ACCSCIENCE PUBLISHING

Microbes & Immunity

ORIGINAL RESEARCH ARTICLE

Big data and systems biology for biomarker discovery in chronic hypersensitivity pneumonitis and drug repurposing using deep learning-based DTI models

Sung-Yu Lin, Chung-Kuan Liang, and Bor-Sen Chen*

Department of Electrical Engineering, College of Electrical Engineering and Computer Science, National Tsing Hua University, Hsinchu, Taiwan

Abstract

Chronic hypersensitivity pneumonitis (CHP) is commonly associated with repeated inhalation of organic dust-containing microbial agents. The identification and mitigation of exposure to these microbes are key steps in managing and preventing CHP. Therefore, to elucidate the pathogenetic mechanisms underlying fibrosis and identify biomarkers for targeted therapeutics, a systems biology approach integrating large-scale microarray data mining and whole-genome RNA-sequencing data is crucial. In this study, systems biology methodologies were employed to construct genome-wide genetic and epigenetic networks for fibrotic lung cells from CHP patients and healthy controls. Protein-protein interaction networks and gene/microRNA/long non-coding RNA regulatory networks were established using RNA sequencing data, followed by refinement through a system identification framework. By applying principal network projection and the Kyoto Encyclopedia of Genes and Genomes pathway annotation, core signaling pathways and their downstream cellular dysfunctions were identified, revealing key signaling pathways, such as the mitogen-activated protein kinase, phosphoinositide 3-kinase, chemokine (C-X-C motif) ligand (CXCL)/CXC chemokine receptors family, and chemokine (C-C motif) ligand (CCL) family signaling pathways. In addition, significant biomarkers of pathogenetic mechanisms of CHP, such as AKT1, CCL20, cell division cycle 23, CXCL1, nuclear factor-kappa B, and tumor necrosis factor, were identified as potential drug targets of CHP. Significant biomarkers associated with cell inhibitors, cellular dysfunction, and cellular disorders were identified as drug targets in the core disease-causing signaling pathways. Finally, using a deep neural network-based drug-target interaction model and drug design specification, three molecular drugs, azathioprine, masitinib, and primaquine, were screened and proposed as candidates for a multiple-molecular drug for the treatment of CHP.

Keywords: Chronic hypersensitivity pneumonitis; Systems biology; Genome-wide genetic and epigenetic networks; Deep neural network-based drug-target interaction model

***Corresponding author:**
Bor-Sen Chen
(bschen@ee.nthu.edu.tw)

Citation: Lin S, Liang C, Chen B. Big data and systems biology for biomarker discovery in chronic hypersensitivity pneumonitis and drug repurposing using deep learning-based DTI models. *Microbes & Immunity*. 2025; 2(2):76-105.
doi: 10.36922/mi.4620

Received: August 21, 2024

1st revised: December 11, 2024

2nd revised: January 6, 2025

Accepted: January 20, 2025

Published online: March 28, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Chronic hypersensitivity pneumonitis (CHP), also known as chronic bird fancier's lung or chronic pigeon breeder's lung, is a complex interstitial lung disease characterized by an

aberrant inflammatory response to the inhalation of various organic antigens.[1,2] This condition represents a subtype of hypersensitivity pneumonitis, a group of immunologically mediated disorders caused by sustained exposure to a diverse array of environmental agents, including animal proteins, bacterial and fungal components, and low-molecular-weight chemical compounds. CHP is a relatively uncommon disease, with an estimated prevalence ranging from 1.6 to 6.8 cases/100,000 population.[3] The disease can affect individuals of all ages, but it is more commonly diagnosed in adulthood, with a mean age of onset between 50 and 60 years old. Some studies suggest a slight predilection for females, although the reasons for this potential gender difference remain unclear.[1]

The pathogenesis of CHP is multifactorial, involving a dysregulated[4] interplay between genetic predisposition, environmental exposures, and immunological dysregulation. Repeated inhalation of the offending antigens leads to the recruitment and activation of various inflammatory cells, including lymphocytes, macrophages, and neutrophils, within the pulmonary interstitium. This sustained inflammatory response culminates in the formation of non-caseating granulomas, interstitial fibrosis, and, ultimately, irreversible lung parenchymal destruction. Through bidirectional RNA sequencing analysis of lung tissue samples from CHP patients, we trimmed false positives from these networks, yielding refined core genome-wide and epigenetic interaction networks (GWGENs).[4] From these curated networks, we identified key dysregulated signaling pathways associated with the pathogenesis of CHP, such as the mitogen-activated protein kinase (MAPK) signaling pathway, phosphoinositide 3-kinase (PI3K) pathway, and chemokine signaling pathways involving the chemokine (C-X-C motif) ligand (CXCL)/chemokine (C-X-C motif) receptor (CXCR) and chemokine (C-C motif) ligand (CCL) families.[5] Accumulating evidence implicates these signaling pathways in the aberrant inflammatory response, oxidative stress, apoptosis, and fibrosis that characterize CHP. For instance, the dysregulated MAPK signaling pathway leads to the altered phosphorylation and activation of transcription factors such as nuclear factor-kappa B (NF-κB), AP-1, and their downstream pro-inflammatory mediators.[5] Similarly, the PI3K/AKT signaling pathway regulates cellular processes such as proliferation, survival, and metabolism, contributing to lung epithelial cell dysfunction in CHP. By integrating our systems biology methodologies with annotations from curated pathway databases (e.g., Kyoto Encyclopedia of Genes and Genomes [KEGG] pathways), we identified critical nodes and regulatory hubs within these disease-associated signaling pathways. Leveraging principal network projection (PNP) techniques, we extracted core signaling pathways of CHP and non-CHP to identify putative biomarkers, such as kinases (e.g., MAPK, PI3K), transcription factors (NF-κB, AP-1), and chemokine receptors (CXCR, CCR), as therapeutic drugs for the treatment of CHP for further validation and drug development efforts.[5]

Despite significant advancements in understanding CHP, diagnostic challenges persist due to its reliance on multimodal assessment combining clinical evaluation, radiological interpretation, histopathological findings, and exclusion of other interstitial lung diseases. Current management remains primarily supportive, focusing on antigen avoidance and cautious administration of immunosuppressive agents, such as corticosteroids and antifibrotic therapies.[2,3] To address this need, our study employs a systematic multi-omics approach to elucidate the key molecular signaling pathways and biomarkers implicated in CHP pathogenesis, with the ultimate goal of identifying potential therapeutic drug targets. Specifically, we leveraged large-scale data from transcriptomics, proteomics, and interactomics to construct comprehensive systems models, including protein-protein interaction networks (PPINs), gene regulatory networks (GRNs), and epigenetic interaction networks (EIN).[6] Furthermore, we employed pre-trained deep neural networks (DNNs) for drug-target interaction (DTI) prediction, utilizing established DTI databases coupled with large-scale drug databases to virtually screen for potential multi-target drug combinations.[6] These candidate compounds are further prioritized based on their pharmacological properties, including regulatory mechanisms, toxicity profiles, and target specificities, paving the way for potential drug repurposing or *de novo* design strategies tailored to CHP (Figure 1).

Consequently, there is an urgent need for a deeper comprehension of the intricate molecular mechanisms underlying CHP, with the ultimate goal of identifying novel diagnostic biomarkers and developing targeted therapeutic strategies. In this context, the integration of systems biology approaches, leveraging large-scale multi-omics data and network-based modeling, holds significant promise in unraveling the complex regulatory networks and signaling cascades implicated in CHP pathogenesis.

2. Materials and methods

2.1. Overview of systems biology methods and systemic drug discovery of CHP

This paper employs a systems biology approach to investigate core signaling pathways in CHP and identify significant biomarkers as drug targets, ultimately facilitating multi-target drug discovery via DNN-based

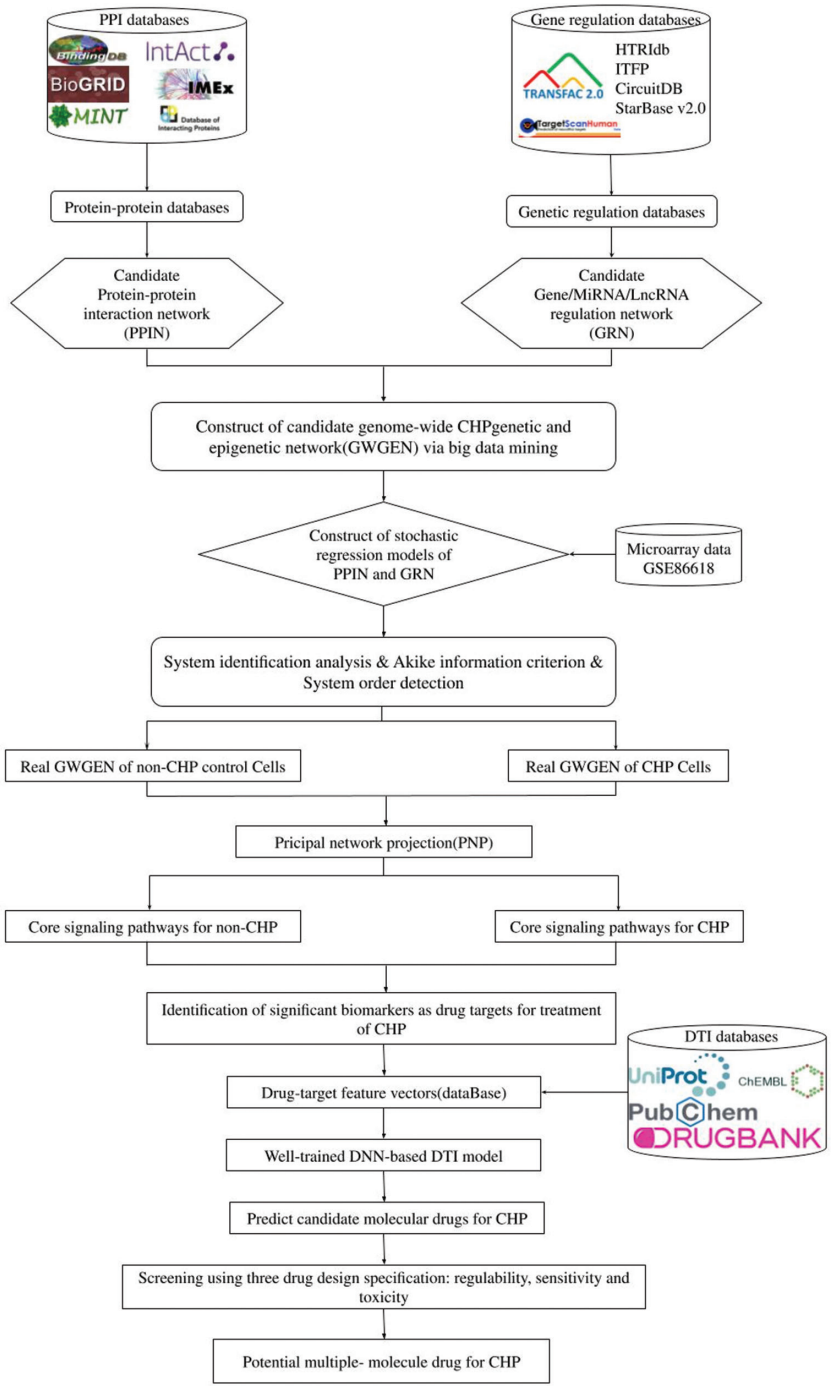

Figure 1. Flowchart of the proposed systems biology method and systematic drug discovery design for a multi-molecular drug for chronic hypersensitivity pneumonitis (CHP). The candidate core genome-wide and EIN (GWGEN) for CHP is composed of a gene regulation network (GRN) and protein-protein interaction network (PPIN). The GRN was constructed by integrating data from gene regulation databases, while the PPIN was built using data from protein-protein interaction databases. To refine the GWGEN and prune down false positives, the real GWGENs for both CHP and healthy controls were pruned using microarray data from the GSE86618 dataset. This process involves the application of system parameter identification and system order detection methods. Then, core GWGENs for both CHP and healthy control were extracted from real GWGENs using the principle network projection method. The core signaling pathways for both non-CHP and CHP were identified by annotating the core GWGENs with Kyoto Encyclopedia of Genes and Genomes pathway information. The significant pathogenetic biomarkers of CHP were identified by comparing the core signaling pathways and their downstream abnormal cellular functions between non-CHP and CHP. A well-trained deep neural network (DNN)-based drug-target interaction (DTI) model, trained on DTI databases, was used to predict candidate molecular drugs for these pathogenetic biomarkers. Based on this prediction, drug design specifications were proposed to select a multi-molecule drug for the treatment of CHP.

models. Three design specifications were used to select adequate molecular drugs as potential multi-molecular drugs for CHP, with an emphasis on inhibiting and regulating disease symptoms. Lung fibrosis cells from CHP patients and normal lung tissue cells from healthy control groups, specifically the A549 cell line, were sampled to construct a GWGEN for the identification of core signaling pathways involved in the pathogenesis of CHP. Candidate GWGEN networks, including candidate PPINs and GRNs encompassing regulatory networks of fibrosis genes and genes from healthy controls, were constructed using big data mining and preprocessing of gene/microRNA (miRNA)/long non-coding RNA (lncRNA) expression data from fibrosis cells in online databases.

By leveraging big data mining and systems modeling techniques, we aimed to investigate and compare the molecular mechanisms of CHP and non-CHP. This involved constructing a GWGEN, core GWGEN, and core signaling pathways for both CHP and non-CHP. The steps involved are depicted in Figure 1. First, big data mining was used from the corresponding database and preprocessing of gene, miRNA, and lncRNA expression data. Second, candidate PPI networks and GRNs were employed to construct GWGEN candidates. Next, whole-genome microarray data from different samples of CHP and non-CHP (fibrosis cells and healthy control cells) were used to identify the pathogenetic mechanism of CHP and compare them with the healthy control group. Subsequently, the real GWGENs of CHP and non-CHP were constructed, and the PNP method was employed to extract core proteins, genes, miRNA, and lncRNA from the real GWGENs to construct the core GWGEN of CHP and non-CHP. Finally, core signaling pathways were constructed by annotating the KEGG pathways of CHP and health control from their core GWGENs. Finally, the core signaling pathways of CHP and healthy control and their downstream cellular dysfunction were compared to investigate the pathogenetic mechanisms.

By comparing the core GWGEN of fibrosis cells from CHP patients with normal lung cells from the healthy control group at different stages of the cell repair cycle (damage, repair, and fibrosis), we extracted the differential core signal pathways between CHP patient fibrosis cells and normal lung cells. This helped to unravel the genetic and epigenetic mechanisms underlying fibrosis in CHP patients. In addition, based on our findings, we selected significant biomarkers of the pathogenetic mechanisms as drug targets for the treatment of CHP. Finally, utilizing DTI databases, we trained a DNN-based DTI model to predict candidate molecular drugs, from which potential molecular drugs were selected as multi-molecular drugs based on three design specifications for the treatment of CHP.

2.2. Data preprocessing of CHP and healthy controls microarray data

In this study, microarray data (GSE86618) from the National Center for Biotechnology Information (NCBI) were analyzed. The dataset, divided into the fibrosis lung cell and the healthy lung tissue cell control groups, consists of 325 CHP samples and 215 non-CHP samples. The significant variability in gene expressions and PPIs under different biological conditions highlights the risk of overlooking individual differences if only significant genes or proteins are considered. To uncover the cellular mechanisms of CHP, the corresponding core signaling pathways must be identified from GWGEN. The candidate GWGEN, constructed using experimental data and computational predictions from multiple databases, includes both the PPIN and GRN. A binary Boolean matrix was used to represent the candidate GWGEN, with "1" indicating interaction or regulation between nodes and "0" indicating no interaction or regulation.

To establish the candidate PPIN, we referenced various databases, including the Transcription Factor Database (TRANSFAC),[7] the Biological General Repository for Interaction Datasets (BioGRID),[8] the Database of Interacting Proteins (DIP),[9] the Molecular INTeraction database (MINT),[10] and IntAct.[11] Similarly, we utilized TargetScanHuman[12] and the Human Transcriptional Regulation Interaction Database (HTRIdb)[13] to identify candidate regulatory interactions between human transcription factors and their targets. These interactions were based on integrated transcription factor platform databases (ITFP),[14] CircuitDB,[15] and StarBase v2.0.[16] By integrating candidate PPI, candidate regulatory interactions between transcription factors (TFs)/miRNA/lncRNA and their target genes/miRNA/lncRNA, whole-genome microarray gene/miRNA/lncRNA expression data, and DNA methylation profiles, we obtained the candidate GWGEN using MATLAB text files. String manipulation tools in large text mining were used to standardize these gene names according to the standard gene name rules in the NCBI Gene Database, achieving an automatic extraction method for key biomarkers.

2.3. Stochastic regression models construction of candidate PPIN and gene regulation network of candidate genome-wide and EINs

In the previous section, based on text mining and comprehensive big data mining from GSE86618 microarray gene databases, we constructed a candidate GWGEN. However, the candidate GWGEN built from

numerous databases and experimental datasets contains many false positives, including some seemingly reasonable but incorrect information. Therefore, these false positives should be removed to obtain the real GWGEN of CHP and non-CHP using systems biology methods.[17] Therefore, we needed to trim these false positives from the candidate GWGEN based on gene/miRNA/lncRNA expression data of CHP and non-CHP.

We first constructed a stochastic regression interaction/regulation model of human cells to characterize the interactions and regulations in the candidate GWGEN, including PPI, gene regulation, miRNA regulation, lncRNA regulation, and epigenetic regulation. To identify the real GWGEN for each lung tissue cell condition, we applied system identification and system order detection methods to the interaction/regulation models of the candidate GWGEN using gene/miRNA/lncRNA expression data and epigenetic profiles for each lung tissue cell condition. Significant interactions and regulations beyond the system order are considered false positives in the candidate GEN, which were trimmed to obtain the real GWGEN for CHP fibrosis cells and healthy lung tissue cells. The stochastic regression protein interaction model for candidate PPIN in the candidate GWGEN is represented in Equation I for the protein interaction of the *k*-th protein in lung cells of the sample *i*:

$$S_k[i] = \sum_{\substack{w=1 \\ w \neq k}}^{W_k} \alpha_{kw} S_k[i] S_w[i] + \beta_k + \gamma_k[i]$$

for $k = 1,2,\ldots,K$ and $i=1,2,\ldots,I$ (I)

where S_k[i] and $S_w[i]$ denote the expression level of the *k*-th protein and the *w*-th protein for the *i*-th sample, respectively. α_{kw} is the interaction ability between the *k*-th protein and the *w*-th protein, which is an interactive protein of the *k*-th protein. W_k represents the number of proteins interacting with the *k*-th protein and K is the number of proteins with candidate PPIN. *i* denotes the number of data samples (pneumonitis lung cell and non-pneumonitis lung cell).β_k represents the basal level of protein k due to some unknown effects, such as phosphorylation, methylation, and ubiquitination. $\gamma_k[i]$ is the stochastic noise of the *k*-th protein for the sample *i* due to model uncertainty and measurement noise. The protein interaction model in Equation I can be interpreted as follows: the expression level of the *k*-th protein is related to the interactions with other proteins, denoted as W_k, in the candidate PPIN.

Subsequently, we created a regulatory system model that includes interactions between genes and their regulators, such as TFs, miRNAs, and lncRNAs. The gene regulatory model of candidate GRN, describing the transcriptional regulation of the *l*-th gene of lung slice cells for sample *I,* is given in Equation II[18]:

$$t_l[i] = \sum_{x=1}^{X_l} \delta_{lx} e_x[i] P_x[i] + \sum_{y=1}^{Y_l} \zeta_{ly} f_y[i] P_l[i] - \sum_{z=1}^{Z_l} \omega_{lz} g_z[i] P_l[i] + v_l P_l[i] + u_l[i]$$

for $l = 1,2,\ldots,L$ and $i=1,2,\ldots,I$ (II)

where $t_l[i]$, $e_x[i]$, $f_y[i]$, and $g_z[i]$ denote the expression level of the *l*-th target gene, the *x*-th TF, the *y*-th lncRNA and the *z*-th miRNA for the *i*-th sample, respectively. δ_{lx} and ζ_{ly} are the transcription regulatory ability of the *x*-th TF and the *y*-th lncRNA on their corresponding binding target gene *l*. ω_{lz} indicates the post-transcriptional regulatory ability of the *z*-th miRNA to inhibit the *l*-th target gene ($-\omega_{lz} \leq 0$). X_l, Y_l, and Z_l represent the number of TFs, lncRNAs and miRNAs binding to the *l*-th target gene. L and I denote the number of genes with candidate GRN and the number of data samples. v_l represents the basal level of the target gene *l*. $u_l[i]$ is the stochastic noise of the *l*-th target gene for the sample *i* due to model uncertainty and data noise. $P_l[i]$ denotes the methylation regulation of the *l*-th gene through its effect on the binding affinities of TFs, miRNAs, lncRNAs, and RNA polymerase on the target gene. The terms $\delta_{lx} e_x[i] P_x[i]$, $\zeta_{ly} f_y[i] P_l[i]$, $\omega_{lz} g_z[i] P_l[i]$, and $v_l P_l[i]$ denote the effect of methylation, phosphorylation, or ubiquitination on the binding sympathy of TFs, miRNAs, lncRNAs, and RNA polymerase to the *l*-th target gene, respectively. Furthermore, to evaluate the methylation regulation direction of the *l*-th target gene $t_l[i]$ using DNA methylation profile $P_l[i]$ can be defined as follows[19] in Equation III.

$$P_l[i] = \frac{1}{1 + (\frac{P_l[i]}{0.5})^2} \quad \text{(III)}$$

where $P_l[i]$ indicates the DNA methylation profile of the *l*-th gene for the sample *i*. In the equation above, the range of effects of DNA methylation on the *l*-th target gene $P_l[i]$ is 1 to 0.2, while the DNA methylation profile, $p_l[i]$, ranges from 0 to 1. From the biological system aspect, the equation above suggests that the higher the DNA methylation level, the weaker the binding between TFs, miRNAs, lncRNAs, and RNA polymerases and their target genes. In contrast, the lower the DNA methylation level, the stronger the binding between TFs, miRNAs, lncRNAs, and RNA polymerases and their target genes. The methylation regulation $P_l[i]$ in Equation III has a regulation value ($P_l[i] = 0.2$), which corresponds to the DNA methylation

profile of 1, indicating that the bindings of TFs, miRNAs, lncRNAs, and RNA polymerase to the *l*-th target gene still exist. The expression level of the *l*-th gene results from transcriptional regulations of TFs and lncRNAs, the post-transcriptional regulations of miRNAs, and the basal level of the *l*-th gene with measurement noise in Equation II. Furthermore, the transcriptional regulations of TFs and lncRNAs, the post-transcriptional regulations of miRNAs, and basal expression level (RNA polymerase binding) can be influenced by DNA methylation in Equation III.

In addition, the miRNA regulatory system model of candidate GRN, describing the transcriptional expression of the *m*-th miRNA of lung slice cells for sample *i*, is constructed by Equation IV:

$$d_m[i] = \sum_{x=1}^{X_m} \eta_{mx} e_x[i] + \sum_{y=1}^{Y_m} \kappa_{my} f_y[i] - \sum_{z=1}^{Z_m} \lambda_{mz} g_z[i] d_m[i] + c_m + h_m[i]$$

for $m = 1,2,\ldots,M$ and $i=1,2,\ldots,I$ (IV)

where $d_m[i]$ represents the expression level of the *m*-th miRNA in the *i*-th sample. X_m, Y_m, and Z_m, respectively, represent the total numbers of TFs, lncRNAs, and miRNAs binding to the *m*-th miRNA. $e_x[i]$, $f_y[i]$, and $g_z[i]$, respectively, denote the expression levels of the *x*-th TF, the *y*-th lncRNA, and the *z*-th miRNA in the *i*-th sample. η_{mx} denotes the transcriptional regulatory ability of the *x*-th TF on the *m*-th miRNA. k_{my} denotes the transcriptional regulatory ability of the *y*-th lncRNA on the *m*-th miRNA. $\lambda_{mz} > 0$ represents the post-transcriptional regulatory ability of the *z*-th miRNA in degrading the *m*-th miRNA. M is the total amount of miRNAs in the candidate GWGEN, and I is the total number of data samples (fibrosis slice cells or non-fibrosis slice cells). c_m represents the basal level due to phosphorylation, ubiquitination, acetylation, or other gene regulatory effects that are considered to have unknown effects on the miRNA. h_m[i] represents the random noise of the *m*-th miRNA in the *i*-th sample due to the residuals from model establishment and experimental measurement noise.

Furthermore, the lncRNA regulatory system model of candidate GRN, describing the transcriptional expression level of the *n*-th lncRNA of lung cells for sample *i*, is presented in Equation V:

$$q_n[i] = \sum_{x=1}^{X_n} o_{nx} e_x[i] + \sum_{y=1}^{Y_n} \rho_{ny} f_y[i] - \sum_{z=1}^{Z_n} \tau_{nz} g_z[i] q_n[i] + u_n + v_n[i]$$

for $n = 1,2,\ldots,N$ and $i=1,2,\ldots,I$ (V)

where $q_n[i]$ represents the expression level of the *n*-th lncRNA in the *i*-th sample. X_n, Y_n, and Z_n, respectively, represent the total amount of TFs, lncRNAs, and miRNAs binding to the *n*-th lncRNA. $e_x[i]$, $f_y[i]$, and $g_z[i]$, respectively, denote the expression levels of the *x*-th TF, the *y*-th lncRNA, and the *z*-th miRNA in the *i*-th sample. o_{mx} denotes the transcriptional regulatory ability of the *x*-th TF on the *n*-th lncRNA. ρ_{ny} denotes the transcriptional regulatory ability of the *y*-th lncRNA on the *n*-th lncRNA. $\tau_{nz} > 0$ represents the post-transcriptional regulatory ability of the *z*-th miRNA and the *n*-th lncRNA. N is the total amount of lncRNAs in the candidate GWGEN, and I is the total amount of data samples (fibrosis slice cells or non-fibrosis slice cells). u_n represents the basal level, that is, the change in expression of the *n*-th lncRNA due to phosphorylation, ubiquitination, and acetylation. $v_n[i]$ represents the random noise in the *i*-th sample due to the residuals of modeling and measurement noise.

2.4. System identification and order detection in stochastic regression models for real gene regulation network of candidate genome-wide and EINs

After constructing the protein interaction regression model for candidate PPINs in Equation I and the regulatory models of genes/miRNAs/lncRNAs for candidate GRNs in Equations II, IV, and V, we applied system identification method[17] and system order detection scheme[19] to identify protein interactions in the protein interaction model in Equation I and genetic regulations of genetic regulation models in Equations II, IV, and V. Using parameters α_{kw}, β_k for the protein interaction model and δ_{lx}, k_{my}, and τ_{nz} for the gene/miRNA/lncRNA regulatory models, we combined lncRNA expression data and DNA phosphorylation, ubiquitination, and methylation of each lung slice cell with the protein interaction model and the gene/miRNA/lncRNA regulatory models to identify real GWGENs of CHP and non-CHP using their microarray data. Next, our goal was to use system identification and order detection methods to filter out false-positive interactions and regulations[17] from the candidate GWGENs. In addition, Equations II, IV, and V can be represented as the following linear regression equations (Equations VI-IX) to obtain the parameter vectors for protein interactions and gene regulations in candidate GWGEN of CHP and non-CHP.[18]

$$S_k[i] = \left[S_k[i] S_1[i] S_k[i] S_2[i] \cdots S_k[i] S_{Kk}[i] 1 \right] \times \begin{bmatrix} a_{k1} \\ a_{k2} \\ \vdots \\ a_{Kk} \\ b_k \end{bmatrix}$$

$$+\varphi_k[i] = \sigma_k[i] * \beta_k + \varphi_k[i]$$

for $k = 1,2,\ldots,K$ and $i=1,2,\ldots,I$ (VI)

$$t_l[i]=\begin{bmatrix} e_1[i]\cdots e_{Xl}[i] f_1[i]\cdots f_{Yl}[i] g_1[i] t_l[i] \\ \cdots g_{Zl}[i] t_l[i] 1 \end{bmatrix} \times \begin{bmatrix} \delta_{l1} \\ \vdots \\ \delta_{XI} \\ \zeta_{lI} \\ \vdots \\ \zeta_{YI} \\ -\omega_{I1} \\ \vdots \\ -\omega_{ZI} \\ \nu_I \end{bmatrix} + \varphi_l[i]=\sigma_l[i]^*\beta_l+\varphi_l[i]$$

for i = 1,2,…,L and i=1,2,…,I (VII)

$$d_m[i]=\begin{bmatrix} e_1[i]\cdots e_{Xm}[i] f_1[i]\cdots f_{Ym}[i]\ \mathrm{g}_1[i] f_m[i] \\ \cdots g_{Zm}[i] d_m[i] 1 \end{bmatrix} \times \begin{bmatrix} \eta_{m1} \\ \vdots \\ \eta_{Xm} \\ \kappa_{u1} \\ \vdots \\ \kappa_{Ym} \\ -\lambda_{u1} \\ \vdots \\ -\lambda_{Zm} \\ c_m \end{bmatrix} + \varphi_m[i]=\sigma_m[i]^*\beta_m+\varphi_m[i]$$

for m = 1,2,…,M and i=1,2,…,I (VIII)

$$q_n[i]=\begin{bmatrix} e_1[i]\cdots e_{Xn}[i] f_1[i]\cdots f_{Yn}[i]\ \mathrm{g}_1[i] g_n[i] \\ \cdots g_{Zn}[i] d_m[i] 1 \end{bmatrix} \times \begin{bmatrix} o_{n1} \\ \vdots \\ o_{Xn} \\ \rho_{n1} \\ \vdots \\ \kappa_{Yn} \\ -\tau_{n1} \\ \vdots \\ -\tau_{Zn} \\ u_n \end{bmatrix} + \varphi_n[i]=\sigma_n[i]^*\beta_n+\varphi_n[i]$$

for n = 1,2,…,N and i=1,2,…,I (IX)

Then, with i microarray data samples, linear Equations VI-IX can be represented as follows[18]:

$$\begin{bmatrix} S_k[1] \\ S_k[2] \\ \vdots \\ S_k[I] \end{bmatrix} = \begin{bmatrix} \sigma_k[1] \\ \sigma_k[2] \\ \vdots \\ \sigma_k[I] \end{bmatrix} \beta_k + \begin{bmatrix} \varphi_k[1] \\ \varphi_k[2] \\ \vdots \\ \varphi_k[I] \end{bmatrix}, for\, k=1,2,\ldots,K \quad (X)$$

$$\begin{bmatrix} t_l[1] \\ t_l[2] \\ \vdots \\ t_l[I] \end{bmatrix} = \begin{bmatrix} \sigma_l[1] \\ \sigma_l[2] \\ \vdots \\ \sigma_l[I] \end{bmatrix} \beta_l + \begin{bmatrix} \varphi_l[1] \\ \varphi_l[2] \\ \vdots \\ \varphi_l[I] \end{bmatrix}, for\, l=1,2,\ldots,L \quad (XI)$$

$$\begin{bmatrix} d_m[1] \\ d_m[2] \\ \vdots \\ d_m[I] \end{bmatrix} = \begin{bmatrix} \sigma_m[1] \\ \sigma_m[2] \\ \vdots \\ \sigma_m[I] \end{bmatrix} \beta_m + \begin{bmatrix} \varphi_m[1] \\ \varphi_m[2] \\ \vdots \\ \varphi_m[I] \end{bmatrix}, for\, m=1,2,\ldots,M \quad (XII)$$

$$\begin{bmatrix} q_n[1] \\ q_n[2] \\ \vdots \\ q_n[I] \end{bmatrix} = \begin{bmatrix} \sigma_n[1] \\ \sigma_n[2] \\ \vdots \\ \sigma_n[I] \end{bmatrix} \beta_n + \begin{bmatrix} \varphi_n[1] \\ \varphi_n[2] \\ \vdots \\ \varphi_n[I] \end{bmatrix}, for\, n=1,2,\ldots,N \quad (XIII)$$

Equations X-XIII can be rewritten into linear regression in Equations XIV-XVII.

$$S_k=\sigma_k{}^*\beta_k+\Delta_k, for\, k=1,2,\ldots,K \quad (XIV)$$

$$T_l=\sigma_l{}^*\beta_l+\Delta_l, for\, l=1,2,\ldots,L \quad (XV)$$

$$D_m=\sigma_m{}^*\beta_m+\Delta_m, for\, m=1,2,\ldots,M \quad (XVI)$$

$$Q_n=\sigma_n{}^*\beta_n+\Delta_n, for\, n=1,2,\ldots,N \quad (XVII)$$

where σ_k, σ_l, σ_m, and σ_n denote the regression vectors for i protein/gene/miRNA/lncRNA expression data and DNA methylation profiles. β_k, β_l, β_m, and β_n are the parameter vectors associated with the protein interactions/gene/miRNA/lncRNA regulations, respectively, containing protein interaction abilities, transcriptional regulatory abilities, post-transcriptional regulatory abilities, and basal levels.

Hence, we can use the constrained least square estimation to estimate the parameter vectors β_k, β_l, β_m, and β_n, containing protein interaction abilities, transcriptional regulatory abilities, post-transcriptional regulatory abilities, and basal levels by the corresponding microarray data and DNA methylation, phosphorylation, and ubiquitination profiles. The constrained least square estimation problem of parameter vectors β_k, β_l, β_m, and β_n in Equations XIV–XVII can be solved by the following constrained least square parameter estimation problems (Equations XVIII-XXI), respectively.[18]

$$\widehat{\beta_k} = \min_{\beta_k} \frac{1}{2} || \sigma_k * \beta_k - S_k ||_2^2 \quad \text{(XVIII)}$$

$$\widehat{\beta_l} = \min_{\beta_l} \frac{1}{2} || \sigma_l * \beta_l - T_l ||_2^2 \quad \text{(XIX)}$$

$$subject\ to \begin{bmatrix} 0 & \cdots & 00 & \ldots & 01 & \ldots & 00 \\ \vdots & \ddots & \vdots\vdots & \ddots & \vdots\vdots & \ddots & \vdots\vdots \\ 0 & \cdots & 00 & \ldots & 00 & \ldots & 10 \\ \underbrace{\quad}_{X_l} & & & & \underbrace{\quad}_{Y_l} & & \underbrace{\quad}_{Z_l} \end{bmatrix}$$

$$\widehat{\beta_m} = \min_{\beta_m} \frac{1}{2} || \sigma_m * \beta_m - D_m ||_2^2$$

$$subject\ to \begin{bmatrix} 0 & \cdots & 00 & \ldots & 01 & \ldots & 00 \\ \vdots & \ddots & \vdots\vdots & \ddots & \vdots\vdots & \ddots & \vdots\vdots \\ 0 & \cdots & 00 & \ldots & 00 & \ldots & 10 \\ \underbrace{\quad}_{X_m} & & & & \underbrace{\quad}_{Y_m} & & \underbrace{\quad}_{Z_m} \end{bmatrix} \quad \text{(XX)}$$

$$\widehat{\beta_n} = \min_{\beta_n} || \frac{1}{2} \sigma_n * \beta_n - Q_n ||_2^2$$

$$subject\ to \begin{bmatrix} 0 & \cdots & 00 & \ldots & 01 & \ldots & 00 \\ \vdots & \ddots & \vdots\vdots & \ddots & \vdots\vdots & \ddots & \vdots\vdots \\ 0 & \cdots & 00 & \ldots & 00 & \ldots & 10 \\ \underbrace{\quad}_{X_n} & & & & \underbrace{\quad}_{Y_n} & & \underbrace{\quad}_{Z_n} \end{bmatrix} \quad \text{(XXI)}$$

The inequality constraints in the above constrained least square parameter estimation problem can ensure that the post-transcriptional regulatory abilities of miRNA on genes/miRNAs/lncRNAs are always non-positive. Therefore, we could solve the constrained least square parameter estimation problem to obtain protein interactive parameters, that is, $\widehat{\beta_k}$ and gene/miRNA/lncRNA regulatory parameters, that is, $\widehat{\beta_l}$, $\widehat{\beta_m}$, and $\widehat{\beta_n}$ through the system identification method in the MATLAB optimization toolbox based on a reflective Newton method for minimizing a quadratic function.[18] Through this process, we are able to get the optimal estimated parameter vectors for PPIs and genes, lncRNAs, and miRNA regulations in GWGENs of fibrosis slice cells of CHP patients and non-fibrosis slice cells of healthy control.

Since large-scale measurement of expressed cellular proteins is not yet feasible, and mRNA abundance explains 79% of the variance in protein abundance, we can use gene expression microarray data as a proxy for protein expressions. By incorporating gene, miRNA, and lncRNA expression data, along with DNA cross-talk between methylation, phosphorylation, and ubiquitination profiles in CHP patients and healthy controls, we can solve S_k, T_l, D_m, Q_n, σ_k, σ_l, σ_m, and σ_n from the least square parameter estimation problems in Equations XVIII-XXI to identify the protein interactive abilities, transcriptional regulatory abilities, post-transcriptional regulatory abilities, and basal levels in β_k, β_l, $\beta_{m,}$ and β_n.

Since the candidate GWGEN interactions and regulations were constructed using big text data mining from a substantial variety of databases and experimental datasets, which may include some plausible information, it is possible that many false positives of protein interactive abilities, transcriptional regulatory abilities, and post-transcriptional regulatory abilities are included in candidate GWGEN. To remove false positives in the estimated protein interactive abilities, transcriptional regulatory abilities in Equations XVIII-XXI, and post-transcriptional regulatory abilities, we applied a system order detection scheme[17-19] to exclude these insignificant abilities as false positives to obtain real GWGENs of CHP and healthy control.

Akaike Information Criterion (AIC) is a system order detection method[20] based on the system identification method in solving Equations XVIII-XXI to detect the real system order. The AICs of the k-th protein of PPIN, the l-th gene of GRN, the m-th miRNA of GRN, and the n-th lncRNA of GRN are shown in Equations XXII-XXV, respectively.

$$AIC(W_k) = \log\left(\Theta_k^2\right) + \frac{2(1+W_k)}{I} \quad \text{(XXII)}$$

$$AIC\left(X_l, Y_{l,} Z_l\right) = \log\left(\Theta_l^2\right) + \frac{2(1+X_l+Y_l+Z_l)}{I} \quad \text{(XXIII)}$$

$$AIC\left(X_m, Y_{m,} Z_m\right) = \log\left(\Theta_m^2\right) + \frac{2(1+X_m+Y_m+Z_m)}{I} \quad \text{(XXIV)}$$

$$AIC\left(X_n, Y_{n,} Z_n\right) = \log\left(\Theta_n^2\right) + \frac{2(1+X_n+Y_n+Z_n)}{I} \quad \text{(XXV)}$$

$\Theta_k^2, \Theta_l^2, \Theta_m^2$, and Θ_n^2 are the estimated residual errors obtained from the least square identification method in Equations XVIII-XXI of the *k*-th PPIN, the *l*-th gene, the *m*-th miRNA, and the *n*-th lncRNA, respectively. Θ_k, Θ_l, Θ_m, and Θ_n, represent the residual parameters estimation of protein, gene, miRNA, and lncRNA, respectively, as follows in Equations XXVI-XXIX.

$$\Theta_k^2 = \sqrt[2]{\frac{(\sigma_k * \widehat{\beta_k} - S_k)^T (\sigma_k * \widehat{\beta_k} - S_{k)}}{I}} \quad \text{(XXVI)}$$

$$\Theta_l^2 = \sqrt[2]{\frac{(\sigma_l * \widehat{\beta_l} - T_l)^T (\sigma_l * \widehat{\beta_l} - T_{l)}}{I}} \quad \text{(XXVII)}$$

$$\Theta_m^2 = \sqrt[2]{\frac{(\sigma_m * \widehat{\beta_m} - D_m)^T (\sigma_m * \widehat{\beta_m} - D_{m)}}{I}} \quad \text{(XXVIII)}$$

$$\Theta_n^2 = \sqrt[2]{\frac{(\sigma_n * \widehat{\beta_n} - Q_n)^T (\sigma_n * \widehat{\beta_n} - Q_{n)}}{I}} \quad \text{(XXIX)}$$

Based on the system order detection theory of system identification,[18,19] the true number of parameters W_k^*, ($X_l^* + Y_l^* + Z_l^*$), ($X_m^* + Y_m^* + Z_m^*$), and ($X_n^* + Y_n^* + Z_n^*$) could minimize AIC (W_k), AIC (X_l, Y_l, Z_l), AIC (X_m, Y_m, Z_m), and AIC (X_n, Y_n, Z_n) in Equations XXII-XXV, respectively, to obtain the true system order of real GWGEN. In other words, the minimum AIC (W_k^*), AIC ($X_l^* + Y_l^* + Z_l^*$), AIC ($X_m^* + Y_m^* + Z_m^*$), and AIC ($X_n^* + Y_n^* + Z_n^*$) could be achieved by true number of parameters (system order) W_k^*, (X_l^*, Y_l^*, Z_l^*), (X_m^*, Y_m^*, Z_m^*), and (X_n^*, Y_n^*, Z_n^*), respectively, by the tradeoff between residual error in the first term of Equations XXVI-XXIVV and parameter association number in the second term of Equations XXII-XXV in the following expression (Equations XXX-XXXIII).

$$W_k^* = arg \min_{W_k} AIC(W_k) \quad \text{(XXX)}$$

$$X_l^*, Y_l^*, Z_l^* = arg \min_{X_l, Y_l, Z_l} AIC(X_l, Y_l, Z_l) \quad \text{(XXXI)}$$

$$X_m^*, Y_m^*, Z_m^* = arg \min_{X_m, Y_m, Z_m} AIC(X_m, Y_m, Z_m) \quad \text{(XXXII)}$$

$$X_n^*, Y_n^*, Z_n^* = arg \min_{X_n, Y_n, Z_n} AIC(X_n, Y_n, Z_n) \quad \text{(XXXIII)}$$

Therefore, we can prune the false positives of protein interaction abilities, transcriptional regulation abilities, and post-transcriptional regulations in candidate GWGENs to obtain real GWGENs for CHP and non-CHP cells. This is done by removing the insignificant interactions and regulations that fall outside the true system order in Equations XXX-XXXIII, processing one protein, one gene, one miRNA, and one lncRNA at a time. The real GWGEN for CHP and non-CHP are pruned from the candidate GWGEN using the proposed AIC method, as shown in Figure 2A and B.

2.5. Core genome-wide and EINs extraction using the PNP method

The real GWGENs remain large and complex, making it challenging to extract significant information and gain insights into the pathogenetic mechanisms of CHP progression. This complexity hinders the comparisons and investigations between lung slice cells from CHP patients and healthy controls. Recently, KEGG pathway annotation has become a powerful tool for analyzing signaling pathways.[19] However, KEGG pathways currently support annotations for only 6,000 molecules. Therefore, the real GWGENs will be truncated to core GWGENs with 6,000 molecules for KEGG pathway annotation. To achieve this, we applied the PNP method to extract 6000 significant proteins, genes, miRNA, and lncRNA as the core GWGENs from the real GWGENs in CHP and non-CHP. The system models describing the PPIs, gene regulations, miRNA

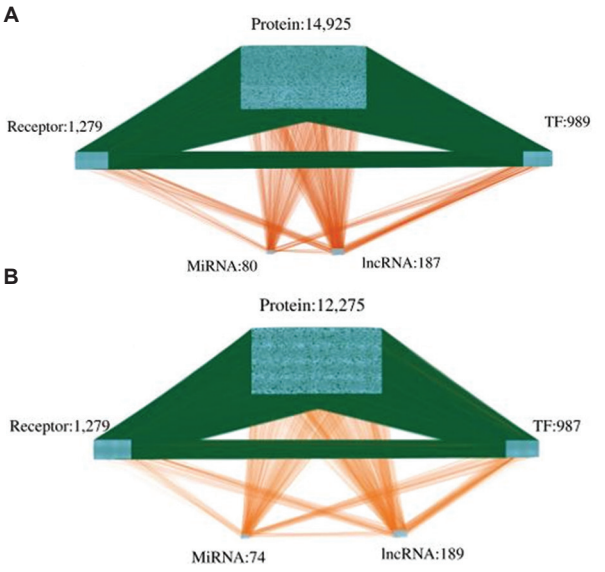

Figure 2. (A) The real GWGEN of CHP and (B) the real GWGEN of non-CHP are identified using system identification and order detection methods via the corresponding microarray data. The numbers indicate the node numbers of proteins, TFs, receptors, lncRNAs, and miRNAs. The green lines represent the protein-protein interactions, and the orange lines represent the gene regulations.
Abbreviations: CHP: Chronic hypersensitivity pneumonitis; GWGEN: Genome-wide and epigenetic interaction networks; lncRNA: Long non-coding RNA; TF: Transcription factor.

regulations, and lncRNA regulations, and epigenetic regulations by DNA methylation, phosphorylation, and ubiquitination of real GENs can be represented in Equations XXXIV-XXXVII:

$$S_k[i] = \sum_{\substack{w=1 \\ w \neq k}}^{W_k} \widehat{\alpha_{kw}} S_k[i] S_w[i] + \widehat{\beta_k}$$

for $k = 1,2,\ldots,K$ and $i = 1,2,\ldots,I$ (XXXIV)

$$t_l[i] = \sum_{x=1}^{X_l} \widehat{\delta_{lx}} e_x[i] P_x[i] + \sum_{y=1}^{Y_l} \widehat{\zeta_{ly}} f_y[i] P_l[i] - \sum_{z=1}^{Z_l} \widehat{\omega_{lz}} g_z[i] P_l[i] + \widehat{v_l} P_l[i]$$

for $l = 1,2,\ldots,L$ and $i = 1,2,\ldots,I$ (XXXV)

$$d_m[i] = \sum_{x=1}^{X_m} \widehat{\eta_{mx}} e_x[i] + \sum_{y=1}^{Y_m} \widehat{\kappa_{my}} f_y[i] - \sum_{z=1}^{Z_m} \widehat{\lambda_{mz}} g_z[i] d_m[i] + \widehat{c_m}$$

for $m = 1,2,\ldots,M$ and $i = 1,2,\ldots,I$ (XXXVI)

$$q_n[i] = \sum_{x=1}^{X_n} \widehat{o_{nx}} e_x[i] + \sum_{y=1}^{Y_n} \widehat{\rho_{ny}} f_y[i] - \sum_{z=1}^{Z_n} \widehat{\tau_{nz}} g_z[i] q_n[i] + \widehat{u_n}$$

for $n = 1,2,\ldots,N$ and $i = 1,2,\ldots,I$ (XXXVII)

where W_k, X_l, Y_l, Z_l, X_m, Y_m, Z_m, X_n, Y_n, and Z_n represent the interaction number of protein k in PPIs, TF regulations on gene *l*, lncRNA regulations on gene *l*, miRNA regulations on gene *l*, TF regulations on miRNA (m), lncRNA regulations on miRNA (m), miRNA regulations on miRNA (m), TF regulations on lncRNA (n), lncRNA regulations on lncRNA (n), and miRNA regulations on lncRNA (n) of real GWGENs obtained by AIC, respectively. $\widehat{\alpha_{kw}}$ indicates the estimated protein interactive abilities and $\widehat{\delta_{lx}}$, $\widehat{\eta_{mx}}$, and $\widehat{o_{nx}}$ denote the estimated transcriptional regulatory abilities of TFs on gene/miRNA/lncRNA, respectively. $\widehat{\zeta_{ly}}$, $\widehat{\kappa_{my}}$, and $\widehat{\rho_{ny}}$ are the estimated transcriptional regulatory abilities of lncRNAs on gene/miRNA/lncRNA and $\widehat{\omega_{lz}}$, $\widehat{\lambda_{mz}}$, and $\widehat{\tau_{nz}}$ represent the estimated post-transcriptional regulatory abilities of miRNAs on gene/miRNA/lncRNA. These estimated protein interactive abilities, estimated transcriptional regulatory abilities, and estimated post-transcriptional regulatory abilities can be obtained by applying system identification methods in Equations XVIII-XXI through the corresponding microarray data.

After the protein interactive abilities, transcriptional regulatory abilities, and post-transcriptional regulatory abilities of real GWGEN were identified, we integrated all interactive abilities of proteins, transcriptional regulatory abilities of TFs and lncRNAs, and post-transcriptional regulatory abilities of miRNAs of the real GWGEN in Equations XXXIV-XXXVII as the following network matrix *R* of real GWGEN (Equation XXXVIII):

$$R = \begin{bmatrix} R_{p-p} & 0 & 0 \\ R_{t-g} & R_{l-g} & R_{m-g} \\ R_{t-l} & R_{l-l} & R_{m-l} \\ R_{t-m} & R_{l-m} & R_{m-m} \\ & & \end{bmatrix} \quad \text{(XXXVIII)}$$

where submatrix R_{p-p} stands for the matrix associated with the interactive abilities of all proteins as Equation XXXIX.

$$\begin{bmatrix} \widehat{\alpha_{11}} & \cdots & \widehat{\alpha_{1w}} & \cdots & \widehat{\alpha_{1W}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\alpha_{k1}} & \cdots & \widehat{\alpha_{kw}} & \cdots & \widehat{\alpha_{kW}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\alpha_{K1}} & \cdots & \widehat{\alpha_{Kw}} & \cdots & \widehat{\alpha_{KW}} \end{bmatrix} \quad \text{(XXXIX)}$$

R_{t-g}, R_{t-m}, and R_{t-l} stand for the matrices associated with transcriptional regulatory abilities of TFs on genes, miRNAs, and lncRNAs, respectively, as Equation XL.

$$\begin{bmatrix} \widehat{\delta_{11}} & \cdots & \widehat{\delta_{1x}} & \cdots & \widehat{\delta_{1X}} \\ \vdots & \ddots & \vdots & & \vdots \\ \widehat{\delta_{l1}} & \cdots & \widehat{\delta_{lx}} & \cdots & \widehat{\delta_{lX}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\delta_{L1}} & \cdots & \widehat{\delta_{Lx}} & \cdots & \widehat{\delta_{LX}} \end{bmatrix} \begin{bmatrix} \widehat{\eta_{11}} & \cdots & \widehat{\eta_{1x}} & \cdots & \widehat{\eta_{1X}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\eta_{m1}} & \cdots & \widehat{\eta_{mx}} & \cdots & \widehat{\eta_{mX}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\eta_{M1}} & \cdots & \widehat{\eta_{Mx}} & \cdots & \widehat{\eta_{MX}} \end{bmatrix} \begin{bmatrix} \widehat{o_{11}} & \cdots & \widehat{o_{1x}} & \cdots & \widehat{o_{1X}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{o_{n1}} & \cdots & \widehat{o_{nx}} & \cdots & \widehat{o_{nX}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{o_{N1}} & \cdots & \widehat{o_{Nx}} & \cdots & \widehat{o_{NX}} \end{bmatrix} \quad \text{(XL)}$$

R_{l-g} R_{l-m}, and R_{l-l} stand for the matrices associated with transcriptional regulatory abilities of LncRNAs on genes, miRNAs, and lncRNAs, respectively, as shown in Equation XLI.

$$\begin{bmatrix} \widehat{\zeta}_{11} & \cdots & \widehat{\zeta}_{1y} & \cdots & \widehat{\zeta}_{1Y} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\zeta}_{l1} & \cdots & \widehat{\zeta}_{ly} & \cdots & \widehat{\zeta}_{lY} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\zeta}_{L1} & \cdots & \widehat{\zeta}_{Ly} & \cdots & \widehat{\zeta}_{LY} \end{bmatrix} \begin{bmatrix} \widehat{\rho}_{11} & \cdots & \widehat{\rho}_{1y} & \cdots & \widehat{\rho}_{1Y} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\rho}_{m1} & \cdots & \widehat{\rho}_{my} & \cdots & \widehat{\rho}_{mY} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\rho}_{M1} & \cdots & \widehat{\rho}_{My} & \cdots & \widehat{\rho}_{MY} \end{bmatrix} \begin{bmatrix} \widehat{\kappa}_{11} & \cdots & \widehat{\kappa}_{1y} & \cdots & \widehat{\kappa}_{1Y} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\kappa}_{n1} & \cdots & \widehat{\kappa}_{ny} & \cdots & \widehat{\kappa}_{nY} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\kappa}_{N1} & \cdots & \widehat{\kappa}_{Ny} & \cdots & \widehat{\kappa}_{NY} \end{bmatrix} \quad \text{(XLI)}$$

R_{m-g}, R_{m-m}, and R_{m-l} stand for the matrices associated with transcriptional regulatory abilities of miRNAs on genes, miRNAs, and lncRNAs, respectively, as shown in Equation XLII.

$$\begin{bmatrix} \widehat{\omega}_{11} & \cdots & \widehat{\omega}_{1z} & \cdots & \widehat{\omega}_{1Z} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\omega}_{l1} & \cdots & \widehat{\omega}_{lz} & \cdots & \widehat{\omega}_{lZ} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\omega}_{L1} & \cdots & \widehat{\omega}_{Lz} & \cdots & \widehat{\omega}_{LZ} \end{bmatrix} \begin{bmatrix} \widehat{\lambda}_{11} & \cdots & \widehat{\lambda}_{1z} & \cdots & \widehat{\lambda}_{1Z} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\lambda}_{m1} & \cdots & \widehat{\lambda}_{mz} & \cdots & \widehat{\lambda}_{mZ} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\lambda}_{M1} & \cdots & \widehat{\lambda}_{Mz} & \cdots & \widehat{\lambda}_{MZ} \end{bmatrix} \begin{bmatrix} \widehat{\tau}_{11} & \cdots & \widehat{\tau}_{1z} & \cdots & \widehat{\tau}_{1Z} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\tau}_{n1} & \cdots & \widehat{\tau}_{nz} & \cdots & \widehat{\tau}_{nZ} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \widehat{\tau}_{N1} & \cdots & \widehat{\tau}_{Nz} & \cdots & \widehat{\tau}_{NZ} \end{bmatrix} \quad \text{(XLII)}$$

In the network matrix R in Equation XXXVIII or real GWGENs of CHP and non-CHP, if an interaction between any two proteins or regulation between any two elements such as TFs, genes, miRNAs, and lncRNA is deleted by AIC in Equations XXII-XXV, its corresponding component in R will be reduced to zero. To employ KEGG pathways annotation for the core signaling pathways, real GWGEN should be truncated to a core GWGEN with 6,000 significant components based on their significance from the energy perspective. Based on singular value decomposition, the network structure projection method, PNP, can be described with Equation XXLIII:

$$R = \Lambda \Sigma \Theta^T \quad \text{(XLIII)}$$

where $R \epsilon \mathbb{R}^{(K+L+M+N)*(X+Y+Z)}$, $\Lambda \epsilon \mathbb{R}^{(K+L+M+N)*((K+L+M+N)}$, $\Theta \epsilon \mathbb{R}^{(X+Y+Z)*(X+Y+Z)}$, and Σ is a diagonal matrix (i.e. $\Sigma = \text{diag}[\sigma_1,\sigma_2,\ldots,\sigma_i,\ldots,\sigma_{X+Y+Z}]$), including X+Y+Z non-negative singular values of R with descending order $\sigma_1 \geq \ldots \geq \sigma_I \geq \ldots \geq \sigma_{X+Y+Z} \geq 0$. In this context, $\text{diag}(\sigma_1, \sigma_2)$ indicates the diagonal matrix of σ_1 and σ_2 (i.e., $\begin{matrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{matrix}$). The eigenvalue expression (network energy) fraction (E_I) can be defined by the normalization in Equation XLIV.

$$E_I = \frac{\Sigma_{i=1}^{I} \sigma_i^2}{\Sigma_{c=1}^{X+Y+Z} \sigma_c^2} \geq 0.85 \quad \text{(XLIV)}$$

We selected the top I singular vectors of Θ, such that the $\Sigma_{i=1}^{I} \sigma_i \geq 0$ accounts for at least 85% of the total energy. The minimal I was chosen to ensure this proportion. These selected singular vectors are then used to construct the principal network structure, which captures 85% of the network's energy. Subsequently, the projection of each row in R onto the top I singular vectors is performed. This means that all edges of each node (i.e., each protein, gene, miRNA, and lncRNA) in real GWGENs should be projected to the top I singular vectors as the following equation (Equation XLV):

$$Project_R^{(b,i)} = \mathrm{r_b} \cdot \theta_i^T$$

for $b = 1,2,\ldots,K + L + M + N$ and $i = 1,2,\ldots,I$ (XLV)

where $\mathrm{r_b}$ represents the b-th row vector of R and i denotes the i-th column vector of Θ, which is the right-singular vector of Θ. We further define the 2-norm projection value of each node to the top I right singular vectors using Equation XLVI.

$$D_R^{(b)} = \left[\sum_{i=1}^{I} (project_R^{(b,i)})^2 \right]^{\frac{1}{2}}$$

for $b = 1,2,\ldots, K + L + M + N$ (XLVI)

In conclusion, we can obtain the projection values of all nodes, including proteins, genes, miRNAs, and

lncRNAs in real GWGENs. A projection value D_R (b) close to zero indicates that the corresponding node is almost independent of the top *I* right-singular vectors. In contrast, if a node of the real GWGEN has a higher projection value, it suggests that the node plays a more significant role in the principal network structure of GWGEN from an energy perspective. Since this study aims to compare the core signaling pathways of CHP and non-CHP lung slice cells to investigate the pathogenetic mechanisms of CHP, we can identify the top 6,000 proteins, TFs, genes, miRNAs, and lncRNAs of real GWGEN of both CHP and non-CHP. These core molecules can then be used for pathway annotation through KEGG pathways. The identified proteins, TFs, genes, miRNAs, and lncRNAs form the core signaling pathways of CHP and non-CHP, as depicted in Figure 3, which will be essential for further investigations into the pathogenetic mechanisms of CHP.

2.6. Candidate drugs predictions to alleviate CHP symptoms using a DNN-based DTI model

Based on the core signaling pathways and downstream cellular dysfunctions of CHP and non-CHP, as illustrated in Figure 4, significant biomarkers of pathogenesis were identified as drug targets for the treatment of CHP patients. Using DTI data from DTI databases such as KEGG, UniProt,[21] PRISM, DrugBank,[22] PubChem[23], and ChEMBL,[24] a DNN-based DTI model was trained to predict candidate molecular drugs for these important biomarkers. The flowchart of the DNN-based DTI model is shown in Figure 5. After identifying candidate molecular drugs for targeting CHP and considering their regulatory ability, sensitivity, and toxicity as design considerations and selection criteria for potential multi-molecule drugs of CHP, we proceeded with drug rediscovery and synthetic design.

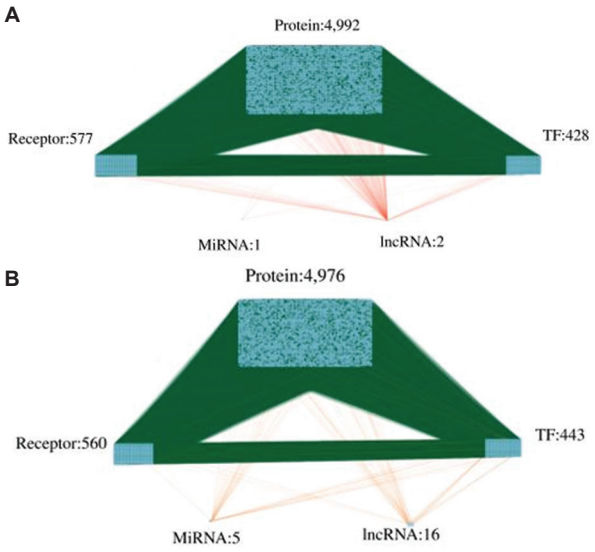

Figure 3. (A) The core GWGEN of CHP and (B) the core GWGEN of non-CHP. The core GWGENs were extracted using the principal network projection method from the real GWGEN to simplify the annotation of core signaling pathways using the Kyoto Encyclopedia of Genes and Genomes pathways to investigate the pathogenetic mechanism of CHP. The numbers indicate the node numbers of proteins, TFs, receptors, lncRNAs, and miRNAs. The green lines represent the protein-protein interactions, and the orange lines represent the gene regulations.
Abbreviations: CHP: Chronic hypersensitivity pneumonitis; GWGEN: Genome-wide and epigenetic interaction networks; lncRNA: Long non-coding RNA; TF: Transcription factor.

We preprocessed the DTI data before training the DNN-DTI model using the DTI data from the databases. We collected relevant data from DTI databases, including KEGG, UniProt,[21] PRISM, DrugBank,[22] PubChem,[23] and ChEMBL.[24] To input the data into the DNN-based DTI model, we used the PROFEAT website and PyBioMed tool in the Python 3.7 environment to convert the DTI data into feature vectors with drug-target pairs. For a drug-target pair, the feature vector can be represented as presented in Equation XLVII[19]:

$$P_{drug-test} = \left[D,T\right] = [d_1,d_2,\ldots\ldots,d_{M-1},t_1,t_2,\ldots\ldots t_{N-1}] \quad \text{(XLVII)}$$

where $P_{drug-test}$ is the drug-target pair in feature vector form. *D* represents the feature vector of the corresponding drug, and *T* represents the feature vector of the drug target (biomarker). *M* is the total number of drug features, and *N* is the total number of drug target features. Since most of the original training data had unknown interactions or negative data, the next step was data preprocessing. To address the class imbalance issue, we reduced the amount of unknown interaction data. Then, to account for variations in units across different features, we standardized each feature vector and normalized their significance. The mathematical formulas (Equations XLVIII and XLIX) for drug and target feature normalization are presented[19]:

$$d_i^* = \frac{d_i - \mu_i}{\sigma_i}, \forall i = 1,2,\ldots,M \quad \text{(XLVIII)}$$

$$t_j^* = \frac{t_j - \mu_j}{\sigma_j}, \forall j = 1,2,\ldots,N \quad \text{(XLIX)}$$

where d_i denotes the *i*-th drug feature and d_i^* expresses the *i*-th drug feature after the standardization. $\propto_i$ and σ_I, respectively, denote the mean and standard deviations of the *i*-th drug feature. t_j represents the *j*-th feature of the targiet and t_j^* represents the *j*-th feature of the target after standardization. j and j separately indicate the mean and standard deviation of the *j*-th target feature. *M* expresses the total number of drug features, and *N* denotes the total number of target features.

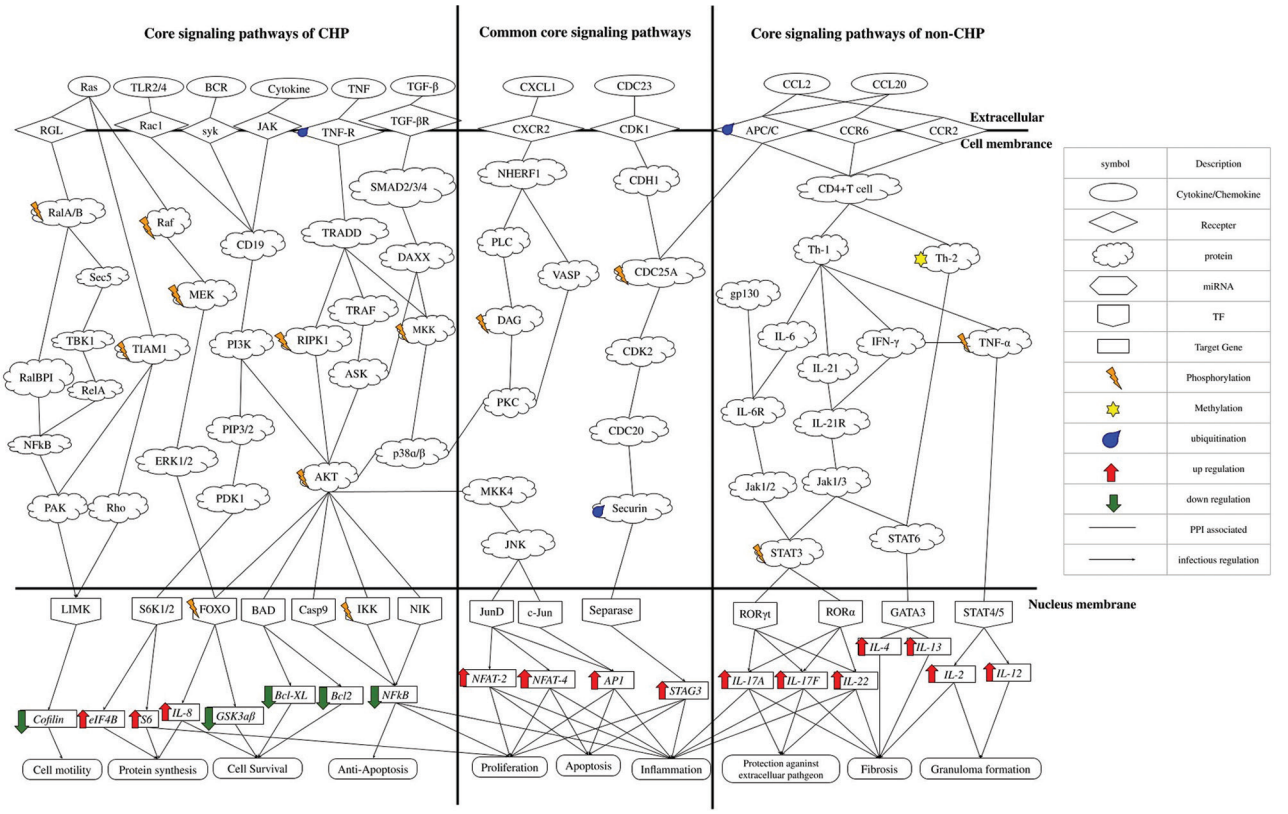

Figure 4. This figure provides annotations of the core genome-wide and epigenetic interaction networks through the Kyoto Encyclopedia of Genes and Genomes pathways, highlighting the specific signaling pathways associated with chronic hypersensitivity pneumonitis (CHP) (shown on the left) and healthy controls (shown on the right). The middle area contains overlapping signaling pathways of both CHP and non-CHP. This figure shows that the transforming growth factor *β*, tumor necrosis factor, and p38-mitogen-activated protein kinase signaling pathways contribute to pneumonitis fibrosis and the chemokine (C-X-C motif) ligand 1, cell division cycle 23, chemokine (C-C motif) ligand 2, and chemokine (C-C motif) ligand 20 signaling pathways are involved in cell apoptosis and proliferation. Based on the pathogenetic mechanisms of CHP, the significant biomarkers above are selected as drug targets for the treatment of CHP.

Since the DNN-based DTI model is designed with only 996 inputs, the dimensions of the drug-target feature vectors in Equations XLVII-XLIX must the reduced to 996 so that they could be inputted into the model. To achieve this reduction, we downsampled the feature vectors of the drug-target pair using the principal component analysis method to reduce their dimension to 996. After data preprocessing, we use 75% of the data as training data and 25% of the data as testing data. Furthermore, we split 20% of the training data as validation data to do the five-fold cross-validation. The DNN-based DTI model in Figure 5 has four hidden layers containing 512, 256, 128, and 64 neurons sequentially. A binary cross entropy method is used as the loss function, with a learning rate of 0.003, and the backward propagation algorithm was used as the optimizer. We set the number of epochs to 100 and the batch size to 100. To prevent the vanishing gradient problem, each hidden layer employed a rectified linear unit activation function. The five-fold cross-validation method is employed to learn data and check the stability and prediction performance of the DNN-based DTI model. The early-stopping approach is applied to avoid overfitting. The activation function of the output layer is a sigmoid function to limit the output value to a range between 0 and 1, as the probability predicted by the DNN-based DTI model. Since DTI is a binary classification problem, the model loss is computed using the binary cross-entropy cost function, as shown in Equations L and LI[19]:

$$C_n(w,b) = -[\widehat{p_n}\log(p_n) + (1-\widehat{p_n})\log(1-p_n)] \quad \text{(L)}$$

$$L(w,b) = \frac{1}{N}\sum_{n=1}^{N} C_n(w,b) \quad \text{(LI)}$$

where p_n denotes the truth label of positive interaction, $\widehat{p_n}$ denotes the predictive probability of positive interaction, and (1- $\widehat{p_n}$) represents the predicted probability of negative

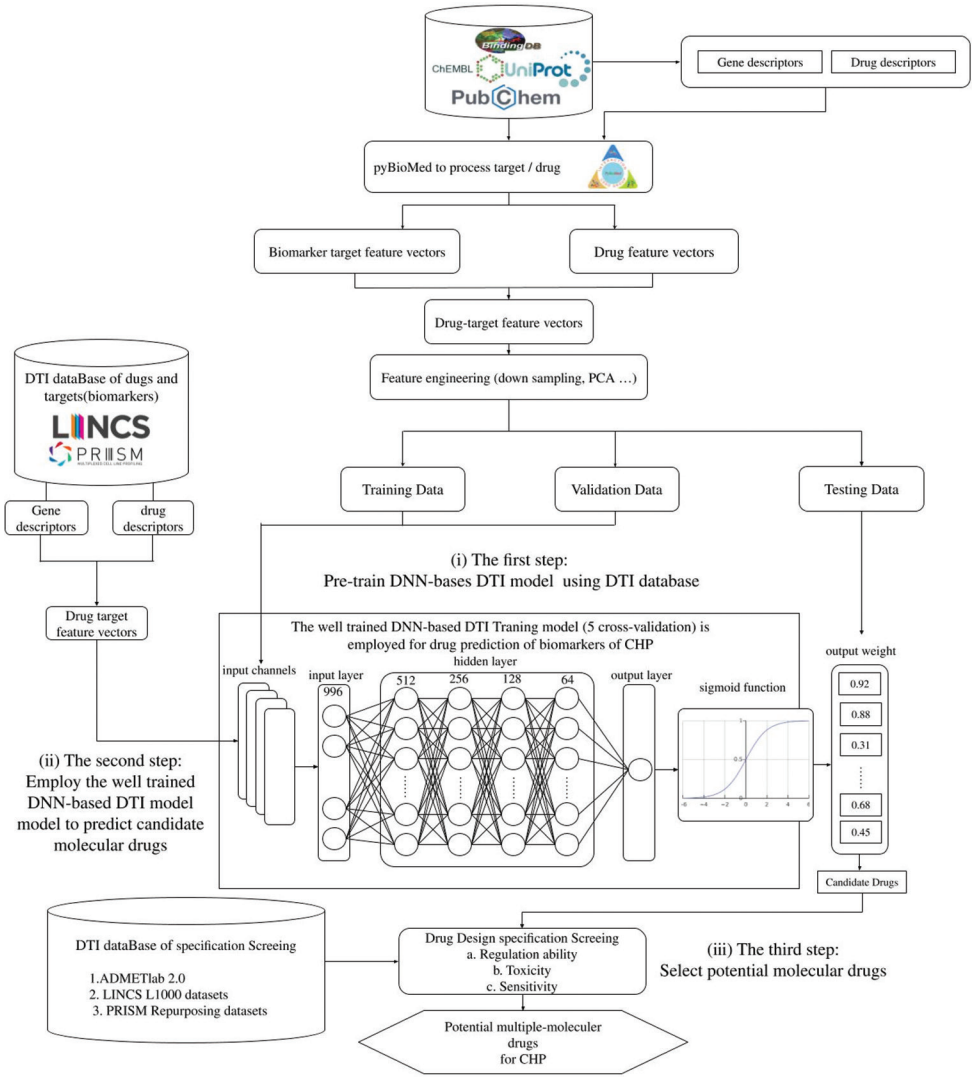

Figure 5. The flowchart of molecular drug prediction process for drug targets (biomarkers) of chronic hypersensitivity pneumonitis. The drug prediction process is divided into three steps. The first step employs DTI data from DTI databases to train the DNN-based DTI model. The second step uses the trained DNN-based DTI model to predict candidate interactive drugs for these drug targets (biomarkers). The third step uses three design specifications to screen potential molecular drugs from candidate interactive drugs for chronic hypersensitivity pneumonitis.
Abbreviations: CHP: Chronic hypersensitivity pneumonitis; DNN: Deep neural network; DTI: Drug-target interaction; PCA: Principal component analysis.

interaction. L(w,b) denotes the average total loss $C_n(w,b)$. According to the cost function, the backward propagation algorithm is applied to update the model parameter set Θ containing the weighting matrix weight and the bias vector by calculating the gradient of the cost function in Equation LI and eventually obtaining the optimal solution Θ*, as in Equations LII-LIV:

$$\Theta = \begin{bmatrix} weight \\ bias \end{bmatrix} \quad \text{(LII)}$$

$$\Theta^{*} = \arg\min L(\Theta) \quad \text{(LIII)}$$

$$\Theta^{e} = \Theta^{e-1} - \eta \Delta L(\Theta^{e-1}) \quad \text{(LIV)}$$

where e is the e-th epoch of the learning procedure, w is the weights, b is bias, Θ is the learning rate, and $\Delta L(\Theta^{e-1})$ is the gradient of $L(\Theta^{e-1})$.

Based on the backward propagation method, the DNN-based DTI model can adjust parameters during each iteration to better fit the DTI data. Moreover, adjusting hyperparameters not only shortens the training time but also achieves the optimal model performance. We used the default optimizer settings and set the learning rate to 0.003 to make the model parameters Θ converge faster and more accurately. We set the number of epochs to 100

and the batch size N to 100. For the data, we used 25% as test data and 75% as training data. In addition, we divided the training data into five equal parts to implement a five-fold cross-validation strategy. Of the five-fold data, four-fifths were used for DNN-based DTI model training, and one-fifth served as the validation data to monitor the model's performance, ensuring it improved over previous epochs. The results of training loss and training accuracy are shown in Figures 6 and 7, respectively. Furthermore, five-fold cross-validation can verify the stability of the data and model. To avoid model overfitting during the training process in Equation LIV, we applied an early stopping strategy to check if the test accuracy decreased while training accuracy continued to increase. Moreover, we embedded dropout layers after each hidden layer to further prevent model parameter overfitting and set the dropout rate to 0.4. After training the DNN-based DTI model in Figure 5, we adopted the performance measures area under the curve (AUC) score and receiver operating characteristic (ROC) curve shown in Figure 8 to check the DNN-based DTI model. The AUC score is a valuable evaluation metric for classification problems, where a higher AUC score (indicating a larger area under the line) reflects better model accuracy in predicting true positive and true negative DTIs. The formulas for the AUC score and ROC curve are presented in Equations LV-LVII[25]:

$$\text{TPR (True Positive Rate)} = \frac{TP}{TP + FN} \quad \text{(LV)}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad \text{(LVI)}$$

$$\text{FPR (False Positive Rate)} = \frac{FP}{TP + FN} \quad \text{(LVII)}$$

where TP represents the correct positive predictions, TN represents the correct negative predictions, FP denotes false positives, and FN denotes false negatives.

The DNN-based DTI model predicted six candidate molecular drugs for CHP biomarkers (Table 1). These were evaluated based on drug design specifications, including regulatory ability, sensitivity, and toxicity, to identify potential drugs for lung fibrosis. Regulatory ability data from the L1000 Phase 5 dataset informed drug selection based on the gene expression of biomarkers – negatively correlated drugs were chosen for upregulated biomarkers and positively correlated drugs for downregulated biomarkers.[26] Sensitivity values near zero, derived from the PRISM repurposing dataset, indicated minimal cellular sensitivity to selected drugs. Toxicity was assessed using ADMETlab 2.0,[27] referencing lethal concentration 50 (LC_{50}) values to prioritize low-toxicity compounds. Based on these criteria, three molecular drugs were proposed as combined multiple molecular drugs for CHP treatment, as shown in Table 2. This combination represents a promising therapeutic approach for CHP.

3. Results

3.1. Overview of the systems biology method and the systematic drug discovery and design for the treatment of CHP

To gain a deeper understanding of the carcinogenic mechanism and identify key carcinogenic biomarkers as drug targets for CHP, we employed a systems biology approach using corresponding microarray data, a DNN-based DTI model trained using DTI databases and a drug screening by drug design specifications to find potential molecular drugs targeting these crucial biomarkers for the treatment of CHP.

The first step involves constructing candidate GWGEN for non-CHP and CHP by mining large datasets from databases such as DIP,[9] IntAct,[11] BioGRID,[8] MINT,[10] HTRIdb,[13] ITFP,[14] TRANSFAC,[7] CircuitsDB,[15] TargetScanHuman,[12] and starBase 2.0.[16] Next, using the system identification methods in Equations I-XXI and the system order detection methods in Equations XXII-XXXIII with microarray data for non-CHP and CHP, we constructed the real GWGENs for non-CHP and CHP, respectively, as shown in Figure 2, by eliminating false positives from the candidate GWGEN.

Since only up to 6,000 molecules in the real GWGENs can be annotated with KEGG pathways, the PNP method in Equations XLIV-XLVI was applied to extract the core GWGEN of CHP and non-CHP, with up to 6,000 key nodes (Figure 3). These core networks highlight the numbers of proteins, TFs, receptors, lncRNAs, and miRNAs within the core GWGEN. The core signal pathways for non-CHP and CHP are constructed by mapping the core GWGEN onto the significant KEGG pathways, as shown in Figure 4.

By analyzing downstream cellular dysfunctions, such as over-expression of apoptosis and proliferation in Figure 4, the pathogenic mechanisms of pulmonary cell fibrosis in CHP were studied, and significant biomarkers were identified as drug targets. In addition, the DNN-based DTI model, trained on DTI interaction data from databases (Figure 5), predicts candidate molecular drugs for these biomarkers. These candidates are evaluated based on drug design specifications, such as adequate regulatory capacity (restoring biomarker expression to normal levels),

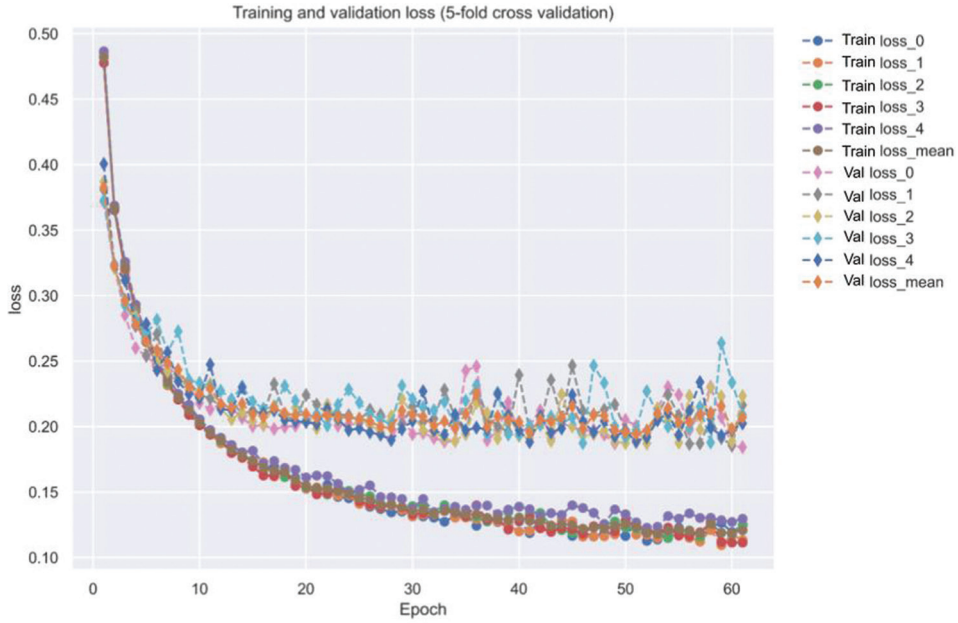

Figure 6. Training loss and validation loss of the deep neural network-based drug-target interaction model (five-fold cross-validation)

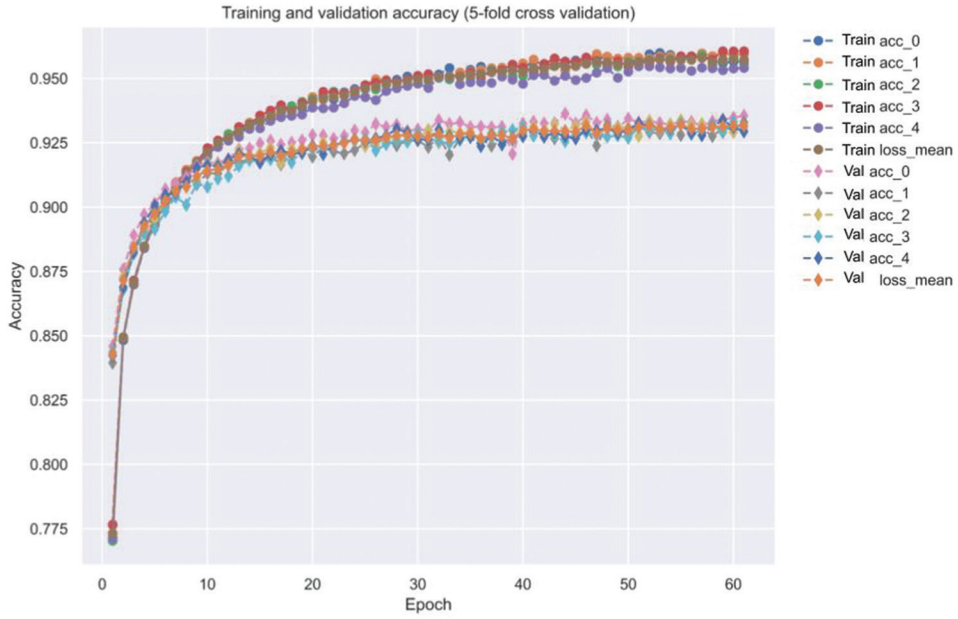

Figure 7. Training accuracy and validation accuracy of the deep neural network-based drug-target interaction model (five-fold cross-validation)

low toxicity, and high sensitivity, to identify potential molecular drugs for CHP treatment.

The investigation of core signaling pathways of CHP and non-CHP, along with their downstream target genes and cellular dysfunctions to uncover the pathogenetic mechanisms of CHP, in which significant biomarkers were identified as drug targets for the treatment of CHP is given in the following subsections.

3.2. The roles of mitogen-activated protein kinase and phosphoinositide 3-kinase signaling pathways in CHP pathogenesis

From the core signaling pathways of CHP in Figure 4, the MAPK signaling cascade emerges as a central regulatory node in the pathogenesis of CHP, orchestrating inflammatory responses, oxidative stress, apoptosis, and fibrosis. Our

Table 1. The regulation ability, toxicity, and sensitivity information of candidate drugs

Biomarker	Candidate drug	Regulation ability (L1000)	Sensitivity (PRISM)	Toxicity (LC_{50}, mL/kg)
AKT Serine/Threonine Kinase 1 (up-regulation)	Azathioprine	0.108309	−0.86865	4.558
	Clobenpropit	0.264513	0.349028	4.223
	Primaquine	0.64545	−0.34658	3.236
Chemokine (C-C motif) ligand 20 (down-regulation)	Zosuquidar	−0.55168	−0.02332	7.732
	Masitinib	−0.81919	0.034718	5.465
	Mycophenolate-mofetil	−0.61481	−0.70094	4.33
Cell division cycle 23 (down-regulation)	Azathioprine	−0.92617	−0.86865	4.558
	Masitinib	−0.20676	0.034718	5.465
	Elesclomol	−0.5743	−2.86051	4.321
Chemokine (C-X-C motif) ligand 1 (down-regulation)	Azathioprine	−0.9438	−0.86865	4.558
	Masitinib	−0.21064	0.034718	5.465
	Primaquine	−0.14379	−0.34658	3.236
Nuclear factor kappa B subunit 1 (up-regulation)	Imatinib	0.224935	−0.80213	5.419
	Masitinib	0.21467	0.034718	5.465
	Primaquine	0.310101	−0.34658	3.236
Tumor necrosis factor (down-regulation)	Azathioprine	−0.54562	−0.86865	4.558
	Masitinib	−0.19718	0.034718	5.465
	Mycophenolate-mofetil	−0.59299	−0.70094	4.33

Abbreviation: LC_{50}: Lethal concentration 50.

Table 2. A potential multi-molecular drug for chronic hypersensitivity pneumonitis

Target Drug	AKT1 (↑)	CCL20 (↓)	CDC23 (↓)	CXCL1 (↓)	NFKB1 (↑)	TNF (↓)
Azathioprine	✓		✓	✓		✓
Masitinib		✓	✓	✓	✓	✓
Primaquine	✓			✓	✓	

Drug	Sensitivity	Toxicity
Azathioprine	−0.86865	4.558
Masitinib	0.034718	5.465
Primaquine	−0.34658	3.236

Note: ✓ denotes the existence of interaction between the molecular drugs and biomarkers, ↑ denotes up-regulation, and ↓ denotes down-regulation.

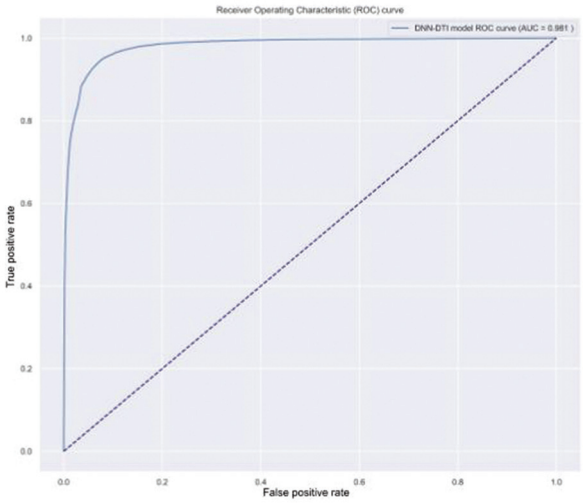

Figure 8. The receiver operating characteristic curve could measure the probability of the prediction accuracy of the deep neural network-based drug-target interaction model, with an area under the curve of the receiver operating characteristic score of 0.981.
Abbreviation: AUC: Area under the curve; DNN: Deep neural network; DTI: Drug-target interaction.

analysis revealed aberrant activation of the three major MAPK pathways – extracellular signal-regulated kinase (ERK), p38, and Jun N-terminal kinase (JNK) – in lung tissue specimens from CHP patients.[28] The p38 MAPK pathway plays a pivotal role in driving the inflammatory milieu.[29,30] Phosphorylation of p38α and p38β isoforms leads to the activation of key transcription factors, such as NF-κB, which subsequently induces the transcription of pro-inflammatory cytokines (e.g., TNF-α) and chemokines.[30] In addition, p38 MAPK signaling mediates transforming growth factor-beta (TGF-β)-induced myofibroblast differentiation and extracellular matrix production, contributing to the fibrotic process.[31]

On the left side of Figure 4, the activation of toll-like receptors (TLR) 2 and TLR4 in the microenvironment leads to the recruitment of guanine nucleotide exchange factors, which activate receptor Ras-related C3 botulinum toxin substrate 1 (Rac1). Activated Rac1 plays a key role in cytoskeletal reorganization, reactive oxygen species (ROS) production, and downstream signaling pathways modulation. In addition, the binding of an antigen to the B-cell receptor (BCR) triggers receptor clustering and spleen tyrosine kinase (Syk) phosphorylation. Once Syk is phosphorylated and activated, it leads to the phosphorylation of downstream signaling molecules. Cytokines bind to their specific receptors, which are associated with JAKs. Subsequently, a cluster of differentiation 19 (CD19) enhances BCR signaling by providing docking sites for signaling molecules that amplify Syk activation. Activated Rac1 promotes cytoskeletal reorganization, which is important for various B-cell functions, while simultaneously activating PI3K. Although CD19 does not directly interact with JAKs, cytokine signaling through JAK-STAT pathways can modulate B-cell responses.[32] Enhanced CD19 signaling may lead to the activation of PI3K. When upstream signals activate PI3K, it produces phosphatidylinositol 4,5-bisphosphate and phosphatidylinositol (3,4,5)-trisphosphate (PIP3) at the cell membrane. PIP3 serves as a binding site for proteins with Pleckstrin homology domains, including phosphoinositide-dependent kinase-1 (PDK1).[33] This recruitment brings PDK1 to the cell membrane, where it can phosphorylate its substrate kinases. Once PDK1 is activated, it phosphorylates TF S6 kinase 1 (S6K1) and S6K2 at a critical threonine residue, which is essential for their activation. Activated TF S6K1/2 phosphorylates our target genes *S6* and *EIF4B*, promoting the translation of mRNAs involved in ribosome biogenesis and protein synthesis.[34]

For the other cytokine proteins in the CHP microenvironment, mutations in Ras or rapidly accelerated fibrosarcoma (Raf) can lead to uncontrolled cell proliferation and cancer. Once activated, Raf phosphorylates and activates MEK, which, in turn, phosphorylates and activates ERK. Activated ERK translocates to the nucleus and regulates the expression of upstream genes involved in cell proliferation, differentiation, and survival.[35] The phosphorylation of ERK1/2 inactivates TF forkhead box O (FOXO) and contributes to the inflammatory response by enhancing the production of pro-inflammatory mediators, such as interleukin-8 (IL-8) chemokine, which subsequently influences protein synthesis pathways. Moreover, ERK signaling promotes lung epithelial cell proliferation and activates pro-fibrotic pathways, including TGF-β signaling.[36] When Ras is activated guanosine triphosphate (GTP)-bound state, it can activate T-cell lymphoma invasion and metastasis-inducing protein 1 (TIAM1), which activates the phosphorylation of p21-activated kinases (PAKs) to cause the phosphorylation of TF LIM domain kinase 1, subsequently inactivating cofilin that affects cytoskeleton dynamics and cell motility. TIAM1 also facilitates Rho GTPase activation. Consequently, it regulates the formation of stress fibers and focal adhesions, contributing to cell contraction and adhesion. Ral guanine nucleotide dissociation stimulator-like (RGL) proteins are direct downstream effectors of Ras. Once activated by Ras, RGL proteins facilitate the exchange of guanosine diphosphate for GTP on RalA and RalB, thus activating these GTPases.[37] Then, Sec5 interacts with RalA and RalB, acting as an effector that links them to the exocytosis complex and thereby influencing vesicle trafficking and exocytosis. Sec5 can also recruit TANK-binding kinase 1 (TBK1) to the exocytosis complex, where TBK1 can be activated. This recruitment is thought to play a role in the spatial regulation of TBK1 activity, integrating signals from vesicle trafficking with innate immune responses.[38] TBK1 can directly phosphorylate components of the NF-κB pathway, including RelA, enhancing its transcriptional activity. Collectively, the sustained MAPK pathway activation perpetuates lung inflammation, oxidative injury, apoptosis of epithelial and endothelial cells, and the accumulation of fibroblasts and myofibroblasts, thereby driving disease symptoms in CHP.[39] In fibroblasts and myofibroblasts, PI3K/AKT signaling enhances proliferation, migration, and resistance to apoptosis, upregulates the expression of extracellular matrix proteins (e.g., collagens and fibronectin), and potentiates TGF-β-induced myofibroblast activation and differentiation.[40]

In summary, the activation of the cytokine TLR2/4 and BCR in the PI3K signaling pathway causes the phosphorylation of our target genes, *S6* and *EIF4B*, through a series of signaling protein transduction mechanisms, which consequently enhances protein synthesis and proliferation. The Ras protein upstreams the target gene *IL8* cytokine through ERK signaling transduction and has an effect on cell survival and cell motility through GTPase activity. The dysregulated PI3K/AKT cascade promotes aberrant epithelial repair, fibroblast activation and proliferation, and persistent inflammation, collectively contributing to the initiation and symptom of pulmonary fibrosis in CHP patients.

3.2.1. The role of transforming growth factor-β signaling pathways in the pathogenesis of CHP

The TGF-β signaling pathway plays a pivotal role in the pathogenesis of CHP, a lung disease characterized by

inflammation and fibrosis. The canonical TGF-β signaling cascade is initiated when the TGF-β ligand binds to the type II TGF-β receptor (TGF-βR), which then recruits and forms a heteromeric complex with the type I TGF-βR.[41] This complex undergoes auto-phosphorylation of serine/threonine residues, allowing TGF-R1 to phosphorylate the suppressor of Mothers against decapentaplegic (SMAD)2 and SMAD3 proteins. The activated SMAD proteins dissociate from the SMAD anchor for receptor activation and oligomerize with SMAD4. This SMAD complex translocates to the nucleus, where it interacts with various transcriptional co-regulators and other factors, modulating gene expression like *DAXX*.[42]

In normal epithelial cells, TGF-β signaling is typically considered an anti-proliferative and pro-apoptotic signal. However, a critical step in the dysregulation of TGF-β signaling in CHP is the loss of this response. TGF-β can exhibit pro-survival and anti-apoptotic effects, leading to mitochondrial dysfunction through SMAD-dependent activation.[43] This subsequently triggers the death-associated protein 6 (DAXX)-mediated transformation growth factor, TNF receptor-associated factor (TRAF)-dependent p38/JNK activation through non-SMAD mechanisms.[44] TRAF proteins facilitate the activation of apoptosis signal-regulating kinase, which, in turn, modulates downstream signaling pathways, including MAPKs and PI3K-AKT. These interactions are crucial for regulating cellular responses to stress, inflammation, and apoptosis, resulting in JNK-mediated cell death. DAXX can modulate the activity of MAPK kinases (MKKs), influencing downstream MAPK signaling pathways involved in cellular processes such as stress response and programmed cell death.[45] In addition to the canonical SMAD pathway, TGF-β signaling can activate the MAPK and PI3K signaling pathways, contributing to the pathogenesis of CHP. The activation of these pathways can promote cell survival, proliferation, and migration, thus exacerbating disease indication and the fibrotic response in the lungs. In the process of epithelial-mesenchymal transition (EMT), a crucial event in fibrosis, TGF-β-regulated fibroblast lineage reprogramming, and cell migration depend on the p38 MAPK signaling pathway. The kinetics and functions of the ERK/MAPK signaling pathway in response to TGF-β stimulation are tissue-specific, varying in epithelial cells, fibroblasts, and cancer cells. The downstream signaling induced by the binding of TGF-β to its serine/threonine kinase receptors is also mediated by SMAD family members.[42] For example, insulin signaling, including the activation of its downstream effector PI3K and AKT, can inhibit TGF-β-mediated apoptosis.[46] Furthermore, the exogenous TGF-β can suppress cytokine-induced human B-cell proliferation, suggesting the potential sensitivity of germinal center B cells to TGF-β and its role in limiting the proliferation of fibroblastic cells in the lung and promoting apoptosis.[46]

Overall, microbes such as *Mycobacterium* spp., *Aspergillus fumigatus*, *Streptomyces* spp., and *Pseudomonas aeruginosa* can interact with AKT1 signaling pathways, contributing to chronic inflammation and immune dysregulation seen in CHP. Their interaction with AKT1 helps sustain inflammation, prevent apoptosis, and promote tissue fibrosis, which are key pathological features of CHP. SMAD proteins act as messengers that relay the signal from the extracellular TGF-β molecule to the interior of the cell, where it can influence various cellular processes and have a crucial effect on the AKT1 signal pathway, including those involved in the pathogenesis of CHP.

3.2.2. The role of tumor necrosis factor signaling pathway in the pathogenesis of CHP: A multifaceted regulator

The TNF signaling pathway is a crucial mediator of various physiological and pathological processes, including cell proliferation, differentiation, apoptosis, immune response modulation, and inflammation induction.[47] Microbes such as thermophilic actinomycetes, *Aspergillus fumigatus*, and endotoxin-producing bacteria are key organic agents that influence TNF responses in CHP. They are commonly found in environments related to farming, moldy buildings, and bird keeping. Chronic exposure to these antigens leads to persistent TNF-mediated inflammation, driving the progression of the disease. TNF, a multifunctional pro-inflammatory cytokine, influences diverse processes, such as lipid metabolism, coagulation, insulin resistance, and endothelial function.[48] Initially discovered as a potential anticancer agent, TNF has since been recognized for its broader roles in cellular signaling. The TNF signaling cascade is initiated by the binding of cytokine TNF to its two receptors, TNF-R1 and TNF-R2. TNF-R1 is ubiquitously expressed in all human tissues and is considered the main signaling receptor for TNF-α, while TNF-R2 is primarily expressed in immune cells and mediates limited biological responses. TNF-R2 can bind to both TNF-α and TNF-β.[49] On TNF-α binding to TNFR1, a rapid formation of a cell-internal protein complex (termed complex I) occurs, comprising receptor-interacting protein kinase 1 (RIPK1), TNF-R1 associated death domain (TRADD), and other signaling molecules. This protein complex activates the expression of inflammatory and survival genes. TRADD acts as an adaptor molecule, binding to activated TNFR1 and recruiting additional signaling molecules into the

complex.[50] Both protein TRADD and protein RIPK1 are major participants in TNFR1 signaling, contributing to NF-κB activation and cell death signaling pathways.[51]

The importance of TRADD in regulating TNF-induced cell signaling is cell type-dependent and may be related to the amount of RIPK1 present in the corresponding cell type. TRADD and RIPK1 are also considered to have redundant or competitive activities in certain contexts. In addition to its role in regulating cell survival and death, the TNF signaling pathway also influences immune cell differentiation and function. For instance, TNF-α suppresses the differentiation of induced regulatory T cells (iTregs) through a signaling cascade involving the induction of TNF-R2 expression and the activation of AKT.[51] Activated AKT interacts with SMAD, leading to the inhibition of TGF-β-induced SMAD3 phosphorylation, which is essential for iTreg differentiation. When AKT phosphorylates TF FOXO, it causes the sequestration of FOXO, preventing TF FOXO from entering the nucleus and transcribing target genes.[52] This leads to the inhibition of TF FOXO's roles in apoptosis, cell cycle regulation, and oxidative stress resistance. Simultaneously, AKT phosphorylates TF BCL2 associated agonist of cell death (BAD) at specific serine residues.[53] This binding sequesters BAD in the cytosol, preventing target genes *BCL2* and *BCLXL,* which play critical roles in regulating cell survival by inhibiting the apoptotic pathways.[54] This inhibits the intrinsic apoptosis pathway, preventing Caspase-9 (Casp9) activation and promoting cell survival. Conversely, when AKT activity is low, TFs FOXO and BAD can promote apoptosis through mitochondrial pathways involving Casp9. The activity of the target gene *GSK3A/B* is regulated by upstream signals, particularly the PI3K/AKT signal pathway. When AKT is activated, it phosphorylates and inhibits *GSK3A/B*,[55] which subsequently promotes cell survival by reducing their pro-apoptotic activity. On the other hand, the activation of AKT stimulates the expression of gene *IL8*, which leads to protein synthesis activation. Furthermore, TNF signaling intersects with other core signaling pathways, such as the PI3K/AKT and NF-κB pathways. PI3K and AKT stimulate gene *NFKB* activation in a dose-dependent manner, suggesting a common link between these pathways. TNF-R1, NF-κB inducing kinase, and inhibitory-κB kinase partially participate in AKT-induced NF-κB stimulation, and PI3K-AKT signaling can activate NF-κB through both TNFR-dependent and independent mechanisms. This interplay between the TNF, PI3K/AKT, and NF-κB pathways may represent a mechanism by which AKT exerts its anti-apoptotic function and cell proliferation in fibrosis cells.

In summary, the TNF signaling pathway has a crucial effect on target genes like *BCL2*, *BCLXL*, *GSK3A/B*, and *NFKB* through the PI3K/AKT signal pathway. In addition, the TNF signaling pathway is a multifaceted regulator of diverse cellular processes with far-reaching implications in various physiological and pathological contexts, including inflammation and immunity in CHP.

3.3. The role of cell cycle control signaling pathways in the pathogenesis of CHP

3.3.1. The role of C-X-C motif chemokine ligand 1-C-X-C motif chemokine receptor 2 chemokines signaling axis in the pathogenesis of CHP

C-X-C motif chemokine receptor 2 is the chemokine receptor for CXCL1, which has been shown to inhibit fibrosis progression and reduce macrophage recruitment in the pneumonitis microenvironment, providing insights into potential therapeutic strategies for treating CHP patients with high CXCL1 expression. Certain organic microbes, particularly bacteria and fungi, interact with CXCR1 and CXCR2 chemokine receptors by promoting IL-8 production and neutrophil recruitment. This interaction plays a crucial role in the pathogenesis of CHP, driving the persistent inflammatory response that leads to lung damage. This chemo synapse comprises multiple proteins that can interact with CXCR2 in both its activated and inactivated conformational states, with some interactions being modulated on receptor activation. Sodium hydrogen exchanger regulatory factor-1 adaptor protein contains a PDZ motif that facilitates its binding to both CXCR2 and phospholipase C-β (PLC-β). This interaction is critical for CXCR2-mediated activation of PLC-β and subsequent downstream signaling cascades. On CXCR2 activation, protein kinase C delta (PKC-δ) phosphorylates vasodilator-stimulated phosphoprotein (VASP), enabling its binding to the receptor.[56] The phospholipase C-diacylglycerol (DAG)-PKC signaling axis is crucial for translating extracellular signals into appropriate cellular responses, allowing cells to adapt to their environment and maintain homeostasis. DAG recruits PKC isoforms to the plasma membrane. The phosphorylated VASP then associates with filamentous actin (F-actin), thereby linking the chemo synapse to the actin cytoskeleton, a process vital for chemotactic responses. In addition, PKC can act upstream of p38 MAPK. For instance, certain stress and inflammatory signals that activate PKC can also lead to the activation of p38 MAPK. This can occur through the activation of upstream kinases that phosphorylate and activate p38, so CXCR2 activation triggers the ERK p38-MAPK pathway through multiple mechanisms, contributing to various cellular processes.[57] Subsequently, MKK4 activation

typically promotes apoptosis and stress responses through JNK and p38α/β in MAPK. In contrast, AKT activation promotes cell survival and growth, also controlling gene expression related to apoptosis, survival, and inflammation. JNK phosphorylates TF c-Jun, a major component of *AP1*.[58] Phosphorylation increases c-Jun's transcriptional activity, enhancing *AP1*-mediated gene expression.[59] While TF JunD is less responsive to JNK phosphorylation compared to TF c-Jun, it can still participate in *AP1* complexes regulated by JNK activity. Moreover, JNK stress-induced signals that activate JNK can enhance the transcriptional activity of *AP1* components that incorporate TF JunD, thereby influencing regulated target genes *NFAT/AP1*.[60] The *NFAT-AP1*-JNK axis plays a critical role in T-cell activation and differentiation. The integration of signals from these pathways ensures the proper transcriptional regulation of cytokines and other immune-related genes. The activation of target gene *AP1* by JNK in response to stress can lead to the expression of inflammation and apoptosis. This response is modulated by the interaction with *NFAT*, allowing for a balanced cellular reaction to stress.[61] Consequently, CXCL1 activation subsequently induces the expression of pro-inflammatory genes, including CXCR2 ligands, and promotes cancer cell migration. CXCL1 can activate these small GTPases through the PI3K pathway, initiating downstream signaling events, such as the activation of p38 MAPK and PAKs.[62]

In summary, the CXCL1 chemokine represents a network of proteins in Figure 4 that interacts with genes *NFAT2*, *NFAT4*, and *AP1*, and TF JunD through JNK signals that orchestrate various downstream signaling pathways, including those involved in chemotaxis, inflammation, and cancer cell migration. Elucidating the intricate interplay within this chemo synapse is crucial for understanding the multifaceted roles of receptor CXCR2 in physiological and pathological processes.

3.3.2. The role of cell division cycle 23 in cell cycle regulation and potential implications in CHP

Cell division cycle 23 is a component of the tetratricopeptide repeat (TPR) subunits in the anaphase-promoting complex/cyclosome (APC/C) complex, involved in regulating mitosis in eukaryotes.[63] Receptor cyclin-dependent kinase 1 (CDK1) not only regulates the timing of APC/C activation to ensure proper mitotic entry, progression, and exit but also controls the initiation of mitosis, whereas cadherin-1 (CDH1) regulates the exit from mitosis and the maintenance of the G1 phase (cell cycle). In late mitosis and G1 phase, CDH1 becomes dephosphorylated and can bind to APC/C, thereby activating it. This leads to the ubiquitination and degradation of mitotic cyclins and other proteins, allowing the cell to exit mitosis and enter the G1 phase.[64] While not a direct substrate of CDH1, the regulation of CDK activity by CDH1 indirectly affects CDC25A function.[65] Proper degradation of cyclins and other regulators ensures that CDKs are only activated by CDC25A at the appropriate time. This process can inhibit CDC25A, preventing the activation of CDKs and halt cell cycle progression. Simultaneously, these checkpoints can influence the activity of CDH1 to ensure that mitotic regulators are properly degraded.[64] Subsequently, CDC25A removes inhibitory phosphates from threonine-14 and tyrosine-15 on CDK2.[65] This dephosphorylation is necessary for the activation of the CDK2/cyclin complexes, which also promotes the initiation of DNA replication, and ensures proper cell cycle progression and genomic stability. CDK2 activity ensures proper DNA replication, while CDC20 regulates the exit from mitosis, ensuring that cells only progress through the cell cycle when all necessary conditions are met.[66] Once CDC20 ubiquitinates Securin, the degradation of Securin releases the enzyme separase from its inhibition. Active separase can then cleave cohesin complexes, allowing the sister chromatids to separate and ensure proper cell division. During interphase and early mitosis, Securin binds to separase and inhibits its protease activity.[63] This binding prevents separase from cleaving cohesin complexes that hold sister chromatids together.[63] ^{Sep}arase is activated at the onset of anaphase and cleaves the cohesin complex to allow chromosome segregation, while *STAG3*-containing cohesin complexes maintain the cohesion of sister chromatids during the early stages of meiosis, ensuring proper alignment and recombination. Abnormal expression of target gene *STAG3* in somatic cells can interfere with the normal cohesin complex function, potentially affecting cell proliferation.[66] Therefore, *in vitro* cell experiments have shown that CDC23 can promote the proliferation of myogenic cells and significantly upregulate the differentiation of skeletal muscle satellite cells. The PPI network shows a high level of interaction between CDC23 and cell cycle pathway-related genes like *STAG3*. Subsequently, in CDC23-overexpressed differentiated myogenic cells, the level of CDC23 was significantly negatively correlated with the expression of key factors in the cell cycle pathway, suggesting that CDC23 may be involved in inhibiting cell cycle signaling pathways, thereby promoting differentiation.

The APC/C, through CDC23 and other components, regulates the cell cycle, ensuring proper cell division and function. Dysregulation in cell cycle control can influence immune cell proliferation and function, potentially impacting the inflammatory and immune responses seen in CHP. Compared to control cells, CDC23,

through the APC/C, indirectly regulates the activity of separase, which cleaves cohesin complexes like those containing the gene *STAG3*. The ubiquitination of the Securin protein, along with CDC23 overexpression in CHP, provides a molecular basis through which gender differences in CDC23 expression may lead to differences in CHP.

3.4. The role of the cytokine chemokine (C-C motif) ligand/chemokine (C-C motif) receptor chemokine family in the pathogenesis of CHP

According to Figure 4, our multi-omics analysis has unveiled the pivotal role of the CCL chemokine family in orchestrating the complex inflammatory cascade and disease symptoms in CHP. Several members of the CCL/CCR family, such as CCL2, CCL20, CCR2, and CCR6, were found to be significantly upregulated in the lung tissue samples of CHP patients.[60] These chemokines express various immune cell populations, including monocytes, macrophages, lymphocytes, and granulocytes.[67] The increased expression of these chemokines in CHP plays a pivotal role in the recruitment and activation of these inflammatory cells, leading to their accumulation within the pulmonary interstitium. The chemokine ligand CCL2 is also known as monocyte chemoattractant protein-1.[68] Elevated levels of CCL2 in CHP promote the recruitment of CCR2-expressing monocytes and macrophages to the site of inflammation.[69] These cells, once activated, release a myriad of pro-inflammatory cytokines, ROS, and proteolytic enzymes, contributing to the perpetuation of the inflammatory response and tissue injury.[59] CCL2 also contributes to the activation and proliferation of fibroblasts, a key event in the fibrotic remodeling process observed in CHP. In addition, CCL2 has been shown to promote the migration and proliferation of lung fibroblasts, contributing to the development of pulmonary fibrosis.[59]

Importantly, the analysis of our core signaling pathway in Figure 4 reveals a complex interplay between the CCL chemokine signaling pathways and other key regulatory cascades implicated in CHP pathogenesis, such as the MAPK and PI3K/AKT pathways. This crosstalk creates a self-sustaining inflammatory loop, wherein the activation of these pathways upregulates the expression of CCL chemokines, which, in turn, potentiate the activation of the cytokine genes *IL17, IL22, IL2,* and *IL12,* perpetuating the inflammatory and fibrotic responses. Furthermore, elevated levels of specific CCL chemokines or their receptors have been proposed as potential diagnostic and prognostic biomarkers for CHP, aiding in the early detection and monitoring of disease symptoms, as well as guiding therapeutic interventions.[70]

In summary, the CCL2 and CCL20 are chemokines that are regulated by target genes *IL17, IL22*, *IL2*, and *IL12*. They orchestrate the recruitment and activation of various inflammatory cell populations, which contribute to the development of granulomatous lesions, tertiary lymphoid structures, and fibrotic remodeling. Targeting specific members of this chemokine family or their receptors may represent a promising therapeutic strategy for attenuating the aberrant inflammatory and fibrotic processes underlying CHP symptoms.

3.4.1. The role of cytokine chemokine (C-C motif) ligand 2 in the pathogenesis of CHP

Chemokine (C-C motif) ligand 2, also known as monocyte chemoattractant protein-1, has emerged as a crucial mediator in the pathogenesis of CHP.[69] In Figure 4, our multi-omics analysis revealed a significant upregulation of CCL2 expression in lung tissue samples from CHP patients compared to healthy controls. CCL2 is a potent chemoattractant for monocytes, memory T cells, and immature dendritic cells, exerting its effects by binding to its cognate receptor, CCR2.[71] The increased production of CCL2 in CHP promotes the recruitment of these CCR2-expressing immune cells to the pulmonary interstitium, perpetuating the inflammatory cascade.[58] On recruitment to the lung parenchyma, monocytes differentiate into inflammatory macrophages, which play a pivotal role in the initiation and propagation of the inflammatory response in CHP. CCL2 can influence the tumor microenvironment by recruiting immune cells that may either support or inhibit tumor growth. The interaction between mutated adenomatous polyposis coli (APC) and increased levels of CCL2 can contribute to a pro-tumorigenic environment.[64] Subsequently, APCs present processed antigens to $CD4^+$ T-cells, initiating the adaptive immune response. The interaction between the T-cell receptors on $CD4^+$ T-cells and the antigen-major histocompatibility complex II complex on APCs is critical for T-cell activation. Moreover, some $CD4^+$ T-cells express CCR2 and can respond to CCL2, which is often produced at sites of inflammation.[72] This allows $CD4^+$ T-cells to migrate to areas where they are needed to provide help to other immune cells. Furthermore, Th1 cells are essential for defending against intracellular pathogens and are involved in promoting inflammation. They originate from naive $CD4^+$ T-cells.[73] Cytokines and other signals present in the immune environment guide the differentiation into Th1 cells. As a subset of $CD4^+$ T-cells, Th1 cells have specialized functions that focus on cell-mediated immunity and the activation of macrophages. They are distinct from other $CD4^+$ T-cell subsets like Th2, which are more involved in humoral immunity and allergic responses.[74]

Interleukin 6 primarily promotes Th17 differentiation in the presence of TGF-β. It can also create a pro-inflammatory environment that indirectly supports Th1 responses.[75] In addition, IL-21 can directly enhance Th1 responses by increasing interferon (IFN)-γ production. This can boost the ability of Th1 cells to activate macrophages and enhance their antimicrobial activities.[76] The presence of IL-6 and IL-21 can help balance different aspects of the immune response. IL-6 can promote a pro-inflammatory environment, while IL-21 can support both humoral and cell-mediated immunity by acting on various immune cells, including Th1 cells. The binding of IL-6 to its receptor induces the dimerization of glycoprotein gp130, bringing associated JAK1 and JAK2 into proximity, which allows for trans-phosphorylation and activation of JAK1 and JAK2. Similarly, the binding of IL-21 to its receptor induces the association of JAK1 and JAK3, allowing for trans-phosphorylation and activation of JAK1 and JAK3. These phosphorylated tyrosines serve as docking sites for signal transducers and activators of transcription (STAT) proteins, particularly STAT3. In the nucleus, phosphorylated STATs act as TF to regulate the expression of target genes involved in immune cell proliferation, differentiation, and function. Activated STAT3 then translocates to the nucleus and promotes the expression of TF retinoic acid-related orphan receptor (ROR)γt and RORα. Once expressed, TF RORγt and RORα directly regulate the transcription of Th17-specific genes, including *IL17A*, *IL17F*, and *IL22*. These target genes are key cytokines involved in the defense against extracellular pathogens, the promotion of inflammation, the regulation of fibrosis, and the direct killing or inhibition of pathogen growth. This enhances the first line of defense at mucosal surfaces.[69]

Furthermore, CCL2 has been implicated in the formation and maintenance of granulomatous lesions, a hallmark of CHP pathology. Th1 cells also produce IFN-γ, which works synergistically with TNF-α to enhance the activation and microbicidal activity of macrophages.[77] TNF-α promotes the expression of adhesion molecules on endothelial cells, facilitating the recruitment of leukocytes to infection or inflammation sites.[48] This enhances the local immune response, as this combination is particularly effective in combating intracellular pathogens. The recruited monocytes and macrophages, along with other inflammatory cells, aggregate and form non-caseating granulomas, perpetuating the local release of cytokines, chemokines, and proteolytic enzymes, leading to the progressive destruction of lung architecture.[78]

Subsequently, TNF-α and target genes *IL2* and *IL12* can work together through the phosphorylated TF STAT4 and STAT5 to enhance the activation and proliferation of T-cells.[70] In healthy lung tissues, the expression of *IL2* and *IL12* are typically low or undetectable, whereas, in CHP lung slice cells, their expression is elevated. In conjunction with *IL2*, it can amplify the inflammatory response by increasing the production of other cytokines and chemokines, leading to the recruitment and activation of additional immune cells. Moreover, *IL12* induces the production of IFN-γ, which, in turn, can stimulate the production of TNF-α by various immune cells. This creates a positive feedback loop that amplifies the Th1 response and enhances the ability to combat intracellular pathogens. The synergy between target genes *IL12* and *IL2* enhances the overall immune response within the granuloma.[79] *IL12* drives the initial Th1 response and macrophage activation, while *IL2* sustains T-cell proliferation and activation, ensuring that the granuloma remains an effective barrier against the pathogen.[80] This crosstalk suggests a self-sustaining inflammatory loop, wherein the activation of these pathways upregulates CCL2 expression, and CCL2, in turn, potentiates the activation of these signaling cascades, exacerbating the inflammatory milieu.[81] The interaction between CCL2 and its receptor, CCR2, on lung fibroblasts has been shown to induce their migration, proliferation, and differentiation into myofibroblasts, which are key mediators of the fibrotic process.[59]

In summary, compared to healthy control lung tissues, hypersensitivity pneumonitis lung tissues show increased expression of *IL2* and *IL12*. The chemokine CCL2 interacts with target genes *IL2* and *IL12* synergistically to regulate immune responses and inflammation, contribute to the formation of granulomatous lesions, and promote fibrotic remodeling through TNF-α and IFN-γ. Targeting the CCL2/CCR2 axis represents a promising therapeutic strategy for modulating the aberrant inflammatory and fibrotic processes underlying CHP symptoms.

3.4.2. The role of cytokine chemokine (C-C motif) ligand 20-cytokine chemokine (C-C motif) receptor 6 in immune response and pathogenesis of CHP

Organic microbes such as *Candida albicans*, *Mycobacterium avium*, and *Saccharopolyspora rectivirgula* are associated with CCL20 production and immune cell recruitment. Their role in CHP could involve the recruitment of CCR6 immune cells, leading to sustained lung inflammation and fibrosis, which are characteristic features of CHP. In Figure 4, $CD4^+$ effector Th17 cells, influenced by the CCL20-CCR6 chemotaxis, play a crucial role in immune responses to mucosal surface infections.[82] They are also associated with the development of various diseases, including asthma, hypersensitivity pneumonitis, and autoimmune diseases, such as rheumatoid arthritis. Chemokine CCR6 directs the migration of $CD4^+$ T-cells to inflammation sites

where CCL20 is expressed.[76] This chemotactic response is crucial for the localization of Th17 cells to mucosal tissues and infection or inflammation sites.

Furthermore, Th2 cells are not only a subset of $CD4^+$ T-cells that are primarily involved in the immune response against extracellular parasites, such as helminths, and allergic reactions but also enhance the production of antibodies that are effective against extracellular parasites and allergens.[73] Naive $CD4^+$ T-cells differentiate into Th2 cells in response to specific cytokines and signals from the immune environment.[76] Target gene *IL4* is the key cytokine driving this differentiation. Phosphorylated STAT6 translocates to the nucleus to promote the expression of Th2-related genes through TF GATA3, which leads to the expression of target genes of Th2-specific cytokines, including *IL4* and *IL13*.[80] Cytokines IL-4 and IL-13 induce the alternative activation of macrophages, promoting an anti-inflammatory phenotype that supports tissue repair and remodeling. Target genes *IL4* and *IL13* of key cytokines also play a role in the pathogenesis of allergic diseases, such as asthma.[82] They contribute to airway hyperresponsiveness, mucus production, and chronic inflammation. Cytokines IL-4 and IL-13 can induce epithelial cells to undergo EMT, a process where epithelial cells acquire mesenchymal, fibroblast-like properties, and contribute to fibrosis.[81] EMT is characterized by the loss of epithelial markers and the gain of mesenchymal markers, leading to increased migration, invasion, and extracellular matrix production by these cells. In diseases such as idiopathic pulmonary fibrosis and asthma, cytokines IL-4 and IL-13 contribute to the development of lung fibrosis by promoting the activation of fibroblasts and the production of extracellular matrix in the lungs.[83]

The right side of Figure 4 aims to characterize the role of the chemotaxis CCL20/CCR6 and ERK signaling pathways in CHP lung slice cells and non-CHP lung slice cells. The expression and production of CCL20/CCR6 in CHP fibrosis samples obtained during surgery were evaluated. In addition, colony formation, ERK signaling, and chemokine production were measured. It was found that CCL20 and CCR6 were highly expressed in most samples from the recurrent group.[59] Overall, these findings suggest that CCR6 and CCL20 may play a role in lung fibrosis, leading to proliferation and migration through autocrine or paracrine mechanisms. Disrupting the CCL20/CCR6 interaction could be a promising strategy for treating fibrosis. CCL20 also has antibacterial properties and is involved in lung innate immunity. The expression of CCL20 was significantly induced by TNF-α and suppressed by dexamethasone. Studies have shown that the abundance of CCL20 in bronchoalveolar lavage fluid from cystic fibrosis patients is nearly 90 times higher than that from healthy volunteers.[84] Furthermore, CCL20 is a bifunctional peptide with both innate and adaptive immune properties, regulated by inflammatory mediators expressed by airway epithelia and increased in cystic fibrosis airway secretions.

In summary, *IL4* and *IL13*-induced CCL20 production recruits $CCR6^+$ cells, potentially enhancing both Th2 and Th17 responses at inflammation sites. Compared to healthy control lung tissues, hypersensitivity pneumonitis lung tissues exhibit increased expression of target genes *IL4* and *IL13*. Through their involvement in immune cell recruitment, modulation of immune responses, promotion of fibrotic processes, and regulation of inflammatory mediators in the CHP lung environment, *IL4* and *IL13* play important roles in immune response and pathogenesis in CHP.

3.5. Selection of biomarkers as drug targets in CHP for systematic drug discovery through a DNN-based DTI model

After studying the pathogenetic mechanism of pulmonary cell fibrosis leading to CHP by observing abnormal apoptosis, immune response, and proliferation in downstream cellular functions of the core signal pathways in Figure 4, we identified important biomarkers of the CHP fibrosis mechanism as drug targets for treatment. Based on the fibrosis mechanism, CHP is characterized by proliferation and apoptosis. According to the aforementioned core signal pathways and their downstream abnormal cellular functions in Figure 4, we selected important biomarkers related to abnormal inflammation, apoptosis, proliferation, angiogenesis, and cell cycle induced by external organic allergens. Therefore, we chose AKT1, CCL20, CDC23, CXCL1, NF-κB, and TNF as important biomarkers, aiming to alleviate or exacerbate their expression levels, restoring them to normal inflammation, apoptosis, proliferation, angiogenesis, and cell cycle levels. In addition, AKT1 and TNF play important roles in the MAPK/PI3K pathway. CCL20, CDC23, and CXCL1 are key regulators of cell migration, growth, and apoptosis in the Th17 pathway, whereas NF-κB is involved in cellular responses to various cellular stimuli. After identifying the important biomarkers as drug targets, we considered their chemical properties. Based on DTI predictions using the DNN-based DTI model, we selected candidate molecular drugs for these drug targets (biomarkers).

The result of the DNN-based DTI model is a probability value, where higher values correspond to more reliable interactions (docking) between the drug and the target. The loss and accuracy during the training process are recorded in Figures 6 and 7, respectively.

Predictions are made based on high-probability biomarkers to find suitable candidate molecular drugs.

Concurrently, candidate drugs are predicted to dock with selected biomarkers. Potent drugs often carry high-risk damage, so considerations of toxicity, allergenicity, and the balance between efficacy and adverse reactions are crucial. To ensure the stability and safety of clinical drug trials, the drug's stability and safety must be considered in the design of clinical guidelines, such as regulatory capacity, toxicity, sensitivity, and specificity. To measure the regulatory capacity of candidate molecular drugs, available regulatory capacity data were downloaded from the L1000 Level 5 dataset, which includes data on 1453 genes treated in 156 different cell lines with 21,043 compounds. Regulatory capacity data, with positive values indicating upregulation and negative values indicating downregulation, were used to identify molecular drugs suitable for cell downregulation. Based on this criterion, we searched for molecular drugs in appropriate cell lines that could reverse the expression of biomarkers leading to pulmonary cell fibrosis, restoring their normal expression. In addition, lower drug toxicity helps reduce side effects, as indicated by the median lethal dose (LD_{50}), typically sourced from DrugBank.[22] The lower the LD_{50} value, the higher the toxicity. Furthermore, higher drug sensitivity (lower half maximal effective concentration [EC^{50}] values) allows for reduced drug dosages. Drug sensitivity data were obtained from the PRISM dataset, which includes information on 4,518 drugs tested in 578 human cell lines. EC_{50} measures drug potency, with lower EC_{50} values indicating greater efficacy at lower doses. By selecting the correct drug, abnormal expressions in diseases can be well-reversed. Predicting candidate drugs through the DNN-based DTI model, we identified several potential drugs for these biomarkers. The DNN-based DTI model, based on identified biomarkers, provided predictions for some candidate drugs along with their regulatory capacity, toxicity, and sensitivity information, as shown in Table 1. Based on the three-drug design criteria (appropriate regulatory capacity, high sensitivity, and low toxicity), potential drugs such as azathioprine, masitinib, and mycophenolate-mofetil were screened and combined into a multi-molecular drug for treating CHP, as shown in Table 2.

4. Discussion

In the stage of recurrence, the most straightforward method to treat CHP is to avoid prolonged exposure to allergens. This primarily includes organic dust, microorganisms, animal proteins, and certain chemicals.[1] However, some allergens are almost unavoidable in general households, such as indoor mold and pet dander. In addition, due to occupational reasons, some patients cannot avoid environments rich in allergens, such as beryllium and other metal dust. Therefore, medication is necessary to alleviate allergic symptoms. At present, corticosteroids like prednisone are commonly used to treat hypersensitivity pneumonitis. They can quickly relieve lung inflammation symptoms, but long-term use can lead to osteoporosis, hypertension, diabetes, peptic ulcers, and increased risk of infection.[85] Clinical evidence indicates that some patients may develop a dependency on corticosteroids, making it difficult to discontinue the medication. In this study, we utilize database mining, CHP whole-genome microarray data, and systems biology methods to identify core GWGEN and establish core signaling pathways of CHP and downstream cellular dysfunctions, as shown in Figure 4. This was done to investigate the pathogenetic mechanism and identify significant biomarkers as drug targets, predicting potential molecular drugs to treat CHP. Using the prediction of the DNN-based DTI model, we aim to find new multi-molecular drugs to treat CHP. Considering sensitivity, toxicity, and regulatory capacity as drug specifications, we combine deep learning and systems biology methods to find suitable molecular drugs for the treatment of CHP. Finally, the molecular drugs azathioprine, masitinib, and primaquine were chosen and formulated into a multi-molecular drug treatment for CHP, as shown in Table 2.

Azathioprine is an immunosuppressant that reduces inflammation by inhibiting immune system activity and helps control disease progression.[86] In the CHP microenvironment, the overactivation of $CD4^+$ T cells in the Th17 signaling pathway leads to abnormal anti-Th1 and Th2 cell signaling, exacerbating the immune response and creating a vicious cycle of uncontrolled cell apoptosis and proliferation.[87] According to the European Respiratory Society, azathioprine shows some efficacy in improving lung function, particularly in enhancing the diffusing capacity for carbon monoxide.[88] While its improvement in forced vital capacity is not significant, azathioprine is valuable in reducing the long-term side effects of steroid use, showing significant clinical value. Azathioprine inhibits the JAK signaling pathway, reducing inflammation and inhibiting TGF-β, thereby reducing B-cell activation and inflammation. It's anti-IL-6 or anti-TNF-α antibodies target specific inflammatory mediators for precise inflammation treatment.

Masitinib can reduce inflammation by inhibiting the activation of mast cells and eosinophils, thereby reducing the release of inflammatory mediators. In addition, by inhibiting multiple tyrosine kinase receptors, masitinib can affect various cell signaling pathways, providing a more comprehensive anti-inflammatory effect. It downregulates CCL20, CDC23, and CXCL1, thereby inhibiting tumor cell growth and immune cell activation. By blocking these cells' activity, masitinib can reduce inflammation and

potentially slow or prevent the progression of lung fibrosis. This is crucial for CHP patients, as managing fibrosis is a major treatment challenge. Studies suggest that masitinib may be beneficial in treating both fibrotic and non-fibrotic forms of HP by regulating immune responses and reducing tissue damage.[89] Compared to traditional broad-spectrum immunosuppressants, masitinib's targeted molecular pathway action offers a more specific treatment approach.

Primaquine is traditionally used to treat and prevent malaria by disrupting the mitochondrial function of *Plasmodium*. In addition, it has been found to have potential in regulating the immune system, particularly by inhibiting T-cell overactivation and reducing inflammation, which is significant in treating immune responses involved in HP. Primaquine also has antifibrotic effects, which can slow or prevent the progression of lung fibrosis in HP patients.[90] It effectively downregulates AKT1 and NF-κB expression levels in the MAPK signaling pathway, which play crucial roles in the metabolic changes of fibrotic cells. Although more clinical research and trials are needed to verify primaquine's specific effects in HP treatment, its potential as a molecular drug is promising, providing new ideas for developing more effective HP treatments.

There are a few limitations to this study. First, it relies on GSE86618 microarray data, and the quality and completeness of this data directly affect the accuracy of the results. If the dataset contains errors or incomplete data, it may lead to biased conclusions. Second, the study's conclusions are mainly based on mathematical models and bioinformatics analysis, lacking clinical experimental support. Predictions not validated by clinical and biological experiments may face challenges in practical drug applications, necessitating further clinical and biological experiments to ensure their effectiveness and safety. In addition, although multiple drug combinations were predicted, their actual effects and interactions need to be confirmed through *in vivo* and *in vitro* experiments. Finally, the model predictions may not fully consider potential synergistic or antagonistic effects between different drugs. These limitations need to be addressed in future research to improve the reliability and applicability of drug repurposing strategies.

5. Conclusion

In this study, we developed and validated a novel approach that combines systems biology to investigate the complex pathogenetic mechanisms underlying CHP-induced cellular fibrosis. Utilizing extensive database mining, we analyzed whole-genome data along with genetic and epigenetic networks to identify significant biomarkers of pathogenetic mechanisms as drug targets from a systemic drug design perspective for multi-molecular drugs of CHP treatment. Our methodology included system order detection methods for systematic identification of GWGEN, which can effectively eliminate false-positive interactions and regulations within candidate GWGEN, resulting in the accurate system identification of the real GWGEN for both CHP and non-CHP. Through big data exploration, we constructed comprehensive genetic and epigenetic biological networks of CHP and non-CHP. Employing the PNP method, we extracted core GWGEN for the annotation of KEGG signaling pathways to obtain the core signaling pathways of CHP and non-CHP. By comparing these core signal pathways of CHP and non-CHP, we investigated the pathogenetic mechanism of CHP and then identified key biomarkers crucial to CHP-induced cellular fibrosis. To discover candidate drugs and their drug targets (biomarkers), we trained a DNN-based DTI model using the DTI database, predicting interaction probabilities using the drug-target feature vectors. For drug targets, we applied stringent criteria – drug regulation capacity, low toxicity, and high sensitivity – to screen potential molecular drugs from the predicted candidate molecular drugs. Consequently, we identified azathioprine, masitinib, and primaquine as an optimal multi-molecular drug combination for CHP treatment, targeting critical biomarkers including AKT1, TNF, CCL-20, CDC23, and CXCL1. Our findings demonstrate that the integration of multiple types of genome-wide genomic data of CHP and non-CHP with the systems biology method can significantly enhance our understanding of key biomarkers in the role of pathogenetics involved in CHP-induced cellular fibrosis. The proposed combination of the systems biology method and systematic drug discovery design offers a promising new direction for the treatment of chronic lung fibrosis progression. Future studies will benefit from incorporating more diverse genomic data types for epigenetic and epigenomic regulation, which will further refine our understanding and improve therapeutic strategies for CHP.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they do not have competing interests.

Author contributions

Conceptualization: Sung-Yu Lin, Bor-Sen Chen
Formal analysis: Sung-Yu Lin, Bor-Sen Chen

Investigation: Sung-Yu Lin, Bor-Sen Chen
Methodology: Sung-Yu Lin, Bor-Sen Chen
Writing – original draft: Sung-Yu Lin
Writing – review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The microarray datasets of CHP patients and non-CHP lung slice cells were accessed from GSE86618 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE86618, accessed on 10 May 2023).

References

1. Selman M, Pardo A, King TE Jr. Hypersensitivity pneumonitis: Insights in diagnosis and pathobiology. *Am J Respir Crit Care Med*. 2012;186(4):314-324.

   doi: 10.1164/rccm.201203-0513CI

2. Vasakova M, Morell F, Walsh S, Leslie K, Raghu G. Hypersensitivity pneumonitis: Perspectives in diagnosis and management. *Am J Respir Crit Care Med*. 2017;196(6):680-689.

   doi: 10.1164/rccm.201611-2201PP

3. Costabel U, Miyazaki Y, Pardo A, *et al*. Hypersensitivity pneumonitis. *Nat Rev Dis Primers*. 2020;6(1):65.

   doi: 10.1038/s41572-020-0191-z

4. Gaxiola M, Buendía-Roldán I, Mejía M, *et al*. Morphologic diversity of chronic pigeon breeder's disease: Clinical features and survival. *Respir Med*. 2011;105(4):608-614.

   doi: 10.1016/j.rmed.2010.11.026

5. Patel AM, Ryu JH, Reed CE. Hypersensitivity pneumonitis: Current concepts and future questions. *J Allergy Clin Immunol*. 2001;108(5):661-670.

   doi: 10.1067/mai.2001.119570

6. Yeh SJ, Chen BS. Systems medicine design based on systems biology approaches and deep neural network for gastric cancer. *IEEE/ACM Trans Comput Biol Bioinform*. 2022;19(5):3019-3031.

   doi: 10.1109/tcbb.2021.3095369

7. Wingender E, Chen X, Hehl R, *et al*. TRANSFAC: An integrated system for gene expression regulation. *Nucleic Acids Res*. 2000;28(1):316-319.

   doi: 10.1093/nar/28.1.316

8. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: A general repository for interaction datasets. *Nucleic Acids Res*. 2006;34(suppl_1):D535-D539.

   doi: 10.1093/nar/gkj109

9. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The database of interacting proteins: 2004 update. *Nucleic Acids Res*. 2004;32(suppl_1):D449-D451.

   doi: 10.1093/nar/gkh086

10. Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G. MINT: A molecular INTeraction database. *FEBS Lett*. 2002;513(1):135-140.

    doi: 10.1016/s0014-5793(01)03293-8

11. Orchard S, Ammari M, Aranda B, *et al*. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*. 2014;42(D1):D358-D363.

    doi: 10.1093/nar/gkt1115

12. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 2015;4:e05005.

    doi: 10.7554/eLife.05005

13. Bovolenta L, Acencio M, Lemke N. HTRIdb: An open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics*. 2012;13:405.

    doi: 10.1186/1471-2164-13-405

14. Zheng G, Tu K, Yang Q, *et al*. ITFP: An integrated platform of mammalian transcription factors. *Bioinformatics*. 2008;24(20):2416-2417.

    doi: 10.1093/bioinformatics/btn439

15. Friard O, Re A, Taverna D, De Bortoli M, Corá D. CircuitsDB: A database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinformatics*. 2010;11(1):435.

    doi: 10.1186/1471-2105-11-435

16. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(D1):D92-D97.

    doi: 10.1093/nar/gkt1248

17. Chen BS, Wu CC. Systems biology as an integrated platform for bioinformatics, systems synthetic biology, and systems metabolic engineering. *Cells*. 2013;2(4):635-688.

    doi: 10.3390/cells2040635

18. Yeh SJ, Chen SW, Chen BS. Investigation of the genome-wide genetic and epigenetic networks for drug discovery based on systems biology approaches in colorectal cancer. *Front Genet*. 2020;11:117.

    doi: 10.3389/fgene.2020.00117

19. Chang S, Wang LH, Chen BS. Investigating core signaling pathways of hepatitis B virus pathogenesis for biomarkers identification and drug discovery via systems biology and deep learning method. *Biomedicines*. 2020;8(9):320.

doi: 10.3390/biomedicines8090320

20. Sakamoto Y, Ishiguro M, Kitagawa G. *Akaike Information Criterion Statistics*. Vol. 81. Dordrecht, The Netherlands: D Reidel. 1986. p. 26853.

21. UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(D1):D506-D515.

doi: 10.1093/nar/gky1049

22. Knox C, Law V, Jewison T, *et al*. DrugBank 3.0: A comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*. 2011;39:D1035-D1041.

doi: 10.1093/nar/gkq1126

23. Kim S, Thiessen PA, Bolton EE, *et al*. PubChem substance and compound databases. *Nucleic Acids Res*. 2016;44(D1):D1202-D1213.

doi: 10.1093/nar/gkv951

24. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res*. 2007;35:D198-D201.

doi: 10.1093/nar/gkl999

25. Ting CT, Chen BS. Repurposing multiple-molecule drugs for COVID-19-associated acute respiratory distress syndrome and non-viral acute respiratory distress syndrome via a systems biology approach and a DNN-DTI model based on five drug design specifications. *Int J Mol Sci*. 2022;23(7):3649.

doi: 10.3390/ijms23073649

26. Subramanian A, Narayan R, Corsello SM, *et al*. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171(6):1437-1452.e17.

doi: 10.1016/j.cell.2017.10.049

27. Xiong G, Wu Z, Yi J, *et al*. ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res*. 2021;49(W1):W5-W14.

doi: 10.1093/nar/gkab255

28. Kyriakis JM, Avruch J. Mammalian MAPK signal transduction pathways activated by stress and inflammation: A 10-year update. *Physiol Rev*. 2012;92(2):689-737.

doi: 10.1152/physrev.00028.2011

29. Cowan KN, Heilbut A, Humpl T, Lam C, Ito S, Rabinovitch M. Complete reversal of fatal pulmonary hypertension in rats by a serine elastase inhibitor. *Nat Med*. 2000;6(6):698-702.

doi: 10.1038/76282

30. Kumar S, Boehm J, Lee JC. p38 MAP kinases: Key signalling molecules as therapeutic targets for inflammatory diseases. *Nat Rev Drug Discov*. 2003;2(9):717-726.

doi: 10.1038/nrd1177

31. Cuenda A, Rousseau S. p38 MAP-kinases pathway regulation, function and role in human diseases. *Biochim Biophys Acta*. 2007;1773(8):1358-1375.

doi: 10.1016/j.bbamcr.2007.03.010

32. Engel P, Zhou LJ, Ord DC, Sato S, Koller B, Tedder TF. Abnormal B lymphocyte development, activation, and differentiation in mice that lack or overexpress the CD19 signal transduction molecule. *Immunity*. 1995;3(1):39-50.

doi: 10.1016/1074-7613(95)90157-4

33. Hers I, Vincent EE, Tavaré JM. Akt signalling in health and disease. *Cell Signal*. 2011;23(10):1515-1527.

doi: 10.1016/j.cellsig.2011.05.004

34. Morley SJ, Naegele S. Phosphorylation of eukaryotic initiation factor (eIF) 4E is not required for de novo protein synthesis following recovery from hypertonic stress in human kidney cells. *J Biol Chem*. 2002;277(36):32855-32859.

doi: 10.1074/jbc.C200376200

35. Dhillon AS, Hagan S, Rath O, Kolch W. MAP kinase signalling pathways in cancer. *Oncogene*. 2007;26(22):3279-3290.

doi: 10.1038/sj.onc.1210421

36. Liu Y. Renal fibrosis: New insights into the pathogenesis and therapeutics. *Kidney Int*. 2006;69(2):213-217.

doi: 10.1038/sj.ki.5000054

37. Jaffe AB, Hall A. Rho GTPases in transformation and metastasis. *Adv Cancer Res*. 2002;84:57-80.

doi: 10.1016/s0065-230x(02)84003-9

38. Chien Y, White MA. RAL GTPases are linchpin modulators of human tumour-cell proliferation and survival. *EMBO Rep*. 2003;4(8):800-806.

doi: 10.1038/sj.embor.embor899

39. Johnson GL, Lapadat R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science*. 2002;298(5600):1911-1912.

doi: 10.1126/science.1072682

40. Duronio V. The life of a cell: Apoptosis regulation by the PI3K/PKB pathway. *Biochem J*. 2008;415(3):333-344.

doi: 10.1042/bj20081056

41. Wrana JL, Attisano L, Wieser R, Ventura F, Massagué J. Mechanism of activation of the TGF-beta receptor. *Nature*. 1994;370(6488):341-347.

doi: 10.1038/370341a0

42. Massagué J, Seoane J, Wotton D. Smad transcription factors. *Genes Dev*. 2005;19(23):2783-2810.

    doi: 10.1101/gad.1350705

43. Akhurst RJ, Hata A. Targeting the TGFβ signalling pathway in disease. *Nat Rev Drug Discov*. 2012;11(10):790-811.

    doi: 10.1038/nrd3810

44. Zhang L, Zhou F, ten Dijke P. Signaling interplay between transforming growth factor-β receptor and PI3K/AKT pathways in cancer. *Trends Biochem Sci*. 2013;38(12):612-620.

    doi: 10.1016/j.tibs.2013.10.001

45. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: Update on Toll-like receptors. *Nat Immunol*. 2010;11(5):373-384.

    doi: 10.1038/ni.1863

46. Polivka J Jr., Janku F. Molecular targets for cancer therapy in the PI3K/AKT/mTOR pathway. *Pharmacol Ther*. 2014;142(2):164-175.

    doi: 10.1016/j.pharmthera.2013.12.004

47. Wajant H, Pfizenmaier K, Scheurich P. Tumor necrosis factor signaling. *Cell Death Differ*. 2003;10(1):45-65.

    doi: 10.1038/sj.cdd.4401189

48. Bradley JR. TNF-mediated inflammatory disease. *J Pathol*. 2008;214(2):149-160.

    doi: 10.1002/path.2287

49. Grell M, Wajant H, Zimmermann G, Scheurich P. The type 1 receptor (CD120a) is the high-affinity receptor for soluble tumor necrosis factor. *Proc Natl Acad Sci U S A*. 1998;95(2):570-575.

    doi: 10.1073/pnas.95.2.570

50. Hsu H, Xiong J, Goeddel DV. The TNF receptor 1-associated protein TRADD signals cell death and NF-kappa B activation. *Cell*. 1995;81(4):495-504.

    doi: 10.1016/0092-8674(95)90070-5

51. Zhang DW, Shao J, Lin J, *et al*. RIP3, an energy metabolism regulator that switches TNF-induced cell death from apoptosis to necrosis. *Science*. 2009;325(5938):332-336.

    doi: 10.1126/science.1172308

52. Huang H, Tindall DJ. Dynamic FoxO transcription factors. *J Cell Sci*. 2007;120(Pt 15):2479-2487.

    doi: 10.1242/jcs.001222

53. Zha J, Harada H, Yang E, Jockel J, Korsmeyer SJ. Serine phosphorylation of death agonist BAD in response to survival factor results in binding to 14-3-3 not BCL-X(L). *Cell*. 1996;87(4):619-628.

    doi: 10.1016/s0092-8674(00)81382-3

54. O'Neill KL, Huang K, Zhang J, Chen Y, Luo X. Inactivation of prosurvival Bcl-2 proteins activates Bax/Bak through the outer mitochondrial membrane. *Genes Dev*. 2016;30(8):973-988.

    doi: 10.1101/gad.276725.115

55. Jope RS, Johnson GV. The glamour and gloom of glycogen synthase kinase-3. *Trends Biochem Sci*. 2004;29(2):95-102.

    doi: 10.1016/j.tibs.2003.12.004

56. Newton AC. Protein kinase C: Poised to signal. *Am J Physiol Endocrinol Metab*. 2010;298(3):E395-E402.

    doi: 10.1152/ajpendo.00477.2009

57. Creamer TP. Calcineurin. *Cell Commun Signal*. 2020;18(1):137.

    doi: 10.1186/s12964-020-00636-4

58. Karin M, Liu Z, Zandi E. AP-1 function and regulation. *Curr Opin Cell Biol*. 1997;9(2):240-246.

    doi: 10.1016/s0955-0674(97)80068-3

59. Wynn TA. Integrating mechanisms of pulmonary fibrosis. *J Exp Med*. 2011;208(7):1339-1350.

    doi: 10.1084/jem.20110551

60. Liu J, Lin A. Role of JNK activation in apoptosis: A double-edged sword. *Cell Res*. 2005;15(1):36-42.

    doi: 10.1038/sj.cr.7290262

61. Garcia-Garcia T, Poncet S, Derouiche A, Shi L, Mijakovic I, Noirot-Gros MF. Role of protein phosphorylation in the regulation of cell cycle and DNA-related processes in bacteria. *Front Microbiol*. 2016;7:184.

    doi: 10.3389/fmicb.2016.00184

62. Vanhaesebroeck B, Alessi DR. The PI3K-PDK1 connection: More than just a road to PKB. *Biochem J*. 2000; 346(Pt 3):561-576.

63. Peters JM. The anaphase promoting complex/cyclosome: A machine designed to destroy. *Nat Rev Mol Cell Biol*. 2006;7(9):644-656.

    doi: 10.1038/nrm1988

64. Lubischer J. The cell cycle, principles of control. David O. Morgan. *Integr Comp Biol*. 2007;47:794-795.

    doi: 10.1093/icb/icm066

65. Shen T, Huang S. The role of Cdc25A in the regulation of cell proliferation and apoptosis. *Anticancer Agents Med Chem*. 2012;12(6):631-639.

    doi: 10.2174/187152012800617678

66. Van Leuken R, Clijsters L, Wolthuis R. To cell cycle, swing the APC/C. *Biochim Biophys Acta*. 2008;1786(1):49-59.

    doi: 10.1016/j.bbcan.2008.05.002

67. Bacon KB, Schall TJ. Chemokines as mediators of allergic inflammation. *Int Arch Allergy Immunol*. 1996;109(2):97-109.

doi: 10.1159/000237207

68. Shaw M, Collins BF, Ho LA, Raghu G. Rheumatoid arthritis-associated lung disease. *Eur Respir Rev*. 2015;24(135):1-6.

doi: 10.1183/09059180.00008014

69. Bettelli E, Oukka M, Kuchroo VK. T(H)-17 cells in the circle of immunity and autoimmunity. *Nat Immunol*. 2007;8(4):345-350.

doi: 10.1038/ni0407-345

70. Smith KA. Interleukin-2: Inception, impact, and implications. *Science*. 1988;240(4856):1169-1176.

doi: 10.1126/science.3131876

71. Charo IF, Ransohoff RM. The many roles of chemokines and chemokine receptors in inflammation. *N Engl J Med*. 2006;354(6):610-621.

doi: 10.1056/NEJMra052723

72. Sariban E, Imamura K, Luebbers R, Kufe D. Transcriptional and posttranscriptional regulation of tumor necrosis factor gene expression in human monocytes. *J Clin Invest*. 1988;81(5):1506-1510.

doi: 10.1172/jci113482

73. Mosmann TR, Cherwinski H, Bond MW, Giedlin MA, Coffman RL. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J Immunol*. 1986;136(7):2348-2357.

74. Zhu J, Paul WE. Peripheral CD4$^+$ T-cell differentiation regulated by networks of cytokines and transcription factors. *Immunol Rev*. 2010;238(1):247-262.

doi: 10.1111/j.1600-065X.2010.00951.x

75. Jones SA, Jenkins BJ. Recent insights into targeting the IL-6 cytokine family in inflammatory diseases and cancer. *Nat Rev Immunol*. 2018;18(12):773-789.

doi: 10.1038/s41577-018-0066-7

76. Lee AY, Eri R, Lyons AB, Grimm MC, Korner H. CC chemokine ligand 20 and its cognate receptor CCR6 in mucosal T cell immunology and inflammatory bowel disease: Odd couple or axis of evil? *Front Immunol*. 2013;4:194.

doi: 10.3389/fimmu.2013.00194

77. Schroder K, Hertzog PJ, Ravasi T, Hume DA. Interferon-gamma: An overview of signals, mechanisms and functions. *J Leukoc Biol*. 2004;75(2):163-189.

doi: 10.1189/jlb.0603252

78. Arora P, Verma G, Chauhan M, Ahuja A. Atypical protean manifestations of cutaneous sarcoidosis. *Indian J Dermatol*. 2022;67(1):87-89.

doi: 10.4103/ijd.ijd_379_21

79. Seder RA, Gazzinelli R, Sher A, Paul WE. Interleukin 12 acts directly on CD4$^+$ T cells to enhance priming for interferon gamma production and diminishes interleukin 4 inhibition of such priming. *Proc Natl Acad Sci USA*. 1993;90(21):10188-10192.

doi: 10.1073/pnas.90.21.10188

80. Wan YY. GATA3: A master of many trades in immune regulation. *Trends Immunol*. 2014;35(6):233-242.

doi: 10.1016/j.it.2014.04.002

81. Grünig G, Warnock M, Wakil AE, *et al*. Requirement for IL-13 independently of IL-4 in experimental asthma. *Science*. 1998;282(5397):2261-2263.

doi: 10.1126/science.282.5397.2261

82. Wills-Karp M, Luyimbazi J, Xu X, *et al*. Interleukin-13: Central mediator of allergic asthma. *Science*. 1998;282(5397):2258-2261.

doi: 10.1126/science.282.5397.2258

83. Wynn TA. Fibrotic disease and the T(H)1/T(H)2 paradigm. *Nat Rev Immunol*. 2004;4(8):583-594.

doi: 10.1038/nri1412

84. Rittirsch D, Flierl MA, Ward PA. Harmful molecular mechanisms in sepsis. *Nat Rev Immunol*. 2008;8(10):776-787.

doi: 10.1038/nri2402

85. Raghu G, Collard HR, Egan JJ, *et al*. An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med*. 2011;183(6):788-824.

doi: 10.1164/rccm.2009-040GL

86. NHS. *Azathioprine: A Medicine for Conditions Affecting the Immune system*. Available from: https://www.nhs.uk/medicines/azathioprine [Last accessed on 2025 Mar 27].

87. Anvar MT, Rashidan K, Arsam N, *et al*. Th17 cell function in cancers: Immunosuppressive agents or anti-tumor allies? *Cancer Cell Int*. 2024;24:355.

doi: 10.1186/s12935-024-03525-9

88. Oldham JM, Lee C, Valenzi E, *et al*. Azathioprine response in patients with fibrotic connective tissue disease-associated interstitial lung disease. *Respir Med*. 2016;121:117-122.

doi: 10.1016/j.rmed.2016.11.007

89. Morisset J, Johannson KA, Vittinghoff E, *et al*. Mycophenolate mofetil and azathioprine in chronic hypersensitivity pneumonitis. *Eur Respir J*. 2016;48(Suppl 60):PA3885.

doi: 10.1183/13993003.congress-2016.PA3885

90. Alberti ML, Rincon-Alvarez E, Buendia-Roldan I, Selman M. Hypersensitivity pneumonitis: Diagnostic and therapeutic challenges. *Front Med* (*Lausanne*). 2021;8:718299.

doi: 10.3389/fmed.2021.718299