Translating Airway Gene Expression into Biomarkers for Tobacco Smoke Exposure and Lung Cancer Detection

Avrum Spira, M.D., M.Sc.
Division of Computational Biomedicine
The Pulmonary Center
Boston University Medical Center
Disclosure

• Founder and consultant to Allegro Diagnostics Inc.
Outline

• Bronchial airway gene-expression as a biomarker for the *early diagnosis* of lung cancer

• Bronchial Airway gene-expression in the *screening* and *chemoprevention* setting
  – Leveraging transcriptomics to identify new therapeutic opportunities (in silico drug repositioning)

• Extending “field of injury” to *microRNA*

• Moving to the *mouth* and nasal epithelium
  – Biomarkers for *second hand exposure*
  – Biomarker of *response to quitting and PREP*
  – Biomarkers of *other inhaled toxic exposures* (air pollution)
The Airway “Field of Injury” Hypothesis

Smoking alters epithelial cell gene expression throughout the respiratory tract - biomarker of the physiological response to smoking

Variability in epithelial cell genomic response to and damage from smoking linked to tobacco-associated lung cancer
The bronchial airway transcriptome in smoking and lung cancer

Smoking impacts airway gene expression
- PNAS 2004; NAR 2005;

Subset of changes are irreversible upon cessation and can serve as biomarker of past exposure
- Genome Biology 2007;

Airway gene expression can serve as an early diagnostic biomarker for lung cancer

RNA- U133A Affymetrix array (~22,500 genes)
Airway gene-expression as a diagnostic biomarker for lung cancer

1. Bronchoscopy performed as initial diagnostic test
   Brush cytologically-normal airway epithelial cells in mainstem bronchus

2. Extract & process RNA using microarray

3. Cytology obtained has low sensitivity especially for early stage disease. This results in clinical dilemma as to who should go for surgical resection.

4. 80 gene biomarker that can distinguish smokers with and without lung cancer
   - Sensitivity 80%; specificity 84% in two independent cohorts (n=87)
   - 90% sensitive for stage 1 disease
   - 95% sensitivity and 95% NPV when combined with cytology obtained at bronchoscopy
   - Independent of clinical and radiographic predictors of disease


Validation study of gene-expression biomarker on independent multicenter cohort by Allegro Diagnostics Inc.
   - 2100 subjects recruited at 21 centers in US, Canada and Europe for both CLIA and FDA trial
   - IDE filed and approved by FDA
   - CLIA trial results presented at 2012 ACCP mtg

Nature Medicine 2007; CAPR 2008
Outline

• Bronchial airway gene-expression as a biomarker for the early diagnosis of lung cancer

• Bronchial Airway gene-expression in the screening and chemoprevention setting
  – Leveraging transcriptomics to identify new therapeutic opportunities (in silico drug repositioning)

• Extending “field of injury” to microRNA

• Moving to the mouth and nasal epithelium
  – Biomarkers for second hand exposure
  – Biomarker of response to quitting and PREP
  – Biomarkers of other inhaled toxic exposures (air pollution)
A pathway-based approach to airway gene-expression

Adenovirus overexpressing pathway mediator (e.g. PI3K)

HG-U133A microarray (in replicate)  HG-U133A microarray (in replicate)

PI3K Activation Probability Calculated For Each Sample (0-1)

Pathway Signature

Histologically Normal Airway Epithelial Cells

 HG-U133A Array
Lung cancer: 60 No Cancer: 69

Bild et al. Nature 2006
Airway gene expression is altered in high-risk smokers with dysplasia and is reversible with chemoprevention.

Cytologically normal airway epithelium from smokers without dysplasia (n=10) vs. smokers with mild-moderate dysplasia (n=14).

Gene Expression profiling reveals increased activation of PI3K in airway of smokers with dysplasia.

Activity of PI3K gene-expression pathway is significantly reduced post-treatment with myo-inositol in those smokers who had regression of their dysplastic lesions.

Extending this paradigm to other lung cancer chemoprevention studies

Current and ex-smokers with bronchial dysplasia (n=27)

Green Tea extract

2/3 had regression of lesion; 1/3 had persistence or progression of dysplasia regardless of treatment arm

Brushing pre-treatment

Biopsy and brushing at 3 months

Biopsy and brushing at 6 months

Placebo
Airway gene expression is associated with progression of premalignant lesions (independent of treatment)

- Persistent/Progression  Partial/complete regression

- Enrichment of this gene list among genes that change in airway of smokers with lung cancer and in lung cancer tissue itself (TCGA).

Biomarker for stratifying smokers with dysplasia into chemoprevention trials?
The Detection of Early Lung Cancer Among Military Personnel (DECAMP) Consortium

External Advisory Board  Grant Officer

Clinical Consortium Steering Committee  Internal Advisory Board

Coordinating Center
- PI: Spira; co-PI: Schnall
  - M.D. Anderson Pathology Core
    - PI: Wistuba
  - Boston University
    - Biorepository Core
    - Bioinformatics Core/Molecular Database
  - ACRIN
    - Imaging Core Lab
    - Data Mgt Ctr
    - Protocol Mgt
    - Regulatory
  - Brown University/UCLA
    - Biostatistics Core
    - PI: Gastonis/Elashoff

Clinical Sites*
- Military Hospitals (4)
  - Naval Medical Ctr Portsmouth, VA
  - Walter Reed National Military Ctr, MD
  - Naval Medical Ctr San Diego, CA
  - San Antonio Military Ctr, TX
- VA Hospitals
  - VA Boston Healthcare System, MA
  - Dallas VA Medical Ctr, TX
  - Denver VA Medical Ctr, CO
  - Greater Los Angeles VA Healthcare System, CA
  - VA Tennessee Valley Healthcare System, TN
  - Philadelphia VA Medical Ctr, PA
  - VA Pittsburgh Healthcare System, PA
- Academic Hospital
  - Roswell Park Cancer Inst., NY

BU Administrative Core
- PI: Massion
  - Vanderbilt Proteomics Core (Blood/Airway)
    - PI: Dubinett
- BU Molecular Core (Airway)
  - PI: Lenburg
- UCLA
  - Molecular Core (Serum Cytokines)
    - PI: Massion

Project 1: Nodules
- Disease measurable
- Non-invasive diagnosis of lung nodules
- Screening of high risk individuals

Project 2: Risk Assessment
- Disease non-measurable
- Detection of recurrence

Endobronchial Biopsies Proteomics
Endobronchial Brushings Proteomics & Gene Expression
Serum Proteomics (MALDI) & Cytokine Profile
Nasal Epithelium Gene Expression
Project 1

Aim 1: Establish Cohort with Indeterminate Pulmonary Nodules

- **Cohort Inclusion Criteria** (n=500)
  - Age: >50 yrs
  - Smoking status: Current or former (>30 PKY)
  - Indeterminate pulmonary nodules: Size: 0.5-2.0cm

**Data & Sample Collection (year of follow-up)**

- **Clinical Data**
- Imaging
- Biosamples:
  - Blood, Urine, Sputum,
  - Nasal & Buccal Brushings,
  - Bronchoscopy*:
    - Brushings & Biopsies

**Pathology Core**

- **Coordinating Center**
  - Clinical Data Biorepository

- **Clinical Diagnosis**
  - Matched:
    - Cancers (n ~ 75)
    - Non-cancers (n ~ 75)
  - Biosamples used for biomarkers in Aim2
  - Other non-cancers (n=350)
    - Biosamples banked in biorepository

Project 2

Aim 1: Establish Longitudinal Cohort for Identifying Incident Lung Cancer

- **Cohort Inclusion Criteria** (Based on 10 yr Bach Risk Model) (n=1000)
  - Age: 50-79 yrs
  - Smoking status: Current (≥ 25 yrs; 10 cig/day)
    - Former (≥ 20PKY; quit <20 yrs ago)
  - COPD or 1st degree family member with lung cancer

**Longitudinal Data & Sample Collection (years of follow-up)**

- **Baseline**
- **Year 1**
- **Year 2**
- **Year 3**
- **Year 4**

- **Clinical Data**
- CT Scan
- Imaging

- Biosamples:
  - Blood, Urine, Sputum,
  - Nasal & Buccal Brushings,
  - Bronchoscopy*:
    - Brushings & Biopsies

**Clinical Diagnosis**

- Matched:
  - Cancers (n ~ 50)
  - Non-cancers (n ~ 50)
- Biosamples used for biomarkers in Aim2 & 3
- Other non-cancers (n=900)
  - Biosamples banked in biorepository
Development of smoking-related pathway signatures in airway epithelial cells

Leveraging Gene-expression to discover new therapeutic opportunities via the Connectivity Map

**Gene Expression Study**
Signature for exposure to carcinogen, premalignant tissue, or tumor tissue
- List of induced genes
- List of repressed genes

**Connectivity Map**
Cancer cell cultures are treated with multiple compounds
- Compound 1
- Compound 2
- Compound N

Gene expression is profiled for each compound
- Compound 1
- Compound 2
- Compound N

- No Significant Pattern
- Anti-correlated pattern to carcinogen/tumor signature
- No Significant Pattern

Spira et al. Cancer Prev Research 2010
Using the connectivity map to uncover novel treatment for basal-subtype of breast cancer

Breast cancer

ER - ER +

Predicting drug sensitivities from microarray datasets

Cohen et al. Molecular Systems Biology 2011
Developing a Genomic Model of Emphysema Progression using Regional Heterogeneity

8 patients getting single or double lung transplants for severe COPD.

- Cores removed from lung slices – 8 clusters of 4 cores

1. **Micro-CT examination**
   - measure Lm to quantify degree of emphysema

2. **Isolation of mRNA for microarray**
   - Affymetrix Human Exon 1.0 ST array

3. **Isolation of microRNA for microarray**
   - Invitrogen nCODE array (~700 human microRNA)

4. **Isolation of DNA for methylation array**
   - Illumina Golden Gate assays

- before
- after
- High resolution CT
Identification of compound (GHK) that reverses emphysema gene expression signature

- Low LM (no emphysema)
- High LM (severe emphysema)

Integrin important for fibroblast migration and adhesion

Campbell et al. Genome Medicine 2012
GHK and TGFβb restore collagen contraction by lung fibroblasts from smokers with COPD

GHK restores ability of fibroblasts (green) to remodel collagen into fibrillar collagen (purple)

Campbell et al. Genome Medicine 2012
Expansion of cMAP via the LINCS program

- 4,000 small-molecule compounds in **20 different cell types**
- 3,000 human genes perturbed using lentivirally-delivered shRNAs or overexpression in the same set of 20 cell lines
- 978 genes measured on luminex based platform
  - Dollars per sample
  - Can be used to extrapolate all genes on Affy array
- The cell lines will be selected based on their lineage diversity, and will span established cancer cell lines, immortalized (but not transformed) primary cells, and both cycling and quiescent cells

http://www.broadinstitute.org/LINCS/
Outline

• Bronchial airway gene-expression as a biomarker for the **early diagnosis** of lung cancer

• Bronchial Airway gene-expression in the **screening** and **chemoprevention** setting
  – Leveraging transcriptomics to identify new therapeutic opportunities (in silico drug repositioning)

• Extending “field of injury” to **microRNA**

• Moving to the **mouth** and **nasal epithelium**
  – Biomarkers for **second hand exposure**
  – Biomarker of **response to quitting and PREP**
  – Biomarkers of **other inhaled toxic exposures** (air pollution)
Impact of smoking on bronchial airway microRNA expression

MicroRNA regulate part of the gene-expression response to smoking

Schembri et al PNAS. 2009
Microarray vs. RNA-seq

**Microarrays**
- Limited by prior knowledge on what is expressed
- Analog output with limited dynamic range
- Cost $$
- Throughput moderate
- Computation ++

**RNA-seq**
- Pure discovery**
- Digital output with large dynamic range
- Alt splicing
- SNP in exons
- Cost $$$$$$
- Throughput low
- Computation ++++
Deep sequencing the airway transcriptome

- Large RNA sequenced on Illumina Solexa system
  - 36 base pair reads
  - 30 million reads per sample

- Small RNA (15-40 bp) sequenced on ABI SOLID platform
  - 30 base pair sequence
  - 200 million reads per run (4 samples multiplexed in single run)

4 pools: Never smoker, current smoker, smoker with cancer, smoker with benign lung disease
mRNA-seq identifies novel smoking- and cancer-related gene expression changes in the airway

Beane et al. 2011
Discovery of novel airway microRNA associated with lung cancer

Perdomo et al. Submitted
Novel miRNA is expressed almost exclusively in the respiratory tract and localizes to airway epithelium.
Novel miRNA is expressed during airway epithelial cell differentiation

Bronchial epithelium

Cells are raised to the ALI (Day 0)

Cells are differentiated (Day 20)
Overexpression of novel miRNA results in more differentiated ciliated cells in ALI

Day 9
Novel microRNA is downregulated in lung cancer and in the airway of smokers with lung cancer.
Overexpression of novel miRNA can inhibit anchorage independent tumor cell growth

Soft Agar Assay

In collaboration with Carmen Tellez and Steve Belinsky
Aim 1
Develop biomarkers for diagnosis of lung cancer

Smokers with suspect lung cancer (indeterminate nodules)
Allegro Diagnostics: n = 150 (75 cancer)

Aim 1a
Bronchial airway epithelium
BU
microRNA
High-throughput sequencing

Aim 1b
Nasal and buccal airway epithelium
BU
microRNA
High-throughput sequencing

Aim 1c
Peripheral blood
UCLA
Proteins
Multiplex immunoassay

Use results to select additional targets

Develop non-invasive biomarkers for diagnosis of lung cancer

Figure 1: Overview of Aim 1
Developing a microRNA-based airway biomarker for diagnosis of lung cancer in the AllegroDx trial

<table>
<thead>
<tr>
<th></th>
<th>No Cancer (n=53)</th>
<th>Cancer (n=75)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Smoking</strong></td>
<td>19 Current, 34 Former</td>
<td>27 Current, 48 Former</td>
<td>1</td>
</tr>
<tr>
<td><strong>Pack years</strong></td>
<td>39.1 +/- 33.7</td>
<td>45.4 +/- 34.9</td>
<td>0.3609</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>18 Female, 35 Male</td>
<td>30 Female, 45 Male</td>
<td>0.579</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>56.6 +/- 12.8</td>
<td>67.4 +/- 11.6</td>
<td>4.2e-06</td>
</tr>
<tr>
<td><strong>RIN</strong></td>
<td>5.9 +/- 1.4</td>
<td>6.0 +/- 1.6</td>
<td>0.5671</td>
</tr>
</tbody>
</table>
Outline

• Bronchial airway gene-expression as a biomarker for the **early diagnosis** of lung cancer

• Bronchial Airway gene-expression in the **screening** and **chemoprevention** setting
  – Leveraging transcriptomics to identify new therapeutic opportunities (in silico drug repositioning)

• Extending “field of injury” to **microRNA**

• Moving to the **mouth and nasal epithelium**
  – Biomarkers for **second hand exposure**
  – Biomarker of **response to quitting and PREP**
  – Biomarkers of **other inhaled toxic exposures** (air pollution)
Extending the field to the upper airway
Mouth Scraper

FIG. 1

Biotechniques 2004
Technique for Obtaining RNA from Nasal Mucosal Brushings

• Cytosoft® brushings from “interior” of inferior turbinate

• Immerse in RNA later
Nasal gene expression reflects the bronchial airway gene-expression response to smoking

Zhang et al. Physiological Genomics 2010
The nose-bronch relationship
Genes differentially expressed in nose between smokers with and without lung cancer in the AllegroDx trial
Genes associated with cancer in nasal epithelium are similarly up- and down-regulated in bronchial epithelium (GSEA p < 0.001)

Nasal genes **up-regulated** with the presence of cancer

Nasal genes **down-regulated** with the presence of cancer
Upper airway biomarkers developed as part of the GEI

• Nasal and buccal gene-expression as biomarker of ever exposure

• Nasal biomarker of second-hand smoke exposure

• Buccal biomarker of cumulative exposure (i.e. pack-yrs) to tobacco smoke

• Nasal gene-expression signature post smoking cessation

• Indoor air pollution in China (supplement)
Development of a nasal gene-expression biomarker of passive exposure to smoking

<table>
<thead>
<tr>
<th></th>
<th>unexposed</th>
<th>SHS exposed</th>
<th>Active smoker (0-10 CPD)</th>
<th>Active Smoker (10-15 CPD)</th>
<th>Active Smoker (&gt; 15 CPD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>12</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>% female</td>
<td>75%</td>
<td>56%</td>
<td>63%</td>
<td>40%</td>
<td>57%</td>
</tr>
<tr>
<td>Age</td>
<td>24.0</td>
<td>23.9</td>
<td>22.3</td>
<td>25.0</td>
<td>23.3</td>
</tr>
<tr>
<td>CPD</td>
<td>0.0</td>
<td>0.0</td>
<td>5.6</td>
<td>11.2</td>
<td>19.6</td>
</tr>
<tr>
<td>Cotinine</td>
<td>0.03</td>
<td>1.1</td>
<td>84.2</td>
<td>122.5</td>
<td>231.3</td>
</tr>
</tbody>
</table>

A nasal mRNA biomarker
Moving nasal biomarkers of secondhand exposure to children

- Columbia Center for Children’s Environmental Health (CCCEH) cohort

- Disease Investigation Through Specialized Clinically-Oriented Ventures in Environmental Research (DISCOVER) cohort
Pilot study: Columbia Center for Children’s Environmental Health (CCCEH)

<table>
<thead>
<tr>
<th></th>
<th>Exposed (n=11)</th>
<th>Control (n=18)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult / Child</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Adult</td>
<td>Child</td>
<td>Adult</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>1C, 0A</td>
<td>6C, 4A</td>
<td>6C, 0A</td>
</tr>
<tr>
<td>RIN</td>
<td>7.1 (±1.0)</td>
<td>6.9 (±1.2)</td>
<td>0.75</td>
</tr>
<tr>
<td>PM$_{2.5}$ (ug/m$^3$)</td>
<td>25.9 (±24.8)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>SHS UVPM (ug/m$^3$)</td>
<td>2.9 (±2.4)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Air Nicotine (ug/m$^3$)</td>
<td>0.97 (±1.91)</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
CCCEH: Similar gene expression changes associated with exposure status are detected across children and adults.
Upper airway biomarkers developed as part of the GEI

- Nasal and buccal gene-expression as biomarker of ever exposure
- Nasal biomarker of second-hand smoke exposure
- Buccal biomarker of cumulative exposure (i.e. pack-yrs) to tobacco smoke
- Nasal gene-expression signature post smoking cessation
- Indoor air pollution in China (supplement)
Biomarkers of response to smoking cessation

Outline of Study

Patient Recruitment

Active Smokers (n=8) → Smoking-Cessation Program

University of Minnesota Boston University Medical Center

Sample Collection

Blood Carbon monoxide Urine Cotinine → Assess Tobacco Abstinence

Nasal Brushings (n = 33)

Microarray Analysis

RNA Isolation → Affymetrix Human Gene 1.0 ST Array

Computational Analysis

Linear mixed effects model: Expression ~ Time + RIN + random(Patient)

Functional Enrichment → Comparison to Other Datasets

Gene Set Enrichment Analysis
101 Genes that change post-smoking cessation (FDR <0.05)

The expression patterns of these genes indicate that the most changes in gene expression occur between 1 and 2 months of tobacco abstinence.
A strong relationship between nasal epithelial gene expression associated with smoking cessation and cessation-induced changes in cross-sectional bronchial airway gene expression.
Effect of SWITCHING TO PREP on gene expression

Current smoker

Quitter

Baseline  2 weeks  4 weeks  6 weeks  8 weeks
Upper airway biomarkers developed as part of the GEI

- Nasal and buccal gene-expression as biomarker of ever exposure

- Nasal biomarker of second-hand smoke exposure

- Buccal biomarker of cumulative exposure (i.e. pack-yrs) to tobacco smoke

- Nasal gene-expression signature post smoking cessation

- Indoor air pollution in China (supplement)
• Buccal epithelial gene-expression as biomarker of response to indoor air pollution (coal smoke) among Chinese women

Buccal scrapings collected from a cohort of never-smoker women with high rates of lung cancer from Xuan Wei County, China.

In collaboration with Nat Rothman and Qing Lan

Enrichment of these genes among those that change in buccal epithelium of active smokers
Female Residents of Xuan Wei and Personal Filter Metrics for Microarray Pilot Study

<table>
<thead>
<tr>
<th>Category</th>
<th>Low Exposure (n=12)</th>
<th>High Exposure (n=11)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEX</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>RNA Quality</td>
<td>Good</td>
<td>Better</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>PM$_{2.5}$ (µg/m$^3$)</td>
<td>104.77 (±50.34)</td>
<td>283.25 (±131.27)</td>
<td>0.001</td>
</tr>
<tr>
<td>BAP (ng/m$^3$)</td>
<td>15.56 (±6.26)</td>
<td>99.69 (±44.49)</td>
<td>4.054e-05*</td>
</tr>
</tbody>
</table>

- **PM$_{2.5}$**: airborne particulate matter (≤2.5 µm in aerodynamic diameter)
- **BAP**: Benzo[a]pyrene level
Acknowledgements

Boston University

- Dan Brooks
- Marc Lenburg
- Jerome Brody
- Joshua Campbell
- Gang Liu
- Sherry Zhang
- Ji Zhang
- Joe Guerrein
- Adam Gower
- Christina Anderlind
- Catalina Perdomo
- Teresa Wang
- Kahkeshan
- Bozena

UBC: Stephen Lam, Jim Hogg, Don Sin,
Univ of Utah: Andrea Bild
NCI: Eva Szabo, Nat Rothman, Qing Lan
Vanderbilt University: Pierre Massion
UCLA: Steve Dubinett, David Elashoff, Brigitte Gomperts
LRRI: Steve Belinsky
Uminnesota: Dorothy Hatsukami, Stephen Hecht

Funded by NCI/EDRN, NHLBI, NIEHS, DOD