Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Clinical, Genetic, and Single-Cell Landscape of Birt-Hogg-Dubé Syndrome: A Study of 85 Chinese Families

Xianyi Xie
The First Affiliated Hospital of Guangzhou Medical University
Yanchi Li
The First Affiliated Hospital of Guangzhou Medical University
Lincai Yang
The First Affiliated Hospital of Guangzhou Medical University
Yuying Hu
The First Affiliated Hospital of Guangzhou Medical University
Panqiao Liang
The First Affiliated Hospital of Guangzhou Medical University
Dongrui Li
The First Affiliated Hospital of Guangzhou Medical University
Yujiao Ma
The First Affiliated Hospital of Guangzhou Medical University
Leduo Liang
The First Affiliated Hospital of Guangzhou Medical University
Chen Chen
The First Affiliated Hospital of Guangzhou Medical University
Yujuan Wu
The First Affiliated Hospital of Guangzhou Medical University
Jie Liu
ljgird001@163.com

The First Affiliated Hospital of Guangzhou Medical University

---

Research Article

Keywords: Birt-Hogg-Dubé syndrome, Chinese single-center cohort, Clinical phenotype, FLCN gene, Genotype, Single-cell RNA sequencing

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11145003/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: Birt-Hogg-Dubé syndrome is a rare genetic disorder characterized by striking phenotypic heterogeneity. Currently, large-scale BHD cohorts from [removed/redacted] China remain underrepresented, and the molecular mechanisms underlying pulmonary cyst formation awaits further exploration.

Methods: We retrospectively analyzed the clinical and genetic data of 85 BHD families admitted between 2017 and 2025. Additionally, single-cell RNA sequencing (scRNA-seq) was performed on lung tissues from four BHD patients and integrated with healthy control datasets to uncover the cellular mechanisms underlying cystogenesis.

Results: Compared to Caucasian populations, our Chinese cohort predominantly presented with pulmonary cysts and spontaneous pneumothorax, with lower incidences of cutaneous and renal lesions. Genetic testing revealed prominent FLCN mutational hotspots (c.1579_1580insA, c.1285dupC, and c.1429C>T) and identified 26 novel variants. Single-cell transcriptomic analysis demonstrated significant microenvironmental remodeling in BHD lung tissues, characterized by a reduction in normal structural epithelial cells and an expansion of activated fibroblast and immune cell subsets. We identified a previously unreported "Epi-EMT-like" epithelial subpopulation which was highly enriched in BHD lungs. This cell cluster concurrently expresses canonical epithelial markers and mesenchymal/profibrotic-associated markers, and is likely derived from the aberrant differentiation of alveolar type 2 cells. These cells exhibited loss of canonical cellular identity, downregulation of adhesion molecules, and involvement in extensive crosstalk with recruited immune and stromal cells, which may facilitate a hyperactive tissue remodeling microenvironment.

Conclusions: BHD syndrome exhibits pronounced phenotypic and genotypic heterogeneity across ethnicities. The discovery of the pathological "Epi-EMT-like" cells and their associated local microenvironmental remodeling provides a new theoretical basis for understanding the mechanisms of BHD pulmonary cyst formation and may offer novel directions for the future development of targeted therapeutic strategies.

Introduction

Birt-Hogg-Dubé syndrome (BHD) is a rare autosomal dominant inherited disorder first reported over 50 years ago[1]. The classic clinical manifestations include multiple pulmonary cysts with or without spontaneous pneumothorax, cutaneous fibrofolliculomas/trichodiscomas, and various types of renal tumors. These can manifest as a single phenotype or in any combination of mixed phenotypes. Pulmonary cystic lesions typically appear earliest and may present as chest tightness and shortness of breath; however, early-stage patients may remain without obvious symptoms. Diagnostically, it primarily requires differentiation from other diffuse cystic lung diseases, including chronic obstructive pulmonary disease, lymphangioleiomyomatosis, pulmonary Langerhans cell histiocytosis, and lymphocytic

interstitial pneumonia. Its population prevalence remains ambiguous. In 2021, Muller et al.[2] calculated the prevalence of BHD syndrome in the general population to be approximately 1.86 per 1,000,000 based on epidemiological data of spontaneous pneumothorax and Bayes' theorem. Conversely, Savatt et al.[3] utilized whole-exome sequencing and found that approximately 0.03% of 136,000 individuals harbored pathogenic/likely pathogenic (P/LP) FLCN gene variants, while a health care system–based study demonstrate that the frequency of clinically validated loss-of-function FLCN variants is 0.0239% in the general population[4], indicating an extremely high rate of missed diagnoses and misdiagnoses for the BHD population, and a severe underestimation of the actual prevalence. For the Chinese population, the mean interval from clinical presentation to diagnosis is 9.6 years, with an age range of 18–84 years (mean 47.0 ± 13.9 years), and a male-to-female ratio of 1:1.6[5].

The pathogenic gene for BHD is located on chromosome 17p11.2 and is caused by germline mutations in the folliculin (FLCN) gene[6]. This gene is widely considered a tumor suppressor gene[7], comprising 14 exons that collectively encode a long protein consisting of 579 amino acids—folliculin[8]. Most FLCN variants are truncating variants, such as frameshift, splice site, and nonsense variants. Currently, according to the Human Gene Mutation Database (HGMD), over 200 variants have been reported in BHD patients.

At present, research on BHD has predominantly focused on Caucasian populations. Nearly 50 years of studies have shown that over 80% of Caucasian BHD patients develop cutaneous fibrofolliculomas[9]. In contrast, Chinese BHD patients primarily manifest pulmonary abnormalities, with less cutaneous and renal involvement[10]; pulmonary cystic lesions or pneumothorax can emerge as isolated phenotypes of BHD[11,12], while combined pulmonary and renal lesions hold greater diagnostic significance[13]. Data indicates that the highly mutated region of the BHD gene in Caucasians is the C8 tract in Exon 11 of the FLCN gene, with the mutational hotspot being c.1285dup/delC[14,15]. Recently, a Japanese study involving 334 BHD patients identified three predominant loci: c.1285dupC in Exon 11, c.1533_1536delGATG in Exon 13, and c.1347_1353dupCCACCCT in Exon 12[16]. Although several Chinese cohort studies have reported c.1285dup/delC as the predominant mutation hotspot[5,10,12,17–19], these cohorts have primarily included patients from northern and southeastern China. Given the large population and genetic diversity of China, additional studies involving broader geographic populations are needed to further characterize the clinical and genetic landscape of BHD in Chinese individuals.

However, the precise cellular and molecular mechanisms by which germline FLCN mutations drive pulmonary cyst formation remain largely unknown. In recent years, single-cell RNA sequencing (scRNA-seq) has emerged as a powerful approach for dissecting tissue heterogeneity and cell–cell communication at unprecedented resolution. Although scRNA-seq has been successfully applied to investigate BHD-associated renal lesions and revealed extensive microenvironmental remodeling driven by FLCN-deficient cells[20–22], the single-cell transcriptomic landscape of BHD lung remains unexplored.

In this study, we investigated the clinical characteristics and genetic landscape of 107 patients from 85 Chinese BHD families, primarily originating from southern China, to further expand the understanding of phenotypic and genotypic diversity in Asian populations. And to explore genotype-phenotype correlations, we compared our data with a Caucasian cohort. More importantly, we performed scRNA-seq analysis of BHD lung for the first time to characterize pathological alterations within the pulmonary microenvironment and provide new insights into the mechanisms underlying pulmonary cyst formation, potentially facilitating the identification of therapeutic targets for this rare genetic disorder.

Subjects and Methods

Study population and samples

We conducted a retrospective, single-center, cross-sectional study involving patients with BHD syndrome from 85 families diagnosed at the Department of Respiratory Medicine, First Affiliated Hospital of Guangzhou Medical University, between 2017 and 2025. This included 17 males and 90 females. All 85 probands were evaluated by a multidisciplinary expert team at our center, comprising respiratory physicians, dermatologists, nephrologists, radiologists, and pathologists. When BHD was highly suspected based on clinical manifestations and imaging, targeted genetic testing was further performed. We obtained informed consent from all patients. Inclusion criteria: The diagnosis of BHD was based on the criteria proposed by the European BHD Consortium in 2009[23] or the diagnostic criteria for BHD in the 2023 "Chinese Expert Consensus on the Diagnosis and Management of Birt-Hogg-Dubé Syndrome"[24] (Table 1). Exclusion criteria: Age < 18 years; incomplete clinical data. This single-center retrospective study was approved by the Institutional Review Board, and written informed consent was obtained from all participants. Six normal lung samples were obtained from the Gene Expression Omnibus (GEO) dataset GSE220797 (GSM6820277, GSM6820279, GSM6820281, GSM6820283, GSM6820285, and GSM6820287).

Clinical data collection and assessment

Study subjects were selected based on established inclusion and exclusion criteria. Comprehensive clinical data were systematically extracted from registration forms, electronic medical records, and outpatient follow-ups. The collected variables encompassed basic demographics (sex, age, age at diagnosis), a family history of relevant lesions (defined as ≥2 affected members with pulmonary cysts/pneumothorax, or renal tumors), and detailed clinical manifestations. Specifically, we recorded the onset age, clinical characteristics, and management modalities for pulmonary cysts (confirmed via chest High-Resolution Computed Tomography [HRCT]), spontaneous pneumothorax, renal tumors (including histopathology), and cutaneous lesions. Additionally, baseline radiological evaluations, encompassing chest, cranial, and comprehensive abdominal/renal imaging (Computed Tomography [CT], Magnetic Resonance Imaging [MRI], or ultrasound), were documented.

Genetic testing

Genetic testing samples were derived from the peripheral blood leukocytes of 107 patients. Following density gradient centrifugation of the collected peripheral blood to isolate leukocytes, genomic DNA was extracted for analysis. Detection was performed on the Illumina sequencing platform. The acquired data achieved an average sequencing depth of ≥ 90X across known human gene exons and their 5 bp upstream and downstream flanking sequences, with approximately 98% of the target sequences reaching a depth of over 20X, and base calling was performed on all sequenced fragments. The GATK software suite was utilized for sequencing data analysis. The VEP (Variant Effect Predictor) software was used to annotate variants, while simultaneously screening them based on genetic disease databases, variant databases, and large-scale population sequencing databases such as ClinVar, OMIM, HGMD, and gnomAD. Concurrently, various recognized computational algorithms were employed to predict and classify the potential pathogenicity of the variants. FLCN mutations were numbered based on GenBank accession numbers NM_144997.7 according to the HGVS nomenclature guideline (http://www.hgvs.org/mutnomen). Information regarding the folliculin protein was obtained from the Genomic Data Commons (GDC) database, and the corresponding lollipop plot (Figure 2) was generated using the GDC Data Portal. A variant was considered novel if no corresponding variant was identified in ClinVar, LOVD, or the published literature after searching the databases and literature up to 21 August 2026.

Single-cell dissociation and sequencing

Fresh lung tissues were minced and enzymatically dissociated using a Tumor Dissociation Kit (Miltenyi Biotec) at 37°C for 40 min. Cell suspensions were filtered through a 70 μm strainer, and cell viability was assessed using a Countstar Fluorescence Cell Analyzer. Single-cell libraries were prepared using the 10x Genomics Chromium Controller with the Chromium Single Cell 5' v2 platform (10x Genomics) and sequenced on an Illumina platform using 150 bp paired-end reads.

Data preprocessing and quality control

Raw reads were processed using fastp. Gene expression matrices were generated by alignment to the human reference genome (GRCh38; Ensembl annotation v104) using CellRanger v5.0.1. Cells with >200 detected genes and mitochondrial UMI content <20% were retained. Mitochondrial genes were excluded from downstream analyses. Doublets were removed using DoubletFinder v2.0.4.

Integration, clustering, and cell annotation

Analyses were performed in R (v3.6.2) using Seurat (v4.0.3). Data were normalized using the LogNormalize method (scale factor 10,000, followed by log-transformation). Batch effects across samples were corrected using the mutual nearest neighbors (MNN) method. Principal component analysis was performed using the top 2,000 highly variable genes, and the first 10 principal components were used for graph-based clustering. Cell types were annotated at three hierarchical levels based on canonical marker genes identified using FindAllMarkers (Wilcoxon rank-sum test), including major

lineages, fine-grained subpopulations (identified by re-clustering within each lineage), and intermediate cell types used for cell-cell communication analysis.

Differential expression and functional enrichment

Differentially expressed genes (DEGs) between disease and control groups were identified using FindMarkers in Seurat (Wilcoxon rank-sum test; log2FC > 0.25, P < 0.05, minimum percentage > 0.1). Gene set enrichment analysis was performed using QuSAGE (v2.16.1) with KEGG and Hallmark gene sets. Pathway activity at single-cell resolution was scored using AUCell (v1.18.1). Gene co-expression modules were identified using Monocle3 (v0.2.3.0) followed by Louvain clustering. Transcription factor regulatory networks were inferred using pySCENIC (v0.11.2) with the RcisTarget database and GRNBoost.

Cell-cell communication analysis

Cell-cell communication was inferred at the intermediate cell-type level using CellChat (v2.1.2). The CellChatDB human was used for analysis. Significant ligand-receptor interactions were defined as P < 0.05. Interaction strength was compared between each disease and control to identify shared and disease-specific signaling pathways and ligand-receptor pairs.

Single-cell trajectory and pseudo-time analysis

We applied the Single-Cell Trajectories analysis utilizing Monocle2 (http://cole-trapnell-lab.github.io/monocle-release) using DDR-Tree and default parameter. Before Monocle analysis, we select marker genes of the Seurat clustering result and raw expression counts of the cell passed filtering. Based on the pseudo-time analysis, branch expression analysis modeling (BEAM Analysis) was applied for branch fate determined gene analysis.

Statistical analysis

Statistical analysesof the data were performed using SPSS 26.0 software (IBM SPSS, USA). Continuous variables are presented as the mean and standard deviation, and were compared using analysis of variance (ANOVA). Categorical variables are expressed as frequencies and percentages, and were compared using the Chi-square test and Fisher’s exact test. A two-sided P < 0.05 indicated that the difference was statistically significant. Due to the frequent incompleteness of clinical data for non-probands within the 85 families, we only performed statistical analysis on the clinical characteristics of the probands to avoid potential bias.

Results

Demographic data

Between July 2017 and July 2025, we admitted a total of 107 BHD patients from 85 families, with the age at diagnosis ranging from 18 to 72 years, averaging 44.9 ± 11.6 years. There were 17 male patients and 90 female patients, yielding a male-to-female ratio of 1:5.3. There were 85 probands (79.4%) and 22 screened family members (20.6%), making the ratio of screened relatives to probands 1:3.9. Among the 85 probands, 38 (44.7%) had a family history of pulmonary cysts or pneumothorax.

Clinical characteristics

Pulmonary Manifestations

Among the 85 probands, the most common clinical manifestation was exertional dyspnea in 61/84 (72.6%), followed by cough in 48/83 (57.8%), chest tightness in 43/81 (53.1%), and chest pain in 31/83 (37.3%). A very small minority experienced hemoptysis in 1/83 (1.2%) or back pain in 5/82 (6.1%).

Spontaneous pneumothorax had occurred in 64 (64/83, 77.1%) of the patients, with first-time occurrences and recurrences accounting for 36.5% and 63.5%, respectively. Twenty-four out of 36 patients (24/36, 66.7%) with pneumothorax received only conservative treatment (observation, supplemental oxygen, closed thoracic drainage), while 12 patients (12/36, 33.3%) underwent surgical treatment (pleurodesis or bullectomy).

Pulmonary cystic lesions were observed on chest CT in all the 85 patients (100.0%), and all exhibiting bilateral distribution with lower lobe predominance. Lung tissue biopsy was performed in 45 patients, and 34 (34/45, 75.6%) patients were pathologically diagnosed with pulmonary bullae or bullous emphysema.

Cutaneous Manifestations

Patients with cutaneous lesions accounted for 5/79 (6.3%, 6 cases missing), all distributed on the face and neck, presenting as multiple, millet-sized, light-colored papules. The mean ages of patients with and without cutaneous lesions were 48.4 ± 6.7 years and 46.4 ± 10.9 years, respectively. Skin biopsies were performed in 2 of the patients with skin lesions, with 1 case confirmed as fibrofolliculoma.

Renal Manifestations

Seventy-eight patients underwent renal imaging evaluations (encompassing ultrasound, CT, and MRI). Renal tumors were found in 8 patients (8/78, 10.3%), including 2 cases of chromophobe renal cell carcinoma, 1 case of multilocular cystic renal neoplasm, 1 case of pheochromocytoma, and 1 case of a renal space-occupying lesion detected by MRI pending pathological confirmation. Additionally, CT identified renal angiomyolipoma (AML) in 3 cases, a renal tumor uncommonly observed in BHD patients, characteristically associated with Tuberous Sclerosis Complex (TSC). Other important imaging findings included 1 case of horseshoe kidney and 25 cases (25/78, 32.1%) of renal cysts. Detailed information on the demographic and clinical features of the 107 Chinese BHD patients is summarized in Table 2. Related manifestations are showed in Figure 1.

Germline mutations and their frequencies

Among the 107 patients from the 85 BHD families, genetic testing excluded variants of uncertain significance, identifying likely or definitively pathogenic FLCN gene mutations in 106/107 (99.1%) patients. Of these, 49/106 (46.2%) had definitively pathogenic mutations, and 57/106 (53.8%) had suspected pathogenic variants.

Across the 85 probands, 49 distinct variants were discovered: 58 cases (54.2%) had frameshift variants, 21 cases (19.6%) had nonsense mutations, 9 cases (8.4%) had splice-site variants, 9 cases (8.4%) had missense variants, 6 cases (5.6%) had in-frame deletion variants, and 4 cases (3.7%) had large fragment deletions.

The spatial distribution, frequency, and variant types of the FLCN mutation loci are depicted in Figure 2. We found that FLCN gene mutation frequently occurred at c.1579_1580insA (p.Arg527GLnfs*75) in Exon 14, c.1285dupC (p.His429Profs*27) in Exon 11 and c.1429C>T (p.Arg477*) in Exon 12. Furthermore, our analysis identified 26 novel FLCN mutation loci (Supplementary Table 1).

Cross-ethnic genotype-phenotype correlations

Given the pronounced phenotypic disparities across different ethnicities, we compared our cohort with a large, previously published Caucasian cohort to investigate genotype-phenotype correlations. The comprehensive frequency distribution of genetic mutations for both cohorts is illustrated in Figure 3, revealing a striking divergence in the mutational spectra between the two populations.

Since high-frequency mutational loci are likely the most representative of the underlying genetic background, we conducted a phenotypic comparison between patients harboring these frequent variants in both cohorts. Patients harboring c.1285dupC/delC in the Caucasian cohort were selected. As our Chinese cohort presented with multiple variants of comparable frequencies, we tentatively selected the top three most prevalent loci: c.1579_1580insA, c.1285dupC, and c.1429C>T. Subsequently, a 2 × 4 contingency table (Table 3) was constructed to compare the distribution of clinical phenotypes between individuals carrying these high-frequency mutations in the tow cohorts. For the Chinese cohort, only probands were integrated to prevent potential statistical bias due to the incompleteness of clinical data for their relatives; the Caucasian cohort, however, comprised both probands and their family members. While the Caucasian group showed a wide phenotypic distribution across all four categories, the Chinese group almost exhibited only lung cysts and spontaneous pneumothorax, with just two cases of fibrofolliculoma and no renal tumors.

Given the small sample size and the presence of a zero cell, Fisher's exact test was utilized to ensure statistical robustness. The result demonstrated a statistically significant difference in phenotype distribution between the two groups (two-sided $P = 3.0\times10^{-6}$), suggesting a potential association between genotype and specific clinical presentation.

Single-cell transcriptomic landscape

Epithelial depletion and immune-stromal expansion

To elucidate the cellular mechanisms underlying pulmonary cyst formation in BHD syndrome, single-cell RNA sequencing (scRNA-seq) was performed on lung tissues from four BHD patients and integrated with six healthy donor lung datasets. Following rigorous Quality control and batch-effect correction (Supplementary Figure 1 and 2), Uniform Manifold Approximation and Projection (UMAP) visualization successfully delineated the major cell types based on canonical marker gene expression (Figure 4a-b). Compositional analysis uncovered a dramatic remodeling of the normal architecture in the BHD lung microenvironment (Figure 4c). Detailed cellular composition changes in BHD syndrome are shown in Supplementary Figure 3. Compared to healthy controls, BHD lungs exhibited a significant reduction in the proportion of neutrophil and endothelial cells. Concurrently, there was a pathological surge in recruited immune populations, characterized by the expansion of T_NK_ILC and DC populations. Furthermore, although likely constrained by the limited cohort size, we observed a trend toward epithelial depletion ($p = 0.088$) alongside an expansion of the fibroblast population that approached marginal statistical significance ($p = 0.055$).

Hyperactive communication and transcriptomic remodeling

To elucidate the functional alterations within the BHD microenvironment, we conducted differential enrichment analysis of signaling pathways and transcription factors (TF) networks (Figure 4d-e). The BHD group exhibited a profound activation of pathological remodeling, predominantly driven by epithelial-mesenchymal transition (EMT) programs. This structural shift was anchored by the hyperactivation of key EMT-driving TFs, including TWIST1, PRRX2, and WT1, alongside a concurrent enrichment of the 'Focal adhesion' pathway in both up- and downregulated gene sets, suggesting a potential turnover of cell-matrix adhesions. Alongside these structural alterations, we also observed an intense inflammatory response, with specific activation of the 'NF-kappa B', 'TNF', and 'IL-17' signaling pathways, orchestrated by core inflammatory TFs such as JUNB and NFATC2. Concurrently, immune stress responses were evidenced by the upregulation of stress-related TFs (e.g., ATF3 and CREM) combined with the enrichment of 'NOD-like receptor' and 'Antigen processing' pathways, while active cell death processes were distinctly underscored by the robust activation of 'Apoptosis' and 'Necroptosis' cascades. Conversely, paralleling this structural deterioration, we observed a prominent downregulation of essential structural maintenance programs ('Adherens junction' and 'Tight junction') and fundamental 'Metabolic pathways' (Supplementary Figure 4). At the intercellular level, interactome analysis demonstrated a global hyperactivation of signaling networks in the BHD lung (Figure 4f-h). Bar charts and chord diagrams illustrated an overall increase in both the quantity and total interaction strength of cellular communication compared to controls. This network hyperactivation was uniquely characterized by drastically enhanced epithelial-stromal crosstalk and prominent immune recruitment, indicating a highly interactive, degraded localized tissue microenvironment.

Pathological "Epi-EMT-like" Subpopulation

Given the profound depletion of structural epithelial cells, high-resolution subclustering of the epithelial lineage was conducted, identifying six distinct epithelial subtypes (Figure 5a-c). Strikingly, this analysis revealed a fundamentally aberrant, paradoxical subpopulation exclusively present in the BHD group, termed "Epi-EMT-like" cells. This subpopulation was completely absent in the healthy control group (p=0.008) (Figure 5d-e). At the transcriptomic level, Epi-EMT-like cells lost their true cellular identity, concurrently highly expressing canonical epithelial markers (e.g., KRT8 and MYRF) alongside mesenchymal and pro-fibrotic matrix markers (e.g., COL1A1 and SPARC).

Simultaneously, the normal Alveolar Type 2 (AT2) cell subpopulation was significantly diminished in the BHD group. To determine the developmental origin of the Epi-EMT-like cells, pseudo-time trajectory analysis was performed (Figure 5f). The inferred trajectory was compatible with a branch from AT2 toward the putative Epi-EMT-like cluster; lineage tracing or tissue validation is required.

Networks orchestrating Epi-EMT-like state

To uncover the molecular drivers initiating this severe deterioration, SCENIC analysis was utilized to evaluate up- or down-regulated transcription factor networks (regulons) across the epithelial subpopulations (Figure 5h). Assessment of the Regulon Specificity Score (RSS) isolated the key active transcription factor networks defining the Epi-EMT-like state (Figure 5g). The analysis revealed five prominent regulatory networks, including NR2F1(+), MEOX2(+), MEIS2(+), TWIST1(+), and SOX6(+). Concurrently, we observed a profound silencing of transcription factor networks essential for maintaining normal epithelial lineage identity, notably TP63(-) and FOXB1(-). The synergistic interplay between the hyperactivation of these EMT-driving factors and the loss of epithelial maintainers may contribute to the transcriptional reprogramming of these aberrant epithelial cells, potentially shaping a localized dysregulated transcriptional program that precipitates pathological matrix remodeling and characteristic cyst formation in BHD syndrome. The complete SCENIC regulon activity landscape is presented in Supplementary Figure 5.

Discussion

This study represents the first large-scale BHD cohort from southern China, including 107 patients from 85 unrelated families. Consistent with previous reports describing Asian patients with BHD, our cohort demonstrated a predominance of pulmonary manifestations, with lower frequencies of cutaneous and renal involvement and a high incidence of spontaneous pneumothorax. All patients exhibited pulmonary cysts, and the prevalence of spontaneous pneumothorax was 77.1%, slightly higher than the 61% estimated in a recent systematic review[25]. This difference may reflect referral bias, as our center is a specialized respiratory department with a particular focus on pulmonary manifestations of BHD. In contrast, cutaneous manifestations were observed in only 6.3% of patients, substantially lower than previously reported in other Chinese cohorts (48.7% and 72.3%)[26]. This discrepancy may be partially explained by differences in clinical assessment, as most patients in our cohort were evaluated primarily by respiratory physicians, with dermatological consultation performed when necessary. Establishing

multidisciplinary management involving dermatologists and increasing the use of skin biopsy may improve the may improve the detection rate of cutaneous manifestations in future clinical practice. Additionally, regional variation may contribute to the observed heterogeneity in clinical presentation. The prevalence of renal tumors was 10.3%, comparable to previous Chinese cohorts (7.1%)[26], with chromophobe renal cell carcinoma being the predominant subtype, consistent with previous reports and distinct from sporadic renal cell carcinoma, in which clear cell renal cell carcinoma is the most common histological subtype[27].

However, unlike findings from other Asian BHD cohort studies, our research found that the high-frequency FLCN gene mutation loci occur at c.1579_1580insA, c.1285dupC and c.1429C>T. This contrasts with a Japanese study of 334 BHD patients, which identified high-frequency mutation loci at c.1285dupC, c.1533_1536delGATG, and c.1347_1353dupCCACCCT[16]. Furthermore, compared with other Chinese cohort studies, our high-frequency loci also differed: a 2022 study of 143 BHD families from various regions in China reported that c.1285dupC/delC was the most common mutation, occurring at a frequency far exceeding other loci[10]. In contrast, c.1579_1580insA was the most common in our cohort. This may be attributed to regional differences, as patients at our center primarily originate from Southern China. In summary, the genetic mutation profile of BHD may exhibit variations between different regions within the same racial group. In addition, we discovered 26 novel FLCN gene mutation loci, significantly expanding the known mutational spectrum of the disease.

Among the 107 Chinese patients in our cohort, the most frequent FLCN variants were c.1579_1580insA, c.1285dupC, and c.1429C>T, whereas c.1285dupC/delC represented the predominant mutational hotspot in a previously reported Caucasian cohort of 219 patients, indicating substantial heterogeneity in the FLCN mutational landscape between the two populations. We further compared the clinical manifestations of patients carrying recurrent variants between Chinese and Caucasian cohorts. Given that c.1285dupC/delC accounted for the majority of variants in the Caucasian cohort, whereas multiple recurrent variants with relatively similar frequencies were observed in our cohort, and considering the limited sample size of our cohort, we combined patients carrying major recurrent variants in the Chinese cohort for phenotype comparison. The distribution of clinical manifestations revealed significant phenotypic heterogeneity between populations: Chinese patients were predominantly characterized by pulmonary involvement, whereas Caucasian patients exhibited more frequent cutaneous and renal manifestations accompanied by pulmonary abnormalities. These findings suggest that differences in the FLCN mutational spectrum may partially contribute to the phenotypic variability of BHDS between Chinese and Caucasian populations. However, whether cutaneous lesions and renal tumors are truly less frequent among Chinese patients carrying these recurrent variants requires validation in larger cohorts. Notably, among patients with recurrent variants identified in our cohort, the only two individuals with cutaneous manifestations both carried the c.1285dupC variant, suggesting a potential association between c.1285dupC and cutaneous involvement. Conversely, whether Caucasian patients harboring variants such as c.1579_1580insA or c.1429C>T also exhibit lower frequencies of cutaneous or renal

manifestations remains unknown. Addressing this question could further corroborate the potential impact of FLCN genotypes on BHD syndrome phenotypes.

Perhaps the most significant contribution of this study is the first comprehensive characterization of the single-cell transcriptomic landscape of BHD lungs using scRNA-seq. Although scRNA-seq has been applied to investigate BHD-associated renal lesions [20–22]—revealing the dual cellular origin of BHD-related renal carcinomas, a paucity of non-tumor infiltrating cells within the tumor microenvironment, and the aberrant activation of TFEB signaling in FLCN-deficient epithelial cells driving renal cystogenesis—its application in elucidating BHD pulmonary pathogenesis has remained unexplored.

By integrating lung samples from four patients with BHD with publicly available healthy lung datasets, we identified profound disruption of epithelial homeostasis in BHD lungs. Normal structural epithelial populations were reduced, accompanied by an increase in activated fibroblasts and infiltrating immune cell subsets, indicating extensive remodeling of the pulmonary microenvironment.

A key finding of our study was the identification of a previously unrecognized “Epi-EMT-like” epithelial subpopulation. These cells exhibited a loss of canonical epithelial identity and acquired mesenchymal-like features, suggesting a potential pathological alteration of epithelial cell fate. Against the genetic background of BHD syndrome, alveolar type 2 (AT2) cells may be prone to aberrant differentiation trajectories rather than normal maturation. Pseudotime analysis demonstrated that BHD AT2 cells may deviated from physiological epithelial differentiation and acquired EMT-like characteristics, including downregulation of epithelial adhesion molecules and upregulation of mesenchymal and extracellular matrix-related markers such as COL1A1.

This epithelial identity alteration, potentially regulated by transcriptional programs involving factors such as TWIST1 and SOX6, may contribute to impaired alveolar structural integrity. Loss of epithelial junctional stability and acquisition of matrix-remodeling properties could increase susceptibility of alveolar walls to mechanical stress. Furthermore, cell–cell communication analysis revealed that Epi-EMT-like cells exhibited potential enhanced interactions with stromal and immune compartments, accompanied by activation of hypoxia-related pathways, angiogenic responses, and fibroblast recruitment. The convergence of this localized matrix remodeling and recurrent mechanical stress likely drives the formation, expansion, and subsequent rupture and fusion of the characteristic thin-walled pulmonary cysts in BHD syndrome. While the previously proposed “stretch hypothesis”[28] emphasizes intrinsic alveolar fragility and mechanical forces generated during respiration as major contributors to cyst formation, our findings indicate that aberrant matrix remodeling may also play important roles. Recently, a murine model study reported that *Flcn* deletion in the pulmonary mesenchyme, but not the epithelium, leads to lung cyst formation[29]. However, in human BHD lung tissues, our single-cell analysis indicates that the epithelial compartment may also undergoes notable transcriptomic changes, highlighted by the emergence of the Epi-EMT-like subpopulation. Additionally, our interactome data (Figure 4) revealed enhanced signaling between epithelial and stromal cells. These findings suggest that

human BHD cystogenesis likely involves interactive processes between the altered epithelium and the mesenchyme, rather than acting through a single cell lineage.

This study has several limitations. The single-center, respiratory specialty-based cohort may have introduced referral and selection bias, limiting the generalizability of our clinical findings. Given the rarity of BHD syndrome, the limited number of lung samples available for single-cell RNA sequencing may have reduced statistical power and warrants validation in larger cohorts. The bioinformatic analyses provide primarily associative and inferential evidence; therefore, the proposed cellular and transcriptional mechanisms require further validation using spatial, molecular, and functional experiments.

Conclusions

In conclusion, our comprehensive analysis provides the first baseline clinical characteristics and genetic mutational landscape of BHD syndrome in southern China, expands the known mutational spectrum in Asian populations and underscoring ethnic genotype-phenotype heterogeneities. Notably, by presenting the first single-cell transcriptomic atlas of BHD lung tissues globally, our study provides new insights into the mechanisms underlying pulmonary cyst formation, and may facilitate the development of future targeted therapeutic strategies for BHD-associated pulmonary disease.

Abbreviations

AML: angiomyolipoma ANOVA: analysis of variance AT2: Alveolar Type 2 BEAM: branch expression analysis modeling BHD: Birt-Hogg-Dubé DEGs: Differentially expressed genes EMT: epithelial-mesenchymal transition FF: fibrofolliculoma FLCN: folliculin GDC: Genomic Data Commons GEO: Gene Expression Omnibus HGMD: Human Gene Mutation Database LSDB: locus-specific database MNN: mutual nearest neighbors P/LP: pathogenic/likely pathogenic PT: primary spontaneous pneumothorax RSS: Regulon Specificity Score scRNA-seq: single-cell RNA sequencing TF: transcription factors TSC: Tuberous Sclerosis Complex UMAP: Uniform Manifold Approximation and Projection VEP: Variant Effect Predictor

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the First Affiliated Hospital of Guangzhou Medical University (approval number: ES-2023-197).

Consent for publication

Written informed consent was obtained from all participants.

Availability of data and materials

The control single-cell RNA sequencing data are publicly available in the Gene Expression Omnibus (GEO) under accession number GSE220797. The single-cell RNA sequencing data generated in this study will be deposited in a public repository upon publication and are available from the corresponding author upon reasonable request.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the “Natural Science Foundation of Guangdong Province” (2023A1515010308), “Guangdong Province 2022 Graduate Education Innovation Plan project” (2022ANLK049), the “Guangzhou Rare Respiratory Disease Diagnosis and Treatment Center Project”, and the “Clinical and Epidemiological Research Project of State Key Laboratory of Respiratory Disease” (SKLRD-L-202705).

Authors' contributions

Conceptualization, XY.X, YC.L and J.L; Data curation, XY.X, LC.Y, YY.H and PQ.L; Formal analysis, XY.X, DR.L and YJ.M; Funding acquisition, J.L; Investigation, XY.X, YC.L, LC.Y, LD.L and C.C; Methodology, XY.X and YJ.W; Project administration, XY.X and J.L; Resources, XY.X, LC.Y and YY.H; Supervision, J.L; Validation, XY.X, PQ.L, DR.L, YJ.M, LD.L and C.C; Visualization, XY.X, YC.L and LC.Y; Writing—original draft, XY.X; Writing—review and editing, XY.X, LC.Y and J.L. All authors have read and agreed to the final version of the manuscript.

Acknowledgements

We thank NovelBio Co., Ltd. for support with single-cell sequencing and data analysis. And we thank the generative artificial intelligence assistant (Gemini 3, Google) for linguistic refinement. The final manuscript text was meticulously edited, and all illustrative figures were original works manually created by the authors to ensure scientific accuracy. The authors take full responsibility for the integrity and academic quality of the submitted work.

References

1. Hornstein OP, Knickenberg M. Perifollicular fibromatosis cutis with polyps of the colon--a cutaneo-intestinal syndrome sui generis. *Arch Dermatol Res = Arch Fur Dermatol Forsch* 1975;253(2):161–175.
2. Muller M-E, Daccord C, Taffé P, Lazor R. Prevalence of birt-hogg-dubé syndrome determined through epidemiological data on spontaneous pneumothorax and bayes theorem. *Front Med* 2021;8:631168.
3. Savatt JM, Shimelis H, Moreno-De-Luca A, et al. Frequency of truncating FLCN variants and birt-hogg-dubé-associated phenotypes in a health care system population. *Genet Med: Off J Am Coll Med*

Genet 2022;24(9):1857–1866.

4. Yngvadottir B, Richman L, Andreou A, et al. Inherited predisposition to pneumothorax: estimating the frequency of birt-hogg-dubé syndrome from genomics and population cohorts. *Thorax* 2025;80(8):553–555.
5. Hu X, Zhang G, Chen X, Xu K-F. Birt–Hogg–Dubé syndrome in Chinese patients: a literature review of 120 families. *Orphanet J Rare Dis* 2021;16(1):223–231.
6. Schmidt LS, Warren MB, Nickerson ML, et al. Birt-hogg-dubé syndrome, a genodermatosis associated with spontaneous pneumothorax and kidney neoplasia, maps to chromosome 17p11.2. *Am J Hum Genet* 2001;69(4):876–882.
7. Nickerson ML, Warren MB, Toro JR, et al. Mutations in a novel gene lead to kidney tumors, lung wall defects, and benign tumors of the hair follicle in patients with the birt-hogg-dubé syndrome. *Cancer Cell* 2002;2(2):157–164.
8. Toro JR, Glenn G, Duray P, et al. Birt-hogg-dubé syndrome: a novel marker of kidney neoplasia. *Arch Dermatol* 1999;135(10):1195–1202.
9. Daccord C, Good J-M, Morren M-A, Bonny O, Hohl D, Lazor R. Birt–hogg–dubé syndrome. *Eur Respir Rev* 2020;29(157):200042.
10. Zhou W, Liu K, Xu K-F, Liu Y, Tian X. Clinical and Genetic Comparison of Birt-Hogg-Dubé Syndrome (Hornstein-Knickenberg Syndrome) in Chinese: A Systemic Review of Reported Cases. *Int J Gen Med* 2022;15:5111–5121.
11. Gunji Y, Akiyoshi T, Sato T, et al. Mutations of the birt–hogg–dubé gene in patients with multiple lung cysts and recurrent pneumothorax. *J Med Genet* 2007;44(9):588–593.
12. Ren H-Z, Zhu C-C, Yang C, et al. Mutation analysis of the FLCN gene in chinese patients with sporadic and familial isolated primary spontaneous pneumothorax. *Clin Genet* 2008;74(2):178–183.
13. Furuya M, Yao M, Tanaka R, et al. Genetic, epidemiologic and clinicopathologic studies of japanese asian patients with birt–hogg–dubé syndrome. *Clin Genet* 2016;90(5):403–412.
14. Schmidt LS, Nickerson ML, Warren MB, et al. Germline BHD-mutation spectrum and phenotype analysis of a large cohort of families with Birt-Hogg-Dubé syndrome. *Am J Hum Genet* 2005;76(6):1023–1033.
15. Khoo SK, Giraud S, Kahnoski K, et al. Clinical and genetic studies of birt-hogg-dubé syndrome. *J Med Genet* 2002;39(12):906–912.
16. Namba Y, Ebana H, Okamoto S, et al. Clinical and genetic features of 334 Asian patients with Birt-Hogg-Dubé syndrome (BHDS) who presented with pulmonary cysts with or without a history of pneumothorax, with special reference to BHDS-associated pneumothorax. *PLoS One* 2023;18(7):e0289175.
17. Liu Y, Xu Z, Feng R, et al. Clinical and genetic characteristics of chinese patients with birt-hogg-dubé syndrome. *Orphanet J Rare Dis* 2017;12(1):104.

18. Liu K, Xu W, Tian X, et al. Genotypic characteristics of Chinese patients with BHD syndrome and functional analysis of FLCN variants. *Orphanet J Rare Dis* 2019;14(1):223–235.
19. Wang Y, Cai M, Jiang X, et al. Exons 1–3 deletion in FLCN is associated with increased risk of pneumothorax in chinese patients with birt-hogg-dubé syndrome. *Orphanet J Rare Dis* 2023;18(1):115–124.
20. Jikuya R, Murakami K, Nishiyama A, et al. Single-cell transcriptomes underscore genetically distinct tumor characteristics and microenvironment for hereditary kidney cancers. *iScience* 2022;25(6):104463.
21. Wang X-M, Mannan R, Zhang Y, et al. Hybrid oncocytic tumors (HOT) in birt-hogg-dubé syndrome patients: a tale of two cities. 2024;
22. Shalaby O, Ohmori T, Miike K, et al. Folliculin deletion in the mouse kidney results in cystogenesis of the loops of henle via aberrant TFEB activation. *Am J Pathol* 2025;195(9):1643–1659.
23. Menko FH, Steensel MAM van, Giraud S, et al. Birt-Hogg-Dubé syndrome: diagnosis and management. *Lancet Oncol* 2009;10(12):1199–1206.
24. Birt-Hogg-Dubé Syndrome Diagnosis and Management Chinese Expert Consensus Group, Respiratory Branch of China Alliance for Rare Diseases, Chinese Thoracic Society. Chinese expert consensus on the diagnosis and management of Birt-Hogg-Dubé syndrome [in Chinese]. Chinese Journal of Tuberculosis and Respiratory Diseases. 2023;46:897–908.
25. Zhang Y, Wang Y, Wang J, Li P, Lv R, Chen J. The prevalence of spontaneous pneumothorax in patients with BHD syndrome: a systematic review and meta-analysis. *Orphanet J Rare Dis* 2025;20(1):218.
26. Chen XM, Xie XY, Zhang XH, et al. Clinical and genetic characteristics of Birt-Hogg-Dubé syndrome in Chinese patients [in Chinese]. Chinese Journal of Tuberculosis and Respiratory Diseases. 2026;49:29–37.
27. Sattler EC, Reithmair M, Steinlein OK. Kidney cancer characteristics and genotype-phenotype-correlations in birt-hogg-dubé syndrome. *PLOS One* 2018;13(12):e0209504.
28. Kennedy JC, Khabibullin D, Henske EP. Mechanisms of pulmonary cyst pathogenesis in birt–hogg–dube syndrome: the stretch hypothesis. *Semin Cell Dev Biol* 2016;52:47–52.
29. Cao K, Chen H, Chu L, et al. Hyperactivation of mTORC1 signaling mediates folliculin deficiency-induced pulmonary cyst formation in birt-hogg-dubé syndrome. *J Clin Invest* 2026;136(4):e194300.
30. Lim DHK, Rehal PK, Nahorski MS, et al. A new locus-specific database (LSDB) for mutations in the folliculin (FLCN) gene. *Hum Mutat* 2010;31(1):E1043-1051.

Tables

Table 1. Diagnostic criteria for Birt–Hogg–Dubé (BHD) syndrome
Diagnostic criteria proposed by the European BHD Consortium[23]
Patient should fulfil one major or two minor criteria for diagnosis
Major criteria:
At least five fibrofolliculomas or trichodiscomas, at least one histologically confirmed, of adult onset
Pathogenic FLCN germline mutation
Minor criteria:
Multiple lung cysts: bilateral basally located lung cysts with no other apparent cause, with or without spontaneous primary pneumothorax
Renal cancer: early onset (age <50 years) or multifocal or bilateral renal cancer, or renal cancer of mixed chromophobe and oncocytic histology
A first-degree relative with BHD
Diagnostic criteria□China□[24]
Patient should fulfil one clinical criterion plus genetic criteria or two clinical criteria
Clinical criteria
Multiple lung cysts: diffuse lung cysts with no other apparent cause, mainly at the basal and mediastinum with or without spontaneous pneumothorax
Rash: at least five fibrofolliculomas or trichodiscomas, mostly on face or neck
Renal cancer: early onset (age<50 years), or multifocal or bilateral renal cancer, or renal cancer of mixed chromophobe and oncocytic histology
Genetic criteria
Pathogenic FLCN germline mutation
A first-degree relative with BHD

Table2. Demographic Data and Clinical Characteristics of 85 BHD Probands

Item	case/total (%)	Item	case/total (%)
Male/Female	11/74	Cutaneous lesions	5/79(6.3)
Diagnostic Modality		Renal lesions	36/78(46.2)
Genetic testing	84/85(98.8)	Renal tumors	8/36□25□
Clinical diagnosis (pulmonary cysts + renal cancer)	1/85(1.2)	Chromophobe renal cell carcinoma	2/8(25)
Family History	38/85(44.7)	Multilocular cystic renal	1/8(12.5)
Pulmonary cystic lesions or	38/38(100)	neoplasm	
pneumothorax		Pheochromocytoma	1/8(12.5)
Renal tumors	0(0)	Renal mass pending	1/8(12.5)
Pulmonary Cystic Lesions	85/85(100)	pathological confirmation	
Distribution		Renal angiomyolipoma (AML)	3/8(37.5)
Unilateral	0(0)	Horseshoe kidney	1/36(2.7)
Bilateral	85/85(100)	Renal cysts	25/36(69.4)
Predominantly upper lobe	0(0)		
Predominantly lower lobe	85/85(100)		
Pathology	45/85(52.9)		
Pulmonary bullae	34/45(75.6)		
Pneumothorax	64/83(77.1)		
First episode	23/63(36.5)		
Recurrence	40/63(63.5)		
Conservative treatment	24/36(66.7)		
Pleurodesis	1/36(2.7)		
Bullectomy	11/36(30.6)		

Note: Conservative treatment for pneumothorax includes clinical observation, supplemental oxygen therapy, and closed thoracic tube drainage.

Table 3 Clinical Manifestations of Chinese and Caucasian BHD Patients Carrying Frequent FLCN Mutations

Mutations	FF	Renal tumor	Lung cysts	PT
Caucasian[a]	73	14	49	34
Chinese[b]	2	0	19	14

Abbreviations: FF, fibrofolliculoma; PT, primary spontaneous pneumothorax.

Note: Fisher's exact test was applied due to low cell counts in the Chinese group. [a] Comprise patients carrying c.1733insC and c.1733delC mutations. [b] Include probands carrying c.1579_1580insA, c.1285dupC, c.1429C>T mutations.

Figures

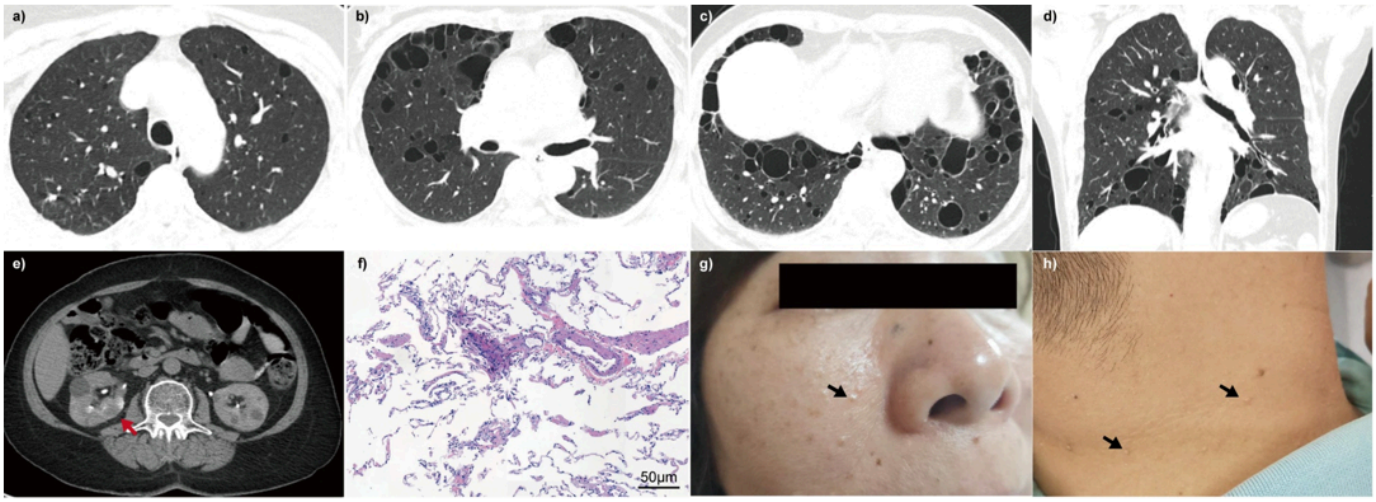


Figure 1

Pulmonary, Renal, and Cutaneous Manifestations in Patients with Birt-Hogg-Dubé Syndrome (BHD).

a-d) CT images of the lungs from a BHD patient demonstrate cystic lesions. e) Abdominal computed tomography scan showing a multilocular cystic renal neoplasm located in the right kidney (arrow), which exhibits progressive enhancement after injection of contrast media. f) Haematoxylin and Eosin stained micrograph of lung tissue from a BHD syndrome patient demonstrating characteristic thin-walled bullous changes and scattered lymphocytic infiltration within the septa. g-h) Multiple millet-sized, flesh-colored papules (fibrofolliculomas) on the neck of a BHD patient (arrows).

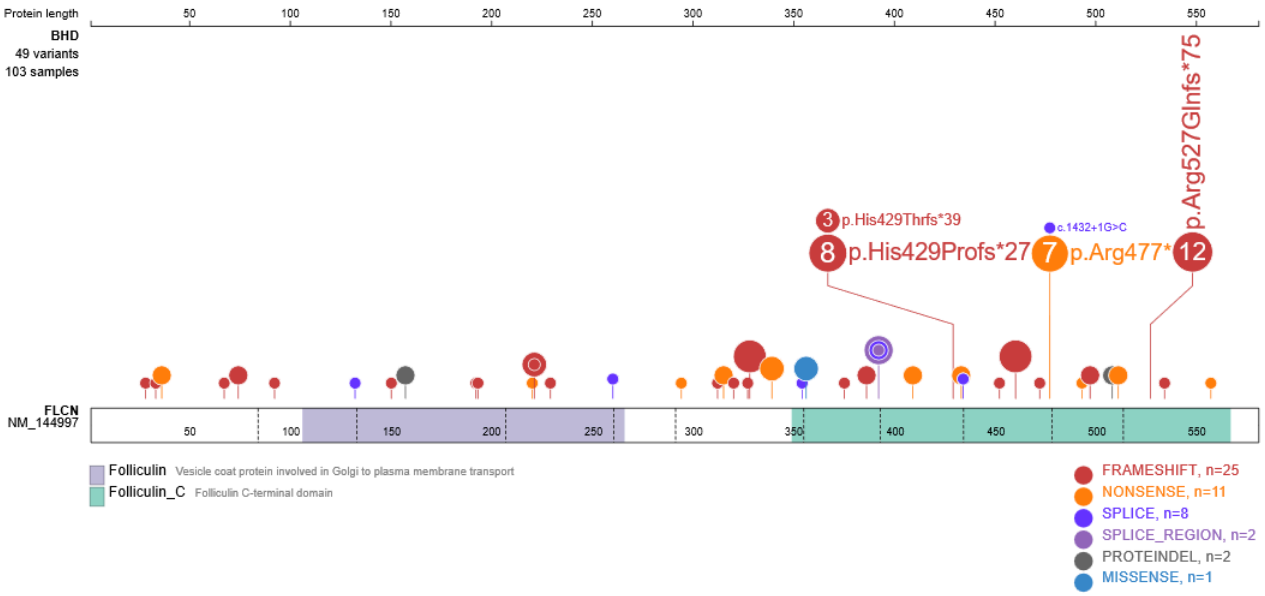


Figure 2

FLCN gene mutations of 85 BHD families.

The lollipop plot illustrates the spatial distribution and frequncy of point mutations, small insertions/deletions (indels), and splice-site variants mapped to the Folliculin protein sequence (RefSeq transcript NM_144997.7). Each circle represents a distinct genetic alteration, with the size of the circle and the inscribed integer denoting the variant recurrence within the WES-diagnosed cohort. Distinct mutation classes are color-coded as follows: frameshift (red), nonsense (orange), splice-site/splice-region variants (purple/light purple), in-frame protein deletions (grey), and missense mutations (blue). The colored boxes along the horizontal axis indicate the conserved functional domains of the Folliculin protein, including the vesicle coat protein domain (purple) and the Folliculin C-terminal domain (green). Note: Large fragment deletions identified in this cohort are not depicted in this amino acid-level schematic.

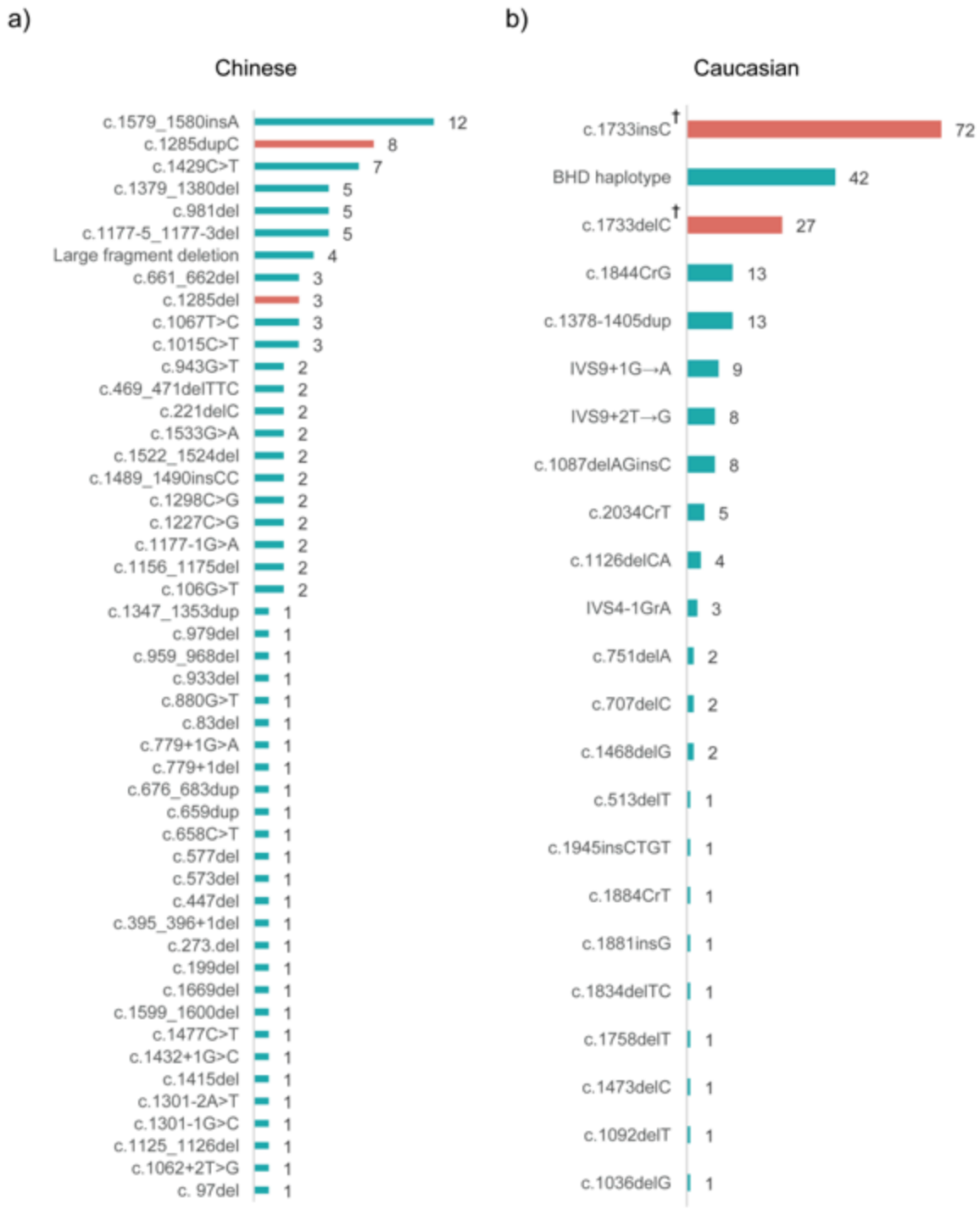


Figure 3

The frequency of FLCN gene mutations of 85 BHD Chinese and 53 Caucasian (Laura S Schmidt et al. 2005 [14] **)families.**

a) The c.1579_1580insA, c.1285dupC, c.1429C>T has been reported predominantly in our Chinese populations, b) while the c.1733insC and c.1733delC mutations are the most commonly observed in

individuals of the Caucasian cohort. The established hotspot c.1285dup/delC is marked in red. †The c.1733ins/delC mutations are renamed as 1285dup/delC currently[30].

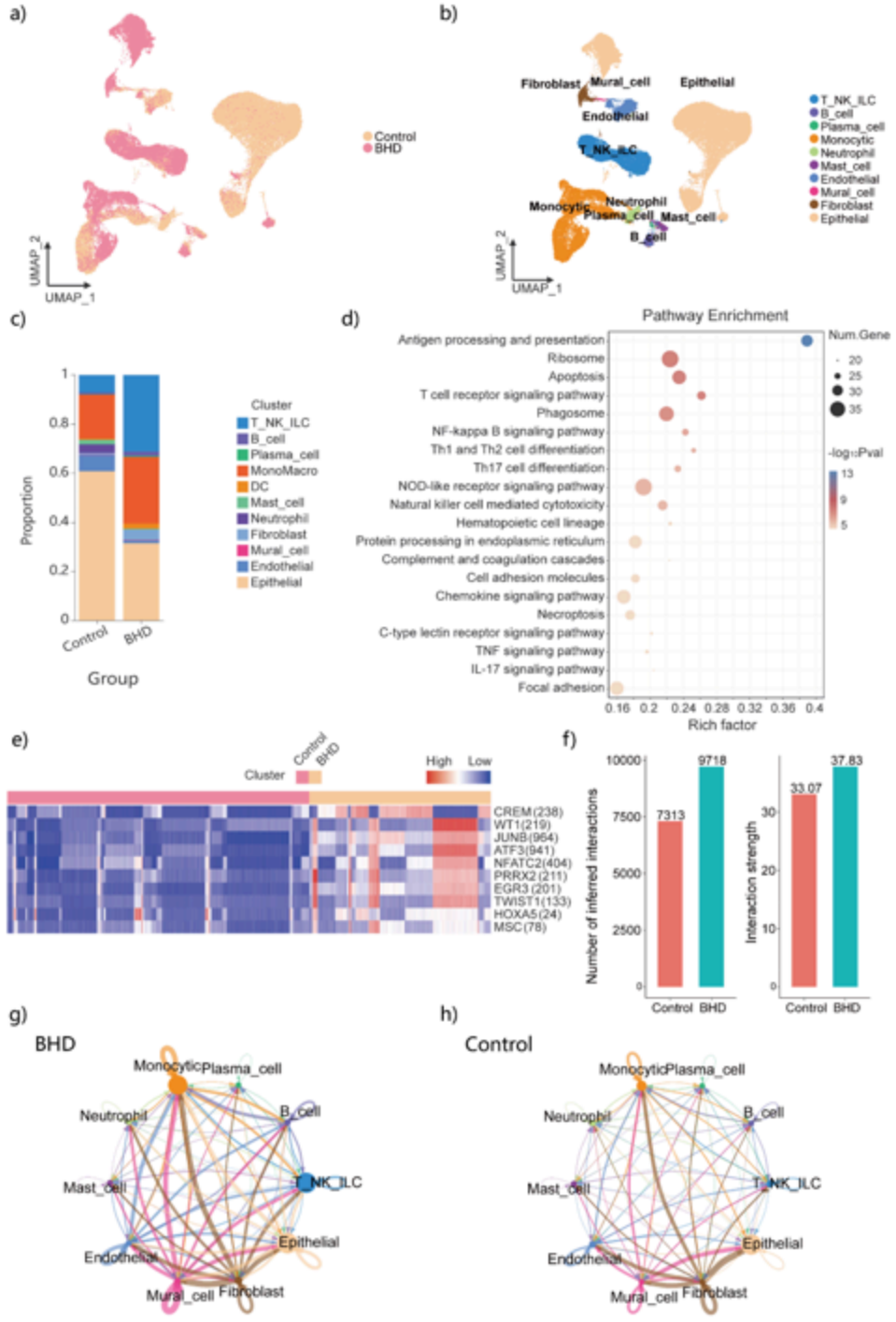


Figure 4

Single-cell transcriptomic landscape reveals cellular composition shifts and hyperactive intercellular crosstalk in BHD lung tissues.

a) UMAP plot of single-cell RNA sequencing (scRNA-seq) data from 4 human BHD lung tissue samples integrated with 6 previously published healthy donor lung datasets as controls, illustrating group distribution following batch-effect correction.

b) UMAP plot highlighting the major cell types identified in the integrated dataset shown in (a), based on canonical marker gene expression.

c) Stacked bar chartdisplaying the comparison of major cell population proportions between BHD lungs and healthy controls. BHD lungs showed significantly lower proportions of neutrophils and endothelial cells, accompanied by increased proportions of T_NK_ILC and dendritic cells. Epithelial cells showed a trend toward depletion ($p = 0.088$), whereas fibroblasts showed a trend toward expansion that approached statistical significance ($p = 0.055$).

d-e) Bubble chart (d) presenting significantly enriched signaling pathways in the BHD group compared to the control group, and heatmaps (e) illustrating significantly enriched transcription factors. Cells in the BHD group demonstrated intense activation states regarding immune stress response, inflammatory signal transduction, apoptosis and stress, and tissue remodeling/epithelial-mesenchymal transition (EMT).

f) Bar charts showing an overall increase in the quantity and intensity of cellular communication in the BHD group compared to the control group.

g-h) Chord diagrams (Circle plots) illustrating the overall interaction strength (communication probability) of the signaling networks in the BHD group (g) and the control group (h). The size of each node correlates positively with the total interaction strength for that specific cell type, while line thickness reflects the relative communication strength between two corresponding cell populations. Arrows indicate the direction of signal transmission (from sender to receiver). This analysis highlights a global hyperactivation of the interaction network in the BHD group, characterized by drastically enhanced epithelial-stromal crosstalk and immune recruitment.

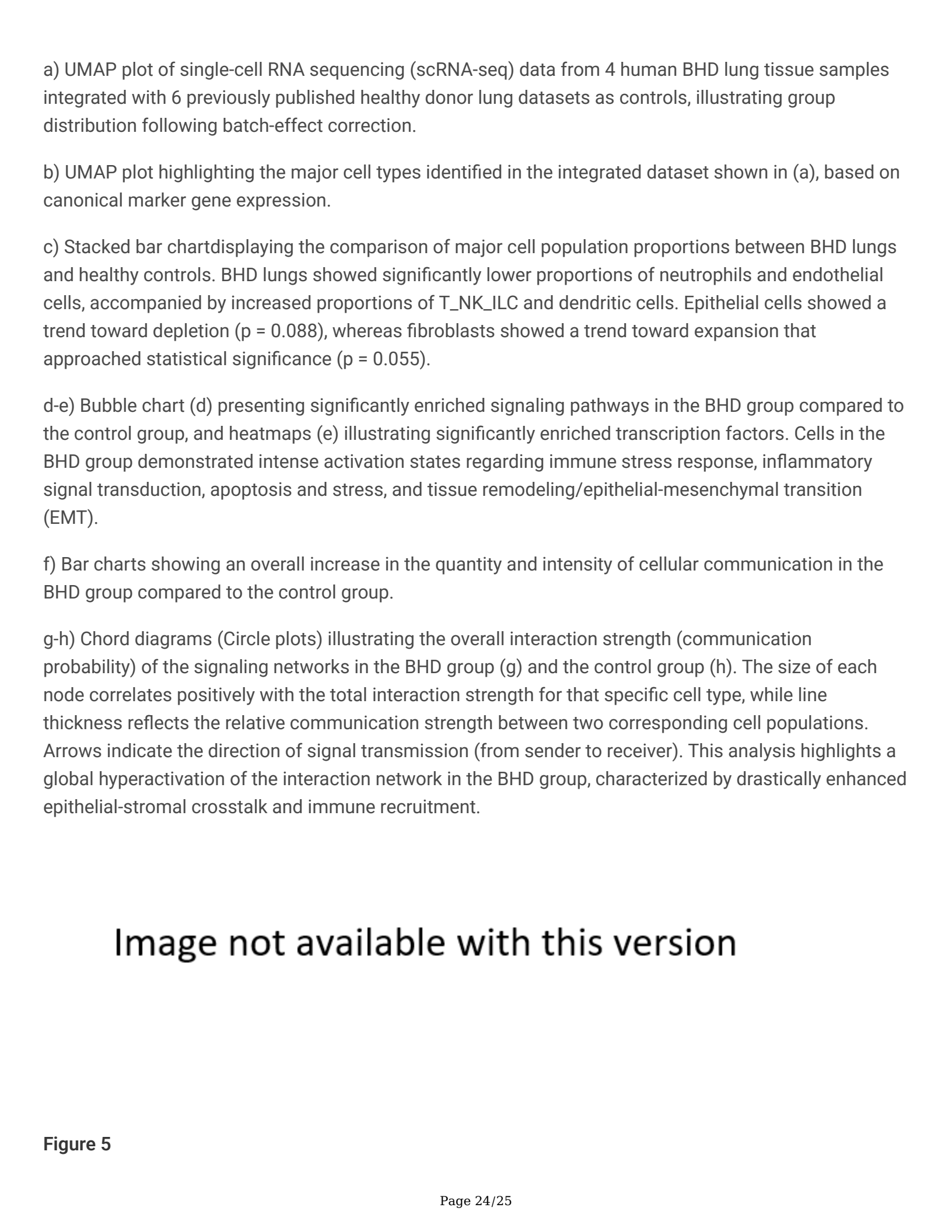


Figure 5

Figure 5. Identification, developmental trajectory, and transcriptional drivers of a novel BHD-specific Epi-EMT-like subpopulation.

a) UMAP plot illustrating the group distribution (BHD vs. Control) of the extracted epithelial cell populations following batch-effect correction.

b) UMAP plot focused on the epithelial population, revealing six distinct epithelial subtypes based on transcriptional clustering.

c) Bubble plots presenting the marker genes defining epithelial subpopulations. We identified a newly discovered Epi-EMT-like subpopulation; these cells highly express epithelial markers such as KRT8 and MYRF, as well as mesenchymal markers like COL1A1 and SPARC.

d-e) Stacked bar charts (d) displaying the proportions of the six epithelial subpopulations between BHD lungs and healthy controls. The AT2 subpopulation was significantly reduced in the BHD group compared to the control group, whereas the Epi-EMT-like subpopulation was exclusive to the BHD group and completely absent in the control group. Box plots (e) demonstrate a statistically significant difference regarding the Epi-EMT-like subpopulation between the two groups ($p=0.008$).

f) Cell developmental trajectory plot based on pseudo-time coloring. (Right) The AT2 cell population was computationally defined as the root of this trajectory. The color gradient from dark blue to light blue represents the inferred developmental chronological sequence (from early state to late state). (Left) The distribution of cell types along the trajectory. This plot highlights a branching point (Node 1) with AT2 cells located at the inferred origin, which subsequently separates towards Ciliated and Epi-EMT-like cell states.

g-h) SCENIC analysis (h) displaying up- or down-regulated transcription factor networks (regulons) across various epithelial subpopulations. Scatter plot (g) of the Regulon Specificity Score (RSS) for the regulatory networks in Epi-EMT-like cells based on SCENIC analysis. The top five regulatory networks (NR2F1(+), MEOX2(+), MEIS2(+), TWIST1(+), and SOX6(+)) are highlighted in red, representing the key active transcription factor networks defining this state, alongside the silencing of epithelial maintainers such as TP63.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarymaterial1.docx