


RESEARCH

Open Access



Integrative analysis of bulk and single-cell RNA sequencing reveals the gene expression profile and the critical signaling pathways of type II CPAM

Fengxia Li^{1†}, Zheng Tan^{2†}, Hongyu Chen¹, Yue Gao², Jie Xia², Ting Huang², Liang Liang², Jian Zhang², Xianghong Zhang³, Xucong Shi⁴, Qiang Chen^{5*}, Qiang Shu^{4,6*} and Lan Yu^{1*} 

Abstract

Background Type II congenital pulmonary airway malformation (CPAM) is a rare pulmonary microcystic developmental malformation. Surgical excision is the primary treatment for CPAM, although maternal steroids and betamethasone have proven effective in reducing microcystic CPAM. Disturbed intercellular communication may contribute to the development of CPAM. This study aims to investigate the expression profile and analyze intercellular communication networks to identify genes potentially associated with type II CPAM pathogenesis and therapeutic targets.

Methods RNA sequencing (RNA-seq) was performed on samples extracted from both the cystic area and the adjacent normal tissue post-surgery in CPAM patients. Iterative weighted gene correlation network analysis (iWGCNA) was used to identify genes specifically expressed in type II CPAM. Single-cell RNA-seq (scRNA-seq) was integrated to unveil the heterogeneity in cell populations and analyze the communication and interaction within epithelial cell sub-populations.

Results A total of 2,618 differentially expressed genes were identified, primarily enriched in cilium-related biological process and inflammatory response process. Key genes such as *EDN1*, *GPR17*, *FPR2*, and *CHRM1*, involved in the G protein-coupled receptor (GPCR) signaling pathway and playing roles in cell differentiation, apoptosis, calcium homeostasis, and the immune response, were highlighted based on the protein-protein interaction network. Type II CPAM-associated modules, including ciliary function-related genes, were identified using iWGCNA. By integrating scRNA-seq data, *AGR3* (related to calcium homeostasis) and *SLC11A1* (immune related) were identified as the only

[†]Fengxia Li and Zheng Tan contributed equally to this work.

*Correspondence:

Qiang Chen

jx-cq@163.com

Qiang Shu

shuqiang@zju.edu.cn

Lan Yu

yulan20@zju.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

two differently expressed genes in epithelial cells of CPAM. Cell communication analysis revealed that alveolar type 1 (AT1) and alveolar type 2 (AT2) cells were the predominant communication cells for outgoing and incoming signals in epithelial cells. The ligands and receptors between epithelial cell subtypes included COLLAGEN genes enriched in PI3K-AKT signaling and involved in epithelial to mesenchymal transition.

Conclusions In summary, by integrating bulk RNA-seq data of type II CPAM with scRNA-seq data, the gene expression profile and critical signaling pathways such as GPCR signaling and PI3K-AKT signaling pathways were revealed. Abnormally expressed genes in these pathways may disrupt epithelial-mesenchymal transition and contribute to the development of CPAM. Given the effectiveness of prenatal treatments of microcystic CPAM using maternal steroids and maternal betamethasone administration, targeting the genes and signaling pathways involved in the development of CPAM presents a promising therapeutic strategy.

Keywords Congenital pulmonary airway malformation, Bulk RNA sequencing, Single-cell RNA sequencing, Cell communication

Introduction

Congenital pulmonary airway malformation (CPAM) was first described by Ch'in and Tang in 1949 as a rare lesion with systemic edema in immature fetuses [1]. It can be diagnosed prenatally through ultrasound screening or postnatally when patients present with pneumonia. While it can be asymptomatic, CPAM carries potential risk of complications, including cyst infection, respiratory distress, failure to thrive, and malignant transformation [2, 3].

According to the Stocker classification, CPAM can be classified into five types based on cyst size and the hamartomatous components of the lesion [4]. Type 0 is incompatible with life due to solid lesions. Type I lesion typically consists of large single or multiple cysts (greater than 2 cm) with pseudostratified columnar epithelium. Type II usually presents with multiple smaller simple and round cysts (less than 1.5–2 cm) lined by ciliated cuboidal to columnar epithelium. Type I and type II lesions are the most frequent forms of CPAM [5]. Type III lesion can expand the entire lung with cuboidal epithelium. Type IV is composed of thin-walled structure lined by alveolar cells and often presents with mediastinal shift, and pneumothorax.

The pathogenic mechanisms of CPAM are not completely understood. Each type originates from a different part of the bronchial tree and may not be related. Type 0, also known as acinar dysplasia (AcDys), is associated with mutations of *TBX4*, *FGFR2*, and *FGF10* [6–8]. Type IV is associated with the inactivation of *DICER1*, caused by germline mutations of *DICER1* [9]. It is suggested to be redesignated as pleuropulmonary blastoma (PPB) type I, as these two entities are indistinguishable based on imaging and share the same pathogenic mechanisms [10, 11]. Type I and some of type III lesions have recently been found to be driven by mosaic *KRAS* mutations that occur in the lung epithelium during early development [12], while type II is negative for *KRAS* mutations and is believed to be the consequence of bronchial atresia [13].

Potential genes and altered expression of factors that control normal lung development have been indicated in CPAM pathogenesis. Fibroblast growth factors (FGFs), including *FGF7* and *FGF10*, are expressed in the embryonic lung mesenchyme [14, 15] and may act on the epithelium through their receptor *FGFR2b* to play a critical role in early lung morphogenesis [16]. Overexpression of *FGF10* has been implicated in the formation of lung cysts [17, 18]. Furthermore, inactivation of a Ribonuclease Dicer or zinc finger factor Yin Yang I (*YY1*) gene in the mouse lung epithelium resulted in a few large epithelial pouches and increased the *FGF10* expression [18, 19]. *Yy1* impacts on Sonic Hedgehog (*SHH*) expression [19], and *SHH* inhibits the expression of *FGF10* [20]. A previous transcriptome analysis of CPAM microcyst focused on the lung epithelium identified significant differences in gene expression between CPAM and healthy controls, indicating dysregulation of TGF-beta signaling in CPAM at both mRNA and protein levels [21]. In the development of CPAM, thyroid transcription factor 1 (*TTF-1*) was highly expressed in the bronchiolar-like epithelial cells of the cysts in CPAM types I, II, and III [22]. The Hox gene *Hoxb-5* was also highly expressed in human CPAM lung tissues during the saccular and alveolar states of lung development [23]. Sex determining region Y-box 2 (*Sox2*) was reported to be involved in the differentiation of endoderm epithelium cells to basal cells by regulating the expression of *Trp63* [24], resulting in cystic lesions [25]. Interestingly, *Sox2* was exclusively detected in the affected area of type II CPAM but not in type I CPAM [24], suggesting distinct cell types in the affected epithelium and disparate origins for type I and type II CPAM. Therefore, it is imperative to conduct separate investigation into the cellular composition and gene expression profiles of these different CPAM types.

Currently, surgical excision of cysts remains the primary postnatal therapy for CPAM due to the lack of molecular therapeutic methods, attributed to the incomplete understanding of its pathogenesis. Maternal

steroids and betamethasone have shown 47–82% effectiveness in prenatal treatment [26, 27], potentially by restoring the balance of cell proliferation and apoptosis during CPAM development [28, 29]. However, there is limited information to identify which patients benefit most from steroid administration and determine optimal timing [29]. Therefore, understanding the genes or signaling pathways involved in different types of CPAM could aid in developing therapeutic targets and tailoring treatment strategies for optimal patient benefit.

Single-cell RNA sequencing (scRNA-seq) has provided new insights into lung development by revealing the heterogeneity in cell populations and the landscape of intercellular interactions [30, 31]. Disturbed communication between epithelial cells and mesenchymal cells is suggested to be involved in the development of CPAM [32, 33]. Using scRNA-seq, Zhang revealed the abnormal proliferation of epithelial cells in CPAM and the pseudo-time trajectory of subpopulation of epithelial cells [34]. However, a comprehensive evaluation of the interactions among cells in CPAM has not yet been determined.

In this study, the cellular landscape of CPAM was mapped, and intercellular communication networks were further analyzed using scRNA-seq based on the type II–III scRNA-seq data [34]. The underlying cellular and molecular events related to CPAM disease were identified, deepening our understanding of its etiology and cell composition, which may provide new insights for treatment. Additionally, we integrated our RNA-seq data of type II CPAM to investigate the expression profile and described the genes or signaling pathways that potentially associated with CPAM pathogenesis. Understanding the genes or signaling pathways that are involved in the CPAM pathogenesis may help to develop potential molecular therapeutic targets and identify patients who would benefit most from precise treatment.

Methods

Subject preparation

Fifteen patients with CPAM from Children's Hospital, Zhejiang University School of Medicine were included for analysis. The diagnosis of CPAM was confirmed by histopathological analysis following surgical procedures, classified according to the Stocker classification [4]. During surgery, specimens from the cystic area and adjacent normal tissue served as case and control samples, respectively. Tissues were promptly snap-frozen and stored in liquid nitrogen until for further analysis. All patients were carefully screened for infection at the time of surgery, and those with detected infections were excluded from the analysis.

This study was approved by the ethics committee of the Children's Hospital, Zhejiang University School of

Medicine (2022-IRB-001), and written informed consent was obtained from the patient's legal guardians.

RNA extraction, cDNA library preparation and RNA sequencing

Total RNA was extracted from frozen tissues of CPAM subjects using TRIzolR Reagent (Invitrogen), as described in the manufacturer's protocols. RNA concentration and integrity were determined with an Agilent Bioanalyzer (Agilent Technologies, CA, USA). A total of 3 µg RNA per sample was used as input for RNA sample preparation. Sequencing library was generated using VAHTS Universal V6 RNA-seq Library Prep Kit for Illumina® (NR604-01/02), according to the manufacturer's recommendations. Index codes were added to attribute sequences to each sample. Eligible libraries were sequenced on the Novaseq 6000 (Illumina) platform, generating 150 base paired-end reads.

Transcriptome analysis of bulk RNA-seq data

The sequencing quality was evaluated by fastqc software, and low-quality reads were filtered by the Trimmomatic tool [35]. After preprocessing, clean reads were aligned to the human hg38 reference genome with Hisat2 software [36]. Gene expression levels were quantified using StringTie [37], with unexpressed or lowly expressed genes filtered out. Batch effects were corrected using ComBat-seq through the sva R package [38] with default parameters. Principal component analysis (PCA) was performed to visualize clustering between case and control samples based on the transcriptome data. Differentially expressed genes (DEGs) between case and control samples were identified using the DESeq2 package in R [39]. DEGs were selected with adjusted P -value < 0.05 and $|\log_2$ fold change (\log_2 FC)| > 1 . DEGs were visualized using volcano plot and heatmap created with the R packages 'ggplot2' and 'pheatmap', respectively.

Functional annotation and protein-protein interaction networks

To predict the biological processes involved in the pathogenesis of CPAM for those DEGs, Gene Ontology (GO) annotation was performed using the clusterProfiler R package [40]. Adjusted P -value of < 0.05 was considered as a significant event. Gene Set Enrichment Analysis (GSEA) was conducted to elucidate molecular mechanism at the whole transcriptome level by comparing the case and controls against the KEGG pathways database. Protein-protein interaction (PPI) based on DEGs enriched in the significant GO terms were predicted by STRING [41]. The PPI network was visualized using Cytoscape software [42].

Identification of gene modules for distinguishing CPAM from control

To identify genes specific to CPAM, iterative weighted gene correlation network analysis (iWGCNA) (<https://github.com/cstoekert/iterativeWGCNA>) was used to construct a co-expression network based on the transcriptome datasets and detect modules or clusters that comprise with highly correlated genes within the network [43]. Random forest analysis with 100 iterations of cross-validation was applied to rank the CPAM-related gene modules in descending order based on the mean decreased Gini index in iWGCNA. For each gene module, an eigengene was calculated and compared between CPAM and control groups, representing a normalized expression value derived from all genes within the module. Correlation coefficients between each module and clinical characteristics were also computed based on the eigengene.

scRNA-seq analysis of CPAM patient

scRNA-seq data from type II-III CPAM samples was obtained from Zhang's published paper [34], and was analyzed with Seurat package (<https://github.com/satijalab/seurat>). Cells expressing fewer than 200 or more than 2500 genes were filtered out, and cells expressing over 20% of genes derived from the mitochondrial genome were removed. Expression data were normalized and scaled by `NormalizedData` and `ScaleData` function in Seurat. Clustering was performed using the k-nearest neighbor method with the `FindClusters` function and visualized using t-distributed stochastic neighbor embedding (t-SNE) with the `RunTSNE` function. Marker genes for each cell cluster were identified by the `FindAllMarkers` function. The expression profiles of these markers in epithelial cells were normalized using Z-score as signature score and was compared between CPAM cases and controls in our bulk RNA-seq data by Wilcoxon test. *P*-value threshold was set as 0.05.

Analysis of cell-cell communication

We conducted quantitative inference and analysis of cellular communication networks using CellChat [44] based on CPAM scRNA-seq data. Seurat data was input into CellChat (<https://github.com/sqjin/CellChat/>; accessed on 01/07/2022) to identify potential ligand-receptor interactions using the default parameters. Human CellChatDB, including 1,939 validated interactions, served as the interaction database. This analysis encompassed cellular interactions and communication categorized by secreted signaling (61.8%), extracellular matrix-receptor interactions (21.7%), and cell-to-cell contact genes (16.5%). The aggregated cell-cell communication network was established by quantifying interactions and visualized in a circular plot format. Following this, we

identified significant cell-cell communication facilitated by interactions. CellChat provided insights into communication patterns among different epithelial cell subtypes, evaluating the number of incoming and outgoing communication patterns using Cophenetic and Silhouette metrics. Prominent ligands involved in outgoing or incoming communication signaling were identified and visualized using heatmaps and bar graphs. Pathway enrichment analysis for these ligands was conducted by EnrichR tool [45] using WikiPathways 2021. Heatmaps displaying the top ten enrichment terms (lowest adjusted *P*-values) were generated with the web tool Morpheus [46]. Finally, the global communication patterns among cell types and signaling pathways were visualized with river plots.

Quantitative real-time PCR

We performed experimental validation of the expression of *AGR3* (Anterior Gradient 3, Protein Disulphide Isomerase Family Member), *COL4A1* (Collagen Type IV Alpha 1 Chain), *COL4A2* (Collagen Type IV Alpha 2 Chain), *COL4A4* (Collagen Type IV Alpha 4 Chain), *EDN1*, *FPR2*, *CHMR1* in 6 cDNA samples from CPAM and matched controls from lung tissues using quantitative PCR (qPCR). The PCR primers targeting the coding regions of these genes were designed and synthesized by Ykang (Hangzhou, China). All qPCR reactions were conducted in a total of 10 μ L volume, comprising 5 μ L 2 \times SYBR Green I Master Mix (Promega), 1 μ L 10 nM of each primer, and 2 μ L of 1:20 diluted cDNA in 96-well plates with QuantStudio™ 7 Flex (Applied Biosystems). All reactions were performed in triplicate under the following conditions: initial denaturation at 95 °C for 5 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 30 s. Relative expression levels were calculated using the standard curve method normalized to the *GAPDH* housekeeping gene. Standard curves for these genes and *GAPDH* were constructed using five-serial 4-fold dilutions of cDNA samples.

Results

Gene expression profiles of CPAM reveals deregulated transcriptional control

Six cystic tissues with type II CPAM as cases and fifteen normal tissues as controls were included into final analysis. As shown in Table S1, the average age at surgery for all subjects was 18 months, with type II CPAM patients averaging 17 months at the time of surgery.

After quality control, an average of 21,382,664 reads per sample was generated, with 97.13% of them mapped to the human reference genome. PCA plot was used to investigate the RNA-seq quality and heterogeneity in the transcriptomes across the dataset. The principal component 1 (PC1) revealed a striking separation between

CPAM and control samples (Fig. S1A), while PC2 suggested a sex difference among these samples with approximate 14% of variance. Sex differences in gene expression have been documented in lung tissue [47], although the impact of these DEGs may be modest [48]. We investigated DEGs between female and male controls, identifying 52 up-regulated and 69 down-regulated genes in female (Fig. S1B). The GO analysis suggested that those up-regulated genes in females were mainly enriched in homophilic cell adhesion via plasma membrane adhesion molecules and cell-cell adhesion via plasma-membrane adhesion molecules related biological process, while those down-regulated genes were enriched in adaptive immune response, lymphocyte, B cell mediated immunity and immunoglobulin mediated immune response (data not shown), consistent with previous findings of sex-biased genes enrichment [47]. We further investigated these DEGs in the case and control set, identifying 29 differently expressed genes (Table S2). However, these findings may be influenced by the small number of male CPAM cases, with only one male case included in the analysis. GO analysis for these 29 genes did not reveal any significant signaling pathways or biological process. All CPAM and control samples were combined for comprehensive DEG analysis.

A total of 2,618 mRNAs met the criteria for DEGs based upon the adjusted *P*-value threshold of 0.05 and a fold change threshold of 2. Among these, 1,856 genes were up-regulated and 762 genes were down-regulated (Fig. 1A-B, Table S3). The heatmap depicting the top 50 up-regulated and top 50 down-regulated genes effectively visualizes these significant DEGs (Fig. 1C). Notably, among these top expressed genes, genes such as *SOX9* (SRY-Box Transcription Factor 9) and *MMP11* (Matrix Metalloproteinase 11) are involved in epithelial-mesenchymal transition (EMT) [49] which is involved in many different lung diseases [50].

Functional variation of differentially expressed genes

To further investigate the DEGs, we performed GO enrichment analysis. The results showed that the up-regulated genes were markedly enriched in several biological processes related to cilia, including cilium organization, cilium assembly, microtubule-based movement, cilium movement, axoneme assembly, microtubule bundle formation, cilium or flagellum-dependent cell motility, cilium-dependent cell motility, and epithelial cilium movement involved in extracellular fluid movement (Fig. 2A, Table S4). These DEGs also exhibited enrichment in molecular functions such as antigen binding, immunoglobulin receptor binding, and cellular components such as immunoglobulin complex and motile cilium (Table S5). The enrichment indicated the potential roles for cilia in the CPAM, possibly involving signaling

transduction or extracellular fluids movement [51], which is consistent with recent transcriptome studies highlighting aberrant ciliary functions in its development [52, 53].

Conversely, down-regulated genes were enriched in processes related to positive regulation of response to external stimulus, regulation of inflammatory response, positive regulation of defense response, regulation of cytosolic calcium ion concentration, and positive regulation of cytokine production (Fig. 2B, Table S4). These down-regulated genes also showed enrichment in molecular functions such as immune receptor activity and carbohydrate binding, as well as cellular components such as integrator complex and secretory granule membrane (Table S5). The dysregulated immune response may contribute to epithelium damage and subsequent airway remodeling [54], which can lead to chronic inflammation and recurrent pulmonary infection in patients with CPAM [55].

Additionally, we utilized GSEA to identify enriched hallmark gene sets based on KEGG pathways. Notably, the genes in type II CPAM were significantly expressed in pathways related to neuroactive ligand-receptor interaction, featuring receptors located on the cell plasma membranes such as *GRIN3B* (Glutamate Ionotropic Receptor NMDA Type Subunit 3B), *LPAR3* (Lysophosphatidic Acid Receptor 3), *GLRB* (Glycine Receptor Beta), and *GPR156* (G Protein-Coupled Receptor 156), which facilitate the signal transduction from extracellular environment into cells [56]. Other enriched pathways included those involving motor proteins, metabolism of xenobiotics by cytochrome P450, and cytokine-cytokine receptor interaction which is associated with inflammation and immune response [57] (Fig. 2C, Table S6).

A total of 867 DEGs that were significantly enriched in GO terms were uploaded to the STRING database to build a PPI network. The cutoff criterion for PPI inclusion was set at a stringent PPI score of >800. Subsequently, the resulting PPI network, composed of 341 nodes and 704 interactions, was visualized using Cytoscape software (Fig. 2D). The top five DEGs that were determined based on their degree of interaction were listed as follows: *EDN1* (Endothelin 1) (degree=25), *MCHR1* (Melanin Concentrating Hormone Receptor 1) (degree=24), *GPR17* (G Protein-Coupled Receptor 17) (degree=24), *FPR2* (Formyl Peptide Receptor 2) (degree=24), and *CHRM1* (Cholinergic Receptor Muscarinic 1) (degree=22).

Gene modules distinguish CPAMs from controls

As a result of iWGCNA, 158 gene modules were identified using the random forest algorithm and ranked based on the mean decreased Gini index. The top 10 gene modules that distinguished CPAMs from controls were presented in Fig. 3A, with module P2_I10_M24 occupying

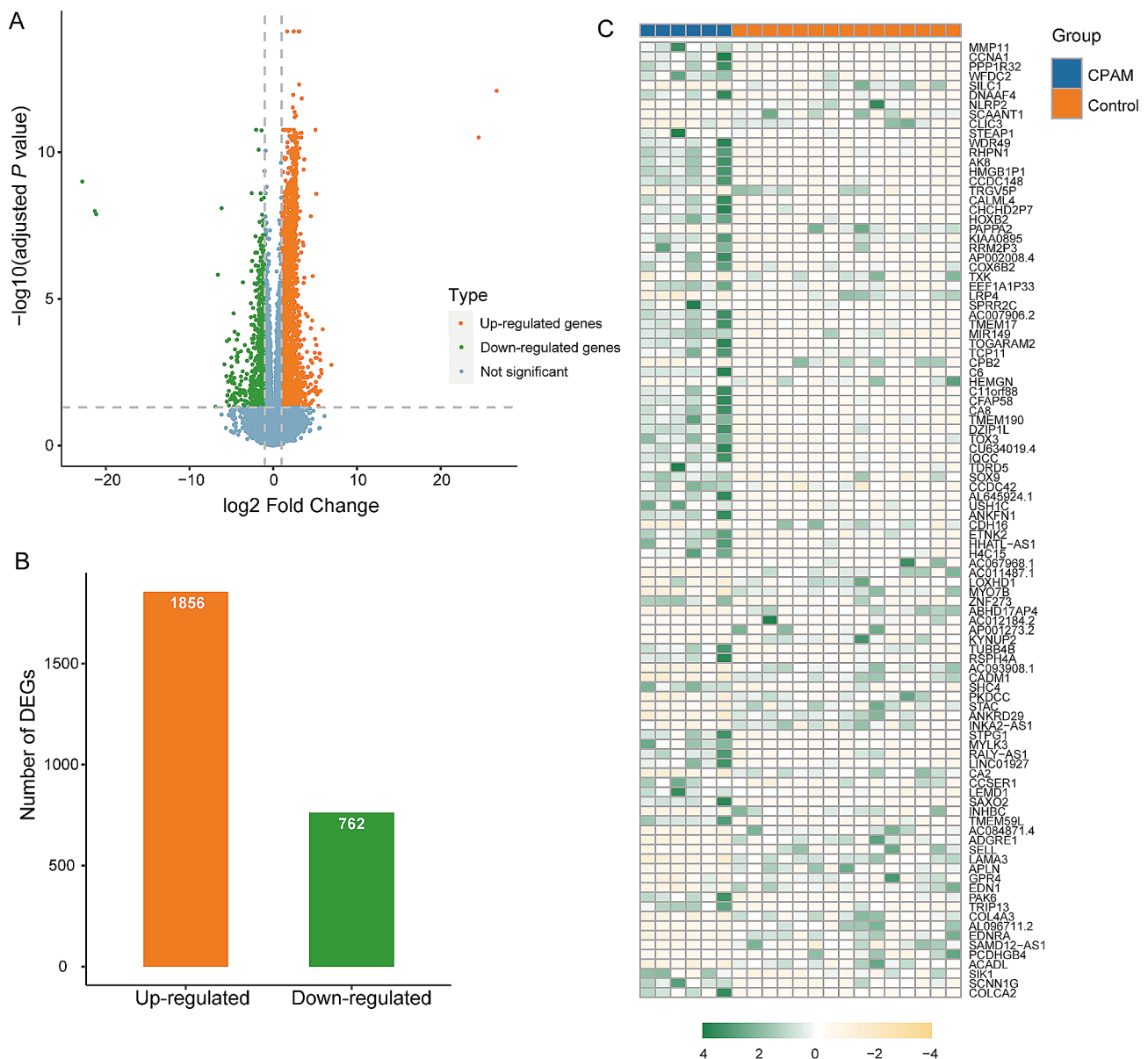


Fig. 1 The differential expression of CPAM. **(A)** Volcano plot showed gene expression profiling between the two groups. Each point represents a gene, with vertical lines indicating a 2-fold change and the horizontal line corresponding to an adjusted P -value of 0.05. **(B)** The amount of up-regulated and down-regulated genes. **(C)** Heatmap of top 50 up-regulated and top 50 down-regulated genes

the highest position, followed by P1_I25_M1 and P3_I8_M9. Hub genes in module P2_I10_M24 such as *PIFO* (*CIMAP3*, Ciliary Microtubule Associated Protein 3), *CFAP221* (Cilia And Flagella Associated Protein 221), and *DNAH7* (Dynein Axonemal Heavy Chain 7), along with genes such as *DRC3* (Dynein Regulatory Complex Subunit 3), *RSPH4A* (Radial Spoke Head Component 4 A), *CFAP43* (Cilia And Flagella Associated Protein 43), *DNAH11* (Dynein Axonemal Heavy Chain 11), *DNAH2*, and *CFAP77* (Cilia And Flagella Associated Protein 77) in P1_I25_M1 module (Table S7), were found to be essential

for ciliary ultrastructure, ciliary stability, and ciliary beating [58–63].

The module eigengene (ME), representing the expression value of P2_I10_M24, was significantly higher in CPAM cases than in controls (Fig. 3B, $P=7.4 \times 10^{-5}$), suggesting an up-regulation of genes in this module in CPAM. Notably, the top 10 modules were all correlated with type II CPAM but independent of factors such as age, sex, cystic position, and disease duration (Fig. 3C), further confirming that sex and other factors did not affect the analysis when all samples were combined. Among these modules, P2_I10_M24, P1_I25_M1,

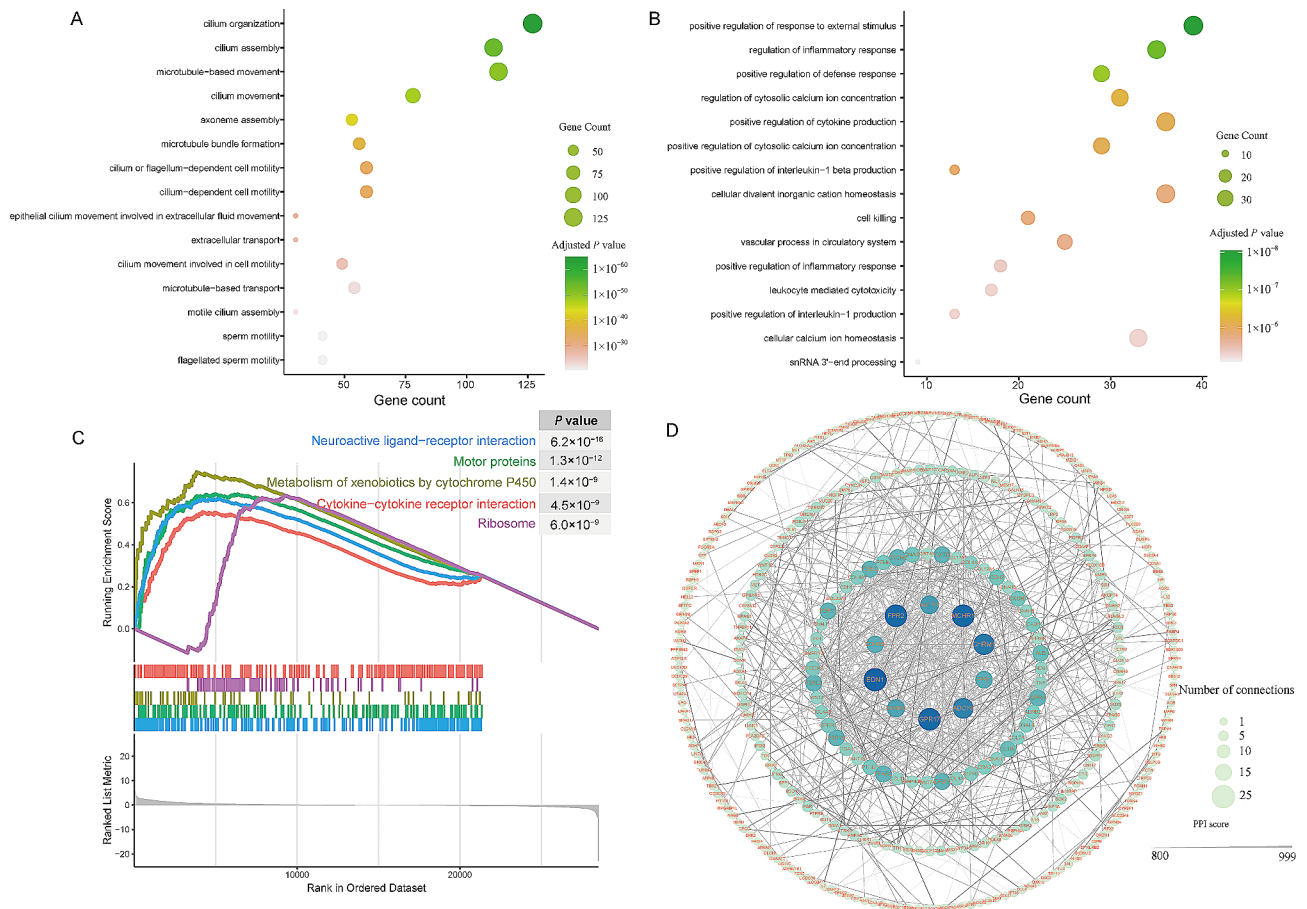


Fig. 2 Functional annotation of DEGs. **(A, B)** GO enrichments of significantly up-regulated and down-regulated genes were displayed by top 15 significantly adjusted P -values, respectively. The size of the bubble indicated the adjusted P -value, and the color of the dot indicated the number of genes enriched in the GO term. **(C)** GSEA of whole transcription from RNA-seq, listing the top5 KEGG pathways. **(D)** Constructed PPI network of DEGs. Nodes represent proteins, and the size of nodes indicates the number of interactions. The edges denote the interactions between two proteins, and the width of an edge denotes the PPI score of the interaction

P5_I8_M9, P5_I8_M1, and P1_I25_M7 showed a positive correlation with CPAM, while the remaining modules were negatively correlated with the disease, indicating that both up-regulated and down-regulated genes could be associated with the development of CPAM (Fig S2) [52, 53, 64]. Furthermore, the CPAM related modules with a positive coefficient were strongly correlated each other (Fig. 3D, coefficient > 0.7), indicating these modules can effectively discriminate CPAM from control groups. We thus defined these top modules as the ‘CPAM-associated module’ for distinguishing CPAMs from controls.

Integrated analysis of RNA-seq and scRNA-seq from CPAM patients

We obtained a total of 14, 814 cells for scRNA-seq analysis after quality control. A high consistency between the bulk RNA-seq and scRNA-seq data was observed (Fig. 4A, $P < 2.2 \times 10^{-16}$). As shown previously, epithelial cells increased and mesenchymal cells decreased in cystic areas of CPAM [34]. The obtained cells could be

divided into 9 cell types, including epithelial cells, monocytes, B-cells, fibroblasts, CD8+T-cells, macrophages, neutrophils, endothelial cells, and NK cells [64]. Given that CPAM is characterized by the proliferation of distal bronchioles leading to the suppression of alveolar formation [65], our primary focus was on the epithelial cells, which are predominant in bronchi and alveoli. We identified 970 marker genes specific to epithelial cells. The signature scores of epithelial cells were calculated based on the normalized expression of each gene using a Z-score. Notably, the signature score of epithelial cells was significantly lower in CPAM cases than in controls (Fig. 4B, $P = 1.8 \times 10^{-3}$), suggesting a disturbance in the regulation of epithelium-specific genes in CPAM. When we overlapped these 970 genes with DEG dataset from bulk RNA-seq data, 38 genes were differently expressed between CPAM and controls (Table S8). Additionally, we investigated these 970 genes within iWGCNA CPAM-associated modules and found 16 overlapping genes (Table S9). Among these 16 genes, *AGR3* and *SLC11A1*

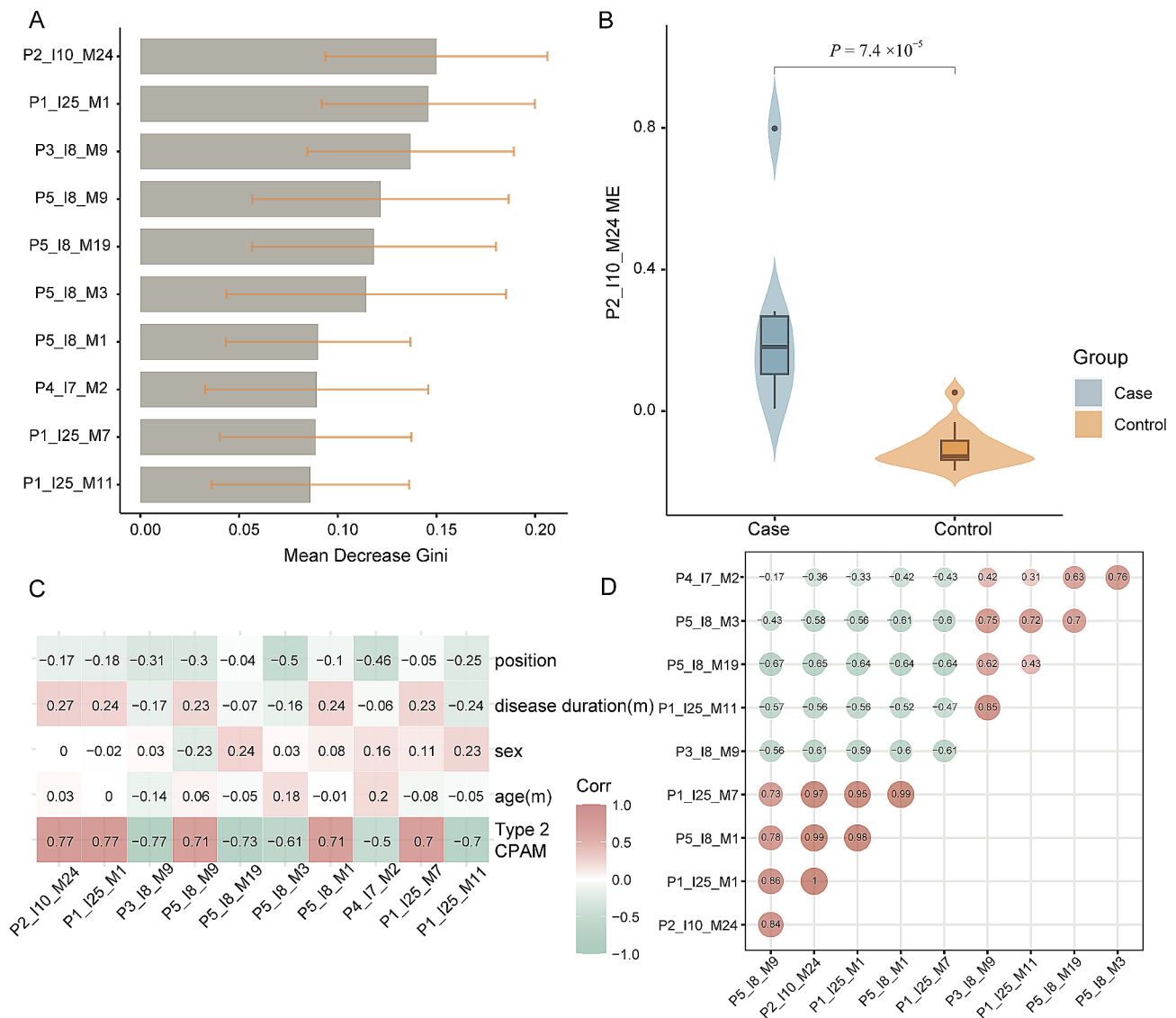


Fig. 3 Identification of essential gene modules for CPAM discrimination. **(A)** Top10 modules ordered by the mean decreased Gini for distinguishing CPAM from control. **(B)** P2_I10_M24 ME in cases and controls (Wilcoxon test, $P = 7.4 \times 10^{-5}$). **(C)** The correlation of top10 module ME and clinical parameters. Numbers in each box represent correlation coefficients, only P values that are significant are displayed by *. **** represented the $P < 0.001$. **(D)** Inter-modular correlation of top10 modules

were the only two genes indicated in DEG dataset as well (Fig. 4C).

We further classified the epithelial cells into alveolar type 1 (AT1), alveolar type 2 (AT2), club cells (club), and ciliated cells according to the expression of marker genes (Fig. 4D and Fig. S3A, Table S10). As indicated in Fig. S3, marker genes of AT1 are primarily enriched in signaling pathways such as focal adhesion and ECM-receptor interaction (Fig. S3B). In contrast, marker genes of AT2 are implicated in pathways such as complement and coagulation cascades, phagosome, and pertussis (Fig. S3C). Marker genes of club and ciliated cells were associated with ribosome and motor proteins (Fig. S3D and Fig. S3E). As shown in Fig. 4E, *AGR3* is mainly expressed AT1,

club, and ciliated cells, consistent with previous findings that *AGR3* is detected in multiciliated organs such as the colon and lung, specifically in the ciliated cells of the airway epithelium [66]. *SLC11A1* is mainly expressed in AT1 cells (Fig. 4F). *SLC11A1* is a proton/divalent cation antiporter that functions in host resistance to certain pathogens through regulating iron and manganese metabolism [67]. It also modulates proinflammatory cytokines and activation of innate lymphocytes [67]. Downregulated *SLC11A1* in CPAM might affect the immune balance and consequently impact lung epithelial development [68].

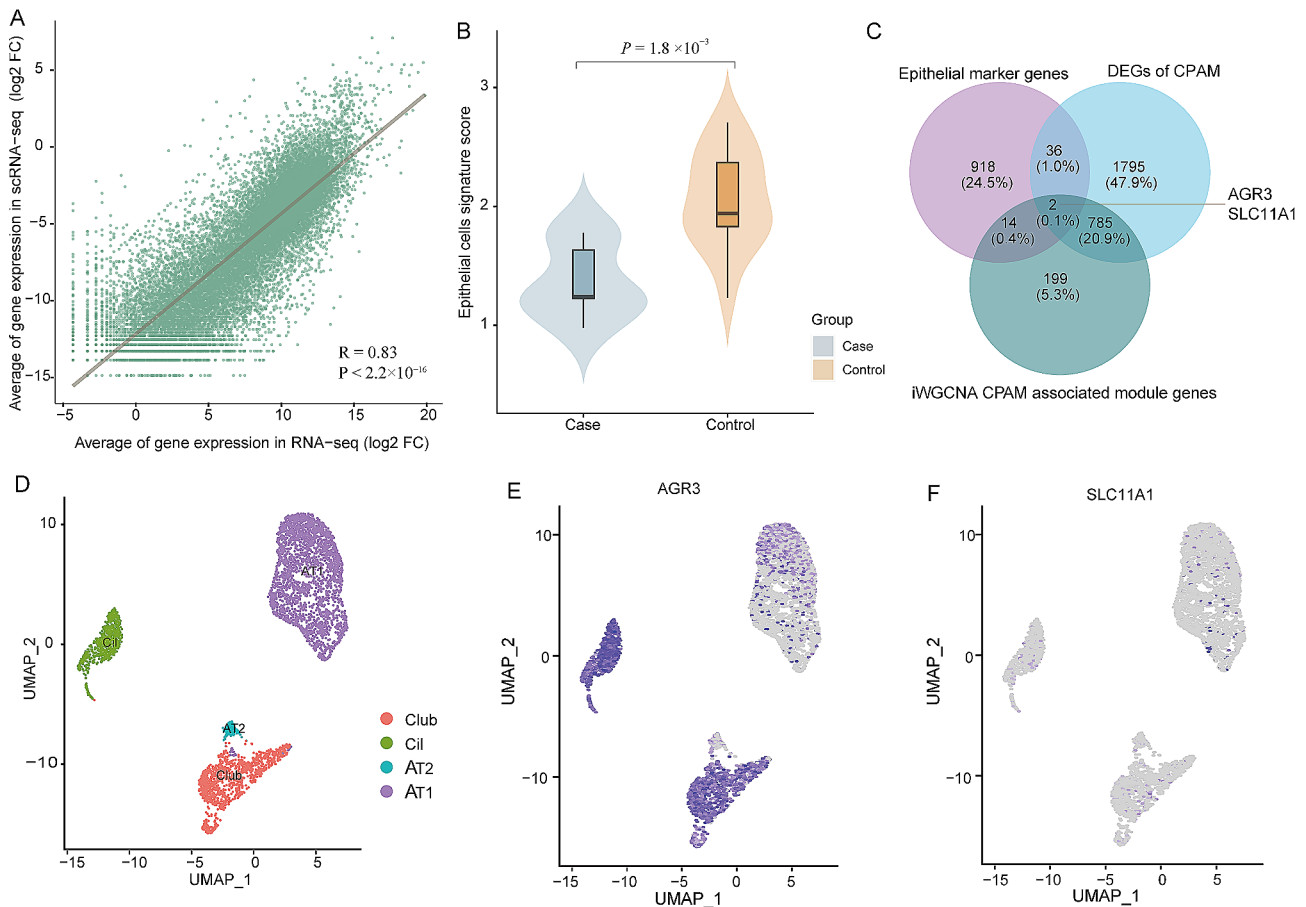


Fig. 4 Integration of scRNA-seq and bulk RNA-seq analysis of CPAM. **(A)** Correlation between the RNA-seq and scRNA-seq in gene expression (log2 FC). The correlation coefficient was calculated by Pearson's test ($R = 0.83$, $P < 2.2 \times 10^{-16}$). **(B)** Comparisons of epithelial cell signature scores in CPAM case and control groups ($P = 1.8 \times 10^{-3}$). P -value was obtained from two-sample t-test. **(C)** Venn diagram of DEGs of CPAM, the iWGCNA CPAM-associated module genes and epithelial cell marker genes. **(D)** Umap of subtypes for epithelial cells from CPAM patients. **(E, F)** Feature plot of *AGR3* and *SLC11A1* in epithelial cells

CPAM-associated epithelial cells receive different communication patterns

We calculated the cell-cell communication network among the epithelial sub-cell types and visualized it using circle plots either by number of interactions (Fig. S4A, upper left) or by interaction weights (Fig. S4A, upper right) between two cell types. As shown in figure, signaling from AT1 cells was more active than that from AT2 cells, club cells, and ciliated cells, based on both interaction net number and interaction weight/strength (Fig. S4A). The results showed that cell-cell communication between AT1 cells and other cells was more pronounced.

We also explored the communication patterns to understand how these epithelial sub-cell types and signaling pathways coordinate. Notably, AT1 cells emerged as the major contributors to outgoing signals (Fig. 5A). Signaling molecules such as LAMININ, COLLAGEN, APP, MIF, and MK exhibited significant involvement in outgoing cellular interactions (Fig. 5A). We identified the correspondence between inferred latent communication

patterns to decode outgoing communication patterns. Four distinct signaling patterns were identified (Fig. S4B), each uniquely associated with specific epithelial cell subtypes that predominantly contributed to outgoing communication: pattern #1 (AT2 cells), pattern #2 (Cil cells), pattern #3 (AT1 cells), and pattern #4 (Club cells) (Fig. 5B). In order to identify the pathways specifically associated with the outgoing communication of AT1, we selected a list of ligands (Table S11) that exert the most influence on communication pattern 3 (Fig. 5B). We found that the ligands to AT1 were mostly enriched in pathways such as PI3K-AKT signaling, focal adhesion, and epithelial to mesenchymal transition (Fig. 5C). The receptors of AT1 cells were enriched in pathways such as gastrin signaling pathway and epithelial to mesenchymal transition (Fig. 5D).

We also identified contributors to incoming signals within cell populations. By evaluating ligands overexpressed across all epithelial cells, we confirmed that AT2 cells were the foremost recipients of incoming signaling

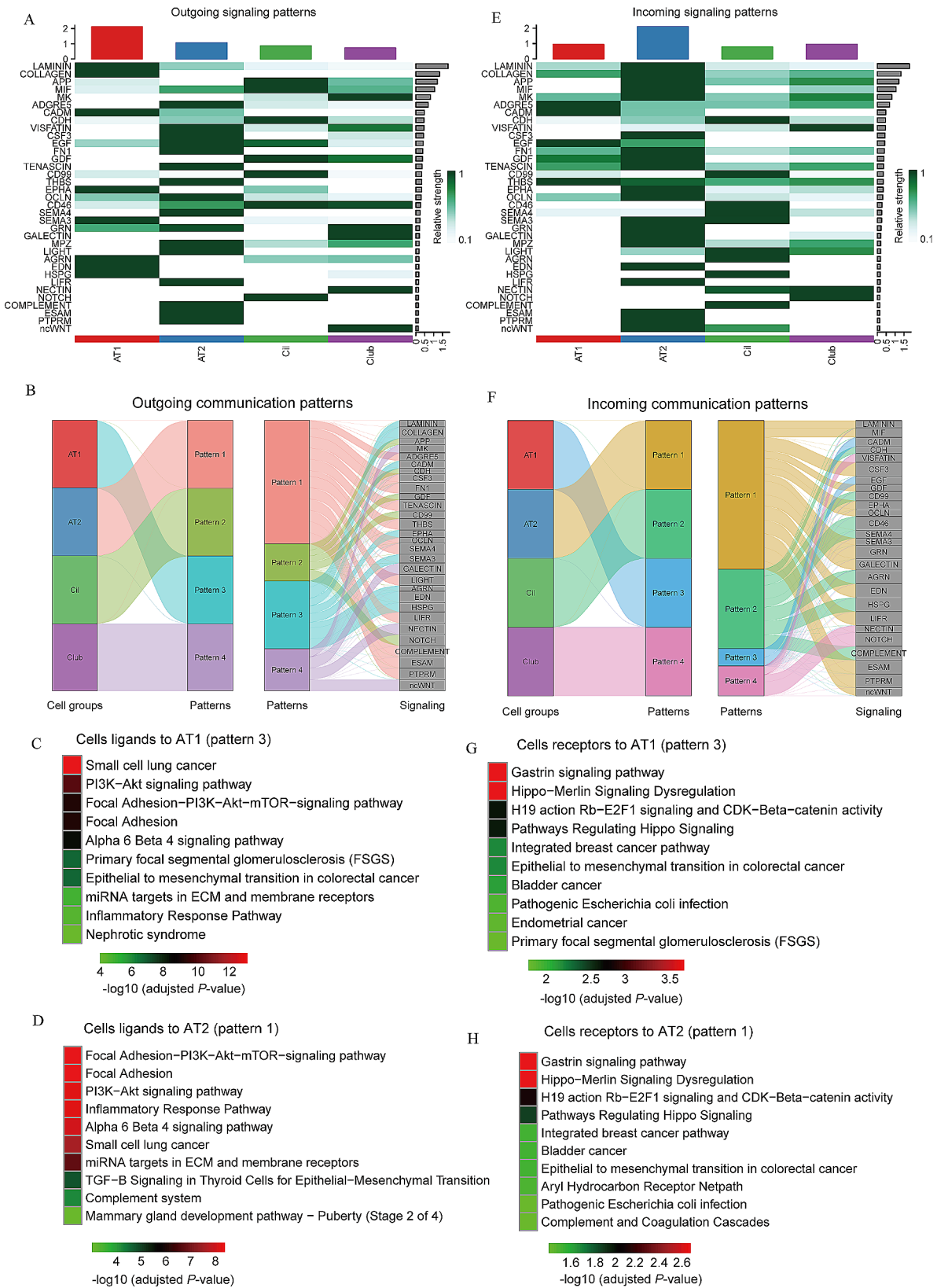


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Outgoing and incoming communication patterns between CPAM-associated epithelial cells. **(A)** Outgoing patterns for signaling of AT1, AT2, club and ciliated cells. **(B)** Outgoing communication patterns of epithelial cell subtypes. The flow thickness represents the contribution of a signaling pathway to each mode of communication. **(C, D)** Pathways identified by outgoing communication pattern 3 for gene sets of AT1 ligands and receptors. The top 10 pathway terms were selected from WikiPathways 2021. **(E)** Incoming patterns for signaling of AT1, AT2, club and ciliated cells. **(F)** Incoming communication patterns of epithelial cell subtypes, which reveals the correspondence between the inferred patterns and cell groups, as well as signaling pathways. **(G, H)** Pathways identified by incoming communication pattern 1 for gene sets of AT2 ligands. The top 10 pathway terms were selected from WikiPathways 2021

(Fig. 5E). Notably, LAMININ, COLLAGEN, APP, MIF, and MK emerged as the most significant molecules involved in incoming signaling (Fig. 5E). Our analysis revealed four distinct incoming signaling patterns (Fig. 5F and Fig. S4C), with AT2 cells exhibiting strong associations with signaling pattern #1 (Fig. 5F). Notably, the ligands interacting with AT2 cells, identified within communication pattern #1 (Fig. 5F), were enriched for processes related to focal adhesion, PI3K-AKT signaling pathway, inflammatory response pathway, and epithelial to mesenchymal transition (Fig. 5G). Furthermore, epithelial to mesenchymal transition was also one of the signaling pathways predominantly enriched in AT2-specific receptors (Fig. 5H). These results suggest that the genes involved in PI3K-AKT signaling pathway and epithelial to mesenchymal transition play a crucial role in the development of CPAM [69] (Table S12).

Top ligand and receptor interactions between epithelial cell subtypes

The numerous ligand and receptor interactions among epithelial cells indicated the complexity of both outgoing and incoming signaling (Fig S5). Thus, to explore the top interactions involved in CPAM, we selected the signaling between AT1, AT2, club, and ciliated cells with a communication probability > 0.05. Among these, we found some strong interactions, such as MIF (ciliated cell) and CD74+CD44 (AT2), APP (ciliated cell) and CD74 (AT2) (Fig. 6A). Notably, collagen-related ligands and their receptors, including *COL4A1*, *COL4A2*, and *COL4A4*, which were activated in the epithelial to mesenchymal transition process and involved in PI3K-AKT signaling pathways [49], were the most predominant ligand-receptor pairs mediating interaction between or within AT1 cells and AT2, cilia or club cells (Fig. 6A and B). AT1 cells serve as the senders in this COLLAGEN signaling network (Fig. 6C). Furthermore, these genes were significantly decreased in our CPAM II cysts samples (Table S3, Fig S6), suggesting that disruption of these genes may impair the PI3K-AKT signaling pathways, resulting in an abnormal EMT process and consequently leading to the formation of CPAM.

Validation of gene expression

To confirm the RNA-seq data, we analyzed the mRNA expression levels of *AGR3*, *COL4A1*, *COL4A2*, *COL4A4*, *EDN1*, *FPR2*, and *CHMR1* in 10 tissue samples from

CPAM and control groups using qPCR (Fig. S6). The mRNA expression level of *AGR3* was significantly higher in CPAM patients than in control tissues. Conversely, *COL4A1*, *COL4A2*, *COL4A4*, *EDN1*, *FPR2*, and *CHMR1* were significantly downregulated in CPAM patients. The qPCR results showed good consistency with the RNA-seq data.

Discussion

Increasing evidence suggests the distinct mechanisms underlying different types of CPAMs. Type II CPAMs are believed to result from bronchial atresia or obstruction [70], with evidence indicating focal lung development arrest during the pseudoglandular stage of gestation (5–17 weeks) [5]. Unlike other types of CPAM associated with specific genetic mutations, no genetic alterations have been identified in type II CPAM. Therefore, investigating the pathogenesis of CPAM type-specifically is crucial for developing effective diagnostic and prognostic assessments. In this study, we focus on exploring the transcriptome expression and cell-cell communications of type II CPAM using RNA-seq and single-cell RNA-seq technologies.

Previous research has identified aberrantly expressed genes and proteins in both fetus and postnatal type II CPAM [5, 24, 71]. Consistent with this, our data revealed significant upregulated and downregulated DEGs between postnatal CPAM cystic and control tissues. Notably, *SOX9* and *SOX2* (SRY-Box Transcription Factor 2) are predominantly expressed in CPAM cysts. In human developing lungs, *SOX9*+progenitors are enriched in distal epithelial tips [72], giving rise to alveolar and airway lineages [73]. *SOX2* is localized in the proximal epithelium [72], with proximal-distal patterning regulated by PI3K signaling critical for lung morphogenesis [72].

It has been demonstrated that reduced expression of *SOX2* is crucial for regulating branching morphogenesis [25]. Ectopic expression of *Sox2* in epithelial cells has been shown to prevent airway branching morphogenesis [24, 25]. Our results corroborate previous studies that exclusively detected *SOX2* expression in type II CPAM but not in control or type I CPAM [24], highlighting the pivotal role of *SOX2* in the development of type II CPAM. This finding also suggests the possibility of distinct origins for type I and type II CPAM.

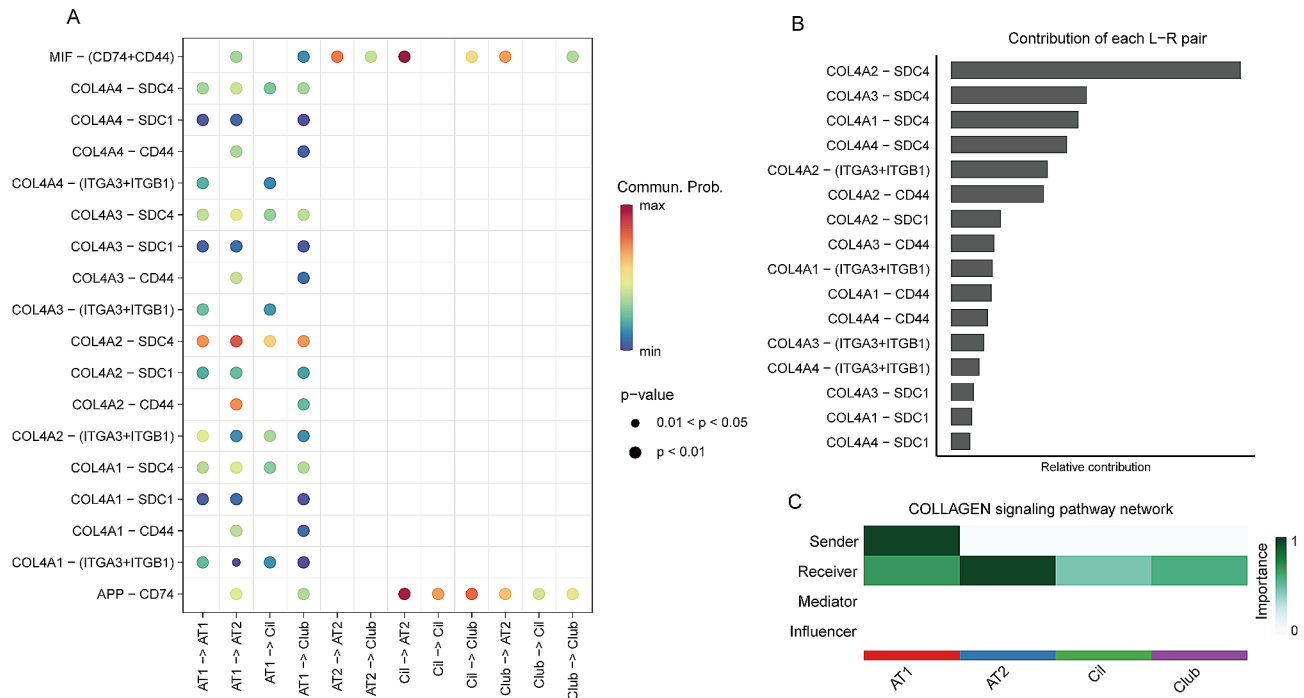


Fig. 6 Top interactions between epithelial cells. **(A)** The key ligand-receptor pairs between AT1, AT2, club, and ciliated cells. **(B)** The contribution of each ligand-receptor pair to the COLLAGEN signaling pathway. **(C)** Identification of the COLLAGEN signaling pathway roles of cell groups.

Previous studies have highlighted the essential role of *FGF10* (Fibroblast Growth Factor 10), its receptor *FGFR2b*, and inhibitor *SHH* (Sonic Hedgehog Signaling Molecule) in the pathogenesis of CPAM [16, 17, 20]. However, our investigation did not reveal any significance in the expression of these genes between the type II CPAM cysts and control samples. This observation is consistent with the findings from studies involving post-natal tissues from type II CPAM and normal controls [16], suggesting that these genes likely exert their essential roles during lung development rather than in the developed CPAM tissues.

Our GO analysis revealed that the top up-regulated transcripts in CPAM were significantly enriched in cilium-related biological process, including cilium assembly and movement. This suggests a potential impairment in cilia function in CPAM patient, which is critical for normal mucociliary clearance in the airway, essential for protecting against various insults. The result is consistent with our previous and other enrichment analysis of RNA expression changes in CPAM [52, 53]. Decreased *SMAD6* in our type II CPAM cysts was observed but TGF- β signaling pathway was not enriched in our dataset, suggesting that expression profiles may vary depending on the CPAM type [53]. This observation aligns with the pathological finding that type II CPAM cysts are typically lined by ciliated cuboidal epithelium. The top down-regulated genes were enriched in inflammatory response process, suggesting disrupted immune response

regulation in cysts with CPAM and a potential presence of chronic inflammation, as supported by histological signs of inflammation in CPAM cysts [74]. Such imbalances in immune process may lead to difficulties in effectively mounting a defense against respiratory pathogens and repairing damage caused by infections, potentially contributing to the increased frequency of respiratory infections in children with CPAM. Furthermore, consistent with previous enrichment analysis, our KEGG analysis demonstrated significant enrichment in gene sets related to cytokine-cytokine receptor interaction pathway and neuroactive ligand-receptor interaction [52], indicating that genes involved in cell growth, differentiation, or inflammation may play pivotal roles in the development of CPAM.

The PPI analysis showed that proteins with top connections are predominantly enriched in down-regulated genes. Among them, *EDN1*, *FPR2*, *GPR17*, and *CHRM1* are involved in G protein-coupled receptor (GPCR) signaling pathways, crucial for various cellular processes such as organ and embryonic development, cell differentiation, apoptosis, tissue homeostasis, and the immune. *EDN1*, the predominant isoform in pulmonary endothelial cells, binds to its receptor EDNRA, activating a diverse network of signaling pathways such as MAPK, PI3K-AKT, PKC pathways to promote cell proliferation, cell growth, and survival [75]. *FPR2*, the most promiscuous member of formyl peptide receptors, can modulate both pro- and anti-inflammatory response [76]. Studies

on *Fpr2/3* knockout mice suggested increased susceptibility to infection, uncontrolled inflammation, and pulmonary dysfunction compared to wild-type animals [77]. *GPR17* negatively regulated pulmonary inflammation induced by house dust mite [78], while *CHRM1*, widely expressed in lungs, regulates bronchoconstriction, airway epithelial cell proliferation, and inflammation [79]. Moreover, *EDN1*, *GPR17*, *FPR2*, and *CHRM1* are involved in calcium ion homeostasis or transportation. Dysregulation of ion balance in airway epithelial cells may impair electrolyte and PH in airway liquid, which is important to maintain ciliary function and mucociliary clearance to protect the host from inhaled pathogens [80]. Consequently, it damages the ability of innate immunity in CPAM patients, resulting in respiratory distress or recurrent pneumonia.

Our iWGCNA analysis identified 10 modules that associated with CPAM cysts, which are either positively regulated or negatively regulated in CPAM. Specifically, ciliary formation or function related genes are indicated in some of these modules, suggesting their important roles in CPAM pathogenesis. Integration of these module genes with DEGs from bulk and epithelial marker genes highlighted *AGR3* and *SLC11A1*. *AGR3* is up-regulated in CPAM cysts and belongs to protein disulphide isomerase (PDI) family which regulates protein folding and calcium homeostasis. Studies on *Agr3* knockout mice have shown reduced calcium-dependent ciliary beat frequency and impaired mucociliary clearance in the airway [66]. Moreover, *AGR3* was overexpressed in many tumor tissues and could promote the stemness of cancer cells by activating Wnt/ β -catenin signaling [81]. Wnt signaling plays a crucial role in the developing lung by contributing to branching morphogenesis and airway formation [82]. Additionally, *AGR3* positively regulates proteins involved in airway epithelial junctions in chronic obstructive pulmonary disease [83], suggesting its role in disrupting normal cell-cell communication in CPAM cysts. On the other hand, *SLC11A1* is associated with the infectious and autoimmune diseases such as tuberculosis and type I diabetes mellitus [84, 85]. It modifies macrophage activation by regulating chemokine/cytokine genes such as tumor necrosis factor (TNF) and IL-1 β [86]. Recent studies have revealed diversity of immune cells and their impact on epithelial maturation during lung development [68], where IL-1 β plays a role in signal transduction in epithelial tip cells [68]. Down-regulation of *SLC11A1* in CPAM may disrupt IL-1 β signal transduction, which in turn affects the expression of *SOX9* and *SOX2*, and consequently leading to abnormal immune-epithelial communication, resulting in lung anomalies. Further experimental investigations are warranted to elucidate the immune functions of *SLC11A1* in CPAM.

Increased cell proliferation and decreased apoptosis during lung development were attributable to the disruption of normal mesenchymal-epithelial interactions, leading to the CPAM [33, 87]. Indeed, focal asynchrony proliferation between epithelial and stromal components was indicated in CPAM [87]. Increased epithelial cells and decreased mesenchymal cells in CPAM cystic area were further confirmed by previous single-cell RNA-seq analysis [34]. We observed significant differences in gene signature scores between CPAM and control groups. Furthermore, both identified epithelial cell ligands and receptors were associated with epithelial to mesenchymal transition pathways, suggesting that signals involved in cell communication and interaction may play a role in the abnormal regulation of CPAM development abnormal regulation of developing of CPAM.

We reanalyzed scRNA-seq data [34] to infer and investigate epithelial cell-cell communications and interactions. We found AT1 and AT2 are the predominant communication cells for outgoing and incoming signals with specific signaling communication patterns. Key ligands involved in these interactions included LAMININ, COLLAGEN, SEMA3, and EDN. LAMININ and COLLAGEN are key components of the extracellular matrix (ECM) and actively mediate interactions between epithelium and mesenchyme during the airway branching process [88, 89]. However, we found the COLLAGEN related genes are decreased in our type II CPAM cyst samples, suggesting disrupted epithelium-mesenchyme interactions.

SEMA3 and its receptor *NRP1* are essential during fetal pulmonary development; disruption of *Sema3-Nrp1* signaling in transgenic *Nrp1^{Sema-}* mice has been shown to impair alveolar-capillary interface formation and result in severe parenchymal anomalies at birth [90, 91]. One isoform of *EDN*, frequently connected with other proteins based on our PPI analysis, can activate the PI3K-AKT pathway and is involved in cell migration. We found that ligands from AT1 cells and those targeting AT2 cells are predominantly enriched in the PI3K-AKT-mTOR-signaling pathway, indicating an important role for genes in this pathway in the pathogenesis of CPAM.

The PI3K-AKT-mTOR signaling pathway is essential to cell cycle, cell proliferation, migration, and apoptosis. Genes in this pathway have been extensively studied and are known to be activated in tumorigenesis and progression of disease [92]. Various novel inhibitors targeting this pathway have been identified in preclinical studies and clinical trials in human cancers [93]. In lung development, Khattar et al. observed that PI3K activation and phosphorylated ATK (pAKT) expression levels are higher in the proximal lung epithelium and distal branch tips of early developing embryonic mice than in distal lung epithelium at later stages. Conditional knockout of PI3K

signaling in the developing lung epithelium resulted in impaired differentiation of airway epithelial cells [94]. Transcriptome analysis across several subtypes of congenital lung lesions, including CPAM, has indicated reduced gene expression in this pathway [69]. Additionally, three-dimensional cell culture studies using epithelial cells isolated from CPAM lesions have suggested intrinsic defects in epithelial function [69], highlighting the requirement of PI3K signaling for proper epithelial patterning during lung organogenesis. Drugs activating/blocking the ligands of the PI3K-AKT-mTOR signaling may be considered as therapeutic targets.

Limitation

Although limited by a relatively small sample size, our bulk RNA-seq data has similar enrichment patterns to previous transcriptome results, suggesting a consistent and reliable gene expression profiles. This reinforces the validity and reliability of our findings. However, it remains challenging to conclusively determine from our postnatal CPAM samples whether these genes play a causal role in initial cyst formation or if their altered expression is a consequence of disease progression. Longitudinal studies tracking gene expression from early fetal development through postnatal stages would provide more dynamic insights into the pathogenesis of CPAM. Despite these limitations, the identified genes and signaling pathways are suggestive of association with cell proliferation, apoptosis, and EMT, all of which are relevant to cyst formation and lung development. To firmly establish a causative link between these genes and cyst formation, further functional studies are needed to elucidate the roles of these molecules and signal pathways in CPAM pathophysiology.

Conclusion

In summary, our study systematically characterized the transcriptome expression pattern and identified gene models associated with type II CPAM. The PPI analysis highlighted several genes in GPCR signaling pathways and the regulation of calcium homeostasis. By integrating scRNA-seq and bulk RNA-seq data, we also identified calcium homeostasis-related gene *AGR3* and the immune-related gene *SLC11A1*. We found that the PI3K-AKT-mTOR signaling pathway may play a crucial role in epithelial cell communication and interactions. The genes and signaling pathways highlighted in this analysis may provide opportunities to explore new molecular mechanisms and develop therapeutic strategies for CPAM.

Abbreviations

AT1	Alveolar type 1
AT2	Alveolar type 2
CPAM	Congenital pulmonary airway malformation
DEG	Differentially expressed gene

FC	Fold change
GO	Gene ontology
GSEA	Gene set enrichment analysis
iWGCNA	Iterative weighted gene correlation network analysis
PCA	Principal component analysis
PPI	Protein-protein interaction
RNA-seq	RNA sequencing
scRNA-seq	Single-cell RNA-seq

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13578-024-01276-8>.

Supplementary Material 1: Supplementary Fig. 1 (A) PCA plot of RNA-seq data shows the characteristics of samples according to the gene expression levels. CPAM and control samples are colored in blue and orange, respectively. Each dot indicates a sample. (B) Volcano plot showed the expression profiling between the female and male in control samples. Each point in the plot indicated one gene, vertical lines refer to 2-fold change and the horizontal line corresponds to the adjusted P-value of 0.05.

Supplementary Material 2: Supplementary Fig. 2 Comparison of epithelial cell signature scores in each module between CPAM case and control groups. P-value was obtained from two-sample t-test.

Supplementary Material 3: Supplementary Fig. 3 (A) Feature plot of selected marker genes of AT1, AT2, club, and ciliated cells. (B-E) GSEA of the marker genes for AT1, AT2, club, and ciliated cells, respectively.

Supplementary Material 4: Supplementary Fig. 4 (A) Interaction net count plot and weight plot of epithelial cell subtypes. The thicker the line represented, the more the number of interactions, and the stronger the interaction weights/strength between the two cell types. (B,C) Cophenetic and Silhouette metrics were used to identify the number of outgoing and incoming communication patterns.

Supplementary Material 5: Supplementary Fig. 5 Ligand-receptor interactions between AT1, AT2, club, and ciliated cells inferred by CellChatDB.

Supplementary Material 6: Supplementary Fig. 6 qPCR validation of *AGR3*, *COL4A1*, *COL4A2*, *COL4A4*, *EDN1*, *FPR2* and *CHMR1* at mRNA level.

Supplementary Material 7: Supplemental Table 1 Clinical characteristics of subjects used for RNA-seq and qPCR validation.

Supplementary Material 8: Supplemental Table 2 Overlapped genes between DEGs related to CPAM and DEGs related to sex.

Supplementary Material 9: Supplemental Table 3 DEGs between CPAM cases and controls.

Supplementary Material 10: Supplemental Table 4 GO enrichment analysis (biological process, BP) of up-regulated and down-regulated genes.

Supplementary Material 11: Supplemental Table 5 GO enrichment analysis (molecular function, MF) and (cell component, CC) of up-regulated and down-regulated genes.

Supplementary Material 12: Supplemental Table 6 KEGG analysis of differentially expressed genes.

Supplementary Material 13: Supplemental Table 7 Gene list for the top 10 modules performed by iWGCNA.

Supplementary Material 14: Supplemental Table 8 Overlapped genes between epithelial cell marker genes and DEGs.

Supplementary Material 15: Supplemental Table 9 Overlapped genes between epithelial cell marker genes and the top 10 module genes.

Supplementary Material 16: Supplemental Table 10 Marker genes of epithelial subtypes.

Supplementary Material 17: Supplemental Table 11 Inference and analysis of CPAM-associated epithelial cells communication from single-cell data using CellChat.

Supplementary Material 18: Supplemental Table 12 Ligands and receptors enriched in PI3K-AKT signaling pathway and epithelial to mesenchymal transition.

Acknowledgements

We would like to thank the patients and their families for their generous contribution.

Author contributions

LFX and TZ conceptualized and designed the analyses, drafted the initial manuscript, and reviewed and revised the manuscript; CHY and GY designed the data collection instruments, collected data and conducted the initial analyses; XJ, HT, LL, ZXH, SXC, and ZJ coordinated and supervised data collection, and critically reviewed the manuscript for important intellectual content; CQ, SQ and YL contributed to study concept and design, critical review, and manuscript revision. All authors read and approved the final manuscript.

Funding

This study was supported by the Natural Science Foundation of Zhejiang province, China (Grant No. LY24H020002).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the ethics committee of the Children's Hospital, Zhejiang University School of Medicine (2022-IRB-001), and written informed consent was obtained from the patient's legal guardians.

Consent for publication

All authors read the manuscript and agreed for publication.

Competing interests

The authors declare no competing interests.

Author details

¹Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, Zhejiang, China

²Department of Thoracic Surgery, Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, Zhejiang, China

³Department of Cardiac Intensive Care Unit, Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, Zhejiang, China

⁴Department of Cardiac Surgery, Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, Zhejiang, China

⁵Department of General Surgery, Jiangxi Provincial Children's Hospital, Jiangxi, China

⁶Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

Received: 8 March 2024 / Accepted: 10 July 2024

Published online: 18 July 2024

References

- Ch'In KY, Tang MY. Congenital adenomatoid malformation of one lobe of a lung with general anasarca. *Arch Pathol (Chic)*. 1949;48(3):221–9.
- Leblanc C, Baron M, Desselas E, Phan MH, Rybak A, Thouvenin G, et al. Congenital pulmonary airway malformations: state-of-the-art review for pediatrician's use. *Eur J Pediatr*. 2017;176(12):1559–71. <https://doi.org/10.1007/s00431-017-3032-7>.
- Parikh DH, Rasiah SV. Congenital lung lesions: postnatal management and outcome. *Semin Pediatr Surg*. 2015;24(4):160–7. <https://doi.org/10.1053/j.sempedsurg.2015.01.013>.
- Stocker JT, Madewell JE, Drake RM. Congenital cystic adenomatoid malformation of the lung. Classification and morphologic spectrum. *Hum Pathol*. 1977;8(2):155–71. [https://doi.org/10.1016/s0046-8177\(77\)80078-6](https://doi.org/10.1016/s0046-8177(77)80078-6).
- Barazzone-Argiroffo C, Lascano Maillard J, Vidal I, Bochaton-Piallat ML, Blaskovic S, Donati Y, et al. New insights on congenital pulmonary airways malformations revealed by proteomic analyses. *Orphanet J Rare Dis*. 2019;14(1):272. <https://doi.org/10.1186/s13023-019-1192-4>.
- Karolak JA, Vincent M, Rupp R, Grazioli S, Wensley D, Jhangiani SN, et al. Phenotypic expansion of TBX4 mutations to include acinar dysplasia of the lungs. *Am J Med Genet A*. 2016;170(9):2440–4. <https://doi.org/10.1002/ajmg.a.37822>.
- Karolak JA, Vincent M, Deutsch G, Gambin T, Cogne B, Pichon O, et al. Complex compound inheritance of Lethal Lung Developmental disorders due to disruption of the TBX-FGF pathway. *Am J Hum Genet*. 2019;104(2):213–28. <https://doi.org/10.1016/j.ajhg.2018.12.010>.
- Barnett CP, Nataren NJ, Klingler-Hoffmann M, Schwarz Q, Chong CE, Lee YK, et al. Ectrodactyly and Lethal Pulmonary Acinar Dysplasia Associated with homozygous FGFR2 mutations identified by Exome sequencing. *Hum Mutat*. 2016;37(9):955–63. <https://doi.org/10.1002/humu.23032>.
- Hill DA, Ivanovich J, Priest JR, Gurnett CA, Dehner LP, Desruisseau D, et al. DICER1 mutations in familial pleuropulmonary blastoma. *Science*. 2009;325(5943):965. <https://doi.org/10.1126/science.1174334>.
- Messinger YH, Stewart DR, Priest JR, Williams GM, Harris AK, Schultz KA, et al. Pleuropulmonary blastoma: a report on 350 central pathology-confirmed pleuropulmonary blastoma cases by the International Pleuropulmonary Blastoma Registry. *Cancer*. 2015;121(2):276–85. <https://doi.org/10.1002/cncr.29032>.
- Dehner LP, Messinger YH, Williams GM, Stewart DR, Harney LA, Schultz KA, et al. Type I Pleuropulmonary Blastoma versus congenital pulmonary airway malformation type IV. *Neonatology*. 2017;111(1):76. <https://doi.org/10.1159/000447992>.
- Nelson ND, Xu F, Chandrasekaran P, Litzky LA, Peranteau WH, Frank DB, et al. Defining the spatial landscape of KRAS mutated congenital pulmonary airway malformations: a distinct entity with a spectrum of histopathologic features. *Mod Pathol*. 2022;35(12):1870–81. <https://doi.org/10.1038/s41379-022-01129-0>.
- Nelson ND, Xu F, Peranteau WH, Li M, Pogoriler J. Morphologic features in congenital pulmonary Airway malformations and Pulmonary sequestrations Correlate with Mutation Status: a mechanistic Approach to classification. *Am J Surg Pathol*. 2023;47(5):568–79. <https://doi.org/10.1097/PAS.0000000000002025>.
- Post M, Souza P, Liu J, Tseu I, Wang J, Kuliszewski M, et al. Keratinocyte growth factor and its receptor are involved in regulating early lung branching. *Development*. 1996;122(10):3107–15. <https://doi.org/10.1242/dev.122.10.3107>.
- Belluscì S, Grindley J, Emoto H, Itoh N, Hogan BL. Fibroblast growth factor 10 (FGF10) and branching morphogenesis in the embryonic mouse lung. *Development*. 1997;124(23):4867–78. <https://doi.org/10.1242/dev.124.23.4867>.
- Lezmi G, Verkarre V, Khen-Dunlop N, Vibhushan S, Hadchouel A, Rambaud C, et al. FGF10 Signaling differences between type I pleuropulmonary blastoma and congenital cystic adenomatoid malformation. *Orphanet J Rare Dis*. 2013;8:130. <https://doi.org/10.1186/1750-1172-8-130>.
- Gonzaga S, Henriques-Coelho T, Davey M, Zoltick PW, Leite-Moreira AF, Correia-Pinto J, et al. Cystic adenomatoid malformations are induced by localized FGF10 overexpression in fetal rat lung. *Am J Respir Cell Mol Biol*. 2008;39(3):346–55. <https://doi.org/10.1165/rcmb.2007-0290OC>.
- Harris KS, Zhang Z, McManus MT, Harfe BD, Sun X. Dicer function is essential for lung epithelium morphogenesis. *Proc Natl Acad Sci U S A*. 2006;103(7):2208–13. <https://doi.org/10.1073/pnas.0510839103>.
- Boucherat O, Landry-Truchon K, Berube-Simard FA, Houde N, Beuret L, Lezmi G, et al. Epithelial inactivation of Yy1 abrogates lung branching morphogenesis. *Development*. 2015;142(17):2981–95. <https://doi.org/10.1242/dev.120469>.
- Morrissey EE, Hogan BL. Preparing for the first breath: genetic and cellular mechanisms in lung development. *Dev Cell*. 2010;18(1):8–23. <https://doi.org/10.1016/j.devcel.2009.12.010>.
- Lezmi G, Vibhushan S, Bevilacqua C, Crapart N, Cagnard N, Khen-Dunlop N, et al. Congenital cystic adenomatoid malformations of the lung: an epithelial transcriptomic approach. *Respir Res*. 2020;21(1):43. <https://doi.org/10.1186/s12931-020-1306-5>.

22. Morotti RA, Gutierrez MC, Askin F, Profitt SA, Wert SE, Whitsett JA, et al. Expression of thyroid transcription factor-1 in congenital cystic adenomatoid malformation of the lung. *Pediatr Dev Pathol*. 2000;3(5):455–61. <https://doi.org/10.1007/s100240010092>.
23. Volpe MV, Pham L, Lessin M, Ralston SJ, Bhan I, Cutz E, et al. Expression of Hoxb-5 during human lung development and in congenital lung malformations. *Birth Defects Res Clin Mol Teratol*. 2003;67(8):550–6. <https://doi.org/10.1002/bdra.10086>.
24. Ochieng JK, Schilders K, Kool H, Boerema-De Munck A, Buscop-Van Kempen M, Gontan C, et al. Sox2 regulates the emergence of lung basal cells by directly activating the transcription of Trp63. *Am J Respir Cell Mol Biol*. 2014;51(2):311–22. <https://doi.org/10.1165/rcmb.2013-0419OC>.
25. Gontan C, de Munck A, Vermeij M, Grosveld F, Tibboel D, Rottier R. Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation. *Dev Biol*. 2008;317(1):296–309. <https://doi.org/10.1016/j.ydbio.2008.02.035>.
26. Peranteau WH, Boelig MM, Khalek N, Moldenhauer JS, Martinez-Poyer J, Hedrick HL, et al. Effect of single and multiple courses of maternal betamethasone on prenatal congenital lung lesion growth and fetal survival. *J Pediatr Surg*. 2016;51(1):28–32. <https://doi.org/10.1016/j.jpedsurg.2015.10.018>.
27. Peranteau WH, Wilson RD, Liechty KW, Johnson MP, Bebbington MW, Hedrick HL, et al. Effect of maternal betamethasone administration on prenatal congenital cystic adenomatoid malformation growth and fetal survival. *Fetal Diagn Ther*. 2007;22(5):365–71. <https://doi.org/10.1159/000103298>.
28. Morris LM, Lim FY, Livingston JC, Polzin WJ, Crombleholme TM. High-risk fetal congenital pulmonary airway malformations have a variable response to steroids. *J Pediatr Surg*. 2009;44(1):60–5. <https://doi.org/10.1016/j.jpedsurg.2008.10.012>.
29. Tsao K, Hawgood S, Vu L, Hirose S, Sydorak R, Albanese CT, et al. Resolution of hydrops fetalis in congenital cystic adenomatoid malformation after prenatal steroid therapy. *J Pediatr Surg*. 2003;38(3):508–10. <https://doi.org/10.1053/jpsu.2003.50089>.
30. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature*. 2020;587(7835):619–25. <https://doi.org/10.1038/s41586-020-2922-4>.
31. Cohen M, Giladi A, Gorki AD, Solodkin DG, Zada M, Hladik A, et al. Lung single-cell Signaling Interaction Map reveals Basophil Role in Macrophage Imprinting. *Cell*. 2018;175(4):1031–e4418. <https://doi.org/10.1016/j.cell.2018.09.009>.
32. Wong KKY, Flake AW, Tibboel D, Rottier RJ, Tam PKH. Congenital pulmonary airway malformation: advances and controversies. *Lancet Child Adolesc Health*. 2018;2(4):290–7. [https://doi.org/10.1016/S2352-4642\(18\)30035-X](https://doi.org/10.1016/S2352-4642(18)30035-X).
33. Cass DL, Quinn TM, Yang EY, Liechty KW, Crombleholme TM, Flake AW, et al. Increased cell proliferation and decreased apoptosis characterize congenital cystic adenomatoid malformation of the lung. *J Pediatr Surg*. 1998;33(7):1043–6. [https://doi.org/10.1016/s0022-3468\(98\)90528-0](https://doi.org/10.1016/s0022-3468(98)90528-0). discussion 7.
34. Zhang S, Ye C, Xiao J, Yang J, Zhu C, Xiao Y, et al. Single-cell transcriptome profiling reveals the mechanism of abnormal proliferation of epithelial cells in congenital cystic adenomatoid malformation. *Exp Cell Res*. 2020;396(2):112299. <https://doi.org/10.1016/j.yexcr.2020.112299>.
35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
36. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–60. <https://doi.org/10.1038/nmeth.3317>.
37. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–5. <https://doi.org/10.1038/nbt.3122>.
38. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform*. 2020;2(3):lqaa078. <https://doi.org/10.1093/nargab/lqaa078>.
39. Love MI, Huber W, Anders S. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
40. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7. <https://doi.org/10.1089/omi.2011.0118>.
41. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–13. <https://doi.org/10.1093/nar/gky1131>.
42. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504. <https://doi.org/10.1101/gr.1239303>.
43. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. <https://doi.org/10.1186/1471-2105-9-559>.
44. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. 2021;12(1):1088. <https://doi.org/10.1038/s41467-021-21246-9>.
45. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90–7. <https://doi.org/10.1093/nar/gkw377>.
46. Starruss J, de Back W, Bruschi L, Deutsch A. Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology. *Bioinformatics*. 2014;30(9):1331–2. <https://doi.org/10.1093/bioinformatics/btt772>.
47. Oliva M, Munoz-Aguirre M, Kim-Hellmuth S, Wucher V, Gewirtz ADH, Cotter DJ, et al. The impact of sex on gene expression across human tissues. *Science*. 2020;369(6509). <https://doi.org/10.1126/science.aba3066>.
48. Lopes-Ramos CM, Chen CY, Kuijjer ML, Paulson JN, Sonawane AR, Fagny M, et al. Sex differences in Gene Expression and Regulatory Networks across 29 human tissues. *Cell Rep*. 2020;31(12):107795. <https://doi.org/10.1016/j.celrep.2020.107795>.
49. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. 2014;15(3):178–96. <https://doi.org/10.1038/nrm3758>.
50. Mottais A, Riberi L, Falco A, Soccia S, De Gohy S. Epithelial-mesenchymal transition mechanisms in Chronic Airway diseases: a common process to target? *Int J Mol Sci*. 2023;24(15). <https://doi.org/10.3390/ijms241512412>.
51. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. *Nat Rev Mol Cell Biol*. 2017;18(9):533–47. <https://doi.org/10.1038/nrm.2017.60>.
52. Zhang G, Cai C, Li X, Lou L, Zhou B, Zeng H, et al. Application of second-generation sequencing in congenital pulmonary airway malformations. *Sci Rep*. 2022;12(1):20459. <https://doi.org/10.1038/s41598-022-24858-3>.
53. Zhang G, Lou L, Shen L, Zeng H, Cai C, Wu R, et al. The underlying molecular mechanism of ciliated epithelium dysfunction and TGF-beta signaling in children with congenital pulmonary airway malformations. *Sci Rep*. 2024;14(1):4430. <https://doi.org/10.1038/s41598-024-54924-x>.
54. Varricchi G, Brightling CE, Grainge C, Lambrecht BN, Chanez P. Airway remodeling in asthma and the epithelium: on the edge of a new era. *Eur Respir J*. 2024;63(4). <https://doi.org/10.1183/13993003.01619-2023>.
55. Lujan M, Bosque M, Mirapeix RM, Marco MT, Asensio O, Domingo C. Late-onset congenital cystic adenomatoid malformation of the lung. Embryology, clinical symptomatology, diagnostic procedures, therapeutic approach and clinical follow-up. *Respiration*. 2002;69(2):148–54. <https://doi.org/10.1159/000056318>.
56. Lauss M, Krieger A, Vierlinger K, Noehammer C. Characterization of the drugged human genome. *Pharmacogenomics*. 2007;8(8):1063–73. <https://doi.org/10.2217/14622416.8.8.1063>.
57. Turner MD, Nedjai B, Hurst T, Pennington DJ. Cytokines and chemokines: at the crossroads of cell signalling and inflammatory disease. *Biochim Biophys Acta*. 2014;1843(11):2563–82. <https://doi.org/10.1016/j.bbamcr.2014.05.014>.
58. Song K, Awata J, Tritschler D, Bower R, Witman GB, Porter ME, et al. In situ localization of N and C termini of subunits of the flagellar nexin-dynein regulatory complex (N-DRC) using SNAP tag and cryo-electron tomography. *J Biol Chem*. 2015;290(9):5341–53. <https://doi.org/10.1074/jbc.M114.626556>.
59. Liang Y, Lu Y, Chen Q, Cheng Y, Ma Y, Huang Y, et al. Identification of long noncoding RNAs that exert transcriptional regulation by forming RNA-DNA triplexes in prostate Cancer. *Int J Mol Sci*. 2023;24(3). <https://doi.org/10.3390/ijms24032035>.
60. Rachev E, Schuster-Gossler K, Fuhl F, Ott T, Tverikhina L, Beckers A, et al. CFAP43 modulates ciliary beating in mouse and Xenopus. *Dev Biol*. 2020;459(2):109–25. <https://doi.org/10.1016/j.ydbio.2019.12.010>.
61. Dougherty GW, Loges NT, Klinkenbusch JA, Olbrich H, Pennekamp P, Menchen T, et al. DNAH11 localization in the Proximal Region of Respiratory Cilia defines distinct outer dynein arm complexes. *Am J Respir Cell Mol Biol*. 2016;55(2):213–24. <https://doi.org/10.1165/rcmb.2015-0353OC>.

62. Zhang Y, Chen Y, Zheng J, Wang J, Duan S, Zhang W, et al. Vertebrate Dynein-f depends on Wdr78 for axonemal localization and is essential for ciliary beat. *J Mol Cell Biol.* 2019;11(5):383–94. <https://doi.org/10.1093/jmcb/mjy043>.
63. Kubo S, Black CS, Joachimiak E, Yang SK, Legal T, Peri K, et al. Native doublet microtubules from *Tetrahymena Thermophila* reveal the importance of outer junction proteins. *Nat Commun.* 2023;14(1):2168. <https://doi.org/10.1038/s41467-023-37868-0>.
64. Tan Z, Li F, Chen Q, Chen H, Xue Z, Zhang J, et al. Integrated bulk and single-cell RNA-sequencing reveals SPOCK2 as a novel biomarker gene in the development of congenital pulmonary airway malformation. *Respir Res.* 2023;24(1):127. <https://doi.org/10.1186/s12931-023-02436-z>.
65. Jiang Y, Luo Y, Tang Y, Moats R, Warburton D, Zhou S, et al. Alteration of cystic airway mesenchyme in congenital pulmonary airway malformation. *Sci Rep.* 2019;9(1):5296. <https://doi.org/10.1038/s41598-019-41777-y>.
66. Bonser LR, Schroeder BW, Ostrin LA, Baumlín N, Olson JL, Salathe M, et al. The endoplasmic Reticulum Resident protein AGR3. Required for regulation of ciliary beat frequency in the Airway. *Am J Respir Cell Mol Biol.* 2015;53(4):536–43. <https://doi.org/10.1165/rcmb.2014-0318OC>.
67. Cunrath O, Bumann D. Host resistance factor SLC11A1 restricts Salmonella growth through magnesium deprivation. *Science.* 2019;366(6468):995–9. <https://doi.org/10.1126/science.aax7898>.
68. Barnes JL, Yoshida M, He P, Worlock KB, Lindeboom RGH, Suo C, et al. Early human lung immune cell development and its role in epithelial cell fate. *Sci Immunol.* 2023;8(90):eadf9988. <https://doi.org/10.1126/sciimmunol.adf9988>.
69. Swarr DT, Peranteau WH, Pogorilic J, Frank DB, Adzick NS, Hedrick HL, et al. Novel molecular and phenotypic insights into congenital lung malformations. *Am J Respir Crit Care Med.* 2018;197(10):1328–39. <https://doi.org/10.1164/rccm.201706-1243OC>.
70. Peranteau WH, Merchant AM, Hedrick HL, Liechty KW, Howell LJ, Flake AW, et al. Prenatal course and postnatal management of peripheral bronchial atresia: association with congenital cystic adenomatoid malformation of the lung. *Fetal Diagn Ther.* 2008;24(3):190–6. <https://doi.org/10.1159/000151337>.
71. Tastekin E, Usta U, Kaynar A, Ozdemir C, Yalcin O, Ozyilmaz F, et al. Congenital pulmonary airway malformation type 2: a Case Report with Review of the literature. *Turk Patoloji Derg.* 2014. <https://doi.org/10.5146/tjpath.2013.01208>.
72. Danopoulos S, Alonso I, Thornton ME, Grubbs BH, Bellusci S, Warburton D, et al. Human lung branching morphogenesis is orchestrated by the spatiotemporal distribution of ACTA2, SOX2, and SOX9. *Am J Physiol Lung Cell Mol Physiol.* 2018;314(1):L144–9. <https://doi.org/10.1152/ajplung.00379.2017>.
73. Nikolic MZ, Caritg O, Jeng Q, Johnson JA, Sun D, Howell KJ, et al. Human embryonic lung epithelial tips are multipotent progenitors that can be expanded in vitro as long-term self-renewing organoids. *Elife.* 2017;6. <https://doi.org/10.7554/eLife.26575>.
74. Hermelijn SM, Mackenbach MJ, van Horik C, Ciet P, Wolf JL, von der Thusen JH, et al. Quantitative CT imaging analysis to predict pathology features in patients with a congenital pulmonary airway malformation. *J Pediatr Surg.* 2022;57(8):1567–72. <https://doi.org/10.1016/j.jpedsurg.2021.10.008>.
75. Rosano L, Spinella F, Bagnato A. Endothelin 1 in cancer: biological implications and therapeutic opportunities. *Nat Rev Cancer.* 2013;13(9):637–51. <https://doi.org/10.1038/nrc3546>.
76. Tylek K, Trojan E, Regulska M, Lácivita E, Leopoldo M, Basta-Kaim A. Formyl peptide receptor 2, as an important target for ligands triggering the inflammatory response regulation: a link to brain pathology. *Pharmacol Rep.* 2021;73(4):1004–19. <https://doi.org/10.1007/s43440-021-00271-x>.
77. Machado MG, Tavares LP, Souza GVS, Queiroz-Junior CM, Ascencio FR, Lopes ME, et al. The annexin A1/FPR2 pathway controls the inflammatory response and bacterial dissemination in experimental pneumococcal pneumonia. *FASEB J.* 2020;34(2):2749–64. <https://doi.org/10.1096/fj.201902172R>.
78. Maekawa A, Xing W, Austen KF, Kanaoka Y. GPR17 regulates immune pulmonary inflammation induced by house dust mites. *J Immunol.* 2010;185(3):1846–54. <https://doi.org/10.4049/jimmunol.1001131>.
79. Maeda Y, Hizawa N, Jinushi E, Honda A, Takahashi D, Fukui Y, et al. Polymorphisms in the muscarinic receptor 1 gene confer susceptibility to asthma in Japanese subjects. *Am J Respir Crit Care Med.* 2006;174(10):1119–24. <https://doi.org/10.1164/rccm.200601-081OC>.
80. Haq IJ, Gray MA, Garnett JP, Ward C, Brodrie M. Airway surface liquid homeostasis in cystic fibrosis: pathophysiology and therapeutic targets. *Thorax.* 2016;71(3):284–7. <https://doi.org/10.1136/thoraxjnl-2015-207588>.
81. Chi J, Zhang H, Hu J, Song Y, Li J, Wang L, et al. AGR3 promotes the stemness of colorectal cancer via modulating Wnt/beta-catenin signalling. *Cell Signal.* 2020;65:109419. <https://doi.org/10.1016/j.cellsig.2019.109419>.
82. Pongracz JE, Stockley RA. Wnt signalling in lung development and diseases. *Respir Res.* 2006;7(1):15. <https://doi.org/10.1186/1465-9921-7-15>.
83. Ye R, Wang C, Sun P, Bai S, Zhao L. AGR3 regulates Airway Epithelial junctions in patients with frequent exacerbations of COPD. *Front Pharmacol.* 2021;12:669403. <https://doi.org/10.3389/fphar.2021.669403>.
84. Paccagnini D, Sieswerda L, Rosu V, Masala S, Pacifico A, Gazouli M, et al. Linking chronic infection and autoimmune diseases: Mycobacterium avium subspecies paratuberculosis, SLC11A1 polymorphisms and type-1 diabetes mellitus. *PLoS ONE.* 2009;4(9):e71109. <https://doi.org/10.1371/journal.pone.0007109>.
85. Asante-Poku A, Morgan P, Osei-Wusu S, Aboagye SY, Asare P, Otchere ID, et al. Genetic Analysis of TB Susceptibility Variants in Ghana Reveals Candidate Protective Loci in SORBS2 and SCL11A1 genes. *Front Genet.* 2021;12:729737. <https://doi.org/10.3389/fgene.2021.729737>.
86. Blackwell JM, Searle S, Mohamed H, White JK. Divalent cation transport and susceptibility to infectious and autoimmune disease: continuation of the *Ity/Lsh/Bcg/Nramp1/Slc11a1* gene story. *Immunol Lett.* 2003;85(2):197–203. [https://doi.org/10.1016/s0165-2478\(02\)00231-6](https://doi.org/10.1016/s0165-2478(02)00231-6).
87. Morotti RA, Cangiarella J, Gutierrez MC, Jagirdar J, Askin F, Singh G, et al. Congenital cystic adenomatoid malformation of the lung (CCAM): evaluation of the cellular components. *Hum Pathol.* 1999;30(6):618–25. [https://doi.org/10.1016/s0046-8177\(99\)90084-9](https://doi.org/10.1016/s0046-8177(99)90084-9).
88. Godoy-Guzman C, San Martín S, Pereda J. Proteoglycan and collagen expression during human air conducting system development. *Eur J Histochem.* 2012;56(3):e29. <https://doi.org/10.4081/ejh.2012.e29>.
89. McGowan SE. Extracellular matrix and the regulation of lung development and repair. *FASEB J.* 1992;6(11):2895–904.
90. Joza S, Wang J, Tseu I, Ackerley C, Post M. Fetal, but not postnatal, deletion of semaphorin-neuropilin-1 signaling affects murine alveolar development. *Am J Respir Cell Mol Biol.* 2013;49(4):627–36. <https://doi.org/10.1165/rcmb.2012-0407OC>.
91. Joza S, Wang J, Fox E, Hillman V, Ackerley C, Post M. Loss of semaphorin-neuropilin-1 signaling causes dysmorphic vascularization reminiscent of alveolar capillary dysplasia. *Am J Pathol.* 2012;181(6):2003–17. <https://doi.org/10.1016/j.ajpath.2012.08.037>.
92. Alzahrani AS. PI3K/Akt/mTOR inhibitors in cancer: at the bench and bedside. *Semin Cancer Biol.* 2019;59:125–32. <https://doi.org/10.1016/j.semcancer.2019.07.009>.
93. Glaviano A, Foo ASC, Lam HY, Yap KCH, Jacot W, Jones RH, et al. PI3K/AKT/mTOR signaling transduction pathway and targeted therapies in cancer. *Mol Cancer.* 2023;22(1):138. <https://doi.org/10.1186/s12943-023-01827-6>.
94. Khattar D, Fernandes S, Snowball J, Guo M, Gillen MC, Jain SS, et al. PI3K signaling specifies proximal-distal fate by driving a developmental gene regulatory network in SOX9+ mouse lung progenitors. *Elife.* 2022;11. <https://doi.org/10.7554/eLife.67954>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.