

A Systems Biology Approach Reveals the Dose- and Time-Dependent Effect of Primary Human Airway Epithelium Tissue Culture After Exposure to Cigarette Smoke *In Vitro*

Carole Mathis¹, Stephan Gebel², Carine Poussin^{1,*}, Vincenzo Belcastro^{1,*}, Alain Sewer^{1,*}, Dirk Weisensee², Arnd Hengstermann², Sam Ansari¹, Sandra Wagner¹, Manuel C. Peitsch¹ and Julia Hoeng¹

¹Philip Morris International R&D, Philip Morris Products S.A., Neuchâtel, Switzerland. ²Philip Morris International R&D, Philip Morris Research Laboratories GmbH, Cologne, Germany. *Contributed equally to this work.

ABSTRACT: To establish a relevant *in vitro* model for systems toxicology-based mechanistic assessment of environmental stressors such as cigarette smoke (CS), we exposed human organotypic bronchial epithelial tissue cultures at the air liquid interface (ALI) to various CS doses. Previously, we compared *in vitro* gene expression changes with published human airway epithelia *in vivo* data to assess their similarities. Here, we present a follow-up evaluation of these *in vitro* transcriptomics data, using complementary computational approaches and an integrated mRNA–microRNA (miRNA) analysis. The main cellular pathways perturbed by CS exposure were related to stress responses (oxidative stress and xenobiotic metabolism), inflammation (inhibition of nuclear factor- κ B and the interferon gamma-dependent pathway), and proliferation/differentiation. Within post-exposure periods up to 48 hours, a transient kinetic response was observed at lower CS doses, whereas higher doses resulted in more sustained responses. In conclusion, this systems toxicology approach has potential for product testing according to “21st Century Toxicology”.

KEYWORDS: air liquid interface, cigarette smoke, mechanistic investigations, organotypic culture, systems biology

CITATION: Mathis et al. A Systems Biology Approach Reveals the Dose- and Time-Dependent Effect of Primary Human Airway Epithelium Tissue Culture After Exposure to Cigarette Smoke *In Vitro*. *Bioinformatics and Biology Insights* 2015;9:19–35 doi: 10.4137/BBI.S19908.

RECEIVED: September 07, 2014. **RESUBMITTED:** December 03, 2014. **ACCEPTED FOR PUBLICATION:** December 04, 2014.

ACADEMIC EDITOR: J.T. Efrid, Associate Editor

TYPE: Original Research

FUNDING: The research described in this article was supported by Philip Morris International. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

CORRESPONDENCE: julia.hoeng@pmi.com

Paper subject to independent expert blind peer review by minimum of two reviewers. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Introduction

The human lung epithelium is continuously exposed to various environmental stressors, including cigarette smoke (CS). In line with “21st Century Toxicology”, new strategies for a systems biology-based risk assessment of environmentally hazardous substances, such as tobacco products, are currently under development.^{1–4} The establishment of reliable *in vitro* systems is extremely important in this because human testing is very limited, and animal studies are expensive and laborious, with disputable translatability to humans.^{5,6}

Cell culture models using primary human cells are valuable screening tools for both risk assessment purposes and mechanistic investigations. The most common *in vitro* models used to investigate the effect of aerosols on human lung epithelium mainly consist of monolayers of lung-derived epithelial cells. However, although transformed (tumor-derived or immortalized) cells are readily available, they have a number of limitations. These include (i) their inherent tumorigenicity (eg, BEAS-2B and A549 cells); (ii) their ability to only partially mimic normal bronchial epithelial cell behavior (eg, A549, a type-2 alveolar cell-like adenocarcinoma-derived



tumor cell line)⁷; (iii) the lack of multiple cell types, which as multicellular *in vitro* systems would better represent the tissue reaction to certain exposures⁸; and (iv) the inability of the cell monolayer, normally covered with culture medium, to achieve organotypic differentiation, including epithelial polarity. Moreover, although submerged cells that have not undergone organotypic differentiation can be exposed directly to complex aerosols such as CS,^{9–12} the nonspecific stress and damage caused by removal of the culture medium and the lack of differentiated features that protect against airborne toxicants (eg, mucociliary or squamous differentiation) cannot be excluded. Therefore, exposure surrogates such as aqueous CS extracts¹³ or specific fractions diluted in solvents such as dimethyl sulfoxide are still used in most models.¹⁴

In recent years, progress has been made in the development of valuable *in vitro* systems that mimic human airway and lung epithelia. Organotypic pseudostratified bronchial epithelium-like tissue derived from primary human bronchial epithelial cells that closely resembles normal human tracheobronchial epithelium after differentiating into ciliated, non-ciliated, goblet, and basal cells is now available.^{15,16} Such tissue cultures have been used to investigate the effects of CS-related substances such as nicotine, formaldehyde, urethane, cadmium,¹⁷ and hydrogen peroxide.¹⁸ Applying these substances as liquid solutions to the air liquid interface (ALI) has identified a number of exposure-specific molecular defense mechanisms.^{17–19} We and others have previously exposed organotypic murine²⁰ and human bronchial epithelial tissue cultures^{21,22} to whole mainstream CS at the ALI. By analyzing different endpoints, including gene expression, the effects of CS were observed on many cellular processes, including xenobiotic metabolism, the oxidant/antioxidant balance, cell proliferation, and DNA damage and repair.

Recently, we exposed human organotypic bronchial epithelial tissue cultures (AIR-100 tissue) at the ALI to mainstream CS using the Vitrocell[®] system. Data generated in this *in vitro* experiment as well as *in vivo* published data were used to perform a comparative analysis of different CS exposure durations (corresponding to different doses) and post-exposure periods. We demonstrated that this *in vitro* system is a reliable model of the bronchial epithelium of human smokers.²¹ In a second analysis of these *in vitro* data, presented herein, we applied a broad systems biology approach, including a novel computational modeling method, to more extensively analyze the gene expression data, and generated additional microRNA (miRNA) profile to further understand the exposure effect. For an enhanced tissue context-specific interpretation of the gene expression changes, we complemented gene set enrichment analysis (GSEA)²³ with a novel reverse causal reasoning (RCR)-based approach,²⁴ including several selected computable causal network models: (i) the cellular stress network model,²⁵ (ii) the cell proliferation network model,²⁶ and (iii) the inflammatory process network model.²⁷

The RCR method²⁴ integrates large-scale transcriptomics data into a biological context using information from differentially expressed genes together with a priori knowledge of cause and effect relationships to identify high-level upstream controllers of the observed gene expression called hypotheses (HYPs) (for details, see Refs. 25–27). HYPs were mapped as functional nodes on specific cellular networks, thereby enabling the identification of molecular perturbations of specific cellular signal pathways within the networks in response to CS exposure.

Another side of this study was to generate from the same samples both gene expression and miRNA profiles in order to integrate them into our systems biology investigation of the effect of CS. A stringent approach was applied based on the high-confidence miRNA targets provided by TarBase.²⁸ Various small regulatory networks were then created where possible around the significantly expressed miRNAs to reveal details of the downregulation of nuclear factor NF- κ B signaling and the activation of cell cycle activation.

Materials and Methods

Cell culture. EpiAirway[™] tissue cultures (AIR-100), purchased from MatTek Corporation, were produced from primary bronchial epithelial cells obtained from a 23-year-old male donor with no history of smoking. Tissue inserts were first cultured for 24 ± 2 hours at 37°C after shipment prior to exposure. The different exposure conditions did not trigger a decrease in cell viability as assessed 24 hours after exposure using the soluble tetrazolium salt resazurin (Sigma) (data not shown).

CS exposure and experimental design. The reference cigarettes 3R4F were obtained from the University of Kentucky (www.ca.uky.edu/refcig) and stored in their original packaging in a cooling chamber at $5 \pm 3^\circ\text{C}$ with uncontrolled humidity conditions before conditioning. Conditioning of the sticks was performed at $22 \pm 2^\circ\text{C}$ with $60 \pm 5\%$ humidity for 48 hours prior to the experiment. AIR-100 cultures were exposed in triplicate (three inserts per exposure condition) for 7, 14, 21, and 28 minutes at the ALI to either synthetic air (SA; 85% nitrogen and 15% oxygen; Praxair) or to 15% (vol/vol) mainstream CS from the 3R4F reference cigarette in the Vitrocell[®] system.²⁹ For all conditions, 1 puff/minute, 2 mL/minute of vacuum rate, 0.395 L/minute of dilution airflow, and 35 mL/puff were used. The smoking regimen was in basic conformity with the International Organization for Standardization (ISO 2000) smoking regimen using the VC 10 smoking robot. After exposure, tissues were incubated in fresh culture medium for 0.5, 2, 4, 24, and 48 hours before further analysis. Table 1 summarizes the various end points, time points, and dose levels.

Array-based transcriptomics. RNA was extracted and hybridized to high-throughput GeneChip[®] Human Genome U133 Array Plates (Affymetrix). For miRNA analysis, total RNA was labeled using the miRCURY LNA[™] miRNA Array

Power Labeling Kit (Exiqon). The Hy3-labeled samples and a Hy5-labeled reference RNA sample were hybridized to the fifth generation of miRCURY LNA Array version (Exiqon). Array data were submitted to ArrayExpress with the following accession numbers: E-MTAB-874 for the mRNA dataset and E-MTAB-877 for the miRNA dataset.

Gene set enrichment analysis. To investigate the main biological effects regulated by CS in AIR-100 tissue cultures, GSEA was performed with entire contrasts computed for the treatment effect independently of the dose.²³ For each post-exposure time, the list of genes was sorted in descending order by the t -values associated with the β_i coefficient estimated from the following linear model:

$$\text{Expression} \sim \beta_0 + \beta_t^* \text{Treatment} + \beta_d^* \text{Dose} + \beta_{td}^* \text{Treatment}^* \text{Dose} + \varepsilon$$

The coefficients of the model reflected the basal gene expression (β_0 , intercept), the whole CS effect (treatment: CS versus sham) independent of the dose (β_t), the dose-effect (7, 14, 21, and 28 minutes) independent of the treatment (β_d), and the interaction effect between the treatment and the dose (β_{td}). The parameter ε corresponds to the residual error. Normalized enrichment scores (NESs) and the associated false discovery rate (FDR) were then computed for each gene set of the molecular signatures database (MSigDB) C2 collection representative of canonical pathways, biological processes, and chemical and genetic perturbations (<http://www.broadinstitute.org/gsea/msigdb/index.jsp>). These were used as prior knowledge for the analysis. Gene sets with significant NESs (NESs with $\text{FDR} \leq 0.05$) for at least one post-exposure time and similar functions were grouped together to represent specific biological processes perturbed by CS.

Reverse causal reasoning. RCR²⁴ is a knowledge-driven approach used to identify high-level upstream controllers, HYPs, which could explain the observed differential gene expression profiles. A priori biological knowledge in the form of possible HYPs and experimental data in the form of significantly regulated genes selected from each treatment-control pairwise contrast gene expression profile (different doses and post-exposure times) were used as inputs for RCR analysis. These pairwise comparisons were computed using limma³⁰

following a global linear model to produce a contrast for each treatment-control pair (CS for a given exposure time/dose and post-exposure time were compared with the corresponding sham condition). Genes that were significantly differentially regulated by CS were selected by an absolute fold change ≥ 1.3 , a significance level (FDR) ≤ 0.05 , and an abundance level $\geq \log_2(100)$.

An HYP corresponds to a simple causal network constituted by an upstream node (subject term or controller) connected to downstream nodes (measured quantities, eg, expression of target genes) by a causal edge. Each edge represents the causal relationship between the controller and the target gene derived from an evidence line extracted from the literature (eg, transcriptional activation of x increases the expression of y , but decreases the expression of w), and each causal statement is coded in the biological expression language (BEL) (<http://www.openbel.org>) (for details, see Ref. 24).

The notion of causal relationship directionality is important in RCR. Indeed, HYPs can be evaluated as potential explanations for the observed differences in measurement sets by calculating the concordance statistic, which measures the consistency of the directions between the observed state changes (corresponding to the pairwise differentially expressed genes) and the expected direction. A P -value was computed for each concordance statistic. RCR analysis of our data was performed on the Gene Technology Platform (Selventa), currently holding more than 2,000 HYPs.²⁴

Network models. The following causal biological network models were applied: cellular stress,²⁵ cell proliferation,^{26,31} and inflammatory process networks.²⁷ Only subnetworks related to lung epithelial cells were considered for the latter model (ie, epithelial barrier defense and epithelial cell proinflammatory signaling); immune-cell-specific subnetworks (eg, neutrophil response) were excluded. Network models and their underlying subnetworks consisted of edges and nodes coded in BEL. Nodes are biological entities such as protein abundance, mRNA expression, and protein activity or can represent biological processes such as response to oxidative stress (ROS). Some nodes are also called HYPs when they can be predicted to increase or decrease by the RCR based on the transcriptomics input to the network. All causal connections or edges were based on manually curated

Table 1. Exposure conditions and biological endpoints. \checkmark shows that all doses and post-exposure time points were measured for the endpoint.

TEST SUBSTANCE	SHAM				CIGARETTE SMOKE (15%)			
Exposure time (min)	7	14	21	28	7	14	21	28
Post-exposure time (hours)	0.5,2,4,24,48							
Viability (test at post exposure time)	4,24	4,24,48	4,24	4,24,48	4,24	4,24,48	4,24	4,24,48
Gene expression (all post-exposure time)	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
miRNAs (all post-exposure time)	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
Reverse phase protein array (test at post-exposure time)	0.5,2,24,48							
Immunostaining/histology (test at post-exposure time)	24	24,48	24	24,48	24	24,48	24	24,48



literature (for details of network model building, see Refs. 25–27).

HYP clustering. To group HYPs that behave similarly with regard to the concordance statistic across doses and post-exposure times, the affinity propagation (AP) clustering algorithm in the form of an R implementation was used (APcluster^{32,33}). Briefly, following RCR, HYPs belonging to one of the subnetworks investigated with at least one significant concordance ($P < 0.05$) across the 14 and 28 minutes exposure time and all post-exposure time contrasts were selected for the clustering process. The Pearson correlation coefficient of the minus log-transformed concordance P -values was computed as a measure of similarity for each pair of HYPs. A pairwise correlation matrix was then taken as input by APcluster using default parameters to generate the clustering. As a result, HYPs were grouped into communities and further leveraged for biological interpretation of data.

Strategy for the integrated mRNA–miRNA analysis. Integrated mRNA–miRNA analyses are usually more sensitive than specific. They produce many candidate miRNA gene interactions, a significant fraction of which are false positive because of high signal-to-noise ratios in the statistical analyses. Here, we adopted a different approach by prioritizing specificity over sensitivity. This enabled our results to be interpreted on the basis of causations rather than correlations or statistical associations, which was in line with the approach taken for the analysis of mRNA expression data. We therefore made the following choices: (i) we used a less-detailed linear model to determine mRNA and miRNA differential expression (see below); (ii) differential expressions of significant miRNAs were compared with the corresponding values of their family members (ie, miRNAs from alternative strand, co-transcribed miRNAs, and paralogous miRNAs) to provide additional confidence in the obtained results; (iii) we used preferential mRNA–miRNA relationships supported by the most reliable experimental methods, as described in the TarBase repository²⁸; and (iv) we restricted the final interpretation to the 24 and 48 hours post-exposure time points, in which the effects of differential miRNA regulation could be reasonably expected to be detected in the differential expression of their mRNA targets (at least two hours were necessary to observe a significant decrease in target transcripts in the most favorable case).³⁴

Calculation of differential expression for integrated mRNA–miRNA analysis. Exiqon-based miRNA raw expression data were preprocessed as described in our previous study.²¹ Because the miRNA data yielded a weaker signal than mRNA data, we replaced the linear model described above by a more robust two-step approach, which we applied to both miRNA and mRNA normalized expression data. First, CS-induced differential expression values were calculated by subtracting the mean of the post-exposure time- and dose-matched control group from the expression value of each sample. Second, for each post-exposure time point (0.5, 2,

4, 24, and 48 hours), the obtained CS-induced differential expression matrix was applied to the following linear model:

$$\text{CS-induced differential expression} \sim \beta_g \cdot \text{Treatment} + \beta_{dg} \cdot \text{Treatment} \cdot \text{Normalized_Dose} + \varepsilon$$

The variables in the model are Treatment $\in \{0,1\}$, which corresponds to sham and treated samples, respectively, and Normalized_Dose $\in \{-1, -1/3, 1/3, 1\}$, which results from centering and normalizing the dose values $\{7, 14, 21, 28\}$. In this model, β_g captures the global CS-induced effect valid for all four doses, while β_{dg} captures the systematic dose-specific deviations around the global CS-induced effect β_g . The coefficients of the linear model are statistically significant when the corresponding FDR values are smaller than 0.05, as calculated using moderated t -statistics and multiple-testing corrections.²¹

Small network assembly for integrated mRNA–miRNA analysis. We assembled small regulatory networks around the significantly differentially expressed miRNAs where possible. We used TarBase as the source of high-confidence miRNA targets,²⁸ and took advantage of the recently established property that mRNA–miRNA interactions are weakly dependent upon the tissue context.³⁵ mRNA differential expression values were assigned to the gene nodes in the networks. This choice was supported by two observations. First, even if miRNAs operate functionally at the protein level, their effects at the mRNA levels are highly representative.³⁶ Second, the identification of mRNA and protein differential expression as a consequence of miRNA changes is expected to be more relevant at the 24 and 48 hours post-exposure time points considered in our integrated mRNA–miRNA analysis than at the earlier time points (see above).

Results and Discussion

To investigate dose-effects over different post-exposure times, the organotypic bronchial epithelial tissue cultures were exposed for 7, 14, 21, or 28 minutes to 15% (vol/vol dilution with air) mainstream CS (corresponding to different doses) with five post-exposure times (0.5, 2, 4, 24, and 48 hours) for each of the doses (Table 1). The systems biology approach applied for the mechanistic investigation of CS effects was based on mRNA and miRNA expression data, and on computational data analysis and interpretation using an integrated mRNA–miRNA analysis. This included small network assembly and the two complementary approaches for gene expression: (i) GSEA and (ii) RCR networks and AP-clustering.

GSEA. To identify the main general biological processes perturbed in AIR-100 tissue culture after CS exposure from the data set of differentially expressed genes, GSEA was performed with entire contrasts (computed for the treatment effect, independent of the dose) using a collection of gene sets representative of canonical pathways (from the MSigDB collection).^{23,37} Briefly, this commonly used approach enabled the



identification of a significant enrichment of coordinated co-regulated genes representative of biological processes (within full gene expression profiles). Focusing on the CS effect, independent of the exposure dose, enabled us to elucidate processes that were regulated over post-exposure times, as shown in Figure 1. This revealed a significant enrichment of genes involved in the oxidative stress response, which was rapidly induced at early post-exposure times (2 and 4 hours). Part of the stress response was repressed at later post-exposure time points (24 and 48 hours), whereas a sustained enrichment of genes involved in processes characteristic of detoxification (eg, metabolism of xenobiotics – metabolism by cytochrome P450) and antioxidant systems (eg, glutathione metabolism) was seen. Genes involved in the machinery for mRNA processing and translation were rapidly upregulated and coordinately induced after CS exposure. Interestingly, gene sets associated with processes leading to mRNA translation and newly synthesized proteins were sequentially enriched over post-exposure times (eg, mRNA capping, elongation, splicing, translation, and protein export). At later post-exposure times, GSEA revealed a consistent enrichment of genes encoding proteins of the mitochondrial electron respiratory chain (at 24 and 48 hours) as well as of energy metabolism (eg, amino acids, pyruvate, and lipids). Strikingly, gene sets representative of interferon and NF- κ B signaling pathways were significantly enriched in the downregulated gene profiles associated with early (2 and 4 hours) and late (24 and 48 hours) post-exposure times. The analysis also captured a late, yet coordinated, induction of genes involved in cell proliferation at 24 hours post-exposure. This observation could be linked to the induction of energy metabolism highlighted after 24 and 48 hours post-exposure.

RCR analysis. To follow up on the GSEA results and obtain a more granular, tissue-specific mechanistic understanding of the molecular effect of CS in a dose- and time-dependent manner, we applied a recently developed method that uses computable network models in combination with the RCR approach.²⁴ This method is based on tissue-specific biological causal network models built to represent different biological processes, such as cell proliferation,²⁶ cellular stress,²⁵ and inflammatory processes,²⁷ in the context of the lung. RCR integrates large-scale transcriptomic data into biological meaning by identifying higher level upstream controllers of specific gene expression. These HYPs can be mapped as functional nodes on specific cellular networks, enabling comparisons of molecular perturbations with signaling pathways and other response mechanisms across diverse doses and post-exposure time points both qualitatively and quantitatively.^{38,39}

RCR analysis was performed using differential gene expression (CS exposure vs. sham exposure) derived from AIR-100 tissue cultures exposed to whole smoke for 14 minutes (low dose) or 28 minutes (high dose) followed by five post-exposure times (0.5, 2, 4, 24, and 48 hours). The 7- and 21-minute exposure times were not considered for RCR because they were found to have insufficient differentially

expressed genes to perform RCR (Supplementary Table 1A). By applying a distinct significance criterion (ie, concordance <0.05), 671 HYPs were identified that showed significant changes in at least one exposure condition (Supplementary Table 1B). Of these, 244 (37%) could be mapped to one or several of the following three different network models: 126 were related to the cellular stress response network model, 54 to inflammatory process subnetworks related to the lung epithelium (ie, epithelial barrier defense and epithelial cell proinflammatory signaling), and 130 to the cell proliferation network model (Supplementary Table 1C).

To gain insights into the time- and dose-dependent activation or repression of HYPs and potential co-regulations, a clustering approach (APcluster) was applied to the minus log-transformed *P*-value concordance associated with the 244 identified HYPs across 10 different exposure conditions. Seventeen different communities of similarly regulated HYPs were identified. Figure 2 shows the clustering with the assignment of the single HYPs to the three different network models (cellular stress, cell proliferation, and inflammatory processes). There was no clear correlation of distinct communities with a single network, suggesting that different cellular processes show, at least in part, similar time- and dose-dependent regulations by CS exposure. The interconnection of the different networks is also obvious from the fact that certain HYPs, such as the transcriptional activity of the activating protein (AP)-1 complex (“taof(AP-1 complex Hs)”), belong to all three networks.

Two major effects are notable. First, an overall dose- and time-dependent response was detected. At the low dose (14 minutes of exposure), no significant HYPs were detected at 0.5 hours post-exposure and only three were detected at 48 hours post-exposure, whereas a stronger response was observed at 2, 4, and 24 hours post-exposure. Thus, at the low dose, an almost complete recovery of the cellular system occurred within 48 hours post-exposure. By contrast, at the high dose (28 minutes of exposure), a much stronger response was already observed 0.5 hours post-exposure (76 significant HYPs), and a sustained perturbation of the system was reflected by a large number of significant HYPs at 2, 4, 24, and even 48 hours post-exposure. A qualitatively similar response was observed at 2 and 4 hours post-exposure for both doses with a clear trend toward a stronger response with the high dose. Second, cluster analysis revealed a strong similarity between 24 hours post-exposure at the low dose and 48 hours post-exposure at the high dose. This suggests that similar cellular processes are active at both doses and that the differences in dose response mainly reflect the kinetics of the cellular response.

Cellular stress response. Most HYPs related to the cellular stress network showed a transient activation with recovery after 24 hours post-exposure for the low dose and after 48 hours post-exposure for the high dose (Fig. 3A). Detailed investigations of subnetworks and related HYPs revealed a strong activation of



Figure 1. GSEA revealed specific patterns of biological processes modulated over post-exposure times in AIR-100 tissue culture exposed to whole CS. Gene sets with significant NESs for at least one post-exposure time (2, 4, 24, or 48 hours) and known biological functions were grouped together. A heatmap was generated to highlight NES patterns over post-exposure time. Red color indicates the enrichment of gene sets at the top of the ranked gene list (upregulated genes), and green color indicates enrichment of gene sets at the bottom of the ranked gene list (downregulated genes). The color scale is proportional to NES values. Values in bold correspond to NES with associated FDR ≤ 0.05 .

the oxidative stress response that was also identified by GSEA analysis. This activation could be deduced from the prediction of HYPs that are directly related to corresponding biological processes, such as “ROS” and “oxidized low-density lipoprotein”, and HYPs that are related to substances known to induce oxidative stress, such as “acrolein”, “hypochlorous acid”, and

“ozone”. Interestingly, the HYP for acrolein, a CS constituent, was downregulated at 24 hours post-exposure for the low dose and at 48 hours post-exposure for the high dose, indicating a recovery from oxidative stress in the system.

Direct analysis of the oxidative stress subnetwork model through mapping the identified HYPs to the respective nodes

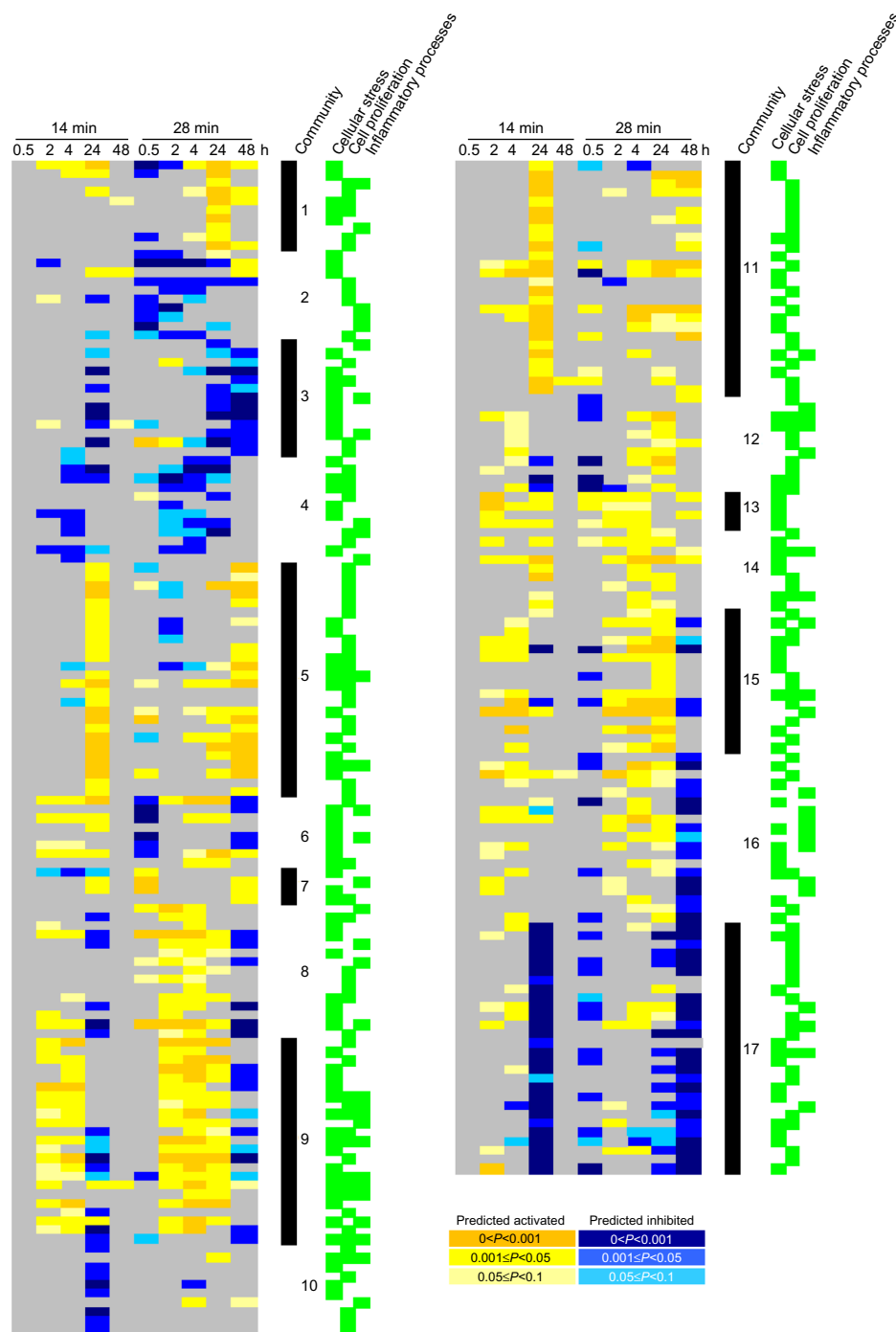


Figure 2. AP clusters using the minus log-transformed P -value of concordance from 244 HYPs with at least one significant concordance ($P < 0.05$) across the two different exposure conditions (14 and 28 minutes) and the five different post-exposure time points (0.5, 2, 4, 24, and 48 hours). All of these HYPs (indicated by a green box) belong to at least one of the three networks (cellular stress, cell proliferation, and inflammatory processes). In all, 17 different communities of similarly regulated HYPs are shown. Color code: yellow, HYPs with predicted increase in abundance or activity; blue, HYPs with predicted decrease in abundance or activity; light yellow/blue, $0.05 \leq P < 0.1$; yellow/blue, $0.001 \leq P < 0.05$; dark yellow/blue, $P < 0.001$; and gray, no significant changes. For more details, see Supplementary Table 1B.

in the network (Fig. 3B and Supplementary Fig. 1A for a higher resolution view) elucidated potential cellular pathways involved in the oxidative stress response. For instance, nuclear factor (erythroid-derived 2)-like 2 (NFE2L2) is the central transcription factor that mediates the oxidative stress response both *in vitro* and *in vivo* when induced by CS.^{13,40,41} As shown

in Figure 3A, NFE2L2 abundance (“NFE2L2”) and activity (“taof(NFE2L2)”) HYPs were predicted to be activated in the NFE2L2_signaling subnetwork, whereas HYPs for “KEAP1” and “Bach1”, which are proteins known to counteract NFE2L2 activity,^{7,42} were predicted to be inhibited (for details, see specific NFE2L2_signaling subnetwork, Supplementary

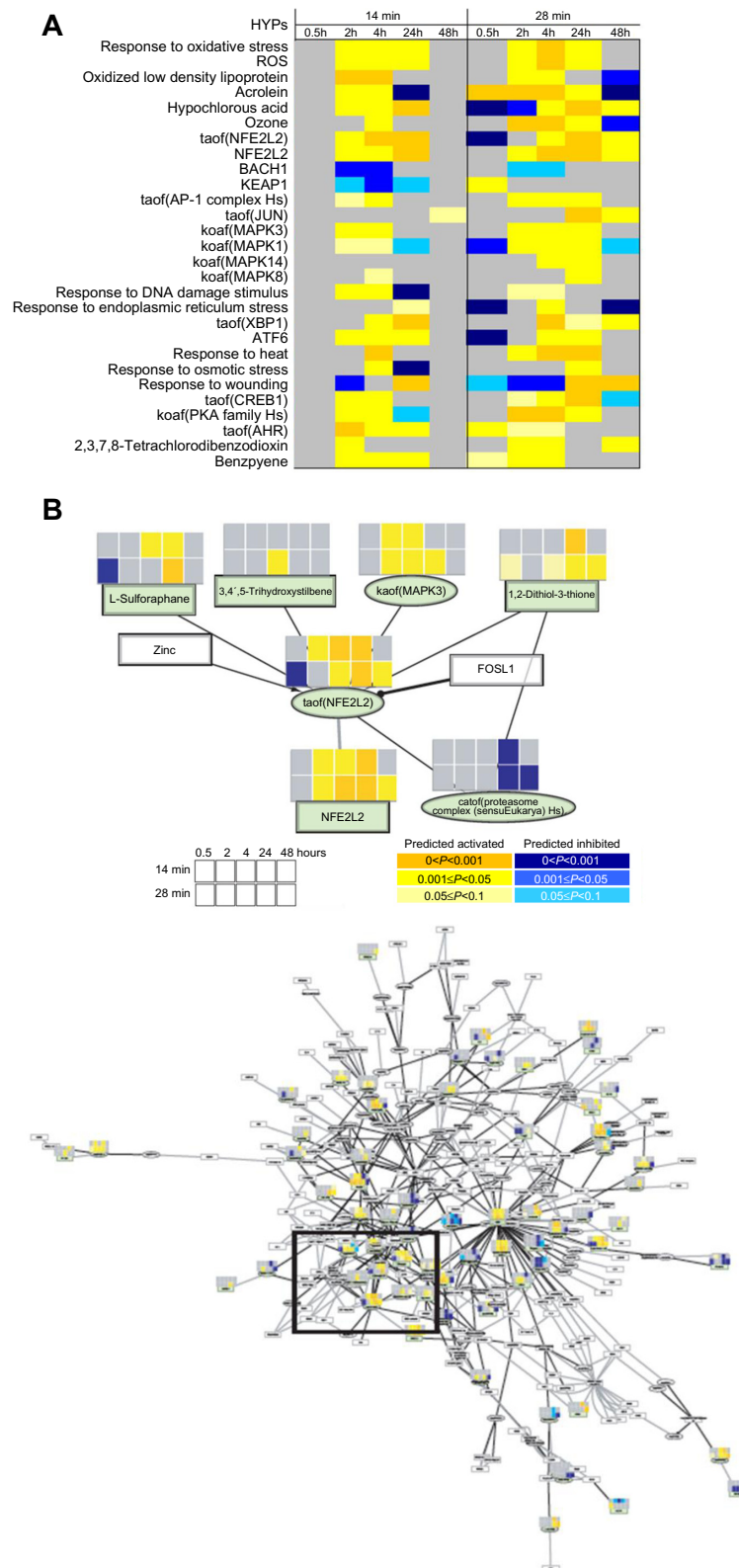


Figure 3. (A) Selection of predicted HYPs within the cellular stress network. ROS, reactive oxygen species; taof(X), transcriptional activity of X; and koaf(X), kinase activity of X. For detailed information, see Supplementary Table 1B. **(B)** High magnification of a schematic representation of perturbations observed in part of the oxidative stress subnetwork. On the upper part of each significant HYP, an eight-folder cassette (shown in light green) depicts the color code for significance of concordance for all post-exposure kinetics (0.5, 2, 4, 24, and 48 hours) after 14 minutes of exposure (top five squares) and after 28 minutes of exposure (bottom five squares). Circle head arrow, inhibition; arrow edge, activation; and gray line, co-regulation. For a high-resolution image of the subnetwork, see Supplementary Figure 1. Color code: yellow, HYPs with predicted increase in abundance or activity; blue, HYPs with predicted decrease in abundance or activity; light yellow/blue, $0.05 \leq P < 0.1$; yellow/blue, $0.001 \leq P < 0.05$; dark yellow/blue, $P < 0.001$; and gray, no significant changes.

Fig. 1B). Additionally, activity of AP-1, another transcription factor known to be activated by oxidative stress,^{41,43} was identified by the activation of related HYPs such as “taof(Jun)”, “taof(AP-1 complex Hs)”, and several mitogen-activated protein kinases (MAPKs), eg, such as extracellular signal-regulated kinase (ERK)1 (“kaof(MAPK3)”) and ERK2 (“kaof(MAPK1)”), which are upstream of AP-1 in cellular signal transduction cascades induced by oxidative stress. These results are in accordance with the activation of pro-proliferative ERK1/2 signaling, and pro-apoptotic MAP kinase p38 and c-Jun N-terminal kinase previously detected by phospho-specific Western blotting in CS condensate-treated submerged bronchial epithelial cells.⁴⁴

Other cellular stress responses were identified from the prediction of related HYPs by RCR. Increased activity of HYPs related to protein damage, such as “response to heat”, which covers the induction of specific heat shock genes; “response to endoplasmic reticulum stress”, including activation of related transcription factors activating transcription factor (“ATF6”) and X-box binding protein 1 (“taof(XBP1)”); and “response to osmotic stress”, which could be observed as early as 2–4 hours after exposure.

Another cellular stress response known to be induced by CS and also identified by GSEA analysis is related to xenobiotic metabolism.⁴⁵ The prediction of HYPs for “taof(AHR)”, “2,3,7,8-tetrachlorodibenzodioxin”, or “benzopyrene” pointed to the activation of the aryl hydrocarbon receptor, the key transcription factor that mediates the induction of phase I enzymes, such as Cyp1A1.⁴⁶ A more detailed description of the evaluation of xenobiotic metabolism subnetwork has been provided elsewhere.⁴⁷ In general, the cellular stress response followed a dose-dependent, transient kinetic response with a stronger and more persistent activation at the high dose; however, on a qualitative level, the responses for both doses were very similar in terms of perturbed biological processes (eg, activating the NFE2L2 pathway to combat oxidative stress, Fig. 3A). Overall, the results indicate a comprehensive, highly regulated cellular stress response to CS exposure, which involves different cellular defense pathways.

Inflammatory processes. In human lung tissue, epithelial cells mediate an initial inflammatory response by actively secreting specific cytokines that recruit and activate inflammatory cells, such as neutrophils or macrophages,⁴⁸ or by the release of damage-associated molecular pattern molecules from necrotic cells.^{49,50} The analysis of significant HYPs that are part of the three subnetworks related to inflammatory response in epithelial tissue revealed distinct perturbations of the main intracellular signal mediator pathways (Fig. 4). In accordance with the GSEA results, inhibition of the interferon (IFN) γ - and NF- κ B-mediated inflammatory responses was visible as the downregulation of related HYPs such as “IFNG” and “taof(RELA)” (Fig. 4). The observed transient inhibition of the NF- κ B/IFN γ response has previously been described in response to aqueous extracts of CS,⁵¹ oxidants in general,⁵² and

acrolein,⁵³ which is a major α,β -unsaturated, highly reactive aldehyde present in CS; this, therefore, links the process of NF- κ B inhibition to the oxidative stress response.

We also observed inhibition of the activity of the transcription factors CCAAT/enhancer-binding protein beta (“taof(CEBPB)”) and signal transducer and activator of transcription 3 (“taof(STAT3)”), which are known to be involved in the inflammatory response. Interestingly, a clear dose-dependent proinflammatory response related to increased cytokine abundance was also predicted (eg, interleukin (IL), “IL1A”, “IL1B”, “IL6”, “IL13”, and “IL24”). Again, this response was mostly reversed at 48 hours post-exposure for the low dose but remained elevated over time for the high dose with respect to several cytokines such as IL2, IL6, and IL13. Allahverdian et al previously reported that IL13 secretion by airway epithelial cells enhanced epithelial repair via heparin binding EGF-like growth factor (HB-EGF),⁵⁴ and potential NF- κ B-independent inflammatory pathways have been suggested to operate through the transcription factors AP-1⁵⁵ and/or early growth response protein-1 (EGR-1)⁵⁶; the activity of both of these transcription factors was predicted in our networks.

Proliferation-related response. In mapping significant HYPs to the proliferation network, we observed a distinct perturbation of the core proliferation subnetwork (ie, cell cycle subnetwork; Fig. 5A and Supplementary Fig. 1C for a higher resolution view). In particular, a coordinated downregulation of HYPs related to cell cycle inhibitors and activation of HYPs for cell cycle inducers was predicted by RCR at 24 hours post-exposure time for the low dose (Fig. 5B). This response fully recovered at 48 hours post-exposure. HYPs describing proteins and/or biological processes that are downregulated in proliferating cells such as “taof(p53)”, “RB1”, “taof(E2F4)”, and “CDKN1A” were clustered in community 17, whereas those describing increased protein abundance or upregulated transcriptional activation involved in cell proliferation, eg, “E2F1”, “E2F2”, “E2F3”, “CDK4”, “CCND1”, and “taof(Myc)”, were mainly clustered in community 11 (Figs. 2 and 5B, and Supplementary Table 1C).

Because CS is known to induce growth factor-related cellular pathways,⁵⁷ we examined a specific subnetwork of the cell proliferation network related to growth factor response more closely (Fig. 6 and Supplementary Fig. 1D for a higher resolution view). Nodes related to the EGF receptor (EGFR) family (eg, “NRG1”, “EGFR:ERBB3”, and “ERBB3”), and related ligands for EGFR activation (eg, “EGF” and “AREG”) were markedly upregulated. Moreover, EGFR ligands (HBEGF, EREG, and AREG) were strongly and dose-dependently induced at the gene expression level by CS exposure (Supplementary Table 1D). Accordingly, HYPs related to downstream cellular pathways, such as RAS (“HRAS mutated at G12V”), phosphatidylinositol 3-kinase (“kaof(PI3K Family Hs)”), and MAPK (“kaof(MAPK3)”) and “kaof(MAPK1)”), as well as downstream transcription factors (“taof(ELK1)”) and “taof(EGR1)”) were also upregulated (Fig. 5B). Additionally,

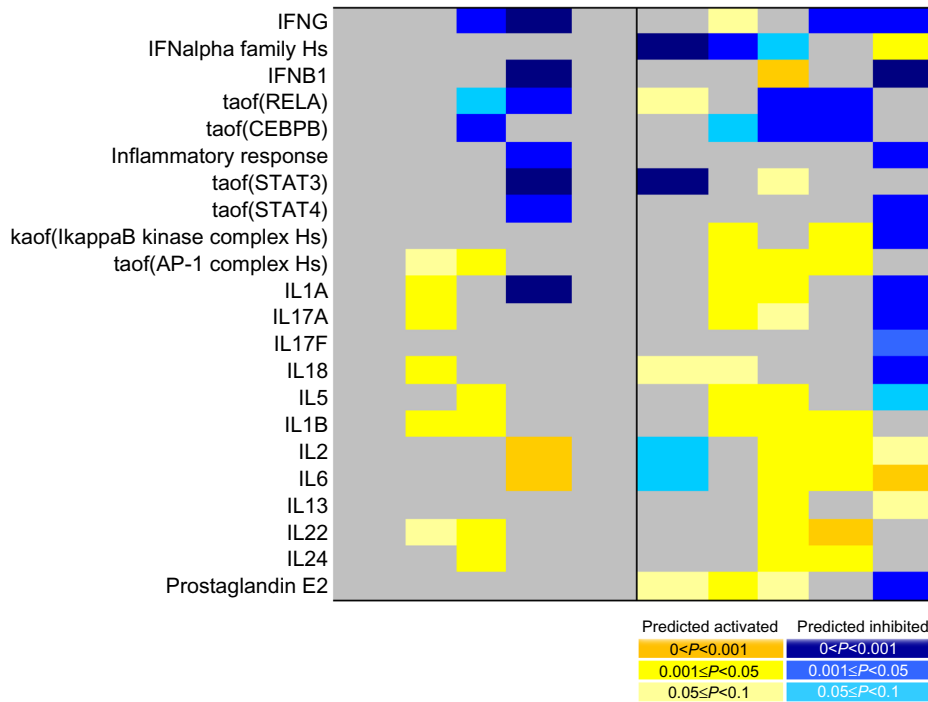


Figure 4. Selection of predicted HYPs within the inflammatory process subnetworks. Color code: yellow, HYPs with predicted increase in abundance or activity; blue, HYPs with predicted decrease in abundance or activity; light yellow/blue, $0.05 \leq P < 0.1$; yellow/blue, $0.001 \leq P < 0.05$; dark yellow/blue, $P < 0.001$; and gray, no significant changes. taof(X), transcriptional activity of X and kaof(X), kinase activity of X. For detailed information, see Supplementary Table 1B.

we observed HYPs for increased hepatocyte growth factor (“HGF”) abundance and increased activity of the related receptor (“kaof(MET)”), pointing to an alternative pathway that may lead to lung cell proliferation. Indeed, Sakamaki et al.⁵⁸ previously reported increased Hgf and Met expression preceding the regenerative proliferation of lung epithelial cells in mice after pneumonectomy.

A pathway related to fibroblast growth factors (“FGF2”, “FGF6”, and “FGF7”) was also identified, pointing to the involvement of these growth factors in the regeneration of AIR-100 tissue.⁵⁹ Additionally, HYPs for mediators of cell migration and tissue differentiation were strongly upregulated (eg, “gtpof(RHOC)”, “IHH”, and “taof(GLI2)”). For the high dose, a similar response was observed at 24 hours post-exposure. In contrast to the low dose, perturbation of the proliferation network persisted for at least 48 hours.

CS effect on the regulation of mucus synthesis in organotypic bronchial tissue cultures. Mucins are major constituents of airway mucus and, therefore, essential for the defense against and clearance of inhaled irritants and toxicants. Excessive mucin overproduction is a distinguishing feature and a risk factor of chronic obstructive pulmonary disease.^{60,61} Mucins have been reported to be expressed in airway epithelial cell cultures,^{62,63} and CS enhances mucin secretion both *in vitro*^{64,65} and in rat airways. Therefore, we investigated the expression of mucin genes in CS-exposed bronchial epithelial tissue cultures. Six membrane-bound

(MUC1, MUC4, MUC13, MUC15, MUC16, and MUC20) and two gel-forming mucins (MUC5AC and MUC5B) were differentially regulated in CS-exposed AIR-100 tissue cultures (Supplementary Table 1D). Most of these were downregulated in CS-exposed tissues compared with sham. For instance, MUC5B mRNA, specifically expressed in goblet cells,⁶³ was only downregulated at 24 hours post-exposure with the longer exposure time, thus corresponding to the observed initial decrease in the proportion of mature goblet cells (data not shown). MUC5AC mRNA, previously shown to be upregulated upon CS exposure both *in vitro* and *in vivo*,^{65–69} was increased at the later post-exposure time in both 14 and 28 minutes CS-exposed tissue cultures. Interestingly, we also observed the upregulation of the gene expression of different positive regulators of MUC5AC,⁷⁰ such as SPDEF and the EGFR ligands AREG, EREG, and HBEGF (Supplementary Table 1D). CS-enhanced MUC5AC expression was shown to be mediated by EGFR signaling and required ERK1/2 and p38 activation in rats,^{65,66,69} which agrees with the predicted ERK1/2 activation (Fig. 3A) in the present study.

CS-induced miRNA differential expression. We identified 11 miRNAs that were differentially expressed following CS exposure (more details are given in Supplementary Fig. 2). Except for hsa-miR-675, all miRNAs were downregulated, which is in agreement with *in vivo* findings⁷¹ and as discussed in our previous article.²¹ Regarding the sensitivity of our approach, the linear model used for the integrated

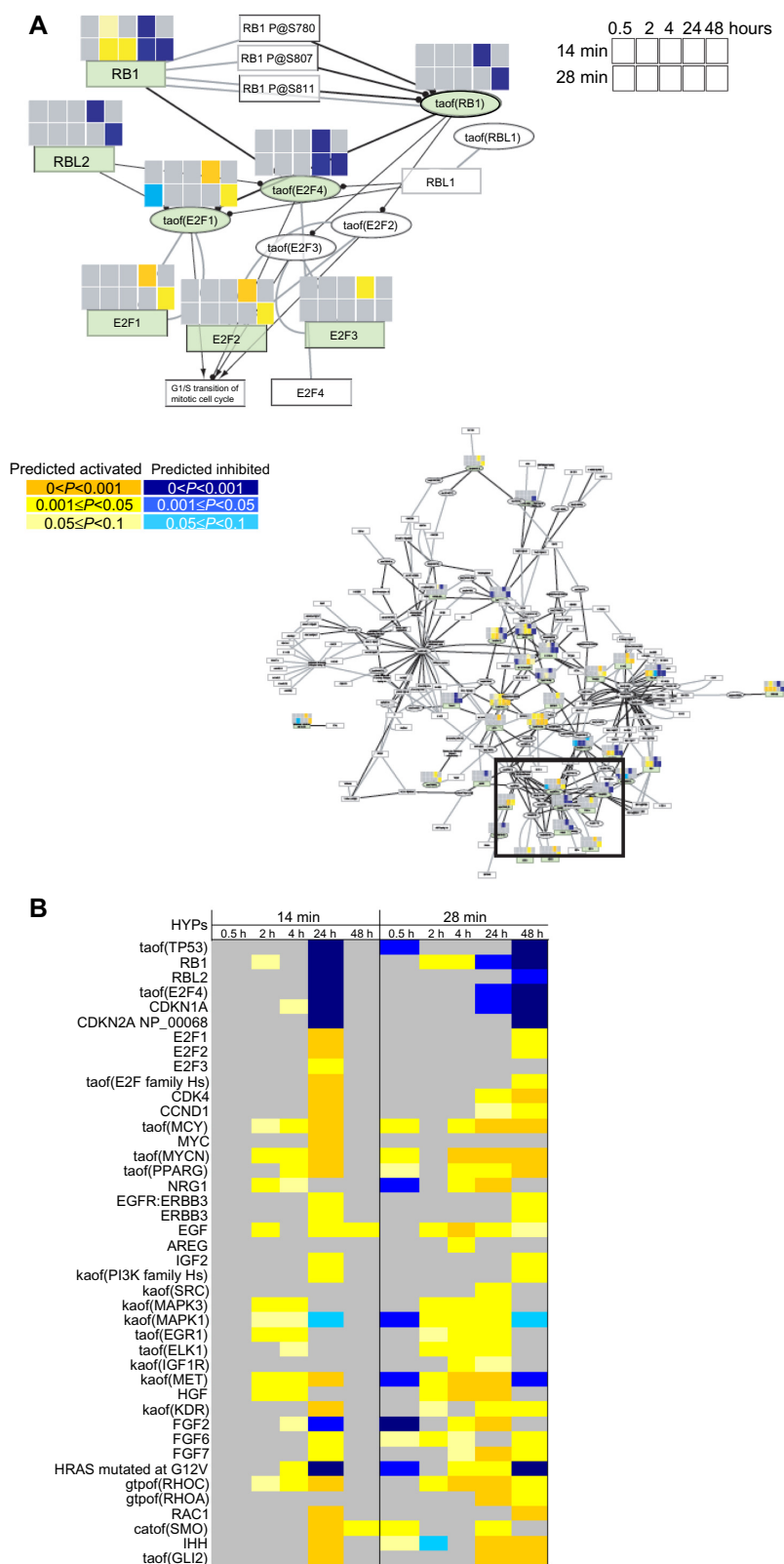


Figure 5. (A) High magnification of a schematic representation of perturbations observed in part of the cell cycle subnetwork. On the upper part of each significant HYP, an eight-folder cassette (shown in light green) depicts the color code for significance of concordance for all post-exposure kinetics (0.5, 2, 4, 24, and 48 hours) after 14 minutes of exposure (top five squares) and after 28 minutes of exposure (bottom five squares). Circle head arrow, inhibition; arrow edge, activation; and gray line, co-regulation. For high-resolution image of the subnetwork, see Supplementary Figure 1. **(B)** Selection of predicted HYPs within the cell proliferation network. taof(X), transcriptional activity of X; kaof(X), kinase activity of X; gtpof(X), GTP-bound activity of X; and catof(X), catalytic activity of X. For detailed information, see Supplementary Table 1B. Color code: yellow, HYPs with predicted increase in abundance or activity; blue, HYPs with predicted decrease in abundance or activity; light yellow/blue, $0.05 \leq P < 0.1$; yellow/blue, $0.001 \leq P < 0.05$; dark yellow/blue, $P < 0.001$; and gray, no significant changes.

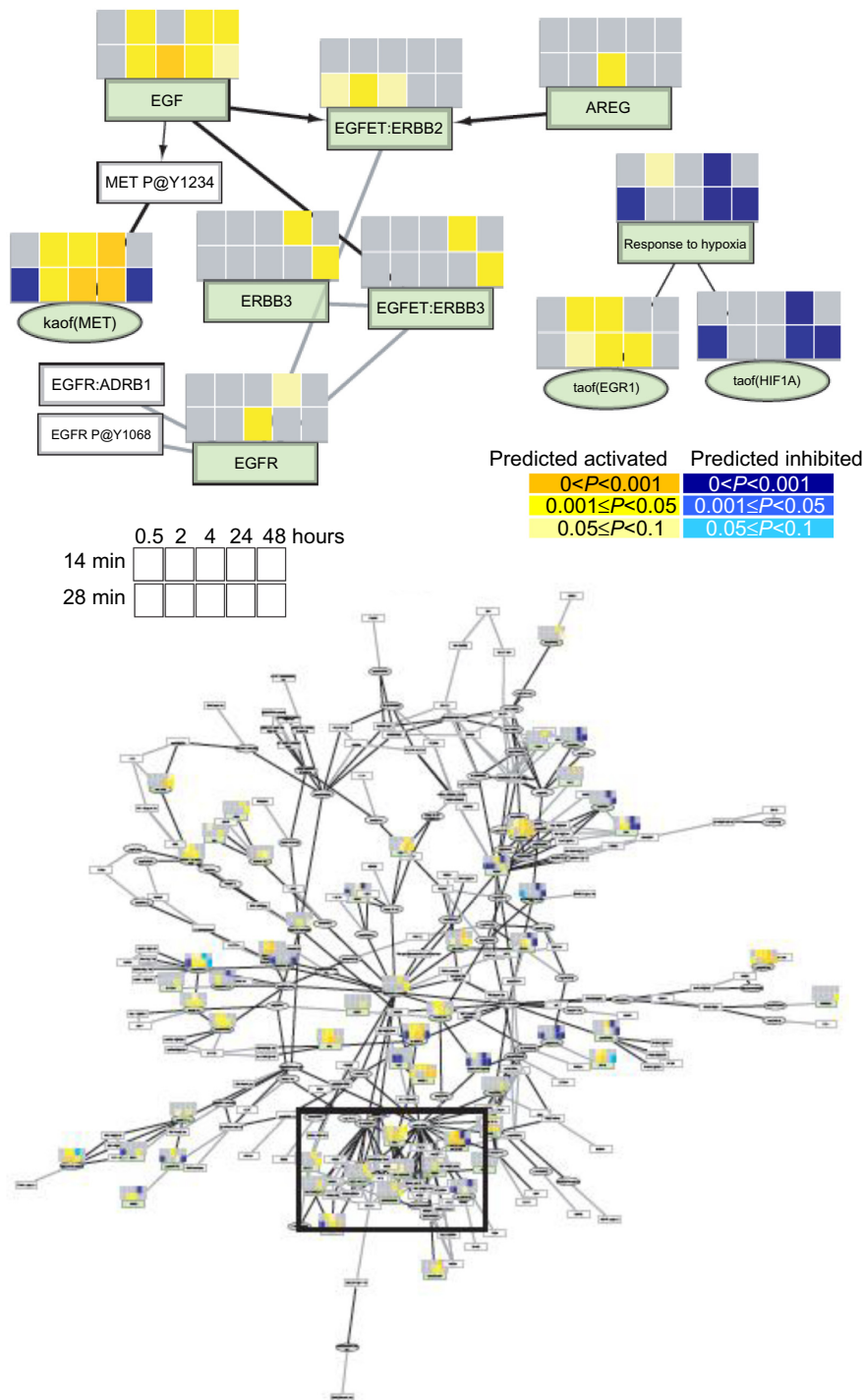


Figure 6. High magnification of a schematic representation of perturbations observed in part of the growth factor response subnetwork. On the upper part of each significant HYP, an eight-folder cassette (shown in light green) depicts the color code for significance of concordance for all post-exposure kinetics (0.5, 2, 4, 24, and 48 hours) after 14 minutes of exposure (top five squares) and after 28 minutes of exposure (bottom five squares). Color code: yellow, HYPs with predicted increase in abundance or activity; blue, HYPs with predicted decrease in abundance or activity; light yellow/blue, $0.05 \leq P < 0.1$; yellow/blue, $0.001 \leq P < 0.05$; dark yellow/blue, $P < 0.001$; and gray, no significant changes. Circle head arrow, inhibition; arrow edge, activation; and gray line, co-regulation. For high-resolution images of the subnetwork, see Supplementary Figure 1.

mRNA–miRNA analysis is not appropriate to capture the dose-dependent modulation of the response to CS exposure, as it is done in the mRNA-only part of the analysis using a more precise model. This does not mean that this effect is absent at the single miRNA level, but rather

that the three replicates per condition are not sufficient to stabilize the signal variance because of the inherently higher variability of the miRNA versus mRNA data. It is also worth mentioning that the dose-dependent modulation of the response to CS exposure was not observed at

the single gene, but at the gene-group level, eg, HYPs or gene sets.

Construction of small regulatory networks. Two small regulatory networks containing 7 of the 11 differentially expressed miRNAs were assembled. Where possible, experimentally high-confidence causal relationships were used to construct the regulatory networks following the same approach adopted during the assembly of the network models used in RCR analysis. Therefore, downstream miRNA targets were extracted from TarBase,²⁸ whereas upstream transcription miRNA factors were found in TransmiR.⁷² In both cases, a relevant structure with negative feedback loops was obtained, confirming the importance of the identified miRNAs in the regulatory mechanisms involved after CS exposure. These findings based on miRNA analysis from the same RNA samples used for gene expression analysis provide important complementary details about the regulatory mechanisms of biological processes, also identified as being perturbed by the mRNA-based evaluation.

The inhibition of NF- κ B found in our previous analysis with GSEA or RCR approaches was further supported by the expression of hsa-miR-146a and has-miR-146b, which are connected to NF- κ B in an auto-regulatory feedback loop.⁷³ The small network model (Fig. 7A) depicts NF- κ B as a transcription factor for both hsa-miR-146a and b, which both repress IL1 receptor-associated kinase (IRAK1), itself an activator of NF- κ B. As shown in Figure 7A, the differential expression of hsa-miR-146a/b is negative at 24 and 48 hours post-exposure, which matches the significant downregulation of the NF- κ B transcript. The positive differential expression of IRAK1 at the same time points is consistent with a weaker effect of hsa-miR-146a/b (de-repression). The connection closing the loop between IRAK1 and NF- κ B is not satisfied (at the transcript level), indicating that the feedback mechanisms bringing NF- κ B expression back to unperturbed levels may not yet be fully effective and that longer post-exposure times may be needed for complete reversibility of the NF- κ B response. It should also be noted that when testing the signed relationships in a negative feedback loop against experimental values, one edge will necessarily be unsatisfied. This does not invalidate the network but rather indicates which mechanisms are dominant at the time point of the measurement. Taken together, these findings demonstrate the importance of hsa-miR-146a/b in regulating inflammatory mechanisms.⁷⁴

Transient activation of the cell cycle was identified by RCR analysis of the proliferation network perturbation. In line with this observation, a second small network model depicting a possible feedback loop involving the downregulated E2F1-inhibiting miRNA hsa-miR-107/hsa-miR-449a/b/c (Fig. 7B and the heat map in Supplementary Fig. 2A) was generated. A similar negative feedback mechanism involving upregulated miR449a/b and CDC20B induced by E2F1 expression, and finally limiting E2F1 levels and excessive E2F1-driven proliferation was previously proposed by Lize et al.⁷⁵ The apparently

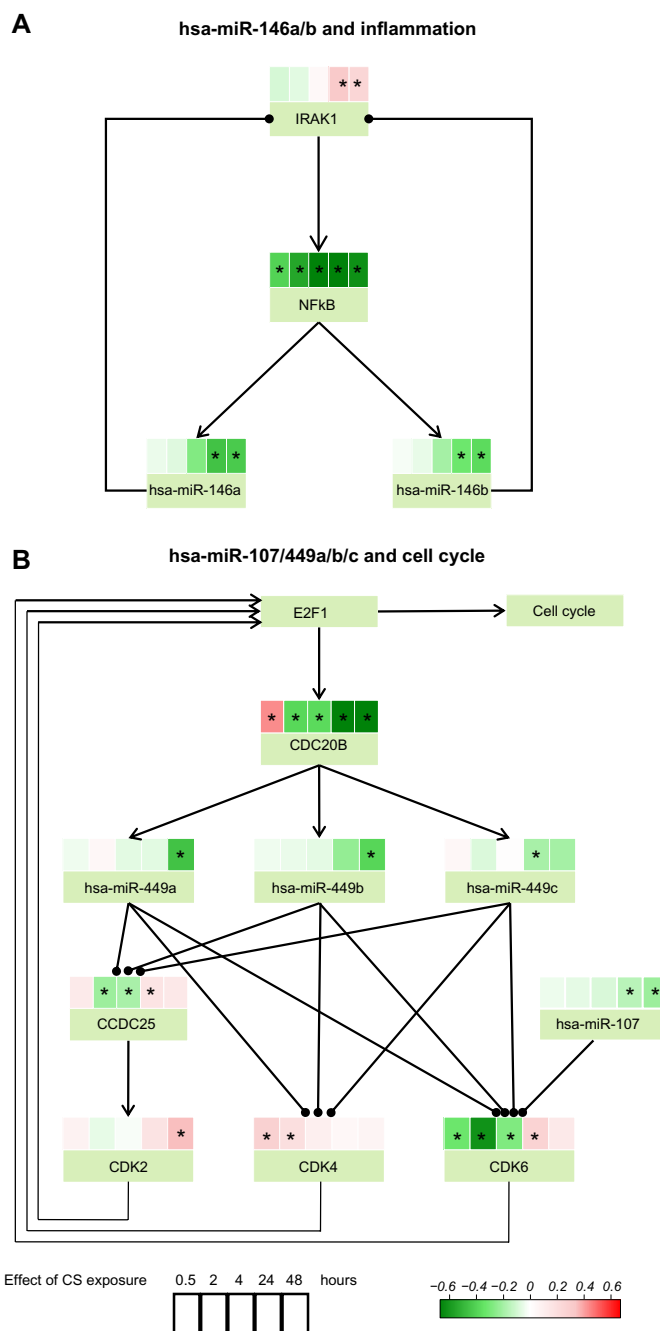


Figure 7. Small regulatory networks containing the differentially expressed miRNAs hsa-miR-146a/b (panel A) and hsa-miR-107/hsa-miR-449a/b/c (panel B). Each node contains a heat map displaying the differential expressions (coefficient β_g of the linear model) and their statistical significance (FDR < 0.05, indicated by an asterisk (*)) for the five post-exposure times (0.5, 2, 4, 24, 48 hours). The signed edges are indicated in the same way as in the other figures: --| represents activation and --| inhibition. The early time points 0.5, 2, and 4 hours were still included to show the complete response although they were not used to select relevant miRNAs.

concomitant transcription of CDC20B, hsa-miR-449a/b/c, and the minor strand hsa-miR-449b-3p (not shown on the network because it is a dead end) is clearly visible in the significant negative differential expression observed at 24- and 48-hour time points (Fig. 7B). This response is fully consistent



with the positive differential expression of CDK2/4/6 and CCDC25, all positively acting on E2F1 to promote the cell cycle. These observations include hsa-miR-107, which was also shown to intervene in the cell cycle regulation via the direct repression of CDK6.^{76–78} Examination of the full feedback loop contained in the small network revealed that the edge E2F1→CDC20B is not satisfied at the late time points because E2F1 is predicted to be positively activated using RCR (see Fig. 5B). This shows that the origin of the negative differential expression of the CDC20B transcript (later followed by its intronic miRNAs, hsa-miR-449a/b/c and hsa-miR-449b-3p) is not a late response but an early event, taking place in the first two hours following CS exposure. Our RCR results show that E2F1 activity was not responsive in the early response; thus, additional investigations are necessary to explain the early behavior of CDC20B expression.

Alternative evaluation of the miRNA–mRNA relationship. For three remaining miRNAs (hsa-miR-342–3p/375/675–5p) that could not be integrated into high-confidence regulatory networks because of a lack of evidence of causal relationships, we used a correlative approach to evaluate their relationship with observed mRNA changes and their potential role in exposure-related biological responses.

For hsa-miR-375, TarBase provided a large quantity (>100) of moderately reliable targets obtained in a microarray experiment. We used a global approach to show that the downregulation of hsa-miR-375 induced a statistically significant anti-correlation effect on its (direct and indirect) downstream targets,⁷² as expected for miRNAs that are expression repressors. This result (shown in the histograms of Supplementary Fig. 2B) indicated that the repressive effect of hsa-miR-375 is present at 24 hours post-exposure (two-sided Kolmogorov–Smirnov test P -value = 6.87×10^{-6}) and disappears at 48 hours (P -value = 0.488). However, the overlap of the distributions of the true targets (red curves in Supplementary Fig. 2B) and random genes (black curves) was too large (~85%) to enable reliable enrichment calculations to be performed, which would have helped in the assigning of functional categories to the subset of anti-correlated target genes.

For hsa-miR-342–3p/675–5p, we relied on isolated literature statements to try to link their differential expression to their target genes. For instance, hsa miR-342–3p was shown to target DNA methyltransferase 1 (DNMT1) and, thereby, inhibits cell proliferation in colorectal cancer cells.⁷⁹ The differential expression values of DNMT1 were consistent with a potential de-repression by the miRNA ($\beta_g > 0$ with FDR values <0.05 for both 24 and 48 hours post-exposure). Because it was the only positively differentially expressed miRNA, hsa-miR-675–5p is an interesting case (also because β_{dg} was the only statistically significant case). Moreover, it was recently shown to target the tumor suppressor runt-related transcription factor 1 (RUNX1).⁸⁰ The observed differential expression values of RUNX1 in the present study agrees with

this relationship ($\beta_g < 0$ with FDR values <0.05 for 24 hours post-exposure; the signal vanishes at 48 hours). The negative differential expression values of this tumor suppressor are consistent with the activation of the cell cycle observed at the late 24 and 48 hours time points, which was already supported by other miRNAs such as the miR-449 family and miR-107 family. Thus, differential expression of miRNAs can still be linked to identifiable effects on their downstream target mRNAs using a less-specific correlative approach when causal relationships are not available.

Summary and Conclusion

In line with the new paradigm shift in the toxicological assessment strategy advising the use of advanced tools and technologies,^{3,81} and with the need to find alternatives to animal testing,⁸² the effect of acute CS exposure was investigated in the present study using (i) a relevant human organotypic bronchial epithelial model that has been previously shown to closely mimic the biological response in the bronchial epithelium of smokers²¹; (ii) a whole smoke *in vitro* exposure system (Vitrocell® system); (iii) a systems toxicology approach based on classical toxicological parameters and the capture of complementary molecular endpoints such as mRNA and miRNA expression, after various exposure times/doses and at different post-exposure times; and (iv) a computational analysis approach combining complementary methods for transcriptomics data interpretation (GSEA²³ and RCR²⁴) and an integrated mRNA/miRNA evaluation.

During recent years, we have developed a network-based approach to mechanistically evaluate the biological impact of various exposures in experimental models.¹ This approach is based on building different biological causal network models relevant to biological processes, such as cell proliferation,²⁶ cellular stress,²⁵ and inflammatory processes,²⁷ in the lung. By applying RCR to identify HYPs or upstream regulators, the vast transcriptomics data generated in this study were interpreted, and signaling perturbations across various doses and post-exposure time points were compared. Our systems biology approach provided an insight into the different cellular responses triggered over time after CS exposure and demonstrated that they were dependent on the CS dose applied.

We believe that the toxicological method applied here in the assessment of CS exposure effect is the first step to a more advanced approach that will allow: (i) the quantification of the biological effect by measuring specific network perturbation amplitudes and the overall biological impact of an exposure using recently developed algorithms,^{38,39} (ii) the mechanistic representation of critical toxicological effects that span over different layers of biological organization (from the initiating molecular events toward the adverse outcome), and (iii) the integration of information derived from proteomics and miRNA analysis into the networks.

Another consideration that should be investigated in the future is the assessment of chronic exposure effect and the

use of tissue inserts derived from different donors. These two challenging aspects are necessary to better understand the long-term effect of CS exposure and the possibility to leverage these observations to a population level.

To conclude, the combination of an *in vitro* model, leveraging human organotypic bronchial epithelial tissue culture, and an *in vitro* ALI exposure model allows for the biochemical analysis of distinct, well-regulated, dose- and time-dependent perturbations to cellular stress responses, inflammatory processes, and cell proliferation. This model and the used system biology approaches have the potential to unravel the underlying molecular mechanisms activated following CS exposure and can be viewed as a paradigm for a systems toxicology approach in “21st Century Toxicology”.^{2–4}

Acknowledgments

The authors thank Lynda Conroy-Schneider for expert editorial support; Marja Talikka, Walter Schlage, and Yang Xiang for helpful discussions and support during the manuscript preparation/revision; and Patrice Leroy for statistical support. They also acknowledge Horst Weiler, Christina Berger, Marek Franitza, Torsten Wilms, Jessica Buechel, Sigrid Lufen, Marco Hebestreit, and Birgit Kurkowsky (all from Philip Morris Research Laboratories) for their skilled technical assistance.

Author Contributions

Conceived and designed the experiments: DW, CM, SW, JH, MCP. Analyzed the data: CM, SG, CP, VB, DW, AH, AS, SA. Wrote the first draft of the manuscript: SG, CM. Contributed to the writing of the manuscript: CP, VB, AS, DW, AH. Agreed with manuscript results and conclusions: all authors. Developed the structure and arguments for the paper: CM, SG, CP, DW, AS, VB. All authors reviewed and approved the final manuscript.

Supplementary Data

Supplementary Figure 1. High-resolution pdf version of the subnetworks presented in Figures 3B, 5A, and 6.

Supplementary Figure 2. Integrated mRNA–miRNA analyses. Panel A displays the miRNA differential expression results obtained from the linear model used in the integrated miRNA–mRNA analysis. The heat map on the left-hand side shows the coefficient β_g . Statistically, significant values are indicated by an asterisk. The differentially expressed miRNAs (with black labels) are complemented by their family members (with gray labels) as discussed in the Materials and Methods. A coherent signal within a miRNA family supports the inclusion of all its members (hsa-miR-449a/b/c beside the minor strand hsa-miR-449b-3p), whereas the absence of a shared signal suggested a probable false positive that was discarded (hsa-miR-30e-3p). The colors in the side bars help in distinguishing the families. The heat map on the right-hand side displays the values for coefficient β_{dg} , for which the inclusion of the full

miRNA families was unnecessary (except for miR-449). Panel B displays two histograms containing the Pearson correlations for the CS-induced differential expression values between hsa-miR-375 and its TarBase targets (red) or all expressed genes (black, labeled as background) (see Materials and Methods).

Supplementary Table 1. (A) Number of significantly differentially regulated genes for each given exposure time (dose) and post-exposure time point. Significantly differentially regulated genes were selected by an absolute fold change ≥ 1.3 , a significance level (FDR) ≤ 0.05 , and an abundance level $\geq \log_2(100)$. (B) A list of HYPs predicted to be significant in at least one of the exposure conditions. Concordance values and network membership are shown. P, present and A, absent. (C) Additional information for Figure 2 (APcluster): names of particular HYPs are shown. Color code for Supplementary Table 1B and C: yellow, HYPs with predicted increase in abundance or activity; blue, HYPs with predicted decrease in abundance or activity; light yellow/blue, $0.05 \leq P < 0.1$; yellow/blue, $0.001 \leq P < 0.05$; dark yellow/blue, $P < 0.001$; and gray, no significant changes. catof(X), catalytic activity of X; gtpof(X), GTP-bound activity of (X); kaof(X), kinase activity of X; paof(X), phosphatase activity of X; taof(X), transcriptional activity of X; P, present; and A, absent. (D) Differential gene expression (CS exposure vs. sham exposure) of selected genes. Green highlight, downregulated genes; red highlight, upregulated genes; and FC, fold change.

REFERENCES

1. Hoeng J, Deehan R, Pratt D, et al. A network-based approach to quantifying the impact of biologically active substances. *Drug Discov Today*. 2012;17(9–10):413–8.
2. Mahadevan B, Snyder RD, Waters MD, et al. Genetic toxicology in the 21st century: reflections and future directions. *Environ Mol Mutagen*. 2011; 52(5):339–54.
3. Stephens ML, Barrow C, Andersen ME, et al. Accelerating the development of 21st-century toxicology: outcome of a human toxicology project consortium workshop. *Toxicol Sci*. 2012;125(2):327–34.
4. NRC. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington DC: Natl Acad Press; 2007.
5. Coggins CR. A further review of inhalation studies with cigarette smoke and lung cancer in experimental animals, including transgenic mice. *Inhal Toxicol*. 2010;22(12):974–83.
6. Fehrenbach H. Animal models of chronic obstructive pulmonary disease: some critical remarks. *Pathobiology*. 2002;70(5):277–83.
7. Singh A, Misra V, Thimmulappa RK, et al. Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer. *PLoS Med*. 2006;3(10):e420.
8. Rothen-Rutishauser B, Blank F, Muhlfeld C, Gehr P. *In vitro* models of the human epithelial airway barrier to study the toxic potential of particulate matter. *Expert Opin Drug Metab Toxicol*. 2008;4(8):1075–89.
9. Aufderheide M, Knebel JW, Ritter D. An improved *in vitro* model for testing the pulmonary toxicity of complex mixtures such as cigarette smoke. *Exp Toxicol Pathol*. 2003;55(1):51–7.
10. Aufderheide M, Scheffler S, Mohle N, Halter B, Hochrainer D. Analytical *in vitro* approach for studying cyto- and genotoxic effects of particulate airborne material. *Anal Bioanal Chem*. 2011;401(10):3213–20.
11. Olivera DS, Boggs SE, Beenhouwer C, Aden J, Knall C. Cellular mechanisms of mainstream cigarette smoke-induced lung epithelial tight junction permeability changes *in vitro*. *Inhal Toxicol*. 2007;19(1):13–22.
12. Phillips J, Kluss B, Richter A, Massey E. Exposure of bronchial epithelial cells to whole cigarette smoke: assessment of cellular responses. *Altern Lab Anim*. 2005;33(3):239–48.
13. Knorr-Wittmann C, Hengstermann A, Gebel S, Alam J, Muller T. Characterization of Nrf2 activation and heme oxygenase-1 expression in NIH3T3 cells exposed to aqueous extracts of cigarette smoke. *Free Radic Biol Med*. 2005;39(11):1438–48.



14. Fields WR, Leonard RM, Odom PS, Nordskog BK, Ogden MW, Doolittle DJ. Gene expression in normal human bronchial epithelial (NHBE) cells following *in vitro* exposure to cigarette smoke condensate. *Toxicol Sci.* 2005;86(1):84–91.
15. Dvorak A, Tilley AE, Shaykhiyev R, Wang R, Crystal RG. Do airway epithelium air-liquid cultures represent the *in vivo* airway epithelium transcriptome? *Am J Respir Cell Mol Biol.* 2011;44(4):465–73.
16. Pezzullo AA, Starner TD, Scheetz TE, et al. The air-liquid interface and use of primary cell cultures are important to recapitulate the transcriptional profile of *in vivo* airway epithelia. *Am J Physiol Lung Cell Mol Physiol.* 2011;300(1):L25–31.
17. Balharry D, Sexton K, BeruBe KA. An *in vitro* approach to assess the toxicity of inhaled tobacco smoke components: nicotine, cadmium, formaldehyde and urethane. *Toxicology.* 2008;244(1):66–76.
18. Yoneda K, Peck K, Chang MM, et al. Development of high-density DNA microarray membrane for profiling smoke- and hydrogen peroxide-induced genes in a human bronchial epithelial cell line. *Am J Respir Crit Care Med.* 2001;164(10 pt 2):S85–9.
19. Sexton K, Balharry D, BeruBe KA. Genomic biomarkers of pulmonary exposure to tobacco smoke components. *Pharmacogenet Genomics.* 2008;18(10):853–60.
20. Lam HC, Choi AM, Ryter SW. Isolation of mouse respiratory epithelial cells and exposure to experimental cigarette smoke at air liquid interface. *J Vis Exp.* 2011;48. <http://www.ncbi.nlm.nih.gov/pubmed/21372793>
21. Mathis C, Poussin C, Weisensee D, et al. Human bronchial epithelial cells exposed *in vitro* to cigarette smoke at the air-liquid interface resemble bronchial epithelium from human smokers. *Am J Physiol Lung Cell Mol Physiol.* 2013;304(7):L489–503.
22. Maunders H, Patwardhan S, Phillips J, Clack A, Richter A. Human bronchial epithelial cell transcriptome: gene expression changes following acute exposure to whole cigarette smoke *in vitro*. *Am J Physiol Lung Cell Mol Physiol.* 2007;292(5):L1248–56.
23. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005;102(43):15545–50.
24. Catlett NL, Bargnesi AJ, Ungerer S, et al. Reverse causal reasoning: applying qualitative causal knowledge to the interpretation of high-throughput data. *BMC Bioinformatics.* 2013;14(1):340.
25. Schlage WK, Westra JW, Gebel S, et al. A computable cellular stress network model for non-diseased pulmonary and cardiovascular tissue. *BMC Syst Biol.* 2011;5:168.
26. Westra JW, Schlage WK, Frushour BP, et al. Construction of a computable cell proliferation network focused on non-diseased lung cells. *BMC Syst Biol.* 2011;5:105.
27. Westra JW, Schlage WK, Hengstermann A, et al. A modular cell-type focused inflammatory process network model for non-diseased pulmonary tissue. *Bioinform Biol Insights.* 2013;7:167–92.
28. Vergoulis T, Vlachos IS, Alexiou P, et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res.* 2012;40(Database issue):D222–9.
29. Weber S, Hebestreit M, Wilms T, Conroy LL, Rodrigo G. Comet assay and air-liquid interface exposure system: a new combination to evaluate genotoxic effects of cigarette whole smoke in human lung cell lines. *Toxicol In Vitro.* 2013;27(6):1987–91.
30. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics.* 2005;21(9):2067–75.
31. Belcastro V, Poussin C, Gebel S, et al. Systematic verification of upstream regulators of a computable cellular proliferation network model on non-diseased lung cells using a dedicated dataset. *Bioinform Biol Insights.* 2013;7:217–30.
32. Frey BJ, Dueck D. Clustering by passing messages between data points. *Science.* 2007;315(5814):972–6.
33. Hu QN, Deng Z, Hu H, Cao DS, Liang YZ. RxnFinder: biochemical reaction search engines using molecular structures, molecular fragments and reaction similarity. *Bioinformatics.* 2011;27(17):2465–7.
34. Bethune J, Artus-Revel CG, Filipowicz W. Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep.* 2012;13(8):716–23.
35. Nam JW, Rissland OS, Koppstein D, et al. Global analyses of the effect of different cellular contexts on microRNA targeting. *Mol Cell.* 2014;53(6):1031–43.
36. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature.* 2010;466(7308):835–40.
37. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 2011;27(12):1739–40.
38. Martin F, Thomson TM, Sewer A, et al. Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks. *BMC Syst Biol.* 2012;6:54.
39. Thomson TM, Sewer A, Martin F, et al. Quantitative assessment of biological impact using transcriptomic data and mechanistic network models. *Toxicol Appl Pharmacol.* 2013;272(3):863–78.
40. Gebel S, Diehl S, Pype J, et al. The transcriptome of Nrf2^{-/-} mice provides evidence for impaired cell cycle progression in the development of cigarette smoke-induced emphysematous changes. *Toxicol Sci.* 2010;115(1):238–52.
41. Reddy SP. The antioxidant response element and oxidative stress modifiers in airway diseases. *Curr Mol Med.* 2008;8(5):376–83.
42. Dhakshinamoorthy S, Jain AK, Bloom DA, Jaiswal AK. Bach1 competes with Nrf2 leading to negative regulation of the antioxidant response element (ARE)-mediated NAD(P)H: quinone oxidoreductase 1 gene expression and induction in response to antioxidants. *J Biol Chem.* 2005;280(17):16891–900.
43. Karin M, Shaulian E. AP-1: linking hydrogen peroxide and oxidative stress to the control of cell proliferation and death. *IUBMB Life.* 2001;52(1–2):17–24.
44. Luppi F, Aarbiou J, van Wetering S, et al. Effects of cigarette smoke condensate on proliferation and wound closure of bronchial epithelial cells *in vitro*: role of glutathione. *Respir Res.* 2005;6:140.
45. Kitamura M, Kasai A. Cigarette smoke as a trigger for the dioxin receptor-mediated signaling pathway. *Cancer Lett.* 2007;252(2):184–94.
46. Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *J Biol Chem.* 2004;279(23):23847–50.
47. Hoeng J, Talikka M, Martin F, et al. Case study: the role of mechanistic network models in systems toxicology. *Drug Discov Today.* 2014;19(2):183–92.
48. Betsuyaku T, Hamamura I, Hata J, et al. Bronchiolar chemokine expression is different after single versus repeated cigarette smoke exposure. *Respir Res.* 2008;9:7.
49. Chen GY, Nunez G. Sterile inflammation: sensing and reacting to damage. *Nat Rev Immunol.* 2010;10(12):826–37.
50. Tolle LB, Standiford TJ. Danger-associated molecular patterns (DAMPs) in acute lung injury. *J Pathol.* 2013;229(2):145–56.
51. Gebel S, Muller T. The activity of NF-kappaB in Swiss 3T3 cells exposed to aqueous extracts of cigarette smoke is dependent on thioredoxin. *Toxicol Sci.* 2001;59(1):75–81.
52. Loukili N, Rosenblatt-Velin N, Rolli J, et al. Oxidants positively or negatively regulate nuclear factor kappaB in a context-dependent manner. *J Biol Chem.* 2010;285(21):15746–52.
53. Lambert C, Li J, Jonscher K, et al. Acrolein inhibits cytokine gene expression by alkylating cysteine and arginine residues in the NF-kappaB1 DNA binding domain. *J Biol Chem.* 2007;282(27):19666–75.
54. Allahverdiyan S, Harada N, Singhera GK, Knight DA, Dorscheid DR. Secretion of IL-13 by airway epithelial cells enhances epithelial repair via HB-EGF. *Am J Respir Cell Mol Biol.* 2008;38(2):153–60.
55. Gensch E, Gallup M, Sucher A, et al. Tobacco smoke control of mucin production in lung cells requires oxygen radicals AP-1 and JNK. *J Biol Chem.* 2004;279(37):39085–93.
56. Shen N, Gong T, Wang JD, et al. Cigarette smoke-induced pulmonary inflammatory responses are mediated by EGR-1/GGPPS/MAPK signaling. *Am J Pathol.* 2011;178(1):110–18.
57. Adcock IM, Caramori G, Barnes PJ. Chronic obstructive pulmonary disease and lung cancer: new molecular insights. *Respiration.* 2011;81(4):265–84.
58. Sakamaki Y, Matsumoto K, Mizuno S, Miyoshi S, Matsuda H, Nakamura T. Hepatocyte growth factor stimulates proliferation of respiratory epithelial cells during postpneumectomy compensatory lung growth in mice. *Am J Respir Cell Mol Biol.* 2002;26(5):525–33.
59. Ware LB, Matthay MA. Keratinocyte and hepatocyte growth factors in the lung: roles in lung development, inflammation, and repair. *Am J Physiol Lung Cell Mol Physiol.* 2002;282(5):L924–40.
60. Nadel JA. Mucous hypersecretion and relationship to cough. *Pulm Pharmacol Ther.* 2013;26(5):510–3.
61. Zhang Y, Jiang Y, Sun C, et al. The human cathelicidin LL-37 enhances airway mucus production in chronic obstructive pulmonary disease. *Biochem Biophys Res Commun.* 2014;443(1):103–9.
62. Cao R, Wang TT, DeMaria G, Sheehan JK, Kesimer M. Mapping the protein domain structures of the respiratory mucins: a mucin proteome coverage study. *J Proteome Res.* 2012;11(8):4013–23.
63. Finkbeiner WE, Zlock LT, Morikawa M, Lao AY, Dasari V, Widdicombe JH. Cystic fibrosis and the relationship between mucin and chloride secretion by cultures of human airway gland mucous cells. *Am J Physiol Lung Cell Mol Physiol.* 2011;301(4):L402–14.
64. Di YP, Zhao J, Harper R. Cigarette smoke induces MUC5 AC protein expression through the activation of Sp1. *J Biol Chem.* 2012;287(33):27948–58.
65. Takeyama K, Jung B, Shim JJ, et al. Activation of epidermal growth factor receptors is responsible for mucin synthesis induced by cigarette smoke. *Am J Physiol Lung Cell Mol Physiol.* 2001;280(1):L165–72.
66. An J, Li JQ, Wang T, et al. Blocking of thromboxane A(2) receptor attenuates airway mucus hyperproduction induced by cigarette smoke. *Eur J Pharmacol.* 2013;703(1–3):11–7.
67. Baginski TK, Dabbagh K, Satjwacharaphong C, Swinney DC. Cigarette smoke synergistically enhances respiratory mucin induction by proinflammatory stimuli. *Am J Respir Cell Mol Biol.* 2006;35(2):165–74.



68. Glader P, Moller S, Lilja J, Wieslander E, Lofdahl CG, von Wachenfeldt K. Cigarette smoke extract modulates respiratory defence mechanisms through effects on T-cells and airway epithelial cells. *Respir Med*. 2006;100(5):818–27.
69. Xiao J, Wang K, Feng YL, Chen XR, Xu D, Zhang MK. Role of extracellular signal-regulated kinase 1/2 in cigarette smoke-induced mucus hypersecretion in a rat model. *Chin Med J (Engl)*. 2011;124(20):3327–33.
70. Boucherat O, Boczkowski J, Jeannotte L, Delacourt C. Cellular and molecular mechanisms of goblet cell metaplasia in the respiratory airways. *Exp Lung Res*. 2013;39(4–5):207–16.
71. Schembri F, Sridhar S, Perdomo C, et al. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proc Natl Acad Sci USA*. 2009;106(7):2319–24.
72. Wang J, Lu M, Qiu C, Cui Q. TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Res*. 2010;38(Database issue):D119–22.
73. O'Neill LA, Sheedy FJ, McCoy CE. MicroRNAs: the fine-tuners of Toll-like receptor signalling. *Nat Rev Immunol*. 2011;11(3):163–75.
74. Leung AK, Sharp PA. MicroRNA functions in stress responses. *Mol Cell*. 2010;40(2):205–15.
75. Lize M, Klimke A, Dobbstein M. MicroRNA-449 in cell fate determination. *Cell Cycle*. 2011;10(17):2874–82.
76. Feng L, Xie Y, Zhang H, Wu Y. miR-107 targets cyclin-dependent kinase 6 expression, induces cell cycle G1 arrest and inhibits invasion in gastric cancer cells. *Med Oncol*. 2012;29(2):856–63.
77. Lee KH, Lotterman C, Karikari C, et al. Epigenetic silencing of MicroRNA miR-107 regulates cyclin-dependent kinase 6 expression in pancreatic cancer. *Pancreatol*. 2009;9(3):293–301.
78. Takahashi Y, Forrest AR, Maeno E, Hashimoto T, Daub CO, Yasuda J. MiR-107 and MiR-185 can induce cell cycle arrest in human non small cell lung cancer cell lines. *PLoS One*. 2009;4(8):e6677.
79. Wang H, Wu J, Meng X, et al. MicroRNA-342 inhibits colorectal cancer cell proliferation and invasion by directly targeting DNA methyltransferase 1. *Carcinogenesis*. 2011;32(7):1033–42.
80. Zhuang M, Gao W, Xu J, Wang P, Shu Y. The long non-coding RNA H19-derived miR-675 modulates human gastric cancer cell proliferation by targeting tumor suppressor RUNX1. *Biochem Biophys Res Commun*. 2014;448(3):315–22.
81. Sturla SJ, Boobis AR, FitzGerald RE, et al. Systems toxicology: from basic research to risk assessment. *Chem Res Toxicol*. 2014;27(3):314–29.
82. European Commission. Alternative testing strategies progress report. In: Commission E, ed. *Replacing, Reducing, and Refining Use of Animals in Research*. Brussels: European Commission; 2009:1–280.