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

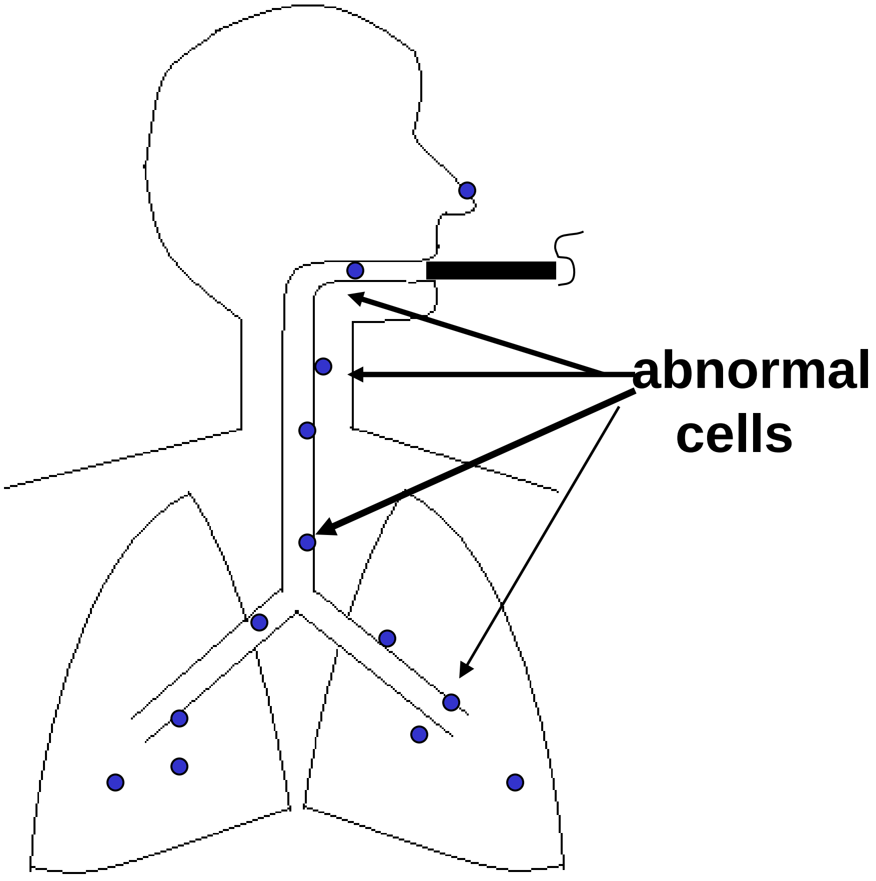
- Moving to the mouth and nasal epithelium
  - Biomarkers for second hand exposure
  - Biomarker of response to quitting and PREP
  - Biomarkers of other inhaled toxic exposures (air pollution)

The Airway “Field of Injury” Hypothesis

Smoking alters epithelial cell gene expression throughout the respiratory tract

-biomarker of the physiological <u>response</u> 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

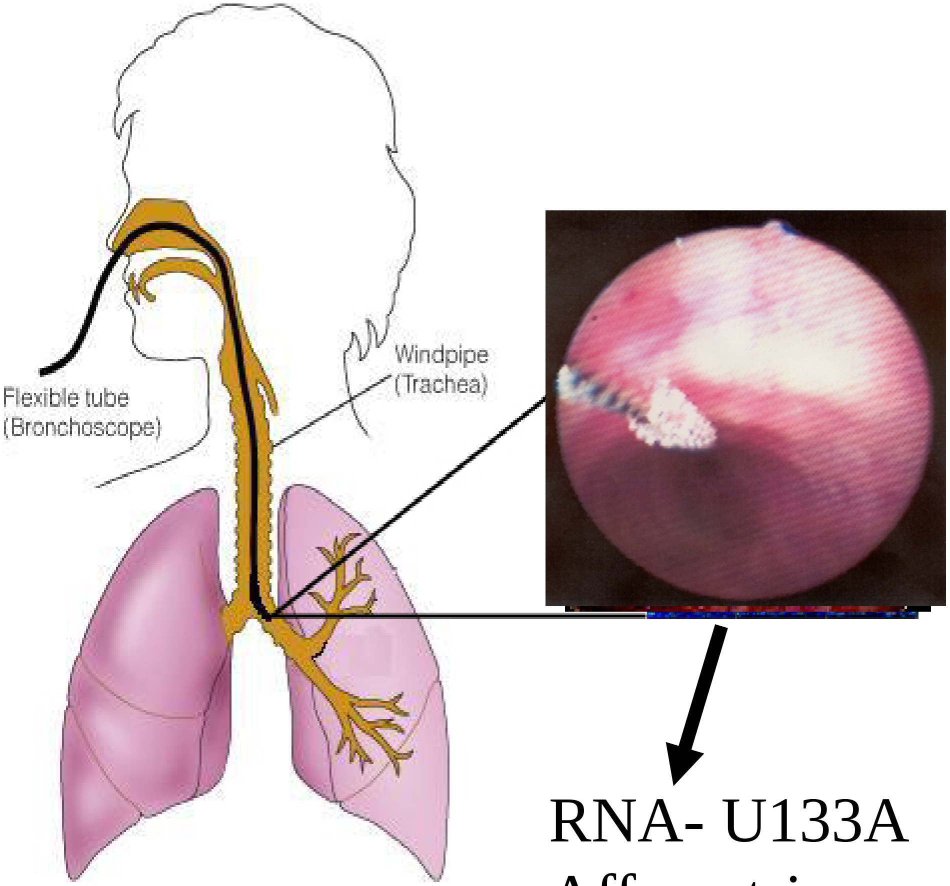
-PNAS 2004; NAR 2005;

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

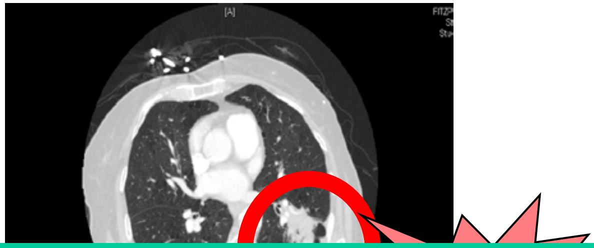
- Nature Medicine 2007; CAPR 2008;
Cancer Research 2009

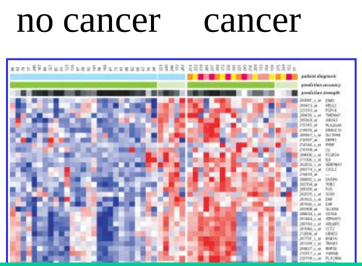


RNA- U133A
Affymetrix array
(~22,500 genes)

Airway gene-expression as a diagnostic biomarker for lung cancer

1.





5.

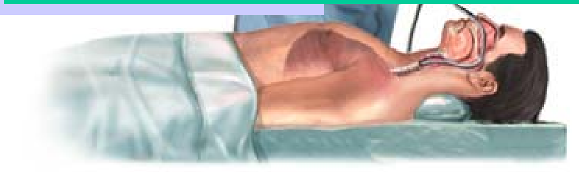
80 gene biomarker that can distinguish smokers with and without lung cancer

-sensitivity 80%; specificity 84% in two independent cohorts (n=87)

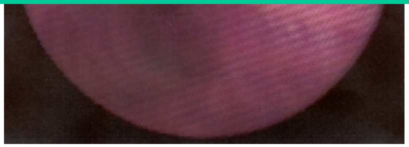
-90% sensitive for stage 1 disease

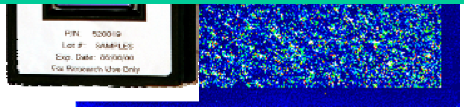
Validation study of gene-expression biomarker on independent multicenter cohort by Allegro Diagnostics Inc.

- -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



A.D.A.M.





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

Human Primary
Epithelial Cell Culture

Adenovirus overexpressing
pathway mediator (e.g PI3K)

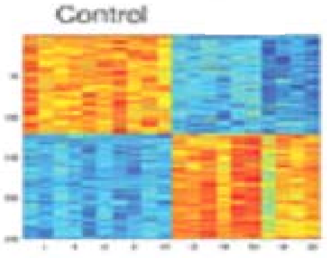
Adenovirus expressing GFP control

HG-U133A microarray
(in replicate)

HG-U133A microarray
(in replicate)

PI3K Off

PI3K On



Pathway Signature

Ras
Myc
E2F3
β-catenin
Src
Np63
PI3K

Histologically Normal
Airway Epithelial Cells

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

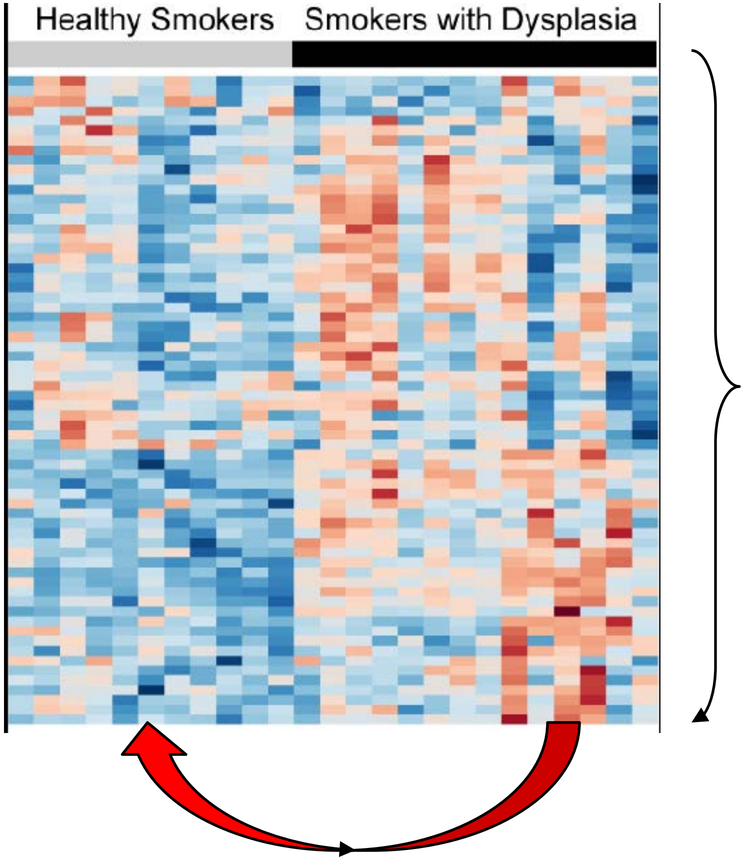
PI3K Activation Probability
Calculated For Each Sample
(0-1)

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



Genes that are highly expressed when PI3K is activated

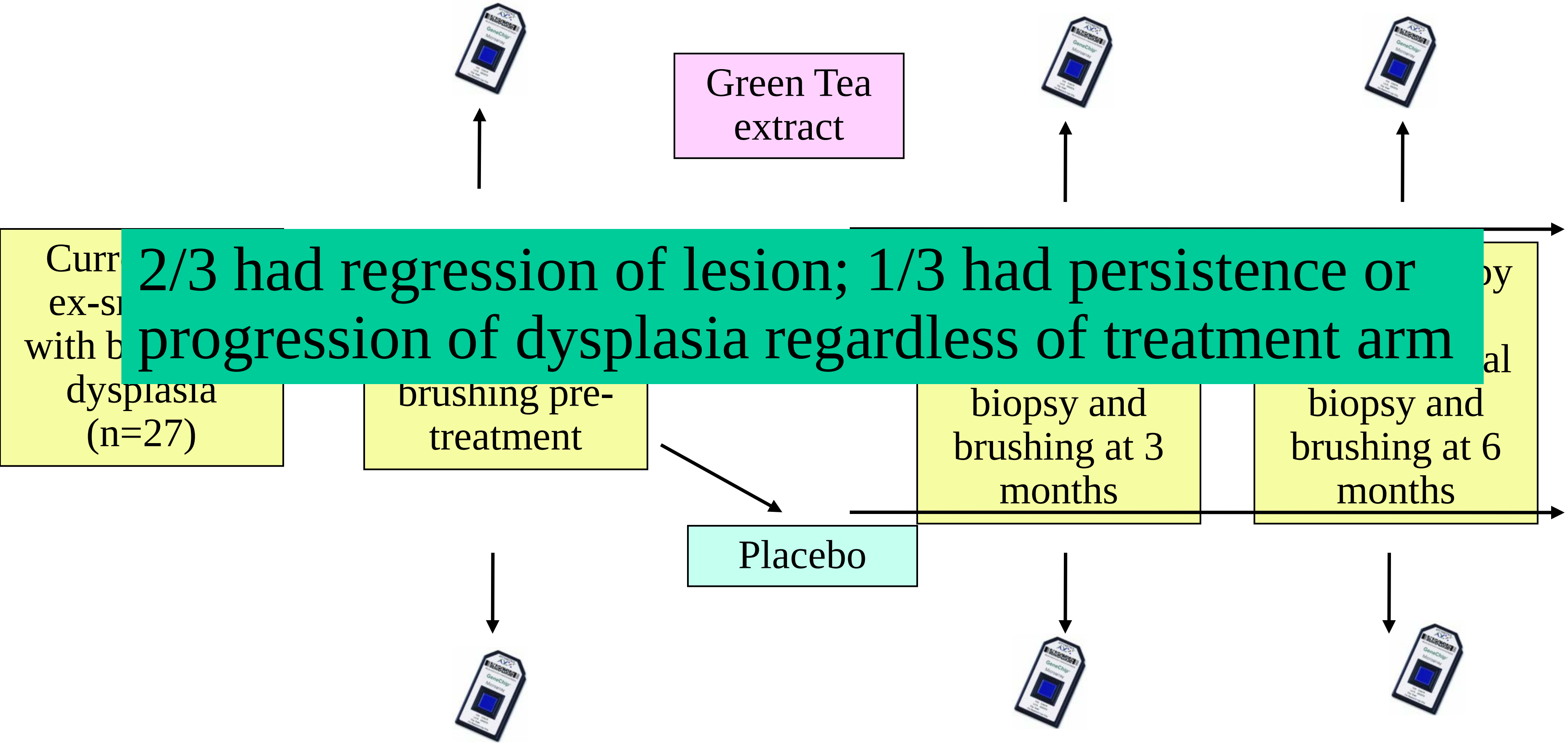
Activity of PI3K gene-expression pathway is significantly reduced post-treatment with myo-inositol in those smokers who had regression of their dysplastic lesions :

Validation in Phase-2 trial at Mayo

y?

Science Translational Medicine. 2010

Extending this paradigm to other lung cancer chemoprevention studies



Curr
ex-s
with b
dysplasia
(n=27)

brushing pre-treatment

Green Tea extract

Placebo

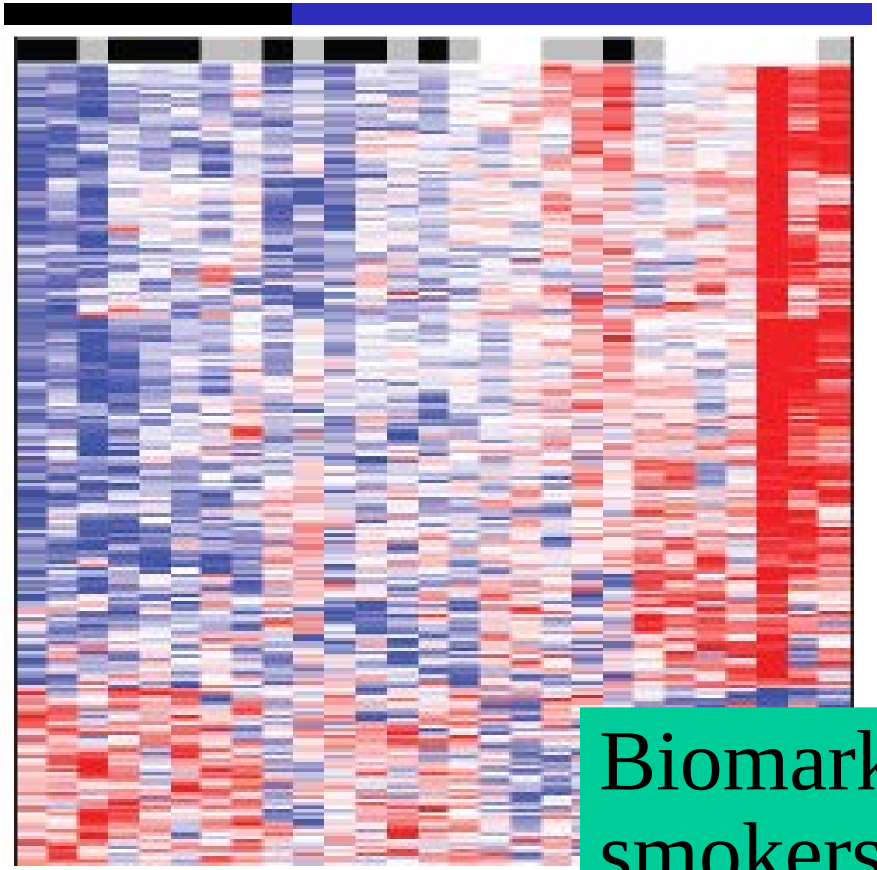
biopsy and brushing at 3 months

py
al
biopsy and brushing at 6 months

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

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

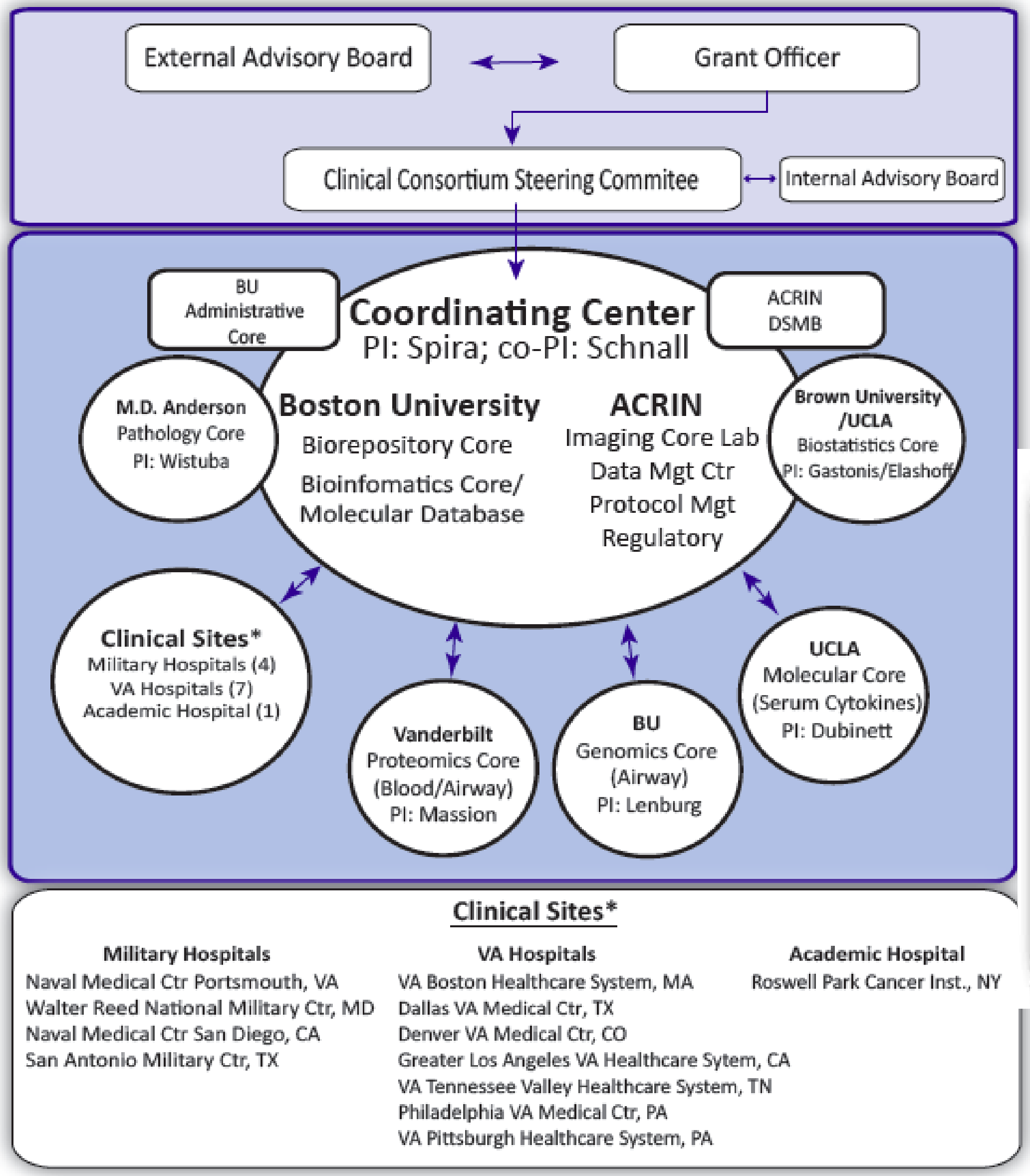
- Persistent/Progression   Partial/complete regression



- 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

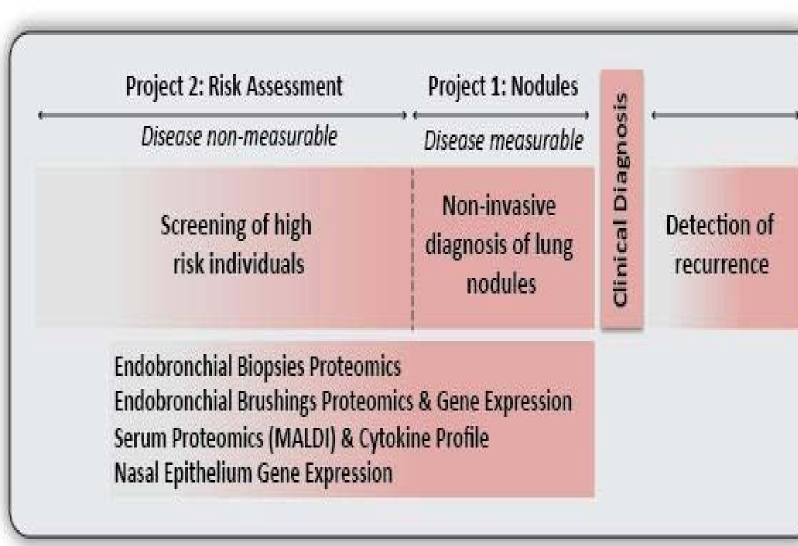


Clinical Sites*

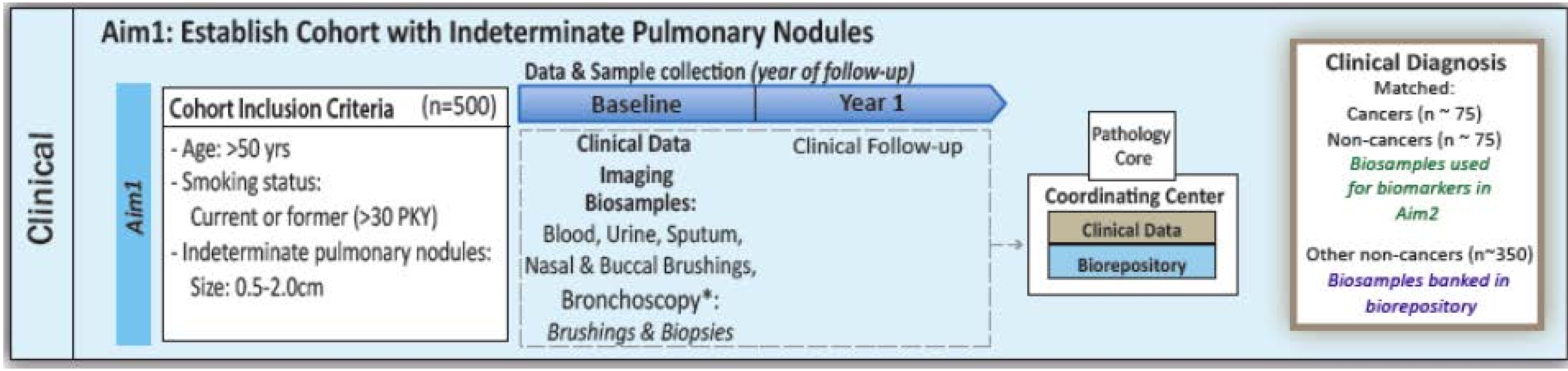
Military Hospitals
Naval Medical Ctr Portsmouth, VA
Walter Reed National Military Ctr, MD
Naval Medical Ctr San Diego, CA
San Antonio Military Ctr, TX

VA Hospitals
VA Boston Healthcare System, MA
Dallas VA Medical Ctr, TX
Denver VA Medical Ctr, CO
Greater Los Angeles VA Healthcare Sytem, CA
VA Tennessee Valley Healthcare System, TN
Philadelphia VA Medical Ctr, PA
VA Pittsburgh Healthcare System, PA

Academic Hospital
Roswell Park Cancer Inst., NY



Project 1



Clinical

Aim1: Establish Cohort with Indeterminate Pulmonary Nodules

Aim1

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)*

Baseline | Year 1

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

Clinical Follow-up

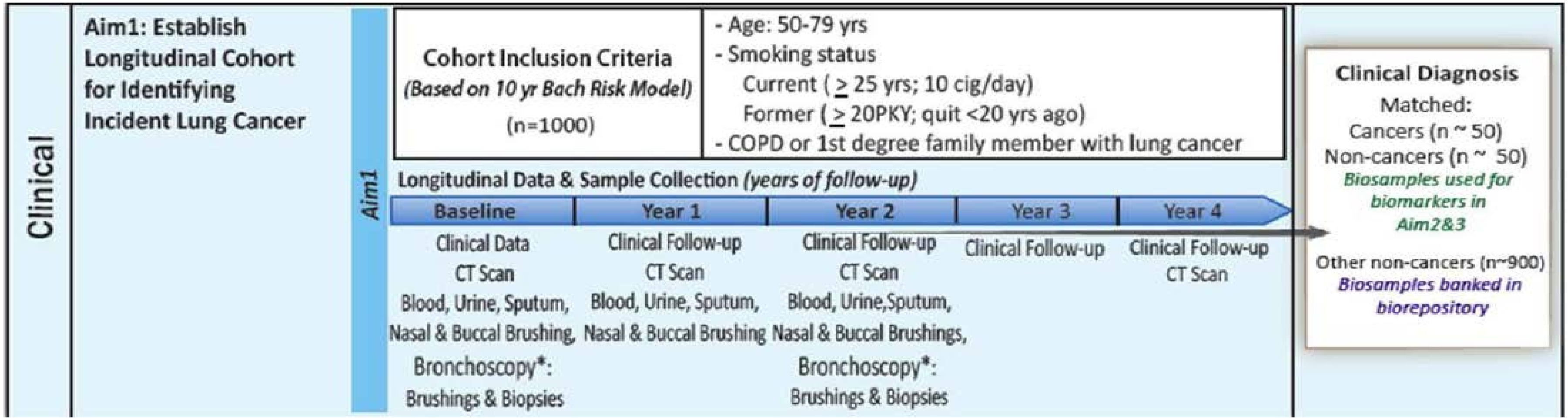
Pathology Core

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



Clinical

Aim1: Establish Longitudinal Cohort for Identifying Incident Lung Cancer

Aim1

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

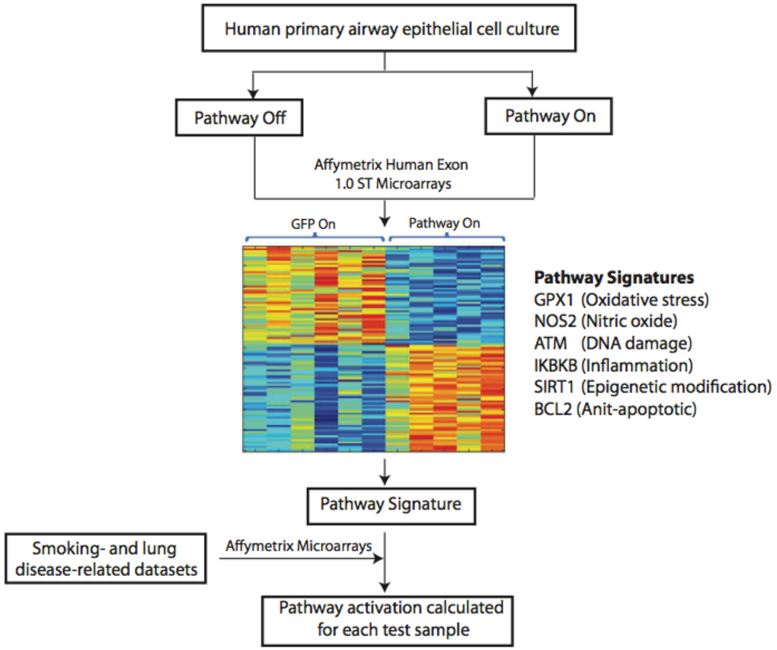
Longitudinal Data & Sample Collection *(years of follow-up)*

Baseline	Year 1	Year 2	Year 3	Year 4
Clinical Data CT Scan Blood, Urine, Sputum, Nasal & Buccal Brushing, Bronchoscopy*: Brushings & Biopsies	Clinical Follow-up CT Scan Blood, Urine, Sputum, Nasal & Buccal Brushing	Clinical Follow-up CT Scan Blood, Urine,Sputum, Nasal & Buccal Brushings, Bronchoscopy*: Brushings & Biopsies	Clinical Follow-up	Clinical Follow-up CT Scan

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



Human primary airway epithelial cell culture

Pathway Off

Pathway On

Affymetrix Human Exon 1.0 ST Microarrays

GFP On

Pathway On

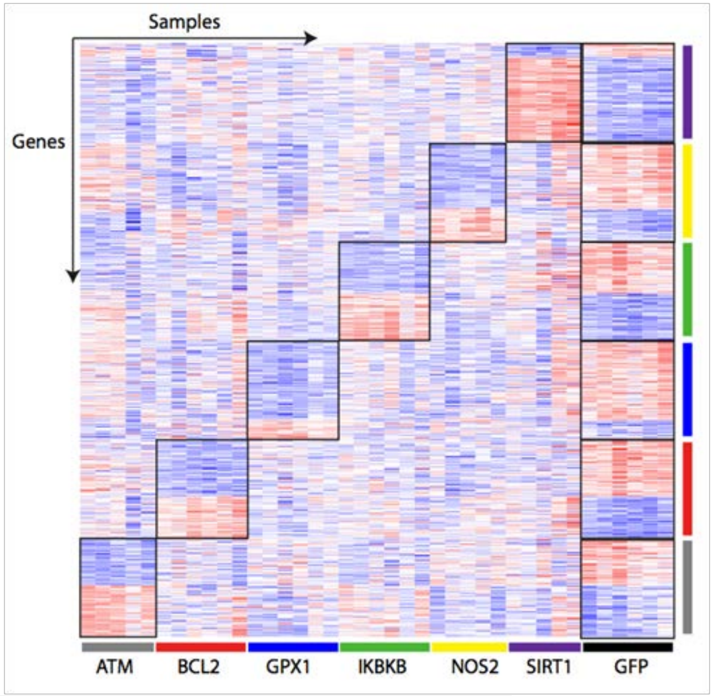
Pathway Signatures
GPX1 (Oxidative stress)
NOS2 (Nitric oxide)
ATM (DNA damage)
IKBKB (Inflammation)
SIRT1 (Epigenetic modification)
BCL2 (Anit-apoptotic)

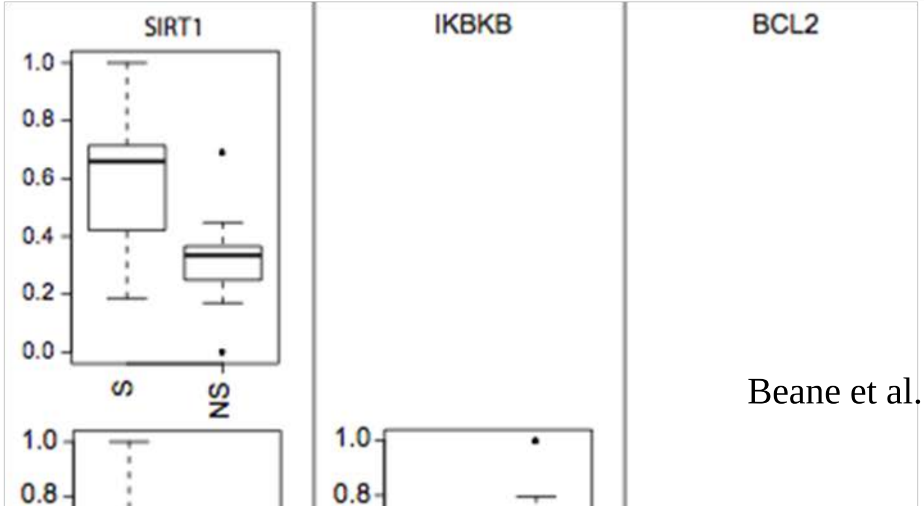
Pathway Signature

Smoking- and lung disease-related datasets

Affymetrix Microarrays

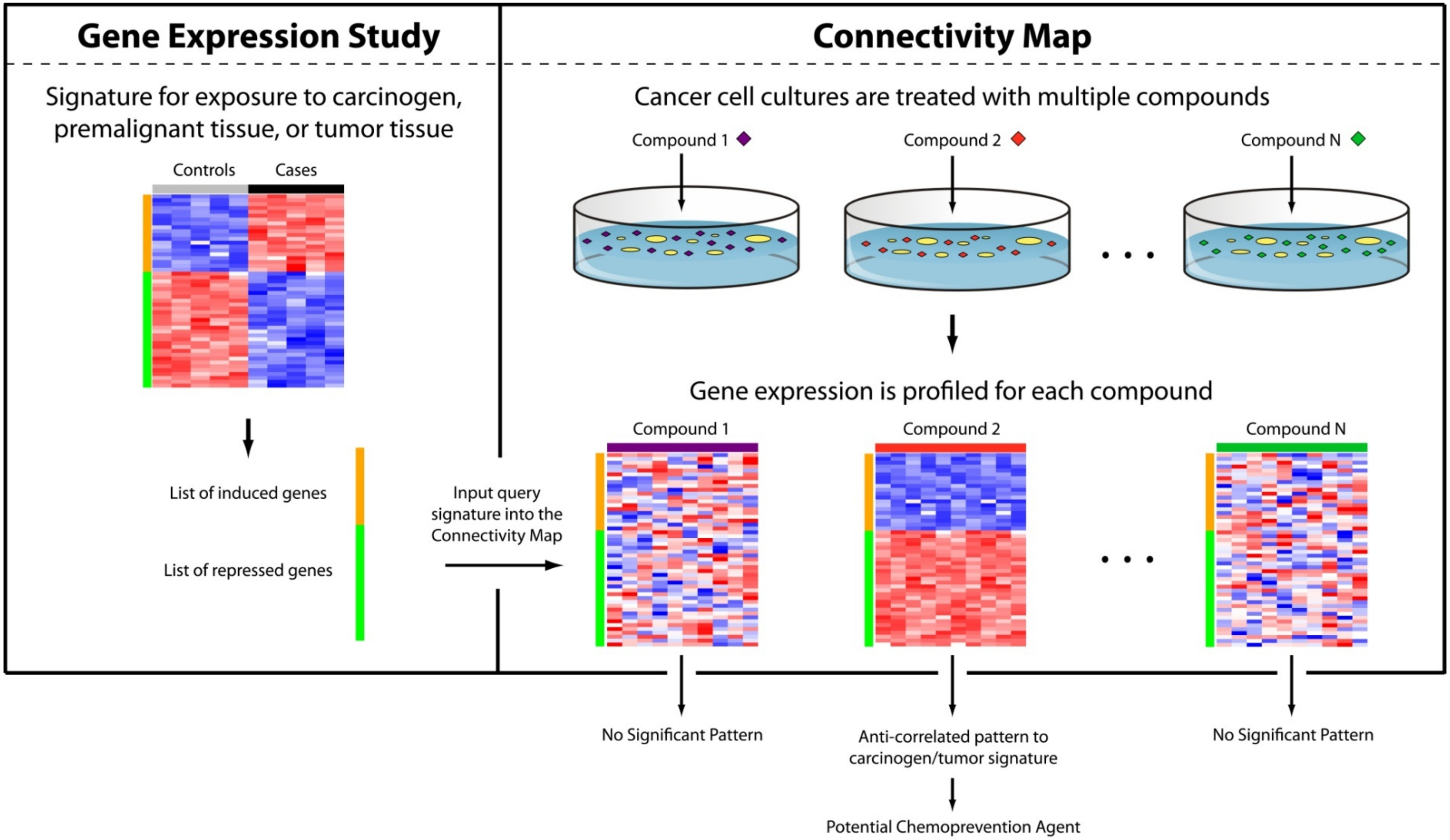
Pathway activation calculated for each test sample





Beane et al. Cancer Research. 2012

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

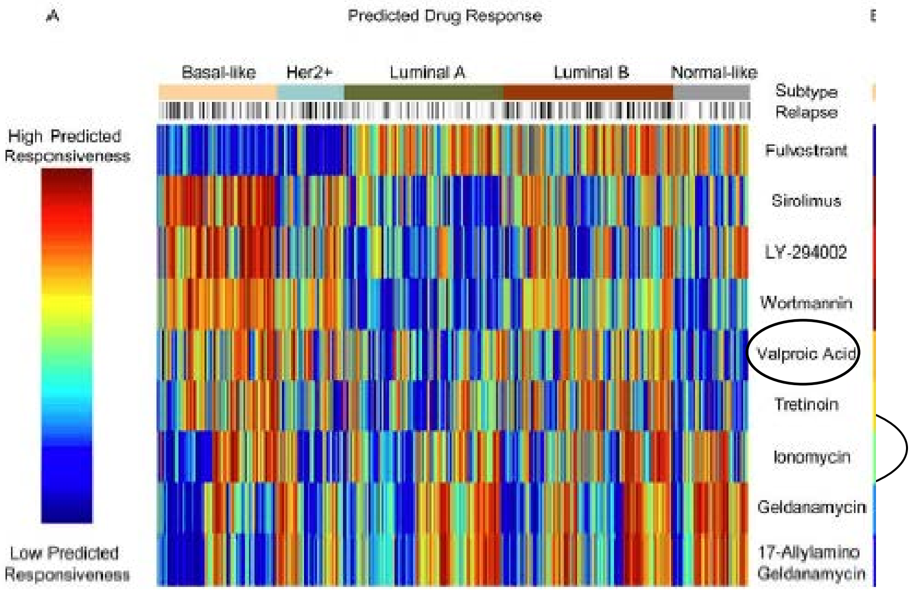


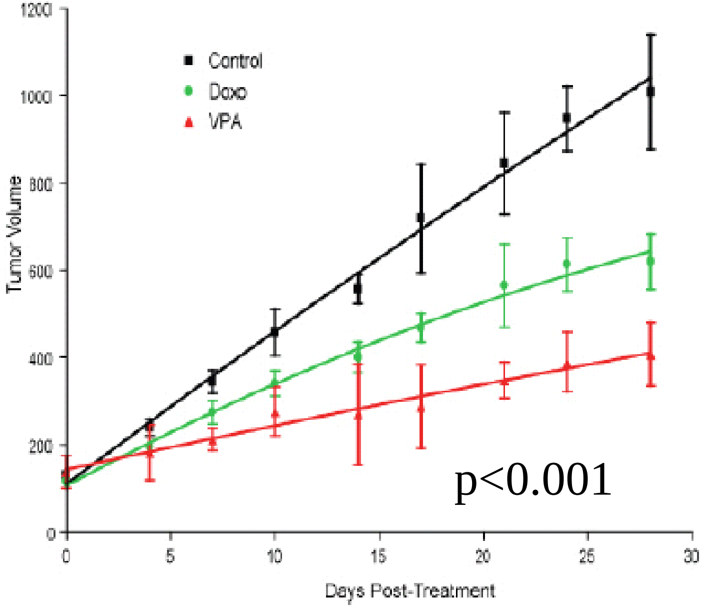
Spira et al. Cancer Prev Research 2010

Using the connectivity map to uncover novel treatment for basal-subtype of breast cancer

Breast cancer

ER -   ER +





$p<0.001$

Testing drug response in vivo (n=10 mice per group

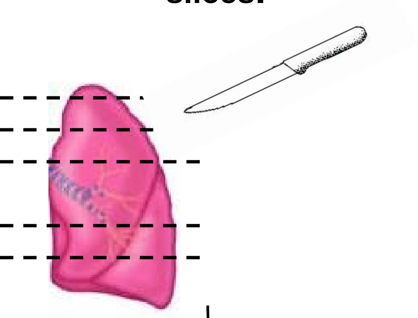
Predicting drug sensitivities from microarray datasets

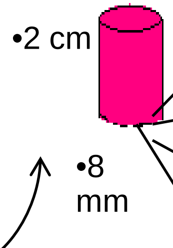
Cohen et al. Molecular Systems Biology 2011

Developing a Genomic Model of Emphysema Progression using Regional Heterogenity

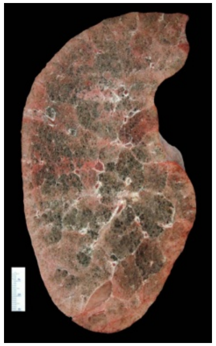
8 patients getting single or double lung transplants for severe COPD.

•Explanted lung inflated, frozen. High resolution CT taken of lung. Cut into slices.

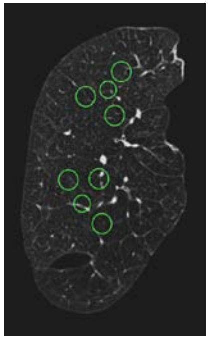




•Cores removed from lung slices – 8 clusters of 4 cores


•before


•after


•High resolution CT

- *University of British Columbia*

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

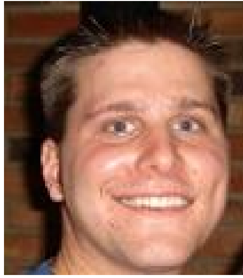
- *Boston University Medical Center*

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)

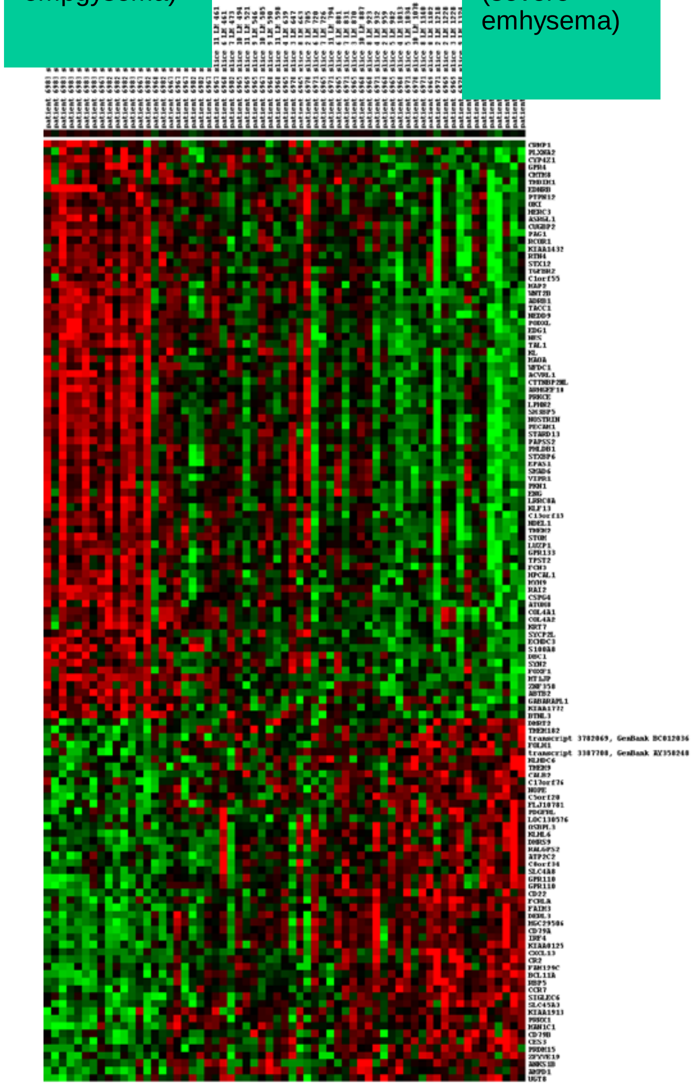
4. **Isolation of DNA for methylation array**
   - Illumina Golden Gate assays

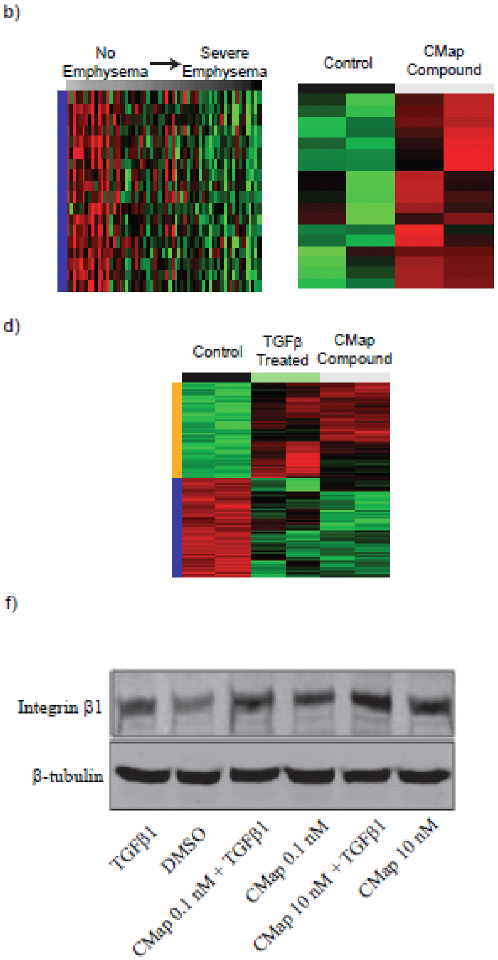
Identification of compound (GHK) that reverses emphysema gene expression signature



- Low LM (no empgysema)

- High LM (severe emhysema)

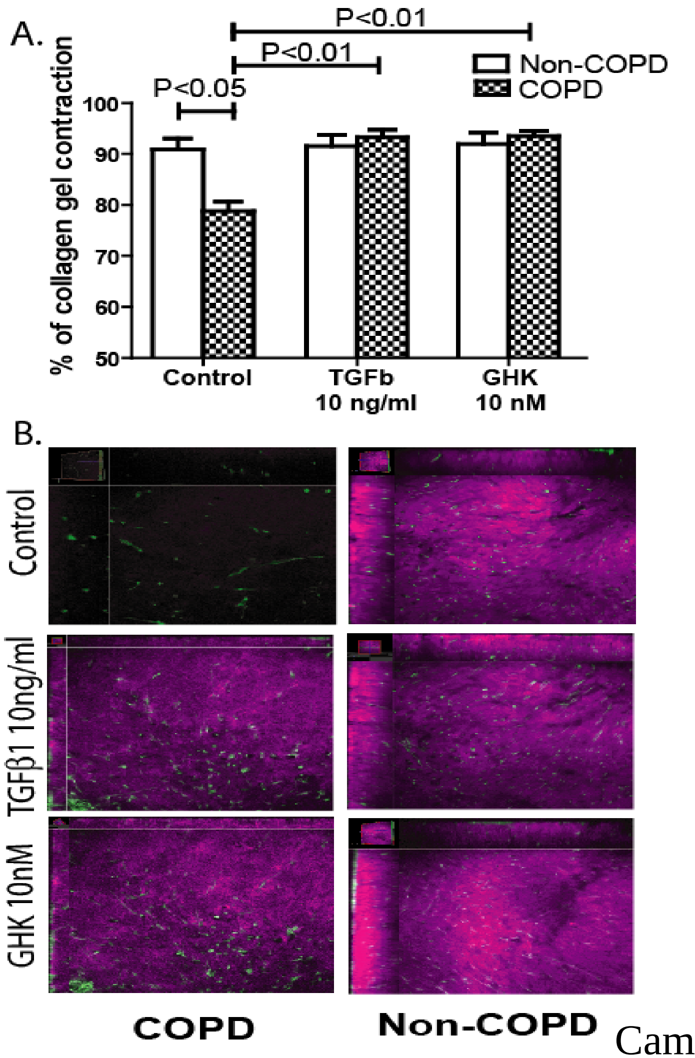




integrin important for fibroblast migration and adhesion

Campbell et al. Genome Medicine 2012

GHK and TGFb restore collagen contraction by lung fibroblasts from smokers with COPD



GHK restores ability of fibrobalsts (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

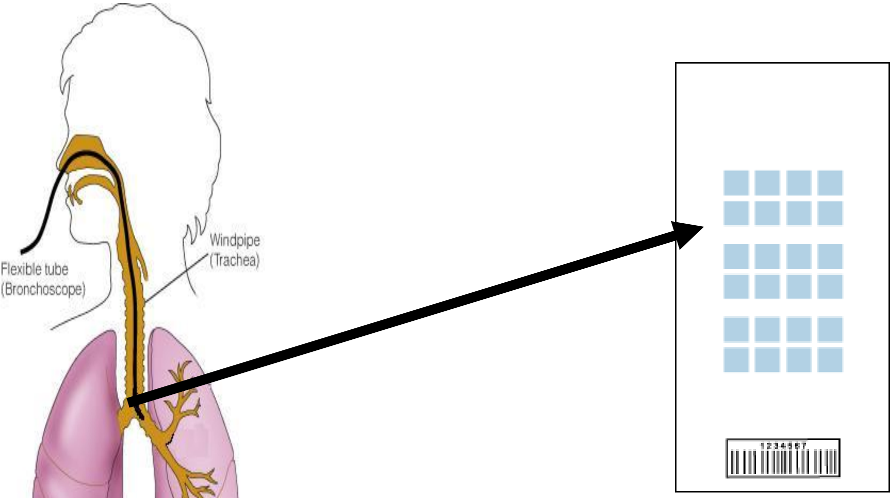
- Bronchial airway gene-expression as a biomarker for the early diagnosis of lung cancer

- 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

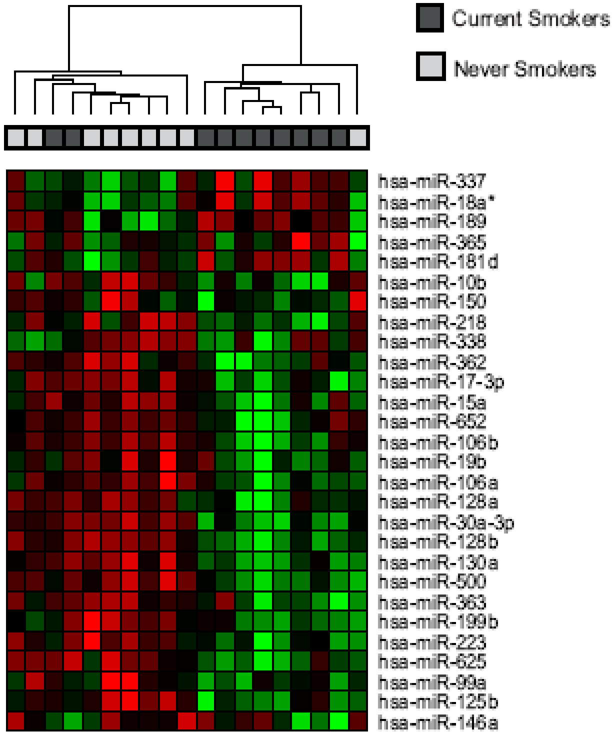
- Moving to the mouth and nasal epithelium
  - Biomarkers for second hand exposure
  - Biomarker of response to quitting and PREP
  - Biomarkers of other inhaled toxic exposures (air pollution)

Impact of smoking on bronchial airway microRNA expression



- -NCODE array

-~500 human miRNA in triplicate



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

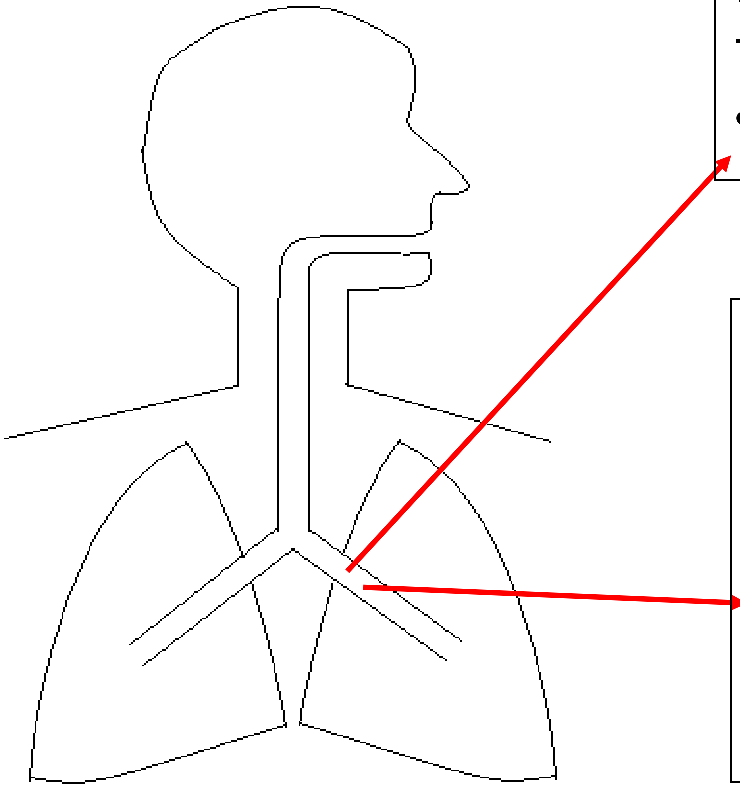
- Cost $$
- Throughput moderate
- Computation ++

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

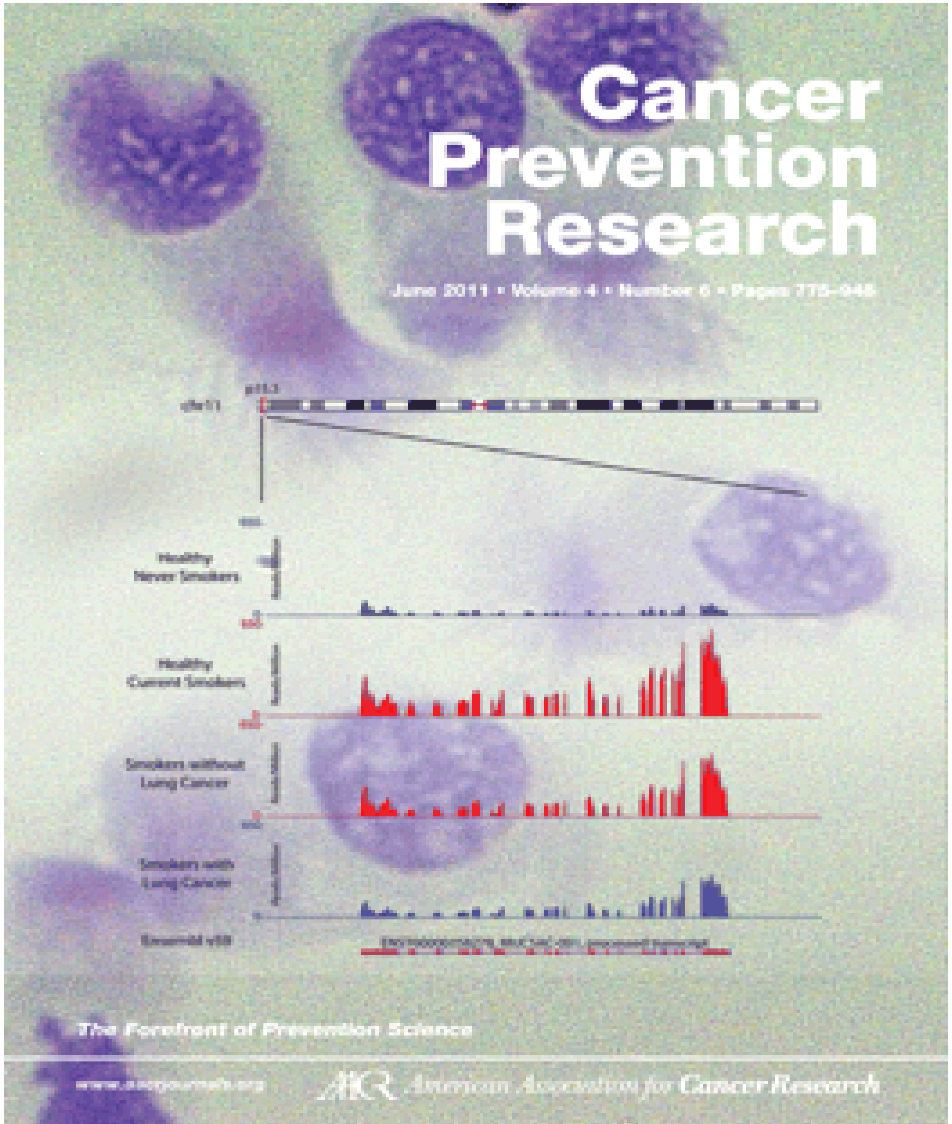
- Small RNA (15-40 bp) sequenced on ABI SOLID platform

- -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



Cancer Prevention Research

June 2011 • Volume 4 • Number 6 • Pages 775–948

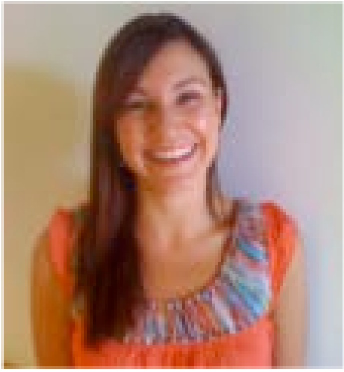
The Forefront of Prevention Science

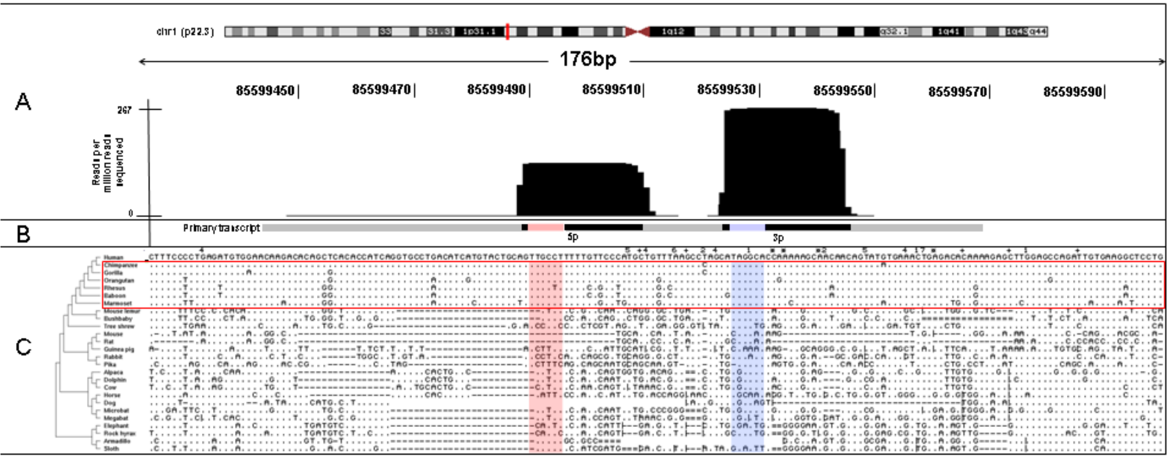
www.aacrjournals.org

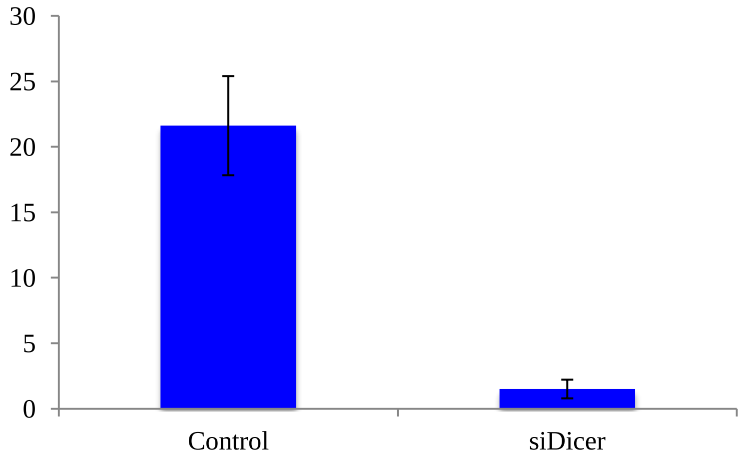
AACR American Association for Cancer Research

Beane et al. 2011

Discovery of novel airway microRNA associated with lung cancer



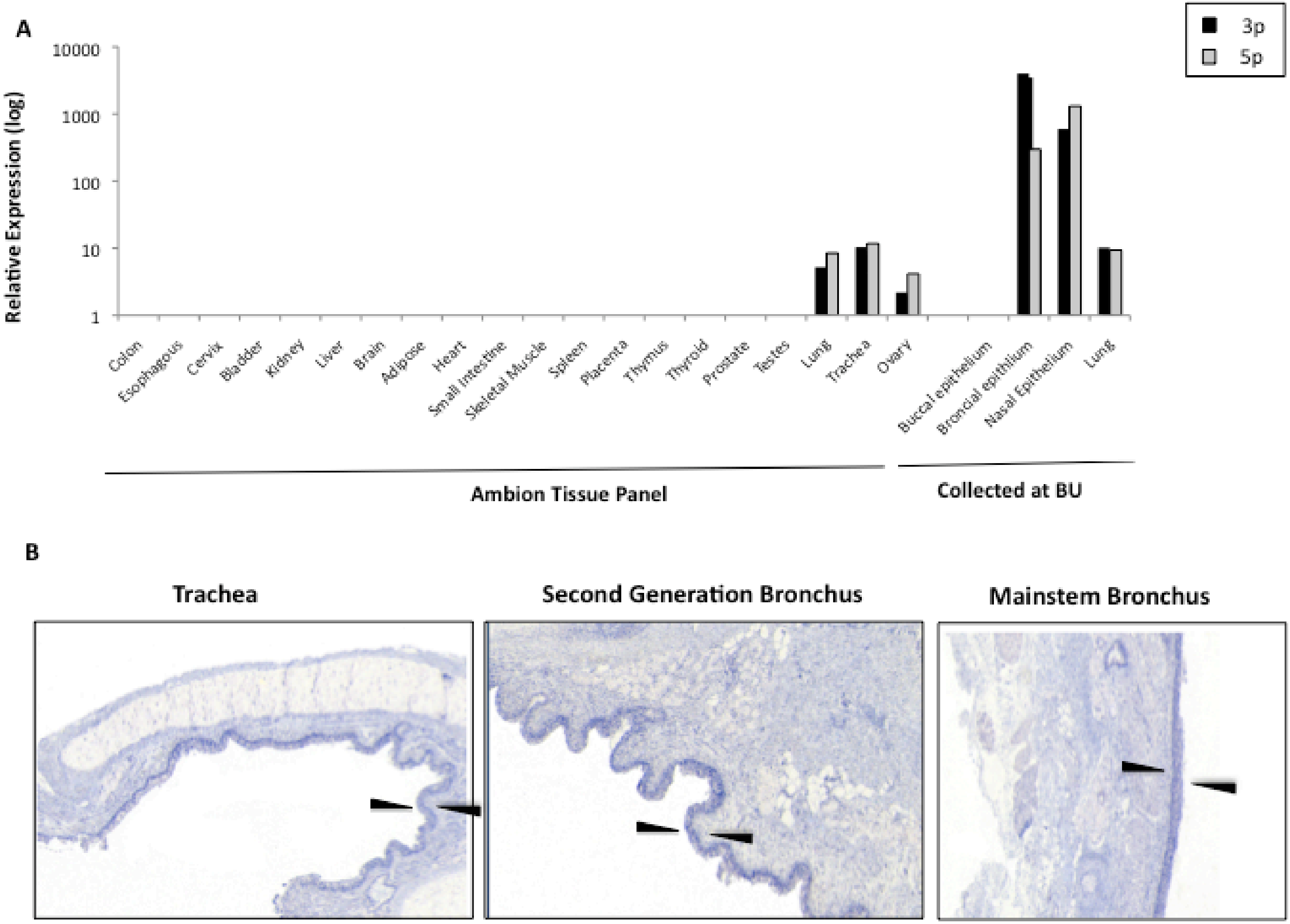




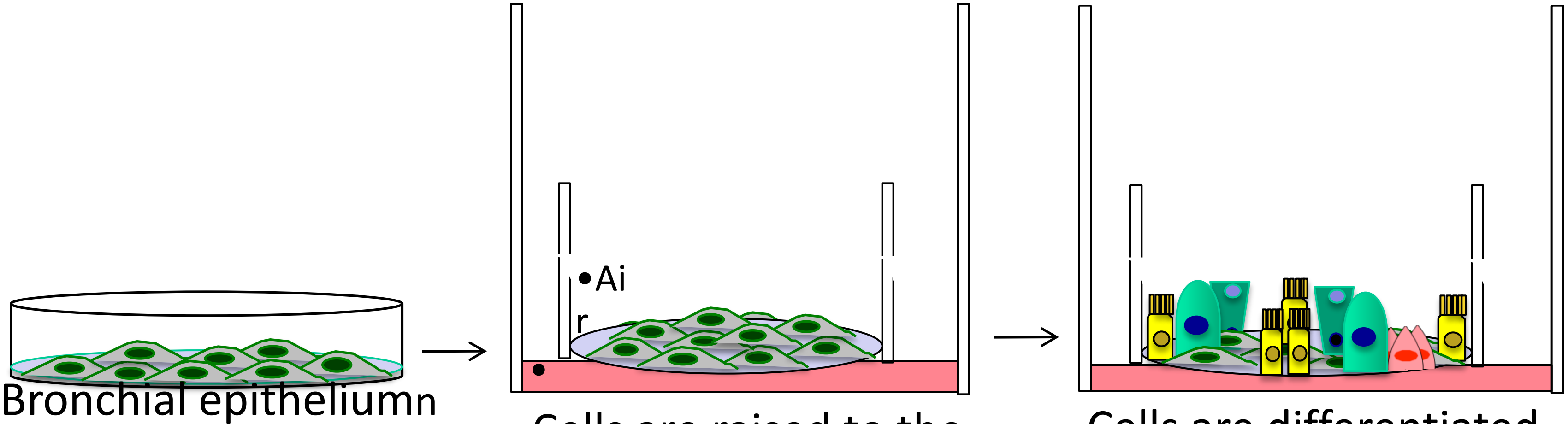
Perdomo et al. Submitted

Novel miRNA is expressed almost exclusively in the respiratory tract and localizes to airway epithelium

Figure 3



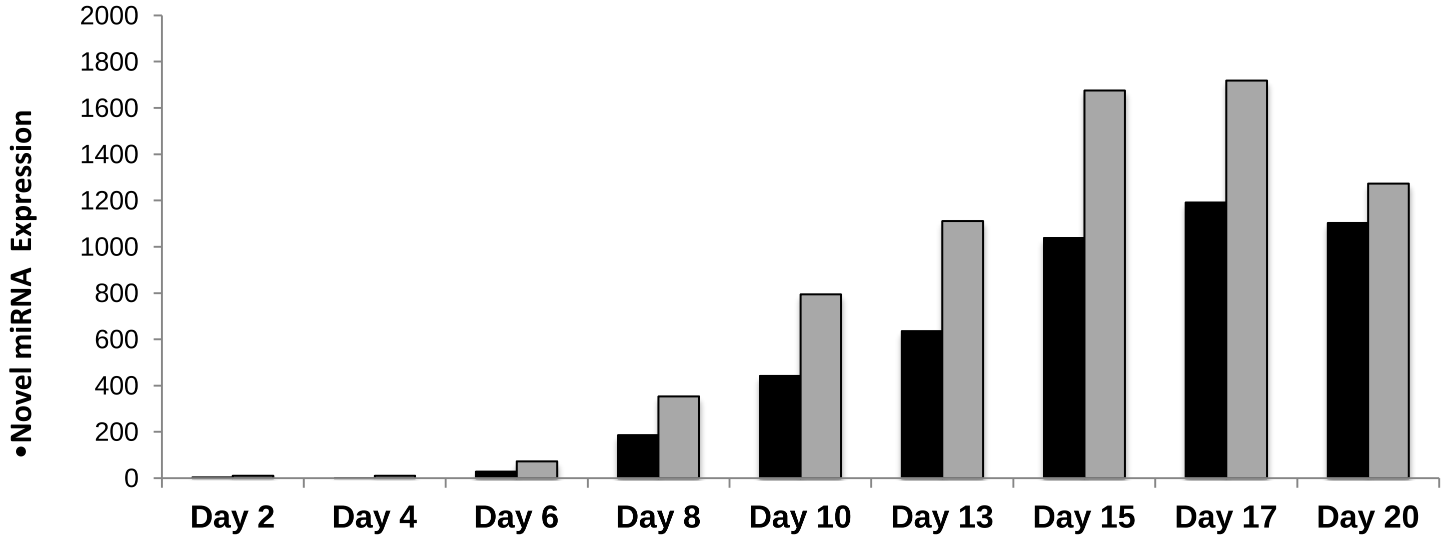
Novel miRNA is expressed during airway epithelial cell differentiation



Bronchial epitheliumn

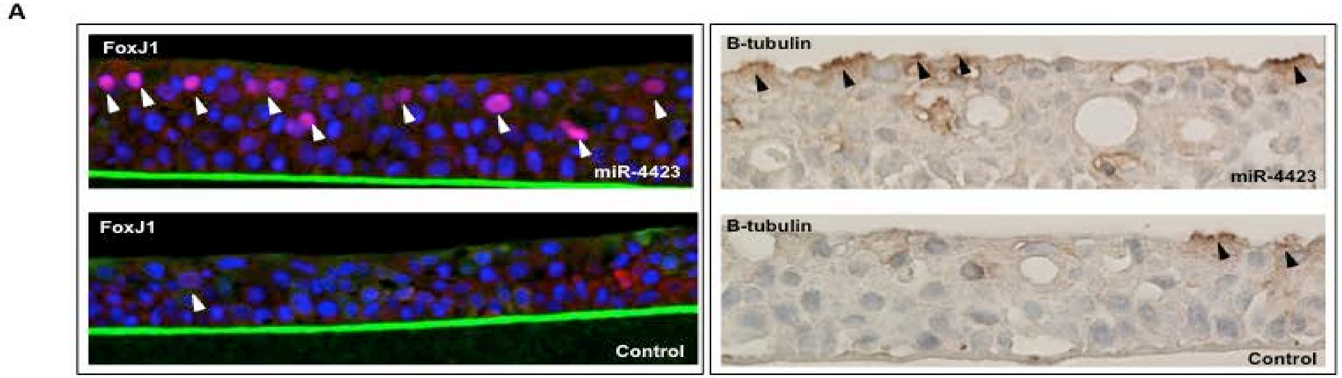
Cells are raised to the ALI (Day 0)

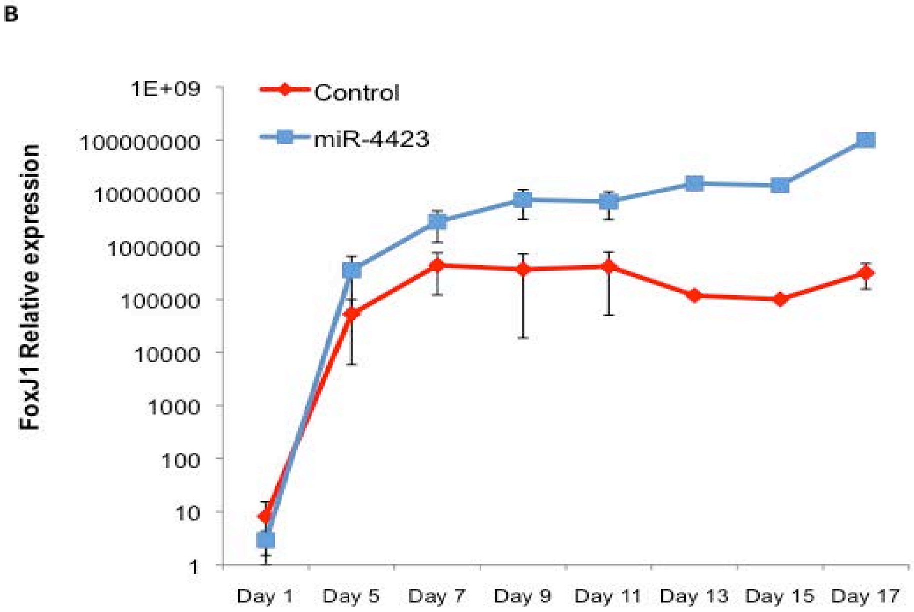
Cells are differentiated (Day 20)



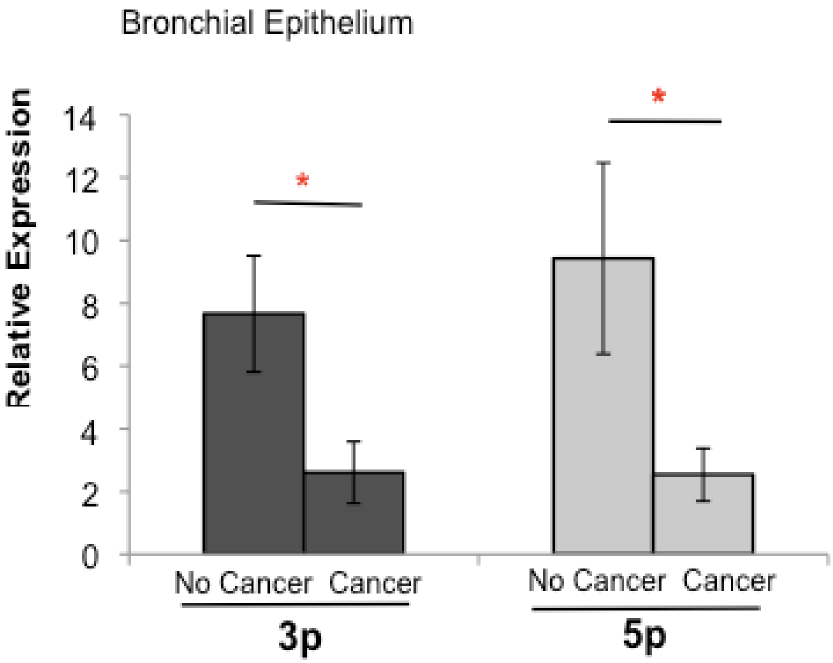
Overexpression of novel miRNA results in more differentiated ciliated cells in ALI

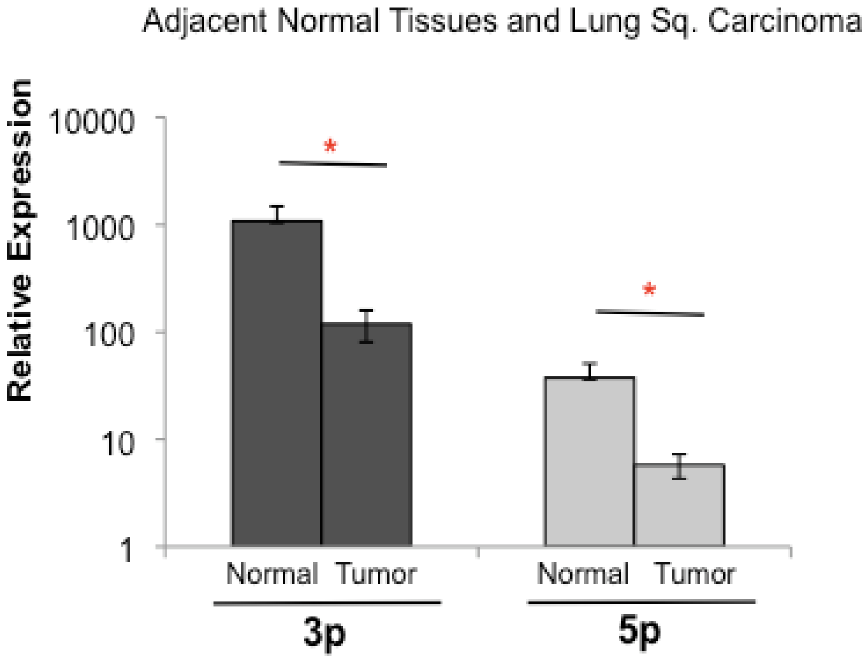
Day 9



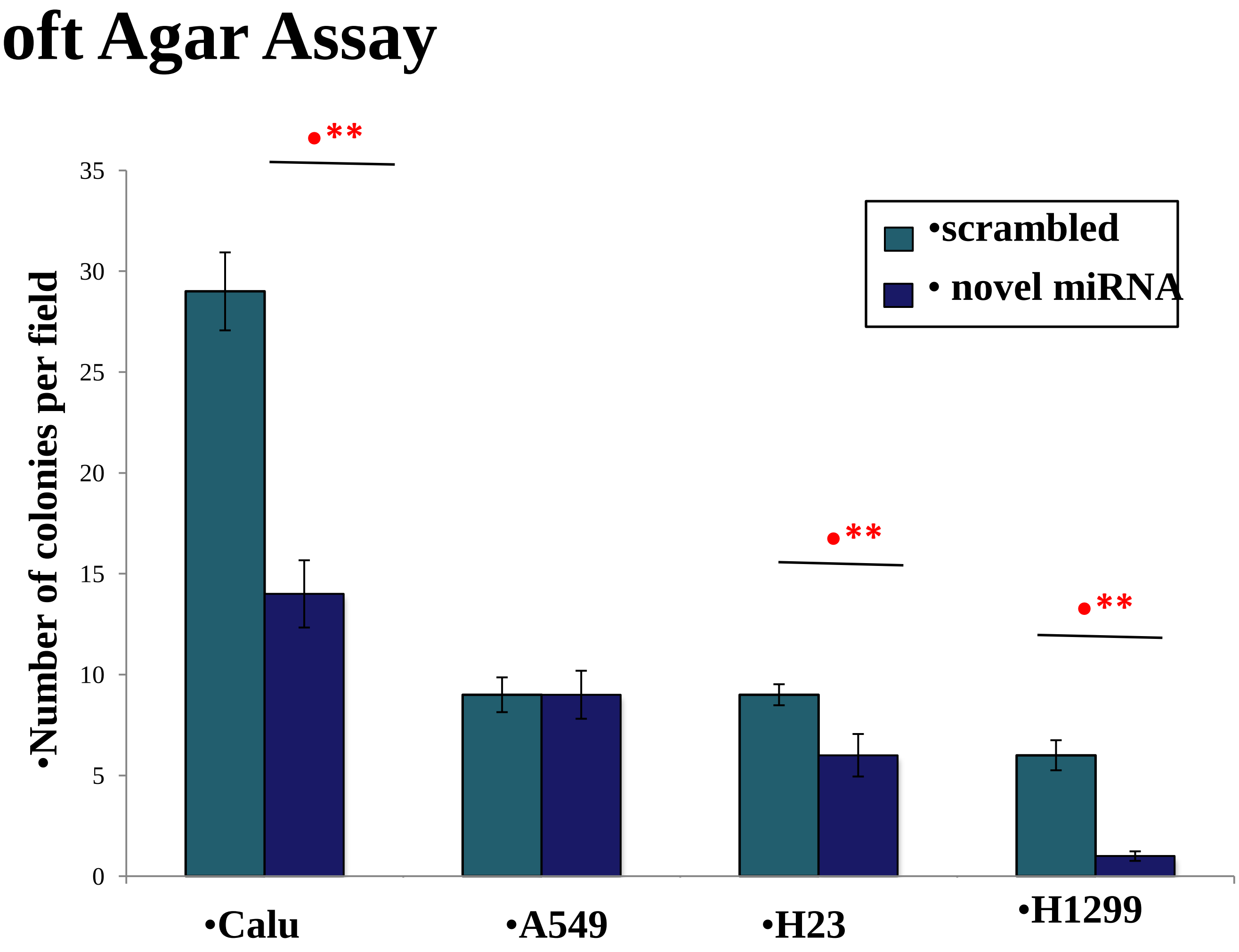


Novel microRNA is downregulated in lung cancer and in the airway of smokers with lung cancer


Bronchial Epithelium

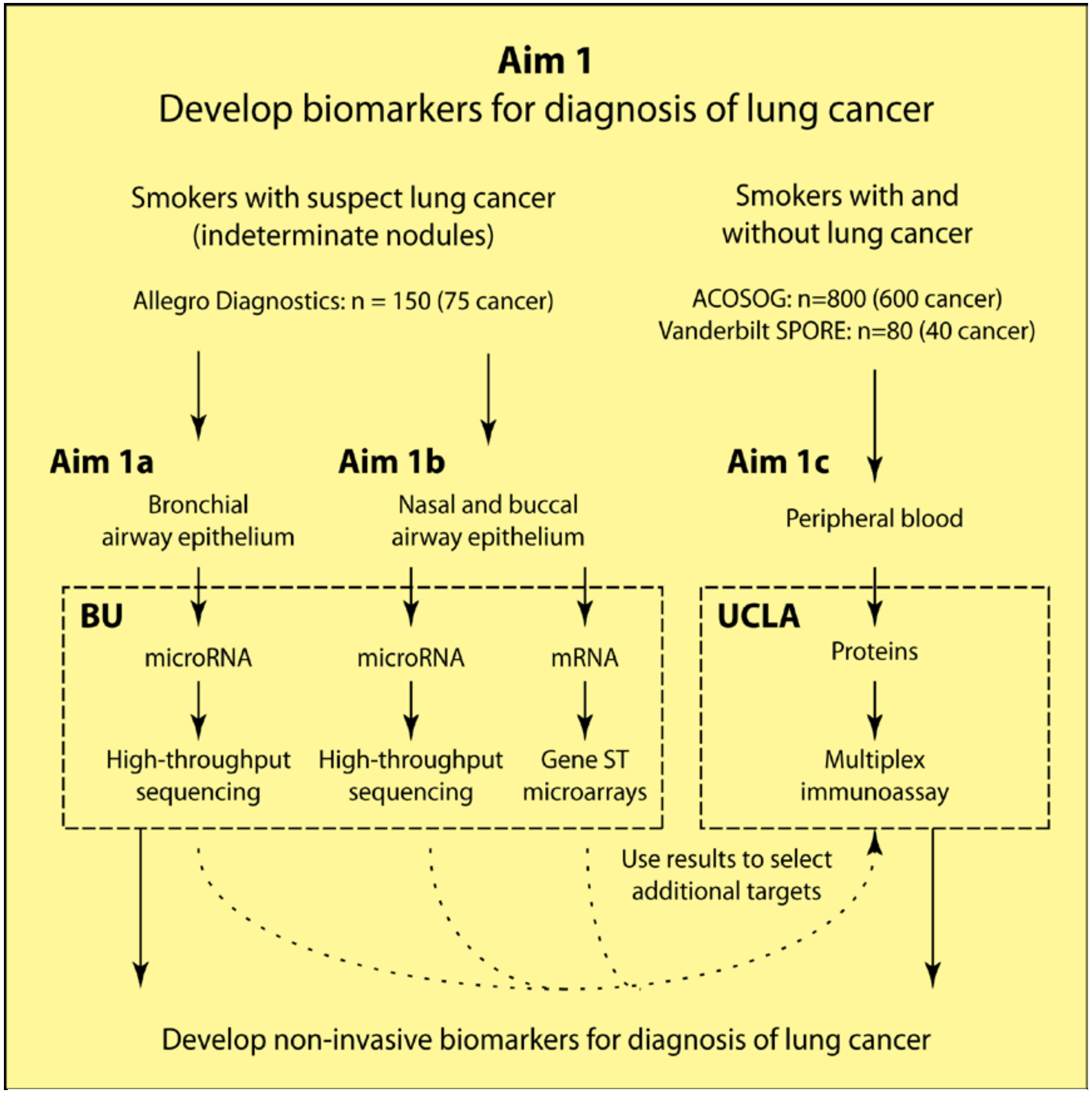

Adjacent Normal Tissues and Lung Sq. Carcinoma

Overexpression of novel miRNA can inhibit anchorage independent tumor cell growth



In collaboration with Carmen Tellez and Steve Belinsky

UCLA-BU EDRN Biomarker Discovery Lab



Aim 1

Develop biomarkers for diagnosis of lung cancer

Smokers with suspect lung cancer (indeterminate nodules)

Allegro Diagnostics: n = 150 (75 cancer)

Smokers with and without lung cancer

ACOSOG: n=800 (600 cancer)
Vanderbilt SPORE: n=80 (40 cancer)

Aim 1a

Bronchial airway epithelium

Aim 1b

Nasal and buccal airway epithelium

Aim 1c

Peripheral blood

BU

microRNA → High-throughput sequencing

microRNA → High-throughput sequencing

mRNA → Gene ST microarrays

UCLA

Proteins → Multiplex immunoassay

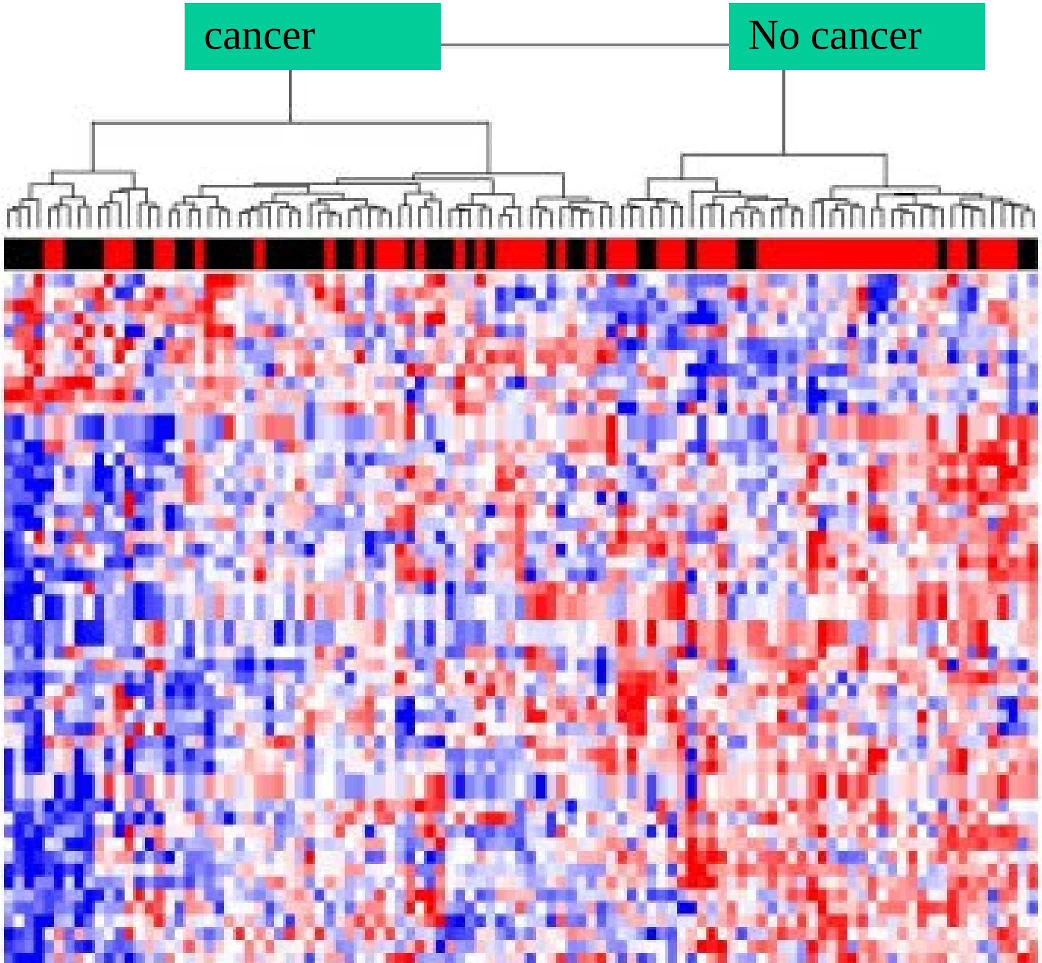
Use results to select additional targets

Develop non-invasive biomarkers for diagnosis of lung cancer

•**Figure 1.** Overview of Aim1

Developing a microRNA-based airway biomarker for diagnosis of lung cancer in the AllegroDx trial

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



Outline

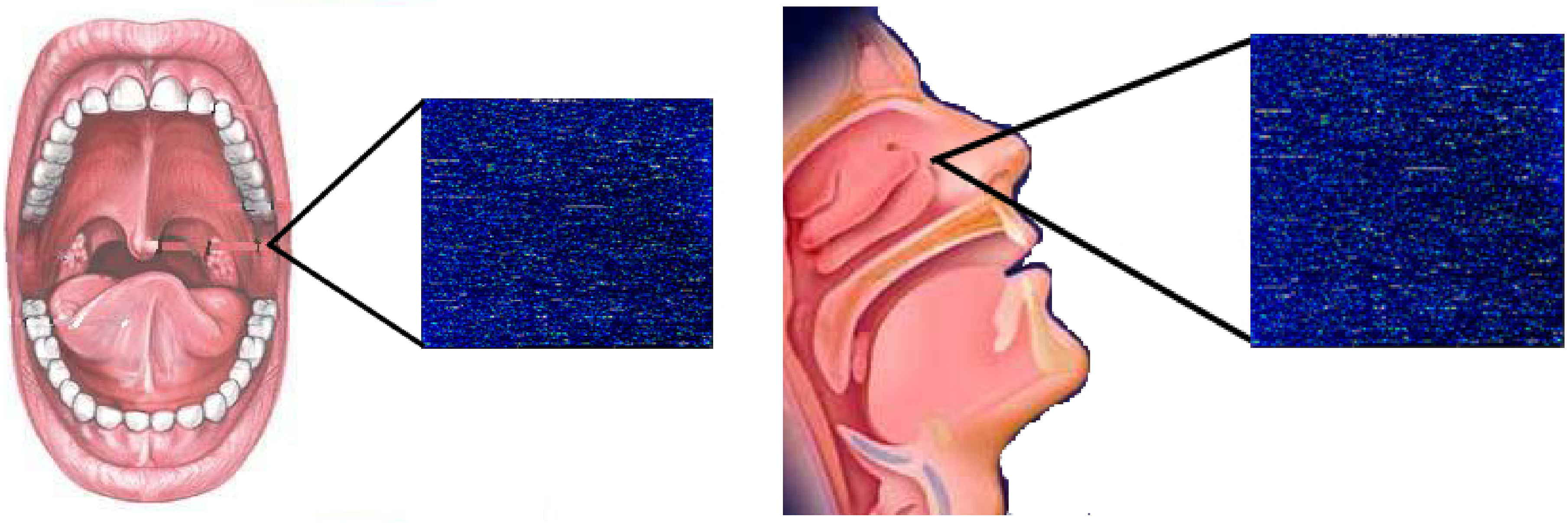
- Bronchial airway gene-expression as a biomarker for the early diagnosis of lung cancer

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

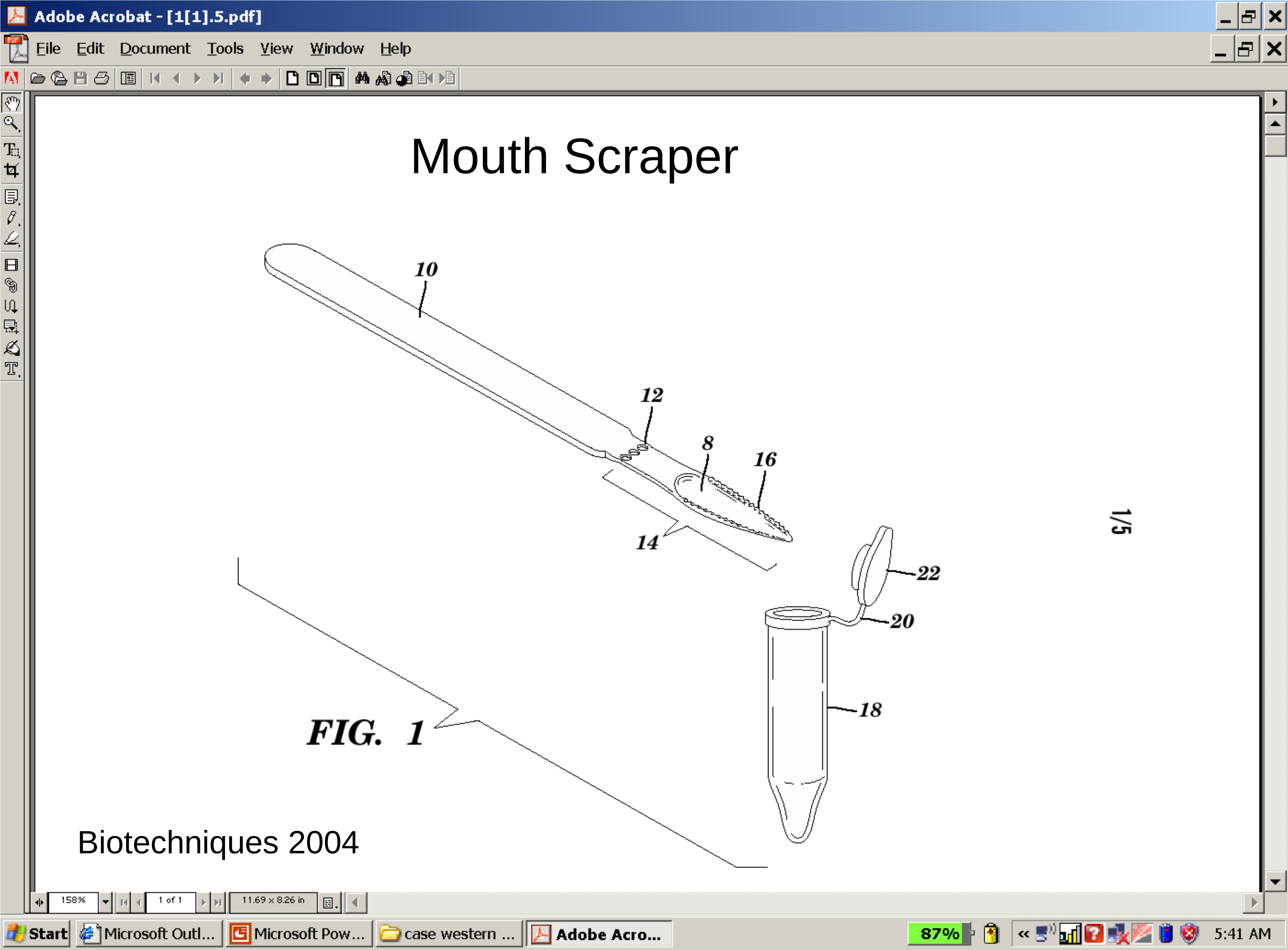
- Extending "field of injury" to microRNA

- 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



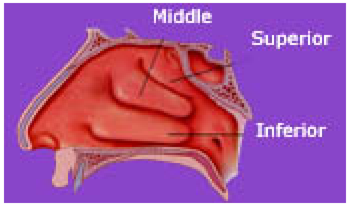
Mouth Scraper

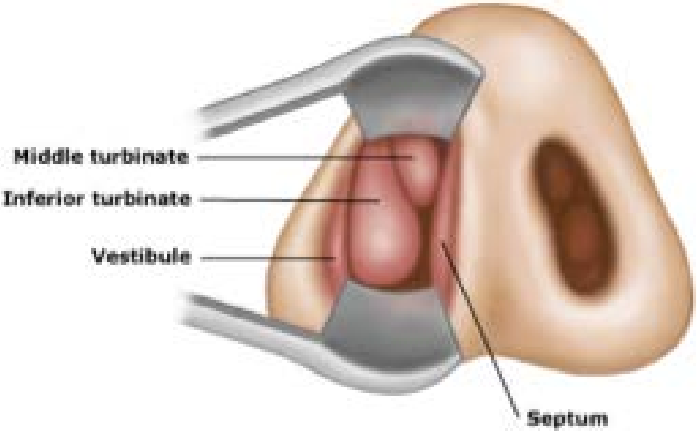


Biotechniques 2004

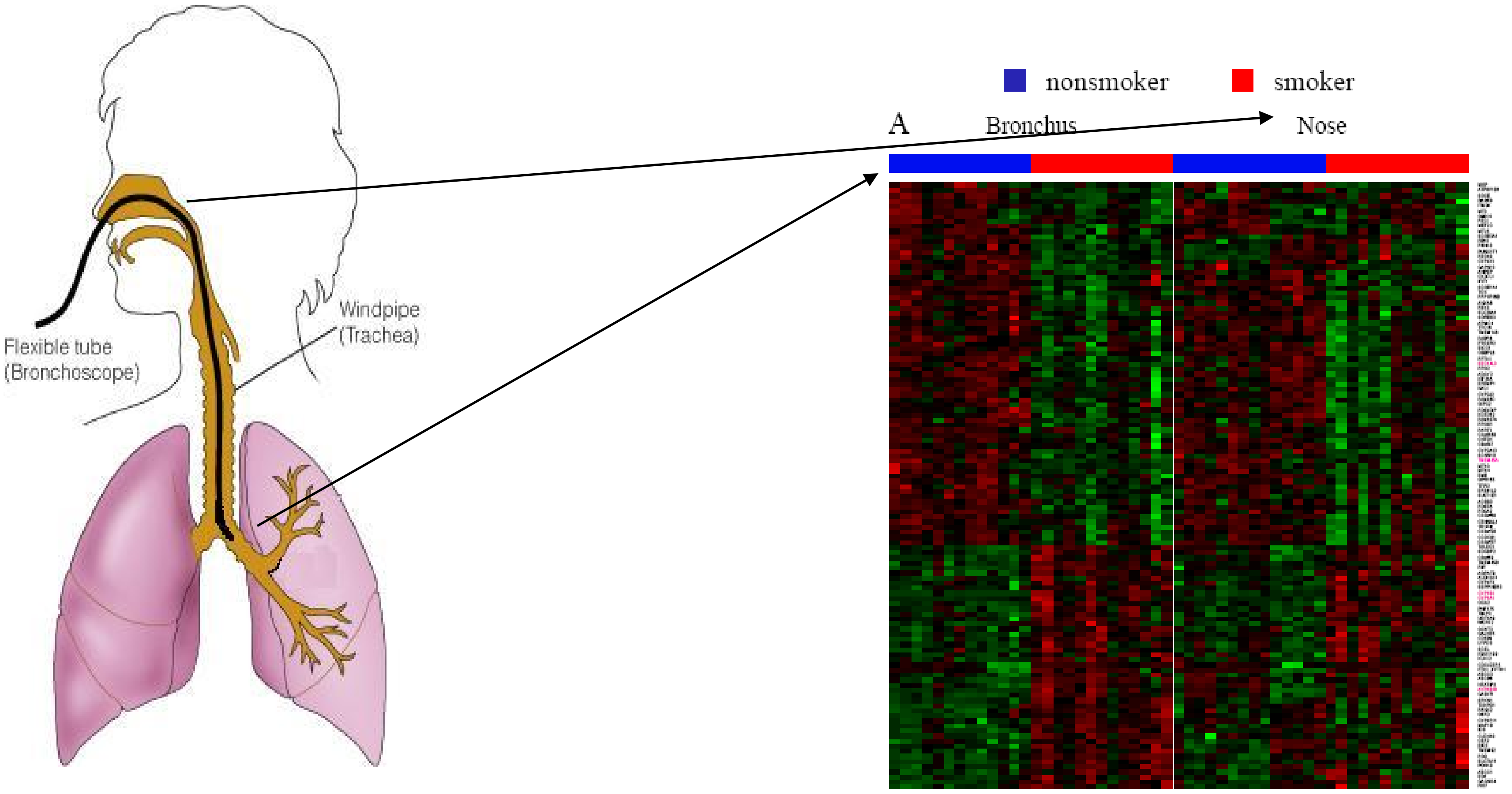
Technique for Obtaining RNA from Nasal Mucosal Brushings

- Cytosoft® brushings from “interior” of inferior turbinate
- Immerse in RNA later



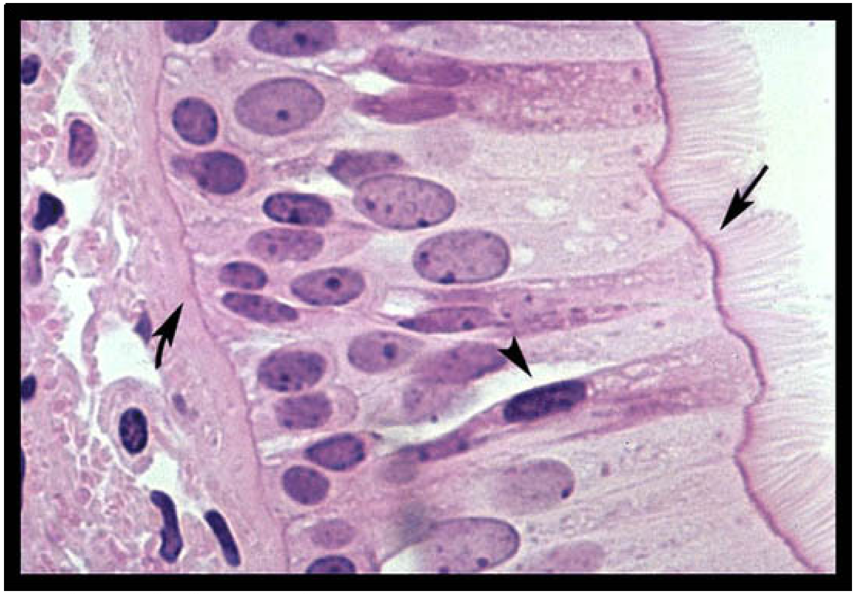


Nasal gene expression reflects the bronchial airway gene-expression response to smoking

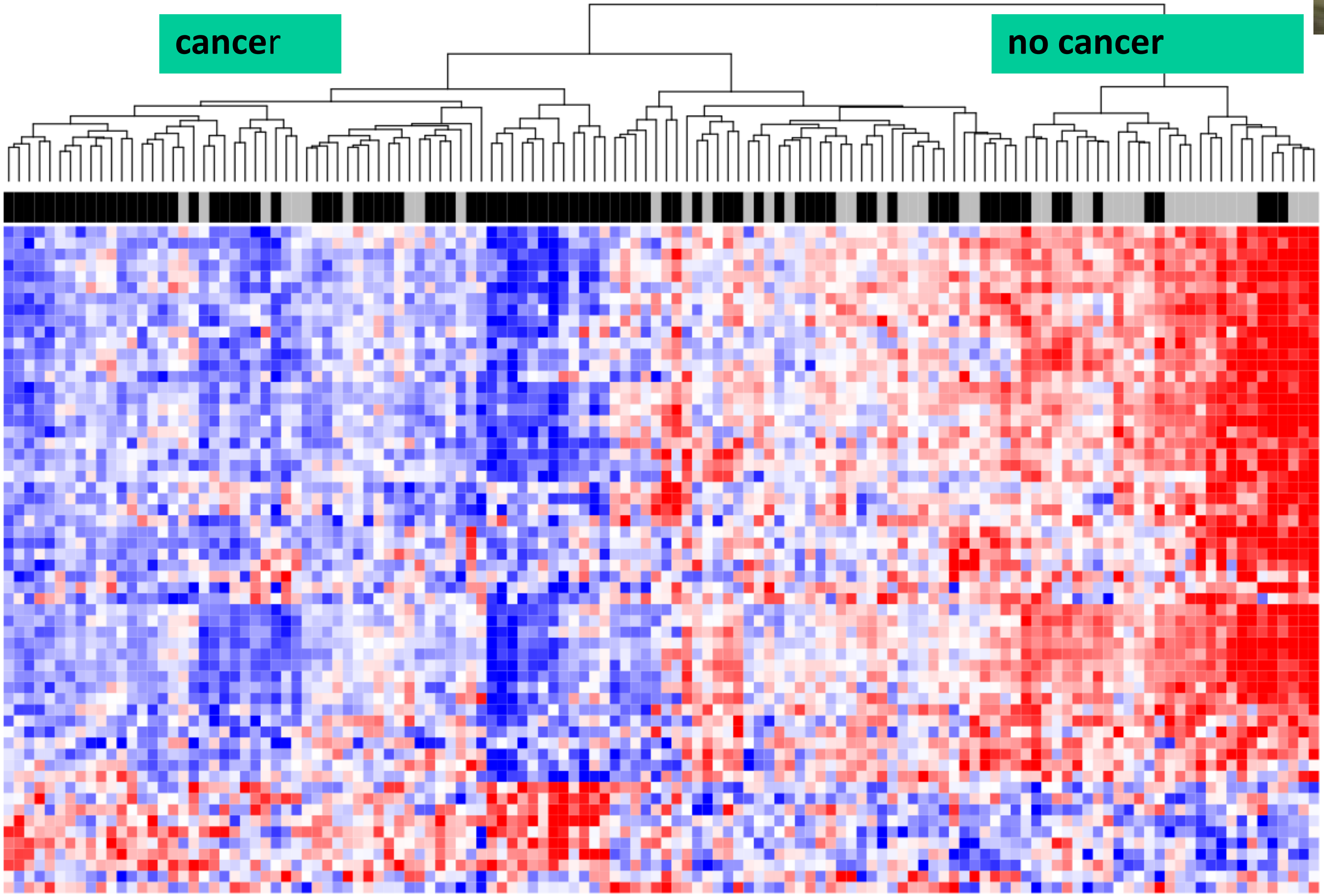


Zhang et al. Physiological Genomics 2010

The nose-bronch relationship



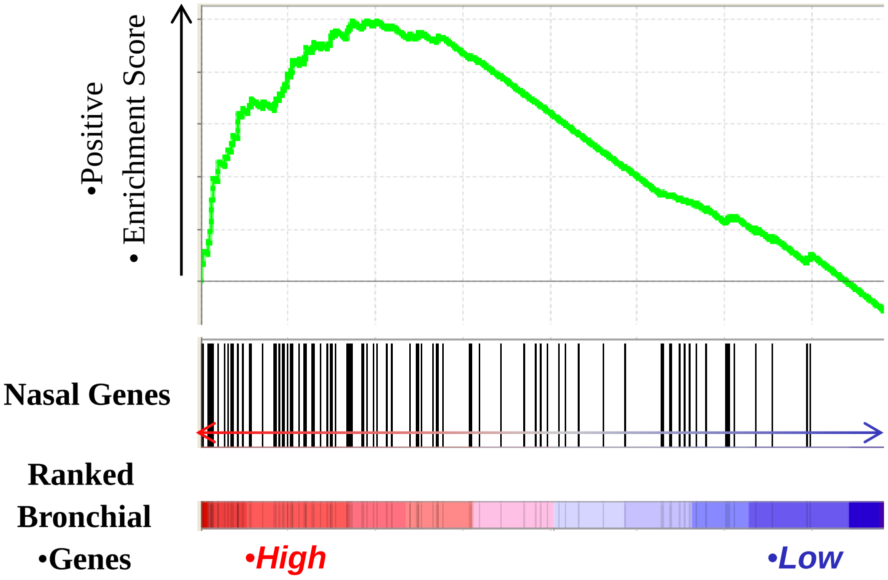
Genes differentially expressed in nose between smokers with and without lung cancer in the AllegroDx trial



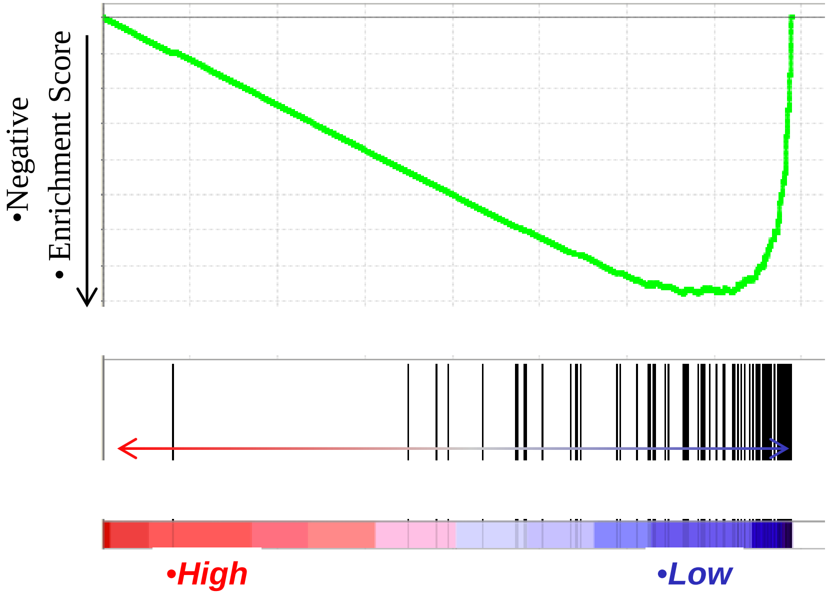
Genes associated with cancer in nasal epithelium are similarly up- and down-regulated in bronchial epithelium

(GSEA $p < 0.001$)

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



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

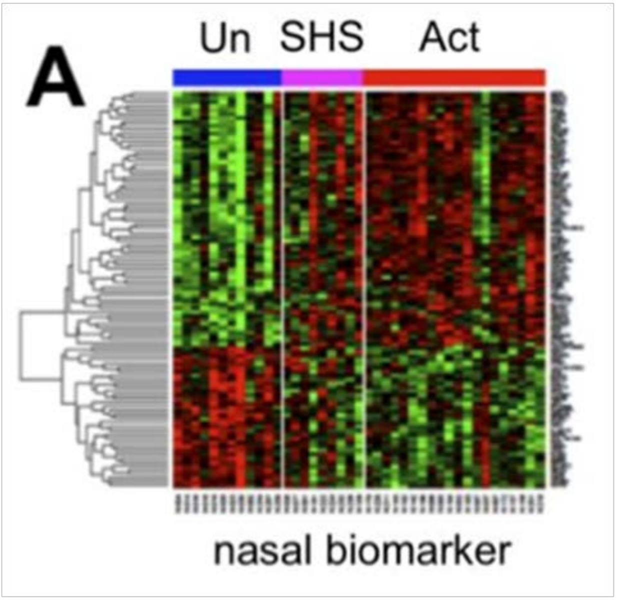


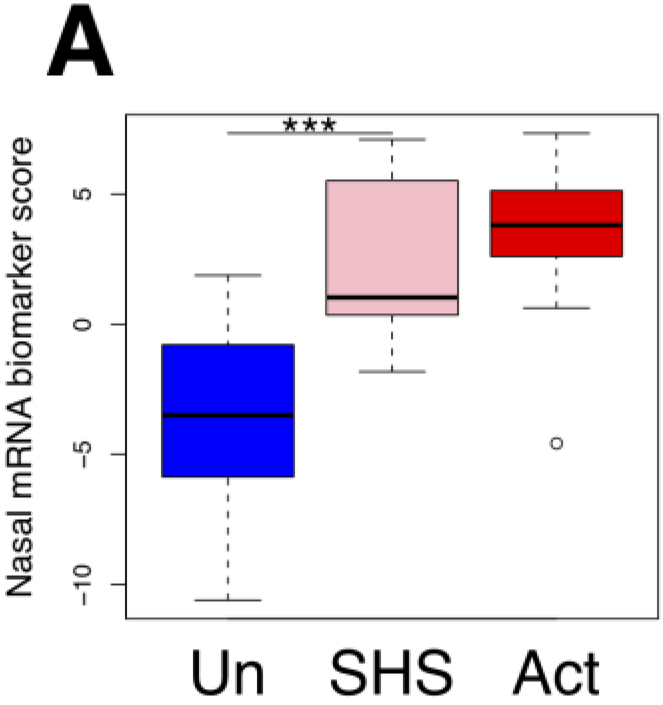
Upper airway biomarkers developed as part of the GEI

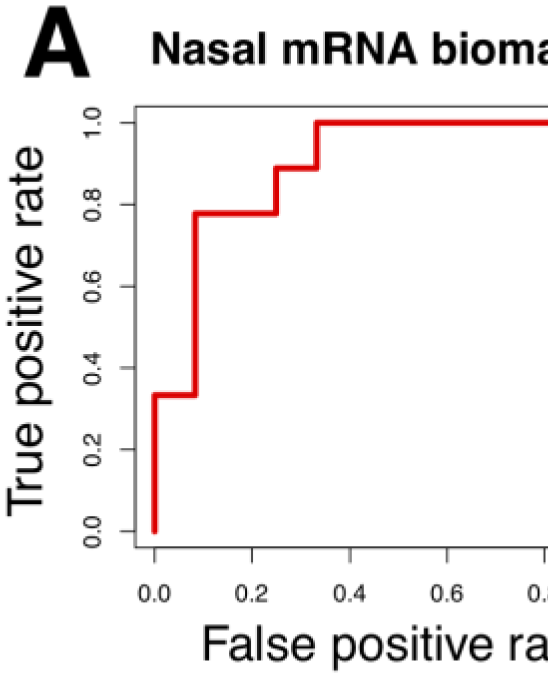
- Nasal and buccal gene-expression as biomarker of ever exposure
- Nasal biomarker of second-hand smoke exposure
- Buccal biomarker of cumulative exposure (i.e. pack-yrs) to tobacco smoke
- Nasal gene-expression signature post smoking cessation
- Indoor air pollution in China (supplement)

Development of a nasal gene-expression biomarker of passive exposure to smoking

	unexposed	SHS exposed	Active smoker (0-10 CPD)	Active Smoker (10-15 CPD)	Active Smoker (> 15 CPD)
n	12	9	8	5	7
% female	75%	56%	63%	40%	57%
Age	24.0	23.9	22.3	25.0	23.3
CPD	0.0	0.0	5.6	11.2	19.6
Cotinine	0.03	1.1	84.2	122.5	231.3







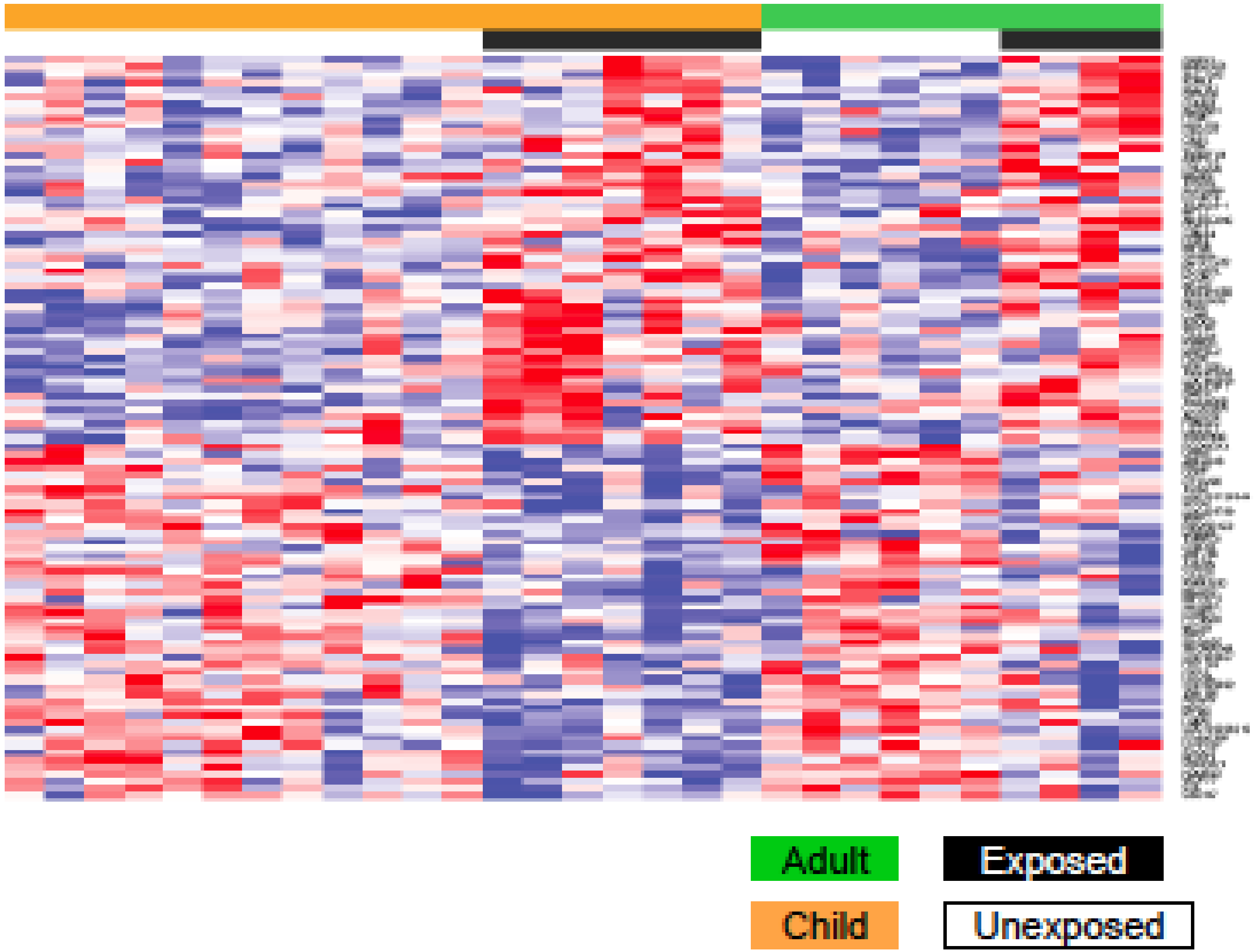
Moving nasal biomarkers of secondhand exposure to children

- 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)

	Exposed (n=11)		**Control (n=18)**		**p-value**
Adult / Child	Adult	Child	Adult	Child	1
	4	7	6	12	
Gender	Male	Female	Male	Female	0.20
	1C, 0A	6C, 4A	6C, 0A	6C, 6A	
RIN	7.1 (±1.0)		6.9 (±1.2)		0.75
$PM_{2.5}$ (ug/m^3)	25.9 (±24.8)		NA		NA
SHS UVPM (ug/m^3)	2.9 (±2.4)		NA		NA
Air Nicotine (ug/m^3)	0.97 (±1.91)		NA		NA

CCCEH: Similar gene expression changes associated with exposure status are detected across children and adults



Upper airway biomarkers developed as part of the GEI

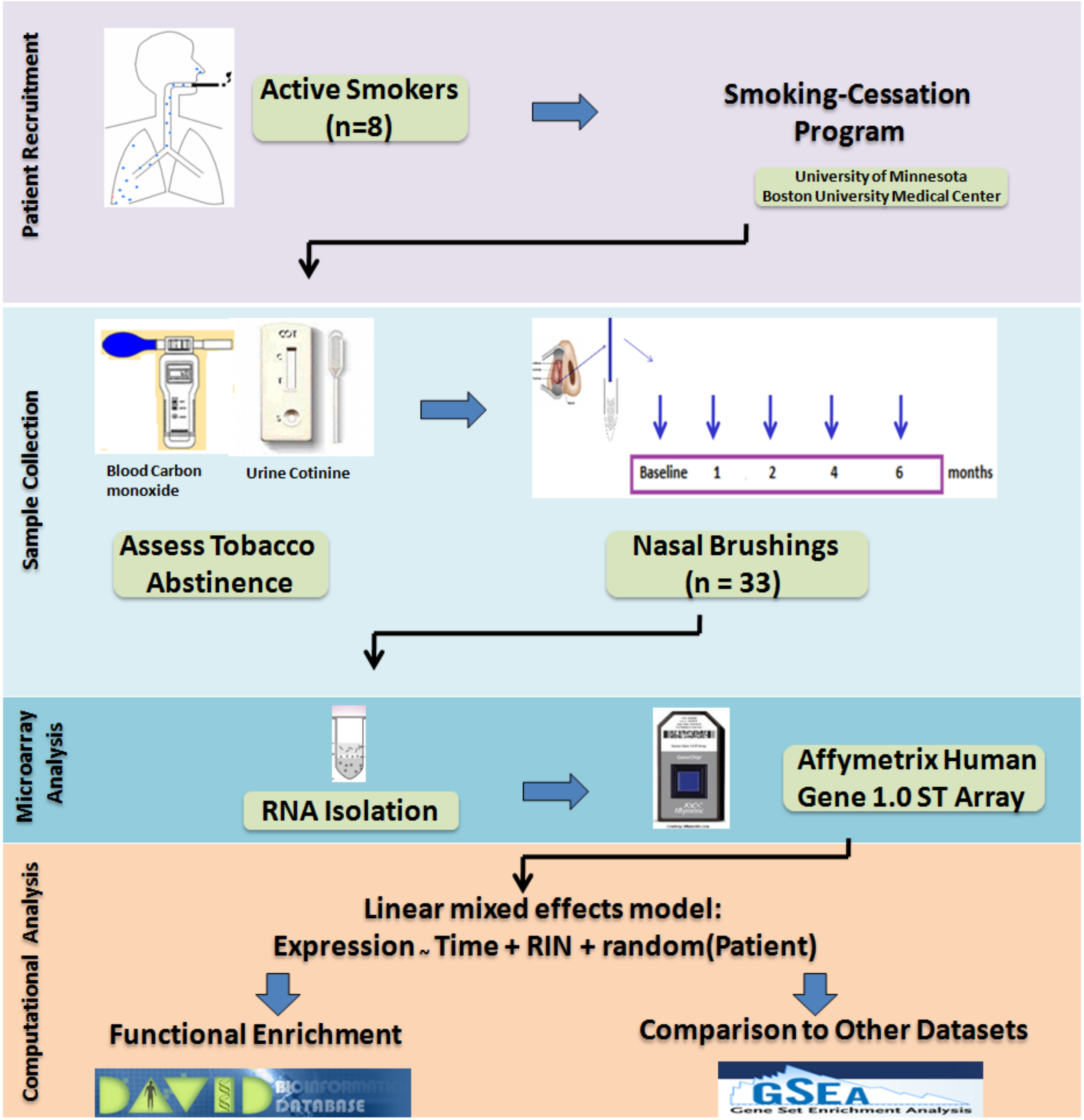
- Nasal and buccal gene-expression as biomarker of ever exposure

- 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



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

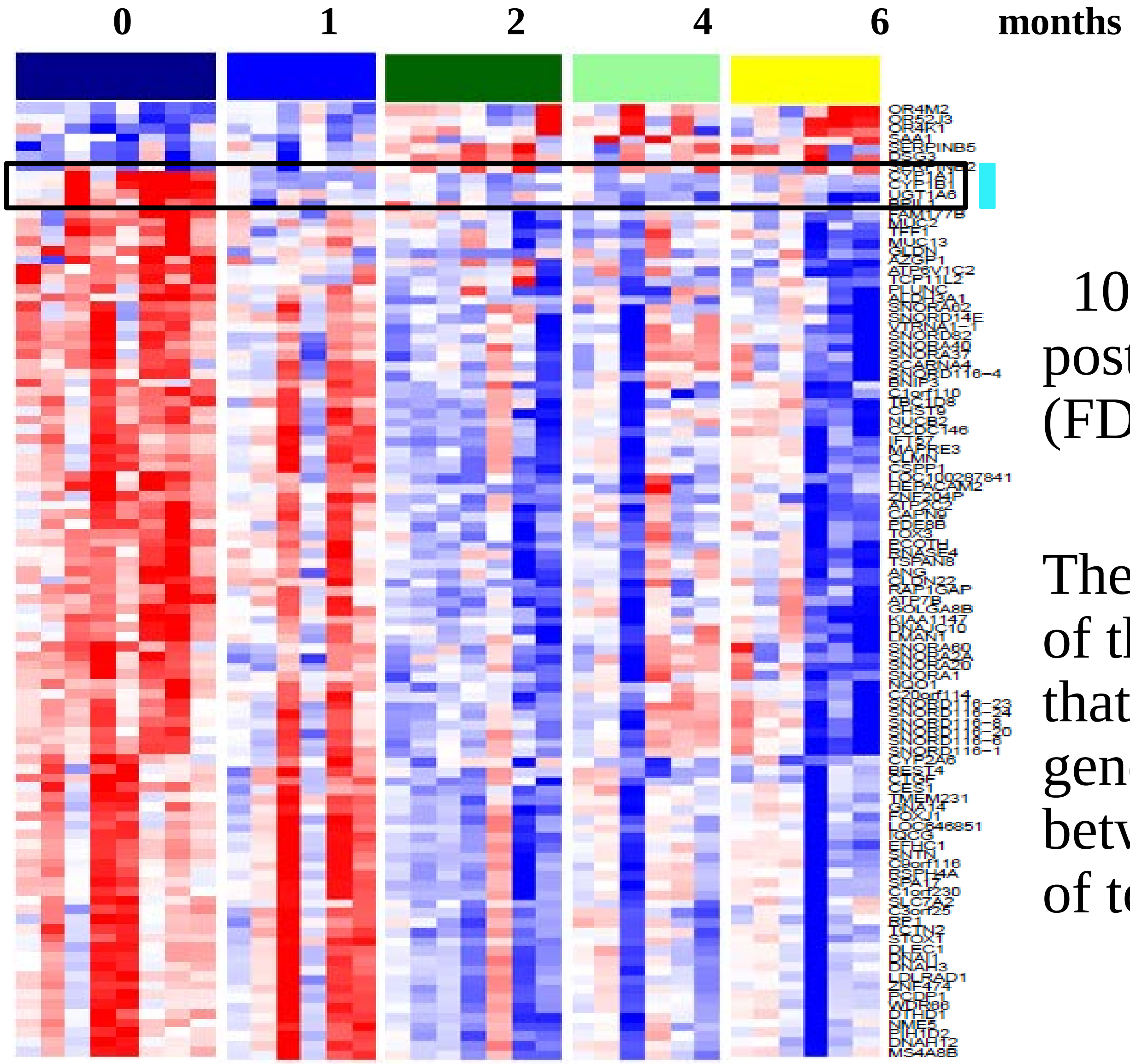
Affymetrix Human Gene 1.0 ST Array

Computational Analysis

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

Functional Enrichment

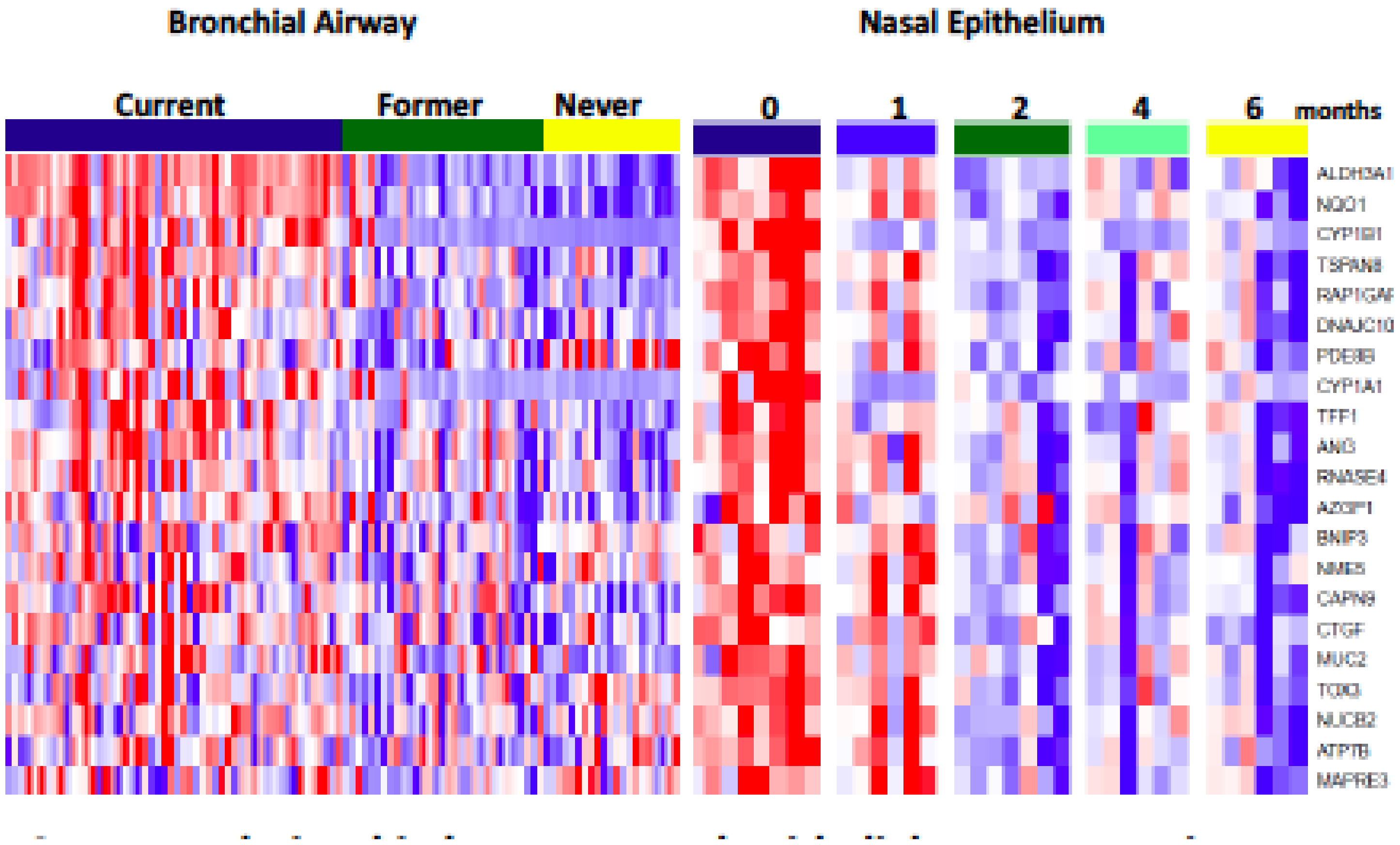
Comparison to Other Datasets



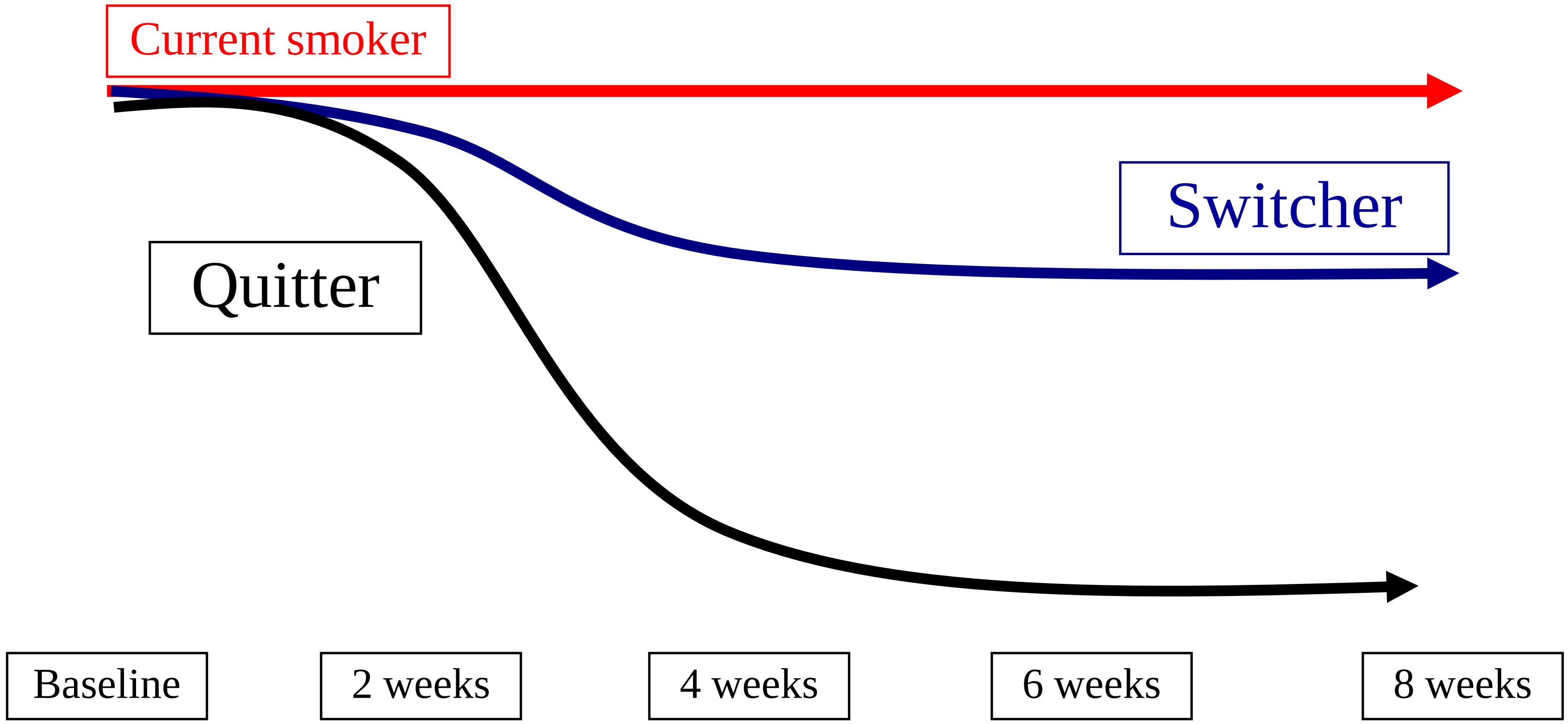
101 Genes that change post- smoking cessation (FDR <0.05)

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



Upper airway biomarkers developed as part of the GEI

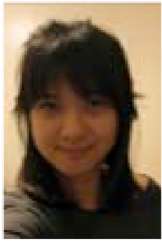
- Nasal and buccal gene-expression as biomarker of ever exposure

- Nasal biomarker of second-hand smoke exposure

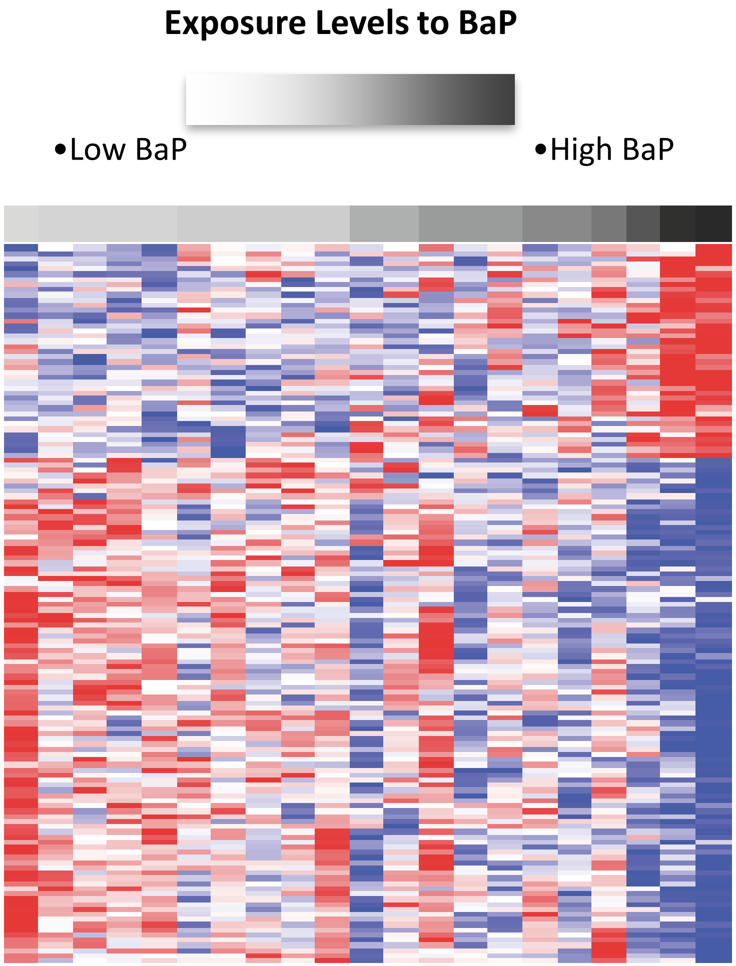
- Buccal biomarker of cumulative exposure (i.e. pack-yrs) to tobacco smoke

- 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



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



Enrichment of these genes among those that change in buccal epithelium of active smokers

In collaboration with Nat Rothman and Qing Lan

Female Residents of Xuan Wei and Personal Filter Metrics for Microarray Pilot Study

Category	Low Exposure (n=12)		High Exposure (n=11)		p-value
SEX	**Male**	**Female**	**Male**	**Female**	1
	0	12	0	11	
RNA Quality	**Good**	**Better**	**Good**	**Better**	1
	5	7	4	7	
PM$_{2.5}$ (µg/m^3)	104.77 (±50.34)		283.25 (±131.27)		0.001
BAP (ng/m^3)	15.56 (±6.26)		99.69 (±44.49)		4.054e-05*

- **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