Cell Reports Medicine, Volume *2*

Supplemental Information

A distinct innate immune signature marks progression from mild to severe COVID-19

Stéphane Chevrier, Yves Zurbuchen, Carlo Cervia, Sarah Adamo, Miro E. Raeber, Natalie de Souza, Sujana Sivapatham, Andrea Jacobs, Esther Bachli, Alain Rudiger, Melina Stüssi-Helbling, Lars C. Huber, Dominik J. Schaer, Jakob Nilsson, Onur Boyman, and Bernd Bodenmiller

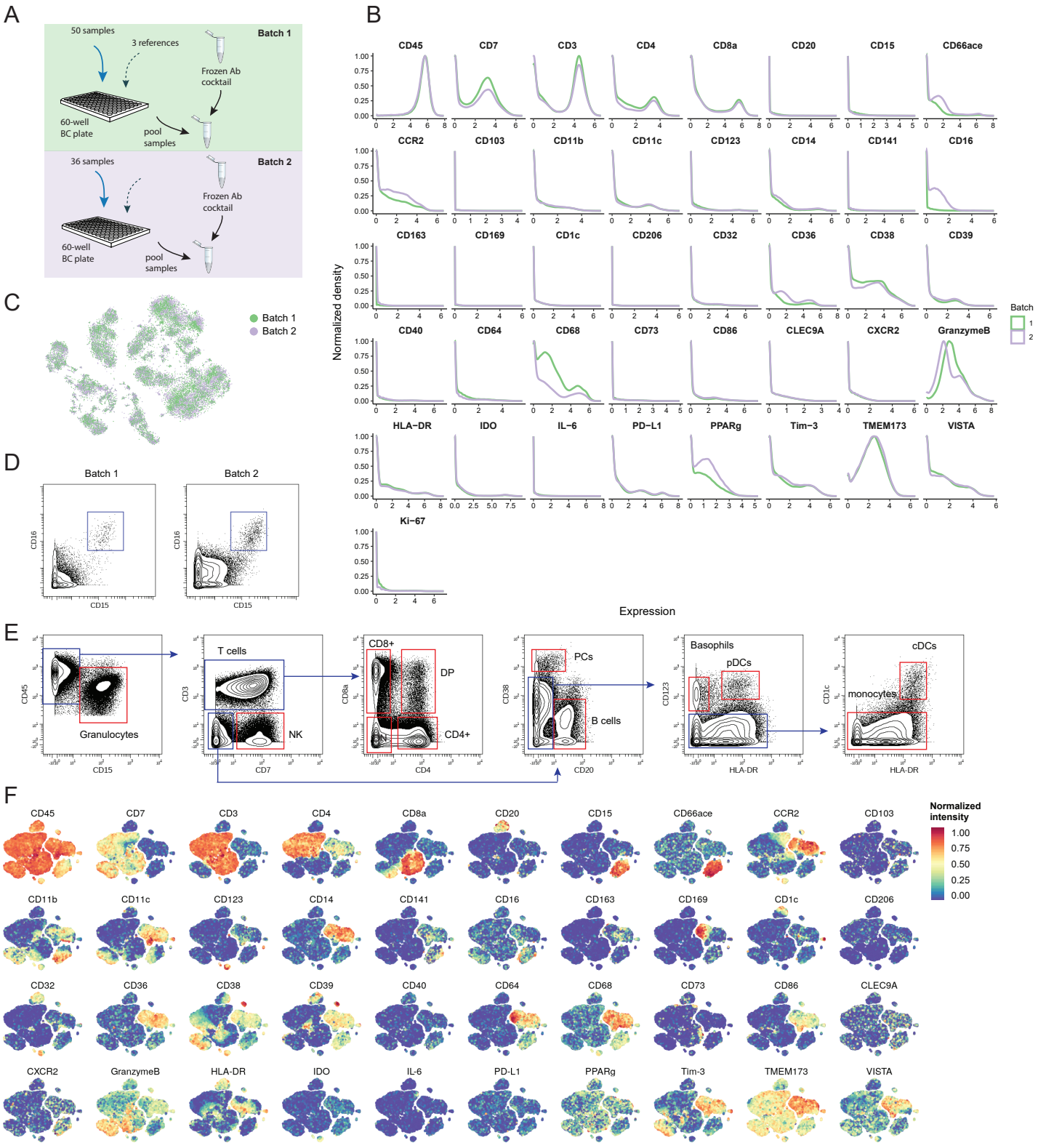

Figure S1: Experimental design for mass cytometry data acquisition in two independent batches and quality assessment. Related to Figure 1.

(A) Schematic of the strategy used to collect mass cytometry data. Samples were analyzed in two batches. The indicated number of samples were barcoded and stained with the same antibody mix stored as frozen aliquots. The staining and the acquisition on the CyTOF were performed independently.

(B) Histograms of signal intensities for each marker included in the panel observed on the pooled references acquired in the two runs. Data were linearly scaled based on the 98th percentile. Shifts in background intensity were observed for CD16, CD66ace, and Granzyme B.

(C) t-SNE plot calculated for the three reference (naive PMC, LPS activated PBMC and PHA activated PBMC) samples included in each CyTOF run based on all markers included in the study, colored by batch.

(D) Dotplot of signal intensities for CD16 versus CD15 in reference sample.

(E) Manual gating strategy used to identify the main cell types used to train the random forest cell classifier.

(F) t-SNE plot as in Figure 1D of a random subset of 1'000 immune cells from each sample colored by normalized marker intensity (n=78 individuals).

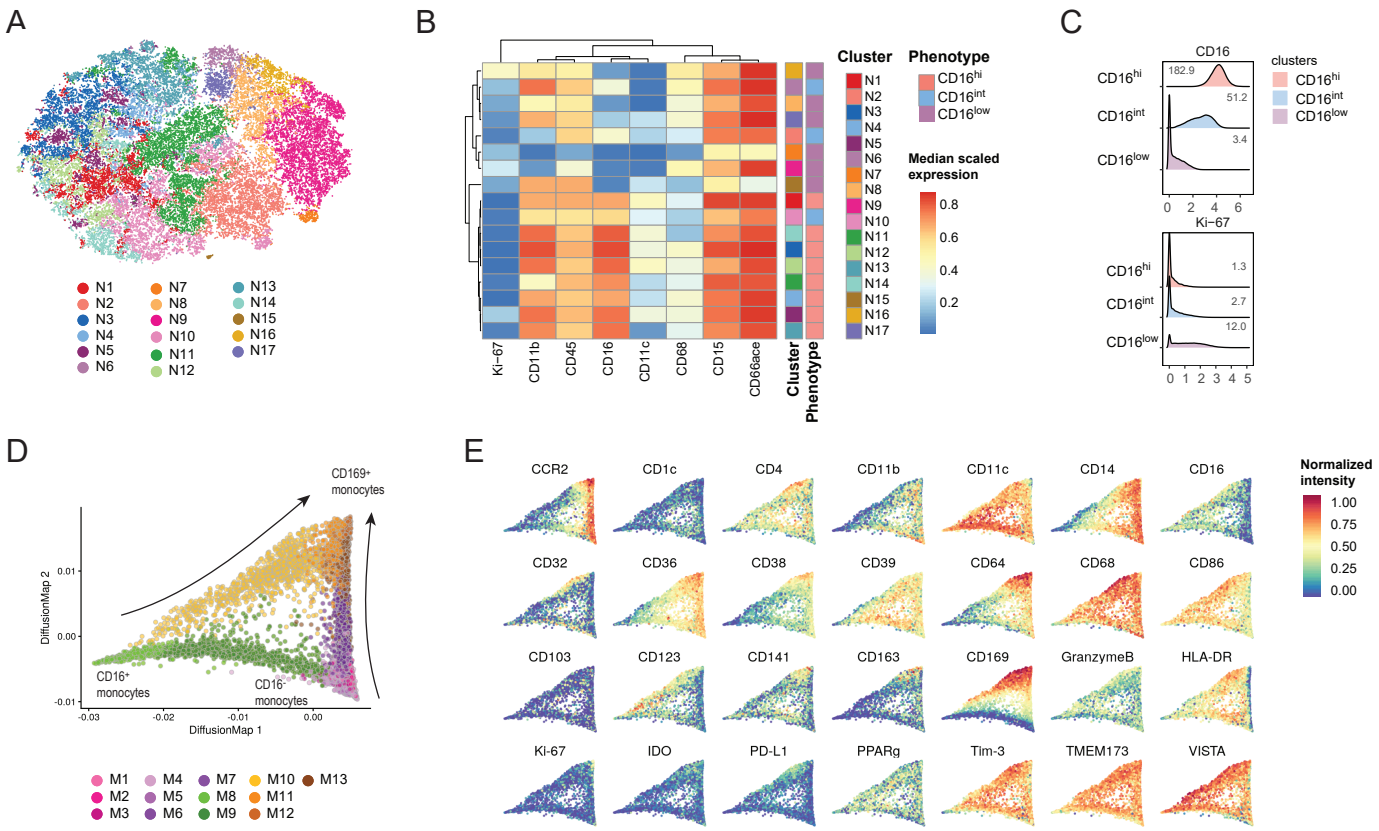

Figure S2: In-depth characterization of the neutrophil and monocyte compartments. Related to Figure 2.
(A) t-SNE plots generated as described in Figure 2A based on a maximum of 1'000 neutrophils per patient colored by clusters identified with the PhenoGraph algorithm (n=78 individuals).
(B) Heatmap of the normalized marker expression for the PhenoGraph neutrophil clusters. Each cluster was manually assigned to three neutrophil subsets based on the negative, intermediate, or high expression of CD16. The assignment to the different neutrophil groups is indicated on the right side of the heatmap as "Phenotype".
(C) Overlaid histogram displaying the expression of CD16 (top panel) and Ki-67 (bottom panel) for the indicated neutrophil subsets. Arcsinh transformed intensities are shown for both markers on a linear axis. Values for each subset and marker correspond to the median ion raw count
(D) Visualization of monocytes clusters using first and second components of a diffusion map. Cells are colored by PhenoGraph clusters. The two branches leading to the CD169+ cells are shown with black arrows (n=78 individuals).
(E) Diffusion map of the monocyte subsets as described in D colored by normalized marker intensities (n=78 individuals).

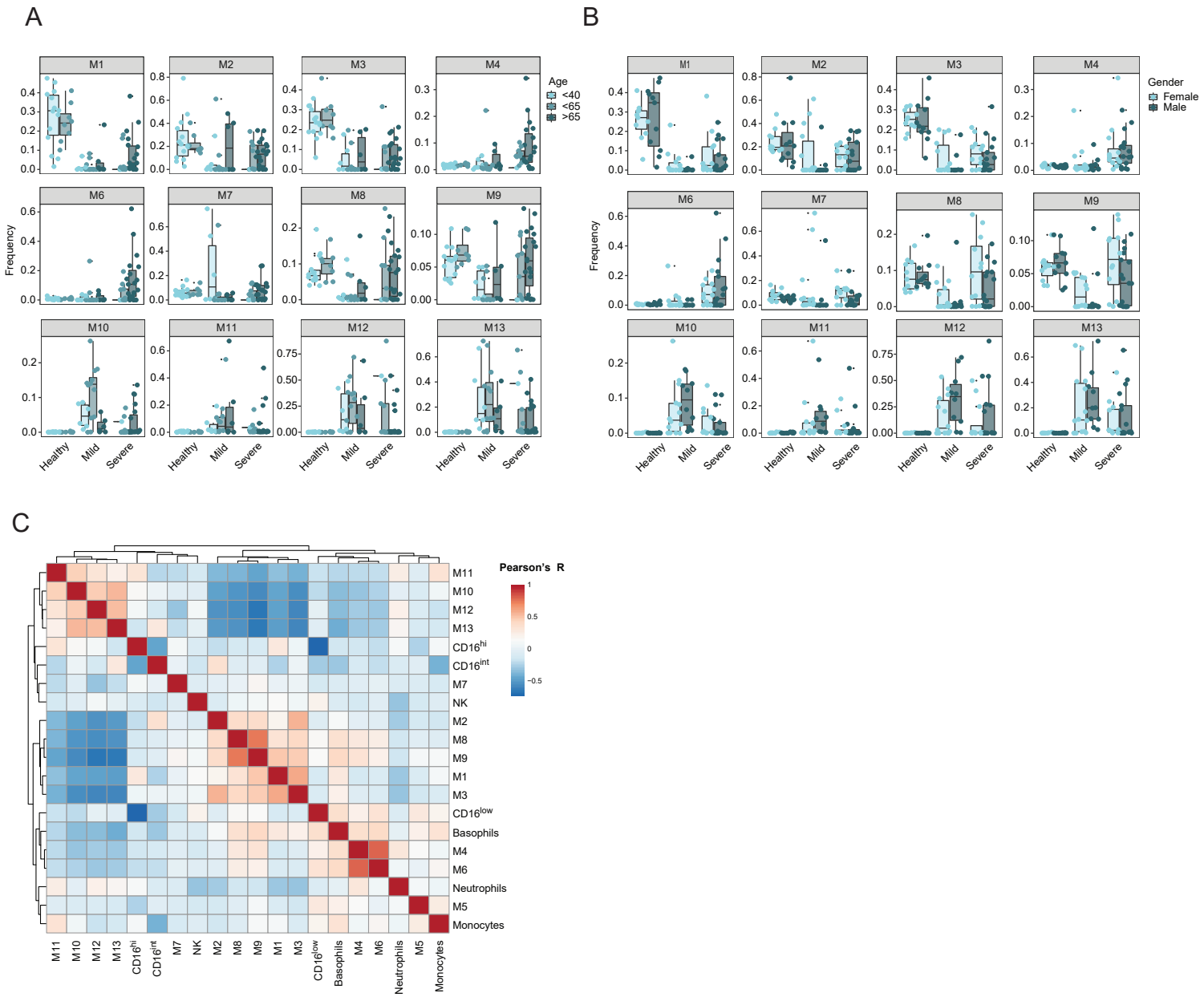

Figure S3: In-depth characterization of the neutrophil and monocyte compartments. Related to Figure 3.
(A) Boxplots of frequencies of the indicated monocyte clusters in the different disease severities splitted by age group as indicated on the right (n=75 individuals).
(B) Boxplots of frequencies of the indicated monocyte clusters in the different disease severities splitted by gender as indicated on the right (n=75 individuals).
(C) Heatmap of correlations between monocyte and neutrophil clusters and myeloid immune subsets across the cohort. The colors indicate Pearson's correlation coefficients (n=75 individuals).

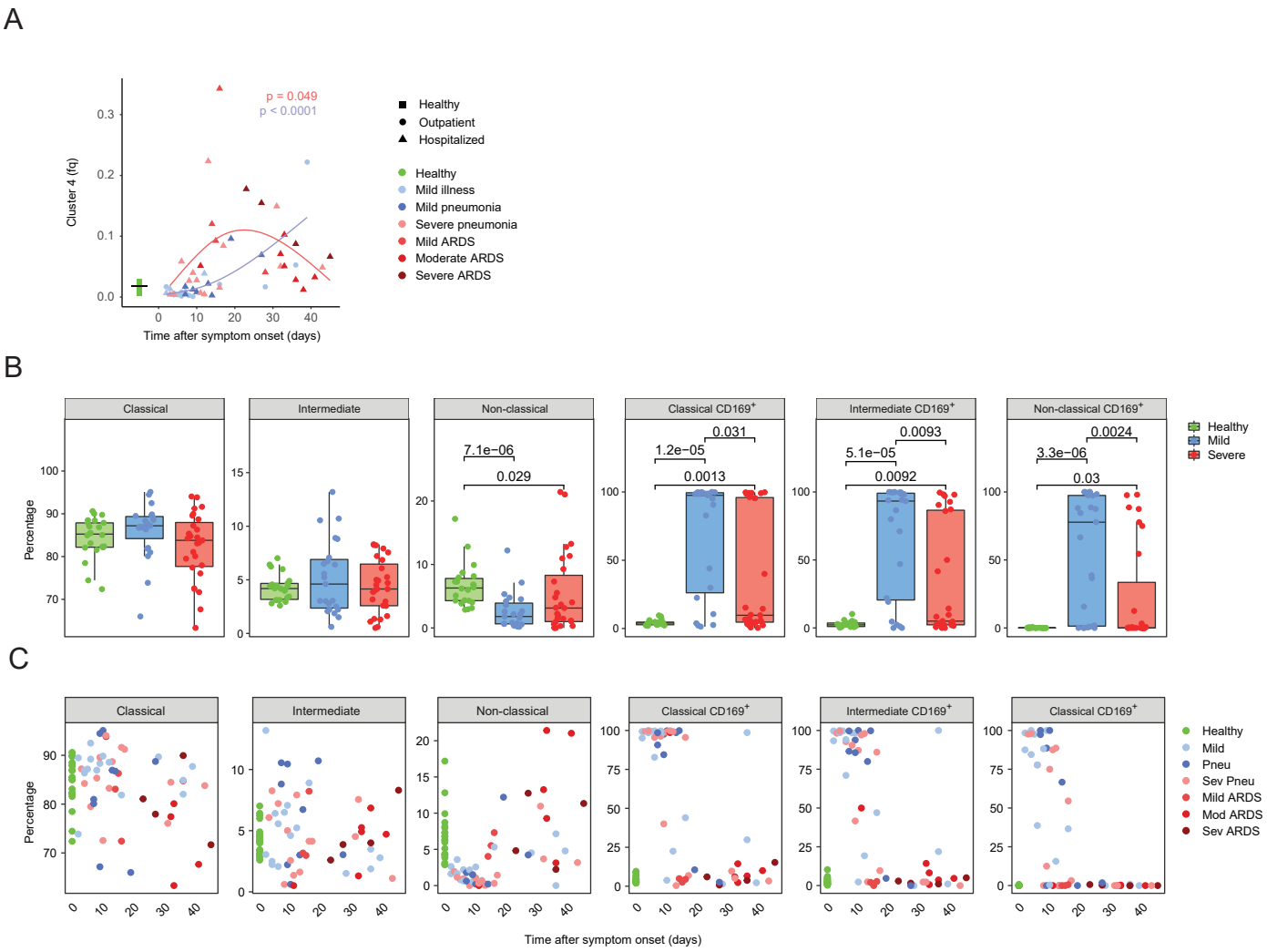

Figure S4: Disease status of patients with high cluster 4 frequency. Related to Figure 4.
(A) Scatter plot of M4 cluster frequencies relative to the time after symptom onset. Dots are colored by disease severity and shapes indicate hospitalization status. The cluster frequency is given in relation to the total monocytes in the PBMCs. The frequencies in healthy controls are shown as a reference on the left. The pseudo-time course was modeled using a generalized additive model for the disease severities separately (mild, blue lines; severe, red lines, n=75 individuals).
(B) Boxplots of frequencies of the indicated monocyte cell population identified based on manual gating as indicated in Figure 4D in the different disease severities. Statistical analyses were performed with a Mann-Whitney-Wilcoxon test corrected for multiple testing based on the Holm method and p-values are shown if the results were significant ($p<0.05$, n=75 individuals).
(C) Scatter plot of manually gated cell subset percentages relative to the time after symptom onset. The dots are colored by disease grade at sampling time. The frequencies in healthy controls are shown as a reference on the left (n=75 individuals).

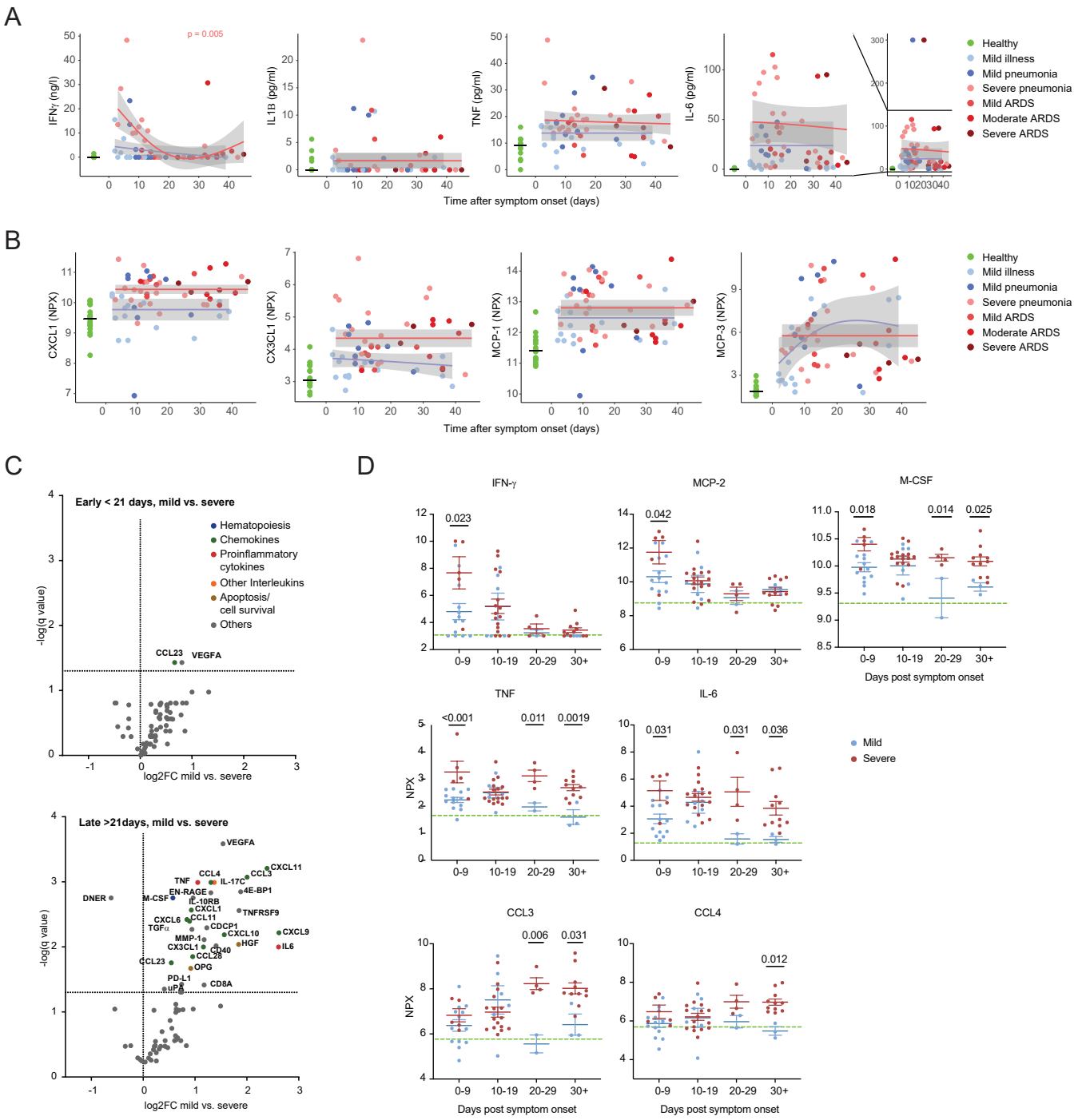

Figure S5: Association between serum protein levels, disease stages, and myeloid subsets. Related to Figure 5.
(A) Scatter plots of expression of selected cytokines as measured by ELISA. The dots are colored by disease grade at sampling time. The expression levels of the healthy controls are shown as a reference on the left, with a horizontal line indicating the median. The pseudo-time course was modeled using a general additive model for the disease severities separately (mild, blue lines; severe, red lines, n=19 healthy, 28 mild and 38 severe).
(B) Scatter plots of the expression of the indicated serum proteins versus the time after symptom onset. The expression is given as NPX on a log2 scale. The dots are colored by disease grade at sampling time. The expression levels in healthy controls are shown as a reference on the left, with a horizontal line indicating the median. The pseudo-time course was modeled using a general additive model for the disease severities separately (mild, blue lines; severe, red lines, 77 individuals).
(C) Volcano plot of proteomics data for COVID-19 patients with mild vs. severe disease at early (top panel) and late (bottom panel) stages. An FDR of 5% was taken as cut-off for significance (n=60 individuals).
(D) Scatter plots of expression of selected serum proteins relative to time after symptom onset binned as indicated. The expression is given as NPX on a log2 scale. The dots are colored by disease severity at sampling time. The mean expression levels of the healthy controls are shown as a dotted horizontal line. For each bin, the median and the standard error of the mean are shown in red (severe) and blue (mild). Statistical analyses were performed with a t-test corrected for multiple testing based on the Holm method and p-values are shown if the results were significant ($p<0.05$, n=60 individuals).

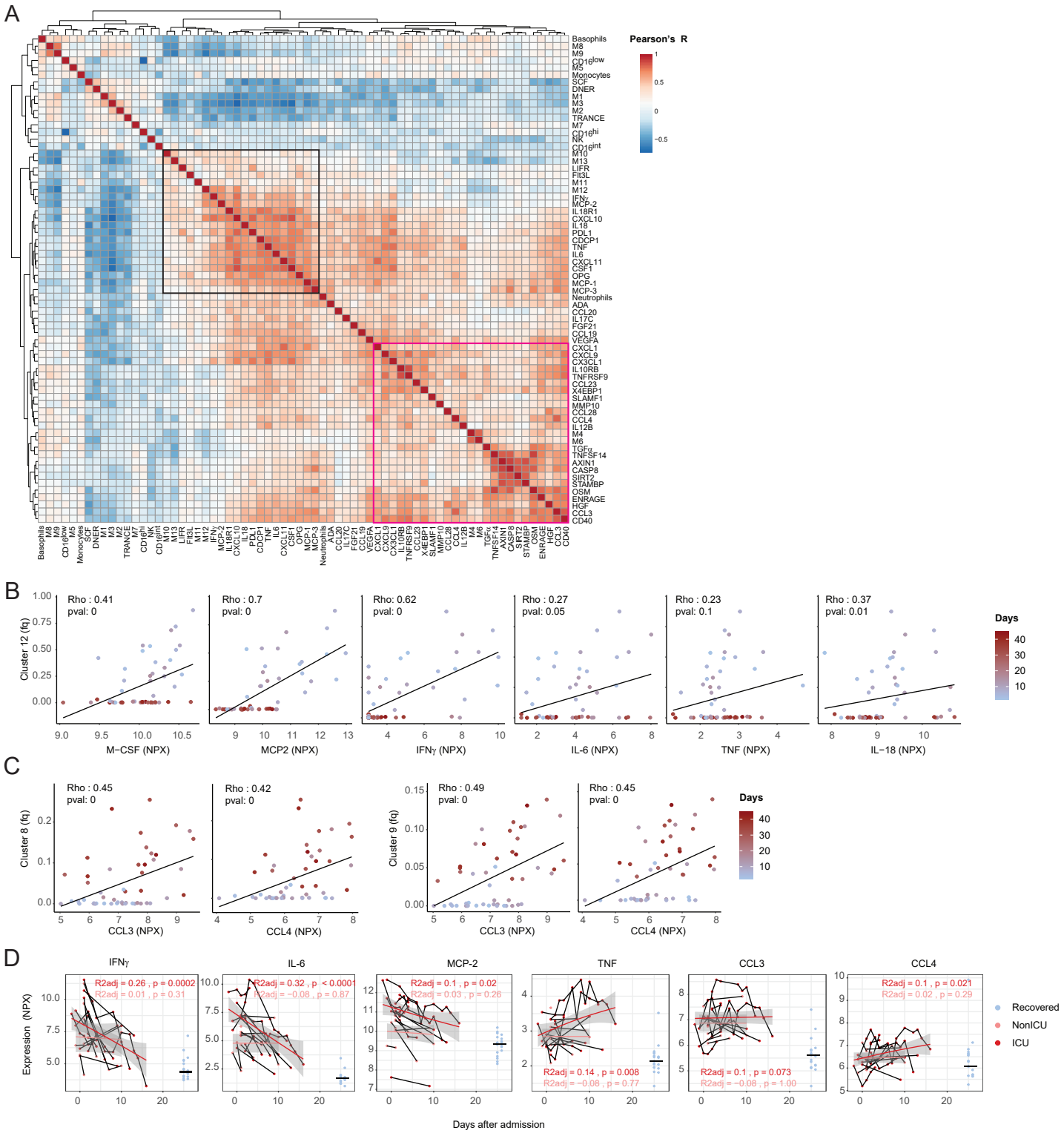

Figure S6 : Association between serum protein levels and myeloid subsets over the course of the disease. Related to Figure 6.
(A) Heatmap of Pearson correlation coefficients for relationships between the different PhenoGraph clusters, myeloid immune cell subsets, and significant differentially expressed serum proteins. Black square identifies a set of serum proteins associated with CD169+ activated monocyte subsets (M10-M13). Pink square identifies a set of serum proteins associated with the disease related monocyte subsets M4 and M6 (n=70 individuals).
(B) Scatter plots of frequencies of the indicated clusters versus expression of selected serum proteins in individual patients as described in Figure 6C, but colored by time after symptom onset. Relationship between the two variables is visualized with a linear regression line and quantified using a Spearman's correlation coefficient (Rho) with the corresponding p- value. Fq, frequency.
(C) Scatter plots of frequencies of the indicated clusters versus expression of selected serum proteins in individual patients as described in Figure 6C, but colored by time after symptom onset. Relationship between the two variables is visualized with a linear regression line and quantified using a Spearman's correlation coefficient (Rho) with the corresponding p- value. Fq, frequency.
(D) Scatter plots of the expression of the indicated serum proteins versus the time after hospitalization of patients from a publicly available dataset (dataset by Rodriguez et al. The expression is given as NPX on a log2 scale. The dots are colored by ICU status. The expression levels in recovered patients are shown as a reference on the right. Dots from individual patients are connected with a black line. The pseudo-time course was modeled using a generalized additive model for the different ICU status separately (non ICU, light red lines; ICU, red lines). Adjusted R2 and p-values of generalized additive models are shown for both the ICU and non ICU groups (n=19 recovered and 17 longitudinal patients).

Supplementary Table 1: Clinical laboratory parameters. Related to Table 1.

Laboratory values	Healthy	Mild cases	Severe cases
C-reactive protein (mean ± SD)	1.21 ± 1.61	29.87 ± 51.60*	89.46 ± 81.23**
Lactate dehydrogenase (% > upper limit of normal)	0%	28%	73.53 %*
Hemoglobin (mean ± SD, [g/l])	139.88 ± 13.31	136.39 ± 15.85	131.61 ± 15.77
Absolute platelet count (mean ± SD, [G/l])	257.25 ± 60.50	203 ± 67.97	219.74 ± 117.14
Total white blood cell count (mean ± SD, [G/l])	5.74 ± 1.52	5.83 ± 2.90	6.65 ± 3.49
Monocytes (mean ± SD, [G/l])	0.42 ± 0.15	0.50 ± 0.36	0.45 ± 0.34
Neutrophils (mean ± SD, [G/l])	3.17 ± 1.03	3.74 ± 2.70	5.28 ± 3.36**
Eosinophils (mean ± SD, [G/l])	0.13 ± 0.08	0.04 ± 0.05*	0.03 ± 0.07**
Basophils (mean ± SD, [G/l])	0.04 ± 0.02	0.02 ± 0.03*	0.01 ± 0.02*
Lymphocytes (mean ± SD, [G/l])	1.95 ± 0.74	1.50 ± 0.69*	0.81 ± 0.44**
CD3- $CD56^{bright}$ $CD16^{dim}$ NK cells (mean ± SD, [cells/ul])	11.14 ± 5.65	8.19 ± 4.98	6 ± 4.39*
CD3- $CD56^{dim}$ $CD16^{bright}$ NK cells (mean ± SD, [cells/ul])	206.29 ± 107.13	165.96 ± 147.96	162.16 ± 103.69
* Indicates significance (p-value threshold <0.05) compared to the healthy, ** in the severe indicates significance in comparison to the healthy and the mild.			