8-1-2019

Temporal gene expression of mesenchymal cells in the pediatric lung

Quinlen F. Marshall
College of Saint Benedict/Saint John's University, qmarshall001@csbsju.edu

Soumyaroop Bhattacharya

Gautam Bandyopadhyay

Ravi Misra

Thomas Mariani

See next page for additional authors

Recommended Citation

Authors

Quinlen F. Marshall, Soumyaroop Bhattacharya, Gautam Bandyopadhyay, Ravi Misra, Thomas Mariani, and Gloria Pryhuber

Temporal Gene Expression of Mesenchymal Cells in the Pediatric Lung

Quinlen F. Marshall,[1,2] Soumyaroop Bhattacharya, M.S.,[3,4] Gautam Bandyopadhyay, Ph.D.,[3]
Ravi Misra, Ph.D.,[3] Thomas Mariani, Ph.D.,[3,4] Gloria Pryhuber, M.D.[3]

[1]College of Saint Benedict and Saint John's University, Collegeville, Minnesota; [2]Strong Children's Research Center, University of Rochester Medical Center, Rochester, New York; [3]Division of Neonatology, Department of Pediatrics, University of Rochester Medical Center, Rochester, New York; [4]Program in Pediatric Molecular and Personalized Medicine, Department of Pediatrics, University of Rochester, Rochester, New York.

ABSTRACT

INTRODUCTION: The newborn lung undergoes vast biochemical and physiological changes during adaptation from the intrauterine to the extrauterine environment. Lung morphogenesis continues from birth into early childhood, mediated by dynamic gene expression and a diversity of pulmonary cell types (Whitsett, JA. et al. *Physiol. Rev*, 2019). Murine models demonstrate that pulmonary mesenchymal cells exhibit remarkable heterogeneity in function and morphology during development, however, confirmation of their role is lacking in human neonates and early childhood (Guo, M. et al. *Nat. Comm*, 2019). In addition, many current human genomic studies of lung maturation suffer from limited sample size, limiting their applicability to longitudinal pediatric lung development. Temporal analysis of gene expression aims to bridge this gap, and the most common analytical approach utilizes Short Time-series Expression Miner (STEM) (Ernst, J. & Bar-Joseph, Z. *BMC Bioinformatics*, 2006). STEM utilizes unique methods to cluster, compare, and visualize short time-series gene expression data.

METHODS: Dissociation of lung cells, sorting into enriched populations, and RNA isolation was performed at the Human Tissue Core of the Molecular Atlas of Lung Development Program (Bandyopadhyay, G. et al. *Am. J. Physiol. Lung Cell Mol. Physiol*, 2018). RNA sequencing (RNAseq) was performed at the University of Rochester Genomics Research Center using the Ilumina NovaSeq6000, and reads were aligned using the Splice Transcript Alignment to a Reference algorithm (STaR). Reads were further normalized using counts per million (CPM) and variance-mean dependence calculated with DESeq as implemented in Bioconductor. Genes not detected in at least 3 time points or exhibiting a minimum fold change of at least 3 across the time series were excluded from further analysis. Time-series analysis was performed with STEM, and profiles were assigned significance by Fisher's exact test (p<0.05). Genes selected from profiles of interest were functionally enriched using ToppGene Functional Gene Enricher (Chen, J. et al. *BMC Bioinformatics*, 2007).

RESULTS: RNAseq was performed using RNA obtained from pulmonary mesenchymal cells, (n=24, (<1 d/o - 8 y/o, 17 m, 7 f) generating 24.3±5.5 million reads at depth of 10 million reads (48.3±4.6% of genome mapped). CPM normalized expression values for repeat donor time points were averaged and then separated into a younger (n=9, <1 d/o - 1 y/o) and older (n=8, 1 y/o- 8 y/o) group. A total of 17,843 genes passed filtering criteria in the younger group and 17,840 passed in the older group. Using STEM, 16 and 20 profiles were found to be significant in the younger and older group, respectively. 7 profiles in the younger group and 8 in the older group were selected for further functional analysis based on significance and directionality of gene expression changes.
Multiple profiles in both groups demonstrated matrix fibroblast associated gene expression increasing in both groups, peaking at 2 years. Next, proliferative fibroblast and cell division associated gene expression decreased from birth to 1 year in the younger group. Detection of multiple mesenchymal- like profiles validates the purity of cells enriched. Additionally, gene expression associated with immune- like pathways increased in both groups. Finally, cell signatures in the older group associated with the Wnt pathway decreased from 1 year until 2 years and then increased from 4 years to 8 years.

CONCLUSIONS: In summary, analysis of dynamic gene expression in isolated cells across a time series demonstrates the unique heterogeneity of pulmonary mesenchymal cells throughout adolescence. In addition, increased gene expression associated with immune signatures during pediatric lung development was noted. Further validation and exploration using this technique may advance understanding of the diversity of pulmonary cell types and pathophysiology of pediatric lung disease.

BACKGROUND

- **Upon leaving the womb, the transition from fetus to newborn is one of the most complex adaptations that occurs during human life** (Hillman, N., et al. *Clin. Perinatol.* (2012).
- Involves transition from fluid environment to air environment where gas exchange must occur.
- Involves clearance of fetal lung fluid, surfactant secretion, and onset of regular air breathing.
- **Development of the lung proceeds through unique phases and involves gene regulation and dynamic cross talk between pulmonary cell types that uniquely contribute to the development of the lung** (Whitsett, JA. *Physiol. Rev.* (2019).
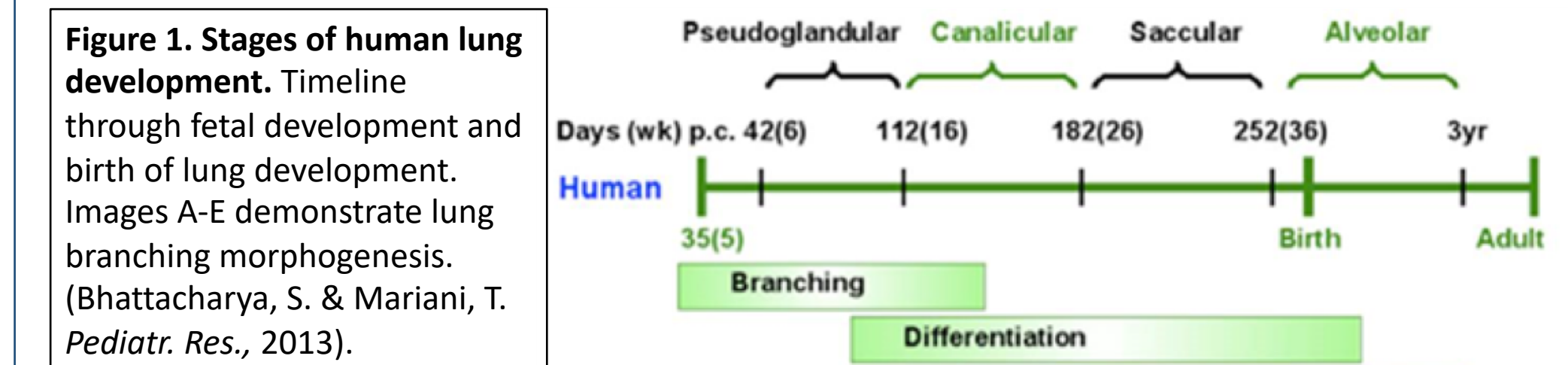
Figure 1. Stages of human lung development. Timeline through fetal development and birth of lung development. Images A-E demonstrate lung branching morphogenesis. (Bhattacharya, S. & Mariani, T. *Pediatr. Res.*, 2013).

- Broadly classified into epithelial, endothelial, mixed immune, and mesenchymal cells with numerous subtypes within each category.
- **Whole transcriptome RNA and single-cell RNA sequencing methods can characterize pulmonary cell populations during development in addition to demonstrating dynamic gene expression patterns.**
- Temporal analysis of genomics data utilizes samples across multiple time points and requires tailored analytical techniques dependent on number of time points.
- An analytical software available for short, time-series genomics data is the Java-based Short Time-series Expression Miner (STEM). STEM uses a method of analysis that takes advantage of large gene datasets and a small number of time points to identify significance.

ACKNOWLEDGEMENTS & REFERENCES

Bandyopadhyay, G. et al. Disassociation, cellular isolation, and initial molecular characterization of neonatal and pediatric human lung tissues. *Am. J. Physiol. Lung Cell Mol. Physiol.* **315**, 576-583 (2018).
Ernst, J. & Bar-Joseph, Z. STEM: a tool for the analysis of short time-series gene expression data. *BMC Bioinformatics.* **7**, 191 (2006).
Guo, M. et al. Single Cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. *Nat. Comm.* **10**, 37 (2019).
Hillman, N., Kallapur, SG. & Jobe, A. Physiology of Transition from Intrauterine to Extrauterine Life. *Clin. Perinatol.* **39**, 769-783 (2012).
Whitsett, JA., et al. Building and Regenerating the Lung Cell by Cell. *Physiol. Rev.* **99**, 513-554 (2019).

METHODS

Figure 2. Subject Demographics. Ages and gender distribution of subjects enrolled in RNAseq study. X-axis represents donor ID arranged in ascending order of age and y-axis represents age. Two adult donors (D071 and D116) were excluded from the current analysis. Pink and blue symbols indicate female and male donors, respectively.
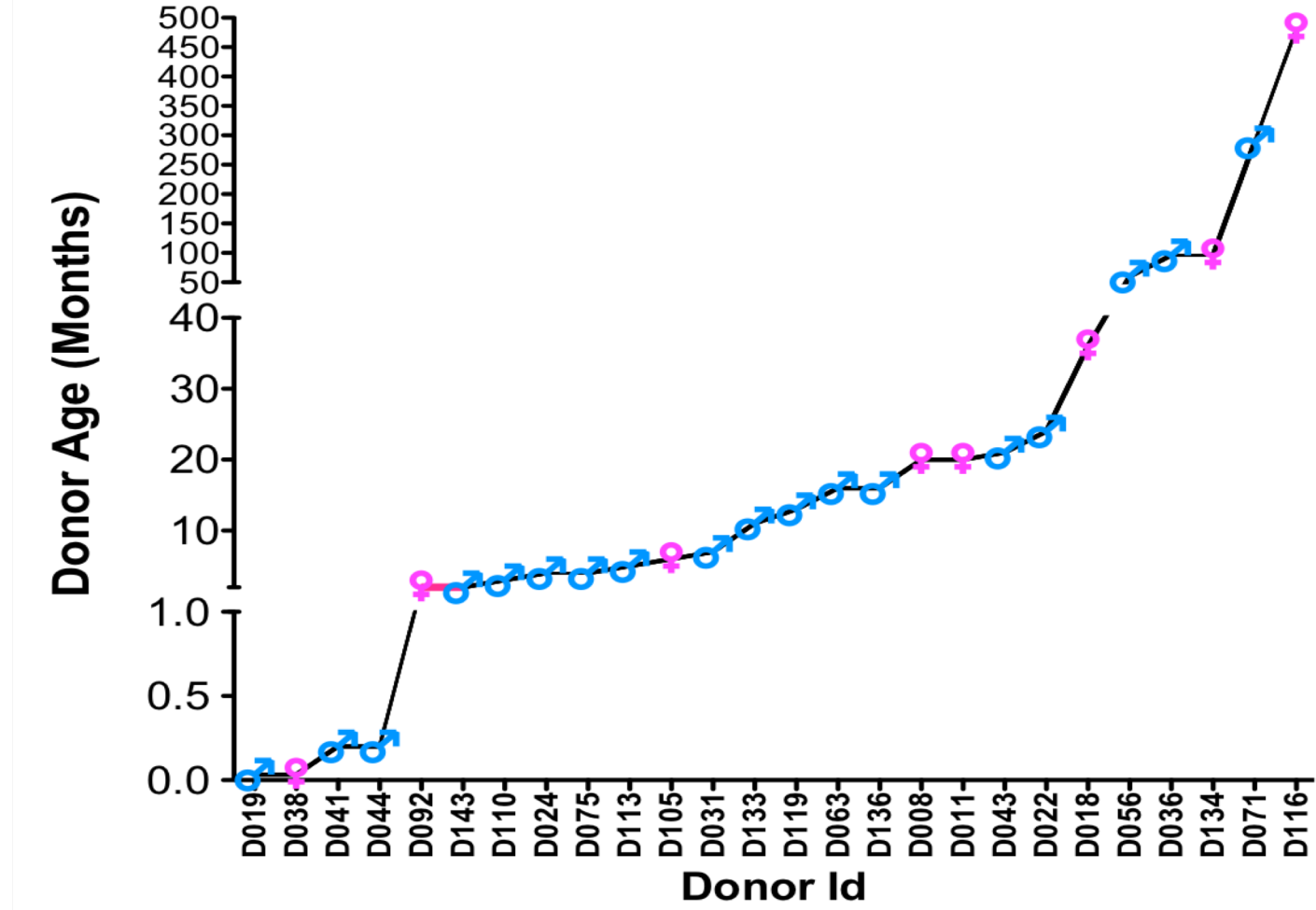
Figure 3. Percent transcriptome detected. Percent gene expression relative to total genes detected for 4 major pulmonary cell populations. X-axis represents donor ID arranged in ascending order of age and y-axis represents percent gene expression relative to total genes detected (n=56,872). Blue diamonds represent endothelial cells, orange squares represent epithelial cells, gray triangles represent mesenchymal cells, and yellow crosses represent mixed immune cells (MIC).
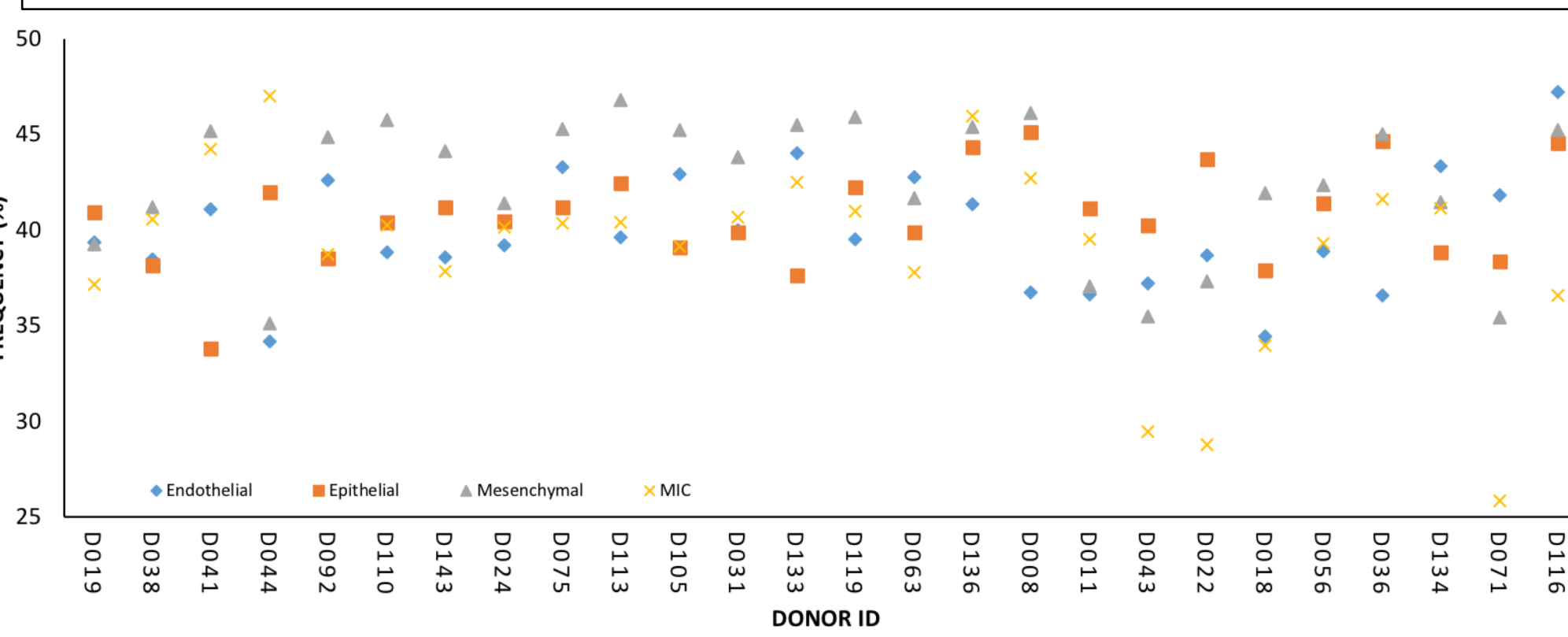
Dissociation of lung cells, sorting into enriched populations, and RNA isolation was performed at the Human Tissue Core of the Molecular Atlas of Lung Development Program by Bandyopadhyay, G. et al.

TISSUE DISSOCIATION
Right-upper and right-middle lung lobes were dissected and then further cut into small tissue. Next, they were placed in digestive cocktail on a GentleMACS Tissue Octo Dissociator. Proteases were neutralized and passed through a 100 μm strainer. Erythrocytes were then lysed and cells were washed with DPBS and counted. Cells were slow-cooled and transferred to N_2 (l) for long-term storage.

CELL SORTING
Thawed, dissociated mixed lung cell populations were incubated with membrane marker specific antibodies. Cell sorts were completed using a 4-laser, 18-color FACSAria flow cytometer. Antibody-stained dissociated cells were enriched for 4 major cell populations: CD45+ mixed immune, EpCam+ epithelial, CD31+CD144+ endothelial, and all negative mesenchymal cells.

RNA SEQUENCING
RNA sequencing was performed using the Ilumina NovaSeq6000 at the University of Rochester Genomics Center, and reads were aligned using the STaR Algorithm. Reads were further normalized using counts per million and variance-mean dependence calculated with DESeq as implemented in Bioconductor. Genes not detected in at least 3 time points or exhibiting a minimum fold change of at least 3 across the time series were excluded from further analysis. Transcripts detected are available on LungMAP.net and the Lung Gene Expression Analysis Web Portal.

DATA ANALYSIS WITH STEM
RNAseq was performed using RNA obtained from lung mesenchymal cells, (n=24, (<1 d/o - 8 y/o (17 m, 7 f) generating 24.3±5.5 million reads (depth of 10 million reads, 48.3±4.6% genome mapped). CPM normalized expression values for repeat time points were averaged and separated into a younger (n=9, <1 d/o – 1 y/o) and older (n=8, 1 y/o – 8 y/o) group. A total of 17,843 genes passed filtering criteria in the younger group and 17,840 passed in the older group. Time-series analysis was performed with STEM, and 16 and 20 profiles were significant in the younger and older group, respectively, using Fisher's exact test (p<0.05).

**WHERE SUMMER ANALYSIS BEGAN**

FUNCTIONAL ENRICHMENT
7 profiles in the younger group and 8 profiles in the older group were selected for further functional analysis based on significance and directionality of change of gene expression. Lists of genes from indicated profiles were functionally enriched using ToppGene Functional Gene Enricher (Chen, J. et al. *BMC Bioinformatics*, 2007) for molecular functions, biological processes, cellular components, pathways, and ToppCell Atlas reference gene lists.

RESULTS I

Figure 4. Cellular purity validation with known mesenchymal cell markers. Heat map of known mesenchymal cell marker genes (Bandyopadhyay, G. et al. *Am. J. Physiol. Lung Cell Mol. Physiol*, 2018), (Whitsett, JA. et al. *Physiol. Rev*, 2019): PDPN, PDGFRA, ELN, COL6A6, CYR61, COL1A1, RORA, COL1A2, A2M, ACTA2, COL5A1, PDGFRB, BNC2, SEMA5A, and COL6A1. Red indicates higher expression while green indicates lower expression.
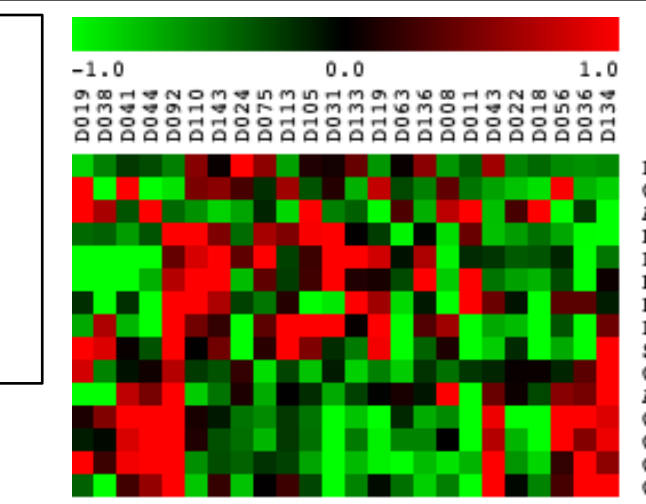
YOUNGER GROUP

Figure 5. Younger group STEM profiles. STEM profiles in decreasing significance order for younger (n=9, <1 d/o – 1 y/o) group. Colored profiles show significance by Fisher's exact test (p<0.05). Red stars indicate selection for functional enrichment.
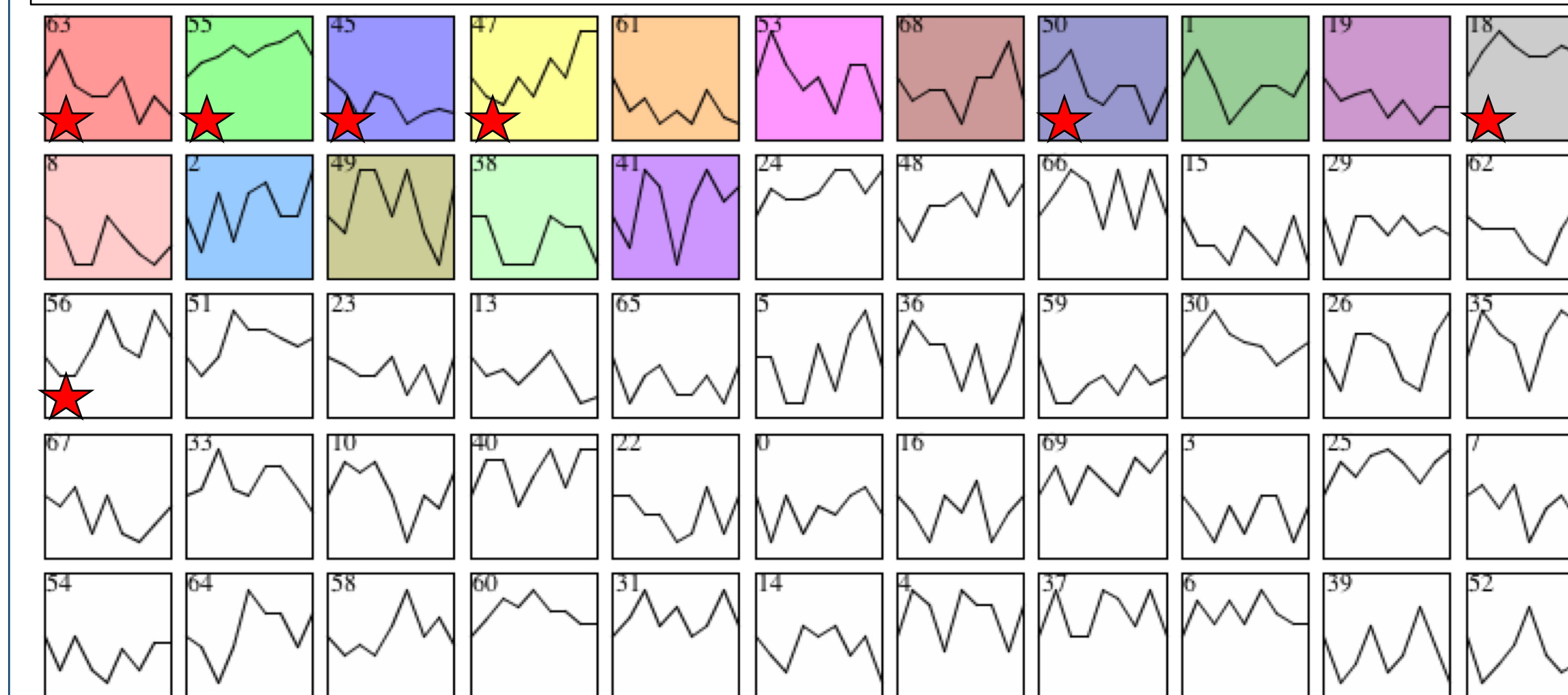
Table 1. Younger group functional enrichment. Functional enrichment results for younger (n=9, <1 d/o – 1 y/o) group. Profile number, general trend significance, and number of genes in a profile are given. Functional enrichment results for molecular functions, biological processes, cellular components, pathways, and ToppCell Atlas reference gene lists are shown.

PATTERN	TREND	SIGNIFICANCE (p-value)	NUMBER OF GENES	GENE ONTOLOGY	PATHWAY	CELL TYPE
63	Decrease	2.0E-219	1,207	Mitotic cell cycle, cell division, sister chromatid segregation, centrosome, kinetochore	Cell cycle and mitotic cell cycle	Proliferative fibroblast, immature myofibroblast
55	Increase	0.00	1,045	-	-	Mature fibroblast, matrix fibroblast subtypes
45	Decrease	7.3E-141	888	ATP synthesis coupled transport, mitochondrial protein complex, respiratory chain complex	Oxidative phosphorylation, respiratory electron transport	Mature fibroblast, matrix fibroblast subtypes
47	Increase	1.6E-58	683	Antigen binding, immune response, leukocyte activation, immunoglobulin complex	Neutrophil degranulation, innate & adaptive immune system	Myeloid cell type, B cell, lymphoid cell, dendritic cell
50	Decrease with variability	1.3E-39	469	Extracellular matrix & structure organization, interstitial matrix	Core extracellular matrix & collagen production	Myofibroblast & general mesenchymal subtypes
18	Increase with peak	1.3E-39	414	-	-	Myeloid cell type, neutrophil, proliferative myeloid
56	Increase with peaks	0.60	300	-	-	Matrix fibroblast associated subtypes

Figure 6. Select genes shared between profiles demonstrating immune signatures. Heat map of select genes found in profiles demonstrating immune signatures in both the younger group and older group. Red indicates higher expression while green indicates lower expression.
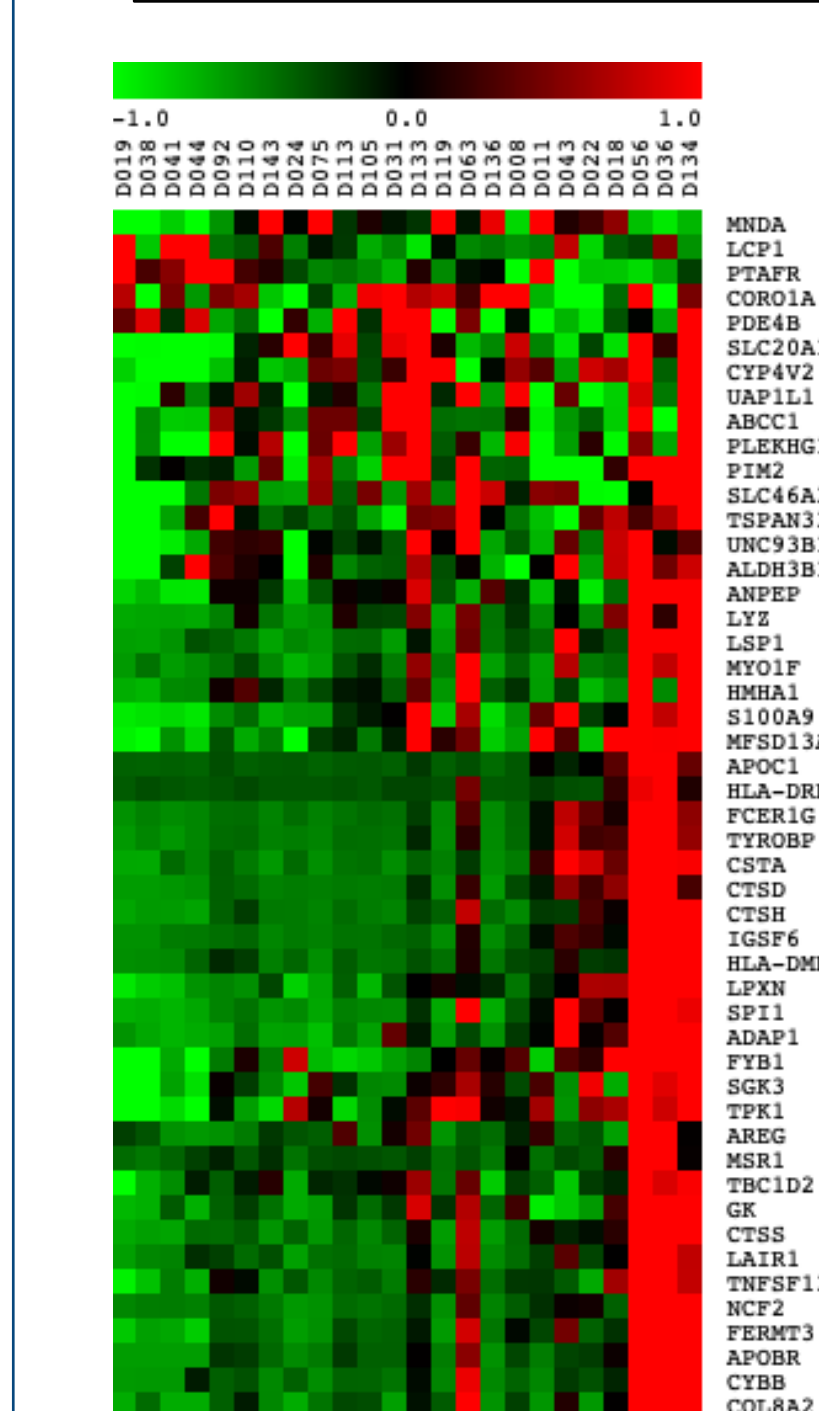
- **Gene expression associated with immune-like pathways (47, 18, 44, 4, 26, 16) increased in both groups. Immune like gene ontologies were also noted.**
  - Suggests dynamic immune response that may decrease after infancy and has considerable variability with individual donors. Contamination from immune cell type population is also possible.

- **Proliferative fibroblast and cell division associated gene expression (63) decreased from birth to 1 year in the younger group.**
  - May be due to decrease in alveolarization of the lung after the rapid fetal growth and development of the lungs.
- **Expression of genes associated with extracellular matrix components and production decreased in the younger group.**
  - These processes may decrease as development of the lung slows after rapid lung growth during growth of the fetus.

RESULTS II

OLDER GROUP

Figure 7. Older group STEM profiles. STEM profiles in decreasing significance order for older (n=8, 1 y/o - 8 y/o) group. Colored profiles indicate significance by Fisher's exact test (p<0.05). Red stars indicate selection for functional enrichment.
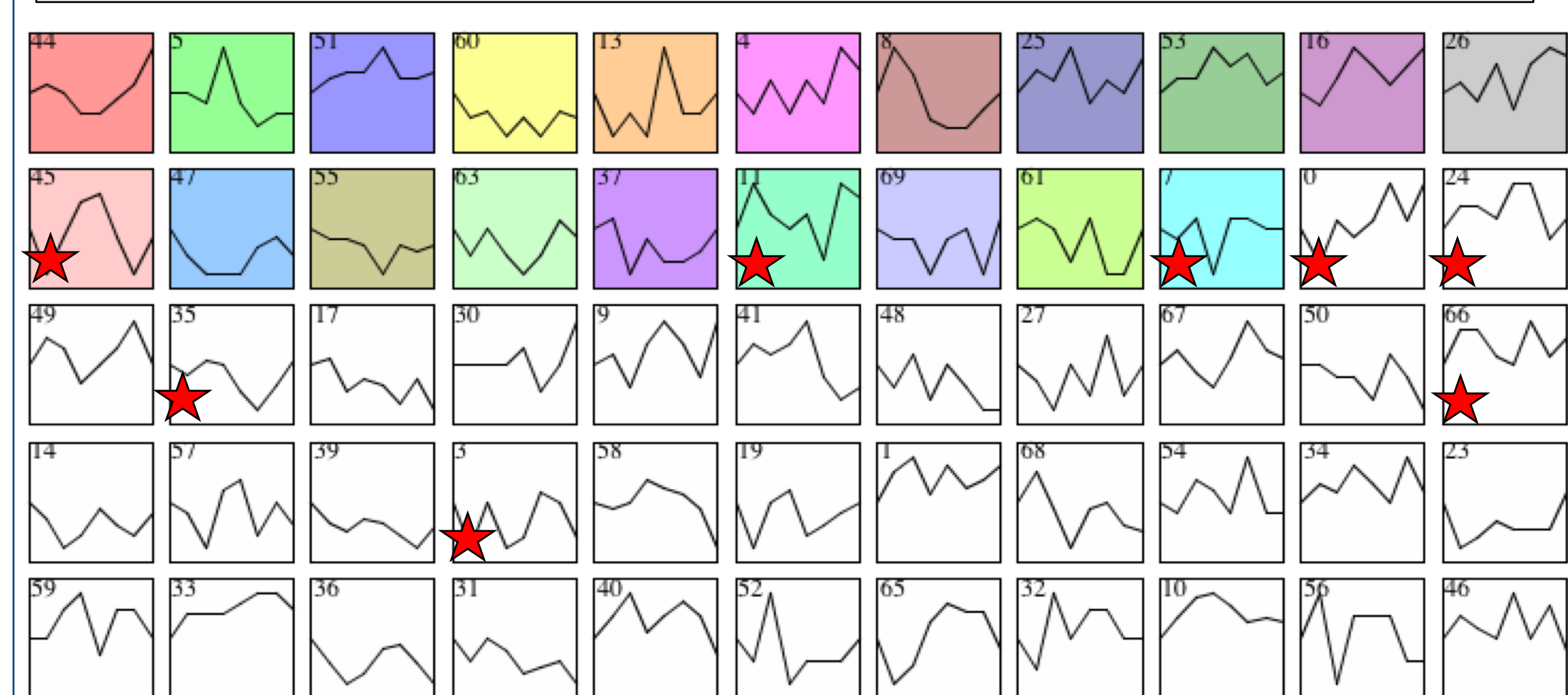
Table 2. Older group functional enrichment. Functional enrichment results for older (n=8, 1 y/o - 8 y/o) group. Profile number, general trend significance, and number of genes in a profile are given. Functional enrichment results for molecular functions, biological processes, cellular components, pathways, and ToppCell Atlas reference gene lists are shown.

PATTERN	TREND	SIGNIFICANCE (p-value)	NUMBER OF GENES	GENE ONTOLOGY	PATHWAY	CELL TYPE
44	Increase with initial peak	3.2E-270	1,754	Antigen binding, immunoglobulin receptor binding, B-cell mediated immunity	Hematopoietic cell lineage	-
4	Increase with variability	1.1E-12	620	Immune response, defense response, inflammatory response, MHC II receptor complex	Neutrophil degranulation, innate immune system, MHC II antigen presentation	Myeloid cell type, neutrophil, proliferative myeloid
53	Increase slightly with peak	1.4E-8	441	-	-	Matrix fibroblast subtypes
16	Increase with peak	2.2E-34	431	Proton transmembrane activity, electron transport chain	Oxidative phosphorylation and TCA cycle	Myeloid cell type, B cell, macrophage, immune monocyte
26	Increase with variability	1.5E-33	412	Immune response, regulation of immune system process, lysosome	Neutrophil degranulation, innate immune system	Myeloid cell type, macrophage subtypes, neutrophil
47	Decrease with later peak	7.9E-5	391	-	-	Wnt signaling pathway	-
24	Middle peak	7.2E-3	318	-	-	Pericyte associated subtypes
30	Increase later	0.32	472	-	-	Core extracellular matrix & enzymes responsible for remodeling matrix	Proliferative fibroblast and immune signatures

Figure 8. Select genes shared between profiles demonstrating fibroblast signatures. Heat map of select genes found in profiles demonstrating fibroblast signatures in both the younger group and older group. Red indicates higher expression while green indicates lower expression.
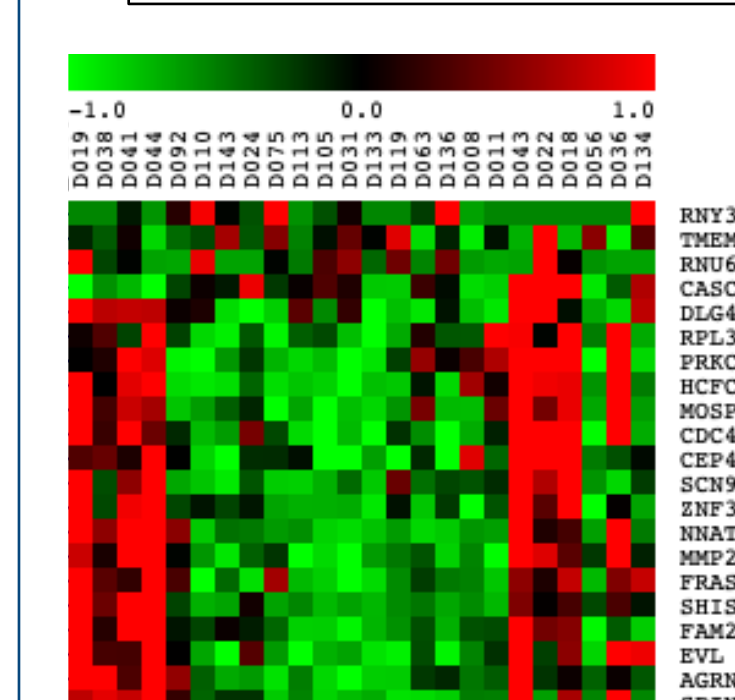
- **Multiple profiles (55, 56) demonstrated matrix fibroblast associated gene expression increasing peaking at 2 years.** Detection of multiple mesenchymal like profiles validates cell enrichment.

- **Cell signatures in the older group associated with the Wnt pathway decreased from 1 year until 2 years and then increased from 4 years to 8 years.**
- Suggests decrease in Wnt signaling pathways after infancy.

SUMMARY

- **Analysis of dynamic gene expression in isolated cells across a time series demonstrates the unique heterogeneity of pulmonary mesenchymal cells throughout adolescence.**
- **Increased gene expression associated with immune signatures during pediatric lung development was noted.**
- **Further validation and exploration using this technique may advance understanding of the diversity of pulmonary cell types and pathophysiology of pediatric lung disease.**