Responses of the human airway epithelium transcriptome to in vivo injury

Adriana Heguy, Ben-Gary Harvey, Philip L. Leopold, Igor Dolgalev, Tina Raman, and Ronald G. Crystal

Department of Genetic Medicine, Weill Medical College of Cornell University, New York, New York

Submitted 28 July 2006; accepted in final form 1 December 2006

The human tracheobronchial tree comprises 23 orders of dichotomous branching airways lined by a 0.2-m² continuous layer of pseudostratified epithelium comprising primarily basal cells, undifferentiated columnar cells, ciliated cells, and secretory cells (5, 31, 31). The airway epithelium plays a central role in maintaining the fluid and electrolyte balance in the epithelial lining fluid and in defending the lung from inhaled xenobiotics, particulates, gases, and pathogens (27, 30, 39). In the normal lung, the airway epithelium turns over slowly, but with injury, the epithelium rapidly regenerates, first to reestablish the continuity of the epithelial layer, and then to differentiate to reestablish the normal cell population (4, 12, 13, 35). In all of the major human airway disorders, including chronic obstructive lung disease, asthma, and bronchogenic carcinoma, there is enhanced, albeit disordered, airway epithelial regeneration (19, 27, 40).

Despite the importance of maintaining the airway epithelium in health and disease and the extensive data regarding the airway epithelium in culture and in experimental animals (11, 21, 22, 26, 32, 37, 48, 49), little is known about the genetic program that mediates the regeneration of the human airway epithelium (32). To help define this genetic program, we have developed a strategy to force the epithelium to undergo regeneration in vivo with fiber-optic bronchoscopy and mechanical brushing of a small region of the airway epithelium of normal individuals. Based on the preliminary studies showing the denuded region of epithelium is fully covered by day 7 following injury, and fully differentiated by day 14, the area of injury was sampled at these time points. To assess the genetic program that participates in the repair reprogramming of the human airway epithelium as it regenerates in situ, we used microarray technology with TaqMan real-time RT-PCR confirmation to compare relative mRNA levels at days 7 and 14 postinjury to that of the resting epithelium. The characterization of the airway epithelial repair transcriptome at these postinjury time points is reported here. This repair signature of healthy individuals provides a baseline of the functional gene categories participating in the process of normal human airway epithelial repair that can be used in future studies of injury and repair in human airway epithelial diseases.

METHODS

Study population. Normal individuals were recruited by posting ads in local newspapers. All were evaluated in the Weill Cornell National Institutes of Health General Clinical Research Center under an Institutional Review Board-approved clinical protocol. Individuals were determined to be phenotypically normal based on standard history, physical exam, complete blood count, coagulation studies, liver function tests, urine studies, chest X-ray, EKG, and pulmonary function tests. The study population consisted of nine phenotypically normal individuals (six males and three females, including five African Americans, and four Caucasians, age 40 ± 2 yr SE). Of the nine, four were nonsmokers and five current cigarette smokers (23.8 ± 1.8 pack-yr SE). The study population and study design, including the pairing of the pairing of samples for microarray analysis, are described in detail in Table 1, which shows age, sex, ethnic group, and smoking history for the study volunteers.

Injury/repair sampling of the airway epithelium. To obtain pure populations of the airway epithelium before and after injury, the epithelium of the large (2nd–3rd order) bronchi of each individual was sampled at rest and then the same individuals were resampled following mechanical (brush) injury. All nine individuals were assessed at rest (one individual was assessed twice at day 0, right and left lung), seven were resampled at day 7 following injury and five at day 14, yielding a total of 22 epithelial samples. The samples obtained and how they were paired for microarray analysis, are described in Table 1.

For all sampling, the subjects were mildly sedated with meperidine and midazolam, and local anesthesia (lidocaine) was applied to the vocal cords and airways. Fiber-optic bronchoscopy was carried out with a Pentax FB-15X fiber-optic bronchoscope (EB-1530T3, Orangeburg, NY), positioned proximal to the opening of the desired lobar bronchus. Disposable brushes (2.0 mm; Ballard Medical, Draper, UT) were advanced through the working channel of the bronchoscope and used to brush-collect airway epithelial cells as previously described by gently gliding the brush back and forth 10 times in different locations in the same general area of the 2nd or 3rd order bronchi of the right or left lower lobe (9, 16, 17, 25).
Brushing of the airway epithelial cells denuded the epithelium, thus stressing the bordering epithelium to repair the injured area. Proximal airway epithelial cells were consistently and sequentially obtained from six different areas of the right lower lobe, these areas include: Site 1, the opening of the superior segment of the right lower lobe at four coordinates: 12, 3, 6, and 9 o’clock; Site 2, the opening of the medial basal segment of the right lower lobe at 12 and 6 o’clock; Site 3, the main lobar bronchus between the superior and the medial basal segment of the right lower lobe at 12, 3, and 6 o’clock; Site 4, the opening of the anterior segment of the right lower lobe, at 12 and 6 o’clock; Site 5, the opening of the lateral segment of the right lower lobe, at 12 and 6 o’clock; and Site 6, the opening of the posterior segment of the right lower lobe, at 12 and 6 o’clock. At day 0, all six areas were sampled. Some individuals had the left lung sampled, and for those individuals, only five sites were sampled, since the medial basal segment is not present in the lower lobe of the left lung. For each coordinate within each site described above, the brush was gently glided back and forth for a distance of 2–3 cm, ~10 times. Digital photographs were taken at each area, and notes were carefully taken; these were reviewed before each sequential sampling. Some of the areas were not brushed but used exclusively for biopsy, which additionally confirmed the injury and the location. The diameter of the brushed areas at sites 1, 2, and 4–6 was ~0.8–1 cm, and at site 3, ~2 cm.

For the three individuals who underwent three bronchoscopies (days 0, 7, and 14), two day 0 samples were collected, one from the right lower lobe and one from the left lower lobe (one served as the area where the day 7 sample was collected, and the other was the area for day 14). For all these individuals, except for smoker S3, only one “day 0” microarray was assessed (the first one collected), since we have previously shown that the transcriptional profile of samples collected from the left and right lungs and over time is very similar (16, 17). For smoker S3, the reproducibility between the resting samples (day 0) obtained from the right lung and left was assessed by correlating expression levels for all probe sets in the right and left lung samples ($r^2 = 0.67$, Supplemental Fig. 1). (The online version of this article contains supplemental material.) All samples were maintained independently of one another through the procedure and sample processing. The collected cells were detached from the brush by flicking into 5 ml of bronchial epithelial basal cell medium (BEBM; Clonetics, Walkersville, MD). An aliquot of 0.5 ml was used for differential cell count. The remainder was processed immediately for RNA extraction. We determined the total number of cells recovered by bronchial brushing from large airways by counting on a hemocytometer.

To determine the morphology of the recovered cells and to quantify the percentage of epithelial and inflammatory cells as well as the cell differentials (proportions of ciliated, basal, secretory, and undifferentiated epithelial cells), the aliquot reserved for differential cell count ($2 \times 10^4$ cells) was prepared by centrifugation (Cytospin 11; Shandon Instruments, Pittsburgh, PA), and the cells were assessed by staining with Diff-Quik (Dade Behring, Newark, NJ).

To visualize the architecture of the epithelium before and after injury, areas different from those used for assessment of the repair transcriptome were assessed by endobronchial biopsy. The biopsies were carried out by advancing a forceps through the working channel of the bronchoscope under direct visualization to the desired areas. Biopsies for day 0 were taken at a site unrelated to the brushing area. Photographs of the brushed areas taken at day 0 were used to identify the injured area, as described above. The tissue was embedded in paraffin, sectioned, and stained with hematoxylin and eosin.

**RNA extraction and microarray procedures.** Total RNA was extracted by a modified version of the TRIZol method (Invitrogen, Carlsbad, CA), where pelleted airway epithelial cells are immediately lysed with the TRIZol reagent, and after the chloroform extraction step, RNA was purified from the aqueous phase, omitting the alcohol precipitation step (which minimizes the risk of losing the sample), by RNaseasy purification (RNeasy MinElute RNA purification kit; Qiagen, Valencia, CA). At the end of the extraction procedure, all RNA samples were stored in RNA Secure (Ambion, Austin, TX), in RNase-free microcentrifuge tubes, at −80°C. An aliquot of each RNA sample was evaluated using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). Three quality control criteria were used for an RNA sample to be accepted for further processing: 1) A260/A280 ratio between 1.7 and 2.3; 2) concentration within the range of 0.2–6 μg/ml; and 3) Agilent electropherogram displaying two distinct peaks corresponding to the 28S and 18S ribosomal RNA bands at a ratio of 28S/18S of >0.5 with minimal or no degradation (Agilent RNA Integrity Number >0.5 for all samples). Double-stranded cDNA was synthesized from 3 μg of total RNA using the GeneChip One-Cycle cDNA Synthesis Kit, followed by cleanup with GeneChip Sample Cleanup Module, in vitro transcription (IVT) reaction using the GeneChip IVT Labeling Kit, and clean-up and quantification of the biotin-labeled cRNA yield by spectrophotometric analysis (all kits were from Affymetrix, Santa Clara, CA). Hybridizations to test chips and to the HG-U133 Plus 2.0 microarray were performed according to Affymetrix protocols, processed by the Affymetrix GeneChip Fluidics Station 450, and scanned with an Affymetrix GeneChip Scanner 3000 7G. Overall microarray quality was determined by taking in consideration the percentage of “present” genes, the 3/5’ signal ratio for control genes and the average background signal, as recommended (http://www.affymetrix.com/support/technical/manual/expression_manual.affx). The following cut-offs for

---

**Table 1. Demographics of the study population and study design**

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Age, yr</th>
<th>Sex</th>
<th>Race</th>
<th>Pack-yr</th>
<th>Day 0</th>
<th>Day 7</th>
<th>Day 14</th>
<th>Day 0</th>
<th>Day 7</th>
<th>Day 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS1</td>
<td>35</td>
<td>M</td>
<td>B</td>
<td></td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>NS2</td>
<td>36</td>
<td>M</td>
<td>W</td>
<td></td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>NS3</td>
<td>40</td>
<td>M</td>
<td>W</td>
<td></td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>NS4</td>
<td>35</td>
<td>M</td>
<td>B</td>
<td></td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>S1</td>
<td>40</td>
<td>F</td>
<td>B</td>
<td>25</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>S2</td>
<td>46</td>
<td>M</td>
<td>W</td>
<td>21</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>S3</td>
<td>44</td>
<td>F</td>
<td>W</td>
<td>30</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>S4</td>
<td>41</td>
<td>M</td>
<td>B</td>
<td>20</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>S5</td>
<td>46</td>
<td>F</td>
<td>B</td>
<td>23</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Smoking history is indicated in pack years (pack-yr). 1 Pack-yr = 1 pack of cigarette per day during 1 year; M, male; F, female; B, black; W, white; Y, yes. Samples were collected and RNA was extracted as described in METHODS and analyzed on HG-U133 Plus 2.0 microarrays; all samples were processed independently. For Smoker S3, two day 0 samples were collected and analyzed by microarray. R, right lung sample; L, left lung sample.
microarray quality control were used: RawQ ≤ 4; average background ≤ 150; % Present calls > 40; and 3’/5’-GAPDH ratio > 3.5. For this set of 22 samples, the noise averaged 2.5, with a range of 1.9 – 4.0; the average background was 70.9 (range 52.1 – 133.5); the percentage of present calls was 48.8 on average (range 44.1 – 52.5%), and the average 3’/5’-GAPDH ratio was 1.6 (range 0.9 – 3.2). In addition, we used Harshlight, an R-package that detects blemishes in high-density oligonucleotide arrays, by calculating an error image based on the variations of a chip’s log-intensities from other chips in an experiment (41). There are no specified cutoffs for this software, but the recommendation is that all arrays within an experiment should be within a comparable range. For this set of 22 samples, compact defects covered an average of 0.02 ± 0.01% of the surface area, while diffuse defects covered an average of 3.77 ± 0.55% of the surface area.

Microarray data analysis and statistics. The starting list included all 54,675 probe sets in the HG-U133 Plus 2.0 array; the cross-gene error model from GeneSpring GX 7.2 was used as a way to estimate measurement precision by combining measurement variation and between-sample variation information based on “deviation from 1,” which assumes a general-purpose array was used, where most genes have little biological variability, as is the case with the Affymetrix HG-U133 Plus 2.0 array. A two-pronged statistical approach was used to analyze the microarray data comparisons of day 0 vs. day 7 and 14 postinjury, including: 1) the Benjamini-Hochberg correction (3, 38) and 2) two different algorithms for generation of expression values from the image files: the MAS5 method (Affymetrix Microarray Suite Version 5 software, which takes into account the perfect match and the mismatch probes) and the RMA algorithm (23) [robust multiarray average (RMA) method that utilizes only perfect match information] using GeneSpring version 7.2 software (Agilent Technologies). This strategy identifies the most robust changes in gene expression by curtailing the number of false positives, although it likely eliminates some potential true “hits” (1, 2). MAS5-analyzed data were normalized using GeneSpring, by setting measurements <0.01 to 0.01, followed by per-chip normalization to the 50th percentile of the measurements for the array, and per-gene by normalizing to the median measurement for the gene across all the arrays in the data set. RMA preprocessed data were normalized to the median measurement for the gene across all the arrays in the data set, because the other background and quantile normalization steps are included in the preprocessing algorithms (2, 23).

Wilcoxon signed rank test for paired samples (day 0 vs. day 7, and day 0 vs. day 14) was conducted in S-PLUS 7.0 (Insightful, Seattle, WA). For each gene, P values were calculated from two-sided exact test, with the normal approximation used to obtain P values when there were ties in the ranks. The Wilcoxon-Mann Whitney test for nonpaired samples was carried out using GeneSpring, for seven day 7 vs. five day 14 samples. To limit the false positive rate, the Benjamini-Hochberg correction (3, 38) was applied to P values calculated by the Wilcoxon signed rank test, for both MAS5- and RMA-generated data, using the 38,176 probe sets that had a MAS5 call of “present” in at least 1 of the 22 arrays in the sample set. To identify robust changes in gene expression, without bias towards any strategy, only genes present in both significant gene lists, generated from the MAS5 and RMA preprocessing algorithms, were considered significant. An additional criterion included a more than twofold change between day 0 and day 7 samples (up- and downregulated genes) in either the MAS5 or the RMA-generated lists of significant genes, and a >1.5-fold change in the other algorithm (i.e., if a greater than twofold change was observed in MAS5, then >1.5-fold was required of the RMA data, and vice versa). To investigate if smoking had an effect on the airway epithelium repair transcriptome, a two-way ANOVA was carried out on the significant genes with time (day 0 vs. day 7) as one parameter and smoking as a second parameter, using a parametric test and assuming equal variances and applying the Benjamini-Hochberg correction (3, 38). The resulting interaction P value was >0.2 for all genes, and although this suggests that smoking may not have a significant effect on the repair transcriptome, it is possible that there are differences that are not detected due to the relatively small number of smoker and nonsmoker samples.

Functional annotation was carried out using the NetAffx Analysis Center (www.affymetrix.com) to retrieve the Gene Ontology (GO) annotations from the National Center for Biotechnology (NCBI) databases. For probe sets with no GO annotations, other public databases [Human Protein Reference Database, Kyoto Encyclopedia of Genes and Genomes (KEGG), PubMed] were searched. All microarray data have been deposited at the Gene Expression Omnibus site (http://www.ncbi.nlm.nih.gov/geo; accession number GSE5372).

Hierarchical clustering was carried out for the 1,196 genes considered significantly changed according to the above-described cutoffs, using the MAS5-analyzed data, with the Spearman correlation as similarity measure, and the complete linkage clustering algorithm, for all samples and the 1,196 genes, using the GeneSpring software.

**TagMan real-time RT-PCR.** To confirm the microarray results, the relative expression of 14 representative genes from the cell cycle and signal transduction categories was measured by TaqMan real-time RT-PCR using the seven paired day 0 and day 7 samples. Premade TaqMan Gene Expression Assays were purchased from Applied Biosystems (Foster City, CA). The relative quantitation method (ΔΔCt) was used, with the ratio of the mRNA level for the gene of interest normalized to the level of ribosomal RNA as the internal control and the average of the day 0 samples as the calibrator value. The PCR reactions were run in an Applied Biosystems Sequence Detection System 7500. The significance of the TaqMan data was calculated by ANOVA using StatView (SAS Institute, Cary, NC); all P values were <0.05 for day 0 vs. day 7 comparison.

**RESULTS**

Morphology of the human airway epithelium during in vivo repair. Endobronchial biopsies obtained immediately following brushing-induced injury showed a completely denuded area with no airway epithelial layer and an exposed interstitial connective tissue layer (Fig. 1A, compare “day 0, pre” to “day 0, post”). Seven days following injury, the injured area was completely covered by epithelial cells, the majority of which appear undifferentiated (Fig. 1A, “day 7”), while at day 14 postinjury (Fig. 1A, “day 14”), the airway epithelial architecture approached that of naive airway epithelium.

Quantification of the samples obtained by brushing demonstrated that the percentage of inflammatory cells was <1% at day 0 preinjury, in agreement with previous observations from our laboratory (16). The percentage of inflammatory cells was also <1% at days 7 and 14 following injury, indicating that at these time points postinjury, any inflammation caused by the brushing-induced injury had subsided (day 7 was chosen as the earliest postinjury time point after preliminary airway epithelial sampling at earlier time points resulted in contamination with inflammatory cells).

Consistent with the biopsy samples, the proportion of undifferentiated cells at day 7 postinjury was much higher than in the resting airway epithelium (P < 0.001) and the proportion of ciliated cells much lower (P < 0.001, Fig. 1B). Whereas the proportion of ciliated cells decreased by approximately half at day 7 postinjury, by day 14 the percentage of ciliated cells was close to resting state. The proportion of basal cells (P < 0.001) also decreased at day 7 postinjury as did the % secretory cells, although the latter was not significant (P > 0.08). By day 14, the percentage of both were close to the normal values (both P > 0.2).
Gene expression profile of repairing human airway epithelium. Global changes in airway epithelial expression in response to mechanical injury were evaluated by comparison of the mean expression value for each of the 54,675 probe sets in the Affymetrix HG-U133 Plus 2.0, at day 7 (Supplemental Fig. S2A) and at day 14 (Supplemental Fig. S2B), each time point vs. that of corresponding paired day 0 samples. This type of analysis shows that, although as expected the majority of expression values do not change postinjury, day 7 samples exhibit a much larger number of gene expression changes greater than twofold compared with day 0 samples than the equivalent comparison of day 14 vs. day 0 samples, where the pattern of gene expression is very similar. Most of the changes that occur at day 7 involve an increase in gene expression levels, suggestive of genes being turned on during airway epithelial repair.

The seven paired day 0 and day 7 samples were compared using the Wilcoxon matched pairs signed rank test. A total of 7,288 (RMA analysis) and 8,934 probe sets (MAS5 analysis) were <0.05, if a false discovery rate [Benjamini-Hochberg (3, 38)] correction was applied to a list of 38,176 genes that were “expressed” in at least one array, with no additional cutoff for fold change. To identify robust changes at day 0 vs. day 7, two independent algorithms for processing image data (MAS5 and RMA), together with the Benjamini-Hochberg false discovery rate correction (3, 38), a P value <0.05 in both MAS5 and RMA-analyzed data (5,515 probe sets), and a greater than twofold change between day 7 and day 0 samples (in either the MASS- or the RMA-generated data) were used as criteria, as described in METHODS. This resulted in a list of 1,196 significant genes, representing the signature profile of the repairing airway epithelium 7 days following mechanical injury in normal human individuals. Application of the same statistical analysis and criteria resulted in an absence of statistically significant changes when day 14 samples were compared with day 0 samples, indicating that the gene expression profile of the airway epithelium 14 days after injury is similar to that of uninjured airway epithelium, as suggested by the global analysis of gene expression. To estimate the variability between samples at each time point, we calculated the coefficient of variation for these 1,196 significantly changed genes, using the RMA-generated data. Variability was low at all three time points, indicating the response to injury is fairly homogeneous for all individuals in this study: at day 0, the average coefficient of variation was 40.1 ± 20.5%; at day 7 it was 36.2 ± 19.1%, and at day 14 it was 40.5 ± 18.3%.

The 1,196 significant day 7 to day 0 genes were classified by functional categories according to the GO classification and other public databases, with 761 genes of known function and 435 genes for which no functional annotations were available. The complete list with fold changes and P values for these 1,196 genes can be found in Supplemental Table S1. The largest categories of significantly different day 7 to day 0 genes were cell cycle, signal transduction, metabolism, transport, and transcription/chromatin (Fig. 2, A and C). The gene expression changes with the highest magnitude were in the cell cycle category, where most genes were upregulated at day 7 postinjury vs. naive epithelium. All the other categories included genes with either increased or decreased expression at day 7 postinjury. By day 14 postinjury, the expression of these signature airway epithelium repair genes was similar to that of resting airway epithelium, with apparent changes of a small magnitude (Fig. 2, B and D). None of these changes of gene expression levels at day 14 to day 0 were significant, indicating that at the transcription level, the restitution to pseudostratified differentiated epithelium was near completion 14 days after injury. Analysis of gene expression changes between the seven day 7 and the five day 14 samples in our study population was carried out to further test this concept. We considered 200 probe sets differentially expressed (P < 0.05, >2-fold change between day 7 and day 14 samples in either the MAS5 or the RMA-generated lists of significant genes, and a >1.5-fold change in the other algorithm). Despite the differences in statistical power and type of test applied (see METHODS), 81% of genes differentially expressed at day 7 vs. day 14 were also differentially expressed at day 7 vs. day 14 (Supplemental Table S2, common genes highlighted). In particular, cell cycle genes were once again the most dominant functional category. This type of analysis further highlights the observation that day 14 samples are similar to the resting day 0 samples.
The pattern of gene expression of the 1,196 genes differentially expressed at day 7 was investigated by hierarchical clustering using all day 0, day 7, and day 14 samples (Fig. 3). This unsupervised analysis shows that the day 7 samples cluster together and separate from day 0 samples, as expected. Interestingly, four out of five day 14 samples clustered with the day 0 samples and one with day 7 samples. This shows that the profile of gene expression of day 14 samples is not identical to day 0 samples and that in spite of the lack of statistical significance between day 0 and day 14 genes, at least some of these 1,196 genes take longer to return to resting levels. In addition, the cluster analysis suggests that there may be interindividual variability in the time frame that it takes to complete the restoration to a completely repaired epithelium.

Expression of cell cycle genes in the repairing airway epithelium. Given the prominent role of cell cycle genes in the repair transcriptome, we further annotated the list of 124 cell cycle probe sets that were significantly changed at day 7, as shown in Fig. 2. The fold-changes were calculated by dividing the geometric mean of day 7 samples by that of paired day 0 samples, and that of day 14 samples by that of paired day 0 samples. The gene expression values were generated using the MAS5 algorithm. Genes downregulated at day 7 or at day 14 vs. day 0 are shown as negative gene expression changes. The genes were categorized according to the Gene Ontology annotations. The abscissa shows the individuals genes, randomly ordered by probe set ID within each category. Categories are listed in descending order according to the number of genes in each category.
postinjury to determine their distribution in the different phases of the cell cycle. For this purpose, we used the KEGG cell cycle pathway (Homo sapiens), as well as the literature utilizing microarray technology describing the temporal program of gene expression in synchronized human fibroblasts and various cell lines (6–8, 24, 50), to categorize the 124 repair-related cell cycle-regulated transcripts within the different cell cycle phases (Fig. 4; see Supplemental Table S3 for the fully annotated list). Interestingly, the majority (60.5%) of these genes belong to the G2 and M phases of the cell cycle, suggesting that at day 7, the cells were in relative synchrony, perhaps because the highly efficient airway epithelial repair has reached its last cell division. This is in agreement with the morphology of the airway epithelium at day 7, which showed complete coverage of the denuded area (Fig. 1).

Confirmation of microarray results by real-time PCR. To confirm the microarray observations, relative genes expression levels at day 0 and day 7 samples were measured using an alternative method to determine mRNA levels, TaqMan RT-PCR, for 14 genes that were differentially expressed at day 7 (Fig. 5). To directly compare the two methodologies, TaqMan expression levels were normalized by dividing individual values by the median expression level of all samples, as was done for microarray analysis. These genes included 12 cell cycle genes [tumor protein p73 (TP73), caspase 3 (CASP3); caveolin 1 (CAV1), cyclin A1 (CCNA1), polo-like kinase 4 (PLK4), growth arrest-specific 2 like 3 (GAS2L3), cyclin B2 (CCNB2), cell division cycle 2 (CDC2), cell division cycle 20 homolog (CDC20), discs, large homolog 7 (Drosophila) (DGL7); kinesin family member 23 (KIF23), and NIMA (never in mitosis a)-related kinase 2 (NEK2)] as well as 2 signal transduction genes [dickkopf homolog 1 (DKK1) and neuromedin U (NMU)].

DISCUSSION

The goal of the present study was to identify genes associated with the response of the human airway epithelium to injury in vivo in healthy individuals. With microarray-based analysis of samples obtained before and after brushing-induced in vivo injury to the airway epithelium, and strict analytic criteria combining two different strategies to assess the microarray-derived expression values, the data define the human airway epithelial repair transcriptome at 7 and 14 days after injury. The expression levels of >1,000 genes were significantly altered in association with airway epithelial repair 7...
days following injury, a time where the epithelium covered the entire denuded area but was still relatively undifferentiated. The transcripational signature at day 7 was dominated by cell cycle genes, most of which belonged to the G2 or M phases of the cell cycle. Other major categories of genes up- or down-regulated 7 days after injury included genes involved in signal transduction, metabolism, transport, and transcription. At day 14 postinjury, there were no major gene expression changes compared with day 0, although cluster analysis revealed differences in pattern of expression between some day 14 and day 0 samples. The similarity of gene expression pattern between day 0 and day 14 samples is in agreement with histology showing that the restitution of the pseudostratified, differentiated airway epithelial structure is almost complete by day 14.

Models to study airway epithelium repair and differentiation. The airway epithelium has a critical function as a barrier against environmental insults and stresses, and its integrity is fundamental for maintaining lung homeostasis (13, 30, 35, 39). In healthy individuals, airway epithelial repair is an ongoing process in response to cell senescence and injury, characterized by dedifferentiation, cell proliferation and subsequent redifferentiation to maintain the normal pseudostratified epithelial lining of the airways (12, 13, 27, 30). Based on animal models, the turnover time of the airway epithelium is estimated at 2–59 days (4).

Several experimental strategies have been used to study the regeneration of airway epithelium following injury. In vitro culture studies are based on the ability to grow primary airway epithelial cells as well as epithelium-derived cell lines as a monolayer. Both mechanical (scrapping) and chemical injury have been used to measure the cellular response to injury in vitro (10, 11, 18, 26, 43, 44, 47, 52, 53). In vivo models of airway injury complement the in vitro models, since they also account for the interaction of airway epithelial cells with the host. Airway epithelial injury has been induced in vivo in rodents or rabbits using toxic substances (i.e., naphthalene, bleomycin, nitrogen dioxide, ozone) or by mechanical means (12, 15, 20–22, 32, 34, 36, 42, 48). Mechanical wounding of guinea pig airway epithelium using a steel brush showed that the first event in repair was dedifferentiation of the cells surrounding the edges of the wound, followed by proliferation and then redifferentiation of the epithelium, which in guinea pigs appeared complete 5 days after inducing the injury (12). Mechanical injury of human bronchial airway xenografts transplanted subcutaneously into immunodeficient mice has also been used to study airway epithelium repair (12).

Although our in vivo model has advantages over in vitro cell culture systems in that it is more physiological, the presence of inflammatory cells at early time points, with consequent inclusion of inflammatory cell transcripts in the analysis, precludes assessment of the earliest changes in airway epithelial gene expression in response to injury. However, the in vivo model offers the unique opportunity to identify genes that are relevant to human airway epithelial repair in a relatively early stage (day 7 postinjury) where proliferation is actively occurring and redifferentiation has initiated but is far from complete.

Genes expressed at day 7 human airway epithelial repair. A striking feature of the gene expression profile at day 7 following human airway epithelial repair is the increased levels of many late stage cell cycle genes encoding factors in G2 and M phases of the cell cycle. This observation suggests either that the proliferative stage of airway epithelial repair at day 7 is partly synchronized or that the proliferative phase of repair is in its last cell divisions, a concept compatible with the morphological assessment of the airway epithelium at day 7 postinjury showing complete coverage of the denuded area.

Among the changes in gene expression at day 7 was upregulation of the wnt/β-catenin pathway inhibitor dickkopf 1 (DKK1) (14), suggesting that the wnt/β-catenin pathway is inhibited at this stage of airway epithelial repair. The wnt/β-
catenin pathway plays an important role in the control of both mesenchymal and epithelial differentiation during lung development (46), and suppression of wnt activity in developing lungs is associated with cell proliferation (28).

Since the airway epithelium at day 7 postinjury consists of a significant proportion of undifferentiated cells, but at day 14 there is differentiation into the different airway epithelial cell types, with histology resembling that of naive epithelium, it is reasonable to assume that redifferentiation of the newly formed airway epithelium probably initiates before day 7 and occurs throughout the 2nd week postinjury. The airway epithelial repair signature at day 7 included several transcription factors and signal transduction molecules with a potential role in airway epithelial differentiation, either previously described to be associated with airway repair and/or epithelial differentiation or belonging to protein families associated with repair or differentiation. For example, in mouse models of naphthalene injury or pneumonectomy, Sox and Fox family members have been implicated in the regeneration of the airway epithelium (29, 32, 33, 51). Among the transcription factors upregulated at day 7 included SOX4 and SOX7, two members of the Sox family of developmentally regulated transcription factors implicated in a wide range of developmental/differentiation processes, including bronchial epithelium differentiation in vitro and in mice (32, 33). FOXN4, a member of the forkhead box transcription factor family was also upregulated consistent with studies implicating members of this family in lung morphogenesis (29, 45, 51).

Fig. 5. Confirmation of microarray data by TaqMan RT-PCR. Expression levels of 14 cell cycle and signal transduction genes identified to be differentially expressed at day 7 postinjury by microarray analysis with the Affymetrix HG-U133 Plus 2.0 chip were confirmed with TaqMan real-time RT-PCR using the same samples that were used for microarray analysis. To allow direct comparisons of values obtained by the two independent methods, we normalized TaqMan expression levels by dividing individual values by the median expression level of all samples, as was done for microarray analysis. Relative expression levels (ordinate) are shown for 14 genes differentially expressed at day 7 postinjury. Top: genes in G1, S, or G2 phases of the cell cycle - tumor protein p73 (TP73), caspase 3 (CASP3), caveolin 1 (CAV1), and cyclin A1 (CCNA1). Middle: genes in G2/M phases of the cell cycle - cyclin B2 (CCNB2), cell division cycle 2 (CDC2), cell division cycle 20 homolog (CDC20), discs, large homolog 7 (Drosophila) (DGL7), kinesin family member 23 (KIF23), NIMA (never in mitosis a)-related kinase 2 (NEK2). Bottom: genes in the M phase of the cell cycle: polo-like kinase 4 (PLK4), cell cycle gene that has not yet been assigned to a phase - growth arrest-specific 2-like 3 (GAS2L3), and signal transduction genes - dickkopf homolog 1 (DKK1), and neuromedin U (NMU). Shown are mean ± SE. White bars are day 0 levels, black bars are day 7 levels; P < 0.05, all comparisons, day 7 vs. day 0 for TaqMan and for microarray data.
downregulated at day 7 included categories that would be expected in actively repairing epithelium, including genes involved in metabolism, stimulation of cell growth, cytoskeletal development, and regulation of the extracellular matrix, such as matrix components, proteases and antiproteases, and adhesion molecules.

It is interesting that there was no clear signature for differentiation of any specific epithelial cell types at day 7, including basal, ciliated, and secretory cells, all of which clearly differentiate after injury or lung growth (12, 21, 22, 32). It is possible that the genes relevant for these specific cell types are upregulated at different times points postinjury but were missed by our analysis of the time points of 7 and 14 days. Alternatively, several different transcriptional programs could be activated in each of these cells types as a result of injury and that therefore, no distinct redifferentiation signature can be distinguished in this in vivo model. It is also possible that the largest increase in expression for airway epithelial differentiation factors occurs at time points earlier than day 7 postinjury and, therefore, could not be identified with our human in vivo model, which is limited by the presence of inflammatory cells at early postinjury time points.

In conclusion, we characterized for the first time the gene expression changes occurring in vivo in the repairing healthy human airway epithelium in response to injury. The hallmark of the transcriptional response to acute airway injury is up-regulation of cell cycle regulators, consistent with a need to cover the denuded area with rapidly proliferating cells to restore the epithelial barrier. In addition, we have shown that factors involved in lung development also play a role in airway epithelial repair, possibly to suppress differentiation while active proliferation occurs. Our study provides a baseline for similar studies in human diseases where defects in airway epithelial repair play a role in their pathophysiology.

ACKNOWLEDGMENTS

We thank Drs. Fei Ji (Laboratory of Statistical Genetics, Rockefeller University) and Martin Lesser (Biostatistics Unit, Feinstein Institute for Medical Research, and General Clinical Research Center, Weill Cornell Medical College) for assistance with statistical analysis, the Microarray Core Facility for help with sample processing, the research coordinators from the Department of Genetic Medicine Clinical Operations and Regulatory Affairs Core for technical support and help with recruiting the volunteers for the study, and N. Mohamed for help in preparing this manuscript.

GRANTS

These studies were supported, in part, by National Institutes of Health Grants R01 HL-074326 and M01RR-00047 and by the Will Rogers Memorial Fund, Los Angeles, CA.

REFERENCES
