

RESEARCH

Open Access



Integrative analysis of gene expression and chromatin dynamics multi-omics data in mouse models of bleomycin-induced lung fibrosis

Zhongzheng Li¹, Mengke Zhang¹, Yujie Zhang¹, Yulong Gan¹, Zhao Zhu¹, Jiawei Wang¹, Yanlin Zhou¹, Guoying Yu^{1*} and Lan Wang^{1*}

Abstract

Background Pulmonary fibrosis is a relentless and ultimately fatal lung disorder. Despite a wealth of research, the intricate molecular pathways that contribute to the onset of PF, especially the aspects related to epigenetic modifications and chromatin dynamics, continue to be elusive and not fully understood.

Methods Utilizing a bleomycin-induced pulmonary fibrosis model, we conducted a comprehensive analysis of the interplay between chromatin structure, chromatin accessibility, gene expression patterns, and cellular heterogeneity. Our chromatin structure analysis included 5 samples (2 control and 3 bleomycin-treated), while accessibility and expression analysis included 6 samples each (3 control and 3 bleomycin-treated).

Results We found that chromatin architecture, with its alterations in compartmentalization and accessibility, is positively correlated with genome-wide gene expression changes during fibrosis. The importance of immune system inflammation and extracellular matrix reorganization in fibrosis is underscored by these chromatin alterations. Transcription factors such as PU.1, AP-1, and IRF proteins, which are pivotal in immune regulation, are associated with an increased abundance of their motifs in accessible genomic regions and are correlated with highly expressed genes.

Conclusions We identified 14 genes that demonstrated consistent changes in their expression, accessibility, and compartmentalization, suggesting their potential as promising targets for the development of treatments for lung fibrosis.

Keywords 3D genome maps, Pulmonary fibrosis, Bleomycin, Chromosome accessibility

*Correspondence:

Guoying Yu
guoyingyu@htu.edu.cn
Lan Wang
wanglan@htu.edu.cn

¹State Key Laboratory of Cell Differentiation and Regulation, Henan International Joint Laboratory of Pulmonary Fibrosis, Henan Center for Outstanding Overseas Scientists of Organ Fibrosis, Pingyuan Laboratory, College of Life Science, Henan Normal University, 46 Jianshe Road, Xinxiang, Henan 453007, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

Introduction

Pulmonary fibrosis (PF) is a type of interstitial lung disease characterized by the formation of scar tissue in the lungs. Although there are over 200 potential causes of PF, they primarily fall into four categories: idiopathic pulmonary fibrosis (IPF), disease-associated PF, familial PF, and exposure-associated PF [1–3]. Moreover, different subgroups within each etiology have been identified based on the affected genes and inheritance patterns. For instance, subgroups can be distinguished between individuals with *MUC5B* mutations and those without [4]. Transcriptomic profiles have delineated a Mitogen-Activated Protein Kinase (MAPK)-active subgroup and an immune-active subgroup [5]. Proteomic studies have identified subgroups associated with cell differentiation, inflammation, and mitochondrial metabolism [6]. Additionally, gene expression changes related to various biological processes such as telomere length [7], immunosuppression [7], oxidative stress [8], and DNA damage [9] are considered markers for distinguishing PF subgroups. Despite extensive research on the pathogenesis of PF, effective treatment options remain elusive.

To better understand the pathogenesis of PF, a variety of animal models have been established. Each model provides a powerful tool for studying certain aspects of fibrotic lung disease [10]. The asbestos-induced fibrosis model primarily emphasizes epithelial cell injury, fibroblast foci, and macrophage oxidative stress [10]. The silica-induced fibrosis model mainly manifests as partial collagen deposition and macrophage *NALP3* inflammasome activation [10]. The Fluorescein Isothiocyanate (FITC) model highlights epithelial injury and vascular leakage, while the TGF β overexpression model emphasizes epithelial-mesenchymal transition (EMT) in the absence of inflammation [10]. The cytokine overexpression fibrosis model focuses on the interaction between TGF β and cytokines [10]. Lung injury (acidic or hypoxic) fibrosis mouse models are used to study fibroproliferative changes in acute lung injury (ALI) [10]. The bleomycin (BLM)-induced murine fibrosis model is the most commonly used experimental model for PF [11]. This model fully embodies the primary alveolar epithelial cell (AEC) injury during fibrosis development [10]. BLM directly causes cell damage by inducing DNA strand breaks, radical production, and oxidative stress, leading to necrosis and/or apoptosis, followed by inflammation and fibrosis [10].

Previous establishment of High-throughput chromosome conformation capture (Hi-C) methods has aided in identifying and visualizing comprehensive whole-genome chromatin interactions in a three-dimensional environment [12, 13], and assay for transposase accessible chromatin with high-throughput sequencing (ATAC-seq) methods have helped to identify changes in

chromatin accessibility sites [14, 15], thereby enabling exploration of the three-dimensional genomic structure and its impact on gene expression regulation. Numerous studies have shown that changes in three-dimensional structure, such as the transition between A/B compartments (The A compartment is linked to open chromatin, while the B compartment is tied to closed chromatin [13].) and changes in chromatin accessibility [16], control gene expression and play a key role in biological processes such as development, cell differentiation, and disease manifestation [17–19]. Epigenetic marks may serve as the elusive link between environmental exposures in genetically predisposed individuals and gene expression changes consistent with disease progression [20–23]. By integrating epigenomic and transcriptomic data, researchers can outline the onset and progression of various diseases, including non-alcoholic fatty liver disease [24] and systemic lupus erythematosus [25]. The pathogenesis of PF is influenced by complex interactions between genetic and environmental factors [1, 3]. Epigenetic marks are emerging as attractive therapeutic targets for PF [21]. Although epigenetic changes during the differentiation of normal alveolar epithelial cells have been depicted [26], the three-dimensional epigenomic landscape of PF in the pathological state remains unclear.

In this study, we conducted a comprehensive multi-omics analysis of the BLM-induced PF mouse model, integrating RNA sequencing (RNA-seq), single cell RNA sequencing (scRNA-seq), ATAC-seq, and Hi-C data. Our analysis identified fibrosis-specific genes and transcription factors (TFs), revealing the regulatory programs underlying the pathogenesis of BLM-induced PF. We present the first reference-quality functional 3D genomic maps of the BLM-induced PF, providing a valuable resource for in-depth, genome-wide, multi-omics analysis of any gene loci. These findings contribute to a better understanding of the complex specific gene expression patterns behind the preclinical model of PF and may facilitate the development of new therapeutic strategies for this devastating disease.

Methods

Generation of BLM-induced PF in mice

Female C57BL/6 mice, aged 9 to 12 weeks, were procured from Veiduoli Hua Company, Beijing, China. Anesthesia was induced in the mice using a 40% isoflurane-oxygen mixture administered via inhalation. Subsequently, the mice were randomly assigned to treatment groups and received an intratracheal instillation of either 1.5 U/kg of BLM dissolved in 50 μ l of a sterile vehicle or an equivalent volume of 0.9% saline. After 21 days, the mice were euthanized by cervical dislocation, and their lungs were promptly excised for subsequent histopathological and molecular analyses.

Hi-C library generation and sequencing

For the Hi-C analysis of lung tissue, we utilized samples from a total of five mice, comprising two control and three BLM-induced fibrosis cases. Approximately 0.5 cm³ of lung tissue from each mouse was meticulously dissected using a razor blade. The tissue fragments were then fixed and subjected to a 20-minute agitation in 10 ml of a 1:10 dilution of 10% neutral buffered formalin (Fisher Scientific). Following fixation, formalin was neutralized by the addition of 0.1 g of glycine (Sigma-Aldrich) and an additional 15 min of agitation. The tissue samples were subsequently dispatched to Frase Genomics, located in Seattle, WA, for Hi-C library construction using their proprietary Mouse Hi-C Kit. Subsequent to library preparation, sequencing was performed on an Illumina HiSeq 4000 platform.

Hi-C data processing

The initial processing of Hi-C paired-end raw data involved the removal of adapter sequences and low-quality reads using the fastp tool (version 0.24) [27]. Subsequently, HiCPro (version 2.7) was employed for mapping, processing, and normalization of the data, incorporating iterative correction strategies [28]. Reads were aligned to the mouse reference genome (mm10) with the bowtie2 algorithm (version 2.5.4) [29]. To ensure data quality, we excluded reads that were indicative of uncut DNA, re-ligated fragments, continuous reads, and PCR duplicates. Reads with a mapping quality score greater than 10 (MAPQ > 10) were utilized to construct the contact matrix. Valid read pairs were assigned to bins across multiple resolutions, resulting in the generation of raw contact matrices at 10, 20, 40, 100, and 200 kilobase (kb) binning resolutions. To correct for biases inherent in the raw contact matrices, such as those associated with GC content, mappability, and effective fragment length, we applied Iterative Correction and Eigenvector-based (ICE) normalization to the Hi-C data.

Heatmap of chromatin interaction frequency and visualization

We combined the Hi-C matrices from our replicate experiments using hicSumMatrices. We then normalized and corrected these matrices using hicNormalize and hicCorrectMatrix, respectively (version 2.1.1) [30]. To visualize the interaction frequencies, we generated heatmaps using hicPlotMatrix at a 500-kb resolution, which was ideal for examining whole-chromosome or whole-genome interactions. To analyze the differences in interaction frequencies between our control and fibrosis samples, we used hicCompareMatrices to determine the log₂ ratios of the interaction frequency matrices [30].

Analyses of A and B compartments

The analysis of A and B chromatin compartments followed the approach of Wang et al. [16]. Tag directories were created using HOMER (version 5.1) [31], and duplicate read pairs were removed. To filter out reads that could artifactually represent contiguous genomic segments or result from re-ligation and self-ligation events, as well as those from high-density regions. Subsequently, the valid contact read pairs were processed to determine PC1 values for the active and inactive compartments using a 40 kb resolution, with an 80 kb window applied for smoothing or aggregating the data. A to B compartments were identified using the hicpca.pl function in Homer software, with significance calculated by getDiffExpression.pl. Significance was determined by an adjusted p-value < 0.05. Contiguous regions of interest were identified by stitching together individual regions exceeding significance threshold.

Processing of ATAC-seq dataset

The data were processed with the PEPATAC pipeline using default settings (version 0.11.3) [32]. Briefly, the reads were trimmed using Skewer (version 0.2.2) [33], and following the exclusion of reads that mapped to mitochondrial and mouse repeat regions, they were aligned to the mm10 mouse genome using Bowtie2 (version 2.5.4) [29]. PCR duplicates were removed, and peaks were called using MACS2 (version 2.2.9). Libraries with TSS enrichment scores below 6 million or fewer than 10 million aligned reads were excluded from further analysis. A consensus peak set was then created by merging peaks across all samples, and a peak count matrix was generated for each individual sample. Subsequently, differential analysis of the consensus peaks was conducted using DESeq2 (version 1.46) [34], and for peaks with significant differences, annotation was performed using Homer's annotatePeaks function.

Bulk RNA sequencing and data analysis

RNA extraction from whole lung tissues was conducted using the Qiagen RNeasy® Mini Kit (#74104), adhering to the manufacturer's recommended protocols. Following extraction, the RNA was enriched for poly-A-containing transcripts and subjected to sequencing of the entire mRNA fraction on the Illumina HiSeq 4000 platform. The next-generation sequencing reads from the bulk RNA of whole lung tissue were aligned to the mouse reference genome mm10 with the STAR aligner [35]. Read count summarization was conducted using the featureCounts tool [36], and the differential gene expression analysis of the bulk lung tissue samples was performed using the DESeq2 package in R (meeting the criteria $|\text{Log}_2\text{FC}| > 1$ and P value < 0.05) [34].

Enrichment analysis

Biological pathway enrichment analysis for gene lists was performed using EnrichR [37] accessed via the enrichr functionality within the gseapy package [38]. To assess the activation of hallmark pathways and fibrosis related pathway activations, hallmark and ontology gene sets were retrieved from the Molecular Signatures Database (MSigDB) version 7.2, provided by the Broad Institute [39]. Inference of TF activity from our RNA-seq data was performed using the decoupleR package (version 2.10.0) [40]. PROGENy model was used to pathway activity inference [41].

Integrated analysis of ATAC-seq and RNA-seq data

To assess the influence of chromatin accessibility on transcriptional activity, we conducted an integrative analysis of ATAC-seq and RNA-seq data. Differential log₂ fold changes in normalized read counts derived from ATAC-seq and RNA-seq experiments for each experimental condition were correlated by aligning the transcription start site (TSS)/promoter regions annotated from ATAC-seq peaks to their respective RNA-seq gene annotations. For subsequent comparative analysis, ATAC-seq peaks and RNA-seq genes with a statistical significance of adjusted p -value < 0.05 were graphically represented in relation to each other to elucidate the correlation between regions of differential chromatin accessibility and genes exhibiting differential expression.

Statistical analysis and data visualization

Bioinformatics analyses for this study were predominantly conducted using R software (version 4.2) and Python (version 3.9). The Wilcoxon rank-sum test was

employed as the default method to evaluate the significance of differences in paired comparisons, unless indicated otherwise. Statistical evaluations and graphical representations were performed using these computational tools. The Spearman correlation coefficient was utilized to assess correlations between two conditions. Analysis outcomes included the calculation of fold changes, p -values, and adjusted p -values. The following criteria were applied to denote statistical significance for all quantitative data: * p < 0.05, ** p < 0.01, *** p < 0.001, and **** p < 0.0001.

Result

Pathological characteristics of BLM-induced model of lung fibrosis

BLM-induced fibrosis in mouse models is widely recognized as the gold standard in preclinical studies due to its histological similarities to human lung fibrosis [11, 42]. Consequently, we have chosen to dissect the epigenomic and chromatin dynamics of this model to gain a deeper understanding of the molecular mechanisms underlying fibrotic progression. The histopathological assessment revealed that the BLM-treated mice exhibited marked alterations in lung architecture, characterized by increased collagen deposition and fibroblast proliferation, as evidenced by hematoxylin and eosin (H&E) and Masson's trichrome staining. These findings were further corroborated by microcomputed tomography (Micro-CT) imaging, which provided a three-dimensional visualization of the lung structure, clearly delineating the decrease of aerated lung compartments and increase of dense fibrotic tissue (Fig. 1A). Our results indicated a significant increase in hydroxyproline levels in the

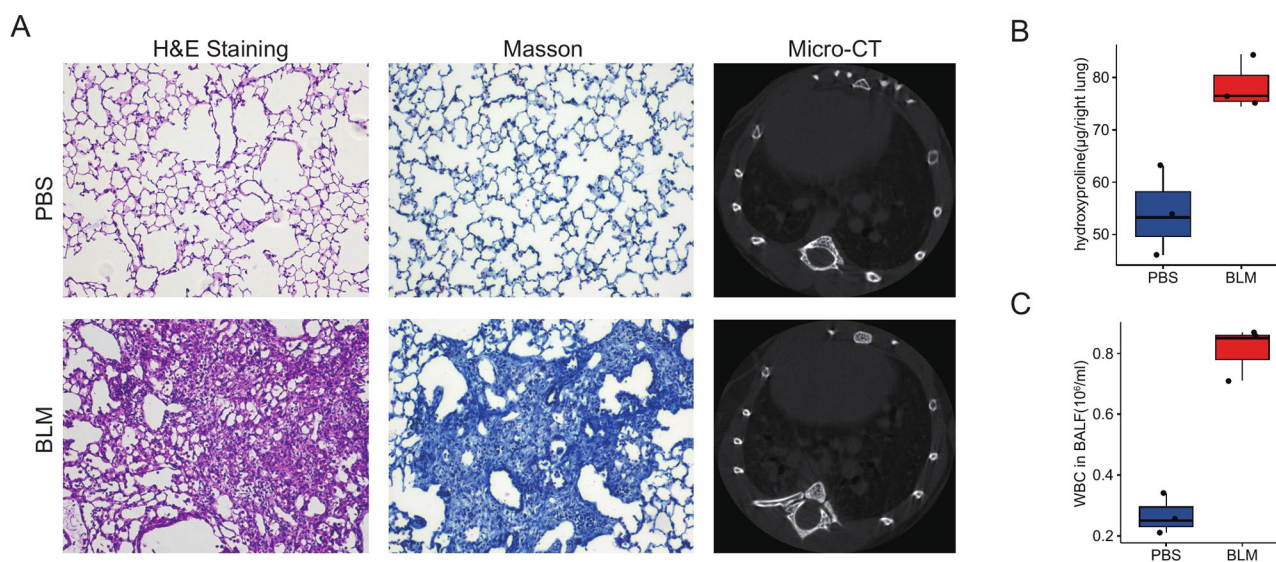


Fig. 1 Pathological characteristics of BLM-induced model of lung fibrosis. **(A)** Representative H&E, Masson-stained lung sections and micro-CT images at d21 after BLM. More collagen deposition/fibrosis in the pulmonary parenchyma was found in the BLM-treated mice. **(B)** Right lung hydroxyproline contents. **(C)** white blood cell count in BALF

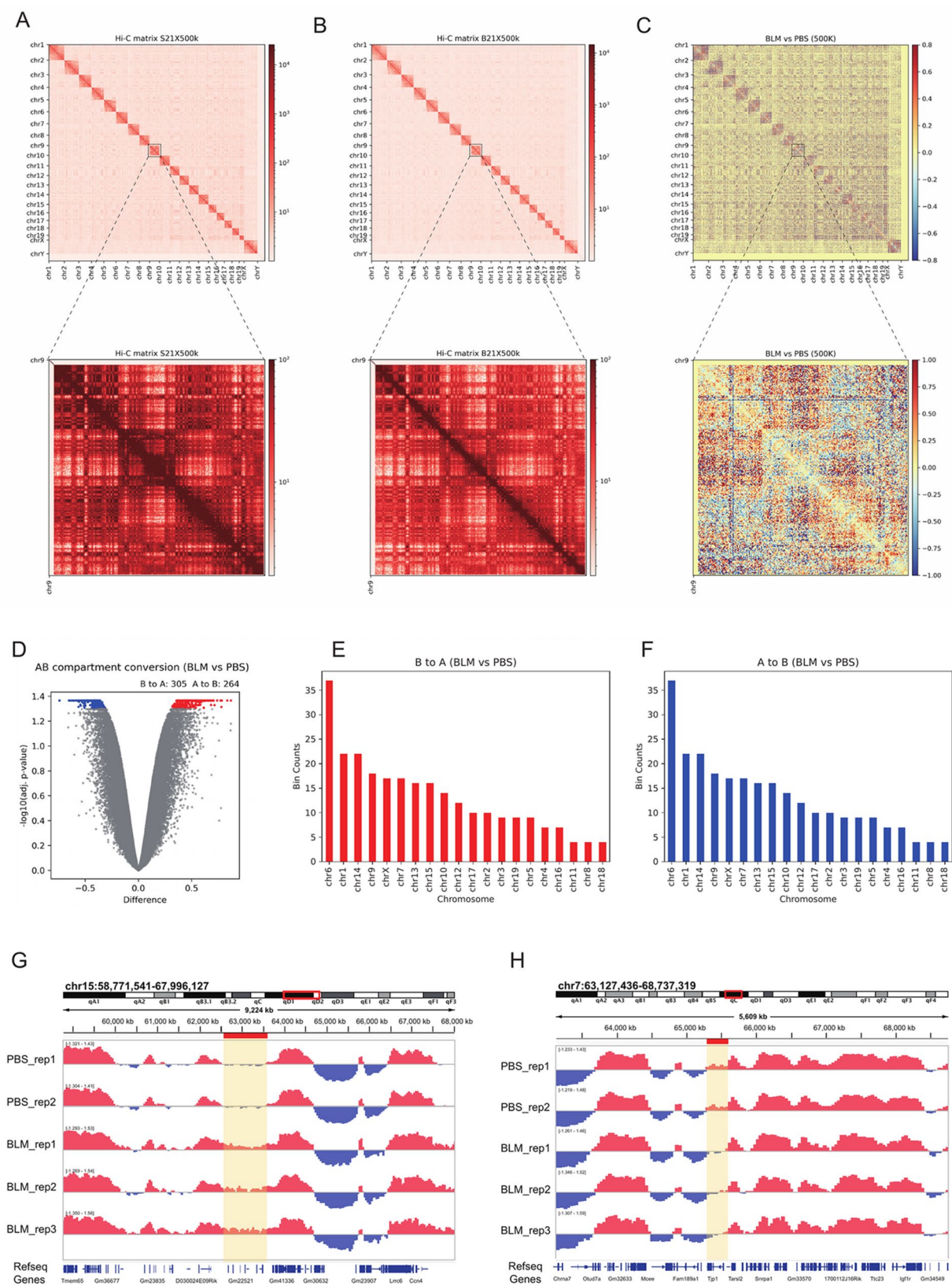


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 BLM-induced alterations in 3D chromatin architecture and compartmentalization in lung fibrosis. **(A,B)** Detailed heatmaps illustrating all-by-all chromatin interactions across the genome at a 500-kilobase resolution were produced for the control **(A)** and fibrotic **(B)** conditions. Chromosomes are displayed in numerical sequence from left to right and top to bottom, alongside a color scale representing the frequency of Hi-C interactions. The lower sections offer zoomed-in perspectives of chr9 at a 500-kb resolution, matching the upper sections. **(C)** The heatmap spans the entire genome, showcasing the log₂ ratio comparisons of Hi-C interaction frequencies at a 500-kilobase resolution in fibrosis as opposed to the parental control. The log₂ ratio values are indicated by the color gradient. The lower section of the image provides a zoomed-in perspective of chr9 at a 500-kb resolution, mirroring the area highlighted in the upper section. **(D)** The scatter plot showcases the transition rates of AB compartments between BLM and PBS treatments. The horizontal axis marks the discrepancies in conversion rates, and the vertical axis represents the log₁₀(adjusted p-value). The data points are distinguished by color, with blue indicating BLM and red signifying PBS. **(E, F)** Histograms illustrate the distribution of BLM (blue) and PBS (red) across individual chromosomes. Each bar indicates the occurrence frequency of BLM versus PBS for a specific chromosome, with the x-axis showing chromosome number and the y-axis depicting the number of instances. **(G-H)** A representative image from the Integrated Genome Browser (IGV) illustrates the compartmentalization of chromosome 15 or 7, with open A-type compartments in red and closed B-type compartments in blue. Regions demonstrating stable (light shading) and differential (dark shading) compartmentalization are highlighted for clarity

BLM-treated group compared to the controls, confirming the presence of excessive collagen accumulation (Fig. 1B). Furthermore, we assessed the inflammatory component of the disease by analyzing the proportion of inflammatory cells in the bronchoalveolar lavage (BAL) fluid. The findings revealed an elevated percentage of inflammatory cells, including neutrophils and macrophages, in the BAL of BLM-treated mice, suggesting an ongoing inflammatory response that contributes to the fibrotic process (Fig. 1C).

Mapping 3D chromatin conformation and identification of alternations in A/B compartment in BLM-induced PF

In order to investigate the alterations in chromatin architecture associated with BLM-induced lung fibrosis, Hi-C libraries were generated from three independent biological replicates of BLM-treated mice and two replicates of control mice. After stringent sequence quality filtering, approximately 119 million and 228 million valid contacts were obtained from the BLM-treated and control samples, respectively (Figure S1A). Within the Hi-C libraries, we observed no significant difference in the number of cis- and trans-interactions between the PBS and BLM groups. The ratio of cis- to trans-interactions indicated that approximately three-quarters of the interactions occurred within chromosomes, while the remaining quarter involved interactions between chromosomes (Figure S1B).

Heatmaps were utilized to represent the normalized Hi-C contact matrices, with each chromosome being compared to another. In these visualizations, a deeper shade of red denotes a greater frequency of interactions, whereas white regions imply a decreased frequency of such interactions (Fig. 2A-B). This analysis highlighted a pronounced prevalence of intra-chromosomal interactions over inter-chromosomal interactions (Fig. 2A-B and Figure S1B), echoing the established concept of “chromosome territories” observed in previous Hi-C data and chromosome imaging investigations [13]. The alternating pattern of high and low interaction frequencies across the heatmap suggests the existence of distinct chromatin compartments, likely corresponding to active

(A) and inactive (B) chromatin in both the control and fibrosis groups (Fig. 2A, B). This compartmentalization reflects the functional organization of the genome, with active regions being more transcriptionally active and accessible.

We identified differences in interaction frequency between specific chromosome loci, as evidenced by color changes in the heatmap from blue in the control group to light red in the fibrosis group, indicating closer proximity and stronger interactions between these loci in the fibrosis group (Fig. 2C). Notably, the intra-chromosomal interaction matrices displayed a clear shift in interaction patterns, particularly evident in chromosome 9, demonstrating that BLM treatment can indeed alter chromatin compartmentalization (Fig. 2C).

The analysis of interaction matrices at a genome-wide level, contrasting the PBS and BLM groups, revealed a spectrum of changes in the interactions and compartmentalization of chromosomes. To gain further insights into the compartment switching associated with BLM-induced fibrosis, we classified the genome into A and B compartments at 40-kb and 80-kb resolutions. Principal component analysis (PCA) revealed distinct differences in eigenvalues between control and fibrosis groups, with samples clustering into two separate groups (Figure S1C). A total of 305 genomic regions were identified that switched from A compartments in the control samples to B compartments in the fibrosis samples, with chromosome 15 showing the highest number of these transitions (Fig. 2D, E). Conversely, 264 genomic regions exhibited the opposite switch from B compartments in the PBS group to A compartments in the BLM group, with the highest number of switches occurring on chromosome 6 (Fig. 2D, F). We also identified 10 contiguous genomic regions that switched from A compartments to B compartments and 5 contiguous regions that switched from B compartments to A compartments (Fig. 2G, H).

Further analysis revealed that 172 genes (133 protein-coding genes) transitioned from A compartments to B compartments, while 196 genes (154 protein-coding genes) transitioned from B compartments to A compartments during BLM-induced lung fibrosis (Figure S1D).

Genes transitioning from B to A compartments were enriched in pathways related to stimulus response, signal transduction, amino sugar metabolism, immune system processes, and biological regulation (Figure S1E). Conversely, genes transitioning from A to B compartments were enriched in pathways related to cell adhesion and junctions, development and morphogenesis, signaling and regulation, cell regulation and differentiation, and vasculature development (Figure S1E).

ATAC-seq identifies differential chromatin accessibility in lung fibrosis

Chromatin remodeling plays a crucial role in coordinating gene transcription, prompting us to analyze chromatin accessibility in lung tissues from control and BLM-induced lung fibrosis mice using ATAC-seq. We obtained a total of 900,292,308 raw reads, which were filtered to yield 900,223,486 cleaned reads. Notably, >82% of these reads mapped to the mouse reference genome (Figure S2A-C, Table S1). The libraries presented multimodal fragment length distributions, including both nucleosome-free and mononucleosome fragments, a signature of open chromatin (Figure S2D, Table S1). We identified 97,344 ATAC-seq peaks from the samples, primarily enriched 3 kb upstream and downstream of the transcription initiation site, indicating that open chromatin regions contribute to transcriptional regulation (Figure S2E, F). Additionally, 34,276 ATAC-seq consensus peaks were identified and distributed across all chromosomes (Fig. 3A). The genome-wide distribution of the consensus peaks revealed a predominant localization in promoter and intron regions (Fig. 3B), suggesting potential involvement in gene expression regulation through interaction with transcription start sites or long-range regulatory elements.

After performing dimension reduction using PCA, we observed the first principal component separated BLM and PBS samples (Fig. 3B-C), indicating data suitability for bioinformatic analysis. This analysis identified 1300 differentially accessible chromatin regions (DARs) with adjusted p -value < 0.05. Among these, 48% (620) exhibited reduced accessibility, while 52% (680) displayed increased accessibility in fibrosis compared to control (Fig. 3E, S2G). Strikingly, 65% of DARs were located within intragenic and intergenic regions, with approximately 35% found within promoters or untranslated regions (UTRs) (Fig. 3E, S2H). This suggests that these specific open regions may regulate gene expression at various stages by influencing either the gene body or promoter region. For example, *Mmp19* exhibited significantly increased peak intensity in BLM group, with chromatin accessibility within its promoter region playing a crucial role in fibrogenesis [43] (Fig. 3G). In contrast, *Bmp4*, which promotes epithelial cell differentiation

[44], showed a specific peak in control, with significant changes in chromatin accessibility within its gene body (Fig. 3H).

Pathway enrichment analysis of the 1300 DAR-related genes revealed that BLM treatment significantly upregulated chromatin accessibility of genes involved in TNF- α signaling via NF- κ B, inflammatory response, IL-2/STAT5 signaling, cytokine production, p53 pathway, hypoxia, adipogenesis, and apoptosis (Fig. 3I). Conversely, BLM downregulated chromatin accessibility of genes related to signal transduction, pluripotency of stem cells, receptor Ser/Thr kinase signaling, RAC1 GTPase cycle, signaling by NOTCH1, TGF- β signaling pathway, signaling by RTK, integration of energy metabolism, and Wnt signaling pathway (Fig. 3J).

To identify potential lung TFs of fibrosis-associated chromatin alterations, we performed motif enrichment analysis on the ATAC-seq datasets using Homer. This analysis revealed significant enrichment of motifs associated with AP-1 family members (*Fosl2* and *Atf1*), interferon regulatory factor 4 (*IRF1*), nuclear factor kappa B (NF- κ B), and Spi-B Transcription Factor (*Spi-1/PU.1* Related/*SpiB*) in the BLM group (Fig. 3K). Conversely, motifs associated with TFs such as *Sp1*, *NFYC*, *Foxj2*, *NFY*, *ZNF263*, and *ZNF711* were significantly enriched in the control group (Fig. 3L).

RNA-seq analysis identifies transcriptional signatures in the fibrotic lungs

To investigate gene expression patterns associated with BLM-induced lung fibrosis, we performed RNA-seq analysis on lung tissues from control and fibrosis groups at two time points, with three biological replicates per group. After processing, 668 million cleaned reads were retained, with over 94% achieving a Q30 quality score, indicating high data quality. The RNA-seq data shows that 95.4–96.9% of reads were aligned to the mm10 reference genome, as detailed in Table S2. PCA revealed distinct clustering of samples from the two groups, indicating significant differences in gene expression patterns (Fig. 4A). Correlation analysis further confirmed high similarity within each group (Fig. 4B). Using a general threshold of $|\log_2 \text{Fold change}| \geq 1$ and adjusted p -value < 0.05, we identified 926 differentially expressed genes (DEGs), including 573 up-regulated genes and 353 down-regulated genes (Fig. 4C).

Our pathway enrichment analysis showed BLM treatment upregulates immune-related pathways, including “Neutrophil Degranulation” and “Innate Immune System,” suggesting immune response disruption. It also induces “Epithelial Mesenchymal Transition” and impacts “Extracellular Matrix Organization,” affecting cell behavior and tissue homeostasis (Fig. 4D). BLM treatment reduced pathways like “Salivary secretion” and

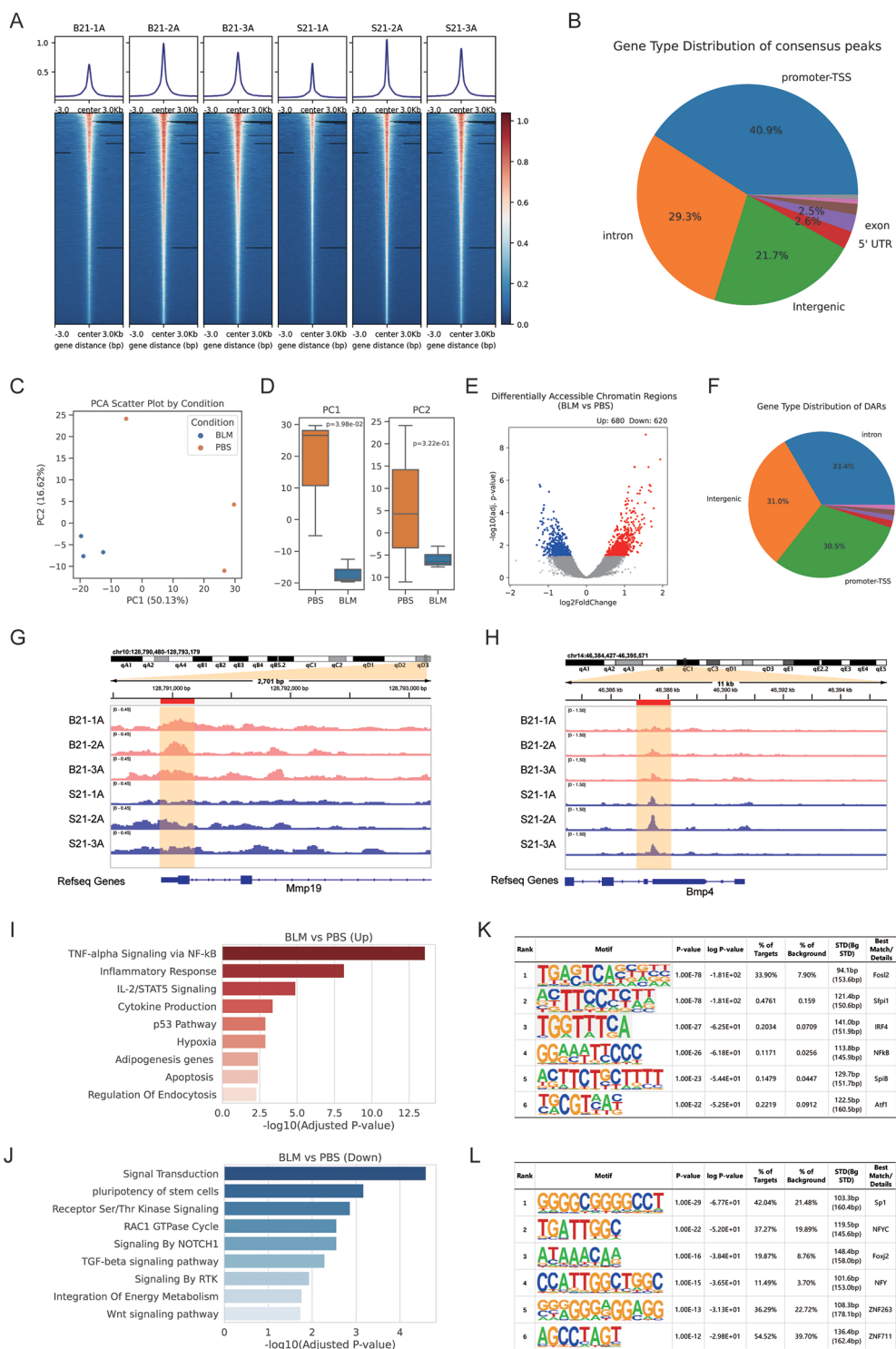


Fig. 3 The landscape of chromatin accessibility in normal and fibrotic mouse lung. **(A)** Genome-wide ATAC-seq consensus peaks in lung tissues. Each column represents one peak. The color represents the intensity of chromatin accessibility (red/more accessible and blue/less accessible). **(B)** Genomic distribution of consensus peaks across all samples. **(C)** Principal component analysis (PCA) of accessible loci as determined by ATAC-seq from control and fibrotic mice, either Sham operated and 21 days after intratracheal BLM delivery. Each dot represents an individual mouse. **(D)** Boxplots display principal component two across PBS samples (orange) and BLM samples (blue) for the study. **(E)** Volcano plot, visualizing the statistical difference of the differentially accessible regions. **(F)** Genomic distribution of differentially accessible regions between control and BLM-induced fibrotic mice. **(G-H)** Visualization of chromatin accessibility region of Mmp19 **(G)** and Bmp4 **(H)** in control and BLM-induced fibrotic mice. **(I-J)** Pathway enrichment analysis of differentially accessible regions between control and BLM-induced fibrotic mice. **(K-L)** De novo DNA motif analysis of differentially accessible chromatin sites between control and BLM-induced fibrotic mice

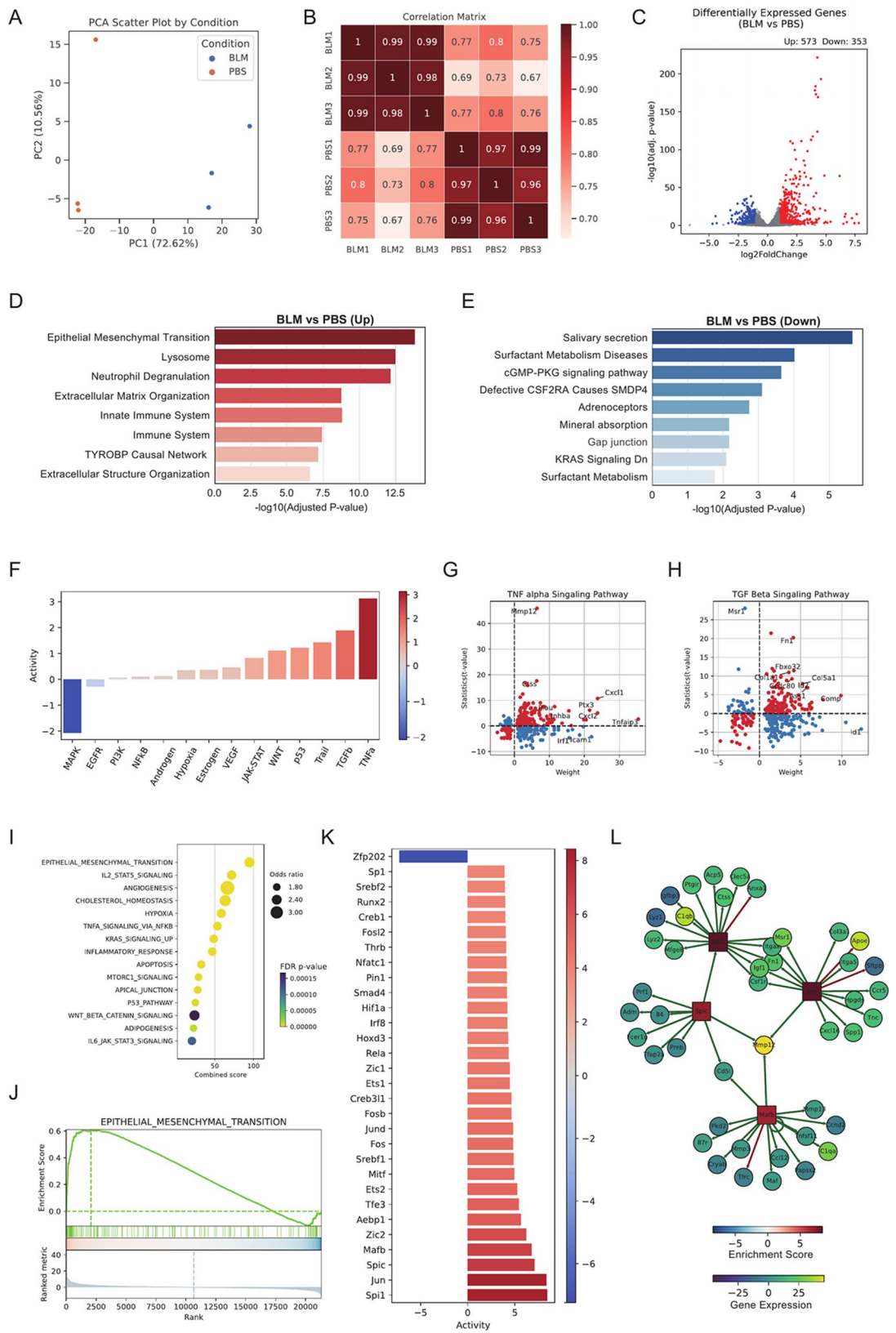


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 RNA-seq analysis identifies unique transcriptional signatures in lungs of BLM-induced fibrotic mice. **(A)** Principal components analysis (PCA) of RNA-seq data from control and BLM-induced fibrotic mice. **(B)** Spearman correlation heatmap across all samples. **(C)** Volcano plot, visualizing the statistical difference of the differentially expression regions. **(D-E)** Pathway enrichment analysis of differentially expression regions between control and BLM-induced fibrotic mice. **(F)** Pathway activity inference of differentially expression regions between control and BLM-induced fibrotic mice. The observed activation of TNF α **(G)** and TGF β **(H)** is due to the fact that majority of its target genes with positive weights have positive t-values (1st quadrant), and the majority of the ones with negative weights have negative t-values (3d quadrant). **(I)** The dotplot illustrates the combined score and corresponding odds ratio for various signaling pathways, with a color gradient indicating the adjusted p-value. Yellow dots represent higher scores and odds ratios, while black dots indicate lower scores and ratios. **(J)** This GSEA plot presents a gene set enrichment analysis for the term “EPITHELIAL_MESENCHYMAL_TRANSITION,” focusing on the enrichment score and ranked metric across different ranks. The enrichment score shows a significant peak at the beginning, indicating high enrichment. **(K)** The barplot illustrates the activity levels of transcript factor in response to BLM, with each column representing a different transcript factor and each row indicating the activity level. Higher activity level is represented by red, while lower activity level is depicted in blue. **(L)** The TF-Gene co-regulatory network for *Spi1*, *Spic*, *Jun* and *Mafb*

“Surfactant Metabolism Diseases,” indicating suppression of secretion and metabolic functions. Alterations in pathways like “cGMP-PKG signaling” highlight the intricate effects of BLM on cellular signaling (Fig. 4E).

To enhance the robustness of pathway enrichment analyses, we applied a multivariate linear model (MLM) incorporating pathway-specific gene weights from the PROGENy database [41]. We observed the activation of several immunologically relevant pathways in response to BLM treatment, including the TNF α , NF κ B, and JAK-STAT signaling pathways (Fig. 4F). Within the TNF signaling cascade, key effectors such as TNF Alpha Induced Proteins (*Tnfaip3*), chemokines (*Cxcl1* and *Cxcl2*), and proteases (*Mmp12* and *Ctss*) exhibited significant upregulation following BLM exposure (Fig. 4G). The TGF β pathway was substantially activated, leading to a marked increase in the expression of extracellular matrix genes, such as *Fn1* and *Col1a1* (Fig. 4H). In addition to validating the results obtained from traditional pathway enrichment analysis, the MLM further expanded our understanding of the molecular pathways implicated in BLM-induced lung fibrosis. Notably, the WNT pathway, which is known to regulate stem cell proliferation, as well as the VEGF and hypoxia pathways, which are crucial for angiogenesis, were identified as prominently activated in response to BLM treatment (Fig. 4F).

To further investigate the functional implications of the DEGs, we performed gene set enrichment analysis using the Hallmark gene sets from the MsigDB database [39]. This analysis revealed significant enrichment of pathways related to epithelial-to-mesenchymal transition (Fig. 4I-J), JAK-STAT signaling, angiogenesis, hypoxia, TNF α signaling, inflammatory response, apoptosis, and WNT signaling (Fig. 4I). These findings were in concordance with the results of our pathway enrichment analysis.

Transcriptome data can also be utilized to infer TF activity. Using the univariate linear model in the CollecTRI network [45], we predicted the enrichment scores of TFs and identified the top 30 active TFs, with 29 of them significantly activated following BLM treatment (Fig. 4K). Notably, the TFs exhibiting significant activation were consistent with those identified by motif enrichment

analysis of ATAC-seq data, further validating our findings. The activated TFs included members of the AP-1 family (*Jun*, *Fos*, *Jund*, *Fosb*, *Fosl2*, etc.), *Spi1* (*Sfpi1*), interferon regulatory factors (*Irf8*), PU.1-related factors (*Spic*), members of the Maf TF family (*Mafb*), ETS TF family members (*Ets1*, *Ets2*), hypoxia-associated factors (*Hif1a*), and TGF β -associated factors (*Smad4*), among others (Fig. 4K). The regulatory interactions network revealed that *Spi1* and *Jun* collaborate to regulate the expression of *Fn1* and promote the expression of *Csf1r* in monocytes and *Itgax*, a marker of interstitial macrophages, suggesting the generation of monocyte-derived macrophages. Additionally, *Mafb*, *Jun*, and *Spic* collectively enhance the gene expression of *Mmp12*, with *Jun* regulating *Col3a1* and *Mafb* regulating *Mmp3* and *Mmp13*, highlighting the crucial role of matrix remodeling in the context of BLM-induced lung fibrosis (Fig. 4L).

Integrated analysis of gene expression, chromatin accessibility and chromosome conformation data

While the switch from A to B compartment in certain genomic regions was evident following BLM treatment, this compartmental reorganization did not correlate significantly with ATAC-seq measured tag density (Fig. 5A-B). The analysis of transcriptomes showed that the majority of DEGs within genomic regions shifted from the B to A chromatin compartment were predominantly upregulated, with significant higher log₂ fold changes. Conversely, DEGs in genomic regions changed from A to B were mainly downregulated (Fig. 5C). Additionally, the expression of DEGs within compartments undergoing transition was observed to be positively correlated with the direction of the switch. In fibrosis, gene expression levels in compartments transitioning from state B to A were significantly higher in the BLM group compared to the PBS, with a median increase of 0.31 (BLM: 9.64, PBS: 9.33) and a mean increase of 0.22 (BLM: 8.50, PBS: 8.28). Conversely, for regions transitioning from state A to B, the BLM group exhibited lower expression levels than PBS, with a median decrease of 0.01 (BLM: 8.09, PBS: 8.10) and a mean decrease of 0.20 (BLM: 7.45, PBS: 7.65).

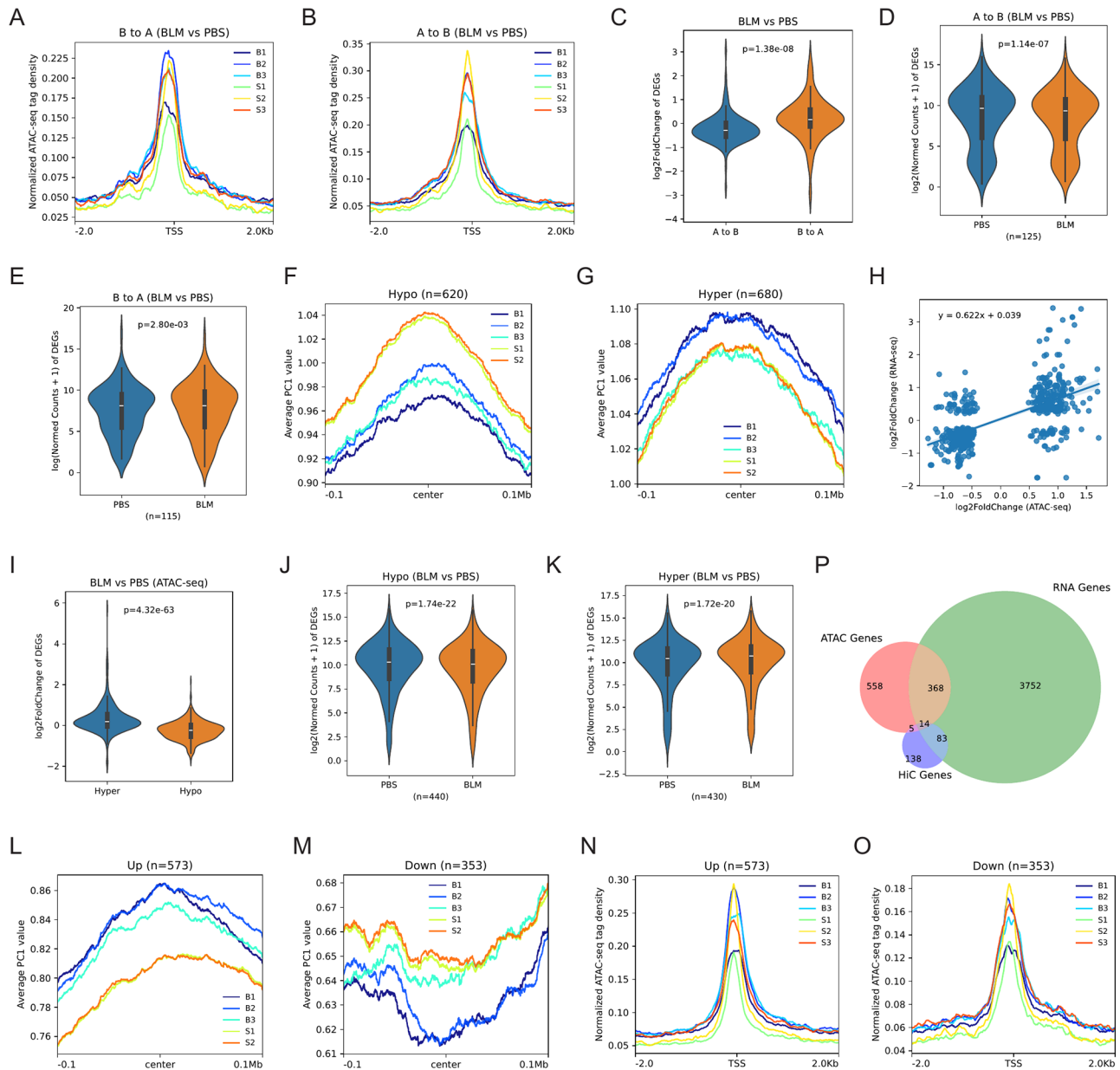


Fig. 5 Integrated analysis of gene expression, chromatin accessibility and chromosome conformation data. The profiles showing the normalized ATAC-seq tag density surrounding the center of the regions of A to B (**A**) and B to A (**B**) across a genomic window of ± 2 kb in mice with BLM-induced fibrosis, contrasted with the control. (**C**) The violin plot represents the log₂ fold change (FC) between fibrosis and control for DEGs situated in regions associated with diverse compartmental switch categories. (**D-E**) Violin plot shows the expression levels [$\log_2(\text{FPKM} + 1)$] of DEGs found in regions associated with different compartmental switch categories. (**F-G**) Profiles illustrate the pattern of average PC1 values, as determined by the HOMER software (version 4.10), in the vicinity of the central areas of hyper- and hypo-accessible regions across a genomic span of ± 100 kb in BLM-induced fibrotic mice, contrasted with the control. (**H**) Scatter plot shows the correlation between the accessibility of chromatin and the expression of genes. ATAC-seq peaks with an adjusted p-value < 0.05 and genes found by RNA-seq with an adjusted p-value < 0.05 were plotted to analyze the variations in accessible peaks and gene expression. (**I**) Violin plot shows the log₂ fold change (FC) between fibrosis and control for DEGs situated in regions linked to distinct accessible peaks. (**J-K**) Violin plot shows expression level [$\log_2(\text{FPKM} + 1)$] of DEGs residing at regions for different accessible peaks. (**L-M**) Profiles illustrate the distribution of average PC1 values surrounding the center of upregulated or downregulated genes across a genomic window of ± 100 kb in BLM-induced fibrotic mice compared to control. (**N-O**) The profiles present the normalized ATAC-seq tag density surrounding the center of the DEGs across a genomic window of ± 2 kb in BLM-induced fibrotic mice compared to control. (**P**) Venn diagram illustrates the distribution of gene categories in three different omics datasets. The overlapping areas represent genes common to multiple omics datasets, while distinct areas indicate unique genes within each dataset

Our research exhibited that the mean PC1 values adjacent to regions of heightened chromatin accessibility were significantly elevated in the fibrosis group compared to the control group. Conversely, for regions with decreased chromatin accessibility, the average PC1 values were lower in fibrosis compared to the control (Fig. 5F-G). We investigated whether changes in gene expression during fibrosis were directly caused by changes in chromatin structure that could impact transcription. By comparing gene expression (using RNA-seq with a cutoff of adjusted $p < 0.05$) with chromatin accessibility (using ATAC-seq with a cutoff of adjusted $p < 0.05$), we found that there was a moderate correlation between the changes in chromatin accessibility and gene expression (with a Spearman correlation coefficient of $R = 0.63$) (Fig. 5H). It demonstrated that the majority of DEGs located in hyper-accessible regions were up-regulated. Conversely, DEGs situated in hypo-accessible regions were mostly down-regulated (Fig. 5I). Additionally, the analysis revealed that the pattern of gene expression alterations within the DARs aligned positively with the direction of accessibility changes. Genes within the more open regions in fibrosis exhibited a marked elevation in expression relative to the control, with a reciprocal pattern observed in less open regions (Fig. 5J-K). The findings indicated a positive correlation between the level of chromatin accessibility and the expression of the related genes ($p\text{-value} = 2.85\text{e-}53$).

We further explored the chromatin structural activity and accessibility surrounding DEGs in response to BLM treatment. Intriguingly, among the 573 genes upregulated by BLM, the average PC1 value in the BLM group was significantly higher than that in the saline group (Fig. 5L), suggesting enhanced structural activity in these regions. Conversely, the average PC1 value was significantly lower in the BLM group compared to the saline group for the 353 downregulated genes (Fig. 5M), indicating reduced structural activity. However, ATAC-seq tag density, a measure of chromatin accessibility, did not show a consistent trend of difference between the BLM and saline groups for these DEGs (Fig. 5N-O).

To further delineate the interplay between chromatin compartmentalization, accessibility, and gene expression in the context of lung fibrosis, we sought to identify genes with coordinated alterations in all three aspects. By intersecting DARs and DEGs with regions exhibiting compartment switching, we identified 14 genes with robust multi-omics support (Fig. 5P; Table 1). These genes represent potential regulatory hubs in BLM-induced lung fibrosis.

Cell-specific expression profiles of 14 genes identified by multi-omics

To link the identified genes to specific cell types and pathogenic mechanisms, we re-analyzed scRNA-seq data from control and fibrotic mouse lungs [48]. Quality metrics, including unique molecular identifier (UMI) counts,

Table 1 The 14 genes with concordant multi-omics support at the levels of differential expression, chromatin accessibility and chromatin structure

Ensembl	Symbol	HiC-seq	ATAC-seq		RNA-seq		Description
		Compartment switch	log2FC	adjusted p-value	log2FC	adjusted p-value	
ENSMUSG00000039607	Rbms3	A to B	-1.023	0.013	-0.817	3.80E-07	RNA binding motif, single stranded interacting protein
ENSMUSG00000054263	Lifr	A to B	-0.924	0.002	-0.453	0.0005953	LIF receptor alpha
ENSMUSG00000030020	Prickle2	A to B	-0.818	0.008	-0.709	9.37E-10	prickle planar cell polarity protein 2
ENSMUSG00000045095	Magi1	A to B	-0.767	0.046	-0.579	0.0002935	membrane associated guanylate kinase, WW and PDZ domain containing 1
ENSMUSG00000022656	Nectin3	A to B	-0.723	0.001	-0.718	5.32E-05	nectin cell adhesion molecule 3
ENSMUSG00000052534	Pbx1	A to B	-0.655	0.010	-0.501	3.21E-06	pre B cell leukemia homeobox 1
ENSMUSG00000025576	Rbfox3	B to A	-1.041	0.017	-1.047	1.70E-06	RNA binding protein, fox-1 homolog (C. elegans) 3
ENSMUSG00000032265	Tent5a	B to A	0.799	0.016	0.567	1.21E-05	terminal nucleotidyltransferase 5 A
ENSMUSG00000032265	Tent5a	B to A	0.679	0.035	0.567	1.21E-05	terminal nucleotidyltransferase 5 A
ENSMUSG00000032265	Tent5a	B to A	0.614	0.042	0.567	1.21E-05	terminal nucleotidyltransferase 5 A
ENSMUSG00000030657	Xylt1	B to A	0.674	0.044	1.089	2.38E-16	xylosyltransferase 1
ENSMUSG00000040537	Adam22	B to A	0.691	0.024	-0.516	0.0023677	a disintegrin and metallopeptidase domain 22
ENSMUSG00000040537	Adam22	B to A	0.757	0.031	-0.516	0.0023677	a disintegrin and metallopeptidase domain 22
ENSMUSG00000074802	Gas2l3	B to A	0.808	0.022	1.295	2.92E-13	growth arrest-specific 2 like 3
ENSMUSG00000055435	Maf	B to A	0.883	0.044	0.464	5.78E-06	MAF bZIP transcription factor
ENSMUSG00000027009	Itga4	B to A	1.048	0.002	0.304	0.018377	integrin alpha 4
ENSMUSG00000003882	Il7r	B to A	1.241	0.003	0.699	4.47E-08	interleukin 7 receptor

detected genes per cell, and reads aligned to the mouse genome, were comparable across samples. However, the initial Uniform Manifold Approximation and Projection (UMAP) map revealed batch effects, with samples clustering by mouse rather than cell type (Figure S3A). After accounting for these batch effects, unsupervised clustering identified 23 distinct clusters corresponding to 20 major cell types, encompassing epithelial, endothelial, mesenchymal, and leukocyte lineages (Fig. 6A–B, S3B). Three clusters with fewer than 100 cells were excluded from further analysis. Additionally, an ambient RNA contamination subset and a small hematopoietic contamination subset were identified and excluded (Figure S3C–D). Notably, even rare cell types, such as lymphatic endothelial cells (152 cells), were identified, highlighting the robustness and accuracy of our computational workflow. Differential gene expression analysis was then used to identify cell type-specific marker genes with high expression levels within each cluster (Fig. 6C, Supplementary Data 1). These clusters were annotated based on known marker genes and literature, providing a comprehensive atlas of cell types in the mouse lung.

Within the mouse lung single-cell atlas that we developed, the expression patterns of 14 genes, ascertained through the integration of differential expression, chromatin accessibility, and chromatin compartmentalization data, revealed distinct cell type-specific biases (Fig. 6D). Upon correlating our gene dataset with the lung cell information from the newly released Human Lung Cell Atlas (HLCA) [49], we observed a high degree of concordance in the cell-type specificity of gene expression (Fig. 6E). For example, *Pbx1*, a homeodomain TF essential for *Fgf10* expression in lung mesenchyme, is preferentially expressed in fibroblasts [50] (Fig. 6F); *Lifr*, implicated in tumor angiogenesis [51], and *Prickle2* are both highly expressed in endothelial cells (Fig. 6G, S3E); *Il7r* is predominantly found in T lymphocytes; *Rbms3* shows high expression in non-immune cell types (Fig. 6H); *Magi1* is highly expressed in lung parenchymal cells (Figure S3E); *Nectin3* is enriched in AT1 and aCap cells (Figure S3E), underscoring its role in governing the air-blood barrier architecture; *Gas2l3* is primarily expressed at high levels in macrophages (Figure S3E); *Maf* is highly expressed in lymphatic endothelial cells (Figure S3E); and *Itga4* is highly expressed across immune cell populations (Figure S3E). Nevertheless, for four genes—*Rbfox3*, *Tent5A*, *Xylt1*, and *Adam22*—expression levels were insufficiently robust in both the human and mouse atlases to evaluate the consistency of their expression profiles (Fig. 6D–E).

Discussion

The ever-expanding knowledge of the molecular underpinnings of preclinical model of PF has yet to translate into substantial clinical improvements. The emergence of Hi-C analysis opens up innovative avenues for exploring the complexities of chromatin organization and its regulatory functions within the genome [52]. The ability of chromatin to be accessed plays a crucial role in both turning on genes and enabling physical interactions [53]. The ATAC-seq technique, known for its efficiency in detecting chromatin accessibility, has been broadly utilized to chart the landscape of chromatin accessibility in a wide variety of tissues and cellular contexts [15]. In this study, high-resolution three-dimensional genome architecture was delineated by integrating Hi-C, ATAC-seq, and mRNA-seq data from BLM-treated fibrotic mouse lung tissues.

As observed in breast cancer cells [16], changes in chromatin accessibility in lung tissue typically led to alterations in chromatin compartmentalization and gene expression. However, our further investigation reveals that the genes identified as differential compartment and those with differential expression do not always correspond directly with the changes in accessibility. This suggests a directed flow in regulatory pathways, where changes in chromatin accessibility may serve as an upstream regulatory event, driving downstream responses such as alterations in AB compartments and gene expression. To illustrate, while the throwing of a stone into a lake causes ripples, the presence of ripples alone does not confirm the stone's impact. It is important to note that the topologically associating domains (TAD) and Chromatin loops, which represent a more refined chromatin structure as indicated by Hi-C data, may influence chromatin accessibility, as previously reported in the literature [16]. Our Hi-C analysis was limited to the level of AB compartment analysis due to the complexity introduced by tissue-level heterogeneity. More detailed analyses of TADs and Loops should be conducted with more specific samples and deeper sequencing to provide a clearer understanding.

Tissue damage and inflammation are important triggers for regeneration and fibrosis [54]. We observed that genomic regions linked to the inflammatory response became activated in lung tissues affected by BLM. Inflammation and cytokine production were also prominent in regions made hyper-accessible by BLM induction and in genes that were upregulated. TNF- α , a cytokine involved in the Th-1 response, has been shown to contribute to fibrosis in various animal studies, as indicated by the use of TNF- α inhibitors or mice lacking TNF receptors [55]. Our findings also showed that BLM triggers the activation of TNF- α signaling at both chromatin accessibility and gene expression levels. This finding is in line

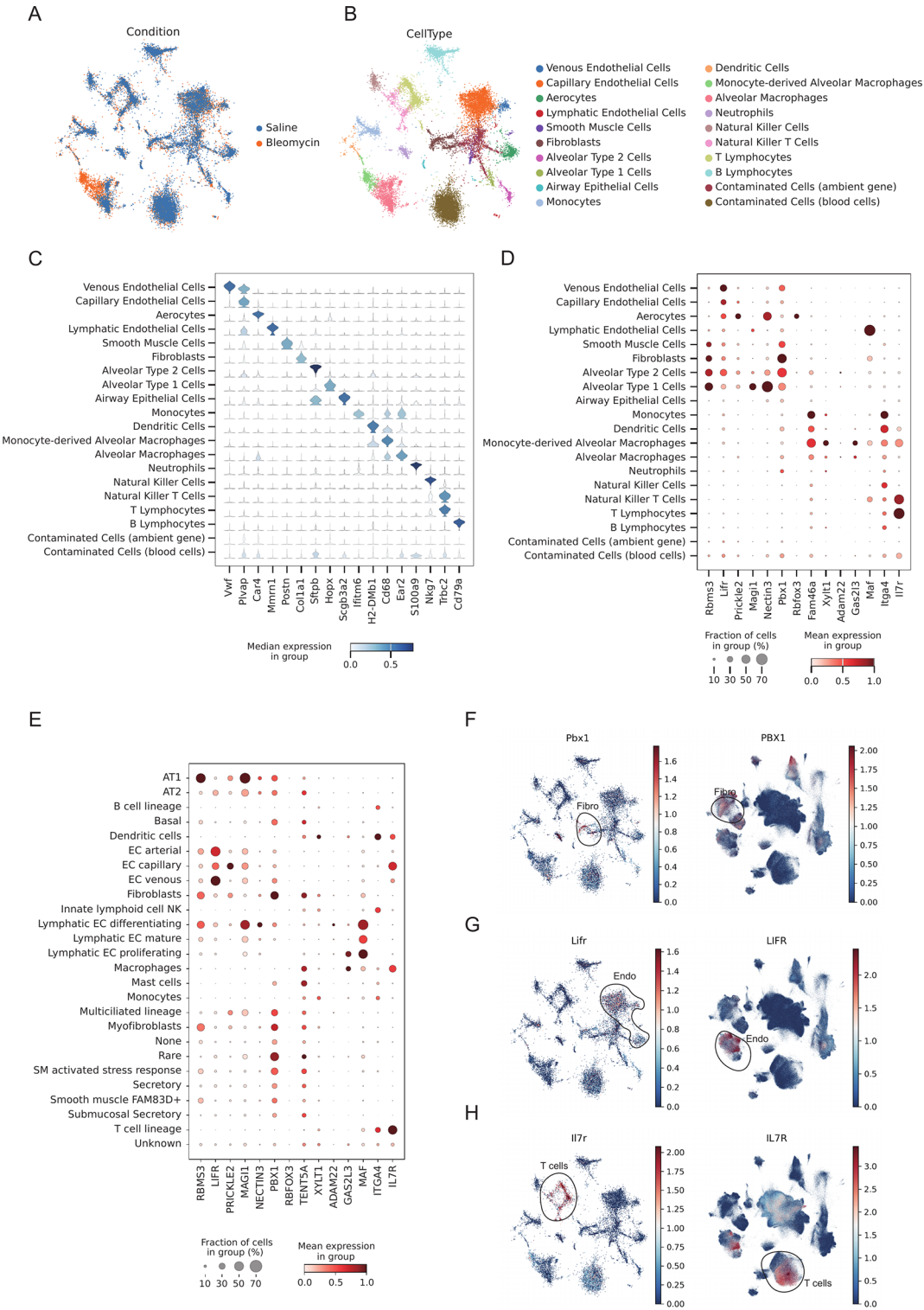


Fig. 6 Identifying pathogenic cell types for epigenetics and Chromatin Dynamics of lung fibrosis. **(A)** The UMAP plot illustrates the distribution of cells under two distinct conditions: saline and BLM. The blue dots represent cells treated with saline, while the orange dots indicate those treated with BLM. **(B)** The UMAP plot of various cell types, each represented by distinct colors. The legend on the right side lists these cell types and their corresponding color codes. **(C)** Violin plot shows the expression levels of the identified marker genes relative to each other. **(D)** Bubble plot reveals how 14 specific genes are expressed across various cell types, based on mouse single-cell RNA sequencing information. **(E)** Bubble plot shows a visual representation of how 14 genes are expressed in different cell types, as documented in the human lung cell atlas. The arrangement features cell types along the rows and genes along the columns. The bubble size is proportional to the number of cells expressing the gene, and the color gradient indicates the degree of gene expression. The UMAP plot showing cell specificity of PBX **(F)**, LIFR **(G)** and IL7R **(H)** expression in human and mouse

with the understanding that interferon- γ (IFN- γ), which is known to promote the differentiation of monocyte-derived macrophages, is a key player in lung inflammation and fibrosis. This is demonstrated by the reduction in lung inflammation and fibrosis in mice where IFN- γ was neutralized after they experienced an acute SARS-CoV-2 infection [56]. The analyses of motif enrichment and TF target gene enrichment both highlight the significant enrichment of interferon regulatory factors, which underscores the importance of the interferon pathway in the development of BLM-induced lung fibrosis.

Our data not only captured the changes in immune characteristics post-BLM treatment but also the alterations in signaling pathways related to tissue regeneration. A significant alteration in the chromatin compartment, linked to organogenesis, is evident following BLM treatment, characterized by a transition from active to inactive chromatin states. This shift suggests a hindering effect on the development and regeneration of alveolar structures [57], processes usually governed by the balance between Wnt and BMP signaling pathways [58]. Furthermore, our ATAC-seq data indicates a marked suppression of BMP signaling following BLM exposure, while Wnt and TGF β signaling-related genes show no significant changes in chromatin accessibility. Interestingly, RNA-seq analysis reveals a significant activation of both Wnt and TGF β signaling pathways. This discrepancy points to the existence of chromatin-independent regulatory mechanisms in BLM-induced fibrosis, potentially involving RNA or protein level regulation that enhances the responsiveness and speed of these signaling pathways to BLM treatment.

An integrated analysis of chromatin compartmentalization, accessibility, and gene expression data identified 14 key regulatory genes. These genes are differentially expressed across various cell types, with the expression profiles widely distributed among them. Specifically, *Il7r*, *Gas2l3*, and *Itga4* are expressed in immune cells; *Magi1* and *Nectin3* in epithelial cells; *Lifr*, *Prickle2*, and *Maf* in endothelial cells; and *Pbx1* and *Rbms3* in mesenchymal cells. Some of these genes have been extensively studied. For instance, *RBMS3* is known to play a role in various aspects of lung pathology in conjunction with *WNT5A* and *MSI2* [47]. The leukemia inhibitory factor receptor (*LIFR*) has been observed to enhance the pathogenic activation of fibrotic fibroblasts [46], although single-cell data suggests that *LIFR* is primarily expressed in endothelial cells and only weakly in fibroblasts, indicating that its regulatory role in endothelial cells may be a more intriguing area of study. Terminal Nucleotidyltransferase 5 A (*TENT5A*), an mRNA poly(A) polymerase, directly affects the conversion of *COL1A1* mRNA [59]. Xylosyltransferase 1 (*XYLT1*) is a key enzyme in the synthesis of proteoglycans and the construction of the extracellular matrix (ECM) [60].

This study provides fundamental information on the epigenomics and chromatin dynamics of BLM-induced fibrosis, albeit with certain limitations that warrant further investigation in the future. The choice to use tissue samples as the subject of study offers only preliminary insights, making it difficult to assess the extent to which cellular composition may have influenced these findings. Additionally, this approach limits the possibility of conducting more detailed analyses of specific cell type phenotypes. Another limitation of this research is that, unlike mature RNA-seq protocols, the quality control standards for Hi-C and ATAC-seq are more challenging to establish, particularly at the tissue level. Future studies should aim to build upon the basic information provided here by selecting samples with reduced heterogeneity, such as individual cells, for more refined sequencing analyses and phenotypic validations.

Conclusion

The study enhances insights into the epigenetic mechanisms and chromatin dynamics underlying BLM-induced PF, offering critical benchmarks for animal model preclinical studies and steering subsequent research endeavors.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13072-025-00579-5>.

Supplementary Material 1: Supplementary Table 1. ATAC-seq data statistics.

Supplementary Material 2: Supplementary Table 2. RNA-seq data statistics

Supplementary Material 3: Supplementary fig. 1 (A) The bar graphs present the distribution of read counts across different tag types for various samples. (B) The histogram presents the distribution of valid pairs across different categories, including duplicates, valid interactions, cis long-range contacts (> 20 kb), cis short-range contacts (< 20 kb), and trans contacts. (C) Principal components analysis (PCA) of PC1 values of Hi-C data from control and BLM-induced fibrotic mice. (D) Gene type distribution of AB compartment conversion regions between control and BLM-induced fibrotic mice. (E) Pathway enrichment analysis of AB compartment conversion regions between control and BLM-induced fibrotic mice.

Supplementary Material 4: Supplementary fig. 2 (A-B) This bar chart illustrates the distribution of reads across four categories: mm10, duplicates, mouse_chrM2x, and unaligned. Samples B21-1, B21-2, B21-3, S21-1, S21-2, and S21-3 are represented with varying percentages in each category. (C) The graph presents the cumulative distribution of unique reads across different samples, with each sample represented by a distinct line. The x-axis shows the total number of reads in millions, while the y-axis indicates the percentage of unique reads. The inset graph zooms into the lower read count range to highlight the detailed trend. (D) The fragment length distribution across all samples. (E) Heatmap and profile of ATAC-seq signals around the ATAC peak center. (F) The bar graph illustrates the TSS enrichment score distribution across all samples. (G) Heatmap and profile of ATAC-seq signals around the BLM-induced upregulated ATAC peak center. (H) Heatmap and profile of ATAC-seq signals around the BLM-induced downregulated ATAC peak center.

Supplementary Material 5: Supplementary fig. 3 (A) The UMAP plot displays the distribution of cells under two distinct conditions before batch effect correction, with blue dots indicating cells treated with saline and orange dots indicating cells treated with BLM. (B) The indicated

marker genes were used to select clusters for subsetting into stromal cells, epithelial cells, endothelial cells, and leukocytes. (C) The UMAP plot shows the contamination factor per cell calculated by decontX. (D) The UMAP plot visualizes the percentage of hemoglobin per cell, illustrating the distribution of cells based on their hemoglobin content. (E) The UMAP plot displays the cell specificity of gene expression for 14 genes in both human and mouse cells.

Acknowledgements

I would like to thank the High-Performance Computing Center of Henan Normal University for their support with data analysis. I am grateful to Professor Linyuan Ru from the Computer Science College for helping to arrange the necessary computing resources.

Author contributions

ZL and GY designed the study and supervised the research project. ZL, MZ and YZ collected and analyzed the data. YG, ZZ and JW interpreted the results and contributed to the theoretical framework. ZL and YZ helped draft the manuscript. LW and GY coordinated the project, substantively revised the manuscript, and ensured the integrity of the work. All authors read and approved the final manuscript.

Funding

Ministry of Science and Technology, PR China grant 2019YFE0119500; Key R&D Program of Henan province grant 231111310400; Zhongyuan scholar 244000510009; Henan Project of Science and Technology grants 232102521025, 232102310011, and GZS2023008.

Data availability

All data from this study are stored at the China National Center for Bioinformation, with ATAC-seq, RNA-seq, and Hi-C data under PRJCA023051 (CRA014630) and scRNA-seq data under PRJCA022530 (CRA014318). Processed data are also available on Figshare (<https://doi.org/10.6084/m9.figshare.28131242>).

Declarations

Ethics approval and consent to participate

The research received the green light from the Animal Care Committee of Henan Normal University, with the approval number HTU2019-02. It was conducted in compliance with all pertinent institutional and national regulations regarding the ethical treatment of animals.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 29 October 2024 / Accepted: 1 March 2025

Published online: 12 March 2025

References

- Henderson NC, Rieder F, Wynn TA. Fibrosis: from mechanisms to medicines. *Nature*. 2020;587:555–66.
- Lederer DJ, Martinez FJ. Idiopathic pulmonary fibrosis. *N Engl J Med*. 2018;378:1811–23.
- Podolanczuk AJ et al. Idiopathic pulmonary fibrosis: state of the Art for 2023. *Eur Respir J* 61(2023).
- Herrerias MM, Budinger GRS. The integrated stress response links Muc5b to pulmonary fibrosis. *Am J Respir Cell Mol Biol*. 2023;68:5–6.
- Liu Y et al., Transcriptome Classification Reveals Molecular Subgroups in Idiopathic Pulmonary Fibrosis. *Genet Res (Camb)* 2022, 7448481 (2022).
- Wang L, et al. Serum proteomics identifies biomarkers associated with the pathogenesis of idiopathic pulmonary fibrosis. *Mol Cell Proteom*. 2023;22:100524.
- Zhang D et al. Telomere length and immunosuppression in non-idiopathic pulmonary fibrosis interstitial lung disease. *Eur Respir J* 62(2023).
- Makena P, Kikalova T, Prasad GL, Baxter SA. Oxidative stress and lung fibrosis: towards an adverse outcome pathway. *Int J Mol Sci* 24(2023).
- Zhu J et al. The role of DNA damage and repair in idiopathic pulmonary fibrosis. *Antioxid (Basel)* 11(2022).
- Moore B. Animal models of fibrotic lung disease. *Am J Respir Cell Mol Biol*. 2013;49:167–79.
- Spagnolo P, Maher TM. A long and winding road: drug development in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2024;209:1072–3.
- Lafontaine DL, Yang L, Dekker J, Gibcus JH, Hi. -C 3.0: improved protocol for Genome-Wide chromosome conformation capture. *Curr Protoc*. 2021;1:e198.
- Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289–93.
- Su Y, et al. Integrative ATAC-seq and RNA-seq analysis of myogenic differentiation of ovine skeletal muscle satellite cell. *Genomics*. 2024;116:110851.
- Grandi FC, Modi H, Kampman L, Corces MR. Chromatin accessibility profiling by ATAC-seq. *Nat Protoc*. 2022;17:1518–52.
- Wang X et al. Reorganization of 3D chromatin architecture in doxorubicin-resistant breast cancer cells. *Front Cell Dev Biology* 10(2022).
- Shaban HA, Gasser SM. Dynamic 3D genome reorganization during senescence: defining cell States through chromatin. *Cell Death Differ* (2023).
- He J, Yan A, Chen B, Huang J, Kee K. 3D genome remodeling and homologous pairing during meiotic prophase of mouse oogenesis and spermatogenesis. *Dev Cell*. 2023;58:3009–e30273006.
- Chen L, et al. Dynamic 3D genome reorganization during development and metabolic stress of the Porcine liver. *Cell Discovery*. 2022;8:56.
- Papazoglou A et al. Epigenetic regulation of profibrotic macrophages in systemic sclerosis-associated interstitial lung disease. *Arthritis Rheumatol* (2022).
- Yang IV, Schwartz DA. Epigenetics of idiopathic pulmonary fibrosis. *Translational Res*. 2015;165:48–60.
- Konigsberg IR et al. Molecular signatures of idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* (2021).
- Zhou S, Wang X, Gao H, Zeng Y. DNA methylation in pulmonary fibrosis. *Adv Exp Med Biol*. 2020;1255:51–62.
- Xu L, et al. 3D disorganization and rearrangement of genome provide insights into pathogenesis of NAFLD by integrated Hi-C, nanopore, and RNA sequencing. *Acta Pharm Sin B*. 2021;11:3150–64.
- Ming, Zhao, Hu. L. 3D genome alterations in T cells associated with disease activity of systemic lupus erythematosus. *Ann Rheum Dis* (2022).
- Zhou B, et al. Comprehensive epigenomic profiling of human alveolar epithelial differentiation identifies key epigenetic States and transcription factor co-regulatory networks for maintenance of distal lung identity. *BMC Genomics*. 2021;22:906.
- Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–90.
- Servant N, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol*. 2015;16:259.
- Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9:357–9.
- Wolff J, et al. Galaxy hicexplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res*. 2020;48:W177–84.
- Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–89.
- Smith JP, et al. PEPATAC: an optimized pipeline for ATAC-seq data analysis with serial alignments. *NAR Genom Bioinform*. 2021;3:lqab101.
- Jiang H, Lei R, Ding SW, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*. 2014;15:182.
- Love MI, Huber W, Anders S. Moderated Estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
- Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
- Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30.
- Kuleshov MV, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90–97.
- Fang Z, Liu X, Peltz G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 39(2023).

39. Liberzon A, et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1:417–25.
40. Badia IMP, et al. DecoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv.* 2022;2:vbac016.
41. Schubert M, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun.* 2018;9:20.
42. Jenkins RG, et al. An official American thoracic society workshop report: use of animal models for the preclinical assessment of potential therapies for pulmonary fibrosis. *Am J Respir Cell Mol Biol.* 2017;56:667–79.
43. Zhao W et al. Endothelial cell-derived MMP19 promotes pulmonary fibrosis by inducing E(nd)MT and monocyte infiltration. *Cell Communication Signal* 21(2023).
44. Guan R et al. Bone morphogenetic protein 4 inhibits pulmonary fibrosis by modulating cellular senescence and mitophagy in lung fibroblasts. *Eur Respir J* 60(2022).
45. Müller-Dott S, et al. Expanding the coverage of Regulons from high-confidence prior knowledge for accurate Estimation of transcription factor activities. *Nucleic Acids Res.* 2023;51:10934–49.
46. Nguyen HN et al. Leukemia inhibitory factor (LIF) receptor amplifies pathogenic activation of fibroblasts in lung fibrosis. *BioRxiv* (2024).
47. Tyler A, Mahoney JM, Carter GW. Genetic interactions affect lung function in patients with systemic sclerosis. *G3 (Bethesda).* 2020;10:151–63.
48. Wang L et al. Single-Cell RNA-seq provides new insights into therapeutic roles of thyroid hormone in the idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* (2023).
49. Sikkema L et al. An integrated cell atlas of the lung in health and disease. *Nat Med* (2023).
50. Li W, et al. Pbx1 activates Fgf10 in the mesenchyme of developing lungs. *Genesis.* 2014;52:399–407.
51. Wu HX, et al. LIFR promotes tumor angiogenesis by up-regulating IL-8 levels in colorectal cancer. *Biochim Biophys Acta Mol Basis Dis.* 2018;1864:2769–84.
52. Kong S, Zhang Y. Deciphering Hi-C: from 3D genome to function. *Cell Biol Toxicol.* 2019;35:15–32.
53. Shashikant T, Ettensohn CA. Genome-wide analysis of chromatin accessibility using ATAC-seq. *Methods Cell Biol.* 2019;151:219–35.
54. Mack M. Inflammation and fibrosis. *Matrix Biol.* 2018;68–69:106–21.
55. Miyazaki Y, et al. Expression of a tumor necrosis factor- α transgene in murine lung causes lymphocytic and fibrosing alveolitis. A mouse model of progressive pulmonary fibrosis. *J Clin Invest.* 1995;96:250–9.
56. Li C 1, 2, W.Q., Xiaoqin Wei 1 2, Harish Narasimhan 1 2 3, Yue Wu 1 2, Mohd Arish 1 2, In Su Cheon 1 2, Jinyi Tang 1 2, Gislane de Almeida Santos 1 2, Ying Li 4, Kamyar Sharifi 1 2, Ryan Kern 5, Robert Vassallo 5, Jie Sun. Comparative single-cell analysis reveals IFN- γ as a driver of respiratory sequelae after acute COVID-19. *Sci Transl Med* (2024).
57. Strunz M, et al. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat Commun.* 2020;11:3559.
58. Aspal M, Zemans RL. Mechanisms of ATII-to-ATI cell differentiation during lung regeneration. *Int J Mol Sci* 21(2020).
59. Devos H, Zoidakis J, Roubelakis MG, Latosinska A, Vlahou A. Reviewing the regulators of COL1A1. *Int J Mol Sci.* 2023;24:10004.
60. Bao L, Chu Y, Kang H. SNAIL-activated long non-coding RNA LINC01711 promotes hepatic fibrosis cell proliferation and migration by regulating XYLT1. *Genomics.* 2023;115:110597.

Publisher's note

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