene expression.* October 2012-Present
•Undergraduate Student with Dr. David Hyde, Department of Biological Sciences, University

of Notre Dame. *Regeneration of the zebrafish retina following Degeneration.* January 2009-May 2010

•Undergraduate Summer Intern with Dr. Romas Kazlauskas, Supercomputing Institute Summer Internship Program, University of Minnesota. *Molecular modeling study of converting an esterase to a hydroxynitrile lyase.* June 2008-August 2008.

•Molecular Cell Biology Independent Research Lab Group Project with Dr. Kevin Vaughan. *RZZ Complex and the Spindle Assembly Checkpoint.* January 2008-April 2008.

TEACHING EXPERIENCE

•Teaching Assistant. Introduction to Life Sciences. Northwestern University. Dr. Steven Anderson, course director. Winter 2013.

•Teaching Assistant. Molecular Cell Biology Independent Research Lab. University of Notre Dame. Dr. Michelle Whaley, course director. Spring 2010.

•Teaching Assistant. Molecular Genetics Lab. University of Notre Dame. Dr. Michelle Whaley, course director. Fall 2009.

•Teaching Assistant. Molecular Cell Biology Independent Research Lab. University of Notre Dame. Dr. Michelle, Whaley course director. Spring 2009.

PRESENTATIONS

• Fossum, S., Mutolo, M. J., Yang, R., Dang, H., O'Neal, W.K., Knowles, M., Leir, S.H., Harris, A. "Ets homologous factor regulates pathways controlling response to injury in airway epithelial cells." Poster Presentation. North American Cystic Fibrosis Conference. Phoenix, AZ. October 2015.

• Fossum, S., Mutolo, M. J., Yang, R., Dang, H., O'Neal, W.K., Knowles, M., Leir, S.H., Harris, A. "Ets homologous factor regulates pathways controlling response to injury in airway epithelial cells." Poster Presentation. Northwestern University Genetics and Genomics Cluster Annual Symposium. Chicago, IL. July 2015.

• Fossum, S., Mutolo, M. J., Yang, R., Dang, H., O'Neal, W.K., Knowles, M., Leir, S.H., Harris, A. "Ets homologous factor regulates pathways controlling response to injury in airway epithelial cells." Poster Presentation. American Physician Scientists Association 2015 Annual Meeting. Chicago, IL. April 2015.

•Fossum, S., Mutolo, M. J., Yang, R., Dang, H., O'Neal, W.K., Knowles, M., Leir, S.H., Harris, A. "Ets homologous factor regulates pathways controlling response to injury in airway epithelial cells." Poster Presentation. Cold Spring Harbor Laboratories Systems Biology: Networks Meeting. Cold Spring Harbor, NY. March 2015.

•Fossum, S., Mutolo, M. J., Yang, R., Leir, S.H., Jones, L. Dang, H., O'Neal, W. Knowles, M. Harris, A. "The regulatory role of EHF in the lung epithelium." Poster Presentation. 2014 Midwest Chromatin and Epigenetics Meeting. Madison, WI. May 2014.

•Fossum, S., Mutolo, M. J., Yang, R., Leir, S.H., Jones, L. Dang, H., O'Neal, W. Knowles, M. Harris, A. "The regulatory role of EHF in the lung epithelium." Poster Presentation. Lurie Children's Research Center Biomedical Research Symposium. Chicago, IL. March 2014.

•"The Role of Endocytosis in Signaling Müller Glia to Proliferate Following Retinal Degeneration." Poster Session. Summer Undergraduate Research Symposium. Notre Dame, IN. July 2009.

•"A molecular modeling study of converting an esterase to a hydroxynitrile lyase." Slide Presentation. Supercomputing Institute Summer Internship Program Presentations. Minneapolis, MN. August 2008.

GRANTS

4/8/2015-4/7/2017. PI: S. Fossum. "The role of the EHF transcription factor in lung epithelial function." NIH/NHLBI F30 HL124925-01A1

PUBLICATIONS: ORIGINAL JOURNAL ARTICLES

Fossum, S.L., Mutolo, M.J., Yang, R., Dang, H., O'Neal, W.K., Knowles, M.R., Leir, S.H., Harris, A. 2014. Ets homologous factor regulates pathways controlling response to injury in airway epithelial cells. *Nucleic Acids Res*. 42, 13588-91358. PMC4267623.

Padhi, S.K., Fujii R., Legatt G.A., **Fossum S.L**., Berchtold R., and Kazlauskas R.J. 2010. Switching from an esterase to a hydroxynitrile lyase mechanism requires only two amino acid substitutions. *Chem. Biol.* 17**,** 863-871.

Bailey, T.J., **Fossum, S.L.**, Fimbel, S.M., Montgomery, J.E., and Hyde, D.R. 2010. The inhibitor of phagocytosis, O-phospho-L-serine, suppresses Müller glia proliferation and cone cell regeneration in the light-damaged zebrafish retina. *Exp. Eye Res.* 91**,** 601-612. PMC2962682.