Emerging Microbes & Infections
2025, VOL. 14, 2494705 (18 pages)
https://doi.org/10.1080/22221751.2025.2494705

Reading of human acute immune dynamics in omicron SARS-CoV-2 breakthrough infection

Haibo Li [a,b*], Hongyu Liu [a,c*], Hongping Wu [a*], Chang Guo [d*,**], Wenting Zuo [a], Ying Zheng [c], Xiaoyan Deng [e], Jiuyang Xu [a], Yeming Wang [a], Zai Wang [f], Binghuai Lu [a], Baidong Hou [g,h] and Bin Cao [a,b,c,f]

[a]National Center for Respiratory Medicine; State Key Laboratory of Respiratory Health and Multimorbidity; National Clinical Research Center for Respiratory Diseases; Institute of Respiratory Medicine, Chinese Academy of Medical Sciences; Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China-Japan Friendship Hospital, Beijing, People's Republic of China; [b]New Cornerstone Science Laboratory, Beijing, People's Republic of China; [c]Department of Respiratory Medicine, Capital Medical University, Beijing, People's Republic of China; [d]Changping National Laboratory (CPNL), Beijing, People's Republic of China; [e]Tsinghua University-Peking University Joint Center for Life Sciences, Beijing, People's Republic of China; [f]Institute of Clinical Medical Sciences, China-Japan Friendship Hospital, Beijing, People's Republic of China; [g]State Key Laboratory of Epigenetic Regulation and Intervention, Institute of Biophysics, Chinese Academy of Sciences, Beijing, People's Republic of China; [h]University of Chinese Academy of Sciences, Beijing, People's Republic of China

ABSTRACT

The dynamics of the immune response to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) breakthrough infections remain unclear, particularly when compared to responses in naive individuals. In this longitudinal prospective cohort study, 13 participants were recruited. Peripheral blood samples were collected every other day until day 7 after symptom onset. Transcriptome sequencing, single-cell sequencing, T-cell receptor (TCR) sequencing, B-cell receptor (BCR) sequencing, Olink proteomics, and antigen-antibody binding experiments were then performed. During the incubation periods of breakthrough infections, peripheral blood exhibited type 2 cytokine response, which shifted to type 1 cytokine response upon symptom onset. Plasma cytokine levels of C-X-C motif chemokine ligand 10, monocyte chemoattractant protein-1, interferon-γ, and interleukin-6 show larger changes in breakthrough infections than naïve infections. The inflammatory response in breakthrough infections rapidly subsided, returning to homeostasis by day 5 after symptom onset. Notably, the levels of monocyte-derived S100A8/A9, previously considered a marker of severe disease, physiologically significantly increased in the early stages of mild cases and persisted until day 7, suggesting a specific biological function. Longitudinal tracking also revealed that antibodies anti-Receptor Binding Domain (anti-RBD) in breakthrough infections significantly increased by day 7 after symptom onset, whereas cytotoxic T lymphocytes appeared by day 5. This study presents a reference for interpreting the immunological response to breakthrough infectious disease in humans.

ARTICLE HISTORY Received 18 December 2024; Revised 19 March 2025; Accepted 13 April 2025

KEYWORDS COVID-19; Breakthrough Infection; Immune response; SARS-CoV-2; Innate immunity; Early stage

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) affects billions of people worldwide. Despite 13 billion vaccines having been administered globally [1], breakthrough infections and reinfections in vaccinated and previously infected individuals are increasingly common [2]. Frequent breakthrough infections also occur with most respiratory viruses, such as influenza [3], rhinovirus [4], and common coronaviruses [5]. The disease burden of breakthrough infections of respiratory viruses is significant, with approximately 1 billion people infected with influenza annually, and countless others infected with various respiratory viruses. Among those with influenza, 3–5 million progress to severe cases, posing a serious societal threat [6]. The transition to severe disease often occurs in the second week after symptom onset [7, 8]. Therefore, whether patients can restore homeostasis through an effective immune response or gradually experience immune dysregulation during the first week post-symptom onset is critical.

Previous studies have primarily focused on comparing immune responses at a single time point between severe and mild cases, as well as on pathogen escape [9, 10], antigen recognition [11–14], and

CONTACT Baidong Hou baidong_hou@ibp.ac.cn; Bin Cao caobin_ben@163.com
*These authors contributed equally to this work.
**Present address: State Key Laboratory of Epigenetic Regulation and Intervention, Institute of Biophysics, Chinese Academy of Sciences, Beijing, People's Republic of China

Supplemental data for this article can be accessed online at https://doi.org/10.1080/22221751.2025.2494705.

vaccine efficacy [15–18]. Understanding of how patients restore homeostasis through an effective immune response during the first week after symptom onset is limited. Human challenge studies have conducted in-depth analyses of the immune response after primary infection onset, including viral kinetics, plasma cytokine levels, and antibody levels [19–21]. Nevertheless, the understanding of the immune changes during the first week of breakthrough infection and how they differ from naïve infections is inadequate [22]. Possible reasons for this gap include the difficulty in determining prior infection history and the challenge in obtaining consecutive clinical samples in the very early stages of illness. Additionally, recent studies have associated elevated S100A8/A9 (also referred to as calprotectin or MRP8/14) expression with severe COVID-19 progression, indicating a potential role in driving excessive inflammatory responses [23,24]. However, S100A8/A9 may also function as a crucial component of normal immune responses, and insufficient levels could be disadvantageous [25].Hence, the precise role of S100A8/A9 in SARS-CoV-2 pathogenesis remains incompletely understood, particularly in the context of early-stage.

In December 2022, anticipating a large-scale infection event in China, we designed an experiment to follow patients who experienced breakthrough infections after vaccination. Peripheral blood was collected and subjected to multimodal immune function testing. Thirteen volunteers were recruited either before or after symptom onset and followed until day 7 from symptom onset. Pre-symptom samples were obtained from six patients. For all 13 participants, samples were collected every other day. Through transcriptome sequencing, single-cell sequencing, T-cell receptor (TCR), B-cell receptor (BCR) sequencing, Olink proteomics, and antigen–antibody binding experiments, a comprehensive longitudinal analysis of the immune response was performed during breakthrough infections at the acute stage. The findings offer a detailed atlas of the recall immune response after infection during the first week.

Methods

Study design

The prospective cohort study was conducted at China-Japan Friendship Hospital. This study employs a longitudinal research design, involving the continuous measurement of the same patients at multiple predetermined time points. During winter 2022 after the end of the zero-COVID policy, majority of Chinese citizens faced the possibility of being infected by SARS-CoV-2. To capture the natural process of immune response for SARS-CoV-2, healthy volunteers not infected or within one week after symptom onset were enrolled. To be eligible for participation, volunteers needed to be at least 18 years of age with no chronic lung diseases, immune system diseases, or pneumonia over the previous six months, no cold or fever symptoms in the past three months, not pregnant or breastfeeding, no history of cancer, and no glucocorticoid use within 6 months were enrolled in this study. All subjects had received inactivated COVID-19 vaccine 3 times. Prior to the Omicron infection, all participants had consistently tested negative for SARS-CoV-2 across multiple rounds of nucleic acid and antibody testing, which were conducted as part of regular, government-mandated surveillance protocols. Nucleic acid testing or antigen testing was used to confirm whether the study participants were infected by SARS-CoV-2 or not during the study period. On the day of enrolment, blood samples were collected immediately. Subsequently, the samples were taken every other day until day 7 after symptoms onset. Sex assigned at birth of participants was determined based on self-report. The sex data indicates that there were 11 female and 2 male participants.

Sample collection

Approximately 6 ml of peripheral blood was collected from each enrolled patient at every sampling time. A total of 47 peripheral blood samples were collected. Blood samples were centrifuged with Ficoll-Paque PLUS (Cytiva, Cat# 17144003) within 30 min of collection. Plasma was collected and stored directly in a −80°C freezer. After the PBMCs were collected, they were washed with PBS (Invitrogen), centrifuged, and a freezing solution containing 90% FBS (Cat# 10099141C) was added containing 10% DMSO (Cat# 67-68-5, Sigma). The samples were aliquoted into 1 ml tubes, frozen by gradient cooling, and stored at −80°C for 3 days. They were transferred to liquid nitrogen. PBMCs were subjected to bulk transcriptome sequencing by directly adding to Trizol for RNA extraction. PBMCs for single-cell sequencing were stained with AO/PI reagent after recovery, followed by counting and enrichment sequencing.

BULK RNA sequencing and data processing

Total RNA from 47 samples from 13 participants was extracted using Trizol reagent (Invitrogen, Cat# 15596026) according to the manufacturer's protocol and reverse-transcribed into cDNA. Bulk RNA-seq was performed using the Illumina Novaseq6000 with PE150 read length. The raw reads were evaluated for quality by Fastp (v.0.23.2) [26] and aligned to human GRCh38 to obtain the raw counts. The Transcripts Per Million (TPM) date were processed by DESeq2 (v.1.42.0) [27]. The gene expression data is filtered to remove genes with low average expression

across samples (lower than the 20th percentile of the average expression values). Hierarchical clustering is performed on the scaled and filtered gene expression data. Based on the silhouette coefficients, clustering was performed with K = 9. Principal Component Analysis (PCA) was employed to explore the variations of different patients. ComplexHeatmap package was used for visualization.

DESeq2 was used to identify the DEGs (different expression genes) for different groups by (P.djust < 0·05) by count Data. The DEGs for each group were compared with the pre-symptoms group and visualized using a Venn plot (https://bioinfogp.cnb.csic.es/tools/venny/index.html). The DEGs were grouped by the hclust of R. The Go terms were enriched by clusterProfiler(v.4.0.2) [28] and visualized by ggplot2.

Olink proteomics and data processing

Total 47 plasma sample from 13 participants was measured using the Olink® Target 96 inflammatory panel (Olink Proteomics AB, Uppsala, Sweden) according to the manufacturer's instructions. The final assay read-out is presented in Normalized Protein eXpression (NPX) values, which is an arbitrary unit on a log2 scale, in which a high value corresponds to higher protein expression. The categorization of the fitting curves is performed using MFuzz [29] (https://github.com/junjunlab/ClusterGVis).

The data from challenge study [20] underwent the following transformation, facilitating comparison with our Olink proteomics data.

$$Y = Log2\left[\frac{Y}{Ybaseline}\right]$$

The results are fitted to a curve by A fourth-order polynomial fit in ggplot2 (v.3.4.1). The comparison of trends between two sets of fitted curves was conducted using Spearman's ρ (rho) coefficient.

The optimal number of clusters is determined based on the degree of match between the fitting curves and their categorization.

Identification of SARS-CoV-2 genotypes

13 Nasopharyngeal swabs from 13 patients were using to identify of SARS-CoV-2 genotypes. Enriched SARS-CoV-2 whole genome amplification was performed with an equimolar mixture of primers using the SARS-CoV-2 Full Length Genome Panel following the manufacturer's protocol (2205, Genskey, Beijing, China). Briefly, for COVID-19 patient samples, total RNA was extracted and reverse-transcribed to synthesize first-strand cDNA. PCR amplification was conducted for 35 cycles using SARS-CoV-2-specific primers. Following amplification, the PCR product was purified using DNA clean beads, and sequencing libraries were constructed by enzyme digestion and PCR-free library preparation reagents. The prepared library was then sequenced on an MGISEQ-2000 platform with 100 bp single-end reads using a sequencing reaction kit (GS-2000-FCS-SE100, Genskey, Beijing, China).

B cell enrichment and mixing process

After the cells were recovered, they were added to ice-cold RPMI1640 culture medium. After AO/PI staining, the following operations were performed according to the number of living cells. For living cells of more than 500,000, negative enrichment of B cells was done using antibody magnetic beads. For living cells less than 500,000, a total of 10,000 cells were directly captured for sequencing. The B cell-negative enrichment operation was done according to the manufacturer's instructions (EasyTM Human Pan-B cell Enrichment Kit, Stem Cell. Cat# 19554). In brief, 50ul Enrichment Cocktail was added to 0·5 ml PBMC cell suspension (0·5-5 x10^6). Mix well and stand at room temperature for 10 min. Then 75ul vortexed Magnetic Particles was added to the mixture and leave it at room temperature for 5 minutes. Then 2·5 mL EasySep™ Buffer (Cat# 20144) was added, and the mixture was gently pipet 2–3 times, and placed on the magnetic stand for 5 minutes. The directly poured the cells into another 5 ml tube, and the tube was centrifuged at 300 g at 4°C for 5 min. Then supernatant was discarded. 0·1 ml 1640 + 2%FBS medium was added and the cell suspension was filtered using a 37um cell Strainers. After enrichment, B cells and PBMC were mixed at a ratio of 1:1. Finally, 10,000 cells were captured and then sequenced.

scRNA library preparation, sequencing, and data processing

Cell suspensions of 25 samples from 5 patients were barcoded using the ReagentSeekOne® Digital Droplet Single Cell 5' library preparation Kit (SeekGene Cat# K00501-24), SeekOne® DD Single Cell TCR Enrichment Kit (Human) (SeekGene Cat# K00601-24), and SeekOne® DD Single Cell BCR Enrichment Kit (Human) (SeekGene Cat# K00701-24). Single-cell RNA libraries were prepared following the manufacturer's instructions. The sequencing libraries had a unique sample index, individually. The libraries were sequenced using an Illumina NovaSeq 6000 with PE150 read length.

The raw scRNA data was aligned to human GRCh38 to obtain a gene expression matrix using SeekSoul®Tools. Low-quality cells were filtered using Seurat (4.3.0) [30] based on cells, in which the number of detected genes <200 or >5000 were deleted. Next,

we used Median Absolute Deviation to filter cells affected by mitochondrial genes and the filtered counts were used for the downstream analysis.

The scRNA doublet data was predicted using Scrublet (v.0.2.3) [31] using a typical workflow for each sample. We next integrated the datasets of different sample cells into a shared space for unsupervised clustering by the harmony algorithm (v.0.1.0) [32] for batch effect correction with default parameters. We then identified the variable genes by FindVariableGenes Seurat function and determined the variability of the numbers of UMIs by ScaleData Seurat function. Dimension reduction was done with RunPCA and RunUMAP Seurat function. We clustered the cells using FindClusters Seurat function by the Leiden algorithm[33] for the 20 dims at a resolution of 0·8. The characteristic genes in these clusters were identified by FindAllMarkers Seurat function, which compares each cluster to the others. These marker genes were expressed in a minimum of 10% of the cells in their cluster and at a minimum log fold-change threshold of 0·25 (FDR < 0·01, logFC > 0·25, Wilcoxon Rank-Sum test). The clusters which expressed two sets of well-studied canonical markers of the major cell types were labelled as doublets and the erythrocytes, which highly expressed *HBA, HBB, and HBD,* were excluded in the following analysis. In total, 6 major cell types were identified, including B cells (*MS4A1, CD79A, CD79B*), Myeloid cells (*CD14, LYZ, CST3*), NK cells (*GNLY, NKG7, CD247*), T cells (*CD3D, CD3E, CD4, CD8A, CD8B*), proliferating cells (*Mki67, TOP2A*) and Plasma cells (*JCHAIN, MZB1*). These major clusters were further classified into 26 subsets representing different cell subpopulations within major cell lineages. Their canonical and signature marker genes for each cluster were deposited in Fig. S4, 5, S6c. The functional enrichment results of the cell subpopulations were analyzed by ClusterGVis (v.0.1.1) and Mfuzz(2.62.0).

Gene ontology (GO) enrichment analysis of the DEGs was implemented through the clusterProfiler (v.4.0.5) R package [34]. GO terms with a corrected *p*-value less than 0·05 were considered significantly enriched for DEGs. KEGG [35] was used to identify the high-level functions and utility of biological systems (e.g. cells, organisms, and ecosystems) from molecular-level information, specifically, the generation of large-scale molecular data sets. Genome sequencing and other high-throughput experimental technologies (http://www.genome.jp/kegg/). The clusterProfiler R package was used to test the statistical enrichment of the DEGs in KEGG pathways.

TCR and BCR analysis

TCR/BCR sequences of the 25 samples from 5 patients were assembled and quantified using the SeekSoul®Tools vdj protocol by GRCh38. The outputs of the assembled contigs for the TCR were filtered and processed by scRepertoire (v.1.7.0) [36] and the clonotype was analyzed for each T cell. The downstream analysis of the TCR was done using scRepertoire. The clonotypes of the BCR were also analyzed using Change-O (v.1.3.0) to examine the BCR clonotype dynamics following the Change-O workflows [37]. The polar plot was used to show the usage of IGHV. The Sankey diagram was used to display αβ VJ pairs. if barcodes are identical, they are used to match the α and β VJ regions.

Enzyme-linked immunosorbent assay (ELISA)

Total 47 plasma sample from 13 participants was measured using ELISA to determine the amount of antigen-specific immunoglobulins in serum, SARS-CoV-2 RBD (Wuhan-Hu-1, homemade), RBD (BA.4/5, SinoBiological, Cat# 40592-V08H130), N(Wuhan-Hu-1, SinoBiological, Cat# HPLC-40588-V07E)and RBD (BF.7, SinoBiological, Cat# 40592-V08H140) were used to coat 96-well plates (CORNING, REF: 3690) overnight at 4°C. Serially diluted sera were added and incubated, and the horseradish peroxidase (HRP)-conjugated anti-human IgG(H + L) (Promega, REF: W4031) was used for detection. 3,3',5,5'-tetramethylbenzidine (Sigma-Aldrich) was used as the HRP substrate, and the optical density at 450 nm was measured using a microplate reader (SpectraMax, Molecular Devices, USA), and analyzed with GraphPad Prism software. The results were plotted using the A450 value of each well. The total AUC was then calculated and compared with the standards.

Area under curve (AUC)

The AUC is calculated as follows: OD450-570 value as the y-axis and the log value of the serum dilution corresponding to the well as the x-axis to establish a curve, and then y = 0 as the baseline to calculate the area under the curve (AUC). Each point represents the data from one patient.

Symptom evaluation

Enrolees were asked about their relevant symptoms. For patients with pre-symptomatic samples, their symptoms were prospectively recorded at the time of each sampling. For patients who only had post-onset samples, their symptoms before enrolment were recorded retrospectively, whereas their symptoms after enrolment were recorded for each sampling time and recorded prospectively. Symptoms included diarrhoea, vomiting, nausea, headache, shortness of breath dyspnoea, muscle or body aches (or soreness),

chills or shaking, fatigue, sore or dry throat, congestion or runny nose, and cough. Patients were first asked if they had any of these symptoms. If so, referring the human challenge experiment, the symptoms were divided into three levels, absence of symptoms, one just noticeable, whereas the other being clearly bothersome some of the time.

Statistical analysis

For descriptions of patient symptom numbers, mean values and standard errors (SE) were utilized. Antibodies titers were characterized using their mean values and SE. The results from Olink proteomics were presented using loess curve fitting. For the comparison between longitudinal symptoms and severity, two-way RM ANOVA mixed-effects models were used. Normality was tested using the Shapiro-Wilk test, and multiple comparisons were conducted using Dunnett's multiple comparisons test and Šídák's multiple comparisons test. The longitudinal comparison of antibodies and the comparison among different variants were conducted using a Mixed-effects model (restricted maximum likelihood, REML). For multiple comparisons, Dunnett's multiple comparisons test was employed. Geisser-Greenhouse correction was used for sphericity. GO enrichment analysis uses hypergeometric distribution tests and controls the false discovery rate by adjusting *P*-values using the Benjamini-Hochberg method, ultimately selecting significantly enriched GO terms (q-value < 0·05). The Spearman rank correlation test and wilcoxon test was employed to compare the trends in cytokine changes between this study and the challenge study. The number and percentage of subgroups for single-cell sequencing results were described using mean values and SE. The top 5 upregulated and downregulated differential genes in myeloid cells, comparing pre-symptom and post-symptom periods, were identified and visualized using volcano plots. Temporal variations in different BCR subtypes were shown by their mean proportion.

For all comparison, two-sided *P* values less than 0·05 were considered statistically significant. GraphPad (version 9.5.1) and R (version 4.2.3) were employed for statistical analyses.

Results

Symptom development and dynamics of antibody response

A total of 13 volunteers, aged 23–35 years, participated in this study. Peripheral blood samples were collected immediately after enrolment, and participants were followed until day 7 from symptom onset. The day on which patients exhibited symptoms was defined as day 1, and the sampling time before symptom appearance was defined as pre-symptoms. Samples were obtained from six volunteers before illness onset (Figure 1(a)). Samples from patients who completed all sampling points were selected for single-cell, TCR, and BCR sequencing. All samples underwent transcriptome sequencing, Olink proteomics, and antigen–antibody binding experiments (number of samples = 47). Symptoms were recorded from symptom onset till day 7 (Figure 1(b)). The average number of symptoms that were just noticeable was 3·35 on day 1, which slowly declined to 3·28 on day 3 and then sharply fell to 1·57 on day 7. For symptoms that were just noticeable, a significant difference was noted on days 5 ($P = 0·0010$) and 7 ($P < 0·0001$) compared with day 1. As for symptoms that are unbearable, a significant numerical difference only appears on day 7 compared with day 1 ($P = 0·0321$). Between these two categories of symptoms, a significant difference was observed until day 7 ($P < 0·0001$, day 1; $P < 0·0001$, day 3; $P = 0·0095$, day 5; Figure 1(c)). The primary symptoms on day 1 were fatigue, sore or dry throat, and fever, and those on day 7 were cough, congestion, or runny nose (Figure 1(d)). On days 1 and 3, the most unbearable symptom was fever (Fig. S1a). Furthermore, compared with naïve infections, a higher proportion of patients had a fever [20]. In addition, all these patients recovered without taking any antiviral medications.

Before the study, all volunteers had received three doses of the inactivated virus vaccine against the wild-type viral strain (Table S1). Among them, 11 volunteers had received their last booster vaccination over 300 days earlier. With BF.7 and BA.5, which were both closely related, as the predominant prevalent strains during this period, the antibody response to the wild type (WT), BF.7, and BA.5 strains was tested. The most apparent characteristic was that these patients had low antibody levels before disease onset, below the antibody levels 2–10 weeks after infection post-vaccination, and comparable to the levels retained 1 year after booster vaccination. However, these levels were significantly higher than those in patients who had neither been vaccinated nor infected $P = 0.0066$, WT; $P = 0.0158$, BA.5; $P = 0.0027$, BF.7. The antibody levels against the three variants were increased after infection, showing a significant difference on day 7 compared with pre-symptoms ($P = 0·0453$, WT; $P = 0·0304$, BA.5; $P = 0·0464$, BF.7). In some patients, antibodies had already increased highly by day 3 (Figure 1(e–g)). In a cross-sectional comparison of the antibody levels on day 1, the levels against WT were significantly higher than those against BF.7 ($P = 0·0242$) and BA.5 ($P = 0·0115$). However, no significant difference was found at later time points, indicating that antibodies against the two variants had been induced more rapidly (Figure 1(h,i)). Dynamic trend of N protein antibody titers is essentially consistent with that of the RBD

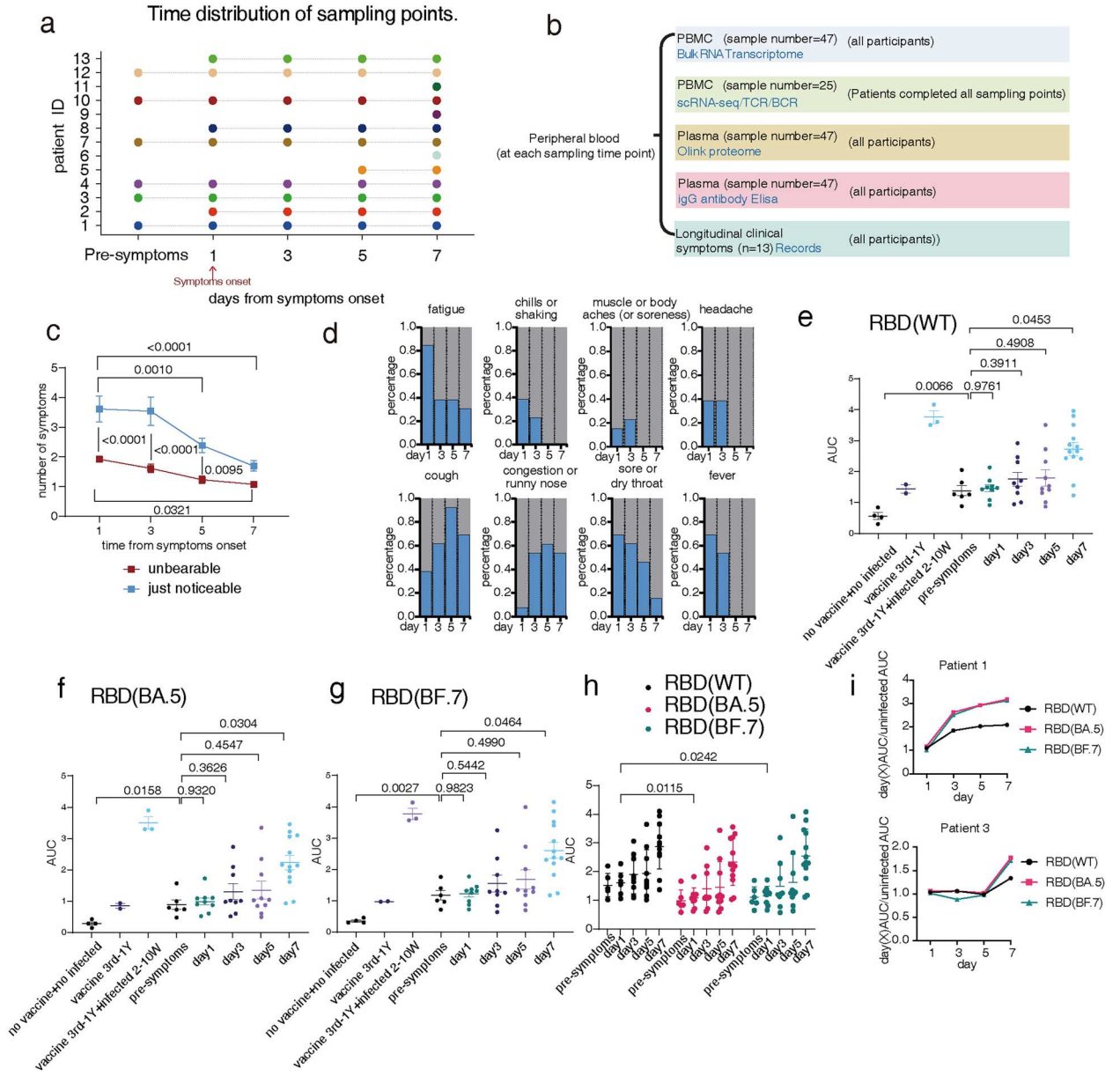

Figure 1. Symptoms development and dynamic of antibody response (a) Sample collection time points. Each coloured row represents a patient, and each point represents a sample collection. (b) The experiments conducted after collecting the samples, sample number = the sample number of for each experiment. *n* = the number of subjects for each experiment. (c) A dotted line graph shows the number of symptoms changing over time. The dot represents the mean. *n* = 13, the lines connect the means, and the error bars represent the standard error. (d) Changes in the percentage of just noticeable symptoms over time. (e-g) Area Under Curve (AUC) of spike RBD–specific IgG against WT (e), BA.5 (f), and BF.7 (g). (h) AUC of spike RBD–specific IgG against WT, BA.5, and BF.7. The line represents the mean and the error bars represent the standard error. (i) The ratio of day(X) AUC to uninfected baseline for RBD(WT), RBD(BA.5), and RBD(BF.7) in Patient 1 and 3 from days 1, 3, 5, and 7 post-symptom onset, illustrating the relative rise in antibody titers across variants.Sample number = 47.

group (Fig. S1b). Considering both the antibody levels and types of antibodies that respond quickly, the reactions of these individuals are consistent with a memory immune response after a breakthrough infection.

Profile of bulk transcriptome during the breakthrough of the infection course

Bulk RNA sequencing was then performed on all samples. Utilizing agglomerative hierarchical clustering, Pre (red) and day 1 (green) were clearly clustered separately, indicating that the peripheral blood mononuclear cell (PBMC) transcriptome also undergoes significant changes associated with symptom onset (Figure 2(a)). Furthermore, samples of pre-symptoms were grouped with some patients of day 7, indicating that many patients had already returned to their pre-symptomatic state. The outlier, patient 4's pre-symptom point, clustered with others' day 1 (Figure 2(a)), indicating that this patient might have already been in an infectious state at that time combined with the sharply high levels of interferon-stimulated genes (ISG)-dominated cytokines during the pre-symptom stage (Fig. S1c).

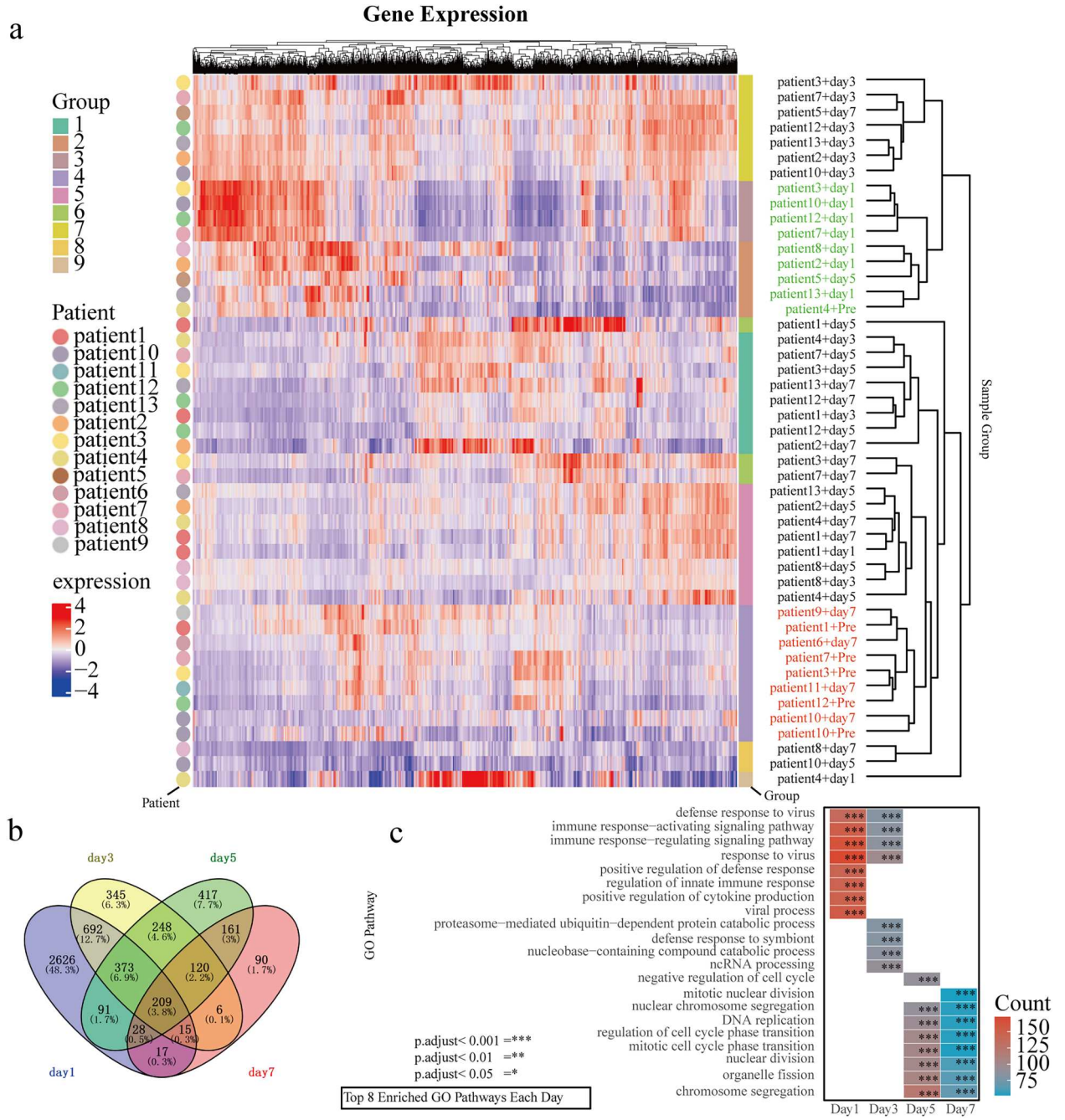

Figure 2. Profile of bulk transcriptome during the breakthrough of the infection course. (a) Heatmap from the gene expression of bulk RNA transcriptomics and hierarchical clustering of samples; sample number = 47. The colour bar on the right represents the group of samples. The groups were defined through hierarchical clustering on scaled and filtered gene expression data, using K = 9 based on silhouette coefficients. (b) Venn diagram comparing different genes of each day to the pre-symptoms by bulk transcriptome sequencing following the exclusion of samples from patient 4. Sample number = 42. Each colour represents one day. The ratio inside the parentheses represents the proportion of that quantity among all differentially expressed gene. (c) Top 8 enriched Gene Ontology (GO) terms in each day. Colours represent enrich gene counts in this term. Red represents a higher quantity, while blue represents a lower quantity. Asterisks denote the *P-adjust* value, with three asterisks indicating an adjusted *P*-value < 0·001. Sample number = 42.

After the removal of outliers, a Venn diagram analysis was conducted. The results indicate that the number of genes specifically expressed on day 1 was the highest, reaching 2,626. Subsequently, a rapid decline was noted, with only 90 specific expression genes remaining by day 7(Figure 2(b)). Then, top 8 enriched Gene Ontology (GO) biological process analysis was conducted in each day (*P-adjust* < 0·001 for all top 8 term in each day, Figure 2(c)). On day 1, the predominant terms were defense response to the virus, innate immune response, and response to the virus. These processes decreased by day 3 and did not rank within the top 20 in days 5–7 (Fig. S1d). Beginning on day 5, cell division and mitotic cell cycle processes were upregulated and continued through day 7, indicating PBMC proliferation

(Figure 1(c), Figure S1d). Overall, the transcriptome exhibits certain regular changes after symptoms appear; however, individual differences occur.

Trends in temporal changes in four types of plasma cytokines

Plasma cytokine levels were quantified using Olink proteomics. By analyzing temporal variations in 92 cytokines, we identified four distinct trends (Figure 3(a)). The first trend, exemplified by monocyte chemotactic protein-1,2 (MCP-1,2), entails a rapid peak after symptom onset, and is followed by a swift decline, with a return to baseline levels between days 3–5. The second trend, exemplified by CX3CL1, is marked by a slow decline after peaking with symptom onset, with a return to baseline levels around day 7. The trend of the first two was consistent with the changes in symptoms. The third trend, exemplified by interleukin (IL-20) and Artemin (ARTN), features a continuous decline after symptom onset. The class 4, with representative cytokines IL17A and IL17C, was characterized by a peak reached on day 3. (Figure 3(b)). After classifying all cytokines, it was found that the cytokines related with type 1 immune responses[38], which focus on a robust proinflammatory response to control infections and activate phagocytic cells, were primarily located within classes

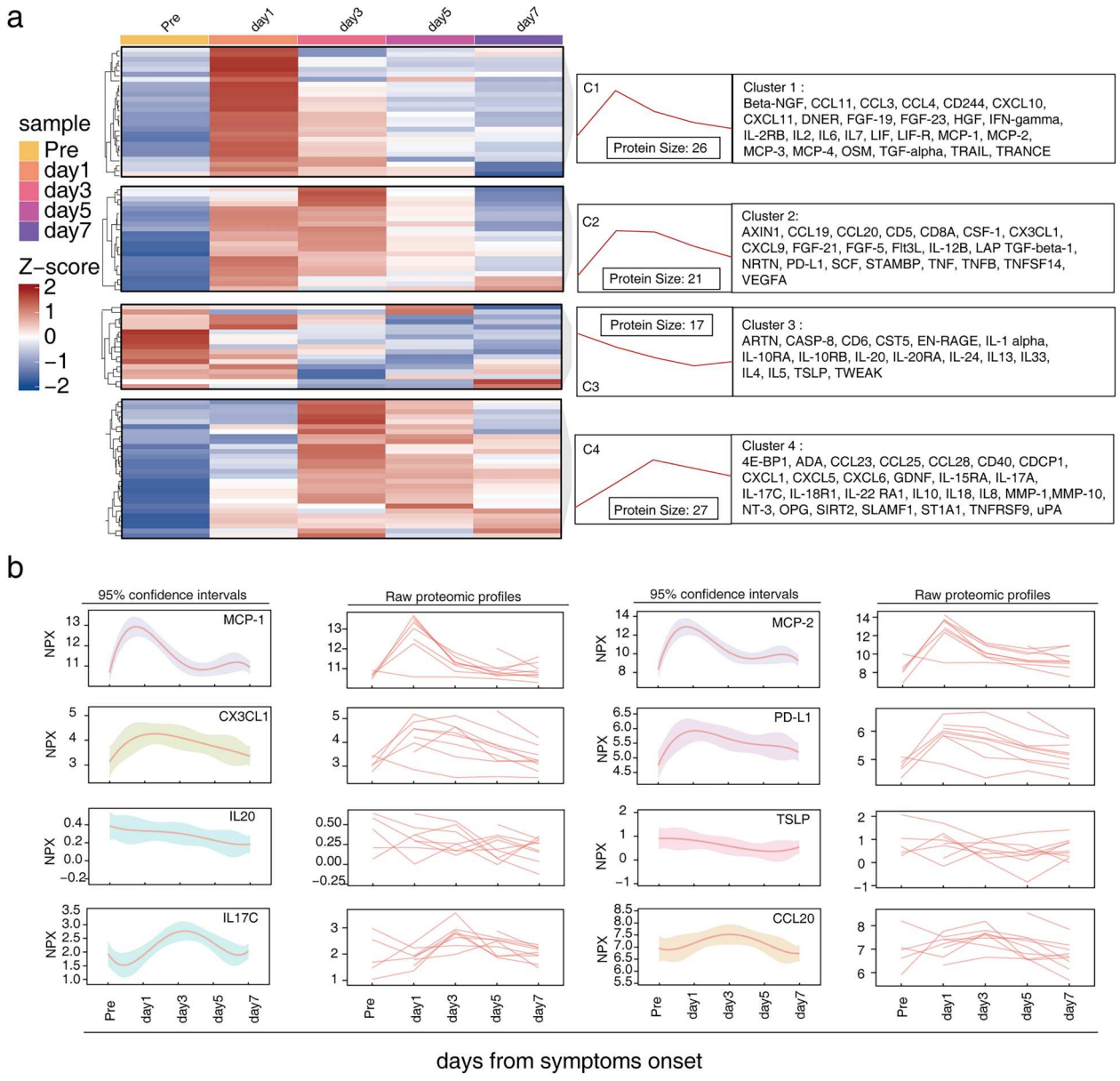

Figure 3. The trends of temporal changes in four types of plasma cytokines. (a) Heatmap displaying the four dynamic expression patterns of genes belonging to inflammatory panel of Olink proteomics during different days. Left. Z-score of different classes of cytokines in each day, Middle, trends in variation among different Class and the number of cytokines they contain. Right, cytokines are categorized from the Class 1(C1) to Class4(C4). Sample number = 42. (b) Representative genes of differing trends. Left, data are presented using fourth-order polynomial fitting. Right, each line represents the data from one patient. Solid lines represent the fitted curves, while the light-coloured areas represent the 95% confidence intervals. Sample number = 42.

1 and 2, whereas those related to type 2 immune responses[39], which promotes antibody production and supports responses that are more anti-inflammatory and often focus on barrier protection and repair rather than on intense phagocytic activity, predominantly align with class 3. Consequently, before symptom onset, the expression of type 2 cytokines[40] was already evident in peripheral blood, potentially corresponding to an early mucosal reaction in the respiratory tract. Subsequently, as a widespread antiviral response emerged in the peripheral blood, accompanied by various symptoms, the expression of type 2 cytokines gradually decreased, transitioning to a predominance of type 1 cytokine[41] (Figure 3(a)). Then, the transcriptomics data from peripheral blood PBMCs were analyzed, and a total of 39 plasma cytokines with corresponding genes expressed in PBMCs were identified. Certain class 1 cytokines, such as CXCL10 and TRAIL, showed expression trends in PBMCs consistent with changes observed in the plasma (Fig. S2).

The challenge study included 34 young adult participants who had neither received a SARS-CoV-2 vaccination nor had a documented infection. These individuals were intranasally inoculated with a pre-Alpha SARS-CoV-2 strain, representing naïve infections.[21]. We compared them with our participants who had been vaccinated against SARS-CoV-2 and later experienced breakthrough infections.The timing of day 1 in breakthrough infections closely aligned with day 4 in naïve infections (Figure 4). The change patterns of eight cytokines under the two infection scenarios were generally consistent, with similar trends observed in the fitted curves (CXCL10, rho = 0·43, $P < 0·001$; MCP-1, rho = 0·33, $P < 0·001$; tumour necrosis factor [TNF], rho = 0·31, $P < 0·001$; CCL4, rho = 0·20, $P < 0·001$; IL10, rho = 0·21, $P < 0·001$; IL6, rho = 0·20, $P < 0·001$; IL18, rho = 0·08, $P = 0·0259$; interferon-gamma [IFN-γ], rho = 0·12, $P = 0·0015$; Figure 4(a–h)). Interestingly, in naïve infections, a subset of cytokines (CXCL10, MCP-1, IL6, TNF, IL10, and IFN-γ) exhibited a secondary rise on day 8 (corresponding to day 5 of breakthrough infections) (Figure 4(a–e, h)). In contrast, breakthrough infections did not exhibit similar changes, except for IL10 (Figure 4(a–h)). The peak levels of CXCL10, MCP-1, CCL4, IL10, IL6, and IFN-γ in breakthrough infections were higher than those in naïve infections (Figure 4(i)). This demonstrates that certain cytokines exhibited a higher level of activation during the early stages of breakthrough infections. In summary, the consistency of the overall trend between naïve infections and breakthrough infections highlights the feasibility and comparability of conducting early-stage immunological research in real-world settings, and the symptoms are associated with cytokine change trends 1 and 2.

Rapid ISG response and shift from antiviral to alarmin response among innate immune cells

To further explore intricate changes in diverse immune cells, single-cell and TCR/BCR sequencing were performed using frozen PBMCs from five patients in whom pre-symptom samples had been collected. To enhance plasma cell yield, B cells were enriched and subsequently added to the original PBMCs at a ratio of 1:1 to enhance plasma cell yield. Then, 10,000 mixed cells per individual were captured and sequenced (Fig. S3a).

Cell clusters were identified using established markers (Table S2). Six major groups were annotated: T, B, plasma, natural killer (NK), myeloid, and proliferating cells (Figures 5(a), S4a, S4b). Myeloid, T, B, and plasma cells were further subdivided into subclusters (Fig. S3b–d) based on marker expression (Figure S4c, S4d, S5a–d). The numbers of myeloid cells increased significantly on day 1, T and NK cells notably decreased (Fig. S3e). Further analysis showed that classical monocytes were the most abundant myeloid cell population (Figure 5(b)). Within proliferating cells, the predominant subpopulations were NK, plasmablasts, and CD4 T cells (Fig. S6a–c).

In the analysis of the enrichment function across all subcluster cells by various terms, all subclusters peaked on day 1 in terms of ISGs, virus response, and immune response; however, these categories rapidly decreased. Most clusters returned to baseline homeostasis levels, except for the monocyte and dendritic cell (DC) subclusters (Figures 5(c), S7a, S7b). The enrichment of cell chemotaxis and inflammatory response increased in myeloid cells and increased gradually (Figure 5(d), S7c). All subclusters were enriched for the term "cytokine production involved in immune response" on day 1. Enrichment within T cells quickly decreased, whereas the B subcluster cells remained continually activated until day 5. Meanwhile, the myeloid subcluster maintained its activation state through day 7 (Fig. S7d). The score of the cell cycle indicated that proliferating cells achieved the highest score on day 5, aligning well with BULK RNA-sequencing results (Fig. S7e).

Given the different trends in changes observed in myeloid cells during the first 7 days, an in-depth analysis of these cells was performed. The top 5 genes expressed by myeloid cells at various time points were analyzed (Figure 5(e), S8a–d). The results indicated that in the first 3 days, the main highly expressed genes belonged to the ISG, including *CXCL10* ($P < 0·01$), *IFIT1* ($P < 0·01$), *IFIT2* ($P < 0·01$), *IFIT3* ($P < 0·01$), and *ISG20* ($P < 0·01$), whereas *S100A8* ($P < 0·01$) and *TNFSF10* ($P < 0·01$) were enriched in a partial subcluster. On days 5 and 7, *S100A8* ($P < 0·01$) and *S100A9* ($P < 0·01$) emerged as the major highly expressed genes.

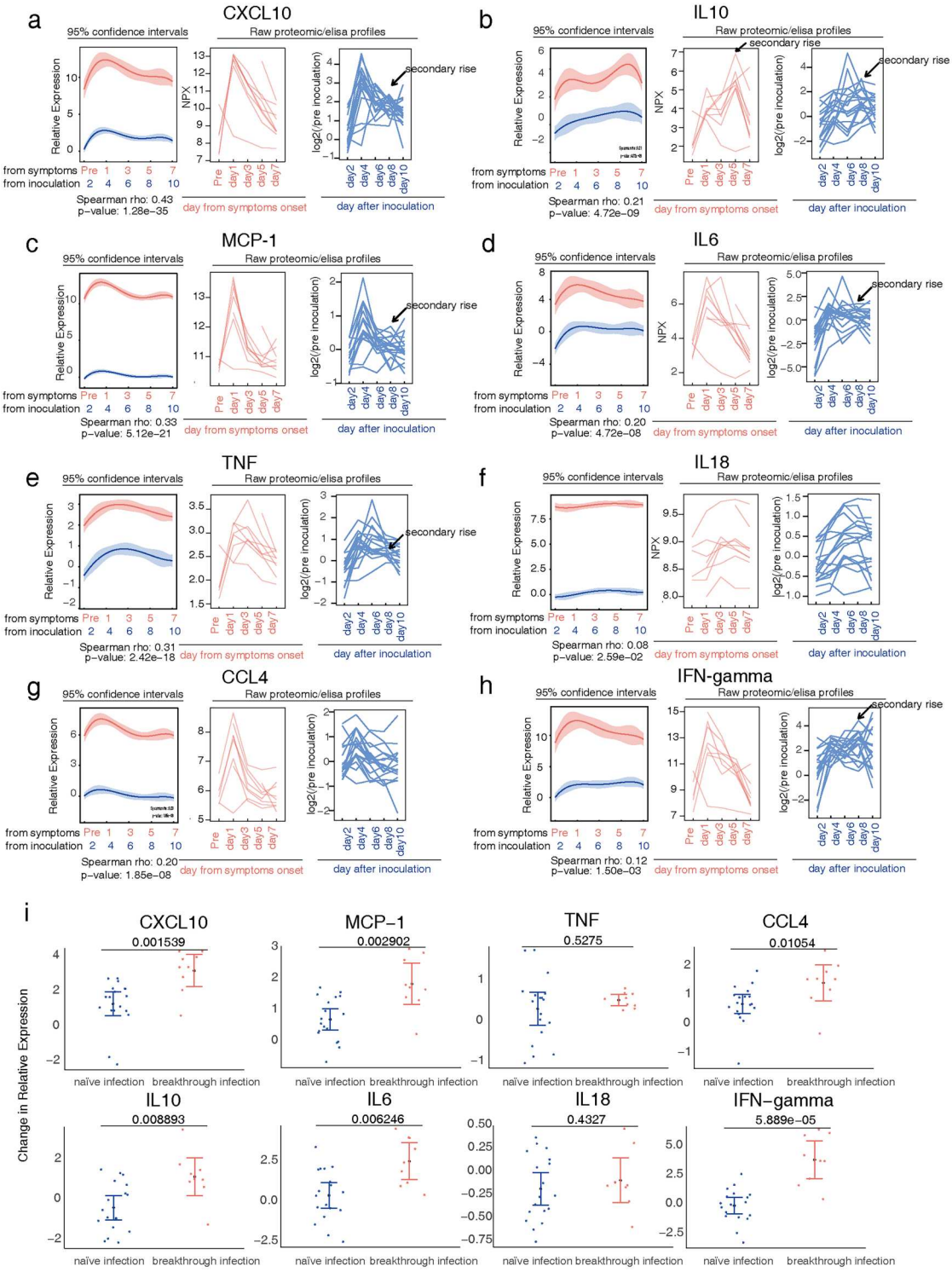

Figure 4. Compare the key cytokine changes between breakthrough and naïve infections. (a – h) Temporal changes in eight plasma inflammatory cytokines among breakthrough and naïve infections. For each graph, red lines represent cytokines from breakthrough infection, and blue lines represent cytokines from naïve infection. Left, Solid lines represent fourth-order polynomial fitting curves for the cytokines corresponding to the two patient groups. Light-coloured areas indicate the 95% confidence intervals. Right, raw Olink proteomic/ELISA profiles for breakthrough and naïve infections. The black arrows indicate a secondary rise. $n = 12$ for breakthrough infection; $n = 18$ for naïve infection. (i) Comparison of the relative expression changes between naïve and breakthrough infections. For naïve infections, the change in relative expression levels was calculated by subtracting day 10 data from day 4 data. For breakthrough infections, the change in relative expression levels was calculated by subtracting day 7 data from day 1 data.

In addition, a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was conducted for these two continuously increasing biological processes (cell chemotaxis and inflammatory response) in five myeloid cell types. On day 1, the highly expressed genes were predominantly those associated with the pro-inflammation immune response pathway, such as Nucleotide-binding Oligomerization

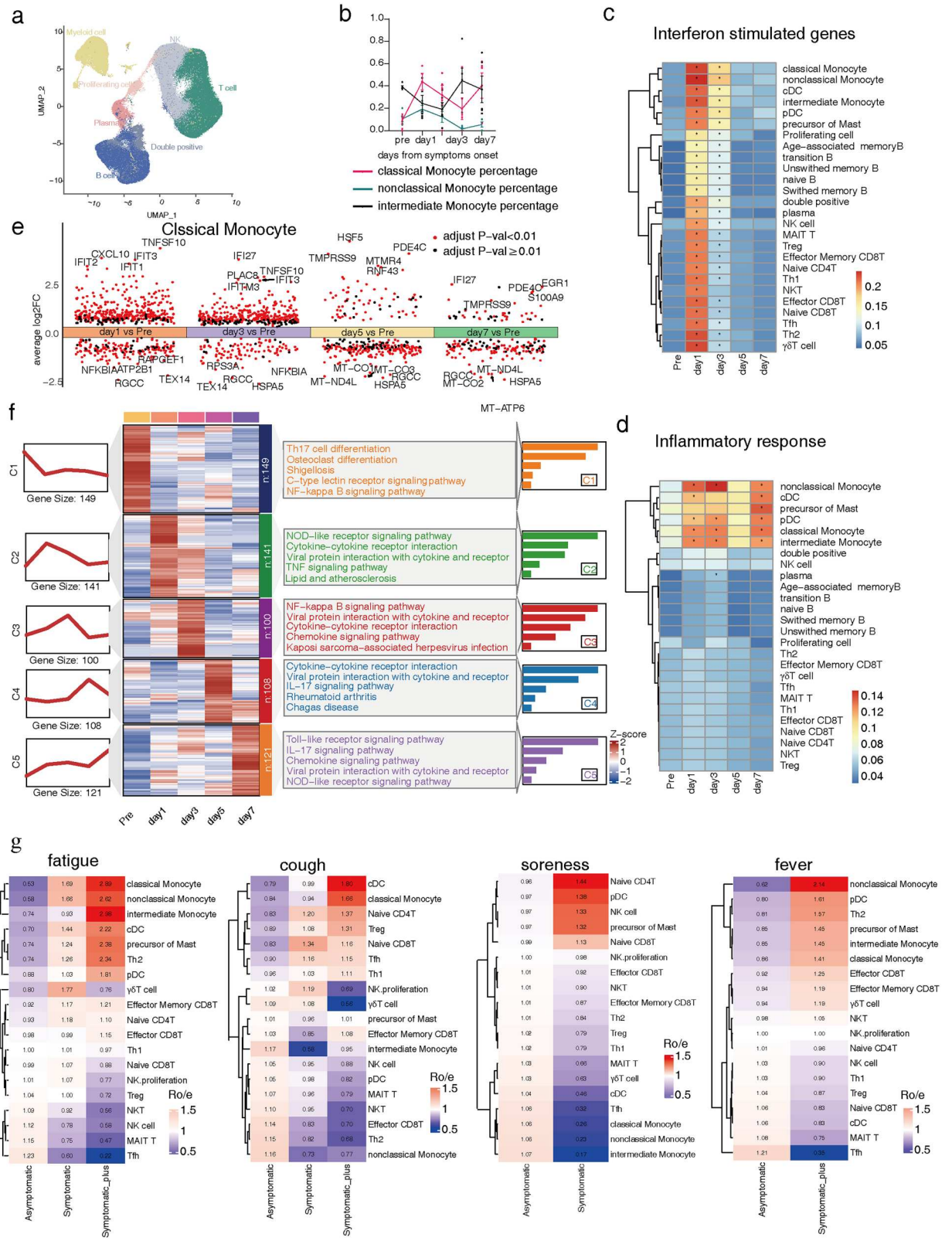

Figure 5. Myeloid cells shift from an antiviral response to alarmin response. (a) UMAP of the classification of myeloid cells and lymphocytes based on the cell marker in Fig. S4a. (b) The line chart of the monocytes sub-classification percentage of myeloid cells. (c) Scores of the interferon stimulated genes between different cell subsets over time. Samples exceeding the $\text{Threshold}_{\text{Activation}}$ marked with black dots, clearly distinguishing between "activated" and "returned to baseline or not activated" states. $\text{Threshold}_{\text{Activation}} = X_{\text{pre-symptom}} + 1 \times \sigma_{\text{pre-symptom}}$, $X_{\text{pre-symptom}}$ represents the mean expression score in the pre-symptom phase, $\sigma_{\text{pre-symptom}}$ represents the standard deviation. (d) Scores of the inflammatory response of GO term between different cell subsets at each time point. (e) The volcano plot shows the top five expressed genes of classical monocytes at each time point. (f) Heatmap showing the five dynamic expression patterns of genes belonging to the inflammatory response GO term in cMono across different days. The middle gray area shows the top five KEGG pathways that are enriched on the corresponding day. The vertical numbers represent the quantity of genes belonging to that pattern. The straight line in the left area represents the fitted curve for this cluster. The bar chart on the right represents the number of genes enriched in corresponding KEGG pathways. (g) Observed-to-Expected Ratio (Ro/e) analysis to quantify the association between immune cell subsets and clinical symptoms.Sample number = 25.

Domain-like Receptor (NOD) and TNF. On day 3, the NF-κB pathway was highly activated. These results are consistent with those observed in cytokine expression and bulk-sequencing analyses. The IL-17 and Toll-Like Receptor (TLR) pathways were activated on days 5 and 7 (Figure 5(f), S9, and S10). Subsequent analysis of the genes involved in IL-17 signaling revealed that multiple branches of the IL-17 pathway were activated on days 5–7, eventually releasing S100A8/A9 (Fig. S11a–c). Meanwhile, the elevation of plasma IL17A and IL17C levels both arise from day 3 (Figure 3, class 4), which is also consistent with the findings of single-cell studies. In addition, TLR4, which interacts with S100A8/A9 [42,43], was activated in myeloid cells (Fig. S12).

To investigate whether immune responses vary with symptom severity, we performed an Observed-to-Expected Ratio (Ro/e) analysis [44] to quantify the association between immune cell subsets and clinical symptoms(Figure 5(g)). This method evaluates the degree of association by comparing the observed proportion of a given immune cell subset in patients exhibiting a specific symptom to its expected proportion under a null distribution. The results revealed distinct immune correlates with different symptoms. Fatigue was most strongly associated with an increased proportion of monocytes, with a higher Ro/e value observed in patients experiencing severe fatigue. Cough was linked to elevated proportions of cDCs and classical monocytes, whereas soreness showed a strong correlation with naïve CD4 T cells. Fever was associated with an increased proportion of non-classical monocytes, pDCs, and Th2 cells, consistent with the role of pDCs in producing TNF-α and IFN-α, which are known to drive febrile responses. Notably, patients with a higher proportion of Tfh cells tended to exhibit milder or no symptoms, suggesting that enhanced B-cell help and early viral neutralization may mitigate excessive immune-inflammatory responses in breakthrough infections.

Overall, myeloid cells underwent the most dramatic changes, with a rapid rise in the proportion of classical monocytes after symptom onset and continued until day 7. All cell subpopulations exhibited transient antiviral and inflammatory responses followed by rapid recovery. However, the response of myeloid cells gradually transitioned to an alarmin response, characterized by the upregulation of S100A8/9 through the IL17 pathway. These results indicate that the increase in monocytes proposes a viral infection but is not significantly associated with symptoms.

Early detection of B-cell and T-cell expansion by BCR and TCR repertoire analyses

BCR analysis revealed that the expanded clones consisted predominantly of plasma cells (Fig. S13a–b). Further analysis of the BCR subtype demonstrated a rapid increase in IgG1 plasma cells over time. On day 7, IgG1 plasma cells comprised more than half of the total plasma cell population (Figure 6(a)). This rapid proliferation of IgG1 plasma cells indicates a rapid recall response. This finding aligns with the results presented in Figure 1(h). Subsequently, the temporal dynamics of biological processes were investigated within plasma cells. The results show that by day 1, ISGs were predominantly activated (response to type I IFN, $P < 0{\cdot}001$; type I IFN signaling pathway, $P < 0{\cdot}001$; cellular response to type I IFN, $P < 0{\cdot}001$), presenting antiviral responses (response to the virus, $P < 0{\cdot}001$; defense response to the virus, $P < 0{\cdot}001$) and type I IFN immune reactions. By day 3, antigen recognition signals became activated (antigen receptor-mediated signaling pathway, $P < 0{\cdot}001$). Day 5 was primarily characterized by enriched signaling for proliferation (covalent chromatin modification $P < 0{\cdot}001$), whereas day 7 mainly involved enhanced endoplasmic reticulum (ER) activity (protein targeting to ER, $P < 0{\cdot}001$), indicating antibody secretion (Figure 6(b)). The changes in this biological process are entirely consistent with previously described variations in the numbers of proliferating cells (Fig. S6), IgG1 plasma cell proportion (Figure 6(a)), and antibody titers (Figure 1(h)). Subsequently, *IGHV* utilized in the plasma were analyzed. Our analysis of B-cell V(D)J data revealed that hypermutated clones – specifically, IGHV3-23 (red) and IGHV4-59 (blue) – were already present in pre-symptom plasma samples with their mutation rates remaining stable during the first 7 days post-infection. This suggests that these clones predominantly derive from memory B cells induced by prior vaccination. On Day 1, IGHV3-23 exhibited the highest usage frequency, indicating its rapid mobilization upon initial viral exposure. However, as infection progressed and antigenic selection pressure from the Omicron variant increased, IGHV3-23 frequency declined while IGHV4-59 became increasingly dominant. Previous studies have associated IGHV3-23 with infections by wild-type and pre-Omicron variants, and IGHV4-59 with Omicron infections[45, 46], supporting the notion of a clonal replacement process: the initially mobilized IGHV3-23 clone, which offers a rapid response, is gradually supplanted by IGHV4-59, which may possess a higher affinity for mutated Omicron epitopes. Alternatively, IGHV4-59 might inherently require a higher antigen concentration or prolonged stimulation due to its initial lower affinity, indicating a delayed activation despite its pre-existing memory origin (Figure 6(c,d)) [47, 48].

The ratio analysis revealed an increase in the proportion of effector CD8 T cells on day 1, followed by a decrease. No obvious change in the proportion of Th1

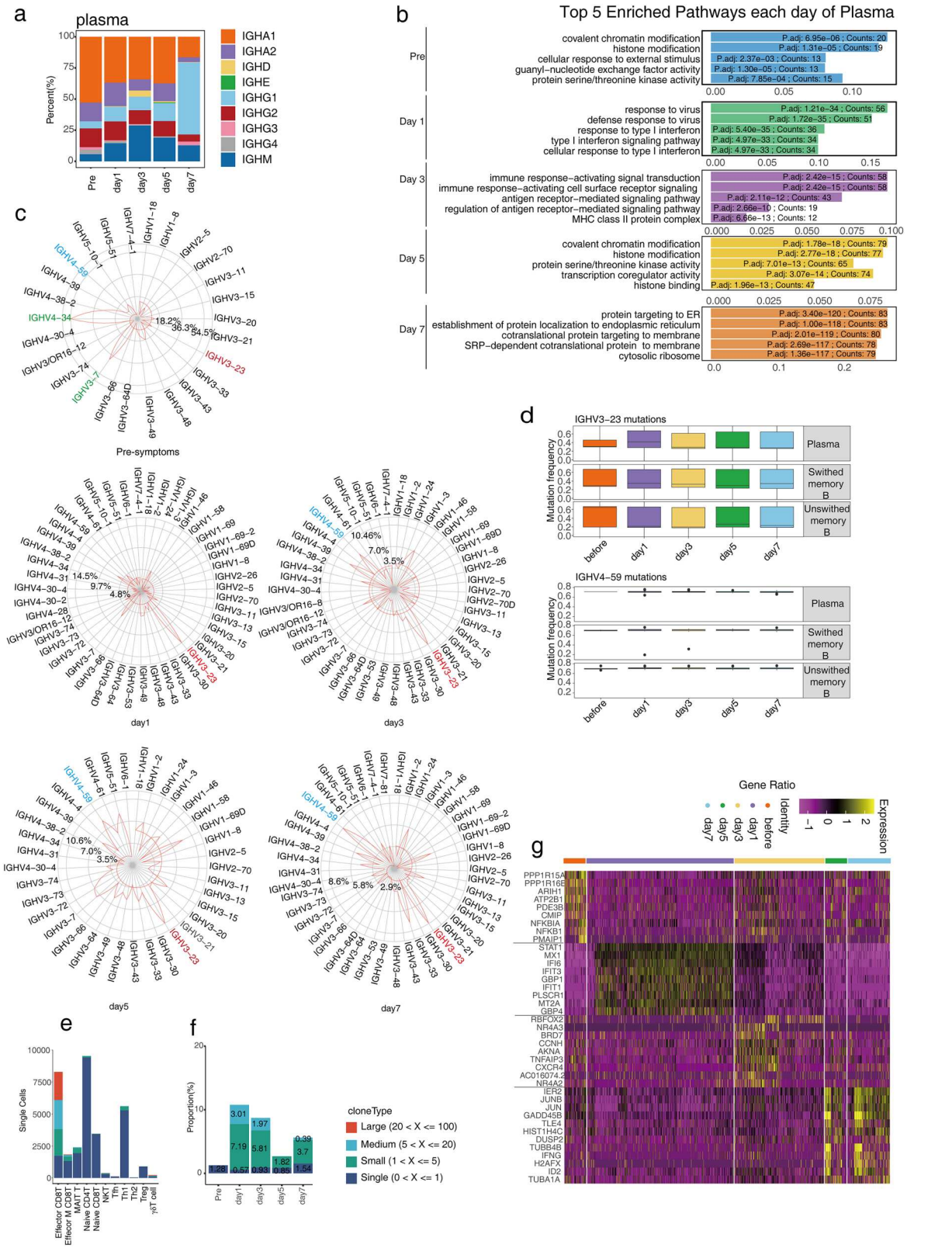

Figure 6. Early detection of B cell and T cell expansion by BCR and TCR repertoire analysis. (a) Proportion of different antibody subtypes in plasma cells over time as determined by BCR analysis. (b) Top five enriched Gene Ontology (GO) terms each day. Counts and *P*-adjust values were labelled in the bar corresponding to the term. The X-axis represents the proportion of all differentially expressed genes. (c) The polar plot shows the mean usage proportion of VH genes on the day of pre-symptoms to day7. (d) Mutation rate analysis of IGHV3-23 and IGHV4-59. (e) Clonotypes of TCRs between different T cell subsets. (f) Changes in cell proportion and clonotypes within clones that were singular prior to symptom onset and began to expand post-symptom emergence. (g) Highly expressed genes at various time points for the clones selected in Figure 5(f). Sample number = 25.

cells was noted during the first 7 days (Fig. S3f). Subsequent analysis focused on the sequencing results of the TCR and T-cell function. The results showed that effector CD8 T cells predominantly underwent clonal expansion. Simultaneously, a clonal expansion of a small number of effector memory CD8, MAIT, Th1,

and γδ T cells occurred (Figure 6(e), S3d, S13c, d). After selecting the single clone at the pre-symptomatic point for analysis, the clone expanded slowly and did not immediately transform into a large one from days 1 to 7 (Figure 6(f), S14a). Analysis of the αβ VJ pairs used by these clones revealed that they predominantly consisted of TRAV/DV4-TRAJ20-TRBV9-TRBJ2-3 (Fig. S14b). Notably, TRAV/DV4 was considered associated with COVID-19 [48]. Similarly, we analyzed the αβ VJ pairs that formed large clones before symptom onset and maintained this status consistently. The analysis showed that they primarily consisted of three pairing combinations, and no evidence links them to COVID-19 (Fig. S14c). Analysis of the enriched GO (Fig. S15a) and gene expression (Figure 6(g)) in clones that were single before symptom onset and then expanded revealed that by day 1, genes related to ISGs were predominantly activated (response to IFN-γ, $P < 0{\cdot}001$, Fig. S15a). By day 3, the activation of genes associated with antigen presentation was observed (antigen receptor-mediated signaling pathway, $P < 0{\cdot}001$, Fig. S15a). Simultaneously, high expression of cell cycle-related genes such as Cyclin H (CCNH*)* indicates that these cells are proliferating (Figure 6(g)). From days 5 to 7, the initiation of specific adaptive immune functions was evident, notably through IFN-γ expression (Figure 6(g)), which signifies a transition to Cytotoxic T Lymphocyte (CTL). Moreover, pathways linked to viral replication were activated at the pre-symptomatic phase ($P < 0{\cdot}001$). This signifies the actual phase of viral activation and replication, with several ribosomal proteins (RPLs) being hijacked for use at incubation periods (Fig. S15a). During days 5–7, pathways related to viral replication were reactivated ($P < 0{\cdot}001$), with an increased number of RPLs observed. This period may correspond to a prelude to a massive activation of effector T cells, resulting in protein synthesis (Fig. S15b).

Overall, clonal expansions of T and B cells were observed in the early stage. The particular expansion of IgG + plasma cells contributed to a rapid increase in IgG antibodies. Effector T cells did not exhibit large clonal expansions, further indicating that these patients were all in the early stages of infection.

Discussion

This study represents the detailed longitudinal investigation of the acute immune response to SARS-CoV-2 breakthrough infection in the real world. Symptom onset was designed as the starting point. Multi-omics technologies were employed to assess the baseline status of pre-symptoms. This study focused on examining various aspects of the immune responses in the first 7 days and analyzed their relationship with clinical symptoms. Thus, this real-world study mimics a breakthrough infection. Through this experimental design, the immune response changes during the transition from incubation infection to breakthrough infection were successfully captured. In the incubation periods (pre-symptoms), peripheral blood cytokine results indicate that this stage of infection induces a type 2 immune response. However, it is ineffective in suppressing the virus, leading to breakthrough infections. The onset of symptomatic breakthrough infection is accompanied by changes in peripheral blood cytokines (mainly ISGs, indicating an IFN response) and types of cells in PBMCs (with changes in their transcription states, transitioning to a type 1 immune response). A primary concern of this study is to ascertain the state before symptom manifestation. Given the heterogeneity within the population, some individuals exhibit significant immunological changes even before symptoms become apparent[49]. Thus, by utilizing transcriptomics to categorize overall gene expression and the blood cytokine dynamics, patient 4, despite being at the pre-symptom stage, was found to be already in a state comparable to that of other patients on day 1 of symptom onset. Moreover, the gene expressions of other pre-symptomatic samples cluster together, effectively illustrating their consistency. In addition, the comparison of the cytokine profiles in naïve infections revealed similarities between these two trends, demonstrating the feasibility of conducting immunological research starting from symptom onset.

Why do breakthrough infections occur? The primary reason is the decline in antibody levels. In this study, the participants demonstrated significantly lower antibody levels before disease onset, consistent with the antibody waning observed in other studies [50, 51]. Despite the well-known protective role of antibodies against symptomatic infections [52], the challenge lies in maintaining stable levels of protective antibodies. This phenomenon has been observed across various types of vaccines[51] and natural infection, and the specific causes remain unclear. According to current research, a notable aspect was that the severity of initial infection was significantly related [50, 53]. Therefore, the variability in individual responses to vaccines may be directly related to the duration of antibody maintenance and the speed at which memory immune responses are elicited, warranting further investigations. Certainly, variants also play a crucial role in breaking through infections, a finding confirmed both in epidemiology[54] and clinical trials[19]. In this study, variants are equally important. Moreover, the rapidness of the memory response plays a significant role in the manifestation of symptoms. In this study, antibody levels significantly increase by day 7, with a shift in CD8 T cells to CTL occurring between days 5 and 7. Thus, before the enhanced activation of the memory immune response, viral invasion was detected by innate immune cells, accompanied by the manifestation of

symptoms. By days 5 and 7, as adaptive immunity was increasingly activated and innate immunity began to wane, a marked decrease in symptoms was observed.

In this study, specific effector CD8 T cells did not form large clones within 7 days of symptom onset, nor was there a significant change in their numbers (Fig. S3f). These results indicate that in the recall immune response, the response of effector CD8 T cells is delayed and does not provide effective protection to prevent infection in the early stages. Numerous animal studies have confirmed that CD8 T cells play a crucial role in preventing severe disease. Therefore, the primary role of CD8 T cells is to prevent severe disease rather than to prevent infection and alleviate symptoms, consistent with the results of previous research[55, 56].

Intriguingly, a notable change, from an antiviral response of the intracellular ISGs to an alarmin response marked by the secretion of S100A8/A9, was observed in myeloid cells. Recent studies have associated the appearance of S100A8/A9/12 with severe COVID-19 progression[23, 24]. In addition, S100A8/A9 have been proposed to act as endogenous TLR4 agonist[42, 57]. However, our results present that the expression of those alarmins in myeloid cells does not directly contribute to systemic inflammatory response, as S100A8/A9 is released via the IL17 pathway[58] and myeloid expression of IL17RA were synchronously detected (Figure 3(a), Fig. S11a, b). Furthermore, S100A8/9 possesses multiple functions such as antibacterial activity[59–62]. A previous study showed that lower levels of S100A8/A9 in neonates were associated with a higher risk of bacterial infections after birth[25]. Building on our findings, we propose that the expression of S100A8/A9 primarily represents a natural response reserved for immune defense against potential secondary bacterial infections, a common complication following respiratory viral infections.

In this study, while the patterns of change in most cytokines were relatively similar between breakthrough and naïve infections, some intriguing differences were noted. Pro-inflammatory cytokines, including CXCL10, MCP-1, IL6, TNF, IL10, and IFN-γ, exhibited a higher proportion of bimodal expression in naïve infection, whereas in breakthrough infection, their expression levels were predominantly unimodal (Figure 4(a–d,h)). This phenomenon becomes clearly discernible only after depicting individual changes[21]. This presents that hosts facing an initial infection might not be able to directly control viral replication and may need to elevate the levels of inflammatory cytokines again to counteract virus spread. This could also indicate a stronger cellular damage in response to a first-time infection, leading to a subsequent rise in inflammatory signaling. Conversely, for breakthrough infection, immune memory and training of innate immunity may lead to a single peak response that controls the disease progression in most patients[63]. Meanwhile, CXCL10, CCL4, IL10, IL6, MCP-1, and IFN-γ have a slightly higher elevation than the first peak in breakthrough infection, which may also indicate the significant role of innate immune memory (Figure 4).

Our study holds significant practical implications and potential translational applications. In breakthrough infections, most proinflammatory cytokines exhibit a single, sharp peak rather than the multiple fluctuations observed in primary infections, suggesting that vaccination can induce a more efficient and rapid inflammatory response. This finding provides valuable insights for developing therapeutic strategies that modulate early inflammation to clear the virus and prevent cytokine storms. Furthermore, we found that even when the initial vaccine was targeted at a strain different from the current one, activated memory B cells can rapidly produce cross-reactive antibodies to effectively counter new variants, underscoring the importance of continued vaccination despite imperfect vaccine matching to emerging strains. Finally, the elevation of S100A8/9 may reflect a normal early immune response rather than disease progression, suggesting that it should not be used as a sole biomarker for severe disease. Recognizing these nuances can help optimize diagnostic strategies, guide the design of adjunct immunotherapies, and reinforce public health policies aimed at reducing breakthrough infections through high vaccination coverage.

This study has some limitations. First, this study has a small sample size. In addition, because of the rapid spread of infection, many participants became infected without sufficient time to obtain baseline samples, resulting in the availability of only post-infection data. Given the widespread expectation of infection, participants were particularly vigilant about symptom onset, which also greatly facilitated the completion of this study. We are also seeking to expand our cohort in future work. Second, the participants were 23–35 years old, all being young adults. Therefore, the generalizability of the findings to older and paediatric populations may be somewhat limited. Third, all participants experienced breakthrough infections; therefore, their immune responses represent those of patients with breakthrough infections. Fourth, variabilities exist between individuals with breakthrough infection from this study and those with naïve infection from the referenced 2024 *Sci Immunol* study. Although we controlled for certain demographic factors, these variables may still introduce bias into the comparative analysis.

In summary, we performed an in-depth dynamic analysis of the acute-phase immune response to SARS-CoV-2 breakthrough infection. This real-world SARS-CoV-2 challenge study provides a

reference for interpreting immunological basis for this virus infectious disease, which are important implications for vaccine development, immune response prediction, and understanding the association between symptoms and immunological response.

Acknowledgments

The authors thank Jie Sun for useful discussions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Elite Medical Professionals Project of China-Japan Friendship Hospital [grant number ZRJY2023-QM08], National Natural Science Foundation of China [grant numbers 82470007, 82241056, 82100009, 81991495], Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences [grant number 2021-I2M-1-048], National High Level Hospital Clinical Research Funding-2024-NHLHCRFLX-01-0101, National Key Research and Development Program of China [grant number 2021YFC2300501], Grant Beijing Research Ward Excellence Program [grant number BRWEP2024W114060103] and New Cornerstone Science Foundation.

Authorship contribution

Haibo Li: Conceptualization, Data curation, Formal analysis, Visualization, Methodology, Writing – original draft, Writing – review & editing. **Hongyu Liu:** Investigation, Formal analysis, Visualization, Writing – review & editing. **Hongping Wu**: Formal analysis, Visualization. **Chang Guo:** Investigation, Formal analysis.

Wenting Zuo: Investigation. **Ying Zheng:** Investigation. **Xiaoyan Deng:** Investigation.

Jiuyang Xu: Writing – review & editing. **Yeming Wang:** Writing – review & editing. **Zai Wang:** Writing – review & editing. **Binghuai Lu:** Writing – review & editing. **Baidong Hou:** Conceptualization, Supervision, Writing – review & editing. **Bin Cao:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Data availability statement

The gene expression data of single cell sequence, TCR/BCR sequence, Bulk RNA sequencing, Olink date and raw data from this study deposited at Genome Sequence Archive under SuperSeries accession numbers PRJCA022808(ngdc.cncb.ac.cn). Code for analysis is available at Github repository

(https://github.com/tsinghuavirgil/cjh_pccm_01_omicron_mild_omics).

Ethics statement

This study was approved by the Ethics Committee of China-Japan Friendship Hospital (2022-KY-058). Written informed consent was obtained from all study participants.

ORCID

Haibo Li http://orcid.org/0009-0004-3276-8425
Binghuai Lu http://orcid.org/0000-0003-3510-2747

References

[1] COVID-19 Vaccines Advice. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines/advice [accessed 2024 May 1].
[2] Lipsitch M, Krammer F, Regev-Yochay G, et al. SARS-CoV-2 breakthrough infections in vaccinated individuals: measurement, causes and impact. Nat Rev Immunol. 2022;22(1):57–65. doi:10.1038/s41577-021-00662-4
[3] Flannery B, Clippard J, Zimmerman RK, et al. Early estimates of seasonal influenza vaccine effectiveness – United States, January 2015. MMWR Morb Mortal Wkly Rep. 2015;64(1):10–15.
[4] Jacobs SE, Lamson DM, St George K, et al. Human rhinoviruses. Clin Microbiol Rev. 2013;26(1):135–162. doi:10.1128/CMR.00077-12
[5] Shah MM, Winn A, Dahl RM, et al. Seasonality of common human coronaviruses, United States, 2014-2021. Emerg Infect Dis. 2022;28(10):1970–1976. doi:10.3201/eid2810.220396
[6] Krammer F, Smith GJD, Fouchier RAM, et al. Influenza. Nat Rev Dis Primers. 2018;4(1):1–21. doi:10.1038/s41572-018-0002-y
[7] Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020;395(10229):1054–1062. doi:10.1016/S0140-6736(20)30566-3
[8] Li H, Ren L, Zhang L, et al. High anal swab viral load predisposes adverse clinical outcomes in severe COVID-19 patients. Emerg Microbes Infect. 2020;9(1):2706–2713. doi:10.1080/22221751.2020.1858700
[9] Cao Y, Yisimayi A, Jian F, et al. BA.2.12.1, BA.4 and BA.5 Escape Antibodies Elicited by Omicron Infection. Nature. 2022;608(7923):593–602. doi:10.1038/s41586-022-04980-y
[10] Kared H, Wolf A-S, Alirezaylavasani A, et al. Immune responses in omicron SARS-CoV-2 breakthrough infection in vaccinated adults. Nat Commun. 2022;13(1):4165. doi:10.1038/s41467-022-31888-y
[11] Kaku CI, Bergeron AJ, Ahlm C, et al. Recall of preexisting cross-reactive b cell memory after omicron BA.1 breakthrough infection. Sci Immunol. 2022;7 (73): eabq3511. doi:10.1126/sciimmunol.abq3511
[12] Lee WS, Tan H-X, Reynaldi A, et al. Durable reprogramming of neutralizing antibody responses following omicron breakthrough infection. Sci Adv. 2023;9(29):eadg5301. doi:10.1126/sciadv.adg5301
[13] Quandt J, Muik A, Salisch N, et al. Omicron BA.1 breakthrough infection drives cross-variant neutralization and memory B cell formation against conserved

epitopes. Sci Immunol. 2022;7(75):eabq2427. doi:10.1126/sciimmunol.abq2427
[14] Chen Y, Tong P, Whiteman N, et al. Immune recall improves antibody durability and breadth to SARS-CoV-2 variants. Sci Immunol. 2022;7 (78):eabp8328. doi:10.1126/sciimmunol.abp8328
[15] Goel RR, Painter MM, Lundgreen KA, et al. Efficient recall of omicron-reactive B cell memory after a third dose of SARS-CoV-2 mRNA vaccine. Cell. 2022;185(11):1875–1887.e8. doi:10.1016/j.cell.2022.04.009
[16] Koutsakos M, Lee WS, Reynaldi A, et al. The magnitude and timing of recalled immunity after breakthrough infection is shaped by SARS-CoV-2 variants. Immunity. 2022;55(7):1316–1326.e4. doi:10.1016/j.immuni.2022.05.018
[17] Wang CY, Hwang K-P, Kuo H-K, et al. A multitope SARS-CoV-2 vaccine provides long-lasting B cell and T cell immunity against delta and omicron variants. J Clin Invest. 2022;132(10):e157707. doi:10.1172/JCI157707
[18] Ibarrondo FJ, Hofmann C, Fulcher JA, et al. Primary, recall, and decay kinetics of SARS-CoV-2 vaccine antibody responses. ACS Nano. 2021;15(7):11180–11191. doi:10.1021/acsnano.1c03972
[19] Jackson S, Marshall JL, Mawer A, et al. Safety, tolerability, viral kinetics, and immune correlates of protection in healthy, seropositive UK adults inoculated with SARS-CoV-2: a single-centre, open-label, phase 1 controlled human infection study. The Lancet Microbe. 2024;5(0):655. doi:10.1016/S2666-5247(24)00025-9
[20] Killingley B, Mann AJ, Kalinova M, et al. Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. Nat Med. 2022;28(5):1031–1041. doi:10.1038/s41591-022-01780-9
[21] Wagstaffe HR, Thwaites RS, Reynaldi A, et al. Mucosal and systemic immune correlates of viral control after SARS-CoV-2 infection challenge in seronegative adults. Sci Immunol. 2024;9(92):eadj9285. doi:10.1126/sciimmunol.adj9285
[22] Koutsakos M, Reynaldi A, Lee WS, et al. SARS-CoV-2 breakthrough infection induces rapid memory and de novo T cell responses. Immunity. 2023;56(4):879–892.e4. doi:10.1016/j.immuni.2023.02.017
[23] Ahern DJ, Ai Z, Ainsworth M, et al. A Blood atlas of COVID-19 defines hallmarks of disease severity and specificity. Cell. 2022;185(5):916–938.e58. doi:10.1016/j.cell.2022.01.012
[24] Mellett L, Khader SA. S100A8/A9 in COVID-19 pathogenesis: impact on clinical outcomes. Cytokine Growth Factor Rev. 2022;63:90–97. doi:10.1016/j.cytogfr.2021.10.004
[25] Ulas T, Pirr S, Fehlhaber B, et al. S100-alarmin-induced innate immune programming protects newborn infants from sepsis. Nat Immunol. 2017;18(6):622–632. doi:10.1038/ni.3745
[26] Chen S, Zhou Y, Chen Y, et al. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–i890. doi:10.1093/bioinformatics/bty560
[27] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol. 2014;15(12):550. doi:10.1186/s13059-014-0550-8
[28] Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021;2(3):100141. doi:10.1016/j.xinn.2021.100141
[29] Kumar L, Futschik E, Mfuzz M. Mfuzz: A software package for soft clustering of microarray data. Bioinformation. 2007;2(1):5–7. doi:10.6026/97320630002005
[30] Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573–3587.e29. doi:10.1016/j.cell.2021.04.048
[31] Wolock SL, Lopez R, Klein AM. Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. Cell Syst. 2019;8(4):281–291.e9. doi:10.1016/j.cels.2018.11.005
[32] Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with harmony. Nat Methods. 2019;16(12):1289–1296. doi:10.1038/s41592-019-0619-0
[33] Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019;9(1):5233. doi:10.1038/s41598-019-41695-z
[34] Yu G, Wang L-G, Han Y, et al. clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–287. doi:10.1089/omi.2011.0118
[35] Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008;36(Database issue):D480–484. doi:10.1093/nar/gkm882
[36] Borcherding N, Bormann NL, Kraus G. scRepertoire: an R-based toolkit for single-cell immune receptor analysis. F1000Res. 2020;9:47. doi:10.12688/f1000research.22139.2
[37] Gupta NT, Vander Heiden JA, Uduman M, et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. Bioinformatics. 2015;31(20):3356–3358. doi:10.1093/bioinformatics/btv359
[38] Tougaard P, Ruiz Pérez M, Steels W, et al. Type 1 immunity enables neonatal thymic ILC1 production. Sci Adv. 2024;10(3):eadh5520. doi:10.1126/sciadv.adh5520
[39] Kopp EB, Agaronyan K, Licona-Limón I, et al. Modes of Type 2 immune response initiation. Immunity. 2023;56(4):687–694. doi:10.1016/j.immuni.2023.03.015
[40] Wynn TA. Type 2 cytokines: mechanisms and therapeutic strategies. Nat Rev Immunol. 2015;15(5):271–282. doi:10.1038/nri3831
[41] Schwartz DM, Bonelli M, Gadina M, et al. Type I/II cytokines, JAKs, and new strategies for treating autoimmune diseases. Nat Rev Rheumatol. 2016;12(1):25–36. doi:10.1038/nrrheum.2015.167
[42] Tsai S-Y, Segovia JA, Chang T-H, et al. DAMP Molecule S100A9 acts as a molecular pattern to enhance inflammation during Influenza A virus infection: role of DDX21-TRIF-TLR4-MyD88 pathway. PLoS Pathog. 2014;10(1):e1003848. doi:10.1371/journal.ppat.1003848
[43] Wang S, Song R, Wang Z, et al. S100A8/A9 in inflammation. Front Immunol. 2018;9:1298.
[44] Yang Y, Chen X, Pan J, et al. Pan-cancer single-cell dissection reveals phenotypically distinct B cell subtypes. Cell. 2024;187(17):4790–4811.e22. doi:10.1016/j.cell.2024.06.038
[45] Lima NS, Musayev M, Johnston TS, et al. Primary exposure to SARS-CoV-2 variants elicits convergent epitope specificities, immunoglobulin V gene usage and public B cell clones. Nat Commun. 2022;13(1):7733. doi:10.1038/s41467-022-35456-2

[46] Liu KJ, Zelazowska MA, McBride KM. The longitudinal analysis of convergent antibody VDJ regions in SARS-CoV-2-positive patients using RNA-Seq. Viruses. 2023;15(6):1253. doi:10.3390/v15061253
[47] He B, Liu S, Xu M, et al. Comparative global B cell receptor repertoire difference induced by SARS-CoV-2 infection or vaccination via single-cell V(D)J sequencing. Emerg Microbes Infect. 2022;11(1):2007–2020. doi:10.1080/22221751.2022.2105261
[48] Zhang F, Gan R, Zhen Z, et al. Adaptive immune responses to SARS-CoV-2 infection in severe versus mild individuals. Sig Transduct Target Ther. 2020;5(1):1–11. doi:10.1038/s41392-020-00263-y
[49] Xin H, Li Y, Wu P, et al. Estimating the latent period of coronavirus disease 2019 (COVID-19). Clin Infect Dis. 2022;74(9):1678–1681. doi:10.1093/cid/ciab746
[50] Vanshylla K, Di Cristanziano V, Kleipass F, et al. Kinetics and correlates of the neutralizing antibody response to SARS-CoV-2 infection in humans. Cell Host Microbe. 2021;29(6):917–929.e4. doi:10.1016/j.chom.2021.04.015
[51] Zhang Z, Mateus J, Coelho CH, et al. Humoral and cellular immune memory to four COVID-19 vaccines. Cell. 2022;185(14):2434–2451.e17. doi:10.1016/j.cell.2022.05.022
[52] Rick A-M, Laurens MB, Huang Y, et al. Risk of COVID-19 after natural infection or vaccination. EBioMedicine. 2023;96:104799. doi:10.1016/j.ebiom.2023.104799
[53] Tomic A, Skelly DT, Ogbe A, et al. Divergent trajectories of antiviral memory after SARS-CoV-2 infection. Nat Commun. 2022;13(1):1251. doi:10.1038/s41467-022-28898-1
[54] Evolutionary insight into the emergence of SARS-CoV-2 variants of concern. Nat Med. 2022;28(7):1357–1358. doi:10.1038/s41591-022-01892-2
[55] Melms JC, Biermann J, Huang H, et al. A molecular single-cell lung atlas of lethal COVID-19. Nature. 2021;595(7865):114–119. doi:10.1038/s41586-021-03569-1
[56] Sette A, Sidney J, Crotty S. T cell responses to SARS-CoV-2. Annu Rev Immunol. 2023;41:343–373. doi:10.1146/annurev-immunol-101721-061120
[57] Ma L, Sun P, Zhang J-C, et al. Proinflammatory effects of S100A8/A9 via TLR4 and RAGE signaling pathways in BV-2 microglial cells. Int J Mol Med. 2017;40(1):31–38. doi:10.3892/ijmm.2017.2987
[58] Christmann C, Zenker S, Martens L, et al. Interleukin 17 promotes expression of alarmins S100A8 and S100A9 during the inflammatory response of keratinocytes. Front Immunol. 2020;11:599947. doi:10.3389/fimmu.2020.599947
[59] Raquil M-A, Anceriz N, Rouleau P, et al. Blockade of antimicrobial proteins S100A8 and S100A9 inhibits phagocyte migration to the alveoli in streptococcal pneumonia. J Immunol. 2008;180(5):3366–3374. doi:10.4049/jimmunol.180.5.3366
[60] Nishikawa Y, Kajiura Y, Lew JH, et al. Calprotectin induces IL-6 and MCP-1 production via toll-like receptor 4 signaling in human gingival fibroblasts. J Cell Physiol. 2017;232(7):1862–1871. doi:10.1002/jcp.25724
[61] Hiroshima Y, Hsu K, Tedla N, et al. S100A8 induces IL-10 and protects against acute lung injury. J Immunol. 2014;192(6):2800–2811. doi:10.4049/jimmunol.1302556
[62] Corbin BD, Seeley EH, Raab A, et al. Metal chelation and inhibition of bacterial growth in tissue abscesses. Science. 2008;319(5865):962–965. doi:10.1126/science.1152449
[63] Fanucchi S, Domínguez-Andrés J, Joosten LAB, et al. The intersection of epigenetics and metabolism in trained immunity. Immunity. 2021;54(1):32–43. doi:10.1016/j.immuni.2020.10.011