Identification of drug candidates to suppress cigarette smoke-induced inflammation via cMap analyses

Gilles Vanderstocken1*, Anna Dvorkin-Gheva1,2*, Pamela Shen1#, Corry-Anke Brandsma3, Ma’en Obeidat4, Yohan Bosse5,6, John A Hassell2, Martin R Stampfl1,7

1Department of Pathology and Molecular Medicine, McMaster Immunology Research Centre, McMaster University, Hamilton, ON, Canada. 2Department of Pathology and Molecular Medicine, Centre for Functional Genomics, McMaster University, Hamilton, ON, Canada. 3University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen, The Netherlands. 4The University of British Columbia Centre for Heart Lung Innovation, St. Paul’s Hospital, Vancouver, BC, Canada. 5Centre de Recherche de l’Institut Universitaire de Cardiologie et de Pneumologie de Québec, Québec City, Québec, Canada. 6Department of Molecular Medicine, Laval University, Québec City, Québec, Canada. 7Department of Medicine, Firestone Institute for Respiratory Health at St. Joseph's Healthcare, McMaster University, Hamilton, ON, Canada.

Author contributions: *G.V. and A.D-G contributed equally to this work. Conception and design (G.V., A.D-G, M.R.S); data acquisition, analysis and interpretation (G.V., A.D-G, P.S., C-A.B, M.O., Y.B., M.R.S); drafting of the manuscript for important intellectual content (G.V., A.D-G, M.R.S).

#Current affiliation: Department of Microbiology, New York University, New York, New York, USA.
Corresponding author: Martin R. Stampfli, McMaster Immunology Research Centre
McMaster University, MDCL-4011, 1280 Main Street West, Hamilton ON Canada L8S 4K1.
Phone: 905-525-9140 ext. 22493, Fax: 905-522-6750. Email: stampfli@mcmaster.ca

Running title: Kaempferol decreases cigarette smoke-induced inflammation

Keywords: Connectivity Map, Cigarette Smoke, Kaempferol, Lung, Gene, Inflammation
ABSTRACT

Cigarette smoking is the main risk factor for COPD, and to date, existing pharmacologic interventions have been ineffective at controlling inflammatory processes associated with the disease. To address this issue, we used the Connectivity Map (cMap) database to identify drug candidates with the potential to attenuate cigarette smoke-induced inflammation. We queried cMap using three independent in-house cohorts of healthy non-smokers and smokers. Potential drug candidates were validated against four publicly available human datasets, as well as six independent data sets from cigarette smoke-exposed mice. Overall, these analyses yielded two potential drug candidates: kaempferol and bethanechol. Subsequently, the efficacy of each drug was validated in vivo in a model of cigarette smoke-induced inflammation. BALB/c mice were exposed to room air or cigarette smoke and treated with each of the two candidate drugs either prophylactically or therapeutically. We found that kaempferol, but not bethanechol, was able to reduce cigarette smoke-induced neutrophilia both when administered prophylactically and therapeutically. Mechanistically, kaempferol decreased expression of IL-1α and CXCL-5 levels in the lung. Our data suggest that cMap analyses may serve as a useful tool to identify novel drug candidates against cigarette smoke-induced inflammation.
INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a term that describes a range of airway and lung disorders characterized physiologically by progressive and largely irreversible airflow limitation, and pathologically by bronchiolitis and emphysema(1). Inherent to COPD is the notion of an abnormal inflammatory response in the airways and lungs to inhaled materials, the most prominent of which is cigarette smoke(2). Notably, current pharmacological interventions have been largely ineffective at treating COPD, and with over 1.1 billion smokers in the world today, the disease remains a major health challenge. This emphasizes the need for the development of novel prophylactic and therapeutic interventions with which to reduce the burden of smoking-related disease.

The Connectivity Map (cMap) database is an in silico drug discovery and hypothesis-generating approach(3). The database is comprised of gene signatures derived of cell lines treated with small molecule compounds to generate a “connectivity map”. This map can be queried to identify similarities between gene expression profiles induced by compounds to gene signatures associated with particular disease statuses. Identified compounds can subsequently be assessed for their ability to ameliorate disease phenotypes. The database was first described in 2006(3), and several studies have since validated the approach(4–6). Thus, this approach has the potential to assist in the identification of novel therapeutics with which to treat COPD. However, the necessity for a disease to be associated with a defined gene expression profile is a challenge for COPD, as the disease is characterized by substantial patient-to-patient heterogeneity. For example, in individual patients with COPD, varying degrees of chronic bronchitis, bronchiolitis and emphysema contribute to chronic airflow limitation. A potential solution to this challenge is the restricted analysis of well-defined disease sub- or endo-phenotypes. Alternatively, another
approach is to use multiple independent datasets and compute interventions that are conserved among cohorts, in order to target treatments that may help the COPD patient population at large.

In the current study, we used cMap to identify drug candidates that attenuate cigarette smoke-induced inflammation. We queried the cMap database using three independent, in house human lung datasets(7). Potential compounds were validated against four publically available human lung datasets. To facilitate in vivo validation of compounds we pursued comparative cMap analyses in six different mouse data sets. Overall, these analyses yielded two potential drug candidates, kaempferol and bethanechol, that were tested for their ability to attenuate cigarette smoke-induced inflammation in mice. While kaempferol attenuated cigarette smoke-induced neutrophilia both when administered prophylactically or therapeutically, bethanechol only attenuated cigarette smoke-induced inflammation in the prophylactic setting. These data show that cMap can be used to identify potential target compounds to attenuate cigarette smoke-induced inflammatory processes.
MATERIALS AND METHODS

Human microarrays
Details of the three internal datasets (GSE23546) are described in (7) and the online supplement. The lists of genes differentially expressed between smokers and healthy non-smokers were obtained separately for each of the subsets (‘limma’ package, R) and only genes with a corrected p-value of <0.05 that were shared by the three lists were used for further cMAP analysis. In addition to the internal datasets, we included gene expression profiles of four publicly available datasets (GEO datasets accession numbers GSE47460 (8), GSE22047 (9), GSE10072 (10), GSE994 (11)). Details of the corresponding data sets are provided in the online supplement. For each of the datasets, we performed a separate comparison between smokers and healthy non-smokers. The differential expression analysis was performed using ‘limma’ package (R), and genes differentially expressed with a fold change of at least 2.0 were selected for cMAP analysis. Genes were considered to be differentially expressed if the comparison yielded a corrected p-value of <0.05.

Mouse microarray data
Mouse lung gene expression profiles (GSE55127, GSE33512, GSE18344, GSE8790, GSE17737, GSE52509, and GSE33561) are described in (12) and the online supplement. The differential expression analysis was performed using ‘limma’ package (R), and genes differentially expressed with a fold change of at least 1.5 were selected for cMAP analysis. This relatively stringent ≥1.5-fold threshold was chosen in accordance with a previous publication (13).
Connectivity Mapping (cMap) analysis

Lists of genes found to be differentially expressed in human and mouse datasets were used to query the cMap database at https://portals.broadinstitute.org/cmap/. For more detailed description see online supplement.

Animals, cigarette smoke exposure, and intervention protocols

Female 6-8 week old BALB/c mice were purchased from Charles River Laboratories (Saint-Constant, PQ, Canada) and housed as described in the online supplement. BALB/c mice were exposed to room air or cigarette smoke for either four days (prophylactic protocol) or five weeks (therapeutic protocol) as described previously and in the online supplement (14, 15). Bethanechol (Sigma Aldrich, St Louis, MO, USA) was administered intranasally in 35 µL of PBS at 1.75 mg/kg (16). Kaempferol (Sigma Aldrich) was administered intranasally in 35 µL of 10% DSMO in PBS at 1.75 mg/kg (17, 18). Vehicle control mice received PBS or 10% DSMO in PBS, respectively. All intranasal instillations were performed daily, 1h prior to the first cigarette smoke exposure. All experiments were approved by the Animal Research Ethics Board at McMaster University (Hamilton, Canada).

Sample processing

Bronchoalveolar lavage (BAL) and differential cell counts were performed as described in detail previously (19). Cells were pelleted and the BAL fluid (BALF) kept at -80°C for cytokine measurements.

Cytokine assessment

Interleukin (IL)-1α, and CXCL-5 were measured by ELISA according to manufacturer’s instructions (R&D Systems, Minneapolis, MN, USA).
Statistical analysis of data obtained from *in vivo* studies

All *in vivo* experiments were performed at least 2 times. Statistical analyses were performed using GraphPad Prism Software (La Jolla, CA, USA). Data are expressed as mean ± SEM. Student’s t-tests and ANOVAs were used as appropriate. A p-value of <0.05 was considered significant.
RESULTS

cMap analyses of human lung gene signatures from the Laval, Groningen, and UBC cohorts

To query the cMap database, we prepared signatures of differentially expressed genes using gene arrays of healthy non-smokers and smokers from the Laval, Groningen, and UBC cohorts (internal datasets). The individual cohorts were described in detail previously(12). All subjects included in the analysis were 18 years or older. All current smokers had a history of ≥10 pack-years. Table 1 shows the clinical characteristics and number of subjects included in the analysis. Differential gene expression analyses were performed separately on each cohort. We focused on gene transcripts with ≥2.0-fold change between the smokers and non-smokers. While age, smoking history, and lung function were comparable across cohorts, we observed a marked variability in numbers of differentially expressed genes between the cohorts with most genes modulated in the Laval and fewest in the UBC cohort (Table 2). Venn diagrams show the number of shared genes that were upregulated (Figure 1A) and downregulated (Figure 1 B) among the cohorts. Among all three cohorts, we observed 90 genes that were commonly upregulated and 4 genes that were downregulated. A summary of the genes shared among all three cohorts is provided in Supplemental Table 1. As cMap requires the signature to be represented by probe sets available on a specific array, genes were translated into probe sets before querying the cMap database (see methods for details). Only the probe sets that were shared across the 3 cohorts were used to create the signature. cMap analysis predicted 39 compounds to revert the gene expression profile observed in smokers (Figure 1C and Supplemental Table 2).

Validation of candidate compounds using external cohorts
To validate candidate compounds predicted from the internal cohorts, we used four independent publicly available cohorts (GSE47460(8), GSE22047(9), GSE10072(10), GSE994(11)). The GSE22047 cohort was divided into two separate groups, based on the source of the samples: (1) samples obtained from large airways, and (2) samples obtained from small airways. We performed differential gene expression analyses between smokers and non-smokers for each of the cohorts separately, thus obtaining five independent signatures. Numbers of differentially expressed genes and probe sets used for the cMap analysis are shown in Table 3. A summary of differentially expressed genes is provided in Supplemental Table 3. Each of the signatures was used to query the cMap database independently, yielding a list of candidate compounds for each of the data sets (Supplemental Table 2). There was limited overlap of potential candidate compounds between the internal and the four publicly available cohorts. Of the 39 compounds found in the internal cohorts, Kaempferol was predicted in two of the four external cohorts, while Bethanechol, Flumequine, Naftifine, and Pentoxifylline were observed in one of the four external cohorts.

**cMap analyses of mouse data sets**

Next, we used a collection of six mouse datasets to determine drug candidates that were observed in mouse models of cigarette smoke-induced inflammation (20–25). The objective of this analysis was to set the stage for target validation in animal models. We included in the analysis datasets from six different cigarette smoke exposure protocols. All profiles are publicly available in GEO (accession numbers GSE55127, GSE33512, GSE18344, GSE17737, GSE52509 and GSE33561) and comprise a total of 54 samples. Datasets containing room air and cigarette smoke-exposed mice were combined into a collection with pooled control samples, as detailed previously(12). We performed differential gene expression analyses to compare cigarette smoke-exposed mice
from each exposure protocol to the group of pooled controls. The numbers of differentially expressed genes in the six experimental models are shown in Supplemental Table 4. Similar to what we observed across human cohorts, gene expression profiles varied substantially among the different datasets (Table 4). Genes that changed ≥1.5-fold between the smoke- and room air-exposed mice were used in cMap analysis. Supplemental Table 2 shows predicted drug candidates in the different mouse experimental systems. The compound Kaempferol, which was identified in cMap analyses using the internal cohorts as well as two of the four external cohorts, was also observed in two of the mouse models. Similarly, bethanechol and pentoxifylline were predicted in the internal and one of the external cohorts, as well as one of the mouse models. Of note, pentoxifylline is used in the clinic and has been reported to attenuate cigarette smoke-induced emphysema (26–29), hence validating the experimental approach. Flumequine and Naftifine, while present in the internal and one of the external cohorts, did not appear in any of the cMap analyses of the mouse models and were not further investigated. Thus, we chose to assess the efficacy of kaempferol and bethanechol in ameliorating cigarette smoke-induced inflammation in our mouse model moving forward. Table 5 summarizes all the shared drug candidates in the different cohorts and models.

**Experimental validation of predicted drug candidates in cigarette smoke-exposed mice**

To begin, we investigated the efficacy of kaempferol and bethanechol in attenuating cigarette smoke-induced inflammatory processes in a prophylactic setting. BALB/c mice were exposed to room air or cigarette smoke for four days. Bethanechol and kaempferol were administered daily, one hour prior to the first smoke exposure (Fig. 2A-C). Bethanechol and kaempferol attenuated the increase in BAL neutrophils by 19% (p = 0.1503, ns) and 70% (p = 0.0005, ***) respectively, compared to untreated cigarette smoke-exposed mice (Fig 2C). No differences were observed in
BAL mononuclear cells (Fig. 2B). Since bethanechol only moderately impacted BAL neutrophils, we pursued subsequent studies with kaempferol only.

To investigate whether kaempferol was able to attenuate established cigarette smoke-induced inflammatory processes, we exposed BALB/c mice to cigarette smoke for five weeks. Kaempferol was administered daily in the fifth week of smoke exposure (therapeutic intervention) (Fig. 2D-F). BAL neutrophil numbers were reduced by 47% (p<0.05,**), compared to cigarette smoke-exposed mice (Fig 2C). No difference was observed in BAL mononuclear cells, similar to our observations in the prophylactic intervention (Fig 2B). Together, these data show that kaempferol attenuated cigarette smoke-induced pulmonary neutrophilia when administered in both a prophylactic and therapeutic manner.

Mechanisms of attenuated cigarette smoke-induced inflammation

To investigate mechanisms of kaempferol-mediated attenuation of inflammation, we assessed expression of interleukin (IL)-1α and C-X-C motif chemokine ligand (CXCL)5 in the BALF of room air and cigarette smoke-exposed mice. Both inflammatory mediators are critically involved in cigarette smoke-induced inflammatory processes(15, 30, 31). We observed reduced expression of IL-1α and CXCL5 for mice treated with kaempferol in the prophylactic intervention protocol (Fig 3A and B). Similarly, we observed a significant reduction of CXCL5 in the therapeutic protocol following kaempferol treatment, while IL-1α was not impacted under these experimental conditions (Fig 3C and D). Hence, decreased expression of neutrophil chemoattractants likely contributed to the observed attenuation of neutrophilic inflammation. Overall, these experimental data provide proof of concept that kaempferol attenuates cigarette smoke-induced neutrophilia.
DISCUSSION

In this study, we used cMap to discover new drug candidates to attenuate inflammatory processes associated with COPD. We utilized gene expression profiles of human (smokers and never-smokers) and mouse (cigarette smoke- and room air-exposed) pulmonary samples to query the cMap database. The unique feature of cMap is that it allows for the screening of compounds against genome-wide disease signatures, rather than a pre-selected set of target genes. We prioritized candidates that were identified in both human and mouse arrays to have the potential to correct smoking-associated gene expression signatures.

COPD is a complex disease, exhibiting multiple and not always well-defined disease phenotypes of varying frequencies within individual cohorts. In line with this, we observed significant heterogeneity in gene expression in resected lung tissue from smokers and non-smokers across multiple cohorts. However, we were able to identify a number of drug candidates that were shared among multiple datasets, that consequently may have the potential to attenuate cigarette smoke-induced inflammation in diverse patient populations. Overall, the cMap analysis yielded 14 drug candidates conserved in humans and mice. Among them, five candidates were predicted in more than two human datasets. These included, for instance, pentoxifylline, a non-specific phosphodiesterase inhibitor that is recognized as a modulator of immune function and attenuates cigarette smoke-induced emphysema(26–29). Pentoxifylline has been tested in COPD patients, although results are controversial(32). However, this example validates this unbiased, in silico experimental approach. Flumequine, an antibiotic that is no longer used clinically, and naftifine, an antifungal drug, were also predicted in the human datasets. However, these compounds did not appear in any of the cMap analyses of mouse tissues and, hence, were not investigated further.
In contrast, we selected bethanechol and kaempferol, two other compounds identified in both human and mouse screens, to be further tested in vivo. Bethanechol is a parasympathomimetic agent, similar to acetylcholine, that selectively stimulates muscarinic receptors. M<sub>1-3</sub> muscarinic receptors are abundantly expressed in the lungs and are implicated in immune inflammatory responses by modulating cytokine production (33, 34). While a pro-inflammatory role for the M<sub>3</sub> receptor has been shown in cigarette smoke-induced inflammation, M<sub>1</sub> and M<sub>2</sub> receptors have anti-inflammatory properties (35). Bethanechol is known to activate all three receptors. Alternatively, the flavonoid kaempferol is widely distributed in the plant kingdom and is commonly found in fruits, vegetables, and some beverages, such as tea and coffee (36). Intake of flavonols has been inversely associated with smoking-related cancer risk and kaempferol, specifically, has been shown to have anti-inflammatory and anti-allergic activities (37–39). In addition, kaempferol attenuates Th2-driven experimental allergic airway inflammation by suppressing IL-5 and IL-13 production (17, 40). However, to the best of our knowledge, bethanechol and kaempferol have not been studied within the context of cigarette smoke-induced inflammation. With this in mind, we assessed the effects of the prophylactic administration of both compounds in a mouse model of cigarette smoke exposure. We observed that treatment with kaempferol, and to a more modest extent, bethanechol, attenuated cigarette smoke-induced neutrophilia in mice. However, the clinical relevance of the prophylactic intervention protocols is limited given that, clinically, most patients would start therapies after years of smoking. To this end, we investigated whether kaempferol attenuates established cigarette smoke-induced inflammatory processes. We found that kaempferol attenuated cigarette smoke-induced neutrophilia when administered in a therapeutic fashion. Overall, these data suggest that the flavonoid kaempferol may have significant potential in reducing cigarette smoke-
induced neutrophil recruitment to the lungs, a process that is widely considered to contribute to
the development of emphysematous lung destruction.

Kaempferol possesses a wide range of biological activities that may contribute to the
beneficial effects on cigarette smoke-induced inflammation. Of interest, antioxidant properties of
flavonoids are well known(41) and the ability of kaempferol to reduce reactive oxygen species
may therefore play a protective role in cigarette smoke-induced inflammation. Kaempferol also
has direct anti-inflammatory activities(42). Evidence suggest that it inhibits nuclear factor (NF)-
κB and MAPK activities(18, 43), and the expression of tumour necrosis factor (TNF)-α, IL-1β(44) and IL-8 (45). Lung inflammation induced by subchronic cigarette smoke exposure leads
to NF-κB activation(46); hence, kaempferol may attenuate cigarette smoke-induced inflammation
by targeting this pathway. In addition, kaempferol is known to inhibit fMLP-induced neutrophil
chemotaxis in a concentration-dependent manner(47). It is also plausible that kaempferol
decreased the release of chemokines, such as CXCL5, by dampening the epithelial and
endothelial response to cigarette smoke. In favour of this interpretation, kaempferol attenuated
CXCL5 expression, a chemokine mainly released by epithelial cells, in the therapeutic
intervention, while IL-1α release was not affected. Notably, previous work in our laboratory has
demonstrated that neutrophil recruitment to the pulmonary environment is driven by IL-1α
production by alveolar macrophages, which in turn stimulates CXCL5 release from local
epithelial cells(31). Given that kaempferol neither affected mononuclear cell recruitment nor IL-
1α production in this study, we propose that the compound inhibits this axis at the epithelial
stage. Of note, Tsou et al. proposed that kaempferol along with two other drugs prevent COPD
occurrence and disease progression among traditional Chinese medicine(48) through IKK2
inhibition. These results complement our findings and suggest that kaempferol treatment could
have beneficial effects in smokers and patients with COPD.
In addition to bethanechol, flumequine, and naftifine, other candidates among the 14 drugs identified by cMap analyses in humans and mice may be of interest. For instance, bumetanide is a strong loop diuretic that is used to treat edema associated with heart failure\(^{(49)}\). Econazole is a broad-spectrum antimycotic agent that inhibits the biosynthesis of ergosterol, a constitutive component in yeast and fungal cell membranes. Harpagoside is the major chemical component of the plant devil’s claw, and is known to inhibit NF-κB activation, and suppress lipopolysaccharide-induced inducible nitric oxide synthase (iNOS) and cyclooxygenase (COX)-2 expression\(^{(50)}\). However, given that these compounds were each identified in only one or two of the human cohorts, their use as treatments may be more restricted to certain patient populations. Further research is necessary to determine potential contextual applications of each compound.

In summary, our study took advantage of a bioinformatics technology, Connectivity Mapping, to predict and validate novel therapeutic interventions in a murine model of cigarette smoke-induced neutrophilic inflammation. Overall, our cMap analysis of mouse and human lung tissue yielded 14 drug candidates with the potential to return the gene expression profiles of smoke-exposed lungs to a state similar to room air-exposed lungs. Of these compounds, further experimental data suggest that the flavonoid kaempferol may be an effective therapeutic with which to attenuate cigarette smoke-induced neutrophilia. Consequently, it may also serve to inhibit COPD disease progression and improve disease management. This study provides rationale to investigate kaempferol further with regards to its ability to attenuate cigarette smoke-induced inflammatory processes in patients with COPD.
ACKNOWLEDGMENTS

This work was supported by the Canadian Institutes of Health Research (CIHR) (MOP-64390 and MOP-142353). G.V. is supported by the Canadian Respiratory Research Network (CRRN). Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases. We are thankful to the members of The Merck-Laval-UBC-Groningen Lung eQTL consortium. The authors would like to thank the staff at the Respiratory Health Network Tissue Bank of the FRQS for their valuable assistance with the lung eQTL dataset at Laval University. The lung eQTL study at Laval University was supported by the Fondation de l’Institut Universitaire de Cardiologie et de Pneumologie de Québec, the Respiratory Health Network of the FRQS, and the Canadian Institutes of Health Research (MOP-123369).
REFERENCES


### TABLE 1

**Table 1. Clinical characteristics of subjects.**
Continuous variables are mean ± SD. The number of missing values is shown between parentheses.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Groningen</th>
<th>LAVAL</th>
<th>UBC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never smokers (n=12)</td>
<td>Current smokers (n=45)</td>
<td>Never smokers (n=17)</td>
</tr>
<tr>
<td>Gender m:f (%) male</td>
<td>5:7 (42 %)</td>
<td>23:22 (51 %)</td>
<td>6:21 (22 %)</td>
</tr>
<tr>
<td>Age (y)</td>
<td>48.3 ± 15.4</td>
<td>57.5 ± 8.7</td>
<td>55.8 ± 11.6</td>
</tr>
<tr>
<td>Pack year</td>
<td>0</td>
<td>37.9 ± 17.4</td>
<td>0</td>
</tr>
<tr>
<td>FEV1% predicted</td>
<td>98.6 ± 9.38 (7)</td>
<td>73.3 ± 19.8 (17)</td>
<td>94.1 ± 13.2 (4)</td>
</tr>
<tr>
<td>FVC% predicted</td>
<td>98.8 ± 7.3 (7)</td>
<td>87.1 ± 16.3 (18)</td>
<td>96.5 ± 12.5 (6)</td>
</tr>
</tbody>
</table>
### TABLE 2

<table>
<thead>
<tr>
<th></th>
<th>Differentially expressed genes</th>
<th>Probe sets used for cMap analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Laval</strong></td>
<td>Up: 3,062 Down: 2,434</td>
<td></td>
</tr>
<tr>
<td><strong>UBC</strong></td>
<td>Up: 241 Down: 99</td>
<td>Up: 90 Down: 4</td>
</tr>
<tr>
<td><strong>Groningen</strong></td>
<td>Up: 1,691 Down: 1,088</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Numbers of genes found to be differentially expressed between smokers and non-smokers among all 3 internal cohorts.
Table 3. Numbers of genes found to be differentially expressed between smokers and non-smokers among all external cohorts. SA: small airways, LA: large airways.
### TABLE 4

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Differentially expressed genes</th>
<th>Probe sets used for cMap analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE33561</td>
<td>Up: 28, Down: 23</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE55127</td>
<td>Up: 299, Down: 91</td>
<td>358</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE52509</td>
<td>Up: 90, Down: 7</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE33512</td>
<td>Up: 84, Down: 26</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE17737</td>
<td>Up: 289, Down: 35</td>
<td>358</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE18344</td>
<td>Up: 308, Down: 155</td>
<td>370</td>
</tr>
</tbody>
</table>

Table 4. Numbers of genes found to be differentially expressed between smoke-exposed and control groups in mouse models.
**TABLE 5**

<table>
<thead>
<tr>
<th>Human datasets</th>
<th>Mouse datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE23546</td>
<td>GSE22047 SA</td>
</tr>
<tr>
<td>GSE47460</td>
<td>GSE22047 LA</td>
</tr>
<tr>
<td>GSE22047</td>
<td>GSE10072</td>
</tr>
<tr>
<td></td>
<td>GSE994</td>
</tr>
<tr>
<td></td>
<td>GSE52509</td>
</tr>
<tr>
<td></td>
<td>GSE55127</td>
</tr>
<tr>
<td></td>
<td>GSE33512</td>
</tr>
<tr>
<td></td>
<td>GSE18344</td>
</tr>
<tr>
<td></td>
<td>GSE17737</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drug</th>
<th>Human</th>
<th>Mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bethanechol</em></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><em>Bumetanide</em></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td><em>Chenodeoxycholic acid</em></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><em>Cloxacillin</em></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td><em>Econazole</em></td>
<td>✓</td>
<td>✓ ✓ ✓</td>
</tr>
<tr>
<td><em>Flumequine</em></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><em>Harpagoside</em></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><em>Indoprofen</em></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><em>Kaempferol</em></td>
<td>✓ ✓ ✓</td>
<td>✓</td>
</tr>
<tr>
<td><em>Naftifine</em></td>
<td>✓ ✓ ✓</td>
<td>✓</td>
</tr>
<tr>
<td><em>Pentoxifylline</em></td>
<td>✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td><em>Sulfamonomethoxine</em></td>
<td>✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td><em>Timolol</em></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><em>Trihexyphenidyl</em></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Table 5. Drug candidates conserved in humans and mice predicted by the Connectivity Map analysis.** Human internal cohorts (GSE23546) were used as a reference set. SA: small airways, LA: large airways.
FIGURE LEGENDS

Figure 1. Number of differentially expressed genes from the three internal human datasets.
Venn diagrams of the differentially expressed genes. Panel A and B show up-regulated and
down-regulated genes, respectively. Panel C lists proposed targets based on th cMap analysis of
the internal cohort.

Figure 2. Prophylactic and therapeutic administrations of kaempferol but not bethanechol
reduces cigarette smoke-induced inflammation. Prophylactic intervention (A-C), BALB/c
mice were exposed to cigarette smoke (CS) for 4 days and treated with either bethanechol (Beth)
or kaempferol (Kaem). Both drugs were administered intranasally in 35µL of PBS. Sham mice
received vehicle only. All intranasal instillations were performed daily 1h before the first
cigarette smoke exposure. Therapeutic intervention (D-F). BALB/c mice were exposed to room
air or cigarette smoke for 5 weeks. Kaempferol or vehicle was delivered intranasally daily during
the fifth week of CS-exposure protocol. Bronchoalveolar lavage (BAL) was performed and the
number of total cells (A, D), mononuclear cells (B, E), and neutrophils (C, F) were determined.
Beth: bethanechol, Kaem: kaempferol. Data are represented as mean±SEM. N=5 per group.

Figure 3. Interleukin-1α and CXCL5 levels are reduced by kaempferol but not bethanechol.
Prophylactic (A-B) and therapeutic (C-D) interventions. IL-1α (A, C) and CXCL5 (B, D) levels
were determined by ELISA in the BAL. Beth: bethanechol, Kaem: kaempferol; IL: interleukin.
Data are represented as mean±SEM. N=5 per group.
FIGURE 1

A

GRNG

Laval

669

928

1950

4

90

94

53

UBC

3046

B

GRNG

Laval

834

250

2125

0

4

55

40

UBC

3526

C

List of proposed targets based on the cMap analysis of the internal cohorts

<table>
<thead>
<tr>
<th>Bambuterol</th>
<th>Fosfosal</th>
<th>Pentetrazol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bethanechol</td>
<td>Furazolidone</td>
<td>Pentoxifylline</td>
</tr>
<tr>
<td>Breytium tosilate</td>
<td>Furosultiamine</td>
<td>Pepstatin</td>
</tr>
<tr>
<td>Bumetanide</td>
<td>Gemfibrozil</td>
<td>Racecadotril</td>
</tr>
<tr>
<td>Bupivacaine</td>
<td>Harpagoside</td>
<td>Sulfabenzamide</td>
</tr>
<tr>
<td>Chenodeoxycholic acid</td>
<td>Indoprofen</td>
<td>Sulfamonemethoxine</td>
</tr>
<tr>
<td>Cicloheximide</td>
<td>Kaempferol</td>
<td>Tetracycline</td>
</tr>
<tr>
<td>Cloxacillin</td>
<td>Methocarbamol</td>
<td>Timolol</td>
</tr>
<tr>
<td>Colistin</td>
<td>Methoxamine</td>
<td>Trihexyphenidyl</td>
</tr>
<tr>
<td>Dextronanolol</td>
<td>Mometasone</td>
<td>Triprolidine</td>
</tr>
<tr>
<td>Diclofenamide</td>
<td>Naftifine</td>
<td>Vinblastine</td>
</tr>
<tr>
<td>Econazole</td>
<td>Naphazoline</td>
<td>Viomycin</td>
</tr>
<tr>
<td>Flumequine</td>
<td>Netilmicin</td>
<td>Xamoterol</td>
</tr>
</tbody>
</table>
FIGURE 2

A

BAL total cells/mL (x10^6)

CS+Veh  CS+Beth  CS+Kaem

B

BAL mononuclear cells/mL (x10^6)

CS+Veh  CS+Beth  CS+Kaem

C

BAL neutrophils cells/mL (x10^6)

CS+Veh  CS+Beth  CS+Kaem

D

BAL total cells/mL (x10^6)

RA+Veh  RA+Kaem  CS+Veh  CS+Kaem

E

BAL mononuclear cells/mL (x10^6)

RA+Veh  RA+Kaem  CS+Veh  CS+Kaem

F

BAL neutrophils cells/mL (x10^6)

RA+Veh  RA+Kaem  CS+Veh  CS+Kaem

Copyright © 2017 by the American Thoracic Society
FIGURE 3

A

B

C

D

Copyright © 2017 by the American Thoracic Society
Supplemental Materials and Methods

Human microarrays

Internal cohorts (GSE23546)

Methods for the primary collection of three in house human datasets have been described in detail previously (1). Briefly, non-neoplastic lung specimens were collected from patients undergoing lung surgery at three different sites: Institut Universitaire de Cardiologie et de Pneumologie de Québec (IUCPQ), Laval University, Canada; University of British Columbia, Vancouver, Canada; and University of Groningen, Groningen, The Netherlands; henceforth referred to Laval, UBC and Groningen, respectively. Pulmonary parenchyma was harvested and stored at -80°C. Gene expression profiling was carried out using Affymetrix arrays and data were deposited as GSE23546. Due to technical differences between the profiling experiments across the three subsets, the lists of genes differentially expressed with ≥2.0-fold change between smokers and healthy non-smokers were obtained separately for each of the subsets (‘limma’ package, R) and only genes shared by the three lists were used for further cMAP analysis.

External cohorts (GSE47460, GSE22047, GSE10072, GSE994)

We included gene expression profiles of four publicly available human datasets downloaded from the Gene Expression Omnibus (GEO; accession numbers (GSE47460(2), GSE22047(3), GSE10072(4), GSE994(5)). Samples belonging to the GSE47460 dataset (whole lung homogenate) were profiled on two Agilent platforms: Whole Human Genome Microarray 4x44K, and SurePrint G3 Human GE 8x60K Microarray. We selected profiles obtained from the latter platform for further analysis. The profiles were pre-processed as described (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47460). Samples in the GSE22047
dataset were collected from small and large airway epithelium and were profiled on Affymetrix HG-U133 plus 2.0 arrays. Arrays were normalized with frozen Robust Multi-array Analysis (fRMA), a procedure that allows the pre-processing of microarrays individually or in small batches (6). Due to the functional differences between the small and the large airway epithelium, all further analyses were performed on samples separated based on their source. GSE994 (brushings from intra-pulmonary airways) and GSE10072 (non-tumour tissue samples) datasets were profiled on Affymetrix HG-U133A arrays and normalized with fRMA procedure as well.

Mouse microarray data

Mouse lung gene expression profiles from seven publicly available datasets were used in the current study. These profiles were deposited in GEO (accession numbers GSE55127, GSE33512, GSE18344, GSE8790, GSE17737, GSE52509 and GSE33561) and comprised a total of 54 samples. These samples were used to compile a collection, as described in detail previously (7). Specifically, GSE33512, GSE55127, GSE33561 and GSE52509 datasets were profiled on Agilent and Illumina arrays and were pre-processed as described in their source publications(1, 8–12). Since GSE18344, GSE17737 and GSE8790 datasets contained samples profiled on Affymetrix Mouse Genome 430 2.0 arrays, arrays were normalized with fRMA. To remove technical variation between the different datasets in the collection, we used the Distance-Weighted Discrimination (DWD) method(13). Since DWD method performs pair-wise combinations of datasets, we first combined datasets for each platform, and subsequently combined groups into a large collection. For each model, we performed comparisons of the respective smoke-exposed mice with a pooled dataset of control mice from all experiments.
Connectivity Mapping (cMap) analysis

Signatures obtained from differential expression analyses of human and mouse datasets were used to query the cMap database. More specifically, each signature, containing a list of up- and a list of down-regulated genes was used to query the database at https://portals.broadinstitute.org/cmap/. Such cMap queries require gene signatures to be represented by probe sets found on Affymetrix HG-U133A arrays. However, not all human datasets were originally profiled on this type of array. Therefore, translation of signatures obtained from other arrays was performed based on HUGO gene symbols. Profiles from mouse datasets were translated based on MGI and HUGO gene symbols. cMap query with each of the translated signatures yielded a list of compounds shown to be correlated or anti-correlated with the signature used for the query. From each resulted list, significant compounds (p<0.05) with the most strongly anti-correlated signature (i.e. negative connectivity) were selected.

Cigarette smoke exposure

Mice were housed under specific pathogen-free conditions with *ad libitum* access to food and water and subjected to a light-dark cycle of 12h. Mice were exposed to 12 3R4F reference cigarettes (Tobacco and Health Research Institute, University of Kentucky, Lexington, KY) with filters removed, for 50 min, twice daily, for 5 days/week in a whole-body smoke exposure system (SIU-48, Promech, Sweden). Total particulate matter ranged from 600 to 900µg/L. We have previously shown that cotinine and carboxyhaemoglobin levels are similar to those observed in human smokers(4).

Sample processing

Lungs were removed from the thoracic cavity and the trachea cannulated with a polyethylene tube. Subsequently, the right lobes were tied off, and bronchoalveolar lavage (BAL) was
performed by instilling the lungs sequentially with 250µL and 200µL of ice-cold PBS. Total cell number in the BAL was determined using a haemocytometer. Cytospins were prepared and stained with Hema 3 (Biochemical Sciences, Swedesboro, NJ) for differential cell counts. 300 cells were counted per cytospin. Cells were pelleted and the BAL fluid (BALF) kept at -80°C for cytokine measurements.
References


Supplemental Table Legends

**Supplemental Table 1.** Genes significantly regulated between non-smokers and current smokers shared among all three internal cohorts.

**Supplemental Table 2.** List of predicted drug candidates for all the human cohorts and mouse datasets.

**Supplemental Table 3.** Genes significantly regulated between non-smokers and current smokers shared among all five external cohorts.

**Supplemental Table 4.** Genes significantly regulated between smoke-exposed and control groups shared among all six mouse datasets.