Journal of proteome research

Article

pubs.acs.org/jpr

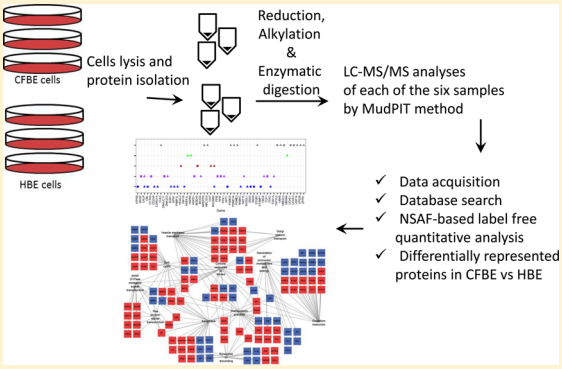
Quantitative Proteomic Profiling Reveals Differentially Regulated Proteins in Cystic Fibrosis Cells

Navin Rauniyar,[†] Vijay Gupta,[‡,⊥] William E. Balch,[‡] and John R. Yates, III*[,†]

[†]Department of Chemical Physiology, [‡]Department of Cell and Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States

Supporting Information

ABSTRACT: The most prevalent cause of cystic fibrosis (CF) is the deletion of a phenylalanine residue at position 508 in CFTR (ΔF508-CFTR) protein. The mutated protein fails to fold properly, is retained in the endoplasmic reticulum via the action of molecular chaperones, and is tagged for degradation. In this study, the differences in protein expression levels in CF cell models were assessed using a systems biology approach aided by the sensitivity of MudPIT proteomics. Analysis of the differential proteome modulation without a priori hypotheses has the potential to identify markers that have not yet been documented. These may also serve as the basis for developing new diagnostic and treatment modalities for CF. Several novel differentially expressed proteins observed in our study are likely to play important roles in the pathogenesis of CF and may serve as a useful resource for the CF scientific community.



KEYWORDS: *Cystic Fibrosis, cystic fibrosis transmembrane conductance regulator, CFTR, MudPIT, mass spectrometry, NSAF, spectral count, label free, proteomics*

INTRODUCTION

Cystic fibrosis (CF) is the most common lethal autosomal recessive disease in the Caucasian population and is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. CFTR is a PKA-regulated chloride channel localized at the apical surface of primary epithelia such as those found in lung, pancreas, intestine, and kidney, where it functions to regulate water and salt homeostasis. More than 1800 individual mutations have been reported in this multidomain 1480-residue polytopic membrane glycoprotein that give rise to a spectrum of differing disease severities and symptoms (http://www.genet.sickkids.on.ca/cftr). The most common mutation in humans (accounting for an estimated 75% of alleles and found in generally 90% of CF patients) is the deletion of a phenylalanine at position 508 (ΔF508-CFTR) in the CFTR protein. The F508 deletion prevents proper folding of CFTR in the endoplasmic reticulum (ER) and impedes its trafficking to its functional site at the cell surface. This mutant version of CFTR is recognized as abnormal and remains incompletely processed in the ER, where it is subsequently degraded.[1] As a consequence, cells expressing the mutant protein are unable to transport chloride ions across the plasma membrane in response to a rise in intracellular cAMP levels.

CF is referred to as a monogenic disease with a broad range of lung disease severity even for patients who are homozygous for the ΔF508 mutation.[2] The non-CFTR genetic variants (modifier genes) and/or environmental influences contribute to the heterogeneity of pulmonary disease severity.[2−4] However, in addition to the identification of modifier genes, a complementary study employing a global proteomics approach that looks at cells expressing wild-type and mutant CFTR holds the potential for discovery of perturbed molecular pathways underlying this complex disease process. The rationale behind this assumption is that the F508 deletion causes proteomic changes that otherwise would not be predicted on the basis of known gene functions. Hence, the interrogation at the level of the proteome and identification of differentially expressed proteins can provide both a useful overview of proteins involved in CF pathogenesis and the opportunity to identify new therapeutic targets. Mass spectrometry-based shotgun proteomics is an effective tool for deciphering differences in biological systems at the proteome level. The current generation of mass spectrometers equipped with high-resolution and rapid-scanning mass analyzers facilitates increased depth of proteome coverage, allowing thousands of proteins to be routinely identified and quantified from biological samples. Here, we used two-dimensional liquid chromatography coupled to tandem mass spectrometry (2D LC−MS/MS), also referred to as MudPIT (multidimensional protein identification technology),[5] in an

Special Issue: Proteomics of Human Diseases: Pathogenesis, Diagnosis, Prognosis, and Treatment

Received: April 13, 2014
Published: May 13, 2014

   dx.doi.org/10.1021/pr500370g | J. Proteome Res. 2014, 13, 4668−4675

LTQ Orbitrap Velos mass spectrometer to determine protein expression changes in bronchial epithelial cell models[6] of CF disease, CFBE41o- (CFBE) and HBE41o- (HBE). We were able to identify 349 differentially expressed proteins using a spectral count label-free quantification approach. A subset of deregulated proteins observed in our study belongs to key biological processes that are of direct relevance to CF pathogenesis, and the others are possibly involved in proteostasis of CFTR processing.

MATERIALS AND METHODS

Materials

Tris(2-carboxyethyl) phosphine (TCEP), Tris, iodoacetamide, sodium chloride, urea, and SDS were obtained from Sigma-Aldrich (St. Louis, MO, USA). NP-40 was from Thermo Fisher Scientific (Rockford, IL, USA). Sequencing grade trypsin was from Promega (Madison, WI, USA).

Sample Preparation for Mass Spectrometry Analysis

Human bronchial epithelial cells stably expressing ΔF508-CFTR (CFBE cells) or isogenic cells stably expressing wild-type CFTR (HBE cells) were cultured as previously described.[7] Cells were harvested by performing two washes in PBS and incubating the plates with lysis buffer on ice for 20 min. The lysis buffer (25 mM Tris, pH 7.6, 150 mM NaCl, 1% NP-40, and 0.1% SDS) was supplemented with 1% protease inhibitor mixture (Roche, Indianapolis, IN, USA). Lysis of cells was also aided by sonication for 5 min in a water bath sonicator. Protein lysates were clarified by centrifugation (13 500 rpm, 30 min at 4 °C). Protein concentration was determined by BCA protein assay kit (Sigma, St. Louis, MO, USA).

A total of 200 μg of total cell lysate was precipitated by adding a 4-fold volume of ice-cold acetone. This was incubated at −20 °C for 2 h and then centrifuged at 10 000 rpm for 10 min at 4 °C. The pellet was air-dried. An initial reaction volume of 100 μL was obtained by resuspending the pellet in Tris buffer 50 mM, pH 8.0, containing an 8 M final concentration of urea. The proteins were reduced with 5 mM TCEP for 20 min and alkylated with 10 mM iodoacetamide for 15 min in the dark. The reaction mixture was diluted to 2 M urea with 25 mM Tris, pH 8.0. Trypsin (Promega, Madison, WI, USA) was added at an enzyme/substrate ratio of 1:50 (w/w). The suspensions were then placed in a Thermomixer (Eppendorf, Westbury, NY) and incubated overnight at 37 °C at 750 rpm. The next day, the sample was acidified with formic acid to a final concentration of 5% and spun at 14 000 rpm for 30 min. Fifty micrograms of tryptic digest was aliquoted for MS analysis.

Mass Spectrometry (MS) Analysis

MS analysis of the samples was performed using multidimensional protein identification technology (MudPIT). Capillary columns were prepared in-house from particle slurries in methanol. An analytical RPLC column was generated by pulling a 100 μm i.d./360 μm o.d. capillary (Polymicro Technologies, Inc., Phoenix, AZ, USA) to 3 μm i.d. tip. The pulled column was packed with reverse-phase particles (Aqua C18, 3 μm diameter, 90 Å pores, Phenomenex, Torrance, CA, USA) until a length of 15 cm was reached. A MudPIT trapping column was prepared by creating a Kasil frit at one end of an undeactivated 250 μm i.d./360 μm o.d. capillary (Agilent Technologies, Inc., Santa Clara, CA, USA), which was then successively packed with 2.5 cm strong cation-exchange particles (Partisphere SCX, 5 μm diameter, 100 Å pores, Phenomenex, Torrance, CA, USA) and 2.5 cm reverse-phase particles (Aqua C18, 5 μm diameter, 90 Å pores, Phenomenex, Torrance, CA, USA). The trapping column was equilibrated using buffer A prior to sample loading. After sample loading and prior to MS analysis, the resin-bound peptides were desalted with buffer A by letting it flow through the biphasic trap column. The trap and analytical columns were assembled using a zero-dead-volume union (Upchurch Scientific, Oak Harbor, WA, USA).

LC−MS/MS analysis was performed on LTQ Orbitrap Velos (Thermo Scientific, San Jose, CA, USA) interfaced at the front end with a quaternary HP 1100 series HPLC pump (Agilent Technology, Santa Clara, CA, USA) using an in-house built electrospray stage. Electrospray was performed directly from the analytical column by applying the ESI voltage at a tee (150 μm i.d., Upchurch Scientific) directly downstream of a 1:1000 split flow used to reduce the flow rate to 250 nL/min through the columns. A fully automated 10-step MudPIT run was performed on each sample using a three-mobile-phase system consisting of buffer A (5% acetonitrile (ACN); 0.1% formic acid (FA) (Sigma-Aldrich, St. Louis, MO, USA)), buffer B (80% ACN, 0.1% FA), and buffer C (500 mM ammonium acetate, 5% ACN, 0.1% FA). The first step was a 60 min reverse-phase run, whereas subsequent steps were of 120 min duration. Each MudPIT run included steps with 10, 20, 30, 40, 50, 70, 80, and 100% buffer C run for 4 min at the beginning of the gradient except for the last step, which included a salt bump of 90% buffer C with 10% buffer B for 4 min.

As peptides were eluted from the microcapillary column, they were electrosprayed directly into the mass spectrometer with the application of a distal 2.4 kV spray voltage. Peptides were analyzed using a top-20 data-dependent acquisition method in which fragmentation spectra are acquired for the top 20 peptide ions above a predetermined signal threshold. For each cycle, survey full-scan MS spectra (m/z range 300−1600) were acquired in the Orbitrap with the resolution set to a value of 60 000 at m/z 400, an automatic gain control (AGC) target of 1×10^6 ions, and the maximal injection time of 250 ms. Each full scan was followed by the selection of the most intense ions, up to 20, for collision-induced dissociation (CID)-MS/MS analysis in the ion trap. For MS/MS scans, the target value was 10 000 ions with an injection time of 25 ms. Once analyzed, the selected peptide ions were dynamically excluded from further analysis for 120 s to allow for the selection of lower-abundance ions for subsequent fragmentation and detection using the following settings: repeat count, 1; repeat duration, 30 ms; and exclusion list size, 500. Charge state filtering, where ions with singly or unassigned charge states were rejected from fragmentation, was enabled. The minimum MS signal for triggering MS/MS was set to 500, and an activation time of 10 ms was used. All tandem mass spectra were collected using a normalized collision energy of 35% and an isolation window of 2 Th.

Data Analysis

Tandem mass spectra were extracted from the Xcalibur data system format (.raw) into MS2 format using RawXtract1.9.9.2. The MS/MS spectra were searched with the ProLuCID algorithm against the human SwissProt database (downloaded March 2014) that was concatenated to a decoy database in which the sequence for each entry in the original database was reversed. The search parameters include 10 ppm peptide precursor mass tolerance and 0.6 Da for the fragment mass

dx.doi.org/10.1021/pr500370g | J. Proteome Res. 2014, 13, 4668−4675

tolerance acquired in the ion trap; carbamidomethylation on cysteine was defined as fixed modification in the search criteria. The search space also included all fully and semitryptic peptide candidates with a length of at least six amino acids. Maximum number of internal miscleavages was kept unlimited, thereby allowing all cleavage points to be considered. ProLuCID outputs were assembled and filtered using the DTASelect2.0[8] program that groups related spectra by protein and removes those that do not pass basic data-quality criteria. DTASelect2.0 combines XCorr and ΔCn measurements using a quadratic discriminant function to compute a confidence score to achieve a user-specified false discovery rate (1% for the current study). We accepted only those proteins that were supported by two or more lines of evidence.

For label-free quantification, normalized spectral abundance factor (NSAF) values were calculated for proteins in each sample to account for protein size and variability between runs.[9] Briefly, the NSAF for a protein k is the number of spectral counts (SpC, the total number of MS/MS spectra) identifying a protein, k, divided by the protein length (L), divided by the sum of SpC/L for all N proteins in the experimental design (eq 1).[9]

$$(\mathrm{NSAF})k = \frac{\left(\frac{\mathrm{SpC}}{L}\right)k}{\sum_{i=1}^{N}\left(\frac{\mathrm{SpC}}{L}\right)i} \quad (1)$$

A critical assumption that must be satisfied for use of statistical approaches is that the data set being analyzed must have a normal/Gaussian distribution.[9] Following elucidation of NSAF values, their natural logarithm (ln(NSAF)) was calculated, and a density plot of the distribution of ln(NSAF) values from replicates of each condition were generated to show the normality of the distribution (Figure S1 in the Supporting Information). After establishing that both CFBE and HBE data sets fit a normal distribution, the data sets were statistically compared to determine the significance of the change between the two groups using Student's t test (two-tailed unpaired t test). To determine the relative abundance of expressed proteins in CFBE cells relative to that in HBE, the data set was first filtered to include only those proteins that were detected in all three replicates for each condition and then the ratio of the mean of the NSAF values from three biological replicates of CFBE cells to the mean of NSAF values from three biological replicates of HBE cells was computed. Proteins were considered to exhibit significant expression changes with $\log_2\mathrm{NSAF}_{\mathrm{CFBE/HBE}} \geq 0.58$ ($p < 0.05$) (overexpressed in CFBE cells) and ≤ -0.58 ($p < 0.05$) (underexpressed in CFBE cells). We discarded the proteins from further quantitative analyses that were identified in both conditions but were found in less than three replicates in each group because of poor reproducibility. The NSAF value, t test, and ratio calculation were performed using Microsoft Excel. The graphs were drawn either in Excel or the R statistical package (http://www.r-project.org/).

Preparation of Cell Lysates and Western Blotting

Cells grown in 6-well dishes were washed twice with ice cold PBS and lysed in lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton X-100, and 2 mg/mL of complete protease inhibitor cocktail) on ice for 30 min with gentle rotation. Protein lysates were cleared by centrifugation at 14 000g for 15 min at 4 °C, and supernatants were collected for further analysis. The protein concentration was assessed by Bradford assay using the Coomassie protein assay reagent (Thermo Fisher Scientific, Rockford, IL, USA). Aliquots of 15–20 μg of total protein were separated by 8% or 4–20% SDS-PAGE, transferred to nitrocellulose, and incubated with the indicated primary antibodies overnight at 4 °C followed by IR-dye labeled secondary antibodies in the dark at room temperature for 1 h. Finally, the blots were scanned using the LiCor Odyssey laser-based image detection method.

RESULTS AND DISCUSSION

In this study, we investigated the differential protein expression between a cell line model of cystic fibrosis (i.e., bronchial epithelial cells expressing ΔF508-CFTR (CFBE cells)) and wild-type CFTR (HBE cells). The experimental approach used in this study is outlined in Figure 1. The experiments were performed in triplicate, that is, the procedure was repeated in parallel for three culture plates of each condition. Each biological replicate was processed in parallel to minimize the effects of systematic errors. The proteins were isolated from HBE and CFBE cells as described in the Materials and Methods. The protein extracts were subsequently digested with trypsin, and the resulting peptides were analyzed by LC–MS/MS using the MudPIT method. The MudPIT technology is an unbiased discovery-based method for rapid, yet nearly comprehensive, proteome analysis where increasing levels of salt are used for stepwise elution of peptides from the strong cation-exchange (SCX) resin onto the reversed-phase resin (vide supra).[5] After each step elution from the cation-exchange column, a reversed-phase gradient elutes the peptides into the mass spectrometer according to their hydrophobicity, and MS/MS are acquired automatically by data-dependent acquisition. All mass spectrometry data collection preceded data analysis.

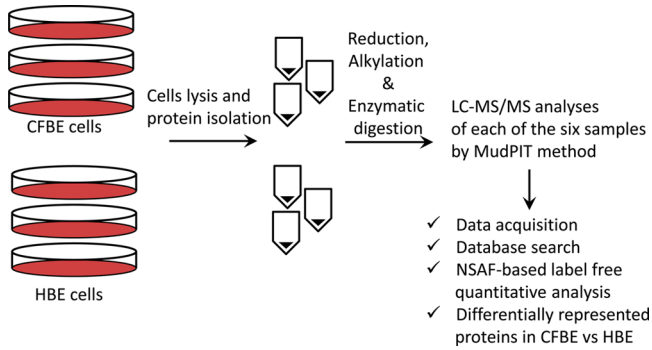


Figure 1. Diagrammatic representation of the experimental approach used for the comparative proteomic analysis of CF cell models, CFBE and HBE.

Three biological replicates were used for each cell type, and from the analyses of replicates of the HBE group, 4307, 4474, and 4415 proteins were identified, whereas 4326, 4372, and 4719 proteins were observed from the CFBE group (Table S1 in the Supporting Information). After consolidating proteins from the replicate samples of each group, a total of 5296 and 5430 proteins were identified from the HBE and CFBE groups, respectively. The bar graph shown in Figure S2 in the Supporting Information shows a summary of the number of proteins identified from the MS analyses of each replicate of the two conditions and the cumulative identifications from three replicates of each condition. The Venn diagram in Figure 2a,b shows the comparison of proteins identified between replicate

dx.doi.org/10.1021/pr500370g | J. Proteome Res. 2014, 13, 4668–4675

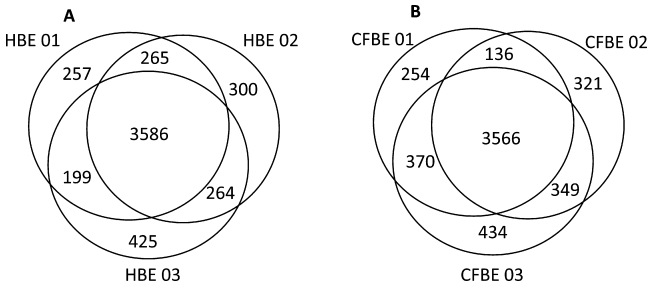


Figure 2. Venn diagrams showing the protein overlap in biological replicates of each cell lines. (A) Venn diagram representing the number of proteins observed for each of the three biological replicate analysis of HBE sample as well as the protein overlap. (B) Venn diagram showing similar observations as those in panel A; however, the data was obtained from the analysis of replicates of CFBE cells.

experiments within each group. The number of times a given protein is detected reflects the reliability of the measurement. As can be seen in the Venn diagram, an overlap of 68% (3586 out of 5296 total identified proteins in HBE cells) and 66% (3566 out of 5430 total identified proteins in CFBE cells) proteins was observed across all three replicates of HBE and CFBE cells, respectively. In addition, a total of 3140 proteins were found to be common to both HBE and CFBE groups among all replicates (Figure S3 in the Supporting Information).

A spectral counting-based label-free approach was used for quantitative profiling of CFBE versus HBE cells. Spectral count, defined as the total number of MS/MS spectra acquired and confidently assigned to a peptide, has proven to be a successful label-free strategy for protein quantification.[10,11] The raw spectral counts were first transformed to yield normalized spectral abundance factor (NSAF) values in order to adjust for the variance in spectral count that occurs because of protein length and the run-to-run variance in total spectral count observed among experimental conditions.[12] NSAF values allow more accurate quantification of both the actual protein abundances in a sample and the expression level changes between multiple samples and experiments. The full lists of proteins that were identified in each sample, three replicates of HBE and three of CFBE, are provided in Table S1 in the Supporting Information. To assess the data quality prior to quantitative analysis, binary comparison of ln(NSAF) values of common proteins identified among the replicate HBE and CFBE samples was performed. The scatter plots shown in Figure S4 in the Supporting Information reveal the distribution of ln(NSAF) values along a diagonal line with very high positive correlation, demonstrating high experimental reproducibility among biological replicates of each group. For relative quantitative analysis, we used only proteins that were common in all runs and identified in both CFBE and HBE cells, which correspond to a total of 3140 proteins (Table S1 in the Supporting Information). The relative abundance of proteins in the CFBE versus HBE comparison was obtained by division of the average NSAF values for each identified proteins in CFBE cells with average NSAF values for the corresponding proteins in HBE cells. The statistically significant differentially regulated proteins in CFBE against HBE pairwise comparison were subsequently selected on the basis of their p value (<0.05) and the magnitude of change in relative abundance ($NSAF_{CFBE/HBE}$) of at least 1.5-fold. On the basis of this threshold, 349 proteins were observed to be perturbed, with 218 proteins upregulated in CFBE cells and 131 proteins downregulated compared to HBE cells. A volcano plot of all 3140 quantified proteins from the CFBE versus HBE data set displaying the relationship between statistical significance (−log p value) and $\log_2$ ratio of each protein is shown in Figure 3. The deregulated proteins

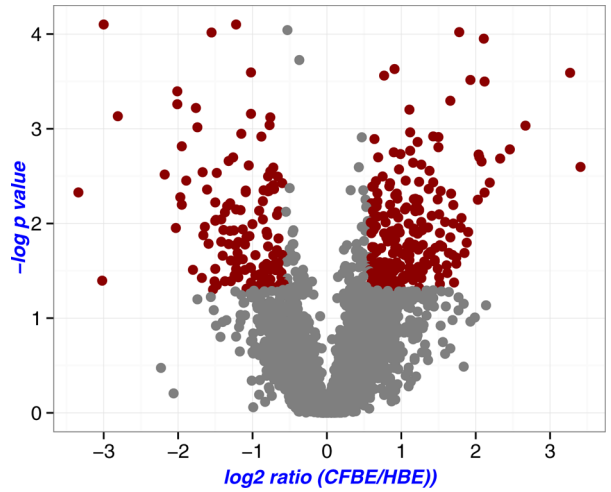


Figure 3. Volcano plot of 3140 quantified proteins from the CFBE versus HBE data set displaying the relationship between statistical significance and fold change of each protein. The $\log_2$ fold change is plotted on the x axis, and the −log p value is plotted on the y axis. The red dots represent the 349 proteins that had statistically significant differential expression, $\log_2$ fold change ≤ -0.58 or ≥ 0.58 ($p < 0.05$).

that are statistically significant ($p < 0.05$) are depicted in red dots in the plot. Table S2 in the Supporting Information provides the list of these proteins along with their UniProt IDs, official gene symbol, NSAF values, p values, fold-change values in logarithm to base 2, and the (raw) spectral counts identified in each of the three biological replicates in each condition. The stochastic nature of data acquisition by mass spectrometry results in instances where proteins are identified in one replicate but not in the other during the replicate sample analysis from the same condition. The proteins that were identified in less than three replicates in both HBE and CFBE cells were discarded from quantitative analysis because of poor reproducibility that would otherwise confound the expression data. However, the proteins that were identified in all three replicates in one or the other group were of particular interest, so these uniquely identified proteins in all three replicates in either HBE or the CFBE samples, labeled as "HBE specific" or "CFBE specific", are listed in Table S2 in the Supporting Information. Because these proteins are observed only in the HBE or CFBE sample, their ratio is not available to report.

The text mining tool Chilibot[13] was used to find the relationship between CFTR or CF and the statistically significant, differentially expressed proteins in our study. Chilibot searches the PubMed literature database (abstracts) for specific relationships among proteins, genes, or keywords. It automatically expands the supplied gene symbols to include its synonyms and then queries PubMed and retrieves relevant records. A subset of proteins detected as differentially expressed in our study has been previously shown to play a role in CFTR biogenesis. One example is protein-glutamine gamma-glutamyltransferase 2 (TGM2) that was upregulated by more than 11-fold in CFBE compared to HBE cells. A significant increase of TGM2 protein and enzymatic activity in CF epithelium and CFTR-defective cell lines have been demonstrated.[14] TGM2 is

dx.doi.org/10.1021/pr500370g | J. Proteome Res. 2014, 13, 4668−4675

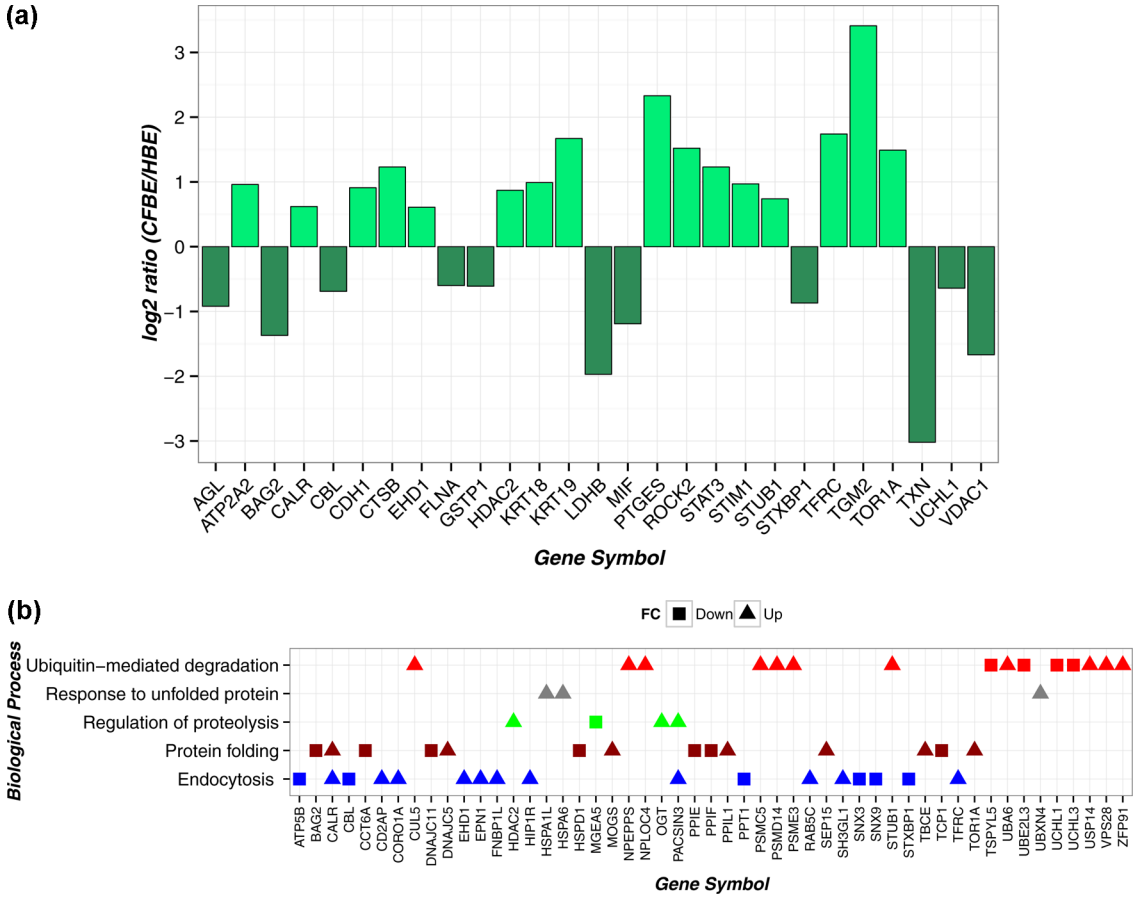


Figure 4. (a) Bar graph displaying a subset of the statistically significant, differentially expressed proteins observed in our data set that have literature evidence for their functional association with CFTR or cystic fibrosis. An online tool, Chilibot (http://www.chilibot.net/), was used to mine PubMed for the relationships. See Table S3 in the Supporting Information for details. (b) Biological functions of a subset of the statistically significant, differentially expressed proteins observed in our data set. The process annotation was obtained from online GO tools, DAVID (http://david.abcc.ncifcrf.gov/), and Enrichnet (http://www.enrichnet.org/). FC, fold change (CFBE/HBE); square and triangular shapes represent down- and upregulated proteins, respectively. The actual fold change value can be obtained from Table S2 (sheet 1) in the Supporting Information.

a pleiotropic enzyme with a calcium-dependent transamidating activity that results in cross-linking of proteins via $\varepsilon(\gamma$-glutamyl) lysine bonds.[15] TGM2 inhibition with cystamine has been shown to rescue mutant CFTR and could become a therapeutic target to control inflammation in CF and possibly in other chronic inflammatory diseases.[16] Figure 4a shows the proteins that were observed to be differentially expressed in our data set and that also have literature evidence for their functional associations with CFTR or CF. Table S3 shows the role of these proteins in CFTR biogenesis or CF pathogenesis and the respective literature references. The list of identified candidates, among which are some confirmations of previous findings, increases the confidence in our data, and we believe that many of the observed new markers could have potential relevance in CF pathogenesis and/or CFTR proteostasis.

The clustering of 349 differentially regulated proteins that were identified was performed according to their biological processes on the basis of the gene ontology (GO) categories. However, only 68% of the uploaded gene symbols were mapped to the biological process GO terms by DAVID[17] analysis. Hence, in addition to DAVID, the assessment of GO categories was also performed using EnrichNet.[18] CFTR, like other membrane proteins, is synthesized and assembled in the ER, where it is core-glycosylated.[19] Once checked for correct folding, this immature form of CFTR migrates to the Golgi complex, where it undergoes further glycosylation. From the Golgi apparatus, only the fully mature form is transported to the plasma membrane, where it functions as a chloride channel. Most of the immature wild-type (∼70%) and approximately 99% of misfolded ΔF508-CFTR are retained in the ER compartment and are degraded via the cytosolic ubiquitin/proteasome pathway.[20−22] Because CF is a protein misfolding disease, some of the enriched biological functions illustrated by these functional annotation tools are of particular interest, especially proteins involved in folding, response to unfolded protein, endocytosis, proteolysis, and ubiquitin-mediated degradation, among others. The differentially expressed proteins that are representative of some of these key biological functions, based on CF biology, are illustrated in Figure 4b. Many molecular chaperones localized to the ER lumen and the cytosol that have been shown to transiently associate with both wild-type CFTR and ΔF508-CFTR were observed in our analysis. Among the differentially expressed members of the molecular chaperones and folding catalysts are heat shock proteins (HSPs), such as HSPD1 (HSP60), HSPA6, HSPA1L, DNAJC5, DNAJC11, TOR1A, and T-complex 1 subunit, TCP1 (aka CCT1). These proteins are known to interact selectively and noncovalently with an unfolded protein, helping them to achieve proper folding and preventing protein aggregation. Peptidylprolyl isomerases are a class of folding catalysts that

dx.doi.org/10.1021/pr500370g | J. Proteome Res. 2014, 13, 4668−4675

accelerate potentially slow steps in the folding process. They increase the rate of cis−trans isomerization of peptide bonds involving proline residues; another class is protein disulfide isomerases, which enhance the rate of formation and reorganization of disulfide bonds.[23] Peptidylprolyl isomerases, such as PPIE and PPIF, were downregulated, whereas PPIL1 was upregulated in our data set. In a cell, a stringent quality control mechanism exists that is capable of discriminating normally folded proteins from abnormally folded proteins.[24] Improperly folded proteins that could otherwise form potentially toxic aggregates can be targeted for degradation by the ubiquitin proteasome system. Degradation of a protein by the ubiquitin system involves two successive steps, conjugation of multiple moieties of ubiquitin and degradation of the tagged protein by the proteolytic activity of the 26S proteasome catalytic core. Several proteins with functional roles in ubiquitin-mediated proteolysis were observed to have differential expression, including ubiquitin-conjugating enzymes (E2), UBE2L3; ubiquitin ligases (E3), CUL5 and STUB1; deubiquitinating enzymes, UCHL1, UCHL3, and USP14; and 26S proteosomal subunits, PSMD14, PSMC5, and PSME3. Calreticulin (CALR), upregulated in CFBE cells compared to HBE in this study, has been shown to assist in CFTR assembly and also to negatively regulate cell surface CFTR by enhancing its endocytosis, leading to proteosomal degradation.[25] The endosome is a membrane-bounded organelle to which materials ingested by endocytosis are delivered. The ΔF508-CFTR has increased endocytic rate in the apical membrane, leading to decreased CFTR-mediated chloride secretion.[26] Endocytic trafficking of CFTR from the membrane surface has implicated several Rab proteins as being active players with distinct roles. Interestingly, in our study, RAB5C was significantly upregulated by more than 2-fold in CFBE cells. RAB5 has been shown to play a role in initial internalization to early endosomes, and the mutant protein can be rescued at the plasma membrane by inhibition of Rab5-dependent endocytosis.[27] RAB1A was also mapped as being differentially regulated in our data set. In addition, our comparative analysis also identified a cohort of proteins with significant expression changes that are involved in biological processes such as endocytosis, oxidation−reduction, homeostasis, response to stress, apoptosis, and response to wounding, among others (Figure S5).

The spectral count quantification results were further validated by confirming the expression level of a subset of differentially expressed proteins utilizing western blotting (Figure S6). CFTR, like other membrane proteins, is synthesized and assembled in the ER, where it is core-glycosylated (also known as band B, ∼145 kDa). Once checked for correct folding, this immature form of CFTR migrates to the Golgi complex, where it undergoes full glycosylation (also known as band C, ∼170 kDa). Approximately 99% of misfolded ΔF508-CFTR protein is degraded before it reaches to the plasma membrane. Only the fully mature form reaches the surface membrane, where it functions as a chloride channel. As seen in the western blot data, band C is the predominant band in HBE cells, whereas only the immature form of CFTR (band B) is observed in CFBE cells. Also observed were down-regulation in the steady-state expression of BAG2 (cochaperone of Hsp70/Hsc70) and elevation in the expression level of calreticulin, thus corroborating the directionality of the fold change observed with spectral counting results. Additional proteins validated by western blot include calnexin, Hsp90, Hsp70, Hsc70, inducible Hsp70 (Hsp70i), Hsp40, Rab5c, Rab7, and Rab11.

To date, only a few studies have examined the proteomic signatures of CF model systems.[28] Using a hybrid approach involving gene transfer and measurement of de novo biosynthetic rates, Pollard et al. have identified 51 significantly changing proteins in CF lung epithelial cells.[29] Davezac et al. have examined the role of misfolded CFTR on global protein expression by comparing the effect of wild-type versus ΔF508-CFTR overexpression in HeLa cells.[30] Their study showed elevation in the expression level of Keratin 8 and 18 (KRT8 and KRT18) in the ΔF508-CFTR cells compared with that from wild-type CFTR controls. A functional assay for CFTR reported in their study revealed that reducing KRT18 expression resulted in increased trafficking of ΔF508-CFTR to the plasma membrane. Interestingly, in our data set, we observed 2-fold upregulation of KRT18 in ΔF508-CFTR cells relative to that of wild type, thus corroborating the observation of Davezac et al. A similar study involving comparison of wild-type versus ΔF508-CFTR and ΔF508-CFTR(4RK), which lacks the four arginine-framed tripeptide (RXR) motif of CFTR, was performed in BHK cells.[31] In addition, a comparison of the proteome of BHK cell lines expressing wild-type or ΔF508-CFTR, grown at 37 °C or low temperature (28 °C), has been reported.[32] Gharib et al. have studied the patterns of protein expression in bronchoalveolar fluid (BALF) samples from CF and control patients to understand the mechanisms in the pathogenesis of CF lung disease.[33] Many of these expression-based studies have relied on an ability to resolve proteins by two-dimensional (2D) gel electrophoresis, which is limited because 2D gels are cumbersome to run, have a poor dynamic range, and are biased toward abundant and soluble proteins. The application of a non-gel, shotgun approach with the MudPIT technique enabled us to delve deeper into the whole-cell proteome to profile thousands of proteins to identify significantly deregulated proteins. This sensitivity is achieved mainly because MudPIT fractionates peptides by 2D liquid chromatography that can be directly interfaced with the ion source of a mass spectrometer. Moreover, label-free measurement of protein expression is simple and does not require prior labeling of proteins or peptides with heavy isotopes. Although not as precise as other methods of quantitative mass spectrometry, the semiquantitative nature of spectral counting enabled us to create a rough estimate of relative protein abundance, but in no way is an indicator of absolute protein concentrations.

In summary, we have observed that the consequence of ΔF508-CFTR expression in bronchial epithelial cells is quite striking in terms of protein deregulation compared to that in wild-type CFTR. Because the experiments were carried out in human airway epithelial cells, the study has more general relevance in the pathophysiology of CF. The quantitative data provide a list of statistically significant proteins with a fold change $\geq$ 1.5, and identification of these proteins in all three biological replicates in each condition provides better confidence in our results. The preliminary findings from this comparison require further validation of the differences observed as well as extension of the study to comparative proteomics analysis of primary cells or tissue biopsies from CF patients and healthy individuals. However, we need to keep in mind that data obtained with cell lines may not be representative of primary samples because cell culture conditions do not always reflect the in vivo microenvironment.

dx.doi.org/10.1021/pr500370g | J. Proteome Res. 2014, 13, 4668−4675

The challenge may also be exacerbated because of the tremendous genetic heterogeneity between individual patients, including modifier gene effects and environmental influences, that affect the disease's severity. Nevertheless, we believe that the differentially regulated proteins identified in this study that are presented both by effect size (i.e., fold change) and by statistical significance (i.e., *p* value) may serve as a useful resource for the CF community. The differentially regulated proteins belong to wide range of biological functions and may be involved in the underlying pathophysiology of a disease or as part of the body's response to the disease. Hence, monitoring the proteins representative of these classes of biological processes, although not specific for CF, are still of great potential utility to track disease progression and therapeutic intervention. The reported results could provide a basis for targeted functional studies of specific proteins and might potentially aid in developing a therapeutic strategy to correct misfolding and/or augment mutant CFTR expression at the plasma membrane. Once translocated to the membrane, the latter is capable of forming cAMP gated chloride channels with nearly normal conduction properties.

■ CONCLUSIONS

Deletion of a phenylalanine residue at position 508 in CFTR protein is the most prevalent disease-causing mutation in CF because it causes the protein to misfold, thereby negatively affecting its intracellular trafficking. In this study, we present a comprehensive comparative proteomic profiling of CFBE versus HBE cells using a LC−MS/MS-based approach, with the aim being to survey the molecular changes associated with expression of ΔF508-CFTR in bronchial epithelial cells. We incorporated analyses of three biological replicates for each of the HBE and CFBE samples to overcome the data-dependent variation in shotgun proteomic experiments and to obtain a statistically significant protein data set with improved quantification confidence. A total of 3140 proteins were identified in all six samples, among which 349 proteins showed statistically significant expression changes. Our data is consistent with the notion that some of the differentially regulated proteins are involved in protein folding and degradation among many other biological processes, and further investigation is needed to determine their relevant roles in CF. CF is a complex disease, and all of the observed changes may not be directly related to the presence of misfolded CFTR protein. Nevertheless, this study provides a scaffold upon which more directed and focused future studies can be built.

■ ASSOCIATED CONTENT

Supporting Information

Table S1 is the result of proteomic analysis of CFBE and HBE cells. The table contains three sheets that list the identified proteins in all three biological replicates from the HBE (1st sheet) and CFBE (2nd sheet) cells and 3140 common proteins among all biological replicates between these two samples that were used for further quantitative analysis (3rd sheet). Table S2 lists the 349 statistically significant differentially expressed proteins (1st sheet) in the CFBE versus HBE comparison; proteins that are identified in all three replicates in one or the other condition are referred to as "HBE specific" (2nd sheet) or "CFBE specific" (3rd sheet). Table S3 lists the differentially regulated proteins identified in our study that have a role in cystic fibrosis or CFTR processing. Figure S1 shows the density plot of distribution of natural log of NSAF values. Figure S2 shows the summary of identified proteins in each run and their cumulative IDs, whereas Figure S3 shows the number of proteins commonly identified among multiple MudPIT runs. Figure S4 illustrates the correlation of natural log of the NSAF values among biological replicates in each condition. Additional GO terms mapped by DAVID analysis of the statistically significant differentially expressed proteins is displayed in Figure S5. Figure S6 is a western blot image validating the expression level of a subset of proteins observed by mass spectrometry. This material is available free of charge via the Internet at http://pubs.acs.org. All of the raw data are available as Thermo.RAW files at http://fields.scripps.edu/published/CFTR_proteomics/.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 858-784-8862. Fax: 858-784-8883. E-mail: jyates@scripps.edu.

Present Address

⊥Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, United States.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Drs. Harsha Gowda and Claire Delahunty for their critical reading of the manuscript. Financial support was provided by NIH grants P41 GM103533 and R01 HL079442-09.

■ REFERENCES

(1) Cheng, S. H.; Gregory, R. J.; Marshall, J.; Paul, S.; Souza, D. W.; White, G. A.; O'Riordan, C. R.; Smith, A. E. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* **1990**, *63* (4), 827−34.

(2) Knowles, M. R. Gene modifiers of lung disease. *Curr. Opin Pulm. Med.* **2006**, *12* (6), 416−21.

(3) 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.; Cutting, G. R. Heritability of lung disease severity in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **2007**, *175* (10), 1036−43.

(4) Boyle, M. P. Strategies for identifying modifier genes in cystic fibrosis. *Proc. Am. Thorac. Soc.* **2007**, *4* (1), 52−7.

(5) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242−7.

(6) Ehrhardt, C.; Collnot, E. M.; Baldes, C.; Becker, U.; Laue, M.; Kim, K. J.; Lehr, C. M. Towards an in vitro model of cystic fibrosis small airway epithelium: characterisation of the human bronchial epithelial cell line CFBE41o. *Cell Tissue Res.* **2006**, *323* (3), 405−15.

(7) Hutt, D. M.; Herman, D.; Rodrigues, A. P.; Noel, S.; Pilewski, J. M.; Matteson, J.; Hoch, B.; Kellner, W.; Kelly, J. W.; Schmidt, A.; Thomas, P. J.; Matsumura, Y.; Skach, W. R.; Gentzsch, M.; Riordan, J. R.; Sorscher, E. J.; Okiyoneda, T.; Yates, J. R., 3rd; Lukacs, G. L.; Frizzell, R. A.; Manning, G.; Gottesfeld, J. M.; Balch, W. E. Reduced histone deacetylase 7 activity restores function to misfolded CFTR in cystic fibrosis. *Nat. Chem. Biol.* **2010**, *6* (1), 25−33.

(8) Tabb, D. L.; McDonald, W. H.; Yates, J. R., 3rd DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **2002**, *1* (1), 21−6.

(9) Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **2006**, *5* (9), 2339−47.
(10) Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76* (14), 4193−201.
(11) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **2005**, *4* (10), 1487−502.
(12) Zybailov, B. L.; Florens, L.; Washburn, M. P. Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. *Mol. BioSyst.* **2007**, *3* (5), 354−60.
(13) Chen, H.; Sharp, B. M. Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinf.* **2004**, *5*, 147.
(14) Maiuri, L.; Luciani, A.; Giardino, I.; Raia, V.; Villella, V. R.; D'Apolito, M.; Pettoello-Mantovani, M.; Guido, S.; Ciacci, C.; Cimmino, M.; Cexus, O. N.; Londei, M.; Quaratino, S. Tissue transglutaminase activation modulates inflammation in cystic fibrosis via PPARgamma down-regulation. *J. Immunol.* **2008**, *180* (11), 7697−705.
(15) Lorand, L.; Graham, R. M. Transglutaminases: crosslinking enzymes with pleiotropic functions. *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (2), 140−56.
(16) Luciani, A.; Villella, V. R.; Esposito, S.; Brunetti-Pierri, N.; Medina, D.; Settembre, C.; Gavina, M.; Pulze, L.; Giardino, I.; Pettoello-Mantovani, M.; D'Apolito, M.; Guido, S.; Masliah, E.; Spencer, B.; Quaratino, S.; Raia, V.; Ballabio, A.; Maiuri, L. Defective CFTR induces aggresome formation and lung inflammation in cystic fibrosis through ROS-mediated autophagy inhibition. *Nat. Cell Biol.* **2010**, *12* (9), 863−75.
(17) Dennis, G., Jr.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **2003**, *4* (5), P3.
(18) Glaab, E.; Baudot, A.; Krasnogor, N.; Schneider, R.; Valencia, A. EnrichNet: network-based gene set enrichment analysis. *Bioinformatics* **2012**, *28* (18), i451−i457.
(19) Amaral, M. D. CFTR and chaperones: processing and degradation. *J. Mol. Neurosci.* **2004**, *23* (1−2), 41−8.
(20) Jensen, T. J.; Loo, M. A.; Pind, S.; Williams, D. B.; Goldberg, A. L.; Riordan, J. R. Multiple proteolytic systems, including the proteasome, contribute to CFTR processing. *Cell* **1995**, *83* (1), 129−35.
(21) Ward, C. L.; Omura, S.; Kopito, R. R. Degradation of CFTR by the ubiquitin-proteasome pathway. *Cell* **1995**, *83* (1), 121−7.
(22) Yang, Y.; Janich, S.; Cohn, J. A.; Wilson, J. M. The common variant of cystic fibrosis transmembrane conductance regulator is recognized by hsp70 and degraded in a pre-Golgi nonlysosomal compartment. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90* (20), 9480−4.
(23) Schiene, C.; Fischer, G. Enzymes that catalyse the restructuring of proteins. *Curr. Opin. Struct. Biol.* **2000**, *10* (1), 40−5.
(24) Kopito, R. R. ER quality control: the cytoplasmic connection. *Cell* **1997**, *88* (4), 427−30.
(25) Harada, K.; Okiyoneda, T.; Hashimoto, Y.; Ueno, K.; Nakamura, K.; Yamahira, K.; Sugahara, T.; Shuto, T.; Wada, I.; Suico, M. A.; Kai, H. Calreticulin negatively regulates the cell surface expression of cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* **2006**, *281* (18), 12841−8.
(26) Swiatecka-Urban, A.; Brown, A.; Moreau-Marquis, S.; Renuka, J.; Coutermarsh, B.; Barnaby, R.; Karlson, K. H.; Flotte, T. R.; Fukuda, M.; Langford, G. M.; Stanton, B. A. The short apical membrane half-life of rescued {Delta}F508-cystic fibrosis transmembrane conductance regulator (CFTR) results from accelerated endocytosis of {Delta}F508-CFTR in polarized human airway epithelial cells. *J. Biol. Chem.* **2005**, *280* (44), 36762−72.
(27) Gentzsch, M.; Chang, X. B.; Cui, L.; Wu, Y.; Ozols, V. V.; Choudhury, A.; Pagano, R. E.; Riordan, J. R. Endocytic trafficking routes of wild type and DeltaF508 cystic fibrosis transmembrane conductance regulator. *Mol. Biol. Cell* **2004**, *15* (6), 2684−96.
(28) Henderson, M. J.; Singh, O. V.; Zeitlin, P. L. Applications of proteomic technologies for understanding the premature proteolysis of CFTR. *Expert Rev. Proteomics* **2010**, *7* (4), 473−86.
(29) Pollard, H. B.; Eidelman, O.; Jozwik, C.; Huang, W.; Srivastava, M.; Ji, X. D.; McGowan, B.; Norris, C. F.; Todo, T.; Darling, T.; Mogayzel, P. J.; Zeitlin, P. L.; Wright, J.; Guggino, W. B.; Metcalf, E.; Driscoll, W. J.; Mueller, G.; Paweletz, C.; Jacobowitz, D. M. De novo biosynthetic profiling of high abundance proteins in cystic fibrosis lung epithelial cells. *Mol. Cell Proteomics* **2006**, *5* (9), 1628−37.
(30) Davezac, N.; Tondelier, D.; Lipecka, J.; Fanen, P.; Demaugre, F.; Debski, J.; Dadlez, M.; Schrattenholz, A.; Cahill, M. A.; Edelman, A. Global proteomic approach unmasks involvement of keratins 8 and 18 in the delivery of cystic fibrosis transmembrane conductance regulator (CFTR)/deltaF508-CFTR to the plasma membrane. *Proteomics* **2004**, *4* (12), 3833−44.
(31) Gomes-Alves, P.; Couto, F.; Pesquita, C.; Coelho, A. V.; Penque, D. Rescue of F508del-CFTR by RXR motif inactivation triggers proteome modulation associated with the unfolded protein response. *Biochim. Biophys. Acta* **2010**, *1804* (4), 856−65.
(32) Gomes-Alves, P.; Neves, S.; Coelho, A. V.; Penque, D. Low temperature restoring effect on F508del-CFTR misprocessing: A proteomic approach. *J. Proteomics* **2009**, *73* (2), 218−30.
(33) Gharib, S. A.; Vaisar, T.; Aitken, M. L.; Park, D. R.; Heinecke, J. W.; Fu, X. Mapping the lung proteome in cystic fibrosis. *J. Proteome Res.* **2009**, *8* (6), 3020−8.