

1 *Article*

2 **Dissecting virus-host interplay dynamics from a response regression perspective**

3 Ruikun Du ^{1,#,*}, Xuran Gu ^{1,#}, Ziwei Li ¹, Xiaoping Zhang ¹, Jichen Yang ¹, Shaobo Wang ^{2,*},

4 Rong Rong ^{3,*}

5 ¹ Qingdao Academy of Chinese Medical Sciences, Shandong University of Traditional Chinese
6 Medicine, Qingdao 266041, China

7 ² Guangzhou National Laboratory, Department of Basic Research, Guangzhou 510005, China

8 ³ College of Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan 250355,
9 China

10 [#] These authors contribute equally to this work.

11 * Correspondence: rosierong@163.com (Rong Rong); wang_shaobo@gzlab.ac.cn (Shaobo
12 Wang) ; ruikun@sdutcm.edu.cn (Ruikun Du)

13

14

Abstract

Dynamic crosstalk between invading pathogens and host immunity determines the course and outcome of acute viral infections. Well-coordinated immune responses mediate effective antiviral defence, while immune dysregulation fuels severe disease. However, standardized criteria for identifying antiviral immune dysregulation remain lacking, especially at early stages of immune disturbance. Previously, we developed a response-regression model to separate viral-load-driven transcriptomic changes from other regulatory signals. In this study, we conceptualize dysregulated antiviral immunity as an independent regulatory input and apply this framework to longitudinal lung transcriptomic data from low- and high-dose influenza A virus-infected mice, identifying functionally distinct gene sets. Some genes track disease progression, while others serve as candidate biomarkers or pathogenic mediators of severe disease. By calibrating host response direction and magnitude relative to viral fluctuations, our method advances the understanding of virus-host interplay dynamics and provides a new criterion to discriminate immune dysregulation from homeostasis.

Keywords: Response regression model, Transcriptome analytical algorithm, Virus-host interplay dynamics

1.1 Introduction

Acute viral infection involves highly dynamic, multifaceted host-pathogen crosstalk, encompassing rapid viral dissemination and subsequent clearance, as well as fluctuating host immune responses that drive either protective or pathogenic outcomes^{1, 2}. In general, the magnitude of host responses scales with the intensity of viral stimulation. High-dose infection or uncontrolled viral replication elevates viral loads and augments host immune activation, whereas spontaneous viral clearance or antiviral intervention reverts these host responses toward homeostatic levels. We term this homeostatic restoration process “response regression”, a biological phenomenon that profoundly shapes the host response state and complicates the distinction between virus fluctuation-driven host responses and signals originating from other regulatory inputs³.

Previously, we developed a transcriptomic analytical pipeline based on the response-regression model to disentangle virus-induced transcriptional changes from alternative regulations, exemplified by pharmacological perturbations arising from host-directed antiviral interventions³. Host-directed antivirals predominantly target viral-responsive host factors, whose expression is dually modulated by viral stimulation and pharmaceutical intervention^{4, 5}. Conventional pairwise comparisons between drug-treated and untreated groups indiscriminately integrate these dual regulatory effects and fail to decouple drug-specific pharmacological impacts from response regression-driven perturbations, thereby introducing analytical bias and compromising mechanistic investigations of host-directed antiviral strategies³.

Given the sequential activation of innate and adaptive immunity during viral infection^{1, 6}, temporal cues of host immunity constitute an independent regulatory factor that shapes immune landscapes over the course of infection. Similarly, uncoordinated or dysregulated immune responses to viral infection that precipitate severe disease may also be regarded as external regulatory inputs^{7, 8}. Proper adjustment for virus-associated confounding factors is critical to reliably discriminate dysregulated virus-host interplay from coordinated protective immune responses, enabling the precise identification of key biomarkers and pathogenic drivers of

severe illness.

In this study, we refined and comprehensively validated this response-regression-based algorithm using longitudinal lung transcriptomic datasets from influenza A virus (IAV)-challenged mice. This response-regression analysis re-examined the dynamics of virus-host interplay, offering novel mechanistic insights into IAV disease progression and uncovering novel potential molecular underpinnings governing infection-associated disease severity.

2 Results

2.1 The host response is highly dynamic across the time course of virus infection

Zheng et al. previously reported longitudinal single-cell transcriptomic profiling of lung tissues from mice infected with non-lethal low-dose (LD) and lethal high-dose (HD) IAV, defining temporal host-response kinetics over infection (Figure 1a) ⁹. We extracted pseudobulk mRNA profiles across all time points from this public dataset for downstream analyses. Figure 1b depicted total DEGs in infected mouse lungs at each time point relative to mock-infected controls. Consistent with the robust early-phase viral replication for both LD (days 1-3) and HD (days 1-2) infections ⁹, rapid expansion of up- and downregulated DEGs mirrored host responses proportional to viral stimulus intensity. At later infection stages, transcriptomic profiles exhibited substantial fluctuations, probably driven by divergent viral load kinetics and progressive tissue pathological damages in combination.

For an overview of host-response dynamics, log₂FC values (relative to mock) of all genes at each time point were plotted against those from LD-day3, which served as a predefined model group (Figure 1c). Linear regression of these scatter plots yielded slopes representing global host-response magnitude at indicated stages. Notably, virus-induced up- and downregulated genes exhibited distinct dynamic features. Dots for upregulated genes shifted leftwards at slopes < 1 and rightwards at slopes > 1, indicating the fastest induction at LD-day2 or HD-day1, followed by dampened upregulation or reversion toward baseline. Conversely, most downregulated-gene dots were displaced rightwards at slopes > 1, displaying sustained deepening repression largely uncoupled from viral clearance.

We next focused on up- and downregulated DEGs obtained from model-versus-mock

comparison, classifying them as viral-responsive genes (VRGs)-up and VRGs-down, respectively. Log₂FC values of VRGs-up (Figure 1d) and VRGs-down (Figure 1e) across all time points were separately plotted against the model group, yielding regression slopes with divergent trends that support the above drifting observations. Heatmaps depicting individual VRG expression further validated these temporal dynamic patterns (Figure 1f). Together, these results highlight the highly dynamic host response during IAV infection, in which discrete gene subsets display markedly divergent temporal kinetics.

2.2 Rationalization and optimization of the response regression-based transcriptome analytical pipeline

Conventional pairwise comparisons evaluate alterations in host factors within test groups relative to mock or model groups separately¹⁰. To overcome this limitation, we previously developed a novel metric, the regression index (RI), to concurrently quantify the relative state of each host factor in test groups against both model and mock counterparts (Figure 2a)³. VRGs-up identified from the model-versus-mock comparison were extracted as an exemplary gene set (Figure 2b), and RI values were computed for each VRG-up at individual time points. An RI of 1 indicates expression matching the model group, whereas RI=0 corresponds to baseline levels in mock controls. Theoretically, the RI of each VRG rises or falls in tandem with increased or reduced viral loads. However, RI sensitivity to viral-load fluctuations varies markedly across genes due to their divergent temporal expression kinetics during IAV infection (Figure 2c). This complicates the discrimination between transcriptomic shifts driven by viral-load changes and those originating from other regulatory sources.

To overcome this analytical barrier, we incorporated an additional reference group having lower viral loads than the model group without extra regulatory inputs. This setup permits prediction of theoretical gene expression status in test groups under perturbations originating solely from viral-load variation. Using LD-day2 samples as the reference dataset (Figure 2c), we validated this framework with the aforementioned model-derived VRGs-up. Initially, RI values of VRGs-up across serial infection time points were plotted against reference expression profiles (Figure 2d).

We next defined coefficient k_1 as the ratio of global host-response intensity between the test and reference groups. Although $\log_2$ FC-derived linear slopes relative to model profile generally reflect overall response magnitude, stage-specific infection dynamics can markedly distort expression patterns when only VRGs-up are considered. We therefore utilized the median RI of VRGs-up at each time point to quantify regression intensity and mitigate temporal confounding (Figure 2d).

Using RI values derived from the reference group as anchors, we computed the viral load-driven estimated RI (est.RI) for each gene in test groups. The normalized RI (norm.RI), defined as the residual between observed and estimated RI, quantifies additional regulatory effects independent of viral-load-driven responses (Figure 2c). Genes with positive and negative norm.RI were categorized as response-regression deviation (RRD) and response-regression transcendence (RRT), respectively. RRD and RRT denote regulatory trends concordant or antagonistic to viral stimulation, which are shaped by the intrinsic viral-response characteristics of individual genes in the model group.

Notably, consistent systematic offsets were observed between actual and estimated RI values (Figure 2e). Genes with high reference RI yielded upward-shifted est.RI values, whereas genes with low reference RI exhibited downward-biased estimation. These biases should have originated from divergent gene-specific temporal kinetics. As high reference-RI genes reach near-peak responses at or before the reference time point and subsequently plateau or decline, low reference-RI genes remain weakly responsive at reference stages but undergo robust activation at the model time point (Figure S1). The mismatch between actual and estimated RI forms an ambiguous interval prone to either false-positive or false-negative signals. Although such ambiguities can be further resolved with denser longitudinal monitoring, we opted to exclude these data from downstream analyses in the present study.

We benchmarked the pipeline by assigning the model group also as the test group (Figure 2f). For data outside the ambiguous interval, we initially applied reciprocal-curve-based thresholds, $y > k_2/(k_1x)$ ($x > 0$) and $y < k_2/(k_1x - k_1)$ ($x < 1$), for RRD and RRT detections respectively. The parameter k_2 governs screening stringency, where smaller k_2 values impose more stringent

filtering (Figure 2g). Reference-group RI values of 0 and 1 serve as critical reliability boundaries. Genes with reference $RI \leq 0$ tend to produce biased RRD signals, while those with reference $RI \geq 1$ are vulnerable to RRT-associated noise. Such low-confidence outputs were further filtered out by the above-mentioned threshold criteria. Note that the low-confidence exclusion ranges can be flexibly tuned by horizontal shifting the threshold curves, producing revised formulas with $y > k_2/(k_1x+a)$ ($x > -k_1/a$) for RRD and $y < k_2/(k_1x-k_1-a)$ ($x < 1+k_1/a$) for RRT (Figure 2h).

In the present study, hook-function-based thresholds were adopted to further reduce subjective bias introduced by manual tuning of low-confidence exclusion ranges. The inequalities $x > (a-y)/k_1+k_2/(k_1y)$ ($y > 0$) and $x < (a-y)/k_1+k_2/(k_1y)$ ($y < 0$) were applied to identify RRD and RRT genes, respectively (Figure 2i). Here, parameter a is redefined as a constant computed as $a = k_{1-test}/k_{1-model}$, whereas k_2 remains a variable controlling screening strictness.

2.3 The response regression model guided algorithm rectifies biases in pairwise comparison

Longitudinal transcriptomic profiles of LD-day3-derived VRGs-up were systematically processed using our optimized analytical pipeline to screen for RRD and RRT genes across successive infection time points, with the filtering stringency parameter k_2 initially fixed to 1. Overall, only a small number of RRD/RRT hits were recovered across the time course in both the LD and HD groups, reflecting tight coupling between viral load dynamics and host transcriptional responses (Figure 3a). On the other hand, it is reasonable to assume that these RRD/RRT genes identified at indicated time points may arise from substantial perturbations imposed by temporal factors. To test this possibility, we examined day-7-derived RRD and RRT genes from LD and HD infections (Figure 3a) and found that most of these genes overlapped between the two infection conditions (Figure S2). Heatmap profiling of these overlapping RRD/RRT genes at LD and HD day7 revealed almost similar temporal kinetic patterns (Figure S2). Moreover, functionally comparable outputs were obtained by defining HD-day2 and LD-day2 as model and reference groups, respectively, confirming the high robustness of our algorithm to alternative model-group configurations (Figure S3). These findings solidify the rationality, robustness, and practical efficacy of our response-regression-based analytical

framework.

In parallel, VRGs-up defined by model-versus-mock comparison were also subjected to conventional pairwise analysis against the model group across longitudinal transcriptomic profiles, with DEG identification thresholds set at $|\log_2FC| > 1$. As shown in Figure 3b, abundant downregulated DEGs were detected at early infection time points including LD-day1/2 and HD-day1, whereas upregulated DEGs were extremely scarce. This skewed distribution reveals a substantial analytical bias originating from attenuated host responses under low viral load, which inevitably distorts downstream biological interpretation.

In contrast, our response regression-based algorithm greatly mitigates these intrinsic biases. Composite scatter plots were generated by integrating pairwise-derived $\log_2FC$ value with response regression k_2 threshold for each gene (Figure 3c), directly illustrating the superior performance of our pipeline. The biased signals from pairwise comparison were largely removed, with hit RRD and RRT genes exhibiting a balanced ratio, which is more consistent with the biological expectation.

We further observed systematic shifts in effective $\log_2FC$ cutoffs dictated by the relative magnitude of host responses between test and model groups, quantified by the k_1 metric. When $k_{1-test} < k_{1-model}$, the functional $\log_2FC$ cutoff was negative and decreased as k_{1-test} fell. By comparison, for $k_{1-test} > k_{1-model}$, the cutoff became positive and rose with increasing k_{1-test} (Figure 3d). Importantly, our algorithm captured RRD/RRT signatures in genes scored as non-DEGs by standard pairwise comparison analysis, exposing their biological importance. More generally, under conditions of moderate bias intensity ($-1 < k_{1-test} < 0$), as a fraction of genes were concordantly detected by both approaches, numerous DEGs from pairwise comparison were validated as false-positives or false-negatives by the response-regression framework (Figure 3e).

Taken together, our pipeline corrects systematic biases embedded in conventional pairwise comparison based differential analysis, producing robust, biologically interpretable transcriptomic outputs.

2.4 Global analysis identifies gene sets with characteristic temporal expression kinetics

To better dissect the temporal dynamics of host-responses upon IAV infection, we performed global response-regression-based transcriptomic analysis on longitudinal LD-group datasets. DEGs identified across all time points relative to mock controls were pooled for downstream processing (Figure 4a). Consistent with previous configurations, LD-day3 was designated the model group and LD-day2 the reference group. All LD-derived DEGs ($n=5678$) were categorized into VRGs-up ($n=1983$) and VRGs-down ($n=3695$) according to their expression shifts between model and mock profiles (Figure 4b).

We then separately screened VRGs-up and VRGs-down for RRD and RRT genes using our established pipeline. Notably, a subset of LD-derived DEGs exhibited $\log_2FC$ values in model group (relative to mock) ranging from -1 to 1 . Such values may yield unreasonably large k_2 and produce spurious RRD/RRT assignments. To mitigate this artifact, we applied an additional filtering criterion to exclude genes with $\log_2FC$ values (relative to mock) falling within this range in both the model and test groups. Results from day 1 and day 7, representing the immediate-early and very late infection stages, respectively, are shown in Figure 4c. The corresponding temporal kinetics of each gene set bearing unique response-regression signatures were subsequently visualized using heatmaps.

As shown in Figure 4d, day1-derived RRD VRGs-up were highly expressed at LD-day1 and gradually declined by day3, after which they followed divergent trajectories. One subset underwent re-upregulation at later infection stages, whereas the remaining subset continued to decrease. In contrast, day-1-derived RRT VRGs-up showed minimal induction or even mild downregulation at day 1 but exhibited robust upregulation by day 3. These genes subsequently segregated into persistently upregulated and baseline-restored subsets.

Regarding the day7 results, more hit genes were identified relative to those presented in Figure 3a, mainly owing to expanded inclusion criteria for VRGs-up in this analytical strategy. Nonetheless, these LD-day7-derived RRD and RRT genes recapitulated their characteristic kinetic profiles (Figure 4d). Specifically, RRD genes maintained sustained upregulation throughout infection, whereas RRT genes peaked at day 3 and gradually decayed to or even below baseline levels, indicative of excessive response regression (Figure 4d). Notably, VRGs-

down bearing RRD or RRT signatures exhibited temporal trajectories analogous to those of VRGs-up, but with inverted regulatory directions (Figure 4e).

Furthermore, these genes with individual transcriptional dynamics are enriched in diverse biological processes and functional pathways (Figure 4d&e). The ability to capture genes with such coordinated signatures further validates the performance and advantages of our analytical algorithm.

2.5 The response regression algorithm reveals dysregulated immune responses at early stages in high-dose IAV infected mice

Divergent host responses following LD and HD IAV infection drive distinct disease trajectories and pathological outcomes^{11, 12}. To uncover HD-specific host-response features underlying severe disease, we applied our response regression-based pipeline to compare longitudinal lung transcriptomic profiles from LD- and HD-challenged mice. For this purpose, the total DEGs relative to mock identified at any time points in either LD or HD groups were collected, generating 7651 VRGs for further analysis (Figure 5a). We first characterized early transcriptional divergence from day 1 datasets (Figure 5b), and RRD/RRT genes uniquely detected in either LD-day1 or HD-day1 were examined (Figure 5c).

Notably, RRD VRGs-up and RRT VRGs-down can be pooled for joint analysis, as can RRT VRGs-up and RRD VRGs-down. Figure 5d shows temporal expression profiles of LD-day1-specific RRD VRGs-up and RRT VRGs-down over the infection time course. These genes were rapidly upregulated at LD-day1 but displayed blunted responses at HD-day1. Cell-type profiling demonstrated specific expression of these genes in neutrophils, macrophages/T cells, fibroblasts or epithelial cells (Figure 5d & S4a).

Conversely, LD-day1-specific RRT VRGs-up and RRD VRGs-down were strongly downregulated on day1 in the LD group but refractory to suppression at HD-day1 (Figure 5e). These genes were enriched in fibroblasts, endothelial cells, epithelial cells or adaptive immune-related dendritic/NK/T/B cells (Figure 5e & S4b). HD-day1-specific RRD VRGs-up and RRT VRGs-down were predominantly enriched in endothelial cells or fibroblasts, with exaggerated day1 upregulation in HD relative to LD mice (Figure 5f & S4c). In parallel, HD-day1-specific

RRT VRGs-up and RRD VRGs-down, largely associated with multiple cell types, exhibited excessive day1 downregulation in HD versus LD animals (Figure 5g & S4d).

Collectively, these results demonstrate divergent host responses to LD and HD IAV infection. For innate immune neutrophils and macrophages, LD infection triggers robust antiviral responses to limit viral spread, while in HD models, the responses are insufficient to successfully contain infection, consistent with the low abundance of innate immune cells according to single-cell dynamic profiling (Figure 5h & S5). For fibroblasts, three distinct modules (F1-F3) exhibit complex infection-associated transcriptional remodeling (Figure 5i). The F1 module, associated with myeloid recruitment and early inflammatory alert, was sharply induced at day1 in LD mice and rapidly resolved later, whereas HD mice showed blunted day1 induction followed by deeper suppression. The F2 module, linked to pro-thrombotic and lipid inflammatory signaling, was actively repressed at day1 under LD infection, but this protective suppression was incomplete in HD mice. The F3 module, marking fibroblast activation and matrix remodeling, was delayed until day2 in LD infection and resolved in a self-limited manner. By contrast, HD infection triggered premature and transient upregulation of F3 at day1, followed by profound collapse. Three epithelial gene modules (E1-E3), representing epithelial alarm signaling and basal epithelial priming, epithelial barrier homeostasis and injury restraint, and mucociliary differentiation, respectively, exhibited temporal trajectories highly concordant with their fibroblast counterparts F1-F3, indicating coupled epithelial-stromal responses during IAV infection (Figure 5i).

In summary, our findings pinpointed the early dyscoordination between virus and innate immunity, as well as epithelium-stroma temporal desynchronization, which may subsequently exacerbate severe pulmonary immunopathology.

2.6 High dose IAV infection elicits accelerated disease progression and compromised T cell immunity

We next analyzed lung transcriptomic profiles from HD-day3 to identify genes with aberrant response-regression signatures (Figure 6a), using LD-day2 and LD-day3 as reference and model groups, respectively. As this workflow intrinsically contrasts LD and HD conditions at

day3, we directly interpreted the resulting outputs from HD-day3.

As shown in Figure 6b, VRGs-up in HD-day3 were dominated by RRT, whereas VRGs-down displayed the opposite pattern. As described in Figure 1, most VRGs-up follow canonical response-regression pattern during IAV infection, whereas VRGs-down preferentially acquire RRD at late time points. We therefore hypothesized that the perturbed regression profiles at HD-day3 reflect accelerated infection progression. To test this hypothesis, we compared the RRD/RRT genes derived from HD-day3 to those from LD-day7 (Figure 6c). As shown in Figure 6d, RRT VRGs-up and RRD VRGs-down identified from LD-day7 were largely recapitulated from HD-day3, whereas those RRD VRGs-up and RRT VRGs-down emerged at LD-day7 were scarce at HD-day3. These hit genes identified from either HD-day3 or LD-day7 were further divided into sub-clusters based on their response regression patterns, and temporal-kinetic analysis was performed for each sub-cluster of hit genes.

Notably, for RRT VRGs-up shared between LD-day7 and HD-day3, transcriptional trajectories in the HD group from days 1 to 3 closely recapitulated LD-group dynamics spanning days 1 to 7 (Figure 6e). These findings validated that HD infection substantially compressed the timescales of host responses and established a dysregulated state predisposing to severe disease. Furthermore, RRT VRGs-up specific to HD-day3 (but absent from LD-day7) exhibited even faster response kinetics in HD relative to LD conditions, whereas LD-day7-specific RRT VRGs-up showed little dependence on infection dose. Similar trends were observed for RRD VRGs-down specifically identified at LD-day7, HD-day3, or shared between the two time points (Figure 6f).

In contrast, LD-day7-specific RRD VRGs-up exhibited sustained upregulation in both groups, albeit with blunted induction in HD mice (Figure 6h). While LD-day7-specific RRT VRGs-down followed a comparable trend with persistently lower expression in HD compared to LD animals, albeit they were downregulated across days 1-3 before rebounding and becoming upregulated (Figure 6i). Functional enrichment linked these two LD-day7-specific gene sets to cell-cycle regulation and T-cell activation (Figure S6), pointing to defective T-cell immunity under HD IAV infection. This aligns with insufficient pulmonary T-cell abundance

in HD compared to LD mice starting at day 2 p.i. (Figure S5b).

Moreover, HD-day3-specific RRD VRGs-up and RRT VRGs-down constitute a gene set exhibiting exaggerated upregulation under HD infection relative to LD conditions (Figure 6i). Follow-up gene-by-gene investigations are warranted to further characterize their roles in disease aggravation and assess their potential as biomarkers or therapeutic targets. To summarize, these findings demonstrate that HD IAV infection induces disorganized host responses and immune imbalance marked by impaired T-cell-mediated immunity.

3 Discussion

The outcome of acute viral infection is shaped by dynamic reciprocal interactions between invading pathogens and host immune responses^{2, 4}. Pathogens drive rapid replication and dissemination, whereas host defensive programs contain and ultimately clear infection^{13, 14}. Nevertheless, over-activated antiviral immunity can shift from protective to pathological^{13, 14}. Accordingly, finely tuned, coordinated host responses are indispensable for antiviral defence, whereas dysregulated immune signaling precipitates severe disease manifestations¹⁵. Discerning perturbed versus coordinated host responses is therefore important to unravelling viral pathogenesis.

Traditional approaches usually rely on absolute response magnitudes to define underactive or overactive immunity¹⁶⁻¹⁹. Although these methods readily detect overt immune dysregulation, they cannot disentangle early immune perturbations, wherein physiological virus-elicited responses and nascent dysregulation remain intertwined. Here, we conceptualized host immune dysregulation as an independent regulatory input and deployed a response-regression model to decouple its effects on host responses from those imposed by viral load fluctuations³. As a proof-of-concept, we validated this analytical pipeline using longitudinal lung transcriptomic profiles from mice infected with LD or HD IAV, two infection conditions that recapitulate balanced and dysregulated immune states, respectively⁹.

Of note, the dynamics of virus-host interplay correlate but do not closely align with, chronological time post-infection. Caution should therefore be exercised when comparing host responses across clinical or experimental conditions. Accordingly, we incorporated temporal

cues as an additional regulatory input and analyzed longitudinal transcriptomic data from lung tissues of LD IAV-infected mice using our response-regression-based algorithm. This analysis identified multiple gene sets bearing distinct temporal kinetic signatures and functional roles, enabling assessment of infection progression irrespective of sampling time in future.

Cross-comparison between LD and HD infection uncovered HD-specific dynamic transcriptional signatures representing potential biomarkers or pathogenic mediators of severe disease. At day 1 p.i., genes associating to neutrophils or macrophages were strongly induced in LD-infected mice but not under HD challenge. This counterintuitive expression pattern aligns with aberrant innate immune responses in HD-infected animals, which provide insufficient early antiviral defence²⁰. In addition, epithelium-stroma dysregulation was also observed at day1 in HD-infected mice compared to the LD counterparts.

Time-course analyses further illuminated mechanisms driving disease exacerbation. At HD-day3, numerous gene sets exhibited response-regression patterns comparable to those observed at LD-day7, supporting that pathologically accelerated host-response trajectories constitute a major driver of severe illness. These aberrant signatures were dominated by RRT VRGs-up and RRD VRGs-down. It can be assumed that quantification of representative genes within these categories may enable objective evaluation of disease progression stages. By contrast, LD-day7-specific RRD VRGs-up and RRT VRGs-down were largely undetectable in HD-day3 profiles. Instead, these genes exhibited persistently repressed expression in HD mice relative to LD counterparts at matched time points. Functional enrichment linked these two gene sets to cell-cycle regulation and T-cell-mediated immunity, implying that HD IAV infection triggers T-cell immune dysfunction that propagates disease severity. It will be interesting to further investigate these genes with unique response regression patterns at different time points and explore the underlying biological implications in the future.

Despite its advantages over conventional pairwise differential analysis, our framework has several intrinsic limitations. First, the model relies on simple linear regression to quantify host-response magnitude relative to viral stimulus, which cannot fully capture gene-specific heterogeneity. Larger datasets together with artificial-intelligence-driven strategies will support

future model refinement. Second, empirical estimation of global host-response intensity across model, reference and test groups may introduce systematic analytical bias. Subsequent work should select typical gene sets that faithfully obey response-regression rules to derive more robust k_1 values.

In summary, this study re-interprets host-response dynamics under viral infection through a response regression-based analytical paradigm. This strategy circumvents systematic biases inherent to standard pairwise comparisons and generates more robust biological readouts. Comparing LD and HD infection establishes a new framework to discriminate immune homeostasis from immune dysregulation, nominating candidate biomarkers, pathogenic mediators and therapeutic targets for severe influenza. Furthermore, our pipeline can be compatible with multi-omics datasets from viral infection models and is broadly extendable to other infectious pathogens.

4 Methods and materials

4.1 Data source and scRNA-seq analysis

Longitudinal transcriptomic profiles of IAV-infected mouse lung tissues were derived from the CNCB single-cell RNA-seq dataset PRJCA034049 (<https://www.cncb.ac.cn/>)⁹. Quality control was completed in Seurat v5.0.1 (R v4.5.0), preserving cells with 300-7000 features, <10,000 RNA counts and <10% mitochondrial reads. Filtered data were merged, normalized and scaled with SCTransform, followed by PCA and UMAP dimensionality reduction (RunUMAP, dims=1:20). A pseudobulk matrix was generated through AggregateExpression to eliminate single-cell technical variation. Genes with raw read counts below 50 across all samples were filtered out prior to subsequent transcriptomic analyses.

4.2 Response regression model-based transcriptome analytical pipeline

4.2.1 Definition of differentially expressed genes (DEGs) and viral responsive genes (VRGs)

Transcriptomic profiles of lung tissues from IAV-infected mice were individually pairwise compared against the mock group or predefined model group. The fold change (FC) of each gene from all IAV-infected groups were calculated. Genes with $|\log_2\text{FC}| > 1$ were identified as DEGs.

For algorithm optimization, DEGs derived from model versus mock comparison were designated as VRGs. These genes were further classified into VRGs-up and VRGs-down according to positive and negative $\log_2FC$ values, respectively. For global analysis, DEGs identified from all IAV-infected groups were pooled as total VRGs. Based on their expression patterns in the model group relative to the mock group, total VRGs were also divided into VRGs-up and VRGs-down.

4.2.2 Relative status examination of each VRG in test group to mock and model profiles

The response regression index (RI) of each gene was calculated as the ratio of $\log_2FC$ from test group vs. mock group to that from model group vs. mock group, as shown in Formula (1):

$$RI_{(geneA, test)} = \frac{\log_2FC_{(geneA, test-vs-mock)}}{\log_2FC_{(geneA, model-vs-mock)}} \quad (1)$$

4.2.3 Evaluation of the overall intensity of host responses

$\log_2FC$ values of all DEGs in each group were plotted against the model group, followed by linear regression analysis. The slope of the fitted regression curve was used to quantify the overall intensity of host responses relative to the mock group. The relative parameter k_1 for each test group was defined as the ratio between the slope of the test group and that of the reference group (Formula 2):

$$k_{1-test} = \frac{slope_{test}}{slope_{reference}} \quad (2)$$

For algorithm optimization where only VRGs-up from model vs. mock comparison were included in downstream analyses, k_1 was alternatively calculated as the ratio between the median RI of VRGs-up in the test group to the median RI of the reference group (Formula 3):

$$k_{1-test} = \frac{median RI_{test}}{median RI_{reference}} \quad (3)$$

4.2.4 Normalization of RI value

To predict the theoretical response regression statuses of host genes in the test group under viral load changes while in the absence of pharmacological effects, the estimated RI (Est.RI) values were calculated using the relative k_1 between test and reference groups (Formula 4):

$$Est. RI_{(geneA, test)} = \frac{k_{1-test}}{k_{1-reference}} \times RI_{(geneA, reference)} \quad (4)$$

The normalized RI (Norm.RI), which reflects pharmacological regulation on target genes, was obtained by subtracting Est.RI from the actual RI of the test group (Formula 5):

$$Norm.RI_{(geneA, test)} = RI_{(geneA, test)} - est.RI_{(geneA, test)} \quad (5)$$

4.2.5 Screening for VRGs with response regression deviation or transcendence

To screen VRGs with response regression deviation (RRD) or transcendence (RRT), norm.RI values of all qualified VRGs in the test group were plotted against RI values of the reference group. Hook functions were applied to set thresholds for screening. Specifically, RRD genes were filtered using the following threshold (Formula 6, $y > 0$):

$$x > \frac{(a-y)}{k_1} + \frac{k_2}{k_1 y} \quad (6)$$

RRT genes were identified based on the criterion below (Formula 7, $y < 0$):

$$x < \frac{(a-y)}{k_1} + \frac{k_2}{k_1 y} \quad (7)$$

where a represents a constant derived from formula 8, while k_2 is variable and used to adjust the stringency of screening.

$$a = \frac{k_{1-test}}{k_{1-model}} \quad (8)$$

4.3 Heatmap visualization

Longitudinal expression patterns of hit genes were visualized in heatmaps using $\log_2FC$ values relative to the mock group. For dynamic curve plotting, the median $\log_2FC$ and interquartile range (IQR; Q25-Q75) were calculated for the indicated gene sets at each time point in both LD and HD groups. Median values were connected to form kinetic trajectories, and IQR ranges were overlaid as shaded bands.

4.4 Functional enrichment analysis of hit gene sets

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed using the OmicShare platform to explore biological functions and pathways associated with hit gene sets ²¹.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgement

This study was supported by the National Natural Science Foundation of China (82574901; 82274397), Natural Science Foundation of Shandong Province, China (ZR2024QH110) and Shandong Provincial Key Research and Development Program (2026CXGC010612).

Data availability statement

All relevant data that support the findings of this study are presented in the manuscript and supplementary information file.

Author Contributions

Ruikun Du, Shaobo Wang and Rong Rong conceptualized the project and design of the research. Ruikun Du, Xuran Gu, Ziwei Li, Jichen Yang and Xiaoping Zhang contributed to data collection and analysis. Ruikun Du wrote the original manuscript. Shaobo Wang and Rong Rong edited the manuscript. Ruikun Du and Rong Rong provided grant support. All authors read and approved the submitted manuscript.

Reference

1. Pommerenke, C. et al. Global transcriptome analysis in influenza-infected mouse lungs reveals the kinetics of innate and adaptive host immune responses. *PLoS One* **7**, e41169 (2012).
2. Shah, P.S. et al. Systems Biology of Virus-Host Protein Interactions: From Hypothesis Generation to Mechanisms of Replication and Pathogenesis. *Annu Rev Virol* **9**, 397-415 (2022).
3. Du, R. et al. Response regression model as a guiding principle for pharmacological mechanistic investigation of host-targeted antivirals. *Acta Materia Medica* **4** (2025).
4. Alanazi, A., Ibrahim, M.N., El Azab, E.F. & Elithy, M.A. Host Cell Virus Interactions: Molecular Mechanisms, Immune Modulation, Viral Pathogenesis, and Emerging Therapeutic Targets. *Viruses* **18** (2026).
5. Wallis, R.S., O'Garra, A., Sher, A. & Wack, A. Host-directed immunotherapy of viral and bacterial infections: past, present and future. *Nat Rev Immunol* **23**, 121-133 (2023).
6. Iwasaki, A. & Medzhitov, R. Control of adaptive immunity by the innate immune system. *Nat*

- 475 *Immunol* **16**, 343-353 (2015).
- 476 7. Blanco-Melo, D. et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of
477 COVID-19. *Cell* **181**, 1036-1045.e1039 (2020).
- 478 8. Chang, D., Dela Cruz, C. & Sharma, L. Beneficial and Detrimental Effects of Cytokines during
479 Influenza and COVID-19. *Viruses* **16** (2024).
- 480 9. Zheng, Y. et al. Influenza A virus dissemination and infection leads to tissue resident cell injury
481 and dysfunction in viral sepsis. *EBioMedicine* **116**, 105738 (2025).
- 482 10. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
483 RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
- 484 11. Liao, M. et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19.
485 *Nat Med* **26**, 842-844 (2020).
- 486 12. Wang, Y. et al. A large-scale single-cell transcriptomic atlas indicates the immune panorama of
487 influenza A infection. *Imeta* **5**, e70121 (2026).
- 488 13. Herro, R. & Grimes, H.L. The diverse roles of neutrophils from protection to pathogenesis. *Nat*
489 *Immunol* **25**, 2209-2219 (2024).
- 490 14. Jung, K.I., McKenna, S., Vijayamahantesh, V., He, Y. & Hahm, B. Protective versus Pathogenic
491 Type I Interferon Responses during Virus Infections. *Viruses* **15** (2023).
- 492 15. Wei, X., Narasimhan, H., Zhu, B. & Sun, J. Host Recovery from Respiratory Viral Infection. *Annu*
493 *Rev Immunol* **41**, 277-300 (2023).
- 494 16. Su, Y. et al. Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate
495 COVID-19. *Cell* **183**, 1479-1495.e1420 (2020).
- 496 17. Liu, J. et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the

- 497 peripheral blood of SARS-CoV-2 infected patients. *EBioMedicine* **55**, 102763 (2020).
- 498 18. Kuri-Cervantes, L. et al. Comprehensive mapping of immune perturbations associated with severe
- 499 COVID-19. *Sci Immunol* **5** (2020).
- 500 19. Channappanavar, R. & Perlman, S. Pathogenic human coronavirus infections: causes and
- 501 consequences of cytokine storm and immunopathology. *Semin Immunopathol* **39**, 529-539 (2017).
- 502 20. Rawat, S., Vrat, S. & Banerjee, A. Neutrophils at the crossroads of acute viral infections and
- 503 severity. *Mol Aspects Med* **81**, 100996 (2021).
- 504 21. Mu, H. et al. OmicShare tools: A zero-code interactive online platform for biological data analysis
- 505 and visualization. *Imeta* **3**, e228 (2024).
- 506

Figures and Figure legends

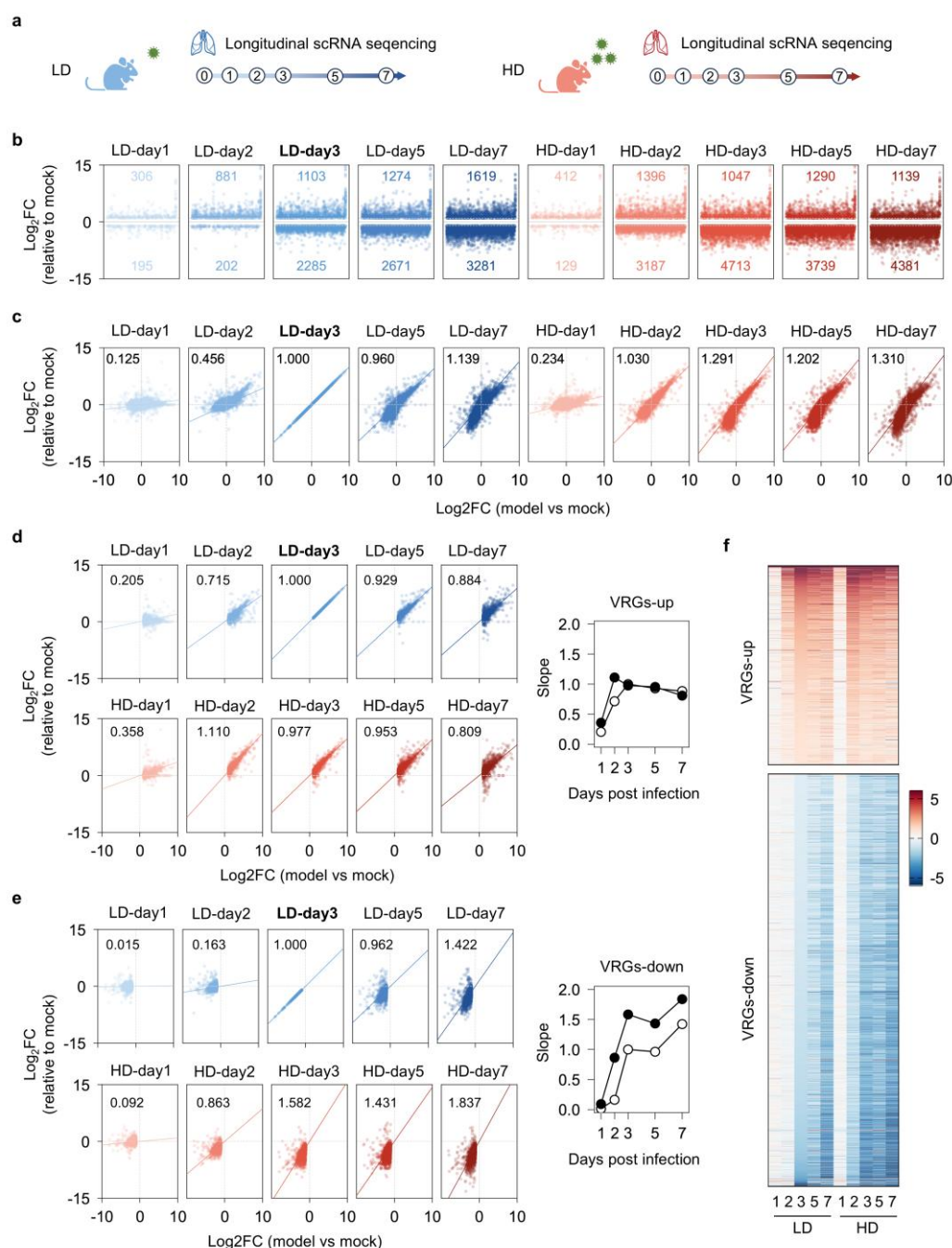


Figure 1 Longitudinal lung transcriptomic responses in mice infected with influenza A virus (IAV). (a) Schematic overview of the longitudinal single-cell RNA-seq experimental design. (b) DEGs at each time point were identified relative to mock controls, with volcano plots generated per time point. (c) Global host response magnitude across infection. Log₂FC values (vs mock) of all genes were plotted against LD day-3 (designated as model group). The slope denotes the regression coefficient from linear fitting of the scatter plot. (d–e) DEGs up-

515 and downregulated in LD-day3 versus mock were defined as viral-responsive genes (VRGs)-
516 up and VRGs-down. Longitudinal kinetics of VRGs-up **(d)** and VRGs-down **(e)** under LD/HD
517 IAV infection are presented. Curves illustrate log₂FC slopes (relative to the model group) over
518 infection for VRGs-up and VRGs-down. **(f)** Heatmaps of VRGs-up and VRGs-down using
519 log₂FC values relative to mock from mouse lung tissues across sequential IAV infection time
520 points.
521

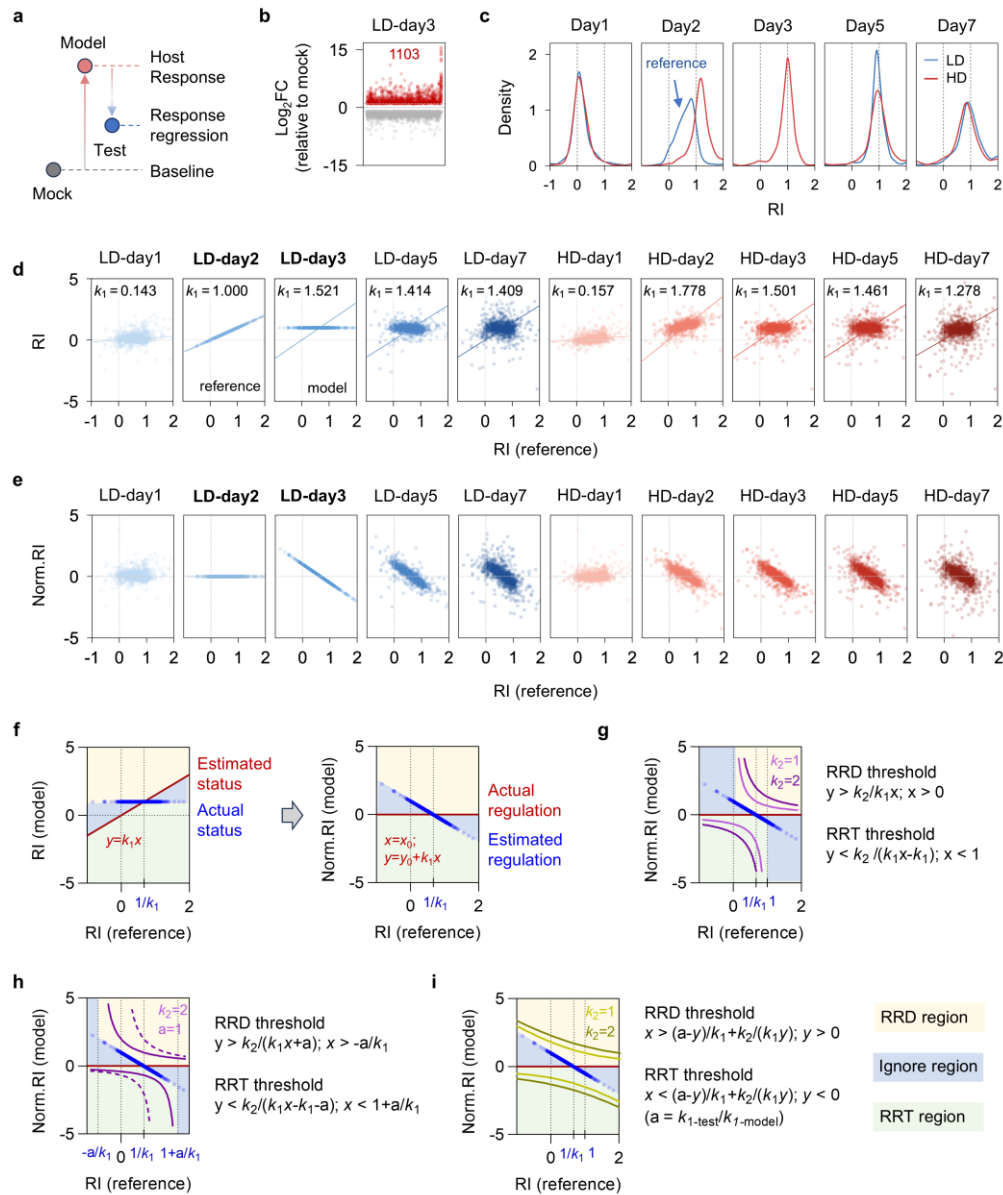


Figure 2 Optimization of the response-regression-based transcriptomic analytical algorithm. (a) Schematic illustration of the response-regression model. (b) Model-group-derived VRGs-up were used to rationalize and optimize the response regression-based transcriptomic analytical pipeline. (c) Distribution of RI values for model-derived VRGs-up in each group. LD-day2 was defined as reference group for subsequent analysis. (d) RI values of VRGs-up across transcriptomic datasets were plotted against reference (LD-day2) RI. Parameter k_1 is the ratio of median RI between test and reference groups. (e) Normalized RI values of VRGs-up across transcriptomic datasets were calculated and plotted against reference (LD-day2) RI. (f) Schematic defining response regression deviation (RRD) and response

532 regression transcendence (RRT), representing perturbations driven by non-viral factors. **(g)**
533 Reciprocal-function-based thresholding. Two core reference RI values (0 and 1) were used to
534 remove low-confidence data. Parameter k_2 controls filtering stringency. **(h)** Modified
535 reciprocal-function thresholding. Horizontal shifts of reciprocal curves adjust the width of the
536 low-confidence range. **(i)** Hook-function-based thresholding. Hook functions establish less
537 subjective thresholds. Parameter k_2 controls filtering stringency.
538

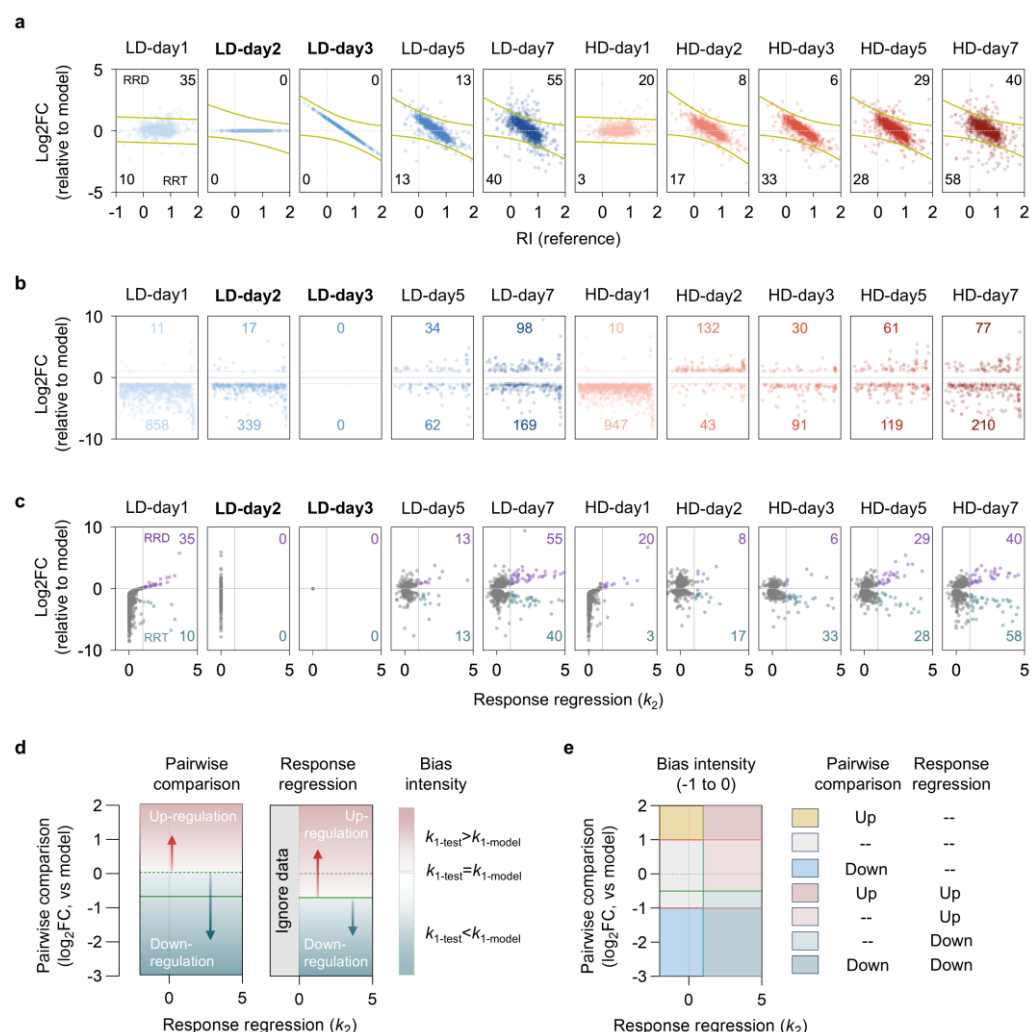


Figure 3 Performance comparison of the response-regression model versus conventional pairwise comparison in transcriptomic analysis. (a) Model-group-derived VRGs-up were screened for RRD/RRT genes at each time point, with k_2 set to 1. **(b)** VRGs-up transcriptomic profiles at each time point were pairwise-compared against the model group to compute log2FC for volcano plots. **(c)** Side-by-side DEG screening results from pairwise comparison and the response-regression model. Log2FC (vs model group) was plotted against response-regression-derived k_2 for all included VRGs-up. **(d)** Schematic of the two screening strategies. Pairwise comparison introduces global biases driven by variable viral loads, whereas the response-regression model mitigates these biases. **(e)** Distribution of gene sets detected by the two approaches. Some gene subsets are concordantly retained or excluded. While certain pairwise-comparison DEGs are bias-associated false positives per the response-regression model, many genes discarded in pairwise tests can be recovered with the response regression framework.

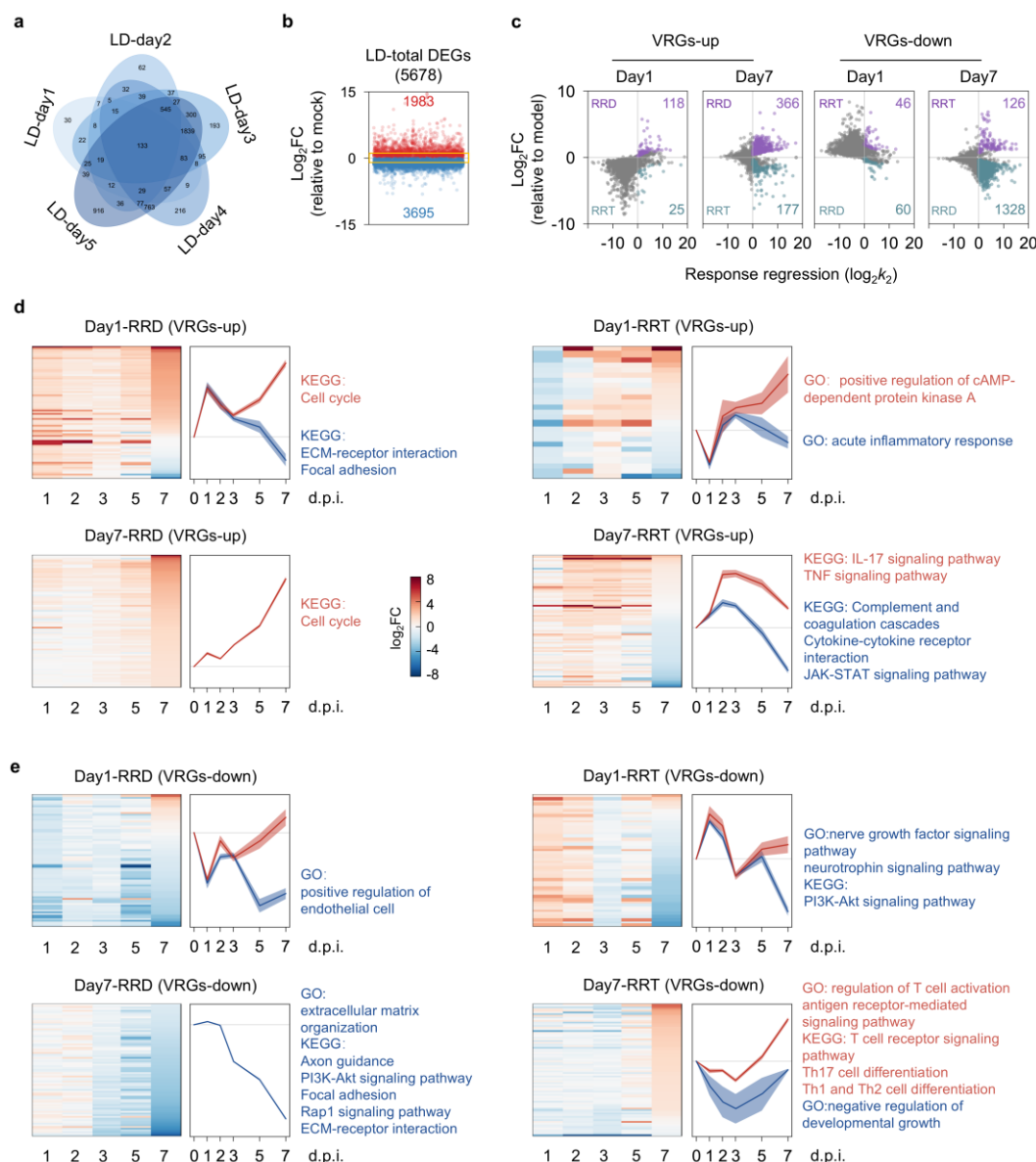


Figure 4 Global screening of host genes with distinct temporal expression kinetics during IAV infection. (a) Venn diagram of total DEGs across LD-group time points relative to mock controls. (b) VRG classification. Total LD-group DEGs (relative to mock) were pooled and assigned to VRGs-up and VRGs-down according to expression differences between the model (LD-day3) group and mock controls. (c) Screening of RRD/RRT VRGs at LD-day1 or LD-day7. VRGs-up and VRGs-down with RRD/RRT at indicated time points are plotted separately. Log₂FC (relative to model group) versus response-regression-derived k_2 values are shown for all included VRGs. (d–e) Heatmaps depicting temporal kinetics of VRGs-up (d) and VRGs-down (e) with characteristic response-regression features. Representative kinetic curves and

562 enriched pathways are displayed in the right panels.

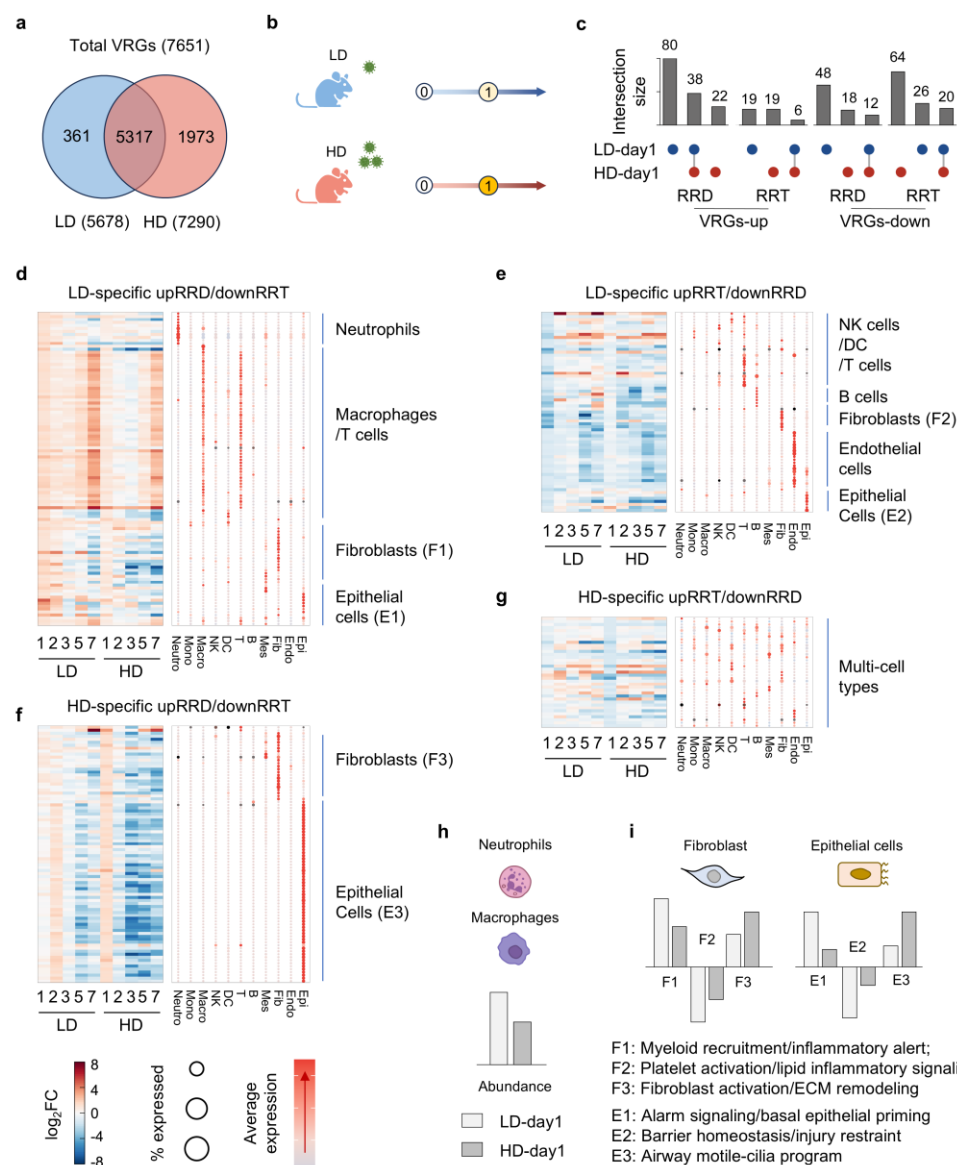


Figure 5 Identification of genes with divergent viral responsiveness at the immediate-early infection stage in LD and HD IAV-challenged mice. (a) Total DEGs across time points from LD or HD groups were pooled for downstream analysis. **(b)** Schematic for LD–HD comparison at day1. RRD/RRT genes from LD-day1 and HD-day1 were compared to uncover immune dysregulation-associated genes. **(c)** Upset diagrams showing counts of hit VRGs with unique response-regression features at LD-day1 or HD-day1. **(d–f)** Heatmaps of temporal expression kinetics for indicated hit VRGs with characteristic response-regression signatures. Highly expressing cell types are annotated. **(h)** Comparison of innate immune response at day1 under LD and HD infection for neutrophils and macrophages. **(i)** Comparison of the coordinated and

573 dysregulated epithelium-stroma response at day1 in LD and HD models, respectively.

574

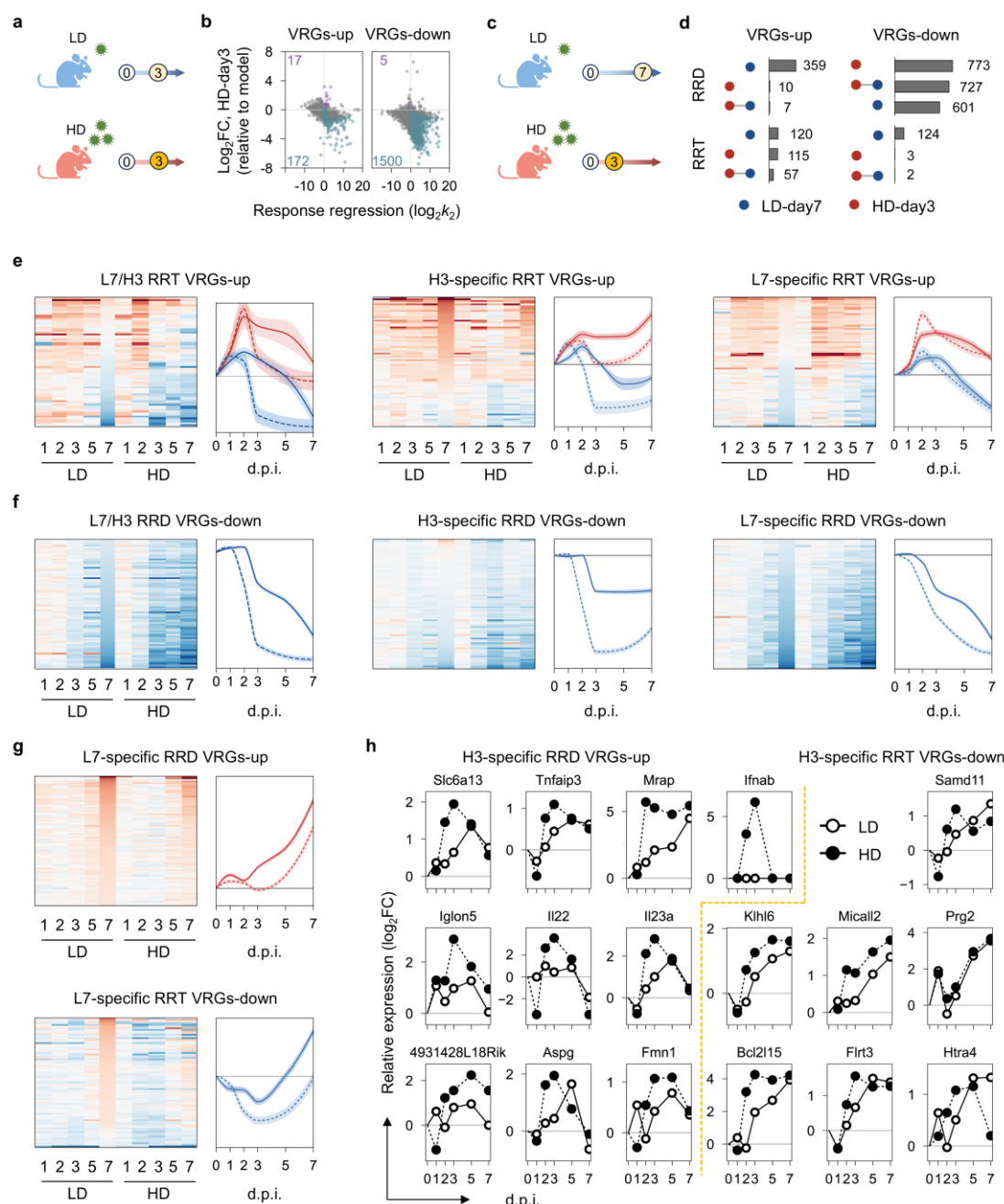


Figure 6 Identification of gene sets potentially associated with disease aggravation during IAV infection. (a) Schematic for LD–HD comparison at day 3. RRD/RRT genes from LD-day3 and HD-day3 were compared to uncover immune dysregulation-associated genes. (b) Screening of RRD/RRT VRGs from HD-day3 using LD-day3 and LD-day2 as model and reference groups, respectively. Log₂FC value (relative to model group) was plotted against response-regression-derived k_2 for all included VRGs. (c) Schematic for comparison between HD-day3 and LD-day7. RRD/RRT genes from LD-day7 and HD-day3 were compared to identify immune dysregulation-associated genes. (d) Upset diagram showing hit genes with

584 RRD/RRT signatures from HD-day3 or LD-day7 groups. **(e–g)** Heatmaps of temporal
585 expression kinetics for indicated hit VRGs with characteristic response-regression signatures.
586 Representative kinetic curves were shown in right panels. **(h)** Temporal expression kinetics of
587 RRD VRGs-up and RRT-VRGs-down specifically identified from HD-day3. Solid lines
588 indicate LD kinetic curves, and dashed lines show HD kinetics.