Transcriptomic analyses reveal regional signatures in lung allograft recipients

Patricia Agudelo-Romero [1,2,3] ‡, Kak-Ming Ling [1,4], Melanie A. Lavender [5], Jeremy P. Wrobel [5,6], Michael Musk [5], Stephen M Stick [1,7,8,9] and Anthony Kicic [1,4,8,9] ‡*

[1]Wal-yan Respiratory Research Centre, Telethon Kids Institute. Perth, WA, Australia.

[2]School of Molecular Science, The University of Western Australia. Perth, WA, Australia.

[3]European Virus Bioinformatics Center, University Jena, Jena, Germany

[4]School of Population Health, Curtin University. Perth, WA, Australia.

[5]Advanced Lung Disease Unit, Fiona Stanley Hospital. Perth, WA, Australia.

[6]Department of Medicine, University of Notre Dame. Perth, WA, Australia.

[7]Medical Faculty, University of Western Australia. Perth, WA, Australia.

[8]Department of Respiratory and Sleep Medicine, Perth Children's Hospital for Children. Perth, WA, Australia.

[9]Centre for Cell Therapy and Regenerative Medicine, School of Medicine and Pharmacology, University of Western Australia. Perth, WA, Australia.

*** Correspondence:**
Anthony Kicic
Anthony.Kicic @telethonkids.org.au

‡ **These authors share senior authorship**

Keywords: Lung allograft, transcriptomics, airway epithelium, small airway epithelial cells, large airway epithelial cells, Cystic Fibrosis, inflammation.

Abstract

Background: Long term outcomes of allograft recipients are compromised by the development of chronic lung allograft dysfunction (CLAD) promoting bronchiolitis obliterans syndrome (BOS). We established baseline transcriptomic profiles of both the large and small airway epithelial cells (referred as LAEC and SAEC, respectively) to identify regional differences irrespective of initiating disease.

Methods: We obtained matched primary LAEC and SAEC from lung allograft recipients (n=4, 42.5 ± 4.2 years) and established primary cultures. Bulk RNA sequencing was performed to determine differentially expressed genes.

Results: We observed differences in the transcriptional program between LAEC and SAEC Transcription factors (TF) were ranked within the top ten differentially regulated genes. The most abundant TF families included C2H2-ZF, homeobox and bHLH. Upstream regulator analyses identified homeobox genes being significantly in LAEC. Protein-protein interaction network analysis emphasised the role of TFs (*ISL1*, *MSX1*, *HOXA1*, *GATA6*, *ZNF423*) in airway modulation. Additionally, functional enrichment analysis revealed the activation of chemotaxis, metalloendipeptidase/metallopeptidase activity and pro-inflammatory signatures (IL17 signalling and RAGE), in LAEC, while SAEC were characterised by elevated expression of surfactant metabolism related genes. Moreover, alveolar and club cells-related genes were expressed in SAEC, suggesting a lower airway-specific signature.

Conclusion: Our analysis shows robust transcriptional differences between LAEC and SAEC. We suggest a potential role for homeobox TF family as well as the activation of the immune system in the biology of LAEC. Conversely, we observed an alveoli-like transcriptional signature in SAEC, including gas-exchange signals and surfactant metabolism; pathways involved in lung homeostasis.

Introduction

Chronic allograft dysfunction (CLAD) due to bronchiolitis obliterans syndrome (BOS) is a major complication following pulmonary transplantation whose aetiology is not fully understood[1]. Recent findings indicate that the pathology of CLAD is not only limited to the small airways, but also manifests in the proximal airways[2]. The underlying mechanism of BOS is suggested to involve injury and inflammation of epithelial and subepithelial cells, which in turn stimulates wound repair and epithelial to mesenchymal transition signaling[3]. Thus, understanding the pathological mechanisms contributing to the onset of BOS, could facilitate tailored interventions aimed at preventing this pathology.

For the purpose of tracking the different processes that might lead to the onset of BOS, we conducted a baseline study, to compare matched samples of proximal and distal lung airways, without any sign of disease. We did RNA-seq to explore regional transcriptional differences in epithelial cultures established from large/proximal and small/distal airway epithelial cells (referred as LAEC and SAEC) from Lung Transplant recipients. Here, we observed that distal airways culture expressed genes related to surfactant metabolism. Whereas proximal airways cultures displaying two proinflammatory pathways and fibrosis pathways that might be related to early dysregulation of BOS; providing evidence that alterations in gene expression begin in the proximal airways.

Materials and Methods

Patient and sampling procedures

Matched samples from proximal and distal lung biopsies were obtained from lung allograft recipients during the routine surveillance bronchoscopy (Figure S1A). Patients did not present symptomatology indicative of bronchiolitis obliterans syndrome (BOS), transplant rejection or infections at the time of sample collection. Specimens were obtained from four patients (3 females, age range 38-47) that underwent double lung transplantation as previously described[4]. The reasons for lung transplantation

were cystic fibrosis (3 cases), and congenital heart disease (1 case). Both, LAEC and SAEC were grown as a monolayer to confluency and cells then harvested for transcriptomics analysis (Figure S1B). The study was approved by the Royal Perth Hospital, Ethics Committee (Registration: EC2006/021) and written consent was obtained from each participant after being fully informed about the premise and purpose of the study. A summary that outlines the analyses performed here is presented in Figure S2.

Total RNA extraction, library preparation and sequencing

Total RNA was extracted using the PureLink® RNA kit (Life Technologies) following their instructions and treated with RNase Inhibitor (Applied Biosystems™) and DNase I (Invitrogen™). RNA concentration and integrity were then determined using a NanoDrop system and an Agilent Bioanalyser, respectively. Libraries were generated using TruSeq Stranded mRNA (Illumina) kit and Poly(A) enrichment. Sequencing was performed using 100 bp single-end configuration (SE; 100 bp; 20 M).

Bioinformatic and statistical analyses

FASTQ files were quality controlled using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were trimmed and adapters removed with Trimmomatic[5] (Figure S3A and S3B). High-quality reads were then mapped to human reference genome (hg19/GRCh37, Ensembl) using Spliced Transcripts Alignment to a Reference (STAR)[6]. Gene-level quantification of raw counts was performed using High-Throughput Sequencing[7] (HTSeq; Supplementary Materials S1) and post-alignment stats were generated with MultiQC[8] (Figure S3C and S3D). Gene expression analysis was done using EdgeR[9] and limma[10]. Latent variance was treated with removeBatchEffect function from limma[10]. Expression values were used for cluster analysis using pvclust R package[11] (Figure S4; Supplementary Materials S2). *P*-values were corrected for multiple-testing using the Benjamini-Hochberg's method, False Discovery Rate (FDR)[12]. Ranking

metric for top regulated genes was calculated as follow: [Rank = -sign(log_2FC) * log_{10}(FDR)]. Differentially expressed genes (DEGs) were then filtered considering an FDR ≤0.05 and a Fold-Chance (FC) ≥|1.5| and annotation was added using the org.Hs.eg.db R package[13] (Supplementary Materials S3). Volcano plot was generated using EnhancedVolcano R package[14].

Mining of transcription Factors (TF) in our data was done following the pipeline described by Lambert *et al*.,[15] (Supplementary Materials S4) Analysis of upstream transcriptional regulators was performed with ChEA3[16] (Supplementary Materials S5). Protein-protein interaction networks were generated using NetworkAnalyst[17] and subsequently analysed using NetworkAnalyzer App with NetworkAnalyzer App (Supplementary Materials S6). Enrichment analyses of gene ontology (GO), KEEG and Reactome databases were done using ClusterProfiler[18] (Supplementary Materials S7). The RNA-sequencing data have been deposited in the National Center for Biotechnology Information's Bioproject PRJNA885405 and Gene Expression Omnibus (GEO) under accession number GSE214434 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE214434).

Results

To determine any regional transcriptional differences between the lung airways, LAEC were compared to SAEC. Hereafter, up-regulated genes will be referred as active in LAEC and down-regulated genes as active in SAEC.

Gene expression differences between lung regional comparison.

After filtering, a total of 355 differentially expressed genes (DEGs), 188 up-regulated and 167 down-regulated genes were identified (Figure 1A). A volcano plot shows the bidirectional distribution patterns of up- and down-regulated genes along with the top genes (Figure 1B). Among the top 10 ranked genes, two transcription factors (TFs) (*ZFHX4* and *MSX1*) belonging to homeobox family were found to be up-regulated, and *ZNF730* belonging to C2H2 Zinc Finger (ZF) which was down-regulated.

Transcript Factors modulation underlies gene expression differences between lung airways

Knowing the essential role TF have in gene modulation, we explored their potential contribution towards the differences seen between the LAEC and SAEC gene transcriptional profiles. For this purpose we made used of the human TFs catalogue[15]. We detected a total of 26 TF belonging to eight TF families and the three most abundant including C2H2 ZF (34.6 %), homeobox (26.9 %) and bHLH (11.5 %) families (Figure 2A; Supplementary Materials S4). Using Venn diagram analysis, two overlapping families were identified from up- and down-regulated genes comparison (C2H2 ZF and homeobox) (Figure 2B). More interesting, four TFs families were uniquely identified in LAEC (BED ZF, DM, Fox and T-box), and two specifics in SAEC (bHLH and GATA) (Figure 2B). Bar plots summarise the 14 up- and 12 down-regulated TFs (Figure 2C).

Homeobox genes represent central modulators of gene expression in the proximal airways

To gain insight into the gene regulatory network of the airways and identify upstream modulators linked to the observed DEGs, we performed a TF enrichment analysis (TFA). We retrieved 161 upstream regulators significantly associated with the up-regulated genes, of which eight TF belonging to homeobox family were observed in the top ten upstream regulators (Figure 3A). In contrast, no upstream regulators for the down-regulated genes were identified. To better understand the interaction between TFs and modulated genes, we then generated networks based on the relevant associations (Figure 3B; Supplementary Materials S5). Interaction between six upstream regulators explained the activation of 18 genes. *HOXD10* and *HOXD9* represented central modules explaining differences in gene expression of five genes, three of which are TF (*MSX1*, *TBX3* and *FOXD1*) (Figure 3B). Lastly, topological analyses were used to identify hub genes from the DEGs. This network analysis employed 823 nodes with at least one connected component, where 97 annotated genes belong to our DEG core. Hub genes were extracted by filtering nodes above 14 degrees, which resulted in a set of 24 genes (17 up-regulated and 7 down-regulated genes) (Figure 3C) and the top five hubs are *SPP1*, *HOXA1*,

S100A8, *KRT14* and *ID2*. Besides, several TFs were identified as hub genes, three up-regulated (*ISL1*; *MSX1*; *HOXA1*) and two down-regulated (*GATA6*; *ZNF423*). Again, the three hub TFs from up-regulated genes belong to homeobox family.

Functional enrichment shows a pro-inflammatory signature in proximal airway and pulmonary surfactant metabolism in the distal airway

We next performed functional enrichment to determine the biological pathways associated with the up- and down-regulated genes. Specifically, we utilised pathway over-representation analysis using GO, Reactome and KEGG databases followed by a category-gene network (Figure 4A-B, Supplementary Materials S7). Several categories were identified enriched from the up- and down-regulated gene list. The top two more statistically significant pathways were taxis/chemotaxis and activity of metalloendipeptidase/metallopeptidase. Besides two pro-inflammatory pathways (RAGE receptor biding and IL17 signalling pathways) were also enriched. Conversely, two pathways representing the down-regulated genes were observed: benzaldehyde dehydrogenase activity and surfactant metabolism.

Discussion

The study of the transcriptional lung regional differences in lung allograft recipients provides an excellent tool to understand baseline molecular mechanisms/signatures that may drive chronic allograft rejection. With this objective in mind, we analysed transcriptional differences between LAEC and SAEC from lung allograft recipients. Our results support the hypothesis that airway transcriptional regional differences exist, where TF genes, specifically homeobox TF may have a potential role in establishing these differences. Protein-protein interaction network analysis emphasised the role of *ISL1*, *MSX1*, *HOXA1* in LAEC, and *GATA6*, *ZNF423* in SAEC. Furthermore, functional enrichment analysis identified activation of chemotaxis, metalloendipeptidase/metallopeptidase and two pro-inflammatory categories in DEGs from LAEC and surfactant metabolism in SAEC DEGs.

Several TF (~7%) were noticed in the top ten genes of the DEG (Figure 1). Detailed analysis about TF families' composition in the airways found homeobox and C2H2 ZF in common (Figure 2). Homeobox is related with patterning in lung branching[19], and although C2H2 ZF plays important roles in development and disease, still remain poorly characterized[20]. Four unique TF families were up-regulated, including BED-ZF, DM, FOX, and T-box, which are involved in development, among other functions. Several TF identified in these families serve primary roles in the lung. For instance, *ZBED2* has been predicted to promote the keratinocyte basal state, inducing differentiation[21]. In addition, *DMRT2* is involved in establishing left–right asymmetry and somitogenesis; whereas *DMRTA2*, also known as *DMRT5*, has been reported in anterior neural tissue development [22]. Moreover, *FOXD1* has been linked as a marker for lung pericytes and *FOXC2* in vascularisation[23]. Finally, *TBX3* has been found to be involved in lung branching morphogenesis with expression in the lung mesenchyme [24]. In contrast, two specific TF families, bHLH and GATA, were identified in the down-regulated genes. The bHLH family is important in regulating embryonic development, and *NPAS1* has been reported to regulate branching morphogenesis in the embryogenic lung[25]. Furthermore, the GATA family plays an important role in both epithelial and smooth muscle cell linage diversity in the lung. Specifically, *GATA6* has been found to induce differentiation of primitive foregut endoderm into respiratory epithelial cell linages, in addition to regulating surfactant protein genes[19].

Next, upstream regulator enrichment analysis of the up-regulated genes predicted that *HOXD9* and *HOXD10* homeobox genes activate *MSX1*, *TBX3* and *FOXD1* (Figure 3B). Interestingly, *MSX1* and *TBX3* are expressed in mesenchyme cells in the single cell (sc) lung map website[26] (LungGENS, 10X sc 24-years-old), and have been shown to be regulated by Hedgehog and Wnt signalling pathways[24,27]. Network analyses identified 24 hub genes (Figure 3C); from the up-regulated hubs we found a conserved signature (*CSF1R*; *MMP1*; *CD8A*; *PLAT*; *S100A8*; *SPP1*) similar to that of the tracheal / proximal airway[28], with genes involved in a variety of processes including cell migration, inflammation

response and chemokine production. Moreover, the three TF detected in the up-regulated gene hubs are related to morphogenesis; two specific to mesenchyme development and the other to proximal airways development[27,29].

Interestingly, several pathways were statically significative in a functional enrichment analysis (Figure 4). From the up-regulated gene list, taxis/chemotaxis processes had the most statistical significative values. Chemotaxis, defined as the directed migration of cells towards a specified entity, are important for cellular pattering and development[30]. In drosophila, chemotaxis has been detected in the tracheal epithelium[31]. Likewise, the metalloendipeptidase activity pathway was also observed to be activated through two matrix metalloproteinase (MMP) genes (*MMP1*, *MMP13*). In skin repair, *MMP1* alters the migratory substratum driving the forward movement of the repairing cells by allowing them to attach, dislodge, then reattach to the wounded matrix[32]. Interestingly, *MM13* has been shown to play a role in the pathogenesis of liver and lung fibrosis [33–35]. Additionally, two proinflammatory (IL17 signalling and RAGE receptor binding) plus an antimicrobial pathway were enriched in the up-regulated genes. Amongst these genes, *S100A7* and *S100A8* are known to expressed in the trachea[36] and have been reported to have a role in innate immune responses to pathogens[37]. Both have gained research interest because they exhibit selectivity towards pathogenic bacteria, while having no effect on beneficial commensal bacteria[38,39].

Interestingly, the up-regulated enriched categories are also related to epithelial repair and innate immunity and may suggest that their activation after lung transplantation may help prevent the entry of undesirable microorganisms via the production of chemokines and antimicrobial peptides[32]. The inflammation category activation also supports this hypothesis, since it can act as a defence mechanism to abiotic[40] and biotic[41] insults. However, a fibrosis-related gene was also observed within the LAEC signature, which in the long term could function to compromise lung allograft outcomes[33–35].

Two hub upstream regulators from the up-regulated genes, *HOXD9* and *HOXD10*, are known to activate TF involved with the chemotaxis process (*FOXD1*; *ISL1*) (Fig 3B and 4B) and genes related to mesenchymal cells (*ISL1; MSX1; TBX3*). In adults, directional movement of chemotaxis modulates different responses depending on whether it is infectious or injurious. In the first, immune cells would be activated and, in the latter, wound healing and tissue regeneration responses are activated. The latter is provided by the regional stem cells that move into damaged areas, produce connective tissue and maintain tissue homeostasis[30]. It is unclear whether this proximal airway signature from LT*x* patients could be due as part of healing/regeneration response after the lung transplantation, or as part of the immune response linked with the proinflammatory processes previously reported[30,42]. Conversely, only two pathways were enriched from the down-regulated gene list. The first, surfactant metabolism, is essential in small airways because it regulated alveolar surface tension and plays a role in protection against oxidants and infection[43]. The second, NAD- and NADP-dependent benzaldehyde dehydrogenase (ALDH) are involved in detoxification and prior work has reported *ALDHA1* and *ALDH3A1* to be involved in the host defence response to toxins in smokers[44].

We acknowledge a number of limitations to this research. Firstly, due to the precious nature of the samples involved, this pilot study has a limited number of biological replicates from LT*x* recipients. Despite this, there was sufficient sensitivity to obtain DEG using the rigorous analysis pipeline outlined. Also, the limited expansion potential of primary AEC is a limitation[45] and its effects remain unknown at the transcriptomic level. Nevertheless, we believe that the utilisation of 'unaltered primary airway cells' are a significant strength of this study. Likewise, monolayer cultures may oversimplify the multicellular interactions, but a robust and repeatable model with low methodological variation is important. All together, we are confident that limitations are minor and that our results provide new insight into the transcriptional regional differences. For prospective studies, these limitations could be solved with the use cell culture of specialized cells.

In conclusion, the data presented here suggests distinctive signatures from LAECs and SAECs at a transcriptomic level, establishing new insights into regional gene expression in the airways of lung transplant recipients (Figure 4C). Proximal airway gene expression changes included a proinflammatory signature which may indicate a defence mechanism since it is the first barrier of defence, as well as a fibrotic signature which may initiate downstream complications such as BOS establishment. In contrast, the small airways reflect a characteristic alveoli's hallmark, including gas-exchange signals and secretion of pulmonary surfactant proteins. These results will seed future large scale works to determine the predictive value of these genes as potential biomarkers aimed at maintaining baseline lung health of allograft recipients.

Data Availability Statement

Raw datasets have been uploaded to GEO, with accession number GSE214434.

Ethics Statement

The study was approved by the Royal Perth Hospital, Ethics Committee (Registration: EC2006/021) and written consent was obtained from each participant after being fully informed about the premise and purpose of the study.

Author Contributions

Conceptualization: SMS, AK and PAR. Data curation: PAR. Formal analysis: PAR. Funding acquisition: SMS and AK. Investigation: PAR, KML, MAL, JPW, MM, SMS and AK. Methodology: PAR, KML, MAL, JPW, MM, SMS and AK. Resources: PAR. Visualization: PAR. Manuscript writing: PAR and AK. Manuscript review: PAR, KML, MAL, JPW, MM, SMS and AK.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This work was supported by grants from the McCusker Foundation and the Heart and Lung Transplant Foundation of Western Australia. A.K. is a Rothwell Family Fellow. S.M.S. is an NHMRC Practitioner Fellow. Analysis was performed in the Pawsey Supercomputing Research Centre with funding from the Australian Government and the Government of Western Australia.

Figure legend

Figure 1. Transcriptomic signatures between proximal and distal lung regions. (A) Bar plot summarises the number of DEGs (FC ≥ |1.5| and FDR < 0,05) of the regional comparison on the y-axis and direction of the change on the x-axis, divided in up- and down-regulated genes, up (red) and down (blue), respectively. (B) Volcano plot displays bidirectional log2 FC on the x-axis and absolute Log10 FDR on the y-axis. Top 10 ranked genes for up- and down-regulated genes are labelled in each case and bold letter highlight Transcription Factor (TF) genes.

Figure 2. Identification of Transcription Factors (TF) regulating proximal and distal lung regions. (A) Pie plot represents the percentage of TF families from the TF in DEGs. (B) Venn diagram highlights the common and uniquely TF families between LEAC and SAEC. Blue and pink represent proximal/LAEC and distal/SAEC samples. (C) Bar plots depict the TF signature found between up- and down-regulated as LAEC and SAEC, respectively. Colour palette refers to the TF families.

Figure 3. Upstream regulator enrichment and transcriptional network analyses of significantly differentially expressed genes. (A) Bar plot visualisation of TF enrichment analysis (TFEA) shows the top upstream regulators enriched from the DEGs core using a Fisher's exact test (FET) from ChIP-X Enrichment Analysis 3 (ChEA3)16. Up- and down-regulated genes are represented in red and blue, respectively. Dash line represents the FDR 5% limit. (B) Network analysis of upstream regulators and associated regulated genes from up-regulated genes. Network was clustered using an Edge-weighted Spring-Embedded Layout method46. Brighter colours illustrate upstream regulators. Shapes depict genes and TF as circles and diamonds, respectively. (C) Protein-protein interaction network of the differentially expressed genes (DEG) was calculated from IMEx Interactome database of InnateDB (https://www.innatedb.com) using NetworkAnalyst17, it was analysed and visualised in Cytoscape46. Colour intensity displays the fold change value (red; up-regulated genes and blue; down-regulated genes), edges are show in light grey; beside node size represents the degree centrality as interconnection.

Figure 4. Functional enrichment analyses and transcriptional hallmarks of the proximal and distal airways from lung transplant (LT*x*) recipients. (A) Scatter plot represents the significant categories enriched in up- and down- regulated genes. Bar side row represents databases used (DB), adjusted P-value (False Discovery Rate, FDR) colours the significance and dot size the numbers of genes involved in each category. (B) Category-gene network associations of enriched terms. Up-regulated genes in red and down-regulated genes in blue. (C) Transcriptional hallmarks overview of the airways. Samples from the proximal airways were collected close to the trachea (1). This region is constantly exposed to abiotic (allergens, cigarettes) and biotic (virus, bacteria, fungi) factors (2). Two proinflammatory pathways (IL-17 signalling and RAGE receptor binding) associated with defence mechanisms were found activated in this region (3). Distal airways specimens were collected near the alveoli (4), where the gas-exchange occurs (5). In agreement, these samples demonstrated abundance of transcripts related to surfactant metabolism, which is crucial to lower the alveolar surface tension during expiration (6). Figure created with BioRender.com.

Supplementary Material

1 **Supplementary Material S1.** Raw counts generated with HTSeq[4], and obtained from an alignment using STAR[3] to the Ensembl human reference genome hg19/GRCh37.

2 **Supplementary Material S2.** Expression values used for cluster analysis and gene expression profiles.

3 **Supplementary Material S3.** List of 355 differentially expressed genes with a cut-off of FC > |1.5| and FDR of 5%.

4 **Supplementary Material S4.** Data mining of transcript factor families based on Lambert et al.,[12].

5 **Supplementary Material S5**. Transcription factor enrichment analysis of the upstream regulators using ChEA3[13].

6 **Supplementary Material S6.** Network analysis of differentially expressed genes generated by Cystoscope[15]. Statistics related to (A) nodes and (B) edges, as well as (C) the subnetwork built in Figure 3C.

7 **Supplementary Material S7**. Enrichment analysis of differentially expressed genes using ClusterProfiler[17].

References

1 Beeckmans H, Bos S, Vos R, Glanville AR. Acute Rejection and Chronic Lung Allograft Dysfunction: Obstructive and Restrictive Allograft Dysfunction. *Clin Chest Med* 2023; **44**:

137–157.

2 Swatek AM, Lynch TJ, Crooke AK, Anderson PJ, Tyler SR, Brooks L *et al.* Depletion of Airway Submucosal Glands and TP63 + KRT5 + Basal Cells in Obliterative Bronchiolitis. *Am J Respir Crit Care Med* 2018; **197**: 1045–1057.

3 Aguilar PR, Michelson AP, Isakow W. Obliterative Bronchiolitis. *Transplantation* 2016; **100**: 272–283.

4 Banerjee B, Kicic A, Musk M, Sutanto EN, Stick SM, Chambers DC. Successful establishment of primary small airway cell cultures in human lung transplantation. *Respir Res* 2009; **10**. doi:10.1186/1465-9921-10-99.

5 Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.

6 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; **29**: 15–21.

7 Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; **31**: 166–169.

8 Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016; **32**: 3047–3048.

9 Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2009; **26**: 139–140.

10 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43**: e47.

11 Suzuki R, Shimodaira H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 2006; **22**: 1540–1542.

12 Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. 1995.

13 Carlson M. org.Hs.eg.db: Genome wide annotation for Human. 2019.

14 Blighe K, Rana S LM. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. 2020.

15 Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M *et al.* The Human Transcription Factors. Cell. 2018; **172**: 650–665.

16 Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res* 2019; **47**: W212–W224.

17 Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res* 2019; **47**: W234–W241.

18 Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi A J Integr Biol* 2012; **16**: 284–287.

19 Warburton D, Schwarz M, Tefft D, Flores-Delgado G, Anderson KD, Cardoso W V. The molecular basis of lung morphogenesis. Mech. Dev. 2000; **92**: 55–81.

20 Han BY, Foo CS, Wu S, Cyster JG. The C2H2-ZF transcription factor Zfp335 recognizes two consensus motifs using separate zinc finger arrays. *Genes Dev* 2016; **30**: 1509–1514.

21 Finnegan A, Cho RJ, Luu A, Harirchian P, Lee J, Cheng JB *et al.* Single-cell transcriptomics reveals spatial and temporal turnover of keratinocyte differentiation regulators. *Front Genet* 2019; **10**: 775.

22 Bellefroid EJ, Leclère L, Saulnier A, Keruzore M, Sirakov M, Vervoort M *et al.* Expanding roles for the evolutionarily conserved Dmrt sex transcriptional regulators during embryogenesis. Cell. Mol. Life Sci. 2013; **70**: 3829–3845.

23 Whitsett JA, Kalin T V., Xu Y, Kalinichenko V V. Building and regenerating the lung cell by cell. *Physiol Rev* 2019; **99**: 513–554.

24 Lüdtke TH, Rudat C, Wojahn I, Weiss AC, Kleppa MJ, Kurz J *et al.* Tbx2 and Tbx3 Act Downstream of Shh to Maintain Canonical Wnt Signaling during Branching Morphogenesis of the Murine Lung. *Dev Cell* 2016; **39**: 239–253.

25 Levesque BM, Zhou S, Shan L, Johnston P, Kong Y, Degan S *et al.* NPAS1 regulates branching morphogenesis in embryonic lung. *Am J Respir Cell Mol Biol* 2007; **36**: 427–434.

26 Du Y, Guo M, Whitsett JA, Xu Y. 'LungGENS': A web-based tool for mapping single-cell gene expression in the developing lung. *Thorax* 2015; **70**: 1092–1094.

27 McCulley D, Wienhold M, Sun X. The pulmonary mesenchyme directs lung development. Curr. Opin. Genet. Dev. 2015; **32**: 98–105.

28 Kicic A, de Jong E, Ling KM, Nichol K, Anderson D, Wark PAB *et al.* Assessing the unified airway hypothesis in children via transcriptional profiling of the airway epithelium. *J Allergy Clin Immunol* 2020; **145**: 1562–1573.

29 Kim E, Jiang M, Huang H, Zhang Y, Tjota N, Gao X *et al.* Isl1 Regulation of Nkx2.1 in the Early Foregut Epithelium Is Required for Trachea-Esophageal Separation and Lung Lobation. *Dev Cell* 2019; **51**: 675-683.e4.

30 Vorotnikov A V., Tyurin-Kuzmin PA. Chemotactic signaling in mesenchymal cells compared to amoeboid cells. Genes Dis. 2014; **1**: 162–173.

31 Sutherland D, Samakovlis C, Krasnow MA. branchless encodes a Drosophila FGF homolog that controls tracheal cell migration and the pattern of branching. *Cell* 1996; **87**: 1091–1101.

32 Parks WC, Wilson CL, López-Boado YS. Matrix metalloproteinases as modulators of inflammation and innate immunity. *Nat Rev Immunol 2004 48* 2004; **4**: 617–629.

33 Uchinami H, Seki E, Brenner DA, Armiento JD'. Loss of MMP 13 Attenuates Murine Hepatic Injury and Fibrosis During Cholestasis. *Hepatology* 2006; **44**: 420–429.

34 Benyon JS, Duffield JPI, Kendall CM, Constandinou R, Christopher JA, Fallowfield M *et al.* Resolution of Murine Hepatic Fibrosis Metalloproteinase-13 and Facilitate the Source of Hepatic Matrix Scar-Associated Macrophages Are a Major. *J Immunol Ref* 2022; **178**: 5288–5295.

35 Nkyimbeng T, Ruppert C, Shiomi T, Dahal B, Lang G, Seeger W *et al.* Pivotal role of matrix metalloproteinase 13 in extracellular matrix turnover in idiopathic pulmonary fibrosis. *PLoS One* 2013; **8**. doi:10.1371/JOURNAL.PONE.0073279.

36 Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A *et al.* Tissue-based map of the human proteome. *Science (80-)* 2015; **347**. doi:10.1126/science.1260419.

37 Kozlyuk N, Monteith AJ, Garcia V, Damo SM, Skaar EP, Chazin WJ. S100 Proteins in the Innate Immune Response to Pathogens. *Methods Mol Biol* 2019; **1929**: 275–290.

38 Gläser R, Harder J, Lange H, Bartels J, Christophers E, Schröder JM. Antimicrobial psoriasin (S100A7) protects human skin from Escherichia coli infection. *Nat Immunol* 2005; **6**: 57–64.

39 Corbin BD, Seeley EH, Raab A, Feldmann J, Miller MR, Torres VJ *et al.* Metal chelation and inhibition of bacterial growth in tissue abscesses. *Science (80-)* 2008; **319**: 962–965.

40 Lajoie S, Lewkowich IP, Suzuki Y, Clark JR, Sproles AA, Dienger K *et al.* Complement-mediated regulation of the IL-17A axis is a central genetic determinant of the severity of experimental allergic asthma. *Nat Immunol* 2010; **11**: 928–935.

41 Iwakura Y, Nakae S, Saijo S, Ishigame H. The roles of IL-17A in inflammatory immune responses and host defense against pathogens. *Immunol Rev* 2008; **226**: 57–79.

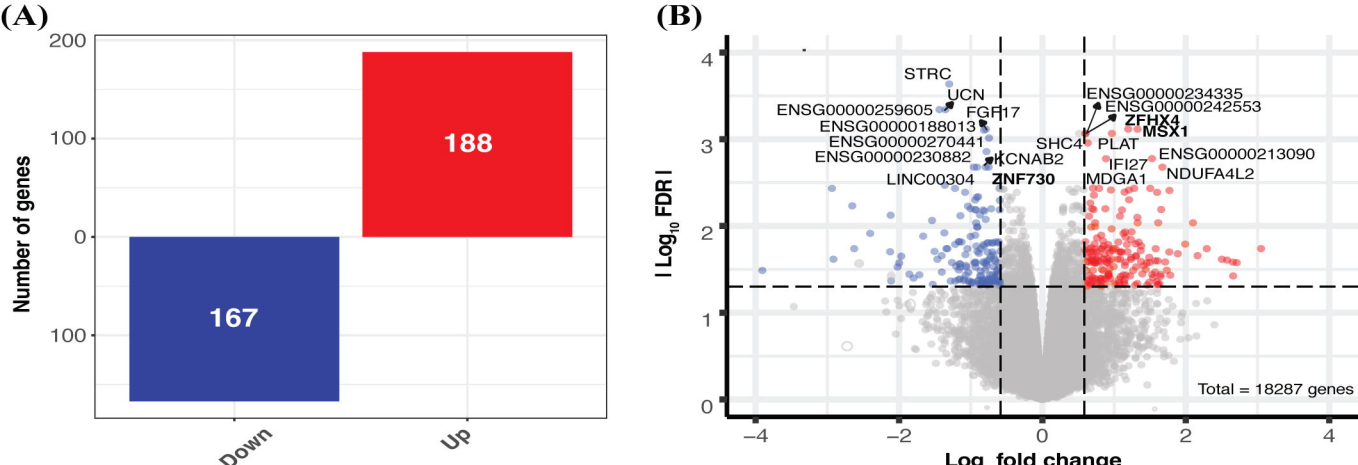
42 Bernardo ME, Fibbe WE. Mesenchymal stromal cells: Sensors and switchers of inflammation. Cell Stem Cell. 2013; **13**: 392–402.

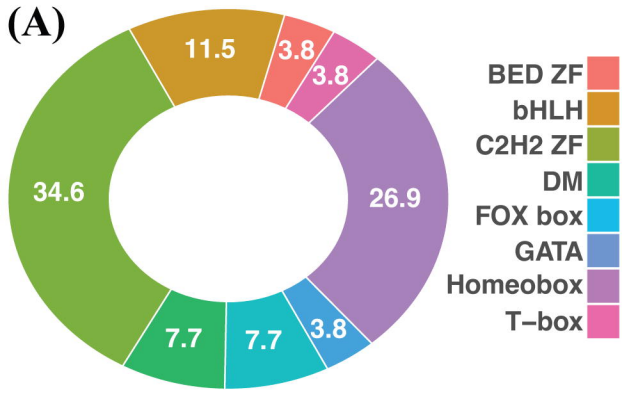
43 Olmeda B, Martínez-Calle M, Pérez-Gil J. Pulmonary surfactant metabolism in the alveolar airspace: Biogenesis, extracellular conversions, recycling. *Ann Anat* 2017; **209**: 78–92.

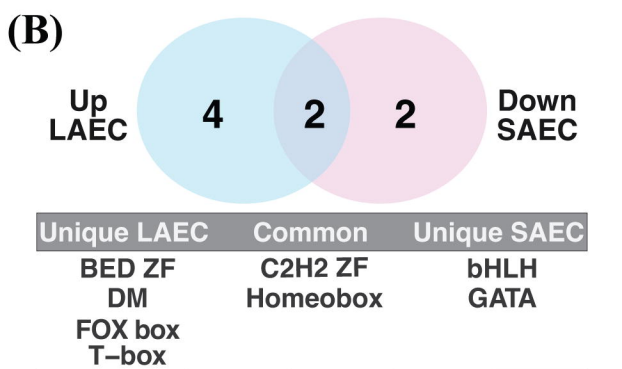
44 Zuo WL, Rostami MR, Shenoy SA, Leblanc MG, Salit J, Strulovici-Barel Y *et al.* Cell-specific expression of lung disease risk-related genes in the human small airway epithelium. *Respir Res* 2020; **21**: 1–11.

45 Martinovich KM, Iosifidis T, Buckley AG, Looi K, Ling KM, Sutanto EN *et al.* Conditionally reprogrammed primary airway epithelial cells maintain morphology, lineage and disease specific functional characteristics. *Sci Rep* 2017; **7**: 17971.

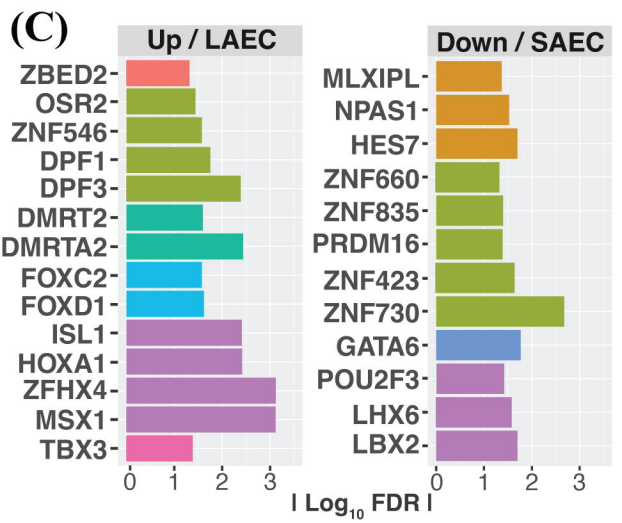
46 Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol* 2019; **20**: 185.

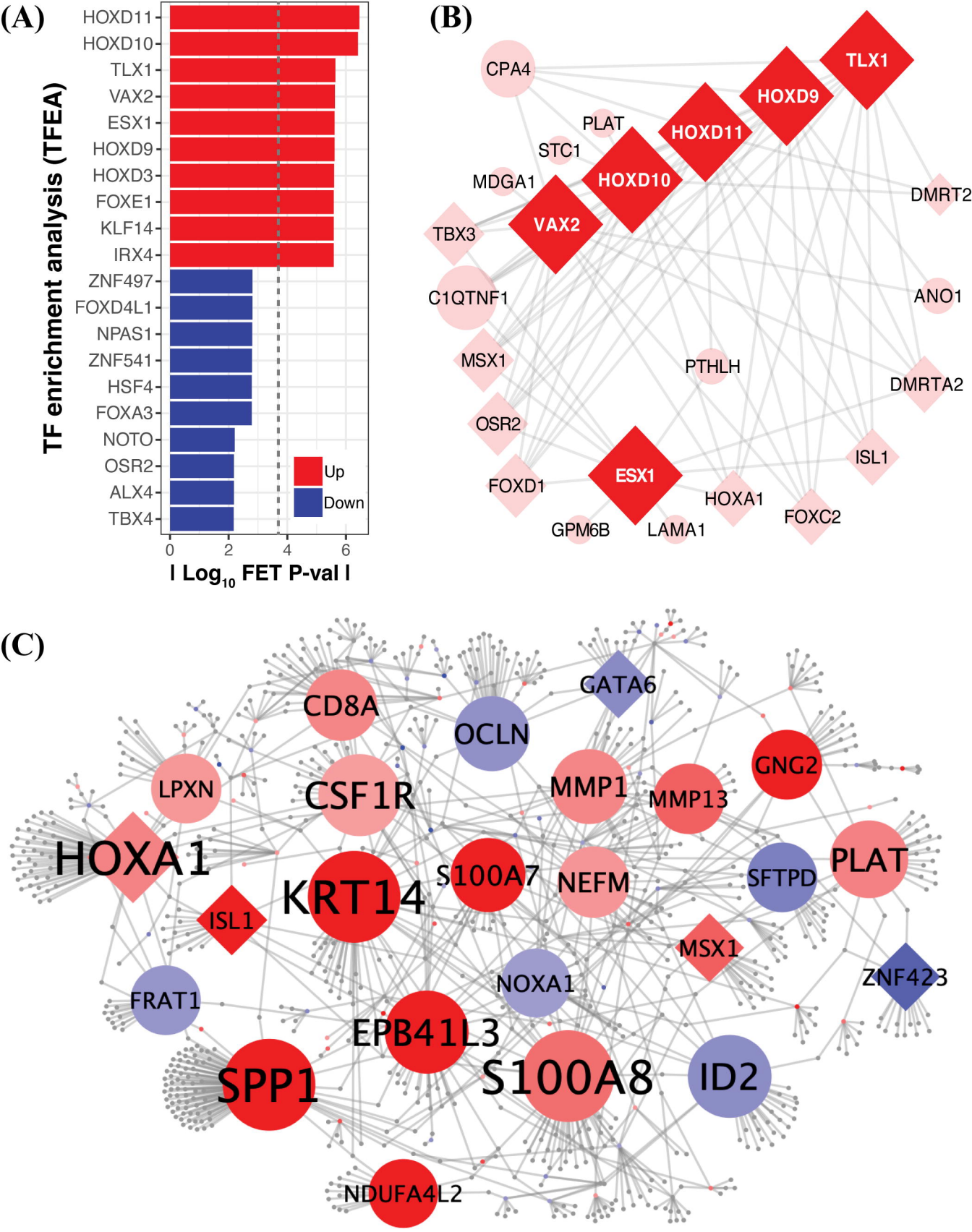

(A)
200
100
0
100
Number of genes
188
167
Down
Up
(B)
4
3
2
1
0
| Log10 FDR |
STRC
UCN
ENSG00000259605
FGF17
ENSG00000188013
ENSG00000270441
ENSG00000230882
KCNAB2
LINC00304
ZNF730
ENSG00000234335
ENSG00000242553
ZFHX4
MSX1
SHC4
PLAT
IFI27
ENSG00000213090
MDGA1
NDUFA4L2
Total = 18287 genes
−4
−2
0
2
4
Log2 fold change

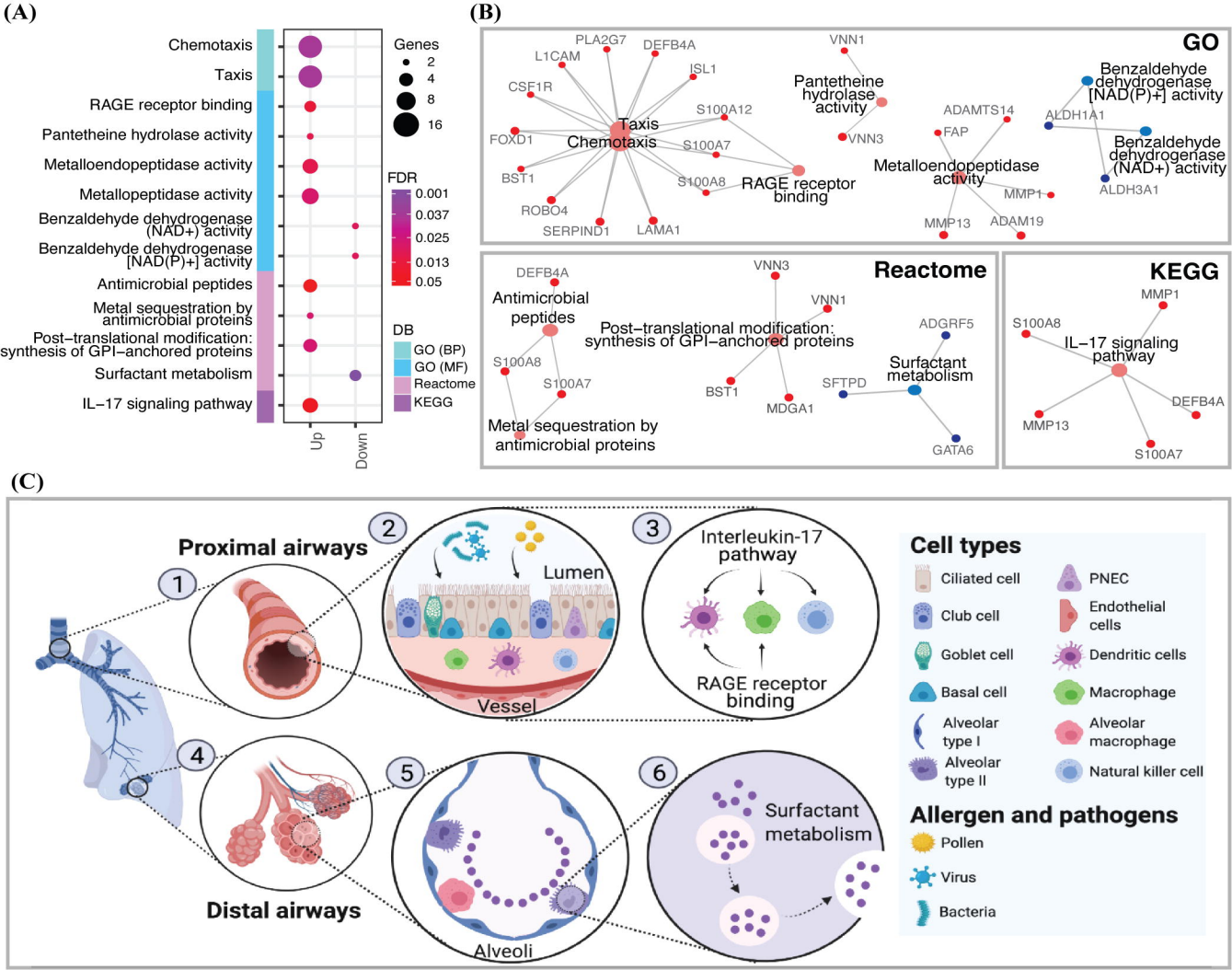




Unique LAEC	Common	Unique SAEC
BED ZF	C2H2 ZF	bHLH
DM	Homeobox	GATA
FOX box		
T–box		




(A)
HOXD11
HOXD10
TLX1
VAX2
ESX1
HOXD9
HOXD3
FOXE1
KLF14
IRX4
ZNF497
FOXD4L1
NPAS1
ZNF541
HSF4
FOXA3
NOTO
OSR2
ALX4
TBX4
Up
Down
TF enrichment analysis (TFEA)
0
2
4
6
| Log10 FET P-val |
(B)
CPA4
TLX1
HOXD9
PLAT
HOXD11
STC1
HOXD10
MDGA1
DMRT2
VAX2
TBX3
C1QTNF1
ANO1
MSX1
PTHLH
DMRTA2
OSR2
ISL1
ESX1
FOXD1
HOXA1
FOXC2
GPM6B
LAMA1
(C)
GATA6
CD8A
OCLN
GNG2
LPXN
CSF1R
MMP1
MMP13
HOXA1
PLAT
S100A7
NEFM
SFTPD
KRT14
ISL1
MSX1
ZNF423
NOXA1
FRAT1
EPB41L3
SPP1
S100A8
ID2
NDUFA4L2


(A)
Chemotaxis
Taxis
RAGE receptor binding
Pantetheine hydrolase activity
Metalloendopeptidase activity
Metallopeptidase activity
Benzaldehyde dehydrogenase (NAD+) activity
Benzaldehyde dehydrogenase [NAD(P)+] activity
Antimicrobial peptides
Metal sequestration by antimicrobial proteins
Post−translational modification: synthesis of GPI−anchored proteins
Surfactant metabolism
IL−17 signaling pathway
Up
Down
Genes
2
4
8
16
FDR
0.001
0.037
0.025
0.013
0.05
DB
GO (BP)
GO (MF)
Reactome
KEGG
(B)
GO
PLA2G7
DEFB4A
L1CAM
ISL1
CSF1R
S100A12
FOXD1
S100A7
BST1
S100A8
ROBO4
SERPIND1
LAMA1
VNN1
VNN3
ADAMTS14
FAP
MMP1
MMP13
ADAM19
ALDH1A1
ALDH3A1
Reactome
DEFB4A
VNN3
VNN1
ADGRF5
S100A8
S100A7
BST1
MDGA1
SFTPD
GATA6
KEGG
MMP1
S100A8
DEFB4A
MMP13
S100A7
(C)
Proximal airways
Distal airways
Lumen
Vessel
Interleukin-17 pathway
RAGE receptor binding
Alveoli
Surfactant metabolism
Cell types
Ciliated cell
Club cell
Goblet cell
Basal cell
Alveolar type I
Alveolar type II
PNEC
Endothelial cells
Dendritic cells
Macrophage
Alveolar macrophage
Natural killer cell
Allergen and pathogens
Pollen
Virus
Bacteria