

# A distinct monocyte transcriptional state links systemic immune dysregulation to pulmonary impairment in long COVID

Received: 17 March 2025

Accepted: 3 December 2025

Published online: 14 January 2026

 Check for updates

Saumya Kumar<sup>1,2</sup>, Chaofan Li<sup>3,4</sup>, Liang Zhou<sup>1,2</sup>, Qiuyao Zhan<sup>1,2</sup>, Ahmed Alaswad<sup>1,2</sup>, Sonja Volland<sup>5</sup>, Bibiana Costa<sup>1,2</sup>, Simon Alexander Krooss<sup>1,2,6,7,8,9</sup>, Isabel Klefenz<sup>10</sup>, Hagen Schmaus<sup>1,2,11</sup>, Antonia Zeuzem<sup>1,2,9,11</sup>, Dorothee von Witzendorff<sup>1,2,11,12</sup>, Helena Lickel<sup>1,2,9,11,12</sup>, Lea Pueschel<sup>9</sup>, Anke R. M. Kraft<sup>1,2,9,11,12</sup>, Markus Cornberg<sup>1,2,9,11,12</sup>, Andreas Rembert Koczulla<sup>7,13,14,15</sup>, Isabell Pink<sup>16,17</sup>, Marius M. Hoeper<sup>16,17</sup>, Cheng-Jian Xu<sup>1,2,18</sup>, Susanne Häussler<sup>12,19,20,21</sup>, Miriam Wiestler<sup>9</sup>, Mihai G. Netea<sup>18,22</sup>, Thomas Illig<sup>5,17,23,24</sup>, Jie Sun<sup>3,4,24</sup> & Yang Li<sup>1,2,12,17,18,23,24</sup> ✉

The mechanisms driving immune dysregulation in long COVID disease remain elusive. Here we integrated single-cell multiome data, immunological profiling and functional assays to investigate immune alterations across multiple cohorts. A transcriptional state in circulating monocytes (LC-Mo) was enriched in individuals with mild–moderate acute infection and accompanied by persistent elevations of plasma CCL2, CXCL11 and TNF. LC-Mo showed TGF $\beta$  and WNT– $\beta$ -catenin signaling and correlated with fatigue severity. Protein markers of LC-Mo were increased in individuals with pronounced fatigue or dyspnea, and those with severe respiratory symptoms showed higher LC-Mo expression. Epigenetically, LC-Mo exhibited AP-1- and NF- $\kappa$ B1-driven profibrotic programs. LC-Mo-like macrophages in bronchoalveolar lavage samples from individuals with severe respiratory symptoms displayed a profibrotic profile, and individuals with a high LC-Mo transcriptional state showed impaired interferon responses after stimulation. Collectively, our findings define a pathogenic monocyte transcriptional state linking systemic immune dysfunction to persistent long COVID disease, providing mechanistic insights and potential therapeutic targets.

Long COVID affects 10–20% of individuals after severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, with symptoms ranging from mild discomfort to severe, long-lasting impairments such as fatigue, respiratory issues and neurological problems. These symptoms can persist for over 3 years (refs. 1–5), representing a substantial health burden and prompting efforts to better characterize long COVID (LC), including biomarker discovery for improved diagnosis<sup>6–10</sup>.

LC presents with diverse symptoms reflecting multiorgan system abnormalities<sup>11–13</sup>. The evidence suggests multiple possible

causes, including persistence of viral remnants or reactivation of latent viruses<sup>7,14–17</sup>. Yet, persistent immune dysregulation is a consistent finding in LC studies<sup>10,11,14,16–19</sup>. Although most LC cases follow mild-to-moderate acute illness, many studies do not stratify individuals by acute infection (AI) severity<sup>6–8</sup>, which is crucial because severe cases, especially those treated in the intensive care unit, develop immune changes due to intensive medical interventions<sup>20,21</sup>. Failing to account for these differences may confound LC-associated molecular signatures, highlighting the importance of refined patient grouping.

A full list of affiliations appears at the end of the paper. ✉ e-mail: [yang.li@helmholtz-hzi.de](mailto:yang.li@helmholtz-hzi.de)

To address this gap, we stratified individuals with LC by acute COVID-19 severity to better resolve immune heterogeneity and identify molecular features underlying chronic symptoms. We applied single-cell multiomics profiling of peripheral blood mononuclear cells (PBMCs) and measured plasma cytokines from individuals with LC with fatigue and respiratory symptoms using longitudinal and cross-sectional samples. We identified a distinct circulating CD14<sup>+</sup> monocyte state associated with LC ('LC-Mo'), which was enriched in individuals with mild-to-moderate AI. This state coincided with persistent elevation of circulating cytokines, indicating systemic inflammation. In two independent cohorts of individuals with LC with severe respiratory symptoms and abnormal lung function, LC-Mo expression was increased in circulating CD14<sup>+</sup> monocyte subsets. In bronchoalveolar lavage (BAL) myeloid cells from individuals with severe respiratory symptoms, LC-Mo-like macrophages showed a profibrotic gene expression profile. Functionally, CD14<sup>+</sup> monocytes from individuals with LC-Mo enrichment showed dysregulated responses to ex vivo stimulation, indicating impaired immune regulation. Together, these findings provide systemic insight into the cellular and molecular basis of LC and highlight potential therapeutic targets.

## Results

### LC has a distinct transcriptome after mild or moderate disease

Individuals presenting with headache, dyspnea or fatigue to the pneumology outpatient clinic at Hannover Medical School (MHH) were recruited according to the German S1 guidelines<sup>22</sup> and the Delphi Consensus Criteria<sup>21</sup> for LC (4–12 weeks) and post-COVID-19 syndrome (>12 weeks). These criteria included symptoms persisting beyond the acute phase of SARS-CoV-2 infection or its treatment, new symptoms emerging after recovery and attributed to prior infection or worsening of pre-existing conditions. Because heterogeneity in LC molecular profiles may be shaped by acute disease severity and treatment, we stratified individuals with acute SARS-CoV-2 infection (AI) and LC into those with mild-to-moderate (WHO score of 1–5) AI (AI<sup>M</sup> and LC<sup>AM</sup>) and those with severe (WHO scores 6–9) AI (AI<sup>S</sup> and LC<sup>AS</sup>).

Cohort 1 included 45 individuals recruited between April 2020 and August 2021 at MHH, of which 9 gave longitudinal samples and 36 gave cross-sectional samples ( $n = 78$  total samples), including 11 donors with AI categorized as AI<sup>M</sup> ( $n = 7$  donors, 42.8% women, median age = 52, range 23–66 years of age, WHO score range 1–5) and AI<sup>S</sup> ( $n = 4$  donors, 50% women, median age = 37, range 32–54, WHO score range 6–9), 37 donors with LC categorized as LC<sup>AM</sup> ( $n = 29$  donors, 8 longitudinal donors with two to three time points and 21 single-time-point donors, 58% women, median age = 49, range 31–84 years) and LC<sup>AS</sup> ( $n = 8$  donors, 3 with two to four time points, 5 single-time-point donors, 25% women, median age = 46, range 19–75) and 8 donors who had recovered after 4–8 months of LC (R<sup>C</sup>; 1 longitudinal donor with two time points and 7 single-time-point donors, 37.5% women, median age = 38, range 19–65), in addition to 6 prepandemic noninfected

control individuals (NI; 50% women, median age = 40, range 24–61). LC and R<sup>C</sup> samples were collected 1.7–10.2 months after infection. Cohort 2 included 117 LC<sup>AM</sup> donors (24 donors with two to four time points, 93 single-time-point donors, 58.9% women, median age = 48, range 19–83) and 25 LC<sup>AS</sup> donors (12 longitudinal donors, 13 single-time-point donors, 20% women, median age = 53, range 18–81), recruited between May 2020 and August 2021 at MHH, along with 33 prepandemic NI samples (48.4% women, median age = 40, range 25–65). Cohort 3 included only LC<sup>AM</sup> donors ( $n = 8$  donors, 62.5% women, median age = 45, range 21–63), all with respiratory postacute sequelae of SARS-CoV-2 infection (PASC) recruited between October and November 2023 at the Pulmonary Rehabilitation Clinic in Schönau am Königssee, Germany. Cohort 4 included LC<sup>AM</sup> donors ( $n = 29$ , 58.6% women, median age = 49, range 33–72), LC<sup>AS</sup> donors ( $n = 11$  donors, 18% female, median age = 57, range 35–81), 8 donors recovered from AI (R<sup>A</sup>) and 2 NI donors (60% women, median age = 41, range 29–67) recruited between August 2020 and June 2022 at MHH. Cohort 5 included LC donors ( $n = 9$  donors, 44.4% women, median age = 64, range 62–83, including 5 with respiratory PASC) and NI donors ( $n = 2$  donors, 50% women, median age = 77, range 73–77), recruited between October 2020 and November 2021 at Mayo Clinic, a previously published study<sup>23</sup> (Fig. 1a and Methods).

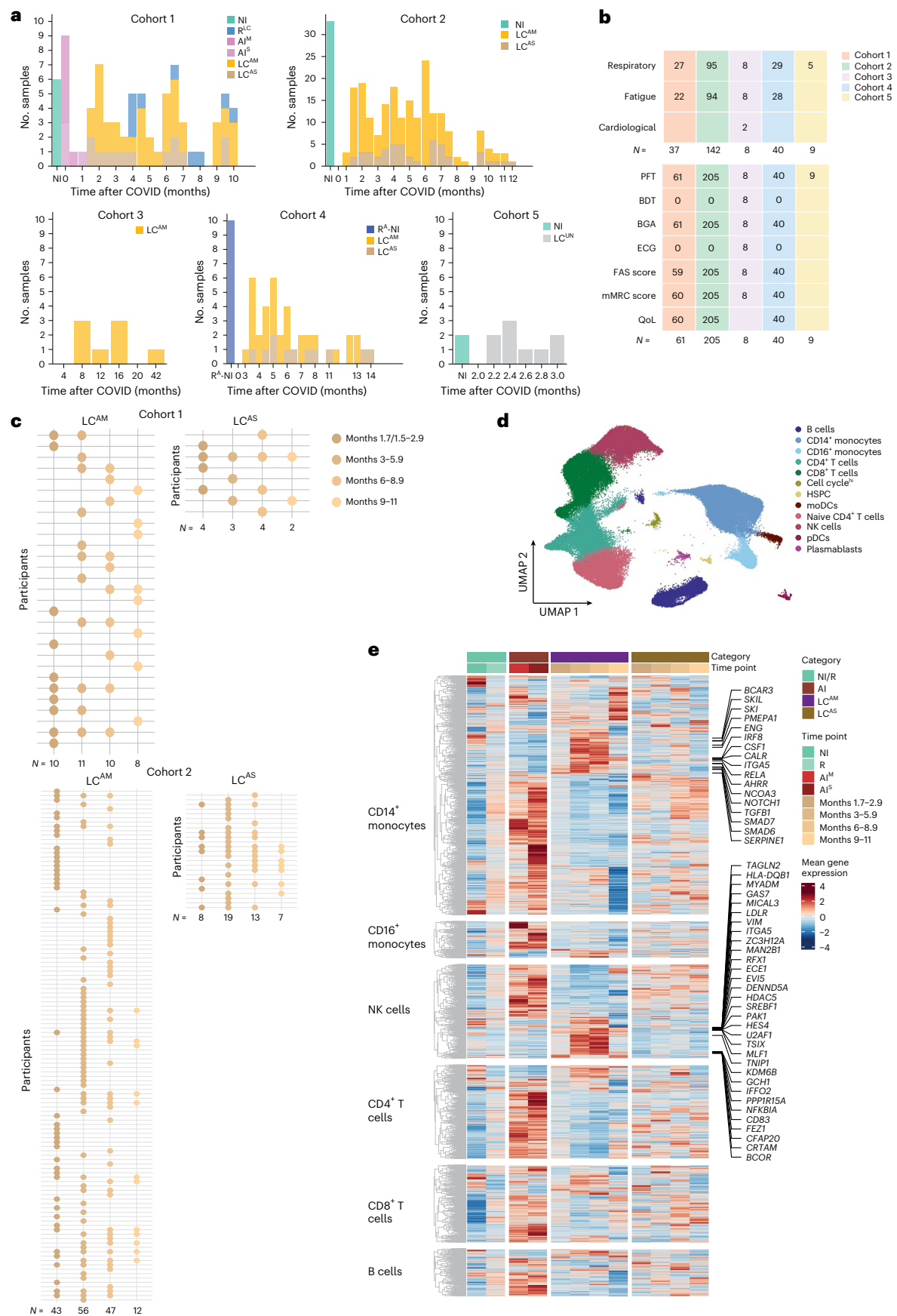
Clinical assessment included blood gas analysis, pulmonary function tests and standardized participant-reported outcome measures: the fatigue assessment scale (FAS), validated in chronic fatigue<sup>24–26</sup> and LC, and the modified medical Research Council (mMRC) dyspnea scale (0–4, where 0 indicates no breathlessness, 1 indicates breathlessness on exertion, 2 indicates breathlessness when hurrying or walking uphill, 3 indicates stopping for breath after ~100 m or a few minutes, and 4 indicates too breathless to leave the house or when dressing), along with quality-of-life metrics<sup>27</sup>. All clinical assessment data were systematically collected at each participant visit for cohorts 1–4 (Fig. 1b and Supplementary Tables 1–5).

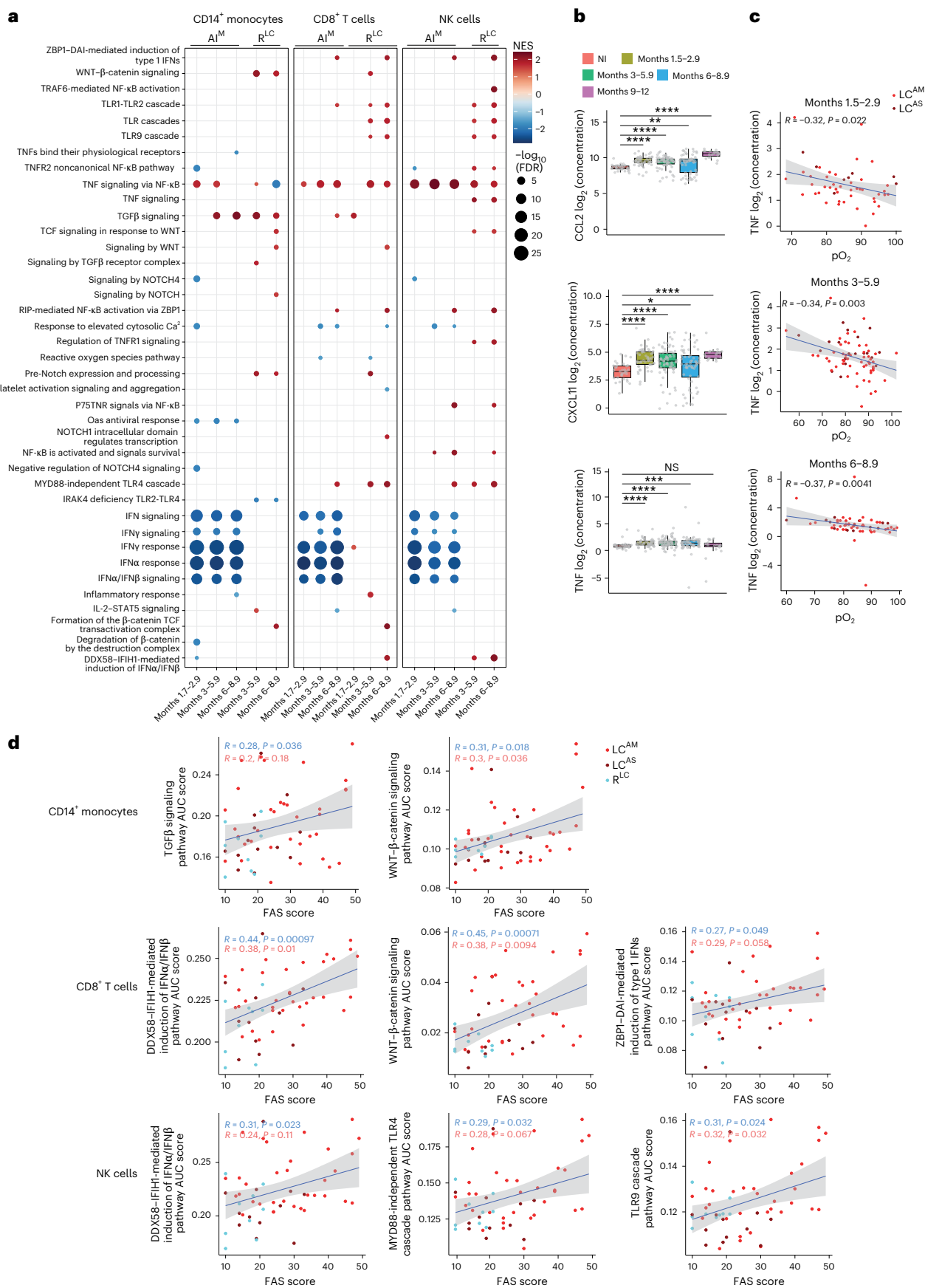
To study molecular signatures of disease progression, we stratified samples in cohorts 1 and 2 by months since AI (months 1.5/1.7–2.9, 3–5.9, 6–8.9 and 9–11; Fig. 1c and Methods). For cohort 1, we generated single-nucleus RNA-sequencing (snRNA-seq) and single-nucleus assay for transposase-accessible chromatin with sequencing (snATAC-seq) data from 78 PBMC samples from NI, R<sup>C</sup>, AI<sup>M</sup>, AI<sup>S</sup>, LC<sup>AM</sup> and LC<sup>AS</sup> donors across all time points. In cohort 2 we measured the concentrations of 14 cytokines in plasma samples from LC<sup>AM</sup> or LC<sup>AS</sup> and NI donors across all time points. Validation was performed using single-cell RNA-sequencing (scRNA-seq; cohort 3), flow cytometry (cohort 4) and a published PBMC/BAL single-cell dataset<sup>23</sup> (cohort 5). All samples, except those from participants with AI, were PCR negative at collection. We used an integrative multistep analysis to identify cell-type-specific immune dysregulation and link and assess relevance in LC (Extended Data Fig. 1a).

Analysis of single-cell data from cohort 1 PBMCs yielded ~118,000 high-quality cells (Fig. 1d). snRNA-seq data showed distinct patterns

**Fig. 1 | Transcriptomes of circulating immune cells show heterogeneity in individuals with LC. a**, Schematic showing the distribution of samples across cohort 1, which included longitudinal and cross-sectional PBMC samples ( $n = 78$ ) from NI donors ( $n = 6$ ) and donors with AI ( $n = 11$ ), LC<sup>AM</sup> (WHO 1–5;  $n = 39$ ), LC<sup>AS</sup> (WHO 6–9;  $n = 13$ ) and R<sup>C</sup> ( $n = 9$ ), collected 1.7–10.2 months after infection; cohort 2, which included longitudinal and cross-sectional samples ( $n = 238$ ) from NI donors ( $n = 33$ ) and donors with LC<sup>AM</sup> ( $n = 158$ ) and LC<sup>AS</sup> ( $n = 47$ ) collected at 1.5–11 months after infection; cohort 3, which included PBMCs from LC<sup>AM</sup> donors ( $n = 8$ ) collected 8–42 months after infection; cohort 4, which included PBMC samples ( $n = 40$ ) from R<sup>A</sup> donors and NI donors ( $n = 10$ ) and donors with LC<sup>AM</sup> ( $n = 29$ ) and LC<sup>AS</sup> ( $n = 11$ ) collected 3–14 months after infection; and cohort 5, which consisted of PBMC ( $n = 11$ ) and BAL ( $n = 9$ ) samples from individuals with LC with unknown acute-phase severity (LC<sup>UN</sup>;  $n = 9$ ) and NI donors ( $n = 2$ ; GEO: GSE263817). **b**, Number of individuals with LC in cohorts 1–5 exceeding thresholds for fatigue (FAS > 21), respiratory symptoms (dyspnea > 0)

or cardiology symptoms (top) and number of samples with pulmonary function tests (PFT), bronchial dilation tests (BDT), blood gas analysis (BGA), electrocardiogram (ECG), FAS and mMRC scores and quality-of-life (QoL) assessments (bottom). Empty boxes denote missing data. **c**, Distribution of LC samples across months 1.7–2.9 (LC<sup>AM</sup>,  $n = 10$ ; LC<sup>AS</sup>,  $n = 4$ ), 3–5.9 (LC<sup>AM</sup>,  $n = 11$ ; LC<sup>AS</sup>,  $n = 3$ ), 6–8.9 (LC<sup>AM</sup>,  $n = 10$ ; LC<sup>AS</sup>,  $n = 4$ ) and 9–11 (LC<sup>AM</sup>,  $n = 8$ ; LC<sup>AS</sup>,  $n = 2$ ) in cohort 1 (top) and months 1.5–2.9 (LC<sup>AM</sup>,  $n = 43$ ; LC<sup>AS</sup>,  $n = 8$ ), 3–5.9 (LC<sup>AM</sup>,  $n = 56$ ; LC<sup>AS</sup>,  $n = 19$ ), 6–8.9 (LC<sup>AM</sup>,  $n = 47$ ; LC<sup>AS</sup>,  $n = 13$ ) and 9–11 (LC<sup>AM</sup>,  $n = 12$ ; LC<sup>AS</sup>,  $n = 7$ ) in cohort 2 (bottom). **d**, UMAP of snRNA-seq data from 78 PBMC samples from all donors and all time points in cohort 1, as in **a** and **c**. **e**, Expression of genes significant by two-sided Wilcoxon test (Benjamini–Hochberg method-adjusted  $P$  value of <0.05 and log<sub>2</sub> (fold change) > 0.8) in CD14<sup>+</sup> monocytes, CD16<sup>+</sup> monocytes, CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, B cells and NK cells, with genes consistently upregulated across labeled LC<sup>AM</sup> time points; HSPC, hematopoietic stem and progenitor cells; moDCs, monocyte-derived dendritic cells; pDCs, plasmacytoid dendritic cells.







**Fig. 2 | TNF and inflammatory pathways in circulating immune cells indicate systemic inflammation in LC.** **a**, GSEA in CD14<sup>+</sup> monocytes, CD8<sup>+</sup> T cells and NK cells from LC<sup>AM</sup> samples compared to AI and R<sup>LC</sup> samples as in Fig. 1e. Pathways are plotted with an adjusted  $P$  of  $<0.1$  (Kolmogorov–Smirnov-based test with permutation-derived  $P$  values, adjusted using the Benjamini–Hochberg method); NES, normalized enrichment score. **b**, Expression of CCL2, CXCL11 and TNF in the plasma of NI control individuals ( $n = 33$ ) and individuals with LC at months 1.5–2.9 ( $n = 51$ ), 3–5.9 ( $n = 75$ ), 6–8.9 ( $n = 60$ ) and 9–12 ( $n = 19$ ) from cohort 2. Data were analyzed by two-sided Wilcoxon rank-sum test; \*\*\*\* $P < 0.00001$ , \*\*\* $P < 0.001$ , \*\* $P < 0.01$  and \* $P < 0.05$ . The box plots show the median (center), first and third quartiles (bounds) and 1.5 times the interquartile range (whiskers). **c**, Correlation between the amount of TNF in the plasma and pO<sub>2</sub> levels in blood in LC donors at

months 1.5–2.9 ( $n = 51$ ), 3–5.9 ( $n = 75$ ), 6–8.9 ( $n = 60$ ) and 9–12 ( $n = 19$ ). Data were analyzed by Spearman correlation.  $P$  values were determined using the exact/permutation-based test. The gray shaded area indicates the 95% confidence interval. **d**, Correlation between AUC score and TGF $\beta$  and WNT– $\beta$ -catenin signaling pathways in CD14<sup>+</sup> monocytes, IFN $\alpha$ /IFN $\beta$  and TLR4 and TLR9 cascade signaling in NK cells and IFN $\alpha$ /IFN $\beta$  and WNT– $\beta$ -catenin signaling in CD8<sup>+</sup> T cells with a FAS score; red, statistics calculated using only LC samples (LC<sup>AM</sup> and LC<sup>AS</sup>); blue, statistics calculated using samples from individuals with LC<sup>AM</sup> ( $n = 39$ ), LC<sup>AS</sup> ( $n = 13$ ) and R<sup>LC</sup> ( $n = 9$ ). Data were analyzed by Spearman correlation.  $P$  values were determined using the exact/permutation-based test. The gray shaded area indicates the 95% confidence interval; NS, not significant.

in LC<sup>AM</sup> and LC<sup>AS</sup> compared to R<sup>LC</sup> and AI across major PBMCs (Fig. 1e). LC<sup>AM</sup> showed downregulated AI genes by months 6–8.9, whereas LC<sup>AS</sup> retained an acute COVID-19-like transcriptomic profile, indicating heterogeneity based on AI history (Fig. 1e). Differential gene expression (DGE) analysis identified 1,737 upregulated genes in CD14<sup>+</sup> monocytes from LC<sup>AM</sup> donors compared to those from AI and R<sup>LC</sup> (Fig. 1e), with upregulation over 1.7–8.9 months, and showed participant-specific heterogeneity (Fig. 1e and Extended Data Fig. 1b). LC<sup>AM</sup> CD14<sup>+</sup> monocytes showed persistent upregulation of proinflammatory (*CSF1*, *IRF8*, *RELA* and *NOTCH1*) and anti-inflammatory (*TGFB1*, *SMADs*, *ENG* and *SERPINE1*) markers (Extended Data Fig. 1c) at all time points, whereas other signature genes showed increased expression from 3 to 8.9 months (Fig. 1e). This signature diminished during months 9–11, possibly due to lower cell numbers (Extended Data Fig. 1d), but showed upregulation of a subset of acute-phase genes, including *IL1B*, *S100A4*, *PDIA3* and *MTRNR2L1*. LC<sup>AM</sup> natural killer (NK) cells also showed distinct increased expression of *SREBF1*, *TAGLN2*, *TNIP1*, *NFKBIA* and *CD83* among others compared to R<sup>LC</sup> and AI NK cells (Fig. 1e). Collectively, transcriptional profiles in individuals with LC reflected differences based on AI severity, with notable molecular changes in LC<sup>AM</sup> monocytes and NK cells, whereas LC<sup>AS</sup> displayed persistent but milder expression of acute-phase genes.

### TNF and TNF signaling genes are upregulated in LC<sup>AM</sup>

We next performed gene set enrichment analysis (GSEA) using pseudo-bulk counts for each cell subset in LC<sup>AM</sup> or LC<sup>AS</sup> samples across all time points, comparing them to the AI and R<sup>LC</sup> cell samples. LC<sup>AM</sup> showed persistent upregulation of the TNF signaling pathway and persistent downregulation of interferon (IFN) signaling and response pathways across all major cell subsets (CD4<sup>+</sup> and CD8<sup>+</sup> T cells, B cells and CD14<sup>+</sup> and CD16<sup>+</sup> monocytes) compared to AI, up to month 8.9 (Fig. 2a and Extended Data Fig. 2a). CD8<sup>+</sup> T cells and NK cells from LC<sup>AM</sup> samples exhibited increased activation of the ‘TLR signaling cascades’ pathway

relative to R<sup>LC</sup> samples at months 3–8.9 (Fig. 2a). In LC<sup>AM</sup> CD14<sup>+</sup> monocytes, the TNF signaling pathway was transiently upregulated at months 1.7–5.9 and downregulated at months 6–8.9, whereas pathways including transforming growth factor- $\beta$  (TGF $\beta$ ), WNT– $\beta$ -catenin and Notch signaling were upregulated at months 3–8.9 compared to in AI and R<sup>LC</sup> CD14<sup>+</sup> monocytes (Fig. 2a). In LC<sup>AS</sup>, the TNF signaling pathway was sparsely activated in CD14<sup>+</sup> monocytes and CD8<sup>+</sup> T cells up to 5.9 months (Extended Data Fig. 2b). LC<sup>AS</sup> CD14<sup>+</sup> monocytes upregulated PD-1 signaling and MHC class II antigen presentation pathways compared to AI, but not R<sup>LC</sup> (Extended Data Fig. 2b, top). CD8<sup>+</sup> and CD4<sup>+</sup> T cells and NK cells from LC<sup>AS</sup> samples displayed increased activation of IFN response pathways compared to CD8<sup>+</sup> and CD4<sup>+</sup> T cells and NK cells from R<sup>LC</sup> samples (Extended Data Fig. 2b).

We also profiled 14 proinflammatory cytokines in cohort 2 plasma using a multiplex bead-based assay (Extended Data Fig. 2c), excluding interleukin-4 (IL-4) and IL-5 due to low detection. CXCL11, CCL2 and TNF were persistently elevated in individuals with LC compared to in NI donors up to month 9 (Fig. 2b). TNF mRNA was also persistently upregulated in individuals with LC<sup>AM</sup> across most immune cell types and time points (Extended Data Fig. 2d). TNF protein exhibited a statistically significant negative correlation with arterial oxygenation (pO<sub>2</sub>) in individuals with LC (Fig. 2c), which remained statistically significant in LC<sup>AM</sup>, but not in LC<sup>AS</sup>, up to month 8.9 (Extended Data Fig. 3a). No other cytokines showed consistent correlations across all time points (Extended Data Fig. 3b, c).

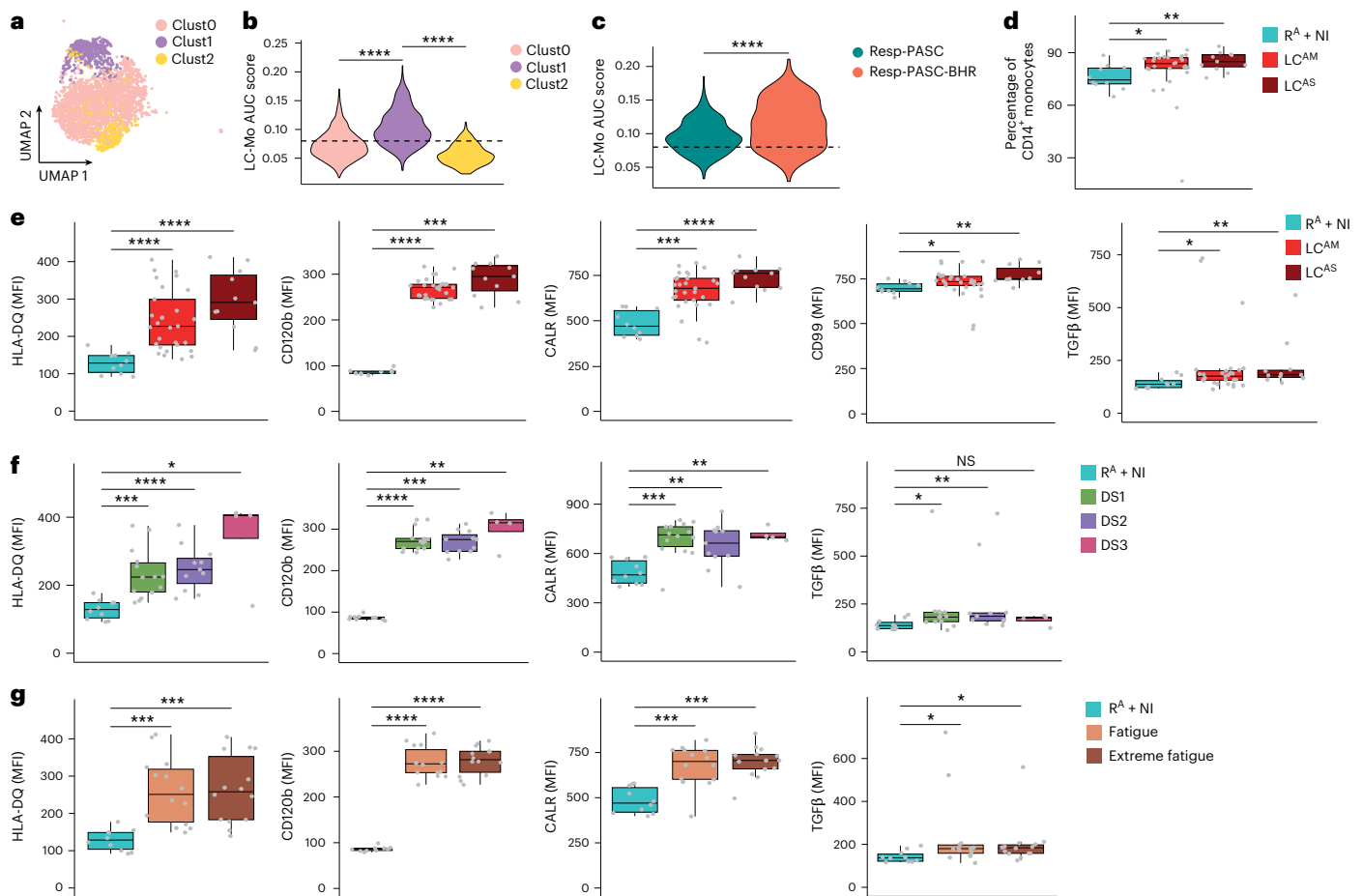
Correlation analysis between key pathways upregulated in CD8<sup>+</sup> T cells, NK cells and CD14<sup>+</sup> monocytes and FAS scores indicated that TGF $\beta$  and WNT– $\beta$ -catenin signaling in CD14<sup>+</sup> monocytes showed modest positive correlations with FAS scores in LC alone and stronger correlations when LC and R<sup>LC</sup> were combined (Fig. 2d). IFN $\alpha$ /IFN $\beta$  induction pathways positively correlated with FAS scores in CD8<sup>+</sup> T cells and NK cells in both LC only or LC + R<sup>LC</sup> combined analyses (Fig. 2d). WNT– $\beta$ -catenin signaling in CD8<sup>+</sup> T cells and Toll-like receptor (TLR)

### Fig. 3 | Distinct cell subclusters drive LC signatures in NK cells, CD8<sup>+</sup> T cells and CD14<sup>+</sup> monocytes.

**a**, UMAP of CD8<sup>+</sup> T cells (left) and violin plots of AUC scores of TNF and TLR1–TLR2 pathways (right) within the identified subclusters CD226<sup>+</sup> CD8<sup>+</sup> T cells (C0), S100A4<sup>+</sup> CD8<sup>+</sup> T cells (C1), CD69<sup>+</sup> GZMK<sup>+</sup> CD8<sup>+</sup> T cells (C2), CD69<sup>hi</sup> GZMK<sup>+</sup> CD8<sup>+</sup> T cells (C3) and KLRC2<sup>+</sup> KLRD1<sup>+</sup> CD8<sup>+</sup> T cells (C4) from all donors and all time points in cohort 1, as in Fig. 1a. Data were analyzed by two-sided Wilcoxon rank-sum test; \*\*\*\* $P < 0.00001$ . **b**, UMAP of NK cells (left) and violin plots of AUC scores of TNF and TLR1–TLR2 pathways (right) in identified subclusters PRF1<sup>+</sup> GZMB<sup>+</sup> NK cells (C0), GZMB<sup>+</sup> KLRF1<sup>+</sup> NK cells (C1), GZMK<sup>+</sup> TGFB1<sup>+</sup> NK cells (C2), IFNG<sup>+</sup> NK cells (C3) and CALR<sup>+</sup> S100A9<sup>+</sup> NK cells (C4) from all donors at all time points as in Fig. 1a. Data were analyzed by two-sided Wilcoxon rank-sum test; \*\*\*\* $P < 0.00001$ . **c**, UMAP of CD14<sup>+</sup> monocytes from all cohort 1 donors at all time points showing subclusters IL1B<sup>+</sup> (MC1), S100A4<sup>+</sup> (MC2), FCN1<sup>+</sup> (MC3) and TGFB1<sup>+</sup> (MC4) cells. **d**, Top significantly upregulated markers in MC1–MC4 CD14<sup>+</sup> monocyte subclusters as in **c**. Plotted genes were significant with a Benjamini–Hochberg method-adjusted  $P$  value of  $<0.05$  (two-sided Wilcoxon test). **e**, Differential enrichment of neighborhoods representing transcriptional states in LC<sup>AM</sup> compared to AI<sup>M</sup> (top) and LC<sup>AM</sup> compared to R<sup>LC</sup> (bottom) at

months 1.7–2.9, 3–5.9, 6–8.9 and 9–11. Each dot represents a neighborhood of ~150–400 cells. Transcriptional states show significant enrichment with a spatial false discovery rate (FDR) of  $<0.1$  ( $F$ -test statistic from the quasiliikelihood  $F$ -test, graph-weighted FDR). **f**, AUC scores of TNF, TGF $\beta$  and WNT– $\beta$ -catenin signaling pathways in MC1–MC4 CD14<sup>+</sup> monocyte subclusters as in **c**. Data were analyzed by two-sided Wilcoxon rank-sum test; \*\*\*\* $P < 0.00001$ . Horizontal dashed lines in **a**, **b** and **f** serve as visual reference for comparison of relative shifts in pathway AUC scores across clusters. **g**, Correlation of the percentage of MC1, MC2, MC3 or MC4 CD14<sup>+</sup> monocyte subclusters with FAS score and pO<sub>2</sub> (LC<sup>AM</sup>,  $n = 38$ ; LC<sup>AS</sup>,  $n = 13$ ; R<sup>LC</sup>,  $n = 9$ ). Data were analyzed by Spearman correlation.  $P$  values were determined using the exact/permutation-based test. The gray shaded area indicates the 95% confidence interval. **h**, Box plot showing FAS score in MC4<sup>hi</sup> ( $>10\%$  of CD14<sup>+</sup> monocytes found within MC4,  $n = 13$ ), MC4<sup>lo</sup> ( $<10\%$  of CD14<sup>+</sup> monocytes within MC4,  $n = 26$ ) and R<sup>LC</sup> ( $n = 7$ ) samples from individuals with LC<sup>AM</sup>, LC<sup>AS</sup> and R<sup>LC</sup>. Data were analyzed by two-sided Wilcoxon rank-sum test; \*\* $P < 0.01$ . The box plot shows the median (center), first and third quartiles (bounds) and 1.5 times the interquartile range (whiskers).





**Fig. 4 | LC-Mo-specific proteins show elevated expression in LC CD14<sup>+</sup> monocytes.** **a**, UMAP of CD14<sup>+</sup> monocytes from individuals with LC<sup>AM</sup> ( $n = 8$ ) from cohort 3 showing *S100A8*<sup>+</sup>*S100A9*<sup>+</sup> CD14<sup>+</sup> monocyte (Clust0), *CTNNB1*<sup>+</sup>*EMP1*<sup>+</sup> CD14<sup>+</sup> monocyte (Clust1) and *FCN1*<sup>+</sup>*CCL3*<sup>+</sup> CD14<sup>+</sup> monocyte (Clust2) subclusters. **b**, LC-Mo signature AUC scores within *S100A8*<sup>+</sup>*S100A9*<sup>+</sup> CD14<sup>+</sup> monocytes (Clust0), *CTNNB1*<sup>+</sup>*EMP1*<sup>+</sup> CD14<sup>+</sup> monocytes (Clust1) and *FCN1*<sup>+</sup>*CCL3*<sup>+</sup> CD14<sup>+</sup> monocytes (Clust2) as in **a**. Data were analyzed by two-sided Wilcoxon rank-sum test; \*\*\*\* $P < 0.00001$ . **c**, LC-Mo AUC scores in Clust1 CD14<sup>+</sup> monocytes from individuals with LC<sup>AM</sup> in cohort 3 with Resp-PASC ( $n = 5$ ) or Resp-PASC-BHR ( $n = 3$ ). Data were analyzed by two-sided Wilcoxon rank-sum test; \*\*\*\* $P < 0.00001$ . Horizontal dashed lines in **b** and **c** serve as visual reference for comparison of relative shifts in pathway AUC scores across clusters. **d**, Percentage of CD14<sup>+</sup>

monocytes among PBMCs in donors recovered from acute COVID-19 ( $R^A$ ) combined with NI ( $R^A + NI$ ,  $n = 10$ ), donors with LC<sup>AM</sup> ( $n = 29$ ) and donors with LC<sup>AS</sup> ( $n = 11$ ) in cohort 4. **e**, MFI of HLA-DQ, CD120b, CALR, CD99 and TGFβ in samples from individuals with LC<sup>AS</sup>, LC<sup>AM</sup> and  $R^A + NI$  as in **d**. **f**, MFI of HLA-DQ, CD120b, CALR and TGFβ in  $R^A + NI$  ( $n = 10$ ) and LC donors in cohort 4 categorized based on mMRC dyspnea score (DS) as DS1 ( $n = 14$ ), DS2 ( $n = 11$ ) and DS3 ( $n = 4$ ) (**f**) or in  $R^A + NI$  ( $n = 10$ ) and LC donors in cohort 4 with fatigue ( $n = 14$ , FAS score 22–34) and extreme fatigue ( $n = 14$ , FAS score 35–47) based on FAS category (**g**).  $P$  values in **d–g** were calculated using a two-sided Wilcoxon rank-sum test; \*\*\*\* $P < 0.00001$ , \*\*\* $P < 0.001$ , \*\* $P < 0.01$  and \* $P < 0.05$ . Box plots show the median (center), first and third quartiles (bounds) and 1.5 times the interquartile range (whiskers).

signaling cascades in NK cells, but not TNF signaling in these cells, also correlated with FAS scores (Fig. 2d). These results indicate that persistent upregulation of inflammatory pathways and cytokines in LC<sup>AM</sup> immune cells might contribute to the clinical symptoms in LC.

### The LC<sup>AM</sup> monocyte signature characterizes a transcriptional state

Next, we performed a reclustering analysis of CD8<sup>+</sup> T cells, NK cells and CD14<sup>+</sup> monocytes from all donor samples. CD8<sup>+</sup> T cells and NK cells each resolved into five clusters (Fig. 3a,b). Differential neighborhood abundance analysis (DA) comparing LC groups at each time point to AI and  $R^{LC}$  was performed. A neighborhood defines a small local group of cells with similar gene expression profiles, representing transitional states. LC<sup>AM</sup> samples exhibited statistically significant increased abundance of neighborhoods in *CD69*<sup>hi</sup>*CD27*<sup>hi</sup> CD8<sup>+</sup> T cells (C3), *GZMB*<sup>+</sup>*KLRF1*<sup>+</sup> NK cells (C1) and *CD69*<sup>+</sup>*TGFB1*<sup>+</sup> NK cells (C2; Extended Data Fig. 4a,b), whereas LC<sup>AS</sup> samples showed increased abundance of neighborhoods in C2 NK cells at months 6–8.9 (Extended Data Fig. 4c,d). C3 CD8<sup>+</sup> T cells and C2 NK cells showed

*GZMK*<sup>+</sup>*GZMB*<sup>lo</sup> signatures (Extended Data Fig. 4e,f), reported to accumulate after SARS-CoV-2 infection and in aging<sup>28–30</sup>. These clusters showed higher expression of TNF and TLR signaling genes (Fig. 3a,b), suggesting the contribution of persistent TNF signaling in the expansion of *CD69*<sup>hi</sup>*CD27*<sup>hi</sup> *GZMK*<sup>+</sup> CD8<sup>+</sup> T cells and *CD69*<sup>+</sup>*TGFB1*<sup>+</sup> *GZMK*<sup>+</sup> NK cells in individuals with LC<sup>AM</sup>.

Within CD14<sup>+</sup> monocytes, four primary clusters (MC1–MC4) were identified (Fig. 3c). MC1 showed high expression of MHC class II molecules, *IL1B* and *NFKB1*; MC2 showed elevated *NFKB1* and *S100A4*; MC3 showed increased expression of *FCN1*, IFN-stimulated genes (*IFI44*, *IFI16* and *IFI30*) and alarmins *S100A8* and *S100A9*; and MC4 displayed higher levels of *IRF1*, *IRF8*, *TGFB1*, *CTNNB1*, *ENG* and *NOTCH1*, among others (Fig. 3d). DA comparing LC samples with AI and  $R^{LC}$  samples across all time points showed a consistent significant increase in MC4 neighborhoods in LC<sup>AM</sup> in both men and women (Fig. 3e and Extended Data Fig. 4g), with this becoming prominent from month 3 onward (Fig. 3e and Extended Data Fig. 5a). By contrast, MC1 neighborhoods showed a marked increase, primarily at months 1.7–2.9, and ‘tapering off’ by month 11 (Fig. 3e). LC<sup>AS</sup> did not exhibit consistent



changes in MC4, except for a small number of neighborhoods at months 6–8.9 attributable to one participant (Extended Data Fig. 5b). Further, area under the curve (AUC) scores of pathways (calculated per cell from all donors) revealed that MC4, which was uniquely abundant in LC<sup>AM</sup>, showed significantly higher expression of TGFβ and WNT-β-catenin signaling genes than MC1, MC2 and MC3 (Fig. 3f). MC1 showed higher expression of the TNF signaling genes (Fig. 3f), whereas MC1 and MC3 showed higher IFNγ response gene expression (Extended Data Fig. 5c). We further performed trajectory analysis (unstratified by disease category or groups) that revealed that lineage 3 overlapped closely with the MC4 immune program (Extended Data Fig. 5d,e), indicating that MC4 cells in LC<sup>AM</sup> have a distinct transcriptional profile compared to MC1–MC3. We next assessed the correlation between the frequency of MC4 within CD14<sup>+</sup> monocytes for all LC and R<sup>LC</sup> samples from all time points with clinical parameters. A modest but statistically significant positive correlation was found between MC4 proportion and FAS score, whereas the correlation with pO<sub>2</sub> was negative (Fig. 3g). By contrast, a higher MC1 proportion was negatively correlated with FAS score (Fig. 3g). The modest MC4–FAS correlation likely reflected participant heterogeneity (Extended Data Fig. 5f). Individuals with LC with a high proportion of MC4 (MC4<sup>hi</sup>) exhibited significantly greater fatigue than those with LC with a low proportion of MC4 (MC4<sup>lo</sup>) or R<sup>LC</sup> (Fig. 3h). These findings indicate that increased MC4 abundance (referred to hereafter as LC monocyte transcriptional state (LC-Mo state)) is associated with LC, as demonstrated by its correlation with both FAS scores and pO<sub>2</sub> levels.

### LC monocytes exhibit increased LC-Mo protein marker expression

To validate the LC-Mo state, we generated and analyzed scRNA-seq data from PBMCs from cohort 3, comprising eight individuals with LC<sup>AM</sup> with LC symptoms reported for 8–42 months at the time of sampling (Supplementary Table 3 and Methods). All individuals with LC reported fatigue and dyspnea (classified as respiratory PASC (‘Resp-PASC’), *n* = 5), and three exhibited bronchial hyper-responsiveness (BHR)<sup>31</sup>, termed ‘Resp-PASC-BHR’ (*n* = 3). Three clusters (Clust0–Clust2) were identified within CD14<sup>+</sup> monocytes (Fig. 4a). Clust1 showed significantly elevated AUC scores for the LC-Mo signature (Fig. 4b). Individuals with Resp-PASC-BHR showed significantly higher expression of the LC-Mo signature in Clust1 than those with Resp-PASC (Fig. 4c), providing independent validation of the LC-Mo state in LC<sup>AM</sup> and suggesting a link with progression to severe respiratory PASC.

We next performed flow cytometry analysis on PBMCs from donors in cohort 4, which included 40 LC samples 3–14 months after acute COVID-19 (Supplementary Table 4 and Methods) and 10 R<sup>A</sup> or NI donors. LC showed a significant increase in CD14<sup>+</sup> monocyte percentages compared to R<sup>A</sup> + NI (Fig. 4d), independent of acute COVID-19 severity. We assessed the expression of 11 proteins (HLA-DR, HLA-DQ, CD105, CD51, TGFβ1, CD99, CD120b, CALR, IRF8, IFNGR1 and CD163) corresponding to LC-Mo transcripts in total CD14<sup>+</sup> monocytes in samples from individuals with LC and R<sup>A</sup> + NI (Extended Data Fig. 6a). Median fluorescence intensity (MFI) of HLA-DQ, CD120b, CALR, CD99

and TGFβ1 was significantly higher in LC compared to in R<sup>A</sup> + NI (Fig. 4e and Extended Data Fig. 6b), whereas HLA-DR, CD51, CD105, IRF8, IFNGR1 and CD163 showed no significant difference (Extended Data Fig. 6c). Stratification by fatigue scores and dyspnea (range 1–3) revealed consistently higher MFI of CALR, CD120b, HLA-DQ and TGFβ1 in those with more severe LC symptoms (Fig. 4f,g), and TGFβ1 MFI inversely correlated with pO<sub>2</sub> (Extended Data Fig. 6d, top). MFI of both TGFβ1 and IRF8 positively correlated with each other (Extended Data Fig. 6d, bottom). Thus, protein markers specific to LC-Mo were elevated in LC, supporting an association between the LC-Mo signature and LC pathology.

### Chromatin profiling reveals AP-1/NF-κB1 activity in LC-Mo

We next investigated epigenetic regulation using snATAC-seq data from individuals with LC in cohort 1. Examination of motif signals in the chromatin landscapes of CD14<sup>+</sup> monocytes, CD8<sup>+</sup> T cells and NK cells from individuals with LC<sup>AM</sup> compared to those with R<sup>LC</sup> at multiple time points identified a persistent positive signal for AP-1 family activity in CD8<sup>+</sup> T cells and NK cells (Extended Data Fig. 7a). In CD14<sup>+</sup> monocytes, AP-1 motif accessibility was elevated up to month 5.9, after which motif enrichment shifted toward transcription factors involved in downstream TGFβ signaling, notably SP1 and KLF family of transcription factors at months 3–8.9 (Extended Data Fig. 7b). MC4 showed the highest number of differentially accessible regions (Fig. 5a). The open chromatin landscape of MC4 showed highest enrichment for motifs for ETS family transcription factors, including GABPA, ETV1, ETV4, SPI1 and SPIC (Fig. 5b). Correlating open chromatin regions with gene expression revealed significant positive associations for proangiogenic and cell adhesion genes (*VEGFA*, *ENG*, *TGFB1*, *RXRα*, *ICAM1* and *ITGA5*) and genes implicated in inflammatory/metabolic diseases (*TTC7A*, *LMNA* and *IER3*) among others (Fig. 5c). AP-1 family, SMADs, NF-κB1 and RELA transcription factor motifs showed a marked increase within the accessible chromatin regions of these genes (Fig. 5d). Within MC4, correlation of transcription factor transcripts and target gene transcripts with accessible motifs enabled pinpointing of noncoding regulatory regions associated with gene expression, such as those for *IER3* and *LMNA*, and establishment of gene–transcription factor relationships (such as NF-κB1 and AP-1 family likely regulators of *LMNA*; Fig. 5e–g and Extended Data Fig. 7c). In summary, these findings indicate that LC-Mo is driven by ETS, AP-1 and NF-κB1 transcription factors.

### BAL myeloid cells show LC-Mo and profibrotic programs

Circulating monocytes contribute to PASC pathogenesis, particularly pulmonary fibrosis<sup>23,32</sup>. To assess whether LC-Mo participates in fibrotic lung remodeling, we analyzed paired PBMC and BAL fluid samples from a public dataset<sup>23</sup> (cohort 5) consisting of nine individuals with LC of unknown severity during AI (LC<sup>UN</sup>), classified based on lung function as Resp-PASC (*n* = 5) or nonResp-PASC (*n* = 4), and PBMCs from NI donors (*n* = 2; Supplementary Table 5). Circulating CD14<sup>+</sup> myeloid cells were reclustered to identify CD14<sup>+</sup>CD16<sup>+</sup> monocytes (Extended Data Fig. 8a), leading to six clusters (CL0–CL5; Fig. 6a). CL5 showed the highest enrichment of LC-Mo signature AUC scores (Fig. 6b and Extended Data Fig. 8b). Within cluster 5,

**Fig. 5 | AP-1 and NF-κB1 transcription factors regulate LC-Mo in CD14<sup>+</sup> monocytes from individuals with LC<sup>AM</sup>.** **a**, Top significant peaks calculated from snATAC-seq data from cohort 1 (*n* = 78), with aggregated peaks from cells in MC1–MC4 subclusters; data were analyzed by two-sided Wilcoxon test (Benjamini–Hochberg method-adjusted *P* < 0.05). **b**, Top ChromVar transcription factor motif enrichment in open chromatin regions in MC4 compared to MC1–MC3; Avg diff, average difference. **c**, Expression of genes significantly correlated with open chromatin regions within the MC4 subcluster (Pearson’s correlation (Benjamini–Hochberg method-adjusted *P* < 0.05)). **d**, Scatter plot of enriched transcription factors motifs (fold enrichment) within open chromatin regions correlated with the expression of genes in **c** against the ChromVar transcription

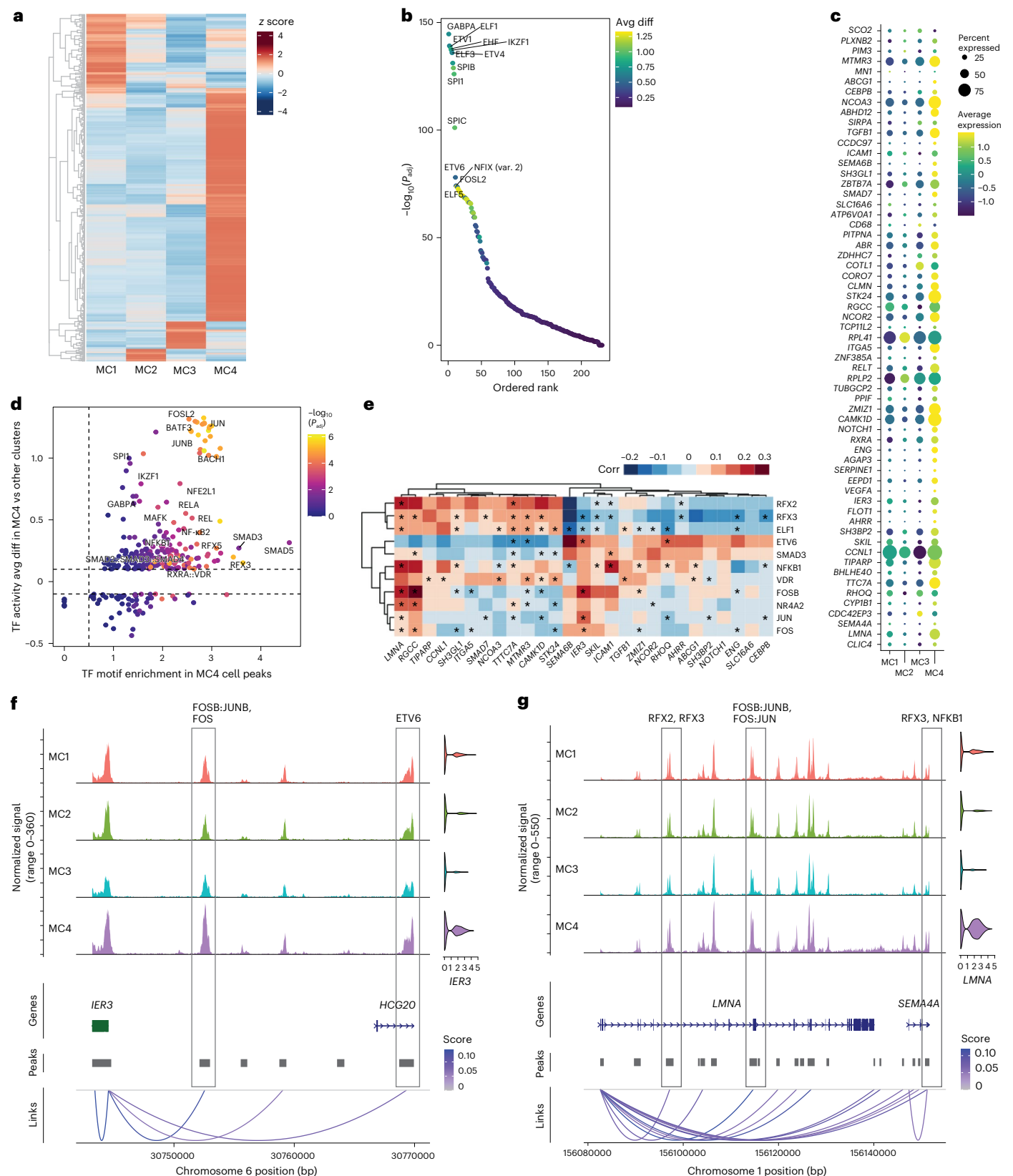
factors motif enrichment as in **b**; TF, transcription factor. Dashed horizontal lines represent –0.1 and 0.1 ChromVAR average difference; dashed vertical line represents 0.5 fold enrichment. **e**, Correlation (Corr) between the expression of genes in **c** with the expression of transcription factors identified in **b** and **d**. Data were analyzed by Pearson’s correlation. Absolute correlations of >0.3 are plotted. The black asterisk indicates transcription factors with motifs in the open chromatin of the correlated gene. **f,g**, Coverage plots showing the chromatin accessibility regions and gene expression of *IER3* (**f**) and *LMNA* (**g**) in subclusters MC1–MC4 and correlations between open chromatin regions and transcription factors with binding sites (gray lined boxes); bp, base pairs; *P*<sub>adj</sub>, adjusted *P* value.

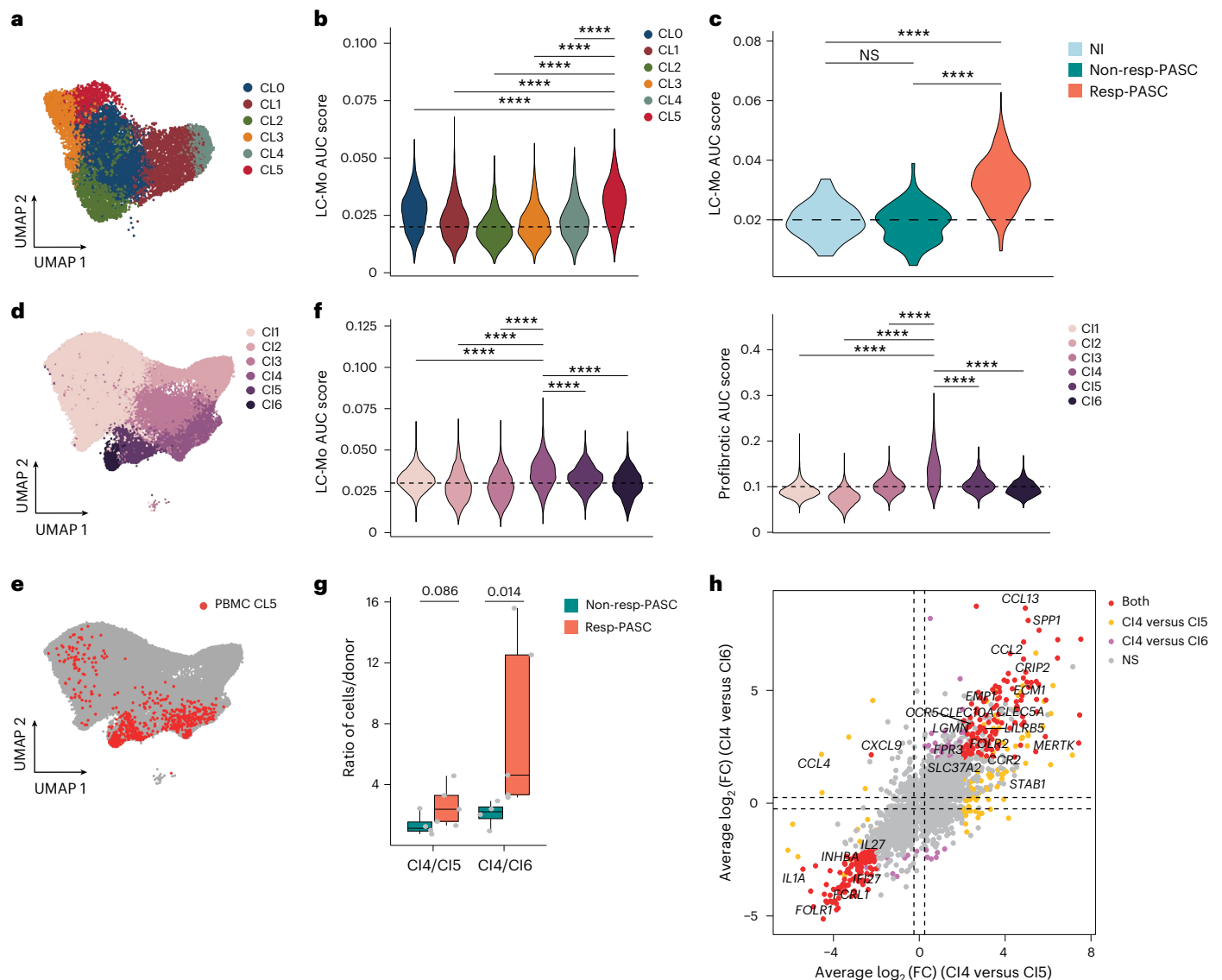


Resp-PASC exhibited significantly higher LC-Mo expression than nonResp-PASC or NI (Fig. 6c).

We next integrated CD14<sup>+</sup> monocytes from PBMCs and CD163<sup>+</sup> or CD14<sup>+</sup> myeloid cells from BAL fluid. This integrated dataset identified C11 with >75% cells from BAL fluid and expressing *MARCO*<sup>+</sup>*FABP4*<sup>+</sup>, markers for tissue-resident alveolar macrophages, two clusters

(C12 and C13) with >65% of cells from PBMCs and expressing *LYZ*<sup>+</sup>*CD14*<sup>+</sup>, markers for circulating monocytes, and three mixed clusters (C14–C16) with comparable proportion of cells from both PBMCs and BAL (Fig. 6d and Extended Data Fig. 8c,d). PBMC monocytes in C15 primarily localized to clusters C14–C16 (Fig. 6e), suggesting a macrophage-polarized phenotype. Among these, cluster C14 had the highest LC-Mo signature





**Fig. 6 | The LC-Mo cluster is enriched in profibrotic monocyte-derived alveolar macrophages in BAL fluid from individuals with LC.** **a**, UMAP of CD14<sup>+</sup> monocytes cells from PBMCs of cohort 5 (GEO: [GSE263817](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE263817))<sup>23</sup> subclusters *FABP4*<sup>+</sup>*CIQA*<sup>+</sup> CD14<sup>+</sup> monocytes (CL0), *NKG7*<sup>+</sup>*GZMB*<sup>+</sup> CD14<sup>+</sup> monocytes (CL1), *FCN1*<sup>+</sup>*S100A9*<sup>+</sup> CD14<sup>+</sup> monocytes (CL2), *KLRC2*<sup>+</sup>*LAG3*<sup>+</sup> CD14<sup>+</sup> monocytes (CL3), *NLRCS*<sup>+</sup> CD14<sup>+</sup> monocytes (CL4) and *TREM2*<sup>+</sup>*CALR*<sup>+</sup> CD14<sup>+</sup> monocytes (CL5) from NI donors ( $n = 2$ ) and LC<sup>UN</sup> donors ( $n = 9$ ). **b**, AUC scores of the LC-Mo signature in CL0–CL5 as in **a**; data were analyzed by two-sided Wilcoxon rank-sum test; \*\*\*\* $P < 0.00001$ . **c**, AUC scores of the LC-Mo signature in CL5 in PBMC CD14<sup>+</sup> monocytes from individuals with LC Resp-PASC ( $n = 5$ ) and nonResp-PASC ( $n = 4$ ) and NI donors ( $n = 2$ ); data were analyzed by two-sided Wilcoxon rank-sum test; \*\*\*\* $P < 0.00001$ . **d**, Integrated UMAP of CD163<sup>+</sup> or CD14<sup>+</sup> myeloid cells from PBMCs and BAL samples of individuals with LC<sup>UN</sup> ( $n = 9$ ) showing *MARCO*<sup>+</sup>*FABP4*<sup>+</sup> macrophages (C1), *LYZ*<sup>+</sup> CD14<sup>+</sup> monocytes (C2–C3) and mix clusters from

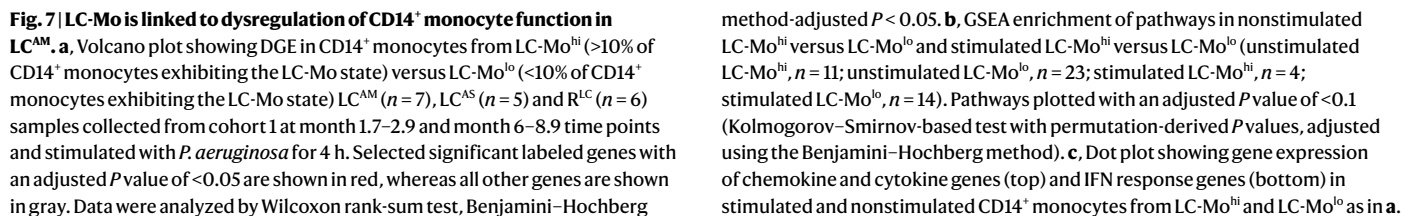
PBMCs and BAL samples with *TREM2*<sup>+</sup>*CCL2*<sup>+</sup> (C14), *CCL23*<sup>+</sup> (C15) and *NUPR8*<sup>+</sup> (C16). **e**, UMAP as in **d** showing CL5 cells. **f**, LC-Mo AUC score within C11–C16 (left) and profibrotic gene signature<sup>33</sup> AUC score as in **d**. Data were analyzed by two-sided Wilcoxon rank-sum test; \*\*\*\* $P < 0.00001$ . Horizontal dashed lines in **b**, **c** and **f** serve as visual reference for comparison of relative shifts in pathway AUC scores across clusters. **g**, Ratio of C14/C15 or C14/C16 cells within each individual with Resp-PASC ( $n = 5$ ) or nonResp-PASC ( $n = 4$ ). Data were analyzed by two-sided Wilcoxon rank-sum test. Box plots show the median (center), first and third quartiles (bounds) and 1.5 times the interquartile range (whiskers). **h**, Scatter plots showing log<sub>2</sub> (fold change) (log<sub>2</sub> (FC)) of DGE in the C14 versus C15 and C14 versus C16 clusters. Genes significant with an adjusted  $P$  value of <0.05 in both comparisons are labeled (two-sided Wilcoxon rank-sum test, Benjamini–Hochberg method-adjusted  $P < 0.05$ ). All data correspond to cohort 5 scRNA-seq data.

enrichment and higher expression of a profibrotic gene set defined in prior COVID-19 BAL studies<sup>33</sup> and including *TREM2*, *CALM1*, *LGMN* and *APOE* (Fig. 6f and Extended Data Fig. 8e). Individuals with resp-PASC showed a higher proportion of C14 cells and higher C14/C15 and C14/C16 ratios than individuals without resp-PASC (Fig. 6g and Extended Data Fig. 8f). Differential expression analysis revealed that C14 cells upregulated the expression of *SPP1*, *CCL13*, *CCL2* and *FOLR2* compared to C15 or C16 cells from both individuals with resp-PASC and non-resp-PASC (Fig. 6h). These results indicate LC-Mo enrichment in

Resp-PASC PBMCs and its association with a profibrotic transcriptional profile in lung myeloid cells.

### Individuals with LC-Mo exhibit dysregulated monocyte function

To assess the functional implications of LC-Mo during immune challenge, we stimulated PBMC samples from cohort 1 (months 1.7–2.9 and 6–8.9) with heat-inactivated *Pseudomonas aeruginosa* for 4 h and performed single-cell multiome profiling in samples from



Although 14 individuals in cohort 1 and 51 in cohort 2 were enrolled before the 3-month National Academies of Sciences, Engineering, and Medicine cutoff for LC, over 70% had symptoms extending beyond this period, aligning with established diagnostic criteria. Overlapping symptoms with post-intensive care syndrome complicate LC heterogeneity. Our data revealed molecular differences in PBMCs based on the severity of the AI. Circulating monocytes have been implicated in severe COVID-19 disease and in resp-PASC<sup>23,32,33,36,37</sup>. Although oxygen saturation in cohort 1 was normal, MC4 cell proportions negatively

correlated with oxygen saturation, suggesting subtle gas exchange defects. Nevertheless, increased LC-Mo expression in PBMCs and BAL fluid from individuals with severe resp-PASC in cohorts 3 and 5, together with a profibrotic BAL phenotype, support a link to lung pathology.

Sustained TNF expression, reported in post-COVID cohorts<sup>8,13,38</sup>, paralleled persistent TNF and/or NF- $\kappa$ B signaling in immune subsets, driving systemic inflammation. Enrichment of *GZMK*<sup>+</sup>CD8<sup>+</sup> T and NK cells in LC<sup>AM</sup>, shown to expand after SARS-CoV-2 infection<sup>30</sup> and in chronic inflammatory diseases<sup>28,29,39,40</sup>, and increased AP-1 accessibility were consistent with TNF-driven activation<sup>41</sup>. Upregulated TLR pathways in CD8<sup>+</sup> T cells and NK cells indicated ongoing viral sensing and potential NF- $\kappa$ B1 activation, fitting with evidence of persistent viral reservoirs or remnants<sup>42–45</sup>.

By contrast, CD14<sup>+</sup> monocytes from individuals with LC<sup>AM</sup> showed transient TNF signaling, with persistent *TGFB1* expression and activation of TGF $\beta$  and WNT- $\beta$ -catenin pathways for up to 11 months. MC4 proportions showed coexpression of TGF $\beta$  and IRF8 mRNA and protein, shown to synergistically drive neuroinflammation in the experimental autoimmune encephalomyelitis mouse model<sup>46</sup>, and displayed motifs for ETS1, AP-1, NF- $\kappa$ B1 and SMAD, transcription factors linked to adhesion and fibrosis<sup>47,48</sup>. MC4 proportions modestly correlated with FAS scores and blood pO<sub>2</sub>, and flow cytometry confirmed elevated TGF $\beta$ 1 in CD14<sup>+</sup> monocytes. LC-Mo enrichment was also validated in two independent cohorts with severe Resp-PASC patients, thereby linking LC-Mo to lung fibrosis.

Although associations between LC-Mo and symptom severity were noted, correlations were modest, leaving causality undetermined. Further functional studies are needed to clarify these mechanisms. Our study focused on respiratory symptoms and fatigue, so the involvement of LC-Mo in other organ systems remains open. Comparison to unstratified recovered controls (due to limited sample size) restricts interpretation; future work with stratified groups and consideration of vaccination or comorbidities will be important.

In conclusion, we provide a systems view of LC with fatigue and respiratory involvement, identifying a pathogenic monocyte state linked to severe symptoms and offering insights into disease mechanisms and heterogeneity.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41590-025-02387-1>.

## References

- Davis, H. E. et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *eClinicalMedicine* **38**, 101019 (2021).
- McCorkell, L. et al. Patient-led research collaborative: embedding patients in the long COVID narrative. *Pain Rep.* **6**, e913 (2021).
- Hastie, C. E. et al. True prevalence of long-COVID in a nationwide, population cohort study. *Nat. Commun.* **14**, 7892 (2023).
- Nalbandian, A. et al. Post-acute COVID-19 syndrome. *Nat. Med.* **27**, 601–615 (2021).
- Archambault, P. M. et al. Post-COVID-19 condition symptoms among emergency department patients tested for SARS-CoV-2 infection. *Nat. Commun.* **15**, 8449 (2024).
- Klein, J. et al. Distinguishing features of long COVID identified through immune profiling. *Nature* **623**, 139–148 (2023).
- Su, Y. et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell* **185**, 881–895 (2022).
- Schultheiß, C. et al. The IL-1 $\beta$ , IL-6, and TNF cytokine triad is associated with post-acute sequelae of COVID-19. *Cell Rep. Med.* **3**, 100663 (2022).
- Lu, S. et al. Early biological markers of post-acute sequelae of SARS-CoV-2 infection. *Nat. Commun.* **15**, 7466 (2024).
- Phetsouphanh, C. et al. Immunological dysfunction persists for 8 months following initial mild-to-moderate SARS-CoV-2 infection. *Nat. Immunol.* **23**, 210–216 (2022).
- Wong, A. C. et al. Serotonin reduction in post-acute sequelae of viral infection. *Cell* **186**, 4851–4867 (2023).
- Altmann, D. M., Whettlock, E. M., Liu, S., Arachchillage, D. J. & Boyton, R. J. The immunology of long COVID. *Nat. Rev. Immunol.* **23**, 618–634 (2023).
- Davis, H. E., McCorkell, L., Vogel, J. M. & Topol, E. J. Long COVID: major findings, mechanisms and recommendations. *Nat. Rev. Microbiol.* **21**, 133–146 (2023).
- Greenhalgh, T., Sivan, M., Perlowski, A. & Nikolich, J. Z. Long COVID: a clinical update. *Lancet* **404**, 707–724 (2024).
- Peluso, M. J. & Deeks, S. G. Mechanisms of long COVID and the path toward therapeutics. *Cell* **187**, 5500–5529 (2024).
- Liew, F. et al. Large-scale phenotyping of patients with long COVID post-hospitalization reveals mechanistic subtypes of disease. *Nat. Immunol.* **25**, 607–621 (2024).
- Cervia-Hasler, C. et al. Persistent complement dysregulation with signs of thromboinflammation in active long COVID. *Science* **383**, eadg7942 (2024).
- Al-Aly, Z. et al. Long COVID science, research and policy. *Nat. Med.* **30**, 2148–2164 (2024).
- Hanson, A. L. et al. Iron dysregulation and inflammatory stress erythropoiesis associates with long-term outcome of COVID-19. *Nat. Immunol.* **25**, 471–482 (2024).
- Yin, K. et al. Long COVID manifests with T cell dysregulation, inflammation and an uncoordinated adaptive immune response to SARS-CoV-2. *Nat. Immunol.* **25**, 218–225 (2024).
- A Clinical Case Definition of Post COVID-19 Condition by a Delphi Consensus (World Health Organization, 2021); [https://www.who.int/publications/i/item/WHO-2019-nCoV-Post\\_COVID-19\\_condition-Clinical\\_case\\_definition-2021.1](https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1)
- Kocuzalla, A. R. et al. S1 guideline post-COVID/long-COVID. *Pneumologie* **75**, 869–900 (2021).
- Li, C. et al. Comparative single-cell analysis reveals IFN- $\gamma$  as a driver of respiratory sequelae after acute COVID-19. *Sci. Transl. Med.* **16**, eadn0136 (2024).
- De Vries, J., Michielsen, H., Van Heck, G. L. & Drent, M. Measuring fatigue in sarcoidosis: the Fatigue Assessment Scale (FAS). *Br. J. Health Psychol.* **9**, 279–291 (2004).
- Drent, M., Lower, E. E. & De Vries, J. Sarcoidosis-associated fatigue. *Eur. Respir. J.* **40**, 255–263 (2012).
- Hendriks, C., Drent, M., Elfferich, M. & De Vries, J. The Fatigue Assessment Scale: quality and availability in sarcoidosis and other diseases. *Curr. Opin. Pulm. Med.* **24**, 495–503 (2018).
- Klok, F. A. et al. The post-COVID-19 functional status scale: a tool to measure functional status over time after COVID-19. *Eur. Respir. J.* **56**, 2001494 (2020).
- Jonsson, A. H. et al. Granzyme K<sup>+</sup> CD8 T cells form a core population in inflamed human tissue. *Sci. Transl. Med.* **14**, eabo0686 (2022).
- Terekhova, M. et al. Single-cell atlas of healthy human blood unveils age-related loss of NKG2C<sup>+</sup>GZMB<sup>+</sup>CD8<sup>+</sup> memory T cells and accumulation of type 2 memory T cells. *Immunity* **56**, 2836–2854 (2023).
- Notarbartolo, S. et al. Integrated longitudinal immunophenotypic, transcriptional and repertoire analyses delineate immune responses in COVID-19 patients. *Sci. Immunol.* **6**, eabg5021 (2021).
- Krooss, S. A. et al. Bronchodilator response in post-COVID-19 patients undergoing pulmonary rehabilitation. *Respiration* **103**, 622–629 (2024).



32. Bailey, J. I. et al. Profibrotic monocyte-derived alveolar macrophages are expanded in patients with persistent respiratory symptoms and radiographic abnormalities after COVID-19. *Nat. Immunol.* **25**, 2097–2109 (2024).
33. Wendisch, D. et al. SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis. *Cell* **184**, 6243–6261 (2021).
34. Däbritz, J. et al. Activation-dependent cell death of human monocytes is a novel mechanism of fine-tuning inflammation and autoimmunity. *Eur. J. Immunol.* **46**, 1997–2007 (2016).
35. Monguió-Tortajada, M., Franquesa, M., Sarrias, M.-R. & Borràs, F. E. Low doses of LPS exacerbate the inflammatory response and trigger death on TLR3-primed human monocytes. *Cell Death Dis.* **9**, 499 (2018).
36. Schulte-Schrepping, J. et al. Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* **182**, 1419–1440 (2020).
37. Zhang, B. et al. Altered and allele-specific open chromatin landscape reveals epigenetic and genetic regulators of innate immunity in COVID-19. *Cell Genom.* **3**, 100232 (2023).
38. Mohd Zawawi, Z. et al. Prospective roles of tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) in COVID-19: prognosis, therapeutic and management. *Int. J. Mol. Sci.* **24**, 6142 (2023).
39. Lan, F. et al. GZMK-expressing CD8<sup>+</sup> T cells promote recurrent airway inflammatory diseases. *Nature* **638**, 490–498 (2025).
40. Zhao, L. et al. Characteristics and functions of an atypical inflammation-associated GZMK<sup>+</sup>GZMB<sup>+</sup>CD8<sup>+</sup> T subset in people living with HIV-1. *Mol. Immunol.* **173**, 40–52 (2024).
41. Qiao, Y. et al. AP-1 is a key regulator of proinflammatory cytokine TNF $\alpha$ -mediated triple-negative breast cancer progression. *J. Biol. Chem.* **291**, 5068–5079 (2016).
42. Talla, A. et al. Persistent serum protein signatures define an inflammatory subcategory of long COVID. *Nat. Commun.* **14**, 3417 (2023).
43. Schultheiß, C. et al. Liquid biomarkers of macrophage dysregulation and circulating spike protein illustrate the biological heterogeneity in patients with post-acute sequelae of COVID-19. *J. Med. Virol.* **95**, e28364 (2023).
44. Peluso, M. J. et al. Plasma markers of neurologic injury and inflammation in people with self-reported neurologic postacute sequelae of SARS-CoV-2 infection. *Neurol. Neuroimmunol. Neuroinflamm.* **9**, e200003 (2022).
45. Zhao, J. et al. Plasma biomarkers for systemic inflammation in COVID-19 survivors. *Proteomics Clin. Appl.* **16**, e2200031 (2022).
46. Yoshida, Y. et al. The transcription factor IRF8 activates integrin-mediated TGF- $\beta$  signaling and promotes neuroinflammation. *Immunity* **40**, 187–198 (2014).
47. Leask, A. Eureka! ETS a target for fibrosis! *J. Cell Commun. Signal.* **5**, 325–326 (2011).
48. Ende, G. et al. TNF- $\alpha$ -mediated adhesion of monocytes to endothelial cells—the role of ephrinA1. *J. Mol. Cell. Cardiol.* **77**, 125–135 (2014).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

<sup>1</sup>Centre for Individualised Infection Medicine (CiiM), a joint venture between the Helmholtz Centre for Infection Research (HZI) and Hannover Medical School (MHH), Hannover, Germany. <sup>2</sup>TWINCORE, a joint venture between the Helmholtz Centre for Infection Research (HZI) and the Hannover Medical School (MHH), Hannover, Germany. <sup>3</sup>Beirne B. Carter Centre for Immunology Research, University of Virginia, Charlottesville, VA, USA. <sup>4</sup>Division of Infectious Disease and International Health, Department of Medicine, University of Virginia, Charlottesville, VA, USA. <sup>5</sup>Hannover Unified Biobank, Hannover Medical School (MHH), Hannover, Germany. <sup>6</sup>Institute of Virology, Hannover Medical School (MHH), Hannover, Germany. <sup>7</sup>Institute for Pulmonary Rehabilitation Research, Schön Klinik Berchtesgadener Land, Schöna am Königssee, Germany. <sup>8</sup>PRACTIS Clinician Scientist Program, Dean's Office for Academic Career Development, Hannover Medical School, Hannover, Germany. <sup>9</sup>Department of Gastroenterology, Hepatology, Infectious Diseases and Endocrinology, Hannover Medical School (MHH), Hannover, Germany. <sup>10</sup>Department of Human Genetics, Hannover Medical School (MHH), Hannover, Germany. <sup>11</sup>German Centre for Infection Research (DZIF), partner site Hannover-Braunschweig, Hannover, Germany. <sup>12</sup>Cluster of Excellence Resolving Infection Susceptibility (RESIST; EXC 2155), Hannover Medical School, Hannover, Germany. <sup>13</sup>Department of Pulmonary Rehabilitation, German Center for Lung Research (DZL), University Medical Center Giessen and Marburg, Philipps-University Marburg (UMR), Marburg, Germany. <sup>14</sup>Department of Medicine, Pulmonary and Critical Care Medicine, University Medical Center Giessen and Marburg, Philipps University Marburg, German Center for Lung Research (DZL), Marburg, Germany. <sup>15</sup>Teaching Hospital, Paracelsus Medical University Salzburg, Salzburg, Austria. <sup>16</sup>Department of Respiratory Medicine and Infectious Diseases, Hannover Medical School, Hannover, Germany. <sup>17</sup>Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), German Center for Lung Research (DZL), Hannover, Germany. <sup>18</sup>Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, the Netherlands. <sup>19</sup>Institute for Molecular Bacteriology, TWINCORE, Centre for Experimental and Clinical Infection Research, Hannover, Germany. <sup>20</sup>Department of Clinical Microbiology, Rigshospitalet, Copenhagen, Denmark. <sup>21</sup>Department of Molecular Bacteriology, Helmholtz Centre for Infection Research, Braunschweig, Germany. <sup>22</sup>Department of Immunology and Metabolism, Life and Medical Sciences Institute (LIMES), University of Bonn, Bonn, Germany. <sup>23</sup>Lower Saxony Center for Artificial Intelligence and Causal Methods in Medicine (CAIMed), Hannover, Germany. <sup>24</sup>These authors jointly supervised this work: Thomas Illig, Jie Sun, Yang Li. ✉e-mail: [yang.li@helmholtz-hzi.de](mailto:yang.li@helmholtz-hzi.de)

## Methods

### Cohorts

**Cohorts 1, 2 and 4.** Sample collection and analyses were approved by the ethics committee of the MHH (ethics vote 9001\_BO\_K). All participants gave written broad consent. Blood was collected from individuals who were admitted to the hospital due to COVID-19, as well as from ambulatory individuals after SARS-CoV-2 infection showing symptoms indicative of LC >4 weeks after infection in accordance with the German S1 guideline for LC and post-COVID syndrome at the MHH. Cohort 1 included individuals with AI<sup>M</sup> ( $n = 7$ , 42.8% women, age median = 52, range 23–66 years) and AI<sup>S</sup> ( $n = 4$ , 50% women, age median = 37, range 32–54 years), individuals with LC symptoms (LC<sup>AM</sup> ( $n = 29$ , 8 longitudinal with two to three time points, 21 single-time-point donors, 58% women, median age = 49, range 31–84) and LC<sup>AS</sup> ( $n = 8$ , 3 longitudinal with two to four time points, 5 single-time-point donors, 25% women, median age = 46, range 19–75)), donors recovered from LC (4–8 months of LC; R<sup>LC</sup> ( $n = 8$ , 37.5% women, median age = 38, range 19–65)) and pre-pandemic NI control individuals ( $n = 6$ , 50% women, median age = 40, range 24–61). LC and R<sup>LC</sup> samples were collected 1.7–10.2 months after infection, and all groups were recruited between April 2020 and August 2021 at MHH. Cohort 2 included individuals with LC<sup>AM</sup> ( $n = 117$ , 24 longitudinal with two to four time points, 93 single-time-point donors, 58.9% women, median age = 48, range 19–83) and LC<sup>AS</sup> ( $n = 25$ , 12 longitudinal, 13 single-time-point donors, 20% women, median age = 53, range 18–81), recruited between May 2020 and August 2021 at MHH, along with pre-pandemic NI samples ( $n = 33$ , 48.4% women, median age = 40, range 25–65). Cohort 4 included individuals with LC<sup>AM</sup> ( $n = 29$ , 58.6% women, median age = 49, range 33–72) and LC<sup>AS</sup> ( $n = 11$ , 18% women, median age = 57, range 35–81) and individuals recovered from AI (R<sup>A</sup>  $n = 8$  and NI  $n = 2$ , 60% women, median age = 41, range 29–67), recruited between August 2020 and June 2022 at MHH. All individuals with AI had a positive SARS-CoV-2 PCR test at admission or externally before admission. All individuals with LC had a prior proven SARS-CoV-2 infection. The SARS-CoV-2 strain in individuals with AI or LC was not recorded. Clinical parameters, including blood gas measurements, pulmonary function tests, FAS and mMRC scores and quality-of-life assessments, were systematically collected at each visit.

**Cohort 3.** Sample collection and analyses were approved by the ethics committee of the Philipps University Marburg (Az.:24-289 ‘Entschlüsselung der molekularen Pathophysiologie des Post-Covid-Syndroms und prädisponierender Faktoren mit Hilfe neuer Sequenzierungstechnologien und Phänotypisierung von Immunzellen’). All participants gave written broad consent. Samples were collected from individuals with LC<sup>AM</sup> ( $n = 8$ , 62.5% women, median age = 45, range 21–63), all with resp-PASC and recruited between October 2023 and November 2023 during their stay at the Pulmonary Rehabilitation Clinic in Schöna am Königssee, Germany. All individuals with LC had prior proven SARS-CoV-2 infection, and samples were collected >6 months after SARS-CoV-2 infection, one sample per participant. The SARS-CoV-2 strain in individuals with AI or LC was not recorded. Clinical parameters, including blood gas measurements, pulmonary function tests, FAS and mMRC scores and quality-of-life assessments, were systematically collected for each participant.

**Cohort 5.** Biosample collection for both PBMCs and BAL fluid is available at ref. 23 and included individuals with LC ( $n = 9$ , 44.4% women, median age = 64, range 62–83), including five with resp-PASC, and NI donors ( $n = 2$ , 50% women, median age 77, range 73–77), recruited between October 2020 and November 2021 at Mayo Clinic.

### Sample processing for PBMCs

Sample processing for cohorts 1–4 and storage was performed following the standard procedures of the Hannover Unified Biobank (HUB) as described by Kopfnagel et al.<sup>49</sup>. PBMCs were isolated from whole

blood using Ficoll gradient centrifugation. Cohort 5 PBMC and BAL sample processing was performed similar to as described previously<sup>23</sup>.

### Cytokine assay (cohort 2)

The Quanterix HD SP-X Imaging and Analysis System was used to measure plasma samples. The following panels were used in this study: Human Corplex cytokine panel 110-Plex array including IL-12p70, IL-1 $\beta$ , IL-4, IL-5, IL-6, IL-8, TNF, IFN $\gamma$ , IL-10 and IL-22. The Simoa chemokine panel 14-plex kit contained four chemokines, including IP-10 (CXCL10), MCP1 (CCL2), MIP1- $\beta$  (CCL19) and ITAC (CXCL11). IL-4 and IL-5 were excluded from further analysis due to being below the limit of detection. The detection values were log<sub>2</sub> transformed. All plasma samples were processed according to standard biobanking protocols and stored at a minimum temperature of –80 °C. For the experiments, the samples were randomized and measured according to the manufacturer’s manual. The study protocol conformed to the ethical guidelines of the Declaration of Helsinki, and the ethics committee of MHH approved this study a priori (9001\_BO\_K, No. 9472\_BO\_K\_2020, broad consent: 2923-2015). Informed consent was obtained from all participants included in the study.

### *P. aeruginosa* stock production

*A. P. aeruginosa* clinical isolate CH5464 was streaked from frozen glycerol stocks onto LB agar plates and incubated overnight at 37 °C. Bacteria from single colonies were used to inoculate an overnight preculture in LB medium. This preculture was then diluted in fresh LB medium and grown at 37 °C with shaking at 180 rpm until reaching the early stationary phase. The culture was centrifuged at 10,000g for 10 min, and the supernatant was discarded. The pellet was washed twice with PBS and incubated at 80 °C for 60 min to inactivate the bacteria in a waterbath. Afterward, the suspension was centrifuged again at 10,000g for 10 min to remove cellular debris. The bacterial suspension was adjusted to a concentration of 10<sup>8</sup> colony-forming units (c.f.u.) per ml and stored at –20 °C for future use. To confirm complete bacterial inactivation, 100  $\mu$ l of the bacterial suspension was plated on blood agar plates.

### In vitro PBMC stimulation

We conducted scMultiome-seq analysis on PBMCs from individuals with LC across five time points: the acute phase; 3, 9 and 12 months after infection and during recovery. Heat-inactivated *P. aeruginosa* strain and a mock stimulation condition were tested over the course of four experimental runs. A pilot study was performed using samples from two healthy individuals, with cells stimulated for 4 and 24 h at four different concentrations to determine the optimal conditions. Based on this pilot study, the 4-h time point and  $2.5 \times 10^6$  c.f.u. per ml were identified as optimal.

For the main experiment, PBMCs were thawed according to an optimized protocol based on 10x Genomics guidelines (CG000365, Rev B). The cells were counted and resuspended at a concentration of  $5 \times 10^6$  cells per ml in warm RPMI. Cell suspension (100  $\mu$ l) was plated into a 96-well, round-bottom plate and rested for 1 h at 37 °C. Following this rest period, the RPMI medium was replaced with 100  $\mu$ l of heat-inactivated *P. aeruginosa* corresponding to a concentration of  $2.5 \times 10^6$  c.f.u. per ml. The cells were incubated at 37 °C for 4 h. After incubation, the plates were centrifuged at 300g for 5 min, and the cells were collected for nuclei isolation and library preparation.

### Isolation of nuclei and library preparation (cohort 1)

scMultiome-seq analysis was performed on both directly thawed and stimulated PBMCs. To manage sequencing costs, cells from three to four donors were pooled together. After pooling, the cells were treated with DNase I to remove free DNA and centrifuged at 300g for 10 min at 4 °C. The cell pellets were resuspended and incubated with 300  $\mu$ l of prechilled 1 $\times$  lysis buffer on ice for 3 min. Lysis was stopped by adding 1 ml of ice-cold wash buffer, followed by centrifugation at 500g for

5 min at 4 °C. The nuclei were washed twice with 500 µl of wash buffer and resuspended in Diluted Nuclei Buffer. To ensure purity and dissociation of single nuclei, the suspension was passed through a 40-µm Flowmi strainer and inspected under a microscope.

Approximately 20,000 nuclei were loaded into a Chromium Controller to produce single-cell gel beads, following the 10x Genomics Chromium Next GEM Single Cell Multiome ATAC + Gene Expression protocol (CG000338, Rev C). After transposition, the nuclei were treated with a transposase enzyme, which selectively fragmented the accessible DNA regions and added adapter sequences to the fragmented DNA ends. The transposed nuclei were loaded onto a Chromium Next GEM ChipJ (PN-1000234), alongside partitioning oil and barcoded gel beads. PCR amplification was performed, targeting approximately 10,000 nuclei per library. Sequencing was performed using the Illumina NovaSeq 6000 platform, with a minimum read depth of 20,000 read pairs per cell for scRNA libraries and 25,000 read pairs per cell for scATAC libraries.

Cell capture and library preparation (cohort 3)

We thawed cells following the 10x Genomics thawing protocol (CG00039, Rev D), cells from four donors were pooled together, and approximately 29,000 cells were loaded into the Chromium X (10x Genomics) to generate single-cell gel beads in emulsion according to the 10x Genomics protocol (CG000731, Rev A). scRNA-seq libraries were prepared using a Chromium GEM-X Single Cell 3' Reagent Kits v4 (10x Genomics) and sequenced on the NovaSeq 6000 platform (Illumina), with a minimum depth of 20,000 read pairs per cell.

Sample preparation for ex vivo flow cytometry experiments (cohort 4)

PBMCs were isolated from fresh whole blood using standard Ficoll density gradient centrifugation and cryopreserved in liquid nitrogen for deferred use. Ex vivo phenotyping of immune cells was performed from cryopreserved PBMCs. In brief, thawed PBMCs were stained with a Zombie NIR Fixable Viability kit (Biolegend, 423106) at room temperature in PBS for 15 min. Nonspecific immunolabeling conferred by Fc receptor binding was blocked by the addition of 10% Gamunex solution (Grifols Deutschland). Surface marker immunolabeling was performed in cell staining buffer (PBS, BSA and EDTA) (Biolegend, 420201) and Brilliant Stain Buffer (BD, 563794) overnight at 4 °C with antibodies to the ontogeny markers anti-human CD3, CD14, CD16 and HLA-DR; macrophage markers CD163 and CD206 and the markers identified from the transcriptomic analysis CD51, CD99, CD105, CD120b and HLA-DQ (see antibody details in the table). After fixation and permeabilization (BD, 554714) for 30 min at room temperature, immunolabeling of intracellular markers was performed for 30 min in Permash buffer (BD, 554714) at 4 °C with antibodies to CALR, IFNGR1, TGFB1 and IRF8. Next, cells were immunolabeled with the secondary antibodies AF488 and AF568 for 30 min in Permash buffer (BD, 554714) at 4 °C to label the unconjugated antibodies CALR and IFNGR1, respectively. All donors were also immunolabeled with the correspondent isotype controls for the used antibodies. Cells were washed with PBS, and data were acquired on a five-laser Sony spectral analyzer (ID7000, Sony) and analyzed with FlowJo software v10.10.0 (Tree Star).

Antibody list

Antibody	Fluorochrome	Clone	Company	Catalog
CD3	SparkBlue	SK7	BioLegend	344852
CD14	PacBlue	63D3	BioLegend	367122
CD16	BUV563	3G8	BD	568289
CD51	APC	NKI-M9	BioLegend	327913
CD99	PE	hec2	BioLegend	398205

Antibody	Fluorochrome	Clone	Company	Catalog
CD105	BUV421	43A3	BioLegend	323219
CD120b	PE-DAZZLE	3G702	BioLegend	358413
CD163	FITC/PE-Cy7	GHI/61	BioLegend	333618/2268070
CD206	APC-Cy7/PE-Cy7	15-2	BioLegend	321120/321124
CALR	Purified		Abcam	ab2907
	AF488		Invitrogen	
HLA-DQ	BB700	Tu169	BD	745976
HLA-DR	AF700	L243	BioLegend	307626
IFNGR1	Purified		Abcam	ab154400
	AF568		Invitrogen	
IRF8	PE	REA516	Miltenyi	130-122-927
TGFB1	PE-CF594	TW4-9E7	BD	562422

Statistical methods

No statistical method was used to predetermine sample size. The samples were randomized before processing for single-cell experiments. The investigators were not blinded to allocation during the experiments or during outcome assessment. All statistics in the manuscript are reported as specified in the figure legends.

Genotyping

Genotyping of DNA samples isolated from participants in the current study was performed using the GSA-MDv3 array (Infinium, Illumina) following the manufacturer's instructions. In total, 725,875 variants of 48 individuals were called by Optical 7.0 with default settings

Genotype processing for demultiplexing

Genotype data were reformatted into PLINK binary format files<sup>50</sup>. Quality control was performed at both the sample and single-nucleotide polymorphism (SNP) levels. Samples were excluded if they exhibited sex mismatches, missing genotyping rates of ≥0.05, heterozygosity rates beyond three standard deviations from the mean or relatedness across samples. A total of 58 samples passed these filters. SNPs were further filtered based on a minor allele frequency of >0.01 and an SNP missingness rate of <0.05. Genotype imputation was conducted using the Minimac4 server<sup>51</sup>, using the TOPMed r3 reference panel<sup>52</sup> and EAGLE v2.4 for phasing. The final analysis included a total of 6,050,031 variants.

Data preprocessing for multiome datasets and demultiplexing

BCL files from each library were converted to FASTQ files using cellranger-arc mkfastq with default parameters and using the respective sample sheet with the 10x barcodes. The 10x Genomics cellranger-arc count pipeline (v2.0.2) was used with default parameters using the human reference genome GRCh38-2020-A-2.0.0 obtained from 10x Genomics website. Demultiplexing was performed using Souporecell (v2.4)<sup>53</sup>. To assess the concordance between the genotypes of each donor in Souporecell-generated VCF and a reference VCF, BCFtools was used to perform a genotype check with parameter 'gtcheck' and the '-u GT' option to compare the genotype fields in the two VCF files.

Quality control and integration of multiome datasets (cohort 1)

Once the donor for each cell was assigned after demultiplexing, only single cells with both RNA and ATAC data were considered. Seurat version 5.0 (ref. 54) was used for downstream analysis. The following filtration criteria were used: 'nCount\_RNA < 6,000 and nCount\_ATAC < 15,000, mitochondrial percentage < 20, RNA features < 3,000, TSS enrichment > 1 and <10, while nucleosome\_signal < 2'. RNA integration across libraries was performed using 'RPCAIntegration' and the top 30 dimensions for both clustering and UMAP generation. Further,



multiple resolutions varying from 0.2 to 0.8 were performed to get clusters, and canonical markers were used for identifying cell subsets (using a combination of known markers and those used by Azimuth celltype. l2). For ATAC integration across libraries, signac version 1.13 (ref. 55) was used. Integration anchors were found by using 'rlsi' and 2–30 top dimensions, followed by integration using the top 30 dimensions.

### Participant and sample stratification and sample category classifications (cohorts 1, 2 and 3)

We first stratified individuals with LC into two groups based on their acute COVID-19 disease WHO scores, where LC samples from individuals with acute COVID-19 WHO scores between 1 and 5 were classified as LC<sup>AM</sup> and LC samples from individuals with acute COVID-19 WHO scores between 6 and 9 were classified as LC<sup>AS</sup>. Further samples were also stratified based on time points of collection resulting in NI, AI, T2: 1.5/1.7–2.9 months, T3: 3–5.9 months, T4: 6–8.9 months and T5: 9–12 months.

For cohort 1, to ensure that our findings were not convoluted by COVID-19 infection imprinting on immune cells and were unique to LC, we performed transcriptome comparisons either to AI samples or to R<sup>LC</sup>. All recovered samples were considered as one category. Consequently, within each group of participants (LC<sup>AM</sup> or LC<sup>AS</sup>), comparisons were performed as T2 versus AI, T3 versus AI, T4 versus AI or T2 versus R<sup>LC</sup>, T3 versus R<sup>LC</sup> and so on. In heat maps, the transcriptome signatures were plotted for all categories, including pre-pandemic healthy controls.

For cohort 3, the LC-MO signature was checked in CD14<sup>+</sup> monocytes of each individual with LC. Participants were further grouped into Resp-PASC-BHR (*N* = 3) and Resp-PASC (*N* = 5) groups based on their pulmonary function test results, as shown in Supplementary Table 3.

### DGE analysis per cell type

For each of the major cell types (CD14<sup>+</sup> monocytes, CD16<sup>+</sup> monocytes, CD8<sup>+</sup> T cells, NK cells, CD4<sup>+</sup> T cells and B cells), DGE analysis was performed for LC samples separately for mild/moderate and severe samples. Comparisons were made against acute COVID-19 samples and against recovered samples using Seurat FindMarkers. Genes upregulated and downregulated in these comparisons with an adjusted *P* value of <0.05 and log<sub>2</sub> (fold change) of >0.8 were considered for each cell-type analysis.

### Pathway analysis per cell type

Pseudobulk of each donor at each time point was calculated, followed by similar comparisons as described in the previous section using DESeq2. GSEA using Hallmark and REACTOME pathways as background was performed using the fgsea R package. Furthermore, immune-related pathways that showed statistical significance in any comparisons were plotted. The whole list of statistically significant pathways resulting from all comparisons is shown in Source Data Fig. 2a and Extended Data Fig. 2a,b.

### Pathway correlations with clinical parameters

For each sample and cell type (CD14<sup>+</sup> monocytes, CD8<sup>+</sup> T cells and NK cells), we computed pseudobulk gene expression profiles. Subsequently, AUC scores for the selected upregulated immune pathways, as described in Fig. 2c, were calculated in each sample. The AUC scores of these pathways were then correlated with clinical parameters using Spearman correlations. Only significant correlations were plotted.

### Cytokine data analysis

Cytokine measurements for each measured cytokine were log<sub>2</sub> transformed. Comparisons and statistical tests against COVID-naïve healthy controls for each measured cytokine were performed. The Spearman correlation test was performed to assess the correlation between transformed cytokine measurements and clinical parameters.

### Subclustering analysis of CD8<sup>+</sup> T cells, NK cells and monocytes

CD8<sup>+</sup> T cells, NK cells and CD14<sup>+</sup> monocytes were subsetted separately and reanalyzed. Libraries contributing less than 60 cells were removed, and integration was performed using 'RPCAIntegration' and k.weight as 60. The top 10 principal components were used for UMAP and FindNeighbors calculation. For clustering CD8<sup>+</sup> T and NK cells, a resolution of 0.4 was used. For CD14<sup>+</sup> monocytes, a resolution of 0.2 was used. An AUC score for Hallmark pathways enriched in pseudobulk analysis was calculated for each cell using the AUCCell R package and raw counts of each cell.

### Neighborhood enrichment analysis

MiloR<sup>56</sup> was used for differential neighborhood analysis. The kNNGraph and neighborhoods were calculated with *k* = 50 and *d* = 50. The design matrix included the sampleID, severity, timePoint and recovered or not as covariates. Differential neighborhood tests were calculated for LC<sup>AM</sup> samples (T2, T3, T4 and T5) from different time points against acute COVID-19 samples (AI) or against R<sup>LC</sup>. The resulting differential neighborhoods were annotated based on cell clusters previously obtained for each cell subset. Neighborhoods with a spatial FDR of <0.1 were considered significant.

### Pseudotime and trajectory analysis

A Seurat RNA assay of CD14<sup>+</sup> monocytes was used to create a singlecell experiment object using scater. Diffusion maps were calculated using the destiny R package<sup>57</sup>. Average dimensionality was calculated using the find\_sigmas function with logCounts of single-cell data. DiffusionMap was calculated using 40 principal components and sigmas calculated in the previous step. The top diffusion components (DCs) were inspected, and DC1 and DC3 were used because DC2 showed sample-dependant bias. Slingshot<sup>58</sup> was used for trajectory calculations. Clustering was calculated using the top 15 DCs and the Mclust package. Clusters with >90% of cells from COVID-naïve healthy controls were chosen as the starting clusters for trajectory calculation, resulting in three lineages. The expression of genes involved in key pathways (from pathway enrichment analysis) and selected upregulated genes from the MC4 cluster were plotted against pseudotime values of each cell. Similarly, AUC scores of key pathways calculated per cell were plotted against pseudotime.

### Peak calling and peak-to-gene linkage

Peaks were called for each major cell subset as identified from RNA-based annotations using Macs3 and Ensembl.Db.Hsapiens.v86. Peaks were linked to RNA assay-based gene expression using the LinkPeaks command. Differential peaks within each cluster were calculated using Seurat function FindMarkers with the 'LR' test and nCount\_peaks as the latent variable.

### Transcription factor motif annotation and enrichment

The Jasp2020 database was used as background for the motif matrix using only the human-specific motif collection. Chromvar was used to calculate transcription factor activity for each cell. Differential transcription factor motif activity for any comparison was calculated using FindMarkers with mean.fxn set to 'rowMeans'. Motif enrichment was assessed by correcting for background peaks using MatchRegionStats.

### Analysis of scRNA-seq datasets (cohort 3)

Once the donor for each cell was assigned after demultiplexing, the doublets were filtered out, and singlets were kept. Seurat version 5.0 was used for downstream analysis. The following filtration criteria were used: 'nCount\_RNA < 8,000 and nFeature\_RNA < 3,500 and mitochondrial percentage < 20'. RNA integration across libraries was performed using 'RPCAIntegration' and the top 30 dimensions for both clustering and UMAP generation. Multiple resolutions varying from 0.2 to 0.5 were carried out to obtain clusters, and canonical markers were used



for identifying cell subsets (using a combination of known markers and those used by Azimuth celltype.l2) to identify CD14<sup>+</sup> monocytes. Further, these cells were subsetted and reintegrated with 15 principal components, and clustering was performed with a resolution of 0.2. LC-Mo/MC4 AUC scores were calculated for each cell using raw counts.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Single-cell multiome data were submitted to European Genome-Phenome Archive and are accessible through the following IDs: [EGAS00000000142](#), [EGAS00000000143](#), [EGAS00000001215](#) and [EGAS00000001216](#). Source data are provided with this paper.

### Code availability

Scripts and code are available at [github.com/CiiM-Bioinformatics-group/LongCOVID](https://github.com/CiiM-Bioinformatics-group/LongCOVID).

### References

49. Kopfnagel, V. et al. The Hannover Unified Biobank (HUB)—centralized standardised biobanking at Hannover Medical School. *Open J. Bioresour.* **8**, 2 (2021).
50. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
51. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).
52. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
53. Heaton, H. et al. Souporecell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).
54. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
55. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
56. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using *k*-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022).
57. Angerer, P. et al. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016).
58. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

### Acknowledgements

This project was supported by an ERC Starting Grant 948207 (ModVaccine) to Y.L., the COVID-19-Research Network of the state of Lower Saxony (COFONI) through funding from the Ministry of Science

and Culture of Lower Saxony in Germany (14-76403-184) to Y.L. and T.I., the Lower Saxony Center for AI and Causal Methods in Medicine (CAIMed) grant (ZN4257) and German Federal Ministry of Education and Research grants (01EQ2302A, FEDCOV, 031L0318A, AID-PAIS) to Y.L. and US National Institutes of Health grants AGO69264, AI147394, HL170961, AI176171 and AGO90337 to J.S. S.A.K. was supported by the PRACTIS Clinician Scientist Program, funded by MHH and DFG (DFG ME 3696/3), and by funding from Julitta und Richard Müller Stiftung (COVIDCODE). The COVID-19 biobank of MHH was funded by the Lower Saxony Ministry of Science and Culture.

### Author contributions

Y.L., T.I. and S.K. conceived and designed the study. I.P., S.V., D.v.W., H.L., S.A.K., I.K., A.R.K. and M.M.H. acquired clinical samples and collected clinical data. L.Z., A.A. and A.Z. generated data. S.K. and Q.Z. performed data analysis and investigation. S.K., Y.L. and S.V. coordinated project administration. B.C., A.R.M.K., H.S., D.v.W., H.L., L.P. and C.L. performed experiments. Y.L., T.I., J.S., S.H., A.R.M.K., M.W., M.C. and C.-J.X. provided resources. Y.L., M.G.N. and J.S. supervised the study. Y.L. and T.I. acquired funding. S.K. and Y.L. wrote the original paper. All authors reviewed and approved the final paper.

### Funding

Open access funding provided by Helmholtz-Zentrum für Infektionsforschung GmbH (HZI).

### Competing interests

M.G.N. is the scientific founder of Biotrip, Salvina, TTxD and Lemba. MHH has received fees for consultations or lectures from 35Pharma, Acceleron, Actelion, Aerovate, AOP Health, Bayer, Ferrer, Gossamer, Inhibikase, Janssen, Keros, MSD and Novartis. The other authors declare no competing interests.

### Additional information

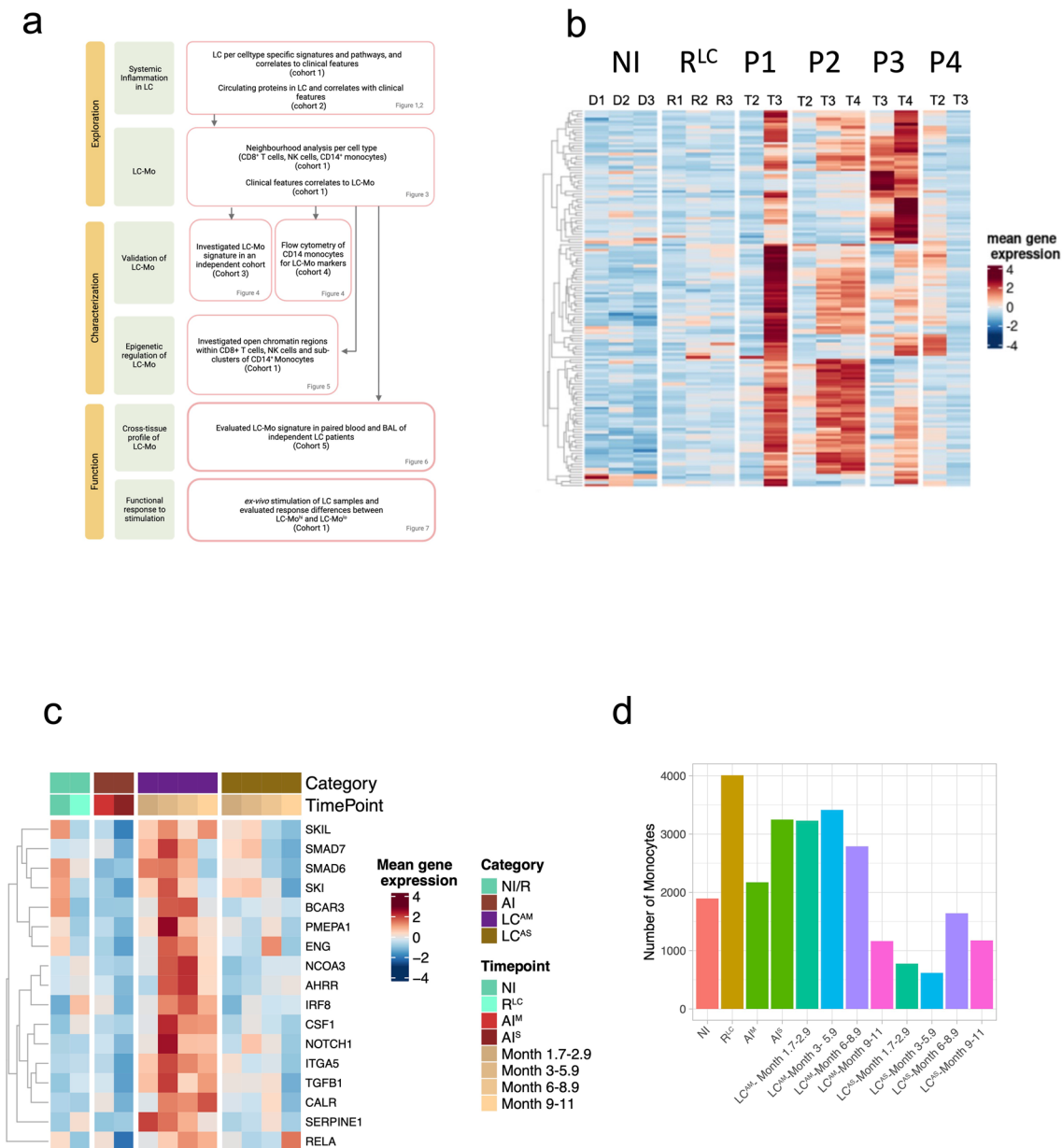
**Extended data** is available for this paper at <https://doi.org/10.1038/s41590-025-02387-1>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41590-025-02387-1>.

**Correspondence and requests for materials** should be addressed to Yang Li.

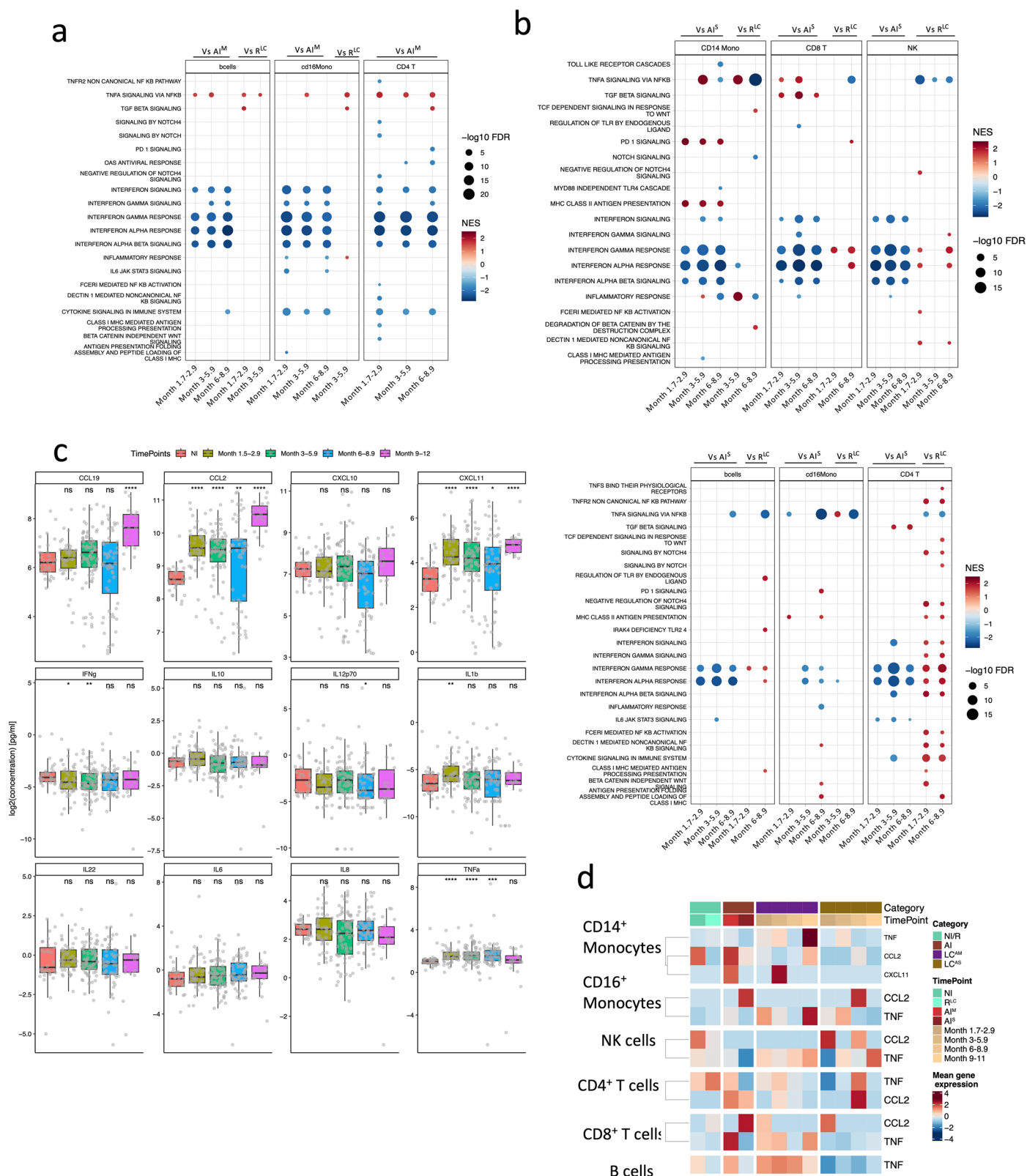
**Peer review information** *Nature Immunology* thanks Ryan Thwaites, Onur Boyman and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Ioana Staicu, in collaboration with the *Nature Immunology* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Extended Data Fig. 1 | LC patients show heterogeneous expression of LC signature.** **a**, Schematic view of analysis flow for the study. **b**, Mean expression of LC signature genes in LC<sup>AM</sup> (n = 4) participants with longitudinal time points

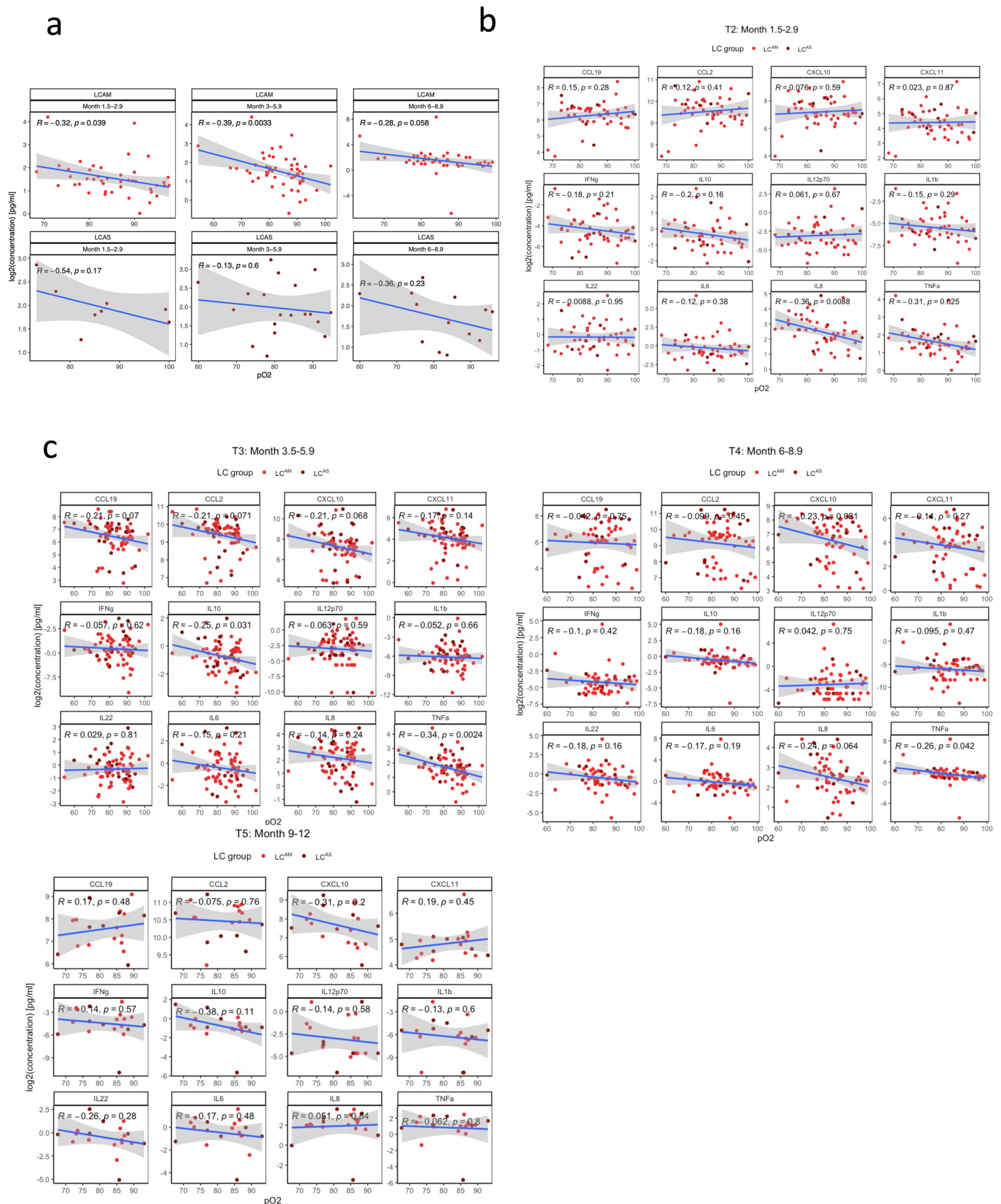
and NI (n = 3) and R<sup>LC</sup> (n = 3) (cohort 1). **c**, Mean expression of LC signature genes showing consistent expression at all timepoints **d**, Number of CD14<sup>+</sup> monocytes in each category at all time points.



### Extended Data Fig. 2 | $LC^{AI^M}$ and $LC^{AI^S}$ show different pathways active in LC.

**a**, Significant GSEA pathways in  $LC^{AI^M}$  at different time points compared to  $AI^M$  or  $R^{LC}$ , in B cells,  $CD4^+$  T and  $CD16^+$  monocytes **b**, GSEA pathways comparing  $LC^{AI^S}$  at different time points from for all major cell subsets (top and bottom) (cohort 1). Pathways plotted with adj Pval < 0.1 (Kolmogorov-Smirnov-based test with permutation-derived p-values, adjusted using the Benjamini-Hochberg method). **c**, comparing measured cytokine levels (CCL19, IFN $\gamma$ , IL10, IL12p70, IL1b, IL22,

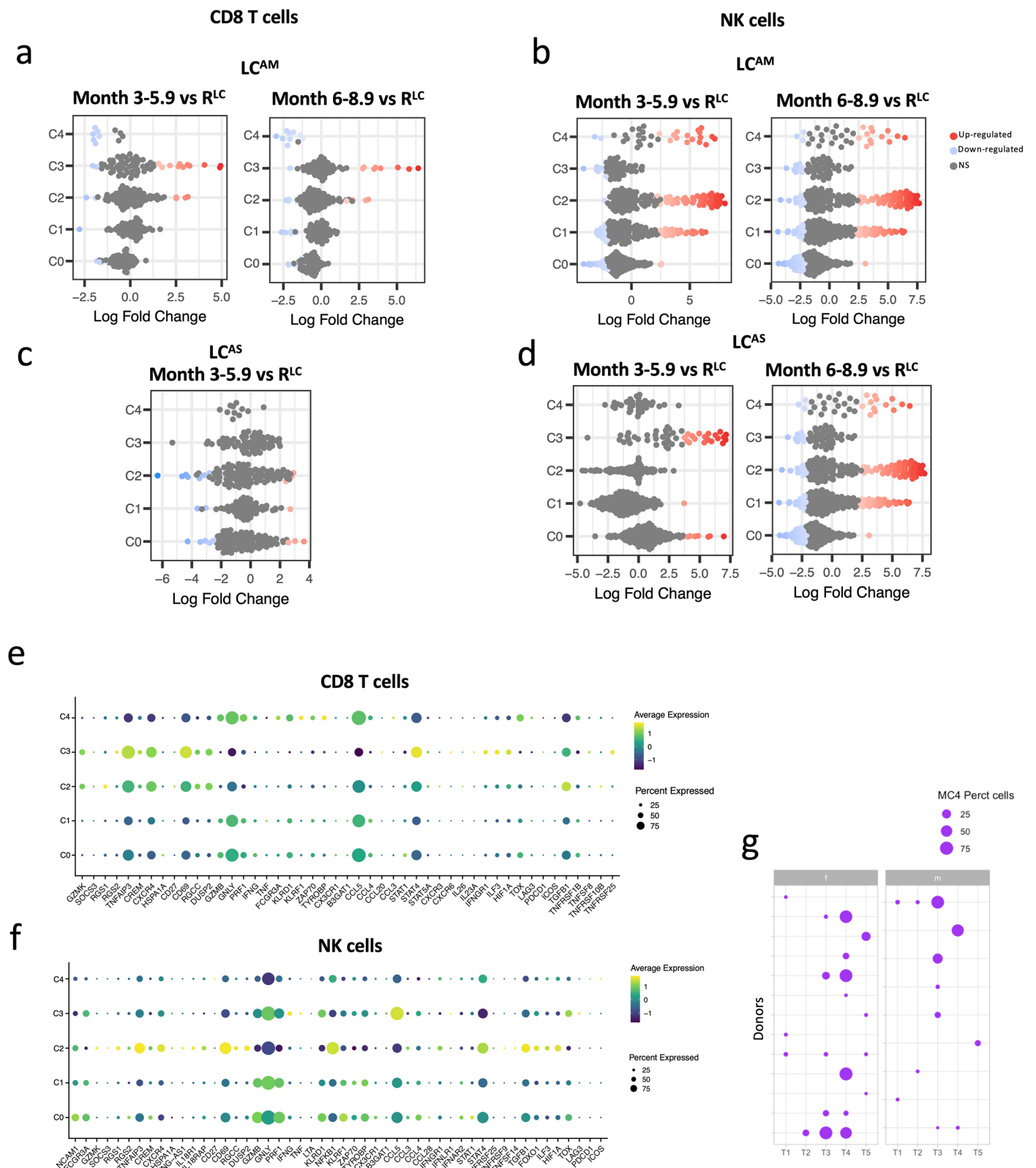
IL6, IL8, IP10) in LC patients at different time points with NI (n = 33), LC: month 1.5-2.9 (n = 51), month 3-5.9 (n = 75) month 6-8.9 (n = 60) month 9-12 (n = 19) (cohort 2). Two-sided Wilcoxon Rank Sum Test \*\*\*\*: p value < 0.00001, \*\*\*: p value < 0.001, \*\*: p value < 0.01, \*: p value < 0.05, ns = not significant. Boxplot shows the median (centre), first and third quartiles (bounds) and 1.5 times the interquartile range (whiskers) **d**, Mean mRNA expression of TNF, CCL2 and CXCL11 across all categories (cohort 1).



**Extended Data Fig. 3 | TNF significantly negatively correlates with pO<sub>2</sub> in LC<sup>AM</sup>.** **a**, TNF correlation with partial pressure O<sub>2</sub> calculated separately for LC<sup>AM</sup> and LC<sup>AS</sup> (cohort 2) at month 1.5-2.9 (LC<sup>AM</sup> n = 43, LC<sup>AS</sup> n = 8), month 3-5.9 (LC<sup>AM</sup> n = 56, LC<sup>AS</sup> n = 19) and month 6-8.9 (LC<sup>AM</sup> n = 47, LC<sup>AS</sup> n = 13) **b** and **c** Correlation of all measured cytokines as for each time point against pO<sub>2</sub>. LC: month 1.5-2.9

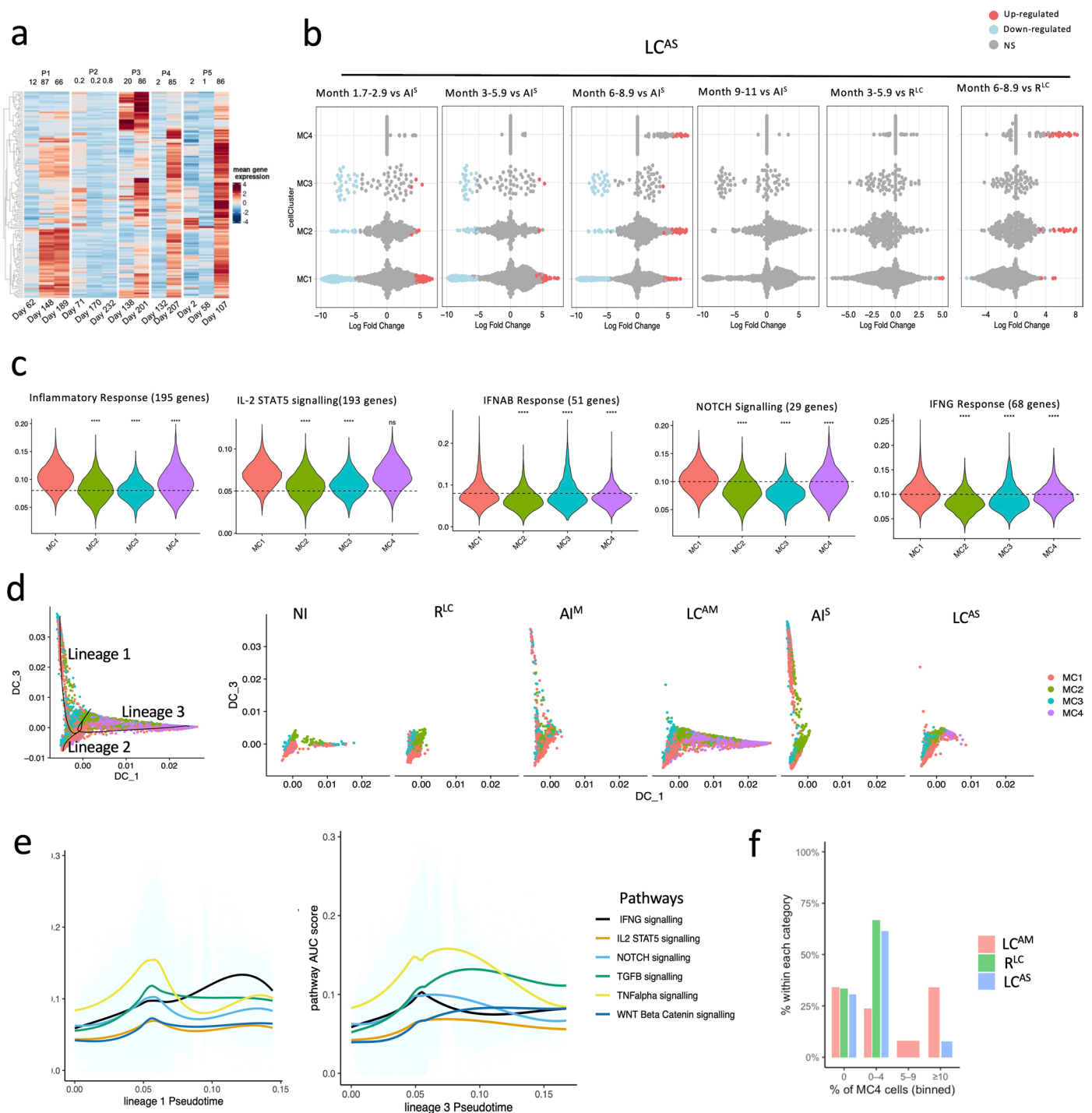
(n = 51), month 3-5.9 (n = 75) month 6-8.9 (n = 60) month 9-12 (n = 19). Spearman correlation P values for Spearman correlation were computed using the exact/permutation-based test. The gray shaded area indicates the 95% confidence interval.





**Extended Data Fig. 4 | LC<sup>AM</sup> CD8<sup>+</sup> T and NK cells show increased abundance of GZMK<sup>+</sup> cells.** **a, b**, Neighbourhood enrichment analysis in LC<sup>AM</sup> (cohort 1) compared to R<sup>LC</sup> in CD8<sup>+</sup> T and **b**, in NK cells. **c, d**, Neighbourhood abundance enrichment analysis in LC<sup>AS</sup> compared to R<sup>LC</sup> in CD8<sup>+</sup> T and **d**, in NK cells. Red dots represent increased significant enrichment of neighbourhoods; blue dots represent significantly decreased enrichment. Transcriptional states showing

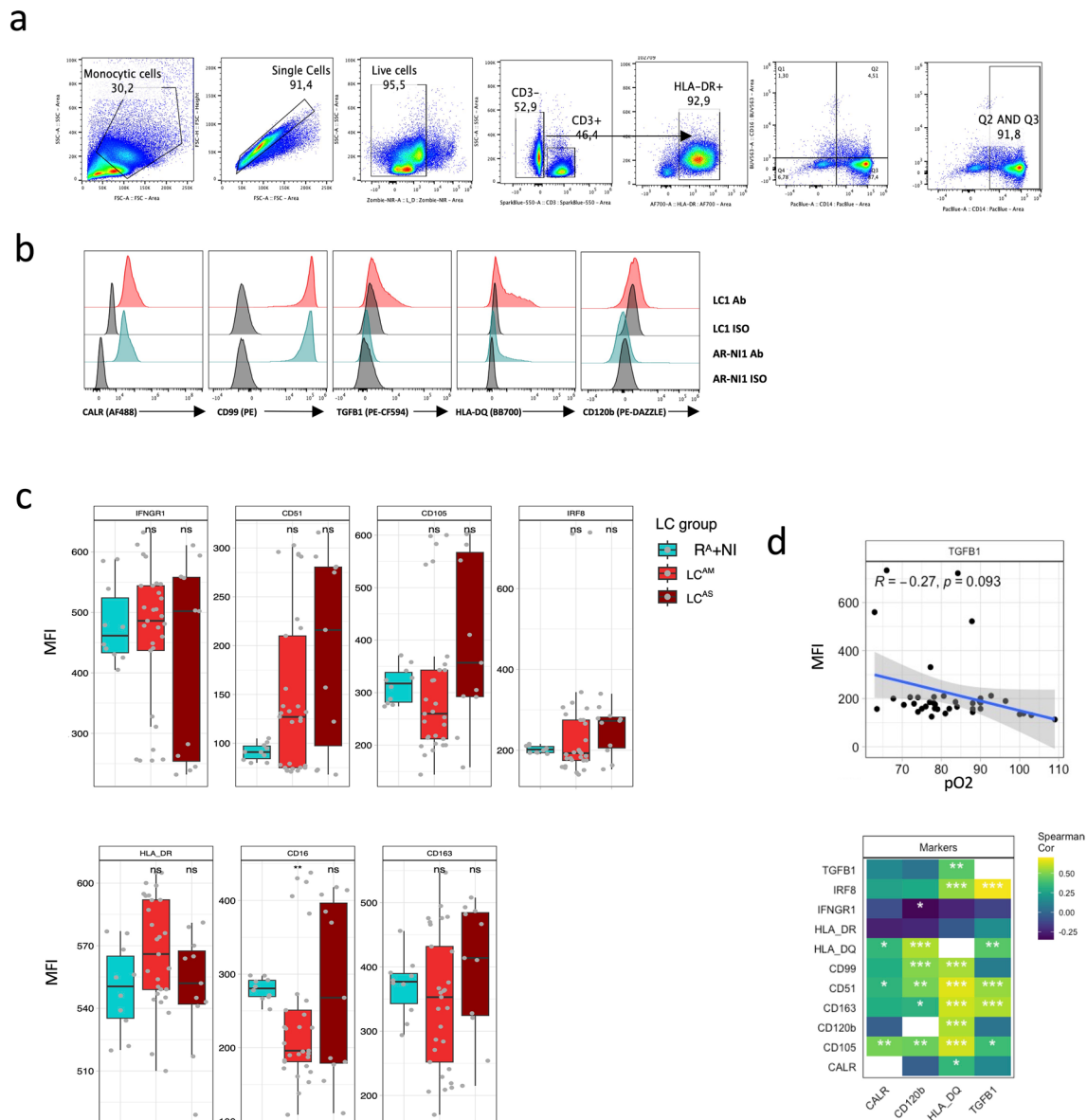
significant enrichment with spatial FDR < 0.1 in red and blue (F-test statistic from the quasi-likelihood F-test, graph weighted FDR). **e**, GZMK<sup>+</sup> cells signature derived from *Jonsson et al.*<sup>28</sup> shown in gene expression profile within CD8<sup>+</sup> T sub-clusters, **f**, within NK sub-clusters. **g**, Proportion of MC4 cells in CD14<sup>+</sup> monocytes of LC<sup>AM</sup> samples from multiple time points stratified on sex.



### Extended Data Fig. 5 | LC-Mo show a distinct transcriptional program.

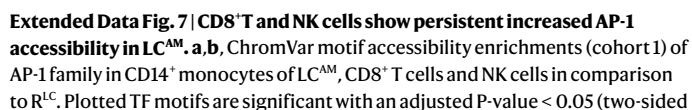
**a**, Mean gene expression of MC4 signature in CD14<sup>+</sup> monocytes of LC<sup>AM</sup> (n = 5) patients with longitudinal samples in cohort 1, with sampling day of each individual and percentage of MC4-like cells in each sample **b**, Differential neighbourhood enrichment in CD14<sup>+</sup> monocyte of LC<sup>AS</sup> compared to AI<sup>S</sup> and R<sup>LC</sup>. Red dots represent increased significant enrichment of neighbourhoods; blue dots represent significantly decreased enrichment. Transcriptional states showing significant enrichment with spatial FDR < 0.1 (F-test statistic from

the quasi-likelihood F-test, graph weighted FDR) **c**, Pathway AUC scores within monocyte clusters. Two-sided Wilcoxon Rank Sum Test \*\*\*\*: pvalue < 0.00001, \*\*\*: pvalue < 0.001, \*\*: pvalue < 0.01, \*: pvalue < 0.05, NS = Non-significant **d-e**, Distinct trajectories of CD14<sup>+</sup> monocytes derived from pseudotime **d**, Diffusion map of all CD14<sup>+</sup> monocytes with predicted lineages (left) cells split on categories of samples (right) **e**, AUC scores of pathways arranged across lineage 1 and lineage 3. **f**, Percentage of MC4 cells within CD14<sup>+</sup> monocytes, categorised and binned.



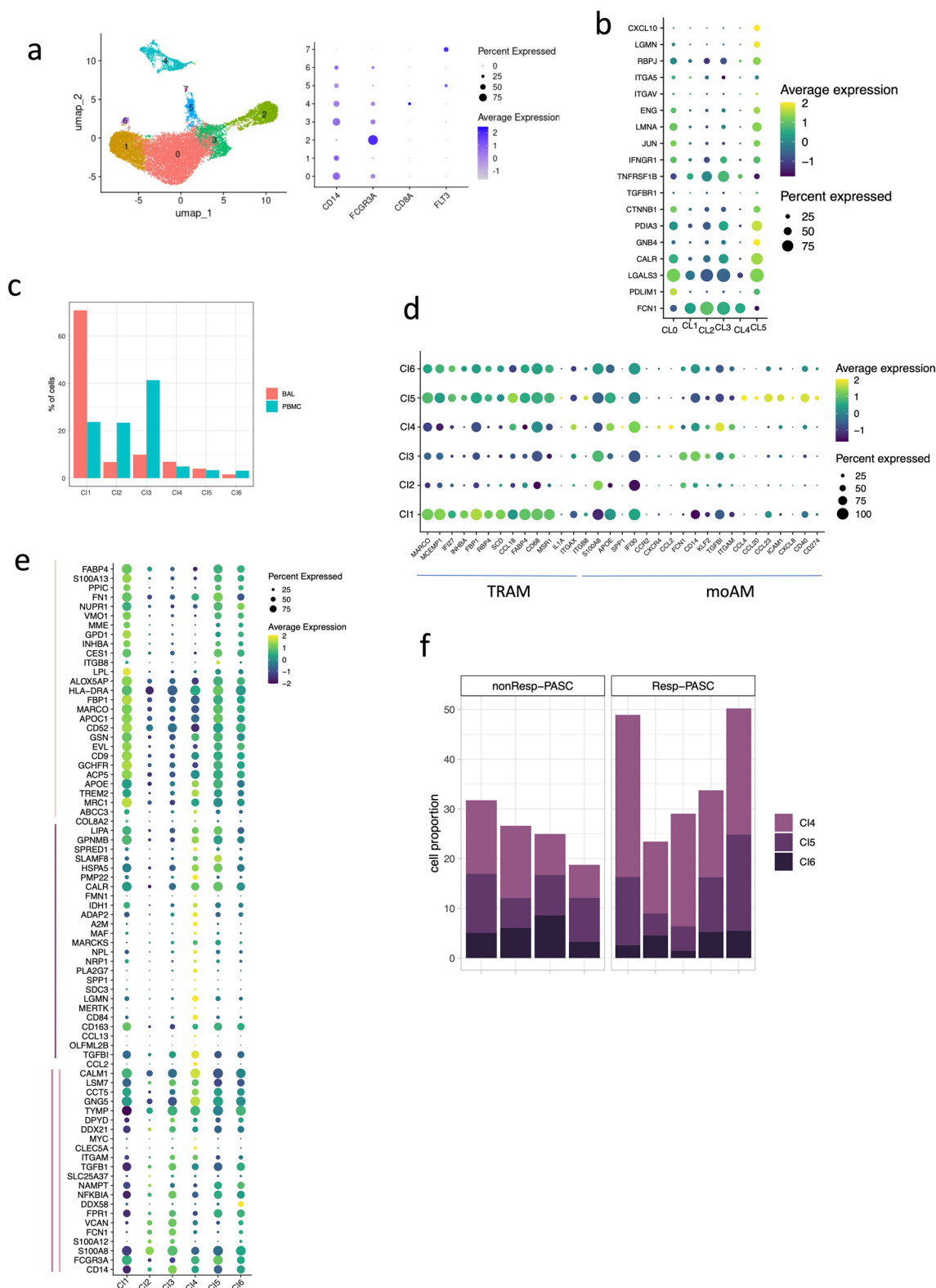
**Extended Data Fig. 6 | Flow cytometry of CD14<sup>+</sup> monocytes shows higher expression of LC-Mo in LC patients. **a**, Gating strategy implemented to identify Monocyte subsets in cohort 4. **b**, Histograms for the flow cytometry analysis of CALR, CD99, TGFβ1, HLA-DQ, and CD120b surface expression enriched for all CD14<sup>+</sup> monocytes from R<sup>A</sup>+NI compared to LC. R<sup>A</sup>+NI, LC antibody, and isotype-stained cells are shown in blue, red, and black, respectively. Shown is a representative donor of biological replicates with similar results **c**, Comparison of Median Fluorescence Intensity (MFI) of measured markers (IFNGR1, CD51, CD105, IRF8, HLA-DR, CD16 and CD163) in CD14<sup>+</sup> monocytes. R<sup>A</sup>+NI (n = 10),**

LC<sup>AM</sup> (n = 29), LC<sup>AS</sup> (n = 11). P-values calculated using two-sided Wilcoxon Rank Sum Test \*\*\*: p value < 0.00001, \*\*: p value < 0.001, \*: p value < 0.01, \*: p value < 0.05. Boxplot shows the median (centre), first and third quartiles (bounds) and 1.5 times the interquartile range (whiskers) **d**, TGFβ1 MFI correlation with blood pO<sub>2</sub> (top), the gray shaded area indicates the 95% confidence interval. correlation of significant markers in LC patients among each other (bottom). Spearman correlation P values were computed using a t-distribution approximation applied to rank-transformed data. \*\*\*: p value < 0.001, \*\*: p value < 0.01, \*: p value < 0.05.



Nature Immunology

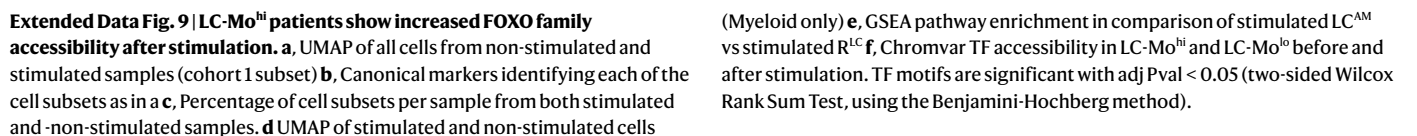




**Extended Data Fig. 8 | LC-Mo like macrophages show pro-fibrotic signature.**

**a**, Blood (cohort 5) CD14<sup>+</sup>, CD16<sup>+</sup> Monocytes and monocyte derived DCs from independent cohort (GSE263817)<sup>23</sup> **b**, Gene expression profile of clusters showing LC-Mo/MC4 like genes in cluster5. **(c)** Proportion of blood and BAL Monocyte/Macrophage cells in each of the cluster as described in Fig. 6d. **d**, Gene expression

as described in (GSE263817)<sup>23</sup> for Tissue resident Alveolar Macrophages (TRAM) and monocyte derived Macrophages (moAM) **e**, Gene expression profile derived from Wendisch et al.<sup>33</sup> of C11-C16 clusters as shown in Fig. 6d **f**, Proportion of cells in either C14, C15 or C16 clusters from each donor classified into PASC categories.



Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted<br><i>Give P values as exact values whenever suitable.</i>                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated   |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	A detailed description of the data and sample collection can be found in the methods section of the manuscript. EDTA blood was collected from patients which were admitted to the hospital due to COVID-19, as well as from ambulatory patients after SARS-COV-2-infection showing Post-COVID19 symptoms (at least 4 weeks after acute COVID-19) at the Hannover Medical school (MHH) or the Siloah hospital. PBMCs were isolated from whole blood using Ficoll gradient centrifugation. 10x Genomics Chromium Next GEM Single Cell Multiome ATAC + Gene Expression protocol and single cell GEM-X 3` Reagent Kits v4 were used to generate single cell multiome data. Sequencing was performed using the Illumina NovaSeq 6000 platform. The Quanterix HD SP-X Imaging and Analysis System™ was used to measure the plasma samples. Human Corplex cytokine panel 1 10-Plex array was used.
Data analysis	For data analysis, we employed the statistical language R(version > 4).A detailed description of the analysis can be found in the methods section of the manuscript. All code for the analysis is made freely available on <a href="#">github.com/CiiM-Bioinformatics-group/LongCOVID</a> .

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Single cell multi-ome data and scRNA-seq data was submitted to EGA and is accessible through following IDs:  
EGAS00000000142, EGAS00000000143, EGAS00000001215 and EGAS00000001216

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Our study included patients of both sexes and therefore applies to both sexes. Sexes were self-reported and later confirmed by DNA genotyping. Single cell dataset from both sexes were equally represented in the datasets.
Reporting on race, ethnicity, or other socially relevant groupings	<i>Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.) Please provide details about how you controlled for confounding variables in your analyses.</i>
Population characteristics	The only relevant population characteristic is the past history of COVID-19 infection. All patients were tested PCR negative at the time of visit and sample collection.
Recruitment	Patients reporting to the pneumological outpatient clinic at Hannover Medical School (MHH, Hannover, Germany) with symptoms such as headaches, dyspnoea and fatigue were recruited based on established LC criteria, 1) Symptoms that persist after acute COVID-19 or its treatment. 2) New symptoms that appear after the end of the acute COVID-19 phase, and can be a consequence of the SARS-CoV-2 infection and 3) Worsening of a pre-existing illness as a result of a SAR-CoV-2 infection. The diagnosis was made in accordance with the German S1 guidelines and the Delphi Consensus Criteria of post COVID-19. The cohort included individuals with all severity of acute COVID-19 (WHO score 1-9).
Ethics oversight	The ethics committee of the Hannover Medical School (MHH) approved the sample collection and analyses (ethics vote 9001_BO_K). Informed consent was obtained from all individual participants included in the study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples were chosen to balance the age, gender and acute COVID-19 severity for samples collected for different time points. Where possible, longitudinal samples from same patients were used.
Data exclusions	Single cells with low quality, high mitochondrial gene expression and doublets were excluded from the data analysis of single cell multiome analysis of discovery cohort.
Replication	Findings from single cell data of cohort 1 were validated with findings from another single cell study of independent cohort 3 and with publicly available datasets.
Randomization	Samples were randomized to include random 4 different donors for each pool during 10X experiments. Additionally, sample timepoint was randomised along with different donors before 10X experiments.
Blinding	The investigators performed the stratification of the long covid samples to identify differences in long covid disease. Therefore, blinding to group allocations was not possible for the further analysis.



# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used

For Human:  
 CD3 SparkBlue SK7 BioLegend 344852  
 CD14 PacBlue 63D3 BioLegend 367122  
 CD16 BUV563 3G8 BD 568289  
 CD51 APC NKI-M9 BioLegend 327913  
 CD99 PE hec2 BioLegend 398205  
 CD105 BUV421 43A3 BioLegend 323219  
 CD120b PE-DAZZLE 3G702 BioLegend 358413  
 CD163 FITC/PE-CY7 GHI/61 BioLegend 333618/2268070  
 CD206 APC-CY7/PE-CY7 15-2 BioLegend 321120/321124  
 CALR purified abcam ab2907  
 AF488 invitrogen  
 HLA-DQ BB700 Tu169 BD 745976  
 HLA-DR AF700 L243 BioLegend 307626  
 IFNGR1 purified ab154400 ab154400  
 AF568 invitrogen  
 IRF8 PE REA516 Miltenyi 130-122-927  
 TGFB1 PE-CF594 TW4-9E7 BD 562422

Validation

All used antibodies are validated/ quality control tested for the analysis of human cells by flow cytometry according to the manufacturer's information.

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.

## Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.

## Flow Cytometry

### Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

Sample preparation	Cryopreserved PBMCs were thawed, stained with the Zombie NIR™ Fixable Viability Kit (Biolegend) at room temperature (RT) in PBS for 15min. Unspecific immunolabeling conferred by Fc receptor binding was blocked by the addition of 10% Gamunex solution (Grifols Deutschland GmbH, Frankfurt am Main, Germany). Surface marker immunolabeling was performed in cell staining buffer (PBS, BSA, EDTA) and Brilliant Stain Buffer (BD), overnight at 4 °C with anti-human CD3, CD14, CD16, HLA-DR, CD163 and CD206, CD51, CD99, CD105, CD120b and HLA-DQ antibodies. For intracellular staining, cells were fixed and permeabilized (BD) for 30min at RT, immunolabeling of intracellular markers was performed for 30min in Permash buffer (BD) at 4°C with anti-human primary CALR, IFNGR1, TGFβ1 and IRF8 antibodies. Next cells were immunolabeled with the secondary antibody AF488 and AF568 for 30min in Permash buffer (BD) at 4°C to label the unconjugated antibodies CALR and IFNGR1, respectively. All donors were also immunolabeled with the correspondent isotype controls for the used antibodies. Cells were washed with PBS.
Instrument	The samples were acquired on a Sony spectral analyzer (ID7000, Sony).
Software	The samples were analyzed with the FlowJo software (10.10.0 Tree Star).
Cell population abundance	Monocytes correspond to roughly 5-10% of the PBMC samples, with a loss of 30% during freeze-thawing process. Purity of monocytes was determined by the expression of CD14+, CD16+ and HLA-DR surface markers.
Gating strategy	SSC-A/FSC-A gate to select monocytic cells -> FSC-H/FSC-A gate to select single cells -> Live/Dead gate to select viable cells -> CD3- gate to exclude T cells -> HLA-DR gate to exclude NK cells -> CD14+/CD16+ gate to select classical (CD14+), non-classical (CD16+), and intermediate (CD14+/CD16+) (see also Supplementary Fig. 5c).

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.