

Supplementary materials

1.1 Systematic Model Construction for miRNAs and lncRNAs in Candidate HPI-GWGEN of COVID-19-associated ARDS and Non-Viral-ARDS Patient

The l^{th} host lncRNA gene in the GRN interaction model can be described by the following equation:

$$l_l^H[n] = \sum_{\tau=1}^{T_l} T_{l\tau}^H t_{\tau}^H[n] - \sum_{\mu=1}^{M_l} M_{l\mu}^H l_l^H[n] m_{\mu}^H[n] + \sum_{\lambda=1, \lambda \neq l}^{L_l} L_{l\lambda}^H l_{\lambda}^H[n] + \beta_{l,GRN}^H + \varepsilon_{l,GRN}^H[n] \quad (1)$$

$$-M_{l\mu}^H \leq 0 \quad \text{for } l = 1 \sim L, n = 1 \sim N$$

where $l_l^H[n], t_{\tau}^H[n], l_{\lambda}^H[n], m_{\mu}^H[n]$ indicate the expression level of the l^{th} host lncRNA gene, the τ^{th} host TF, the l^{th} host lncRNA gene, and the μ^{th} host miRNA gene in the n^{th} sample, respectively; $T_{l\tau}^H, L_{l\lambda}^H, M_{l\mu}^H$ indicate the regulation ability of the τ^{th} host TF, the λ^{th} host lncRNA gene and the μ^{th} host miRNA gene on the l^{th} host lncRNA gene, respectively; $\beta_{l,GRN}^H$ indicates the basal level of the l^{th} host lncRNA gene in the n^{th} sample; $\varepsilon_{l,GRN}^H[n]$ indicates the stochastic noise of the l^{th} host lncRNA gene in the n^{th} sample; T_l, L_l, M_l indicate the total number of host TF, host lncRNA gene and host miRNA gene interacting with the l^{th} host lncRNA gene, respectively; L indicates the total number of the l^{th} host lncRNA gene in candidate GRN; N denotes sample number in candidate GRN, either in COVID-19-associated ARDS or Non-Viral-ARDS group.

The μ^{th} host miRNA gene in the GRN interaction model can be described by the following equation:

$$m_{\mu}^H[n] = \sum_{\tau=1}^{T_{\mu}} T_{\mu\tau}^H t_{\tau}^H[n] - \sum_{x=1, x \neq \mu}^{M_{\mu}} M_{\mu x}^H m_{\mu}^H[n] m_x^H[n] + \sum_{\lambda=1}^{L_{\mu}} L_{\mu\lambda}^H l_{\lambda}^H[n] + \beta_{\mu,GRN}^H + \varepsilon_{\mu,GRN}^H[n] \quad (2)$$

$$-M_{\mu x}^H \leq 0 \quad \text{for } \mu = 1 \sim M, n = 1 \sim N$$

where $m_{\mu}^H[n], t_{\tau}^H[n], l_{\lambda}^H[n], m_x^H[n]$ indicate the expression level of the μ^{th} host miRNA gene, the τ^{th} host TF, the λ^{th} host lncRNA gene, and the x^{th} host miRNA gene in the n^{th} sample, respectively; $T_{\mu\tau}^H, L_{\mu\lambda}^H, M_{\mu x}^H$ indicate the regulation ability of the τ^{th} host TF, the λ^{th} host lncRNA gene and the x^{th} host miRNA gene on the μ^{th} host miRNA gene, respectively; $\beta_{\mu,GRN}^H$ indicate the basal level of the μ^{th} host miRNA gene in the n^{th} sample; $\varepsilon_{\mu,GRN}^H[n]$ indicates the stochastic noise of the μ^{th} host miRNA gene in the n^{th} sample; $T_{\mu}, L_{\mu}, M_{\mu}$ indicate the total number of host TF, host lncRNA gene and host miRNA gene interacting with the μ^{th} host miRNA gene, respectively; M indicates the total number of the μ^{th} host miRNA gene in candidate GRN; N denotes sample number in candidate GRN, either in COVID-19-associated ARDS or Non-Viral-ARDS group.

1.2 Systems identification and model order selection for obtaining real GWGENs of COVID-19-associated ARDS and Non-Viral-ARDS

Equations (1) ~ (2) can be expressed as the following regression form, respectively:

$$l_i^H[n] = \begin{bmatrix} t_i^H[n] & L & t_{T_i}^H[n] & l_i^H[n] m_i^H[n] & L & l_i^H[n] m_{M_i}^H[n] & l_i^H[n] & L & l_{L_i}^H[n] & I \end{bmatrix} \times \begin{bmatrix} T_{iI}^H \\ M \\ T_{iT_i}^H \\ -M_{iI}^H \\ M \\ -M_{iM_i}^H \\ L_i^H \\ M \\ L_{L_i}^H \\ \beta_{i,GRN}^H \end{bmatrix} + \varepsilon_{i,GRN}^H[n] \quad (3)$$

$$= \phi_i^{HL}[n] \theta_i^{HL} + \varepsilon_{i,GRN}^H[n] \quad , \text{ for } i = 1 \sim L, n = 1 \sim N$$

$$\begin{aligned}
m_\mu^H[n] &= \begin{bmatrix} l_l^H[n] & L & t_{T_\mu}^H[n] & m_\mu^H[n]m_l^H[n] & L & m_\mu^H[n]m_{M_\mu}^H[n] & l_l^H[n] & L & l_{L_\mu}^H[n] & 1 \end{bmatrix} \times \begin{bmatrix} T_{\mu l}^H \\ \mathbf{M} \\ T_{\mu T_\mu}^H \\ -\mathbf{M}_{\mu l}^H \\ \mathbf{M} \\ -\mathbf{M}_{\mu M_\mu}^H \\ L_{\mu l}^H \\ \mathbf{M} \\ L_{\mu L_\mu}^H \\ \beta_{\mu,GRN}^H \end{bmatrix} + \varepsilon_{\mu,GRN}^H[n] \\
&= \varphi_\mu^{HM}[n]\theta_\mu^{HM} + \varepsilon_{\mu,GRN}^H[n] \quad , \text{ for } \mu = 1 \sim M, n = 1 \sim N
\end{aligned} \tag{4}$$

where the superscript H , P , HL , and HM denote abbreviation of the host, pathogen, host lncRNA gene, host miRNA gene, and pathogen gene, respectively; $\varphi_l^{HL}[n]$, $\varphi_\mu^{HM}[n]$ denote the regression vector which can be obtained from the corresponding expression data we integrated; θ_l^{HL} , θ_μ^{HM} are corresponding unknown parameter vectors of the l^{th} host lncRNA gene and the μ^{th} host miRNA gene, respectively.

Equation (3) ~ (4) can be further augmented for N samples as follows

$$\begin{bmatrix} l_l^H[1] \\ l_l^H[2] \\ \mathbf{M} \\ l_l^H[N] \end{bmatrix} = \begin{bmatrix} \varphi_l^{HL}[1] \\ \varphi_l^{HL}[2] \\ \mathbf{M} \\ \varphi_l^{HL}[N] \end{bmatrix} \theta_l^{HL} + \begin{bmatrix} \varepsilon_{l,GRN}^H[1] \\ \varepsilon_{l,GRN}^H[2] \\ \mathbf{M} \\ \varepsilon_{l,GRN}^H[N] \end{bmatrix} \quad \text{for } l = 1 \sim L \tag{5}$$

$$\begin{bmatrix} m_\mu^H[1] \\ m_\mu^H[2] \\ \mathbf{M} \\ m_\mu^H[N] \end{bmatrix} = \begin{bmatrix} \varphi_\mu^{HM}[1] \\ \varphi_\mu^{HM}[2] \\ \mathbf{M} \\ \varphi_\mu^{HM}[N] \end{bmatrix} \theta_\mu^{HM} + \begin{bmatrix} \varepsilon_{\mu,GRN}^H[1] \\ \varepsilon_{\mu,GRN}^H[2] \\ \mathbf{M} \\ \varepsilon_{\mu,GRN}^H[N] \end{bmatrix} \quad \text{for } \mu = 1 \sim M \tag{6}$$

Equations (5) ~ (6) above can be simply represented as follows

$$\mathbf{L}_l^H = \Phi_l^{HL} \theta_l^{HL} + \Omega_l^{HL} \quad , \text{ for } l = 1 \sim L \tag{7}$$

$$\mathbf{M}_\mu^H = \Phi_\mu^{HM} \theta_\mu^{HM} + \Omega_\mu^{HM} \quad , \text{ for } \mu = 1 \sim M \tag{8}$$

For each parameter vector θ_l^{HL} , θ_μ^{HM} in equations (7) ~ (8), we can individually

estimate by solving the constrained least-square problem as follows:

$$\theta_l^{HL} = \underset{\theta_l^{HL}}{\operatorname{argmin}} \frac{1}{2} \left\| \Phi_l^{HL} \theta_l^{HL} - L_l^H \right\|_2^2, \text{ subject to } A_l^{HL} \theta_l^{HL} \leq B_l^{HL} \quad (9)$$

where $A_l^{HL} = \begin{bmatrix} O_{M_l \times T_l} & I_{M_l \times M_l} & O_{M_l \times L_l} & O_{M_l \times I} \end{bmatrix}$, $B_l^{HL} = \begin{bmatrix} O_{M_l \times I} \end{bmatrix}$

$$\theta_\mu^{HM} = \underset{\theta_\mu^{HM}}{\operatorname{argmin}} \frac{1}{2} \left\| \Phi_\mu^{HM} \theta_\mu^{HM} - M_\mu^H \right\|_2^2, \text{ subject to } A_\mu^{HM} \theta_\mu^{HM} \leq B_\mu^{HM} \quad (10)$$

where $A_\mu^{HM} = \begin{bmatrix} O_{M_\mu \times T_\mu} & I_{M_\mu \times M_\mu} & O_{M_\mu \times L_\mu} & O_{M_\mu \times I} \end{bmatrix}$, $B_\mu^{HM} = \begin{bmatrix} O_{M_\mu \times I} \end{bmatrix}$

where O and I denote zero matrix and identity matrix, respectively.

It is noted that in the parameter fitting process for the regression model of each protein/gene, what candidate HPI-GWGEN provided is all possible binding molecules and our model will need further parameter trimming process. However, such a model parameter identification process in equations (9) ~ (10) will often result in overfitting conditions when a finite sample of the dataset at hand. Therefore, the systems order detection method, AIC, was employed to detect the system order (i.e., the number of interactions of each protein with other proteins or the number of regulator TFs on each gene by the fact that system order can minimize the corresponding AIC). For each model in HPI-GWGEN, the AIC values of the l^{th} host lncRNA gene in equation (1) and the μ^{th} host miRNA gene in equation (2) are defined as follows:

$$AIC_l^{HL}(\theta_l^{HL}, \Phi_l^{HL}, L_l^H) = \log\left(\frac{\left\| \Phi_l^{HL} \theta_l^{HL} - L_l^H \right\|_2^2}{N}\right) + \frac{2\dim(\theta_l^{HL})}{N}, \text{ for } l = 1 \sim L \quad (11)$$

$$AIC_\mu^{HM}(\theta_\mu^{HM}, \Phi_\mu^{HM}, M_\mu^H) = \log\left(\frac{\left\| \Phi_\mu^{HM} \theta_\mu^{HM} - M_\mu^H \right\|_2^2}{N}\right) + \frac{2\dim(\theta_\mu^{HM})}{N}, \text{ for } \mu = 1 \sim M \quad (12)$$

where $\dim(\theta_l^{HL})$ and $\dim(\theta_\mu^{HM})$ denote parameter vector dimension of each model, respectively. In general, increasing parameter number (system order) will result in good model fit, such that log residual error in the first term of AIC will decrease and the second term of AIC will increase, and vice versa. Therefore, there should be exactly parameter numbers corresponding with the optimal parameter vector to achieve the minimum AIC among all possible binding combinations for each protein/gene. Considering practical computational efficiency for implementation, for each protein/gene, forward and backward stepwise algorithms are both adopted to find the minimum AIC in equation (11) ~ (12), with the corresponding parameter numbers to achieve the minimum AIC in equation (9) ~ (10) with the help of *lsqlin* function in 2021 MATLAB optimization toolbox. Therefore, we trimmed the insignificant parameters in candidate HPI-GWGEN out of system order detected by AIC to obtain real HPI-GWGEN of COVID-19-associated ARDS and Non-Viral-ARDS.

Tables

Table S1. Details of nodes included in *Virus* class used for SARS-CoV-2 in HPI-PPI and HPI-GRN.

Node Network	<i>Virus</i>	Total
HPI-PPI	E, M, N, ORF10, ORF1ab, ORF3a, ORF6, ORF7a, ORF7b, ORF8, S	11
HPI-GEN	3UTR , 5UTR , E, M, N, ORF10, ORF1ab, ORF3a, ORF6, ORF7a, ORF7b, ORF8, S	13

Note: **5UTR** 5'-untranslated region; **3UTR** 5'-untranslated region; **ORF1ab** in HPI-PPI is polypeptide which can be further cleaved to yield 16 nonstructural proteins.

Table S2. KEGG pathway enrichment analysis of COVID-19-associated ARDS core HPI-GWGEN using DAVID website.

Term	Gene Numbers	<i>p</i> -value
Olfactory transduction	101	4.04E-11
Autoimmune thyroid disease	19	6.86E-05
Neuroactive ligand-receptor interaction	57	6.38E-04
Graft-versus-host disease	13	6.70E-04
Cell adhesion molecules (CAMs)	33	1.51E-03
Allograft rejection	13	2.11E-03
Type I diabetes mellitus	14	2.24E-03
Hypertrophic cardiomyopathy (HCM)	19	1.19E-02
Cytokine-cytokine receptor interaction	45	1.77E-02
Dilated cardiomyopathy	19	2.49E-02
Endometrial cancer	13	3.62E-02

Note: KEGG PATHWAY is a collection of molecular pathway diagrams, which are categorized by the annotation terms to represent the current knowledges about the metabolism or cellular function of each species. This table summarizes the functional annotation chart and is obtained by submitting the genes list of 4000 nodes to the DAVID website. Each annotation term summarized total gene counts involved in the corresponding KEGG pathway among the 4000 genes we upload.

Table S3. The KEGG pathway enrichment analysis of Non-Viral-ARDS core HPI-GWGEN using DAVID website.

Term	Gene Numbers	<i>p</i> -value
Olfactory transduction	106	4.30E-11
Amoebiasis	28	1.46E-03
Mineral absorption	15	2.28E-03
Protein digestion and absorption	23	4.95E-03
Bile secretion	19	6.58E-03
Platelet activation	30	8.01E-03

Leukocyte transendothelial migration	27	9.78E-03
Pertussis	19	1.59E-02
ECM-receptor interaction	21	1.83E-02
RIG-I-like receptor signaling pathway	17	3.42E-02

Note: KEGG PATHWAY is a collection of molecular pathway diagrams, which are categorized by the annotation terms to represent the current knowledges about the metabolism or cellular function of each species. This table summarizes the functional annotation chart and is obtained by submitting the genes list of 4000 nodes to the DAVID website. Each annotation term summarized total gene counts involved in the corresponding KEGG pathway among the 4000 genes we upload.

Table S4. Model performance of DNN-DTI model (5-fold cross-validation, epoch=52).

	Validation loss	Validation Accuracy	Testing loss	Testing Accuracy
1	0.185257	0.932791	0.230130	0.932164
2	0.205787	0.930211	0.192034	0.930566
3	0.197644	0.927939	0.201237	0.929030
4	0.214519	0.929585	0.204637	0.932890
5	0.201079	0.934167	0.213735	0.931209
Average	0.200857	0.930938	0.208355	0.931172
Standard Deviation	0.009640	0.002245	0.012915	0.001334

Figures

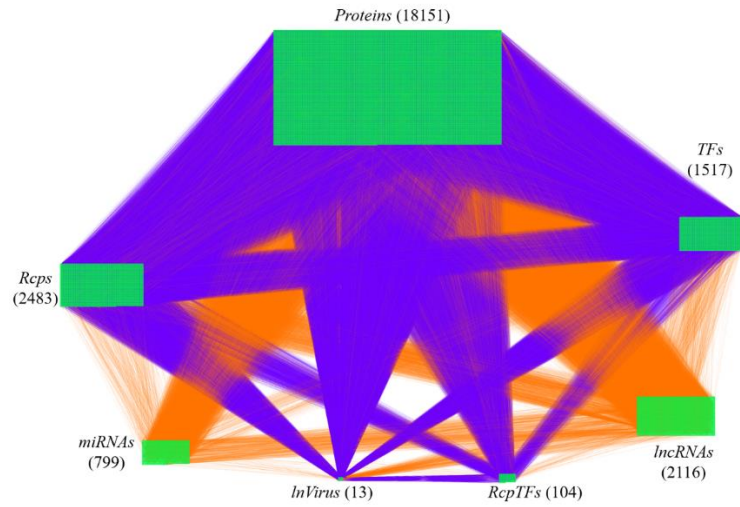


Figure S1. The host-pathogen interspecies real genome-wide genetic and epigenetic network (HPI-GWGEN) of COVID-19-associated ARDS. Purple lines indicate the protein-protein interactions; Orange lines denote the gene regulations. The numbers of *Proteins*, *Rcps*, *TFs*, *RcpTFs*, *miRNA*, *LncRNA*, *Virus* are 18151, 2483, 104, 799, 2116, and 13, respectively.

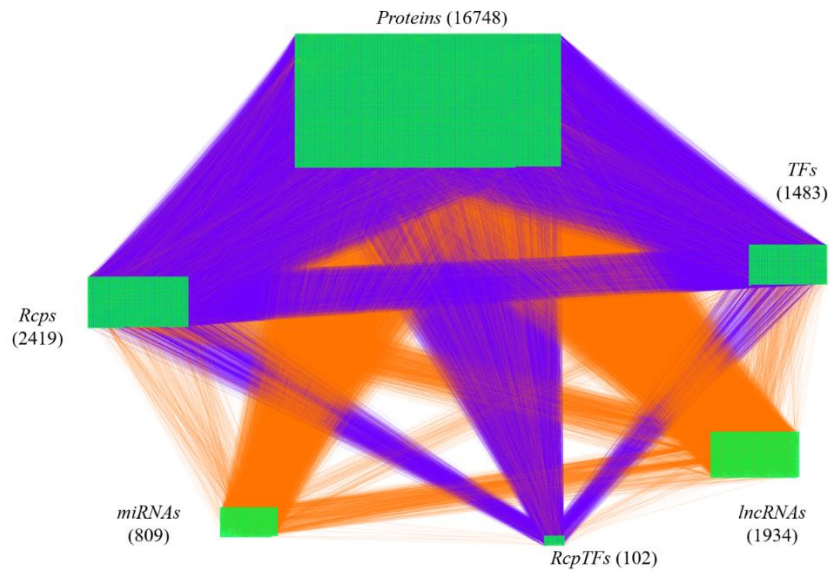


Figure S2. The host-pathogen interspecies real genome-wide genetic and epigenetic network (HPI-GWGEN) of Non-Viral-ARDS. Purple lines indicate the protein-protein interactions; Orange lines denote the gene regulations. The numbers of *Proteins*, *receptors*, *TFs*, *RcpTFs*, *miRNA*, and *LncRNA* are 16748, 2419, 1483, 102, 8099, and 1934, respectively

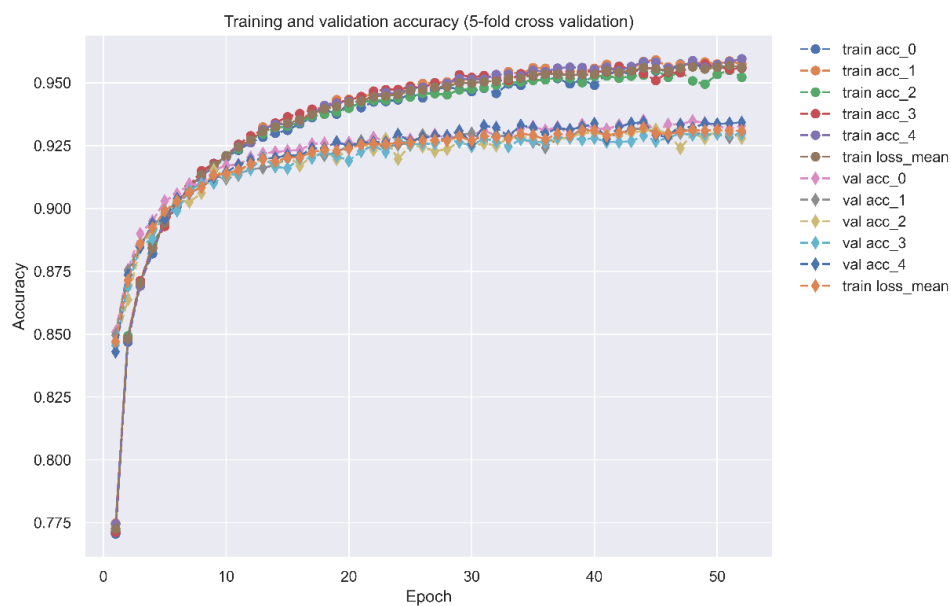


Figure S3. Accuracy learning curves of deep neuron network drug-target interaction model (DNN-DTI). The early Stopping strategy is applied to automatically stop the learning process at the epoch of 52.

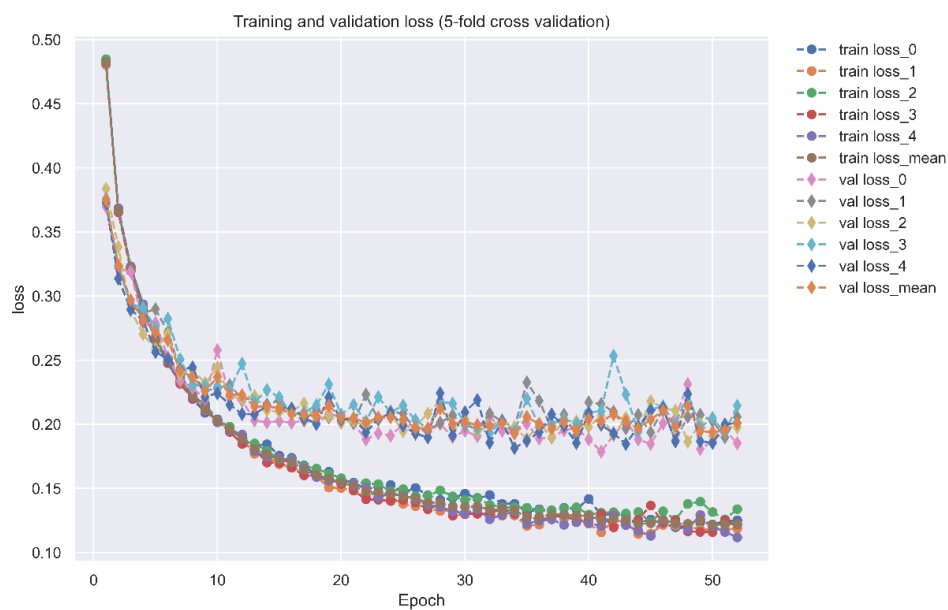


Figure S4. Loss learning curves of deep neuron network drug-target interaction model (DNN-DTI). The early Stopping strategy is applied to automatically stop the learning process at the epoch of 52.

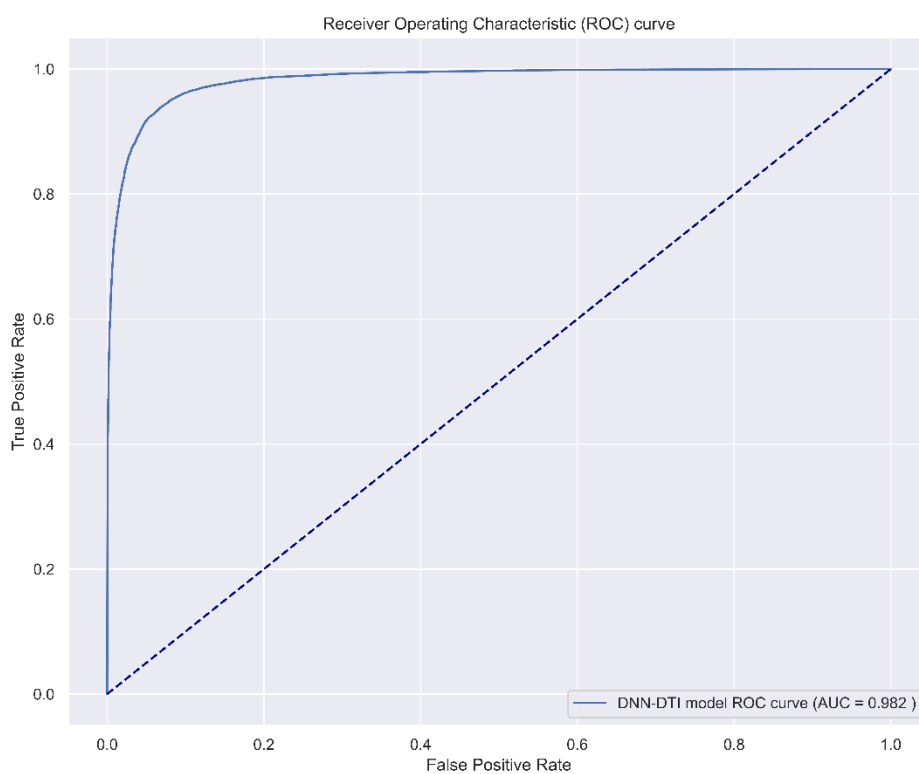


Figure S5. The Receiver Operating Characteristic (ROC) curve and the Area Under the Curve of ROC (AUC-ROC) score of the deep neuron network drug-target interaction model (DNN-DTI). A purple diagonal dot line indicates the worst situation when AUC is approximately 0.50, where the model has no discrimination capacity to distinguish between positive and negative classes. A classifier that predicts at random will appear as the diagonal line. The dotted line represents the ROC curve of a purely random classifier.