

## Past, Present, and Future:

**What we have done...**  
**What we have found...**  
**and**  
**Where we are going**

UAB/UCSD/Yale/CMU

In-person meeting

July 29, 2015

Draft of July 14, 2015

**lungmap.net**



**LungMAP**  
Molecular Atlas of Lung  
Development Program



## Overview of presentation



1. Data or deliverable that has been generated
  2. Results, interpretation of data, and computational models
    - a) Time course mRNA (Ziv/Emre)
    - b) Time course microRNA (Ziv/Emre)
    - c) Time course DNA methylation (Jim Hagood MD)
  3. Where we want to go (or think we should go next)
  4. Plan for integrating the data
  5. Coordination of efforts
    - a) Single cell analysis computational models (Ziv/Aaron)
  6. Translation from mouse to human
  7. Fitting into BREATH and interaction with data
- lungmap.net**

## What is our goal? What are the Aims? LungMAP

- The overall objective of “Alveolar DevMAP” is to generate a compendium of the changes in epigenetic marks, microRNA, mRNA and proteins that happen during alveolar septation, and use this compendium to generate a dynamic temporal regulatory model of normal alveolar septation
- **Specific Aim 1** – to identify changes in coding and non-coding RNAs during alveolar septation
- **Specific Aim 2** – to determine changes in global DNA methylation during alveolar septation
- **Specific Aim 3** – To identify the shifts in transcription factor and proteomic profile during alveolar septation
- **Specific Aim 4** – To use, extend and validate our analytical tools to model dynamic signaling and regulatory networks activated in lung development that will be shared with other members of the consortium

**lungmap.net**

## Data/deliverable generated



	Description	Activity/Deliverable/Milestone	Status
1	Participation in weekly MPI videoconference calls on GoToMeeting	Activity	Ongoing
2	Participation in subcommittees: Ontology: Ambalavanan, Hagood Imaging: Ambalavanan Bioinformatics: Hagood, Kaminski, Bar-Joseph	Activity	Ongoing
3	Uploading of sample data to LungMAP SFTP site	Activity/Deliverable	Completed (9/25/2014)
4	Development of SOPs for mouse breeding, LCM, Proteomics, RNA-Seq, and DNA methylation	Activity/Deliverable	Completed (11/26/2014)
5	Testing of Laser Capture Microdissection (LCM) sample collection, shipping, analysis of LCM sample quality and quantity	Activity/Deliverable	Completed (9/24/2014)
6	Time course analysis of alveolar septation: Collection of mouse lung samples from P1-P28 (q12h from P1-P14, q24 from P15-P28), OCT inflated and frozen	Activity/Deliverable (all samples available: a total of 126 mouse lungs from 42 different time points; 3 per time point)	Completed (10/30/2014)
7	Time course analysis of alveolar septation: LCM of collected samples	Activity/Deliverable (samples shipped to Yale and UCSD)	Completed (1/2/2015)
8	Time course analysis of alveolar septation: mRNA expression and DNA methylation	Deliverable (Nanostring data available; with DCC) Milestone	Completed (2/20/2015)
9	Time course analysis of alveolar septation: Identification of optimal time points	Deliverable (optimal time points identified)/Milestone	Completed (2/20/2015)
10	MicroRNA profiling in mouse lung development	Deliverable (data available)	Completed (7/11/2015)

**lungmap.net**

## What time points should we evaluate LungMAP in mouse alveolar septation?

Objective: Identify optimal time points for analysis during mouse alveolar septation

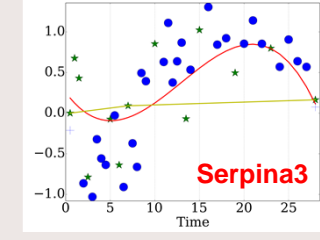
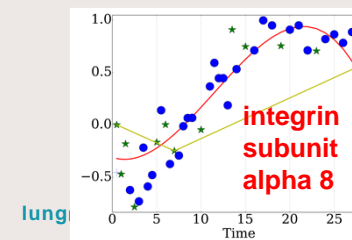
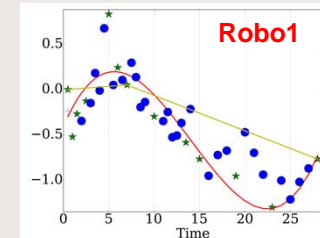
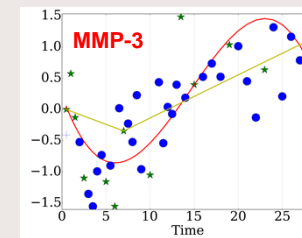
Methods:

- 1) LCM of alveolar regions excluding larger airways and blood vessels
- 2) High resolution time course profiling of selected genes (126) by nanostring (P0.5 to P14 every 12h, P15-P28 every 24h) + E16.5 (43 time points; n=4 per time point for most)
- 3) DNA methylation analysis by targeted next-gen bisulfite sequencing (n=3/time point)

[lungmap.net](http://lungmap.net)

## Why do a time course? 3 vs. 42 time points LungMAP

Movie vs. static images; longitudinal cohort study vs. cross-sectional study



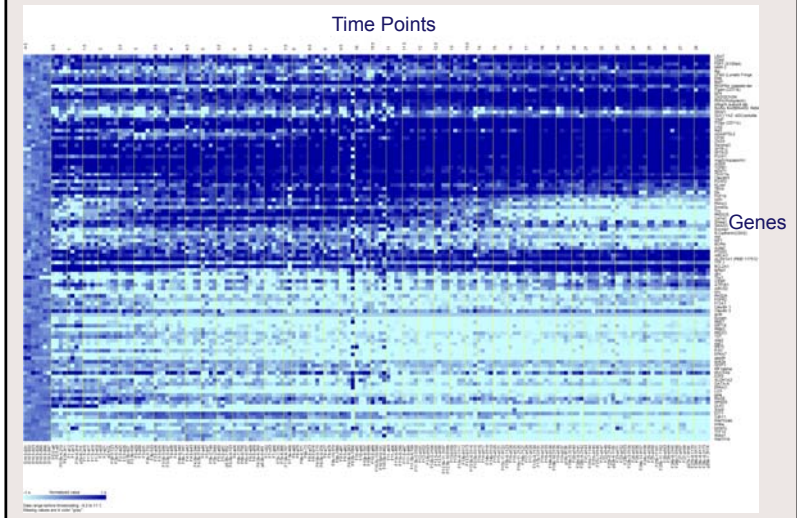
## Gene list for nanostring analyses of time course samples LungMAP

- EC: PECAM1, VCAM1, ICAM2
- EP C: Epcam, E-Cadherin (CDH1), KRT18, CLDN1 (Epithelial-Endothelial Adhesion), MUC5Ac
- MSC: VIM, ABCG2
- Basal: KRT5, P-63
- TJ/Adhesion: N-Cadherin (CDH2) (Heterogeneous cell type), CLDN1
- AT2: SFTPC, SFTPD, ABCA3
- AT1: PDPn (Podoplanin), Aqp5 (Aquaporin), RAGE, AGER
- Pericytes:
  - PDGFRB, FOXD1, HIF1a
- Fibroblast:
  - FSP1 (S100a4), Col1a1, Acta2, Thy1
- Monocyte:
  - CD68, ITGAX (CD11c), ITGAM (CD11b), CHI3L1
- Increased during septation:
  - INMT, Igfbp3, MMP3, LRAT, CHI3L3, KLRA4, SERPINA3
- Decreased during septation:
  - DLK1, PRSS35, HIST1H1B, HIST2AB, MCP14
- Altered in DNA methylation during development:
  - Lrp2, Akt, Igfbp3, Src, Vegf, Wif1, Cdh11, Eln, Tnc, Ctnna2, Robo1, Sox9, DNMT3a, Shisa3, Kcnma1, ADAMTSL2, E2F8, TBX4, EXT1
- Altered DNA methylation in fibrosis:
  - Lox, IGFBP3, Foxf2, Zfp536
- Genes known to be involved in septation:
  - PDGFRA, GUCY1A2 (sGC), LFNG, FGF18, ALDH1A1, RARA, VEGFA, FLG1, KDR, ELN, TGFB1, TGFB2
- miRNA regulated during septation:
  - IGF1, NOL3, PTGS2
- Transcription factors:
  - TTF-1, C/EBP, NFATC3, FOXA2
- Other interesting genes:
  - MMP7, MMP9, GATA-6, ERB, SMAD3, FOXF1, Integrin  $\alpha$ 8

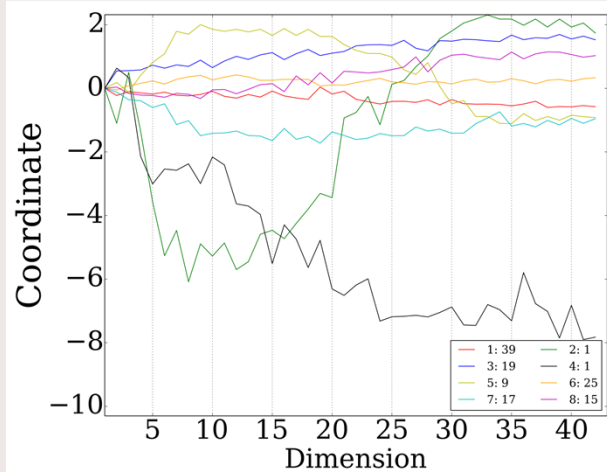
Red = targeted bisulfite Seq

[lungmap.net](http://lungmap.net)

## Heat map of gene expression LungMAP

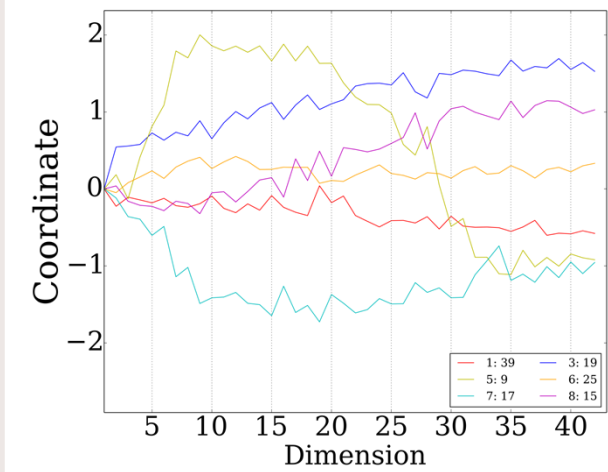


## Clusters of gene expression



lungmap.net

## Stable clusters of gene expression



lungmap.net

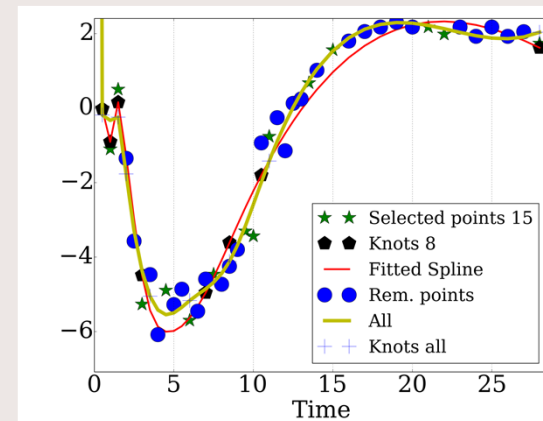
## Time course data



- All genes:
  - <https://github.com/emresefer/geneexpress/tree/master/newdata/splineplots15>
- Clusters:
  - <https://github.com/emresefer/geneexpress/blob/master/newdata/centplot.png>
- Stable clusters:
  - <https://github.com/emresefer/geneexpress/blob/master/newdata/centplotlimit.png>

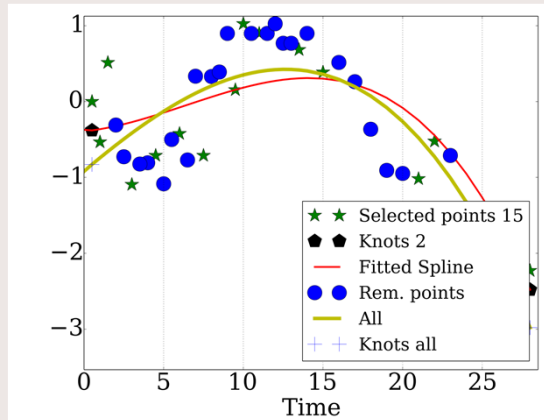
lungmap.net

## Examples of time course data - INMT



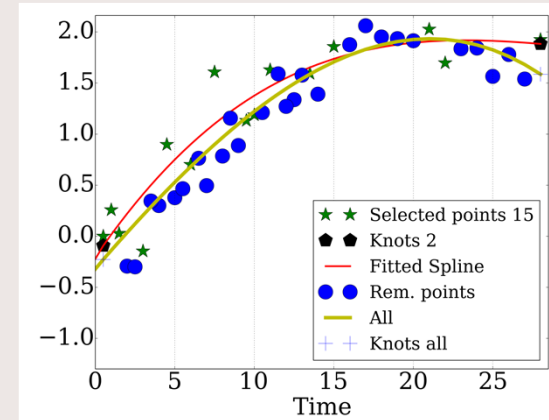
lungmap.net

## Examples of time course data -ELN



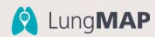
lungmap.net

## Examples of time course data -PDGFRA



lungmap.net

## DNA Methylation Analysis



lungmap.net

Targeted Next-Gen Bisulfite Sequencing (tNGBS)  
Multiplex PCR Optimization

13 genes 42 regions of interest submitted to EpigenDx



Local database established for downstream alignment



42 individual assays (simplex design)



## Multiplex Optimization:

1. Ta gradient
2. [Salt]
3. # of plex and primer concentration
4. Additives and their concentration
5. Input bsDNA titration



## Multiplex Assay Design:

1. PCR primers
  - Nature of primers: Tm, GC contents,  $\Delta G$
  - Primer-primer interactions: dimers, mis-priming
2. PCR Templates: size, GC contents, 2° structures, template-template interactions

lungmap.net

## Targeted Next-Generation Bisulfite Sequencing (tNGBS) Amplicon Information



Gene	Assay ID	Read Tm	Rev Tm	% GC	Size	CpGs	Coordinates (GRCh38/mm10)
Mouse Igfbp3	AD55132	67.8	69.2	28.7	289	9	chr11:7214215-7214499
Mouse Igfbp3	AD55133	68.3	68.5	26.1	241	8	chr11:7212803-7213043
Mouse Igfbp3	AD55134	69.3	69.6	32.1	240	7	chr11:7202949-7204208
Mouse Igfbp3	AD53301	70.5	70	33	176	7	chr11:7208306-7208481
Mouse Src	AD54514	64.8	65.9	35.9	103	26	chr2:15742925-157424027
Mouse Src	AD54515	66.5	68.8	37.6	165	23	chr2:157407351-157457630
Mouse Wif1	AD54516	65.5	65.6	33.7	172	23	chr2:157446741-157449912
Mouse Wif1	AD53302	69.4	68.7	31.3	297	11	chr10:121033395-121033691
Mouse Wif1	AD53303	60.9	60.1	31.8	190	10	chr10:121032800-121032997
Mouse Wif1	AD53304	66.6	67.5	24.8	222	9	chr10:121099752-121099973
Mouse Lox	AD54511	64.7	66.4	21.2	137	2	chr18:52530080-52530216
Mouse Lox	AD54512	69	70.9	31.3	131	10	chr18:52529184-52529315
Mouse Lox	AD54513	65.7	64.8	28.5	137	5	chr18:52528887-52527023
Mouse Vegfa	AD53334	64.7	69.1	36.6	227	12	chr17:46022351-46022577
Mouse Vegfa	AD53335	64.6	63.8	29.5	285	12	chr17:460225336-460225630
Mouse Vegfa	AD53336	67.2	66.9	32.6	138	8	chr17:46018598-46018735
Mouse Dnmt3a	AD53328	68.6	67.7	47.1	155	11	chr12:38040205-38040459
Mouse Dnmt3a	AD5632	64	64.7	32.2	211	5	chr12:3814382-3814952
Mouse Dnmt3a	AD53328	64.7	64	31.8	220	12	chr12:3901545-3901784
Mouse Dnmt3a	AD53329	66.8	66.1	25.4	252	13	chr12:3901514-3901785
Mouse Eln	AD5624	65	67.4	42.6	195	7	chr5:134747412-134747606
Mouse Eln	AD53309	64.1	67.4	37.8	307	9	chr5:134729221-134729526
Mouse Eln	AD53319	67.1	67.9	47.8	255	4	chr5:134731191-134731447
Mouse Cdh11	AD53307	69.1	71.3	29.9	194	7	chr8:102785456-102785649
Mouse Cdh11	AD53318	64.1	69.7	37	154	4	chr8:102784569-102784722
Mouse Cdh11	AD53308	66.8	68.3	36.9	130	4	chr8:102877609-102877766
Mouse Sox9	AD5796	61	66.1	31.4	169	9	chr11:112781641-112781811
Mouse Sox9	AD53311	69.7	68.5	34.7	248	7	chr11:112783358-112783605
Mouse Sox9	AD53310	66.4	63.1	26.2	126	9	chr11:112784760-112784885
Mouse Tnc	AD53322	62.7	63.9	26.7	116	6	chr4:64047034-64047149
Mouse Tnc	AD53323	65	67.5	35.2	244	13	chr4:64017478-64017721
Mouse Tnc	AD53325	62.5	61.6	20.2	188	4	chr4:63982799-63983086
Mouse Tnc	AD53324	63.3	62.2	23	174	6	chr4:63982645-63982818
Mouse Akt1	AD53331	67.3	66.3	32.5	184	5	chr12:112674444-112674627
Mouse Akt1	AD53332	68.3	69.8	38.3	154	5	chr12:112657120-112657273
Mouse Akt1	AD53333	68	65.5	31.5	162	6	chr12:112654548-112654709
Mouse Foxf2	AD54505	63.1	65	42.5	87	5	chr13:31625470-31625558
Mouse Foxf2	AD54506	65.1	64.8	37.9	190	23	chr13:31625904-31626093
Mouse Foxf2	AD54507	68	68.9	28.1	236	12	chr13:31632481-31632716
Mouse Zfp536	AD54508	69.2	67.1	46.5	101	7	chr7:37770516-37770556
Mouse Zfp536	AD54509	68.4	69.9	35.4	158	10	chr7:37567973-37568130
Mouse Zfp536	AD54510	65.4	67.4	35.9	156	6	chr7:37473451-37473606

lungmap.net

## Mouse Zfp536 3'Downstream ADS4510 – Results



Mouse Zfp536 3'Downstream - ADS4510 - Samples Analysis						
	96114	96077	96478	96468	96489	96116
	296493	296490	296461	296461	296423	296489
	chr7:37473476	chr7:37473488	chr7:37473512	chr7:37473522	chr7:37473545	chr7:37473524
Sample ID	CpG131315 (new)	CpG131316 (new)	CpG131317	CpG131318	CpG131319	CpG131320
% Methylation						
JH1	40.2	38.6	26.0	41.6	39.4	38.3
JH2	46.6	45.0	43.1	46.9	51.4	41.1
JH3	46.2	46.5	37.4	43.2	37.3	35.9
JH4	38.3	38.9	30.9	38.4	34.8	25.7
JH5	49.7	54.1	48.1	53.2	48.0	48.3
JH6	32.7	39.7	30.6	40.5	50.0	44.0
JH7	46.1	49.6	44.1	49.8	56.9	36.0
JH8	31.7	41.3	34.4	28.4	27.4	21.8
JH9	52.3	47.7	35.3	49.6	46.9	56.0
JH10	48.8	40.0	54.8	49.6	73.9	52.3
JH11	41.7	43.9	26.1	41.3	40.0	28.3
JH12	50.9	42.8	36.6	44.4	42.2	29.2
JH13	33.8	33.1	26.4	33.9	40.5	35.9
JH14	37.8	38.8	37.2	36.4	44.2	36.2
JH15	57.9	52.8	49.0	50.2	52.4	30.9
JH16	32.7	31.5	36.5	35.9	48.2	28.4
JH17	66.9	55.5	48.4	48.0	44.4	45.1
JH18	45.8	56.6	41.7	54.8	49.4	57.5
# Total Reads						
JH1	189	189	223	149	109	107
JH2	191	191	232	179	146	129
JH3	186	187	238	190	153	135
JH4	167	167	207	151	112	105
JH5	173	172	233	220	179	163
JH6	223	224	278	215	164	150
JH7	219	206	266	235	186	175
JH8	104	104	125	88	62	55
JH9	132	132	167	115	96	91
JH10	125	125	146	123	88	65
JH11	132	132	157	121	98	92
JH12	173	173	213	151	116	106
JH13	139	139	159	109	84	78
JH14	219	219	274	206	165	149
JH15	254	254	300	217	147	136
JH16	107	108	137	103	81	81
JH17	130	131	182	179	151	142
JH18	83	83	115	93	79	73

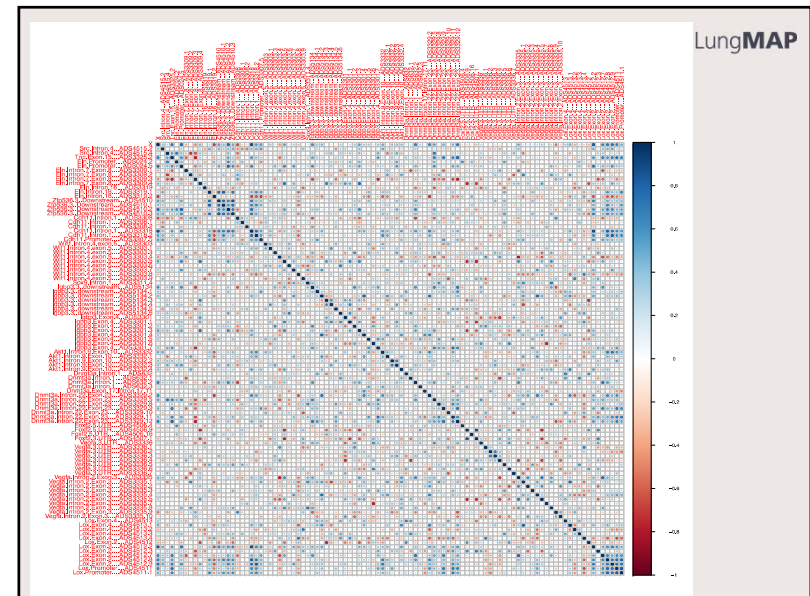
lungmap.net

## Summary of Targeted Bisulfite



- Genes: IGFBP3, Src, Wif1, Lox, VEGFA, Dnmt3, Eln, Cdh11, Sox9, Tnc, Akt1, Foxf2, Zfp536
- CpG regions in promoter, gene body, 3'UTR if possible
- Days 0.5, 2.5, 5, 10, 26
- 3 animals at each time point

lungmap.net







## Time course analysis



- Overall, preliminary data has error (in terms of variance of repeated measurements) that is very good overall, and the data is highly reproducible
- Overall variance of 0.2926 (everything excluding 9.5 and 10.5) = optimal limit of what we can do (obviously cannot go below the experimental / biological noise in our predictions)
- Optimal number of time points calculated based on keeping error at a minimum
  - 15 time points: error ~0.39
  - 9 time points: error ~0.44
  - 3 time points (P0, P7, P28): error ~

lungmap.net

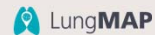
## Optimal time points:



- Combining mRNA and DNA methylation:
  - Fixed “anchor” time points of E16.5, E18.5, P7, P28
  - E16.5, E18.5, P0.5, 1.0, 1.5, 2.5, 4, 5, 7, 10, 13.5, 15, 19, 23 and 28 (error: 0.392)

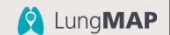
lungmap.net

## Single cell analysis (Ziv/Aaron)



lungmap.net

## Reconstructing developmental networks from time series single cell data



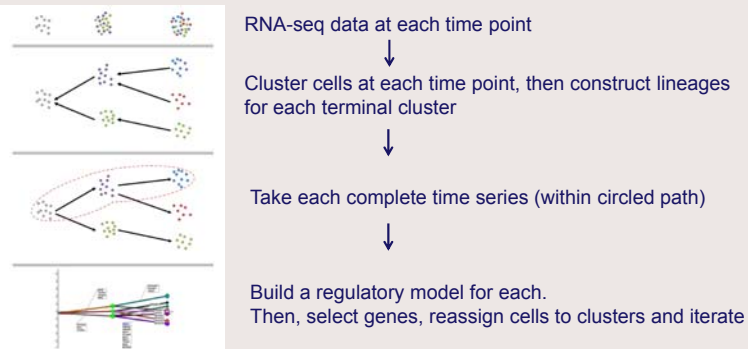
- Single Cell Analysis of REgulatory DYnamics through Computational Assembly of Time series (SCAREDY-CAT)
- Cells at each time point usually come from a mixture of cell types, each of which may be a progenitor of one, or several, specific cell fates
- To determine the regulatory networks controlling cell differentiation we first need to reconstruct the ‘time series’ of cells and then identify key regulatory events that differ between models for different cell fates
- SCAREDY-CAT utilizes expression similarity within and across time and integrates this data with regulatory information using a probabilistic graphical model which iterates between reconstructing different regulatory networks and assigning single cells to these networks (Expectation Maximization (EM) approach to combine clustering and Hidden Markov models)

lungmap.net

## Single cell analysis -2



- Treutlein et al. 2014 : Single cell epithelial data from E14.5, E16.5, and E18.5 used
- CCHMC data at E16.5 used as validation set



## Differentially active transcription factors (Top 20) and their targets in cell types compared to baseline (E14.5)

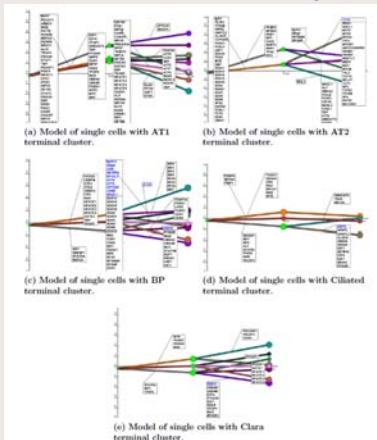


TF	AT1	AT2	BP	Ciliated	Clara
ZFP238	very high	very high	very high	none	none
ZIC2	very high	very high	very high	very high	very high
MAZ	very high	very high	very high	very high	very high
PAX1	very high	high	very high	very high	very high
GTF2I	very high	high	very high	very high	high
ERF	high	very high	high	high	very high
ZBTB7A	high	high	none	none	none
SRF	very high	none	very high	none	none
IRF9	very high	very high	very high	high	high
ASCL1	none	high	high	very low	very low
AR	high	very low	very high	none	none
BCL6	very high	none	very high	none	none
TCF4	none	low	none	very low	very low
STAT5B	high	very low	very low	none	low
PAX8	very low	high	none	none	none
FOS	very low	none	very low	none	none
EGR1	very high	very high	very high	high	high
NR2F1	low	high	low	high	high
TCF3	none	high	none	very low	none
FOSB	low	very high	very low	very high	high

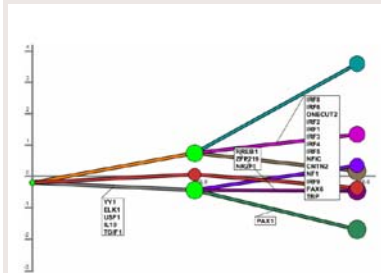
Very high/Very low:  
Log2 fold change >1

High/low:  
Log2 fold change >0.5

## Dynamic regulatory models from supervised marker based cell assignments



Model of single cells built from 16 AT1, 1 AT2, and 9 BP cells at the terminal endpoint



lungmap.net

## MicroRNA profiling



- Nanostring analysis at 43 time points:
  - E16.5 + 42 time points (P0.5-P14 q12; P15-P28 q24)
- At least 3 samples (separate mouse lungs) at each time point
- ~600 mouse miRNA profiled + house-keeping+ spike-in + positive and negative controls

lungmap.net



## MicroRNA profiling



lungmap.net

## MicroRNA profiling



lungmap.net

## MicroRNA profiling



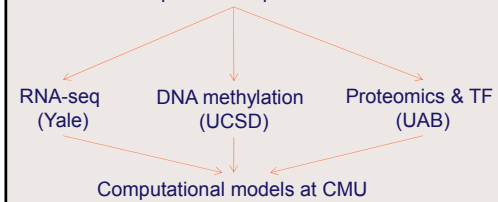
lungmap.net

## What next?



### Experiments and Analyses

LCM at optimal time points at UAB



### Timeline

Completed; July 2015

Aug-Nov 2015

Dec 2015-March 2016

- Data integration (combining RNA-seq, DNA methylation, proteomics, TF)
- Coordination of efforts
- Mouse to human translation
- Integration with BREATH

lungmap.net

## Plan for integrating the data



- Within-center integration
  - Integration of
    - RNA-seq, miRNA data from Yale*
    - DNA methylation data from UCSD*
    - Proteomic and TF data from UAB*
    - Computational models from CMU*
- Between-centers integration
  - Integration of UAB/UCSD/Yale/CMU data and analyses with those from other RCs and HTC/DCC

lungmap.net

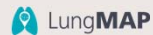
## What our Research Center adds to what is being done by other centers



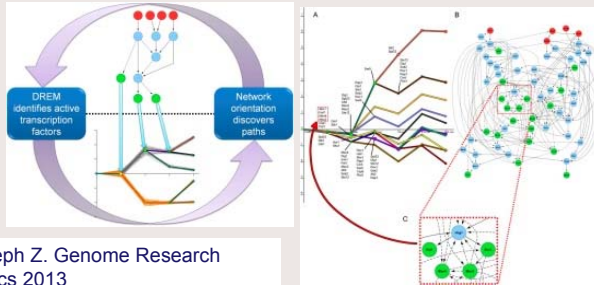
- Low input, high throughput Epigenomics (DNA methylation)
- Focus on noncoding RNAs (microRNAs and lncRNAs)
- Focus on *in situ* lung microenvironment transcriptomics and epigenomics (laser capture microdissection)
- Transcriptional factor profiling in combination with proteomics
- Development of dynamic regulatory models to really understand biological “breakpoints” and watershed moments in alveolarization
- A unique multidisciplinary team that has both genomic, computational and clinical expertise in lung disease from fetus, through child to adulthood

lungmap.net

## Within-center integration



- SDREM (Signaling and Dynamic Regulatory Events Miner):
  - Extension of DREM that allows the joint modeling of signaling and dynamic regulatory networks by linking sources (proteins that initiate signaling cascades) with TF targets that regulate gene expression
  - Able to identify both key signaling proteins (master regulators) and key transcription factors that are activated by paths involving these signaling proteins and in turn activate downstream expression programs



Gitter A...Bar-Joseph Z. Genome Research 2013; Bioinformatics 2013

## Coordination of efforts -1



- With CCHMC
  - E16.5 and E18.5 lungs from CCHMC for LCM and downstream analysis
  - Single cell analysis- computational models (Ziv/Aaron)
  - Cellular localization of molecules identified as master regulators (using single cell analysis data e.g. LungGENS, then validation using flow sorted cells)
- With PNNL & TACC
  - LCM samples sent to PNNL at “anchor” time points for analysis
  - Spatial localization of molecules identified as master regulators (using *in situ* proteomics and ISH)

lungmap.net

## Coordination of efforts -2



- With CHLA & USC
  - Combine the imaging with the systems biology approaches to determine:
    - The regulatory events at (or just before) the time points when specific structural changes are noted
    - The structural changes (e.g. alveolar volume, alveolar complexity, microvascular changes) at (or just after) the time points when specific regulatory events are noted

lungmap.net

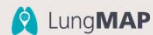
## Translation from mouse to human



- Working with HTC
  - Evaluation of OCT:PBS vs. CMC samples of human lung
    - Will evaluate RNA and DNA quality, mRNA and DNA methylation of selected genes, Proteomics
  - Will do analysis similar to mouse lung samples (RNA-seq, DNA methylation, Proteomics, TF profiling, then computational modeling)
  - Will compare and contrast mouse vs. human lung development in each of the analyses
    - Similar events vs. Dis-similar processes

lungmap.net

## Fitting into BREATH and interaction with data -1



- Data that is/will be available:
  - mRNA (RNA-seq, Nanostring), miRNA (Nanostring), lncRNA (RNA-seq)
  - Proteomics & TF
  - DNA methylation
- Linkage of:
  - mRNA (LCM: time course) with mRNA (single cell: snapshots), with corresponding DNA methylation, protein expression (PNNL and UAB) and histology
  - Corresponding regulatory profile (e.g. miRNA and TF that target a mRNA and how they are expressed)
  - Imaging at time points with regulatory events at those time points

lungmap.net

## Fitting into BREATH and interaction with data -2 (examples)



- One could go from:
  - single cell RNA at specific time points to
  - immunostaining/ISH to
  - time-course mRNA from P0.5 to P28, and
  - determine regulation (miRNA, DNA methylation, TF) of the molecule at different time points
- Viewers can see how the molecule changes expression during the course of alveolar septation, identify potential regulators, and see which other molecules share a similar expression profile.

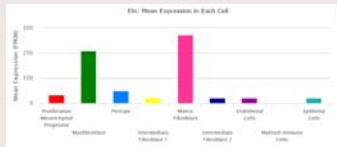
lungmap.net

## Fitting into BREATH and interaction with data -3 (example)

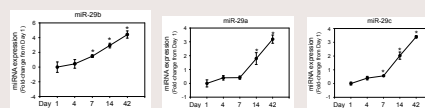


### ■ Elastin (ELN)

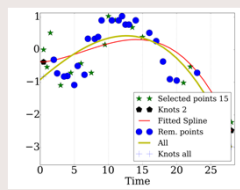
LungGENS E16.5



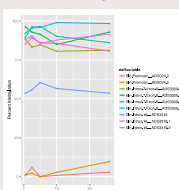
miRNA targeting ELN: miR-29b,a,c



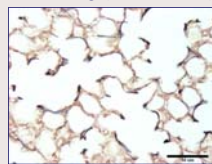
mRNA



DNA methylation



Staining (localization)



lungmap.net

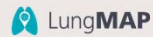
## Issues to think about



Priority: Generation of high-quality reproducible biologically relevant data for our users

1. Do we want to stick with E16.5, E18.5, P7, and P28 for other RCs (or) add selected time points for better integration with the dynamic regulatory models?
2. Do we prioritize specific time points in human lung or do all available samples?
3. Evaluation of DNA methylation and miRNA/lncRNA in sorted cells from different time points?
4. Evaluation of histone modifications (ChIP-seq for histone modifications of choice) (large number to choose from, will drive up cost)
5. Newer technologies and platforms? (Evolution over course of project)

lungmap.net



lungmap.net

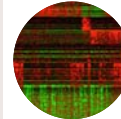
**LungMAP**  
Coordinating Center  
Duke & RTI

lungmap.net

Sponsor  
**NIH** National Heart, Lung, and Blood Institute

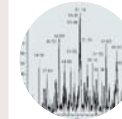
### RESEARCH CENTERS:

**CCHMC**  
Cincinnati



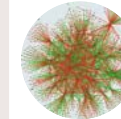
Single Cell  
Transcriptome  
Confocal

**PNNL & TACC**  
Seattle & Austin



Proteomics  
Metabolomics

**UAB & Yale  
& Pitt & UCSD**



Systems  
Biology

**Saban Inst.  
CHLA & USC**



3-D  
Imaging

**URMC**  
Rochester



Human  
Lung

