Check for updates

# Cellular and plasma proteomic determinants of COVID-19 and non-COVID-19 pulmonary diseases relative to healthy aging

Laura Arthur<sup>1,8</sup>, Ekaterina Esaulova<sup>1,8</sup>, Denis A. Mogilenko<sup>1,9</sup>, Petr Tsurinov<sup>1,2</sup>, Samantha Burdess<sup>1</sup>, Anwesha Laha<sup>1</sup>, Rachel Presti<sup>3</sup>, Brian Goetz<sup>4</sup>, Mark A. Watson<sup>1</sup>, Charles W. Goss<sup>5</sup>, Christina A. Gurnett<sup>6</sup>, Philip A. Mudd<sup>10,7</sup>, Courtney Beers<sup>4</sup>, Jane A. O'Halloran<sup>3</sup> and Maxim N. Artyomov<sup>1</sup>

We examine the cellular and soluble determinants of coronavirus disease 2019 (COVID-19) relative to aging by performing mass cytometry in parallel with clinical blood testing and plasma proteomic profiling of ~4,700 proteins from 71 individuals with pulmonary disease and 148 healthy donors (25–80 years old). Distinct cell populations were associated with age (GZMK+CD8+T cells and CD25<sup>low</sup> CD4+T cells) and with COVID-19 (TBET<sup>-</sup>EOMES<sup>-</sup> CD4+T cells, HLA-DR+CD38+CD8+T cells and CD27+CD38+ B cells). A unique population of TBET+EOMES<sup>+</sup> CD4+T cells was associated with individuals with COVID-19 who experienced moderate, rather than severe or lethal, disease. Disease severity correlated with blood creatinine and urea nitrogen levels. Proteomics revealed a major impact of age on the disease-associated plasma signatures and highlighted the divergent contribution of hepatocyte and muscle secretomes to COVID-19 plasma proteins. Aging plasma was enriched in matrisome proteins and heart/aorta smooth muscle cell-specific proteins. These findings reveal age-specific and disease-specific changes associated with COVID-19.

OVID-19 is caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a pandemic virus that rapidly spread worldwide, killing over two million individuals as of February 2021 (World Health Organization<sup>1</sup>). Most individuals infected by SARS-CoV-2 are asymptomatic or have mild to moderate clinical symptoms<sup>2</sup>. However, a notable portion of infected individuals develop severe symptoms, including high fever, shortness of breath and muscle pain. The most severe cases of infection progress to acute respiratory distress syndrome, multiorgan failure and death. COVID-19 severity has been associated with lymphopenia<sup>3-5</sup>, elevated C-reactive protein<sup>6</sup> and increased proinflammatory cytokines such as interleukin (IL)-1<sup>β7,8</sup>, IL-6 (refs. <sup>9-12</sup>), IL-8 (ref. <sup>10</sup>) and tumor necrosis factor (TNF)<sup>9,10</sup>, indicating an ongoing systemic immune response. Several recent studies have characterized the altered composition of the immune cells in patients with COVID-19 compared to healthy or recovered patients<sup>13-15</sup>. In these studies, it remains unclear which emerging features are specific to COVID-19 and how many observations are shared with other inflammatory pathologies.

Compared to other respiratory infections, COVID-19 has several unique features. The risk of progression to severe disease and mortality is greater in individuals with comorbidities like obesity, hypertension and diabetes<sup>16</sup>. Most strikingly, COVID-19 is characterized by a profound age-associated susceptibility; individuals over 65 years old have the highest infection fatality rate and account for more than 70% of COVID-19 deaths<sup>17-19</sup>. It is known that immune cell composition changes significantly with age, as does the environment, for example, the plasma proteome<sup>20</sup>. Therefore, understanding the COVID-19-driven immune response in the context of the aging immune system is critically important in determining why pathogens like SARS-CoV-2 more frequently initiate a severe clinical presentation in older individuals. However, a typical study design for immunophenotyping peripheral blood mononuclear cells (PBMCs) from COVID-19 includes only a comparison between middle-aged healthy or recovered individuals and patients with COVID-19 who are typically 60 years and older<sup>13,21,22</sup>.

In this study, we use clinical blood testing, mass cytometry and unbiased proteomics profiling of ~4,700 proteins to examine the phenotypic characteristics of plasma and PBMCs in nonobese individuals with respiratory distress with or without laboratoryconfirmed infection by SARS-CoV-2 (71 individuals) and compare these cohorts to samples from age-stratified healthy nonobese individuals (148 individuals from 25 to 80 years old).

#### Results

**Study design and clinical cohorts.** First, we considered individuals who presented with respiratory illness symptoms and had a physician-ordered SARS-CoV-2 test performed at the Barnes Jewish Hospital between 26 March 2020 and 28 August 2020 (Washington University 350 (WU350) cohort). Based on nasopharyngeal testing by PCR with reverse transcription (RT–PCR), participants were defined as SARS-CoV-2 positive (CV; 140 females and 173

<sup>1</sup>Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, MO, USA. <sup>2</sup>JetBrains Research, Saint Petersburg, Russia. <sup>3</sup>Department of Medicine, Washington University School of Medicine, Saint Louis, MO, USA. <sup>4</sup>Siteman Cancer Center, Washington University School of Medicine, Saint Louis, MO, USA. <sup>5</sup>Division of Biostatistics, Washington University School of Medicine, Saint Louis, MO, USA. <sup>6</sup>Department of Neurology, Washington University School of Medicine, Saint Louis, MO, USA. <sup>7</sup>Department of Emergency Medicine, Washington University School of Medicine, Saint Louis, MO, USA. <sup>8</sup>These authors contributed equally: Laura Arthur, Ekaterina Esaulova. <sup>Exa</sup>e-mail: martyomov@wustl.edu

#### **NATURE AGING**



**Fig. 1 Study outline and clinical characterization of healthy and COVID-19/non-COVID-19 cohorts.** Blood panels were performed for the following cohorts: A (25–34 years), n=36; B (35–44 years), n=21; C (45–54 years), n=16; D (55–65 years), n=24; E (>65 years), n=25; CV, COVID-19 (32–91 years, 70.8 mean, 11.2 s.d.), n=53; NCV, non-COVID-19 (32–87 years, 52.8 mean, 17 s.d.), n=17. See Extended Data Fig. 1 for statistics related to **b-d. a**, Study outline. An asterisk represents four patients who had a BMI < 33. **b-d**, Selected WBC differentials (**b**); RBC, hemoglobin and platelet differentials (**c**); and clinical blood values (**d**) for cohorts A-E and CV/NCV cohorts. The lower and upper hinges of all box plots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to values that are no further than 1.5 times the IQR from either the upper or the lower hinge. RDW, RBC distribution width; ER, emergency room.

males) or SARS-Cov-2 negative (NCV; 98 females and 40 males; Fig. 1a). The population was heterogeneous for body mass index (BMI), where nearly half of individuals were moderately or severely obese (BMI>33; Extended Data Fig. 1a). Given that obesity is a recognized risk factor for severe COVID-19 (ref. <sup>16</sup>) and known to strongly impact immune and proteomic homeostasis<sup>23</sup>, we chose to minimize these confounding factors in our analysis and excluded participants with moderate and severe obesity. Our selected CV and NCV cohorts consisted of 80 individuals, with age and sex distri-

butions proportional to those of the nonobese individuals (53 CV individuals: median BMI, 25.5; interquartile range (IQR), 21.9–28.4; 27 NCV individuals: median BMI, 27.3; IQR, 25.6–29.8; Extended Data Fig. 1b,c). We cannot conclusively rule out SARS-CoV-2 infection in participants with negative SARS-CoV-2 tests because the false-negative rate of the nasopharyngeal RT–PCR test is reported to be 0.018–0.33 (ref. <sup>24</sup>); however, 13 of 27 NCV individuals were retested, and none of the retests was positive for SARS-CoV-2, and none of the 27 individuals had a subsequent hospital readmission.

The most common diagnoses at discharge were pneumonia and chronic obstructive pulmonary disease (Supplementary Table 1). The majority of nonobese individuals with COVID-19 were males (~70%), and the average age was 71 years. The age of individuals without COVID-19 was distributed more broadly, with an average age of 55 years old (Fig. 1a and Extended Data Fig. 1c). We divided the participants with COVID-19 into three subgroups based on admission to an intensive care unit (ICU) and survival criteria: (1) CV\_moderate, including individuals who were not admitted to the ICU during treatment, (2) CV\_severe, including individuals who were admitted to the ICU, and (3) CV\_deceased, including individuals who did not survive the illness (Supplementary Tables 2 and 3 and Fig. 1a). Most individuals admitted to the ICU were assigned a severity score based on a time-weighted average of discharge readiness<sup>25</sup>. Of note, our ICU-based definition of severity correlated well with known inflammation characteristics such as C-reactive protein levels (Extended Data Fig. 1e,f) and other common parameters of disease severity such as intubation and severity score (Extended Data Fig. 1d). Consistent with the known increase in COVID-19 severity with age, the average age of the deceased cohort was higher compared to individuals with moderate or severe COVID-19 (Supplementary Table 2 and Fig. 1a).

Age is a known susceptibility factor for COVID-19, and it also significantly affects the immune and proteomic homeostasis in healthy individuals<sup>20,26</sup>. Therefore, to discriminate the effect of aging from disease-associated changes, we expanded our study to include a cohort of 148 healthy nonobese individuals aged 25 to 80 years, divided into five age groups (ABF300 cohort; Fig. 1a and Extended Data Fig. 1d). These blood samples were collected before the COVID-19 pandemic as part of an ongoing study of healthy human aging. In total, we analyzed 219 samples using clinical blood tests, complete blood count differentials, mass cytometry immunostaining (CyTOF) and plasma proteomics. Joint analysis of the healthy ABF300 cohort and the WU350 COVID-19 and non-COVID-19 cohorts revealed unique age-specific and disease-specific features of immune and physiological responses to COVID-19.

**Clinical laboratory characteristics.** Complete blood count differential analysis showed a statistically significant increase in the absolute numbers of white blood cells (WBCs) in NCV and non-moderate CV groups (Fig. 1b; see Extended Data Fig. 1f for statistical evaluation between all groups). This increase was attributed to a statistically significant increase in numbers of neutrophils (adjusted *P* value ( $P_{adj}$ ) < 0.001), while changes in lymphocyte and monocytes numbers did not reach statistical significance when comparing NCV and CV groups to the age-matched healthy control groups (Fig. 1b and Extended Data Fig. 1g). This observation is consistent with previous reports<sup>27,28</sup>, including the increase in immature granulocytes with disease severity<sup>14,29</sup> (Fig. 1b).

We observed that red blood cell (RBC) count decreased within the oldest age group (A versus E;  $P_{adj}$ , <0.001) and that RBC count in NCV participants and individuals with moderate COVID-19 did not statistically differ from corresponding age-matched values, while individuals with severe COVID-19 had a statistically lower RBC count compared to healthy individuals of any age (Fig. 1c; see Extended Data Fig. 1g for statistical evaluation between all groups). Similar alterations were observed for hemoglobin levels (Fig. 1c and Extended Data Fig. 1g). Strikingly, RBC distribution width was distinctly associated with COVID-19 at all severity levels relative to both healthy people and individuals without COVID-19 (Fig. 1c), consistent with previous works<sup>30,31</sup>. Lastly, platelet counts demonstrated a decreasing trend that appeared specific to individuals with COVID-19, although it did not reach significance in our cohorts (Extended Data Fig. 1g).

Several biochemical parameters changed dramatically in an inflammation and/or COVID-19-specific manner. Albumin

concentration, indicative of liver health, did not decrease with age, but it significantly decreased during inflammation, particularly in COVID-19 groups of all severity levels (Fig. 1d; see Extended Data Fig. 1h for statistical evaluation between all groups). Calcium significantly decreased in individuals with COVID-19 compared to all ages of healthy controls and individuals without COVID-19, consistent with previous reports<sup>32</sup>, yet our data show that individuals without COVID-19 demonstrated only a nonsignificant decreasing trend compared to healthy individuals (Fig. 1d). Of note, unlike other blood ions (potassium, sodium and chloride), calcium levels did not increase with age (Extended Data Fig. 2a,b). Biochemical measures indicative of kidney function showed patterns that were strikingly specific to individuals with COVID-19 and correlated with disease severity. Specifically, creatinine and urea nitrogen levels did not differ between healthy individuals and participants without COVID-19, while they increased progressively in individuals with COVID-19, with the highest levels reached in the deceased cohort (Fig. 1d). Notably, urea nitrogen levels, but not creatinine levels, were age dependentincreasing with age within the healthy range (Extended Data Fig. 2c,d). However, the significant urea nitrogen level increase in severe and deceased COVID-19 groups was not attributed to age, as the COVID-19-dependent increase was significant even when compared to the oldest age group ( $P_{adj.} < 0.05$ , CV\_severe versus cohort E;  $P_{adi}$  < 0.001, CV\_deceased versus cohort E; Extended Data Fig. 1i). Other age-dependent biochemical properties observed in the healthy control cohort included C-peptide levels<sup>33</sup>, lactic acid dehydrogenase levels<sup>34</sup>, glucose<sup>35</sup>, thyrotropin<sup>36</sup> and DHEA<sup>37</sup> (Extended Data Fig. 2).

CvTOF analysis of peripheral blood mononuclear cells. To understand changes in immune cell populations with the disease, we performed mass cytometry (CyTOF) on PBMCs of 219 blood samples from the healthy and disease cohorts using 28 myeloid and lymphoid markers (Methods). A subset of target proteins was selected based on single-cell RNA sequencing (scRNA-seq) of PBMCs to maximize cellular subset resolution. Specifically, we included mucosal-associated invariant T (MAIT) cell and yo T cell markers (TCRVA7.2 and TCRγδ, respectively) and antibodies to granzymes GZMK and GZMB because we<sup>38</sup> and others<sup>39</sup> have shown that these proteins discriminate two major effector memory T (T<sub>FM</sub>) CD8<sup>+</sup> cell subpopulations. We identified the major cell populations such as T cells (CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, γδ T cells and MAIT cells), B cells, natural killer (NK) cells and myeloid cells (Fig. 2a) using unsupervised clustering and distribution of key lineage markers (Extended Data Fig. 3b and Methods).

Differences between the major cell subpopulations can be appreciated directly from the distributions seen in cell density plots (Extended Data Fig. 3c). B cell proportions significantly increased in both SARS-CoV-2-positive and SARS-CoV-2-negative disease groups in line with previously reported results<sup>13</sup> (Fig. 2b; see Extended Data Fig. 3d for statistical evaluation between all groups), indicating that this increase is a general characteristic of the immune response to pulmonary disease. Proportions of CD4+ T cells for NCV, CV\_moderate and CV\_deceased groups were decreased relative to age-matched healthy controls. A similar decrease in CD4+ T cell proportions during SARS-CoV-2 and influenza infection was recently reported<sup>40,41</sup> (Fig. 2b and Extended Data Fig. 3d). Proportions of CD8<sup>+</sup> T cells were increased in the group with moderate COVID-19 compared to the age-matched healthy group (group E), while there was no statistically significant difference for severe and deceased individuals relative to healthy individuals of any age. Of note, within the healthy cohort, CD8<sup>+</sup> T cells proportions were significantly decreased in the oldest donors (group E; >65 years old) relative to younger groups. Next, we analyzed major immune cell populations individually (Fig. 2b).

#### **NATURE AGING**



**Fig. 2** | **Defining major immune subsets in PBMCs by CyTOF for healthy and COVID-19/non-COVID-19 groups.** CD4<sup>+</sup> T cell activation in participants with COVID-19 comes from age and inflammation signatures. Cohorts: A, n = 38; B, n = 28; C, n = 20; D, n = 29; E, n = 33; NCV, n = 17; CV\_moderate, n = 18; CV\_severe, n = 18; CV\_deceased, n = 12. **a**, Uniform manifold and approximation projection (UMAP) plot of all cell profiles with CyTOF, colored according to identified cell types. **b**, Cell proportions of each cluster across cohorts. **c**, UMAP plot of CD4<sup>+</sup> T cells, colored by the cluster. **d**, Heat map of normalized gene expression for all genes used for CD4<sup>+</sup> T cell analysis, per cluster. **e**, UMAP plots with the expression of selected markers. **f**, UMAP density plots characterizing the distribution of CD4<sup>+</sup> T cells across conditions. **g**, MDS projection for all samples, colored by cohort. For each sample, cluster percentages were used to perform MDS. **h**, Cell proportions of each CD4<sup>+</sup> T cell cluster across cohorts. In **b** and **h**, the lower and upper hinges of all box plots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the values that are no further than 1.5 times the IQR from either the upper or the lower hinge. See Extended Data Fig. 3 for statistics related to **b** and Extended Data Fig. 4 for statistics related to **h**.

**CD4<sup>+</sup> cells.** We performed dimensionality reduction and clustering based on the relevant subset of markers (Methods) and identified 12 CD4<sup>+</sup> T cell subpopulations (Fig. 2c,d). They included three subsets of CD4<sup>+</sup> T<sub>EM</sub> cells (that is, CCR7<sup>-</sup>CD45RO<sup>+</sup>) divided based on EOMES and TBET expression, two subpopulations of central memory T (T<sub>CM</sub>) CD4<sup>+</sup> cells (that is, CCR7<sup>+</sup>CD45RO<sup>+</sup>) distinguished by the level of CD45RO expression (medium or low), two subpopulations of regulatory T (T<sub>reg</sub>) CD25<sup>+</sup> CD4<sup>+</sup> cells (CD45RA positive and CD45RO positive), three subpopulations of naïve CD4<sup>+</sup> T cells

based on the combinatorial expression of CD25 and SELL (CD62L) and two subpopulations with generally low levels of both CD45RA and CD45RO surface markers, which we denoted as RA<sup>low</sup>RO<sup>-</sup> (Fig. 2c,d and Extended Data Fig. 4a). Changes in population structure associated with age and disease were evident from the density plots of individual groups (Fig. 2f). Multidimensional scaling (MDS), computed based on the cluster percentages, also demonstrated distinct age-dependent and disease-dependent sample separation (Fig. 2g).



**Fig. 3 | CD8<sup>+</sup> T cells in COVID-19/non-COVID-19 groups lose the conventional effector memory phenotype, with a COVID-19-specific increase in HLA-DR<sup>+</sup>CD38<sup>+</sup> CD8<sup>+</sup> T cells. Cohorts: A, n = 38; B, n = 28; n = 20; D, n = 29; E, n = 33; NCV, n = 17; CV\_moderate, n = 18; CV\_severe, n = 18; CV\_deceased, n = 12. <b>a**, UMAP plot of all CD8<sup>+</sup> T cells, colored by the cluster. **b**, Heat map of normalized gene expression for all genes used for CD8<sup>+</sup> T cell analysis, per cluster. **c**, UMAP plots with the expression of selected markers. **d**, UMAP density plots characterizing the distribution of CD8<sup>+</sup> T cells across conditions. **e**, MDS projection for all samples, colored by cohort. For each sample, cluster percentages were used to perform MDS. **f**, Cell proportions of each CD8<sup>+</sup> T cell cluster across cohorts. See Extended Data Fig. 5 for statistics related to **f**. The lower and upper hinges of all box plots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to values that are no further than 1.5 times the IQR from either the upper or the lower hinge.

A decrease in naïve CD4<sup>+</sup> T cells was one of the most prominent age-associated features, and this population was further diminished in individuals with pulmonary disease, both in SARS-CoV-2-positive and SARS-CoV-2-negative groups (Fig. 2h; see Extended Data Fig. 4b for statistical evaluation between all groups). Interestingly, the population of naïve CD4+ T cells lacking SELL surface expression was distinctly upregulated (see naïve SELL- population in Fig. 2h; Extended Data Fig. 4b) in disease cohorts, likely comprising a transient population associated with an active immune response. A similar pattern was observed for a subset of T<sub>CM</sub> cells characterized by low levels of CD45RO expression (CM ROlow), which increased specifically in the disease conditions. Among the three subsets of T<sub>EM</sub> cells, the subpopulation lacking both TBET and EOMES (TBET-EOMES-CD4+) expression significantly increased in disease groups, likely indicating effector cells associated with the immune response. Proportions of CD4<sup>+</sup>  $T_{EM}$  cells that expressed both TBET and EOMES were specifically increased in moderate but not severe or deceased COVID-19-infection cohorts. This subpopulation of CD4<sup>+</sup> T cell expresses cytotoxicity markers (GZMB and GZMK), which might be beneficial in disease progression. This population also appeared to accumulate with age, albeit the difference did not reach statistical significance (Fig. 3f and Extended Data Fig. 4b). This population likely corresponds to recently reported cytotoxic CD4<sup>+</sup> T cells that dramatically increase in supercentenerians<sup>42</sup>.

Additionally, we identified a distinct CD4<sup>+</sup> T cell subpopulation, RA<sup>low</sup>RO<sup>-</sup>CD25<sup>low</sup>, which progressively accumulated with age (Fig. 3f and Extended Data Fig. 4b). To our knowledge, this is the first time this cell population has been defined as age dependent. Interestingly, this population was increased in individuals with severe COVID-19 but not in those with moderate or no COVID-19, compared to younger controls (that is, group A or B; Extended Data Fig. 4b).

Taken together, the CD4<sup>+</sup> T cell compartment demonstrates ageassociated (increase in RA<sup>low</sup>RO<sup>-</sup>CD25<sup>low</sup>, loss of naïve cells, increasing trend of TBET<sup>+</sup>EOMES<sup>+</sup> and central memory populations) and inflammation-associated remodeling, where its key features (further loss of conventional naïve cells, increase in TBET<sup>-</sup>EOMES<sup>-</sup>, CD45RO<sup>low</sup> and naïve SELL<sup>-</sup> cells) appear to be associated with the respiratory pathology immune response rather than COVID-19specific responses, with the possible exception of the T<sub>EM</sub> TBET<sup>+</sup> EOMES<sup>+</sup> subpopulation which increases strongly in individuals with moderate COVID-19.

CD8+ cells. CD8+ T cells demonstrated the most striking remodeling in healthy aging and inflammatory contexts (Fig. 3). In total, we identified ten CD8+ T cell clusters (Fig. 3a-c and Extended Data Fig. 5a). In addition to naïve and CD8<sup>+</sup> T<sub>CM</sub> cells, we defined eight distinct subpopulations of the CD8<sup>+</sup>  $T_{EM}$  cells—five subpopulations in healthy individuals and three subpopulations that arise during disease conditions (Fig. 3d-f and Extended Data Fig. 5b). MDS plots and density plots demonstrated distinct CD8<sup>+</sup> compartment remodeling associated with aging and disease (Fig. 3d,e). Consistent with the published scRNA-seq data<sup>39</sup> and our previous observations<sup>38</sup>, CD8<sup>+</sup> T<sub>EM</sub> cells can be divided into two major populations based on expression of GZMK and GZMB (Fig. 3c). In healthy individuals, GZMB-expressing CD8<sup>+</sup>  $T_{EM}$  cells were mostly CD45RA positive, identifying them as  $T_{\text{EMRA}}$ , and were divided into CD27<sup>+</sup>  $(4.1\% \pm 3.7\% \text{ of total CD8}^+ \text{ T cells})$  and CD27<sup>-</sup>  $(9.4\% \pm 10.8\% \text{ of })$ total CD8<sup>+</sup> T cells) subpopulations (Fig. 3b,c). We recently demonstrated that proportions of GZMK+CD8+ T cells among the total CD8<sup>+</sup> T cells increase during healthy aging<sup>38</sup>. However, surface markers distinguishing this population remained unclear. Here, we find that GZMK<sup>+</sup>CD8<sup>+</sup>  $T_{\rm EM}$  cells can be identified by the surface expression of CCR5 and are predominantly CD57 negative (Fig. 3b,c and Extended Data Fig. 5d). These data further extend our previous observation to highlight the gradual age-dependent increase in GZMK+CD8+ T<sub>EM</sub> cells. Additionally, healthy aging was accompanied by a substantial decrease in naïve cells, a significant progressive increase in T<sub>CM</sub> cells and an increasing trend of T<sub>EMRA</sub> cells, although the latter did not reach statistical significance (Fig. 3f; see Extended Data Fig. 5b for statistical evaluation between all groups). This observation extends our previous work, in which the proportion of GZMK+CD8+ T cells among the total CD8+ T cell population was shown to increase with age based on a comparison of young and old populations<sup>38</sup>. In addition to these age-dependent cell populations, two distinct PD-1-positive subsets were present in the healthy individuals, each at ~5% of total CD8<sup>+</sup> T cells: GZMB<sup>+</sup>GZMK<sup>-</sup> and GZMB<sup>+</sup>GZMK<sup>+</sup>  $T_{EM}$  cells (Fig. 3f). These cell subpopulations were characterized by a PD1+CD57+CD45RA- phenotype, yet they differed in the expression of CD27 (Fig. 3c). These cell subpopulations were present at steady levels across the aging subgroups (Fig. 3f).

The disease-associated inflammatory response was accompanied by a pronounced remodeling of the CD8<sup>+</sup> T cell compartment. Three major cell populations emerged in disease groups (Fig. 3f). The largest increase was observed for inflammatory GZMB<sup>+</sup>GZMK<sup>-</sup> and GZMB<sup>+</sup>GZMK<sup>+</sup> T cells that differed from the corresponding healthy counterparts (T<sub>EMRA</sub> and T<sub>EM</sub> GZMK<sup>+</sup> T cells, respectively) in that they lost CD45RA and CD27 surface expression (Fig. 3b,c). Lack of surface expression of CD45RA, CD27, CD28 and PD-1 proteins indicated that these could be effector cells<sup>43</sup>. Appearance of these cell populations was a shared feature of all individuals independent of COVID-19 status. However, an additional inflammatory cell population characterized by expression of HLA-DR, CD38 and PD-1 was found almost exclusively in individuals with COVID-19. The appearance of this cell population was recently reported<sup>13,44</sup>, but specificity to the COVID-19 immune response versus non-COVID-19 respiratory pathology immune response has not yet been established. The increase in these three inflammationspecific cell populations was paralleled by a decrease in the conventional steady-state subpopulations:  $T_{EMRA}$  subpopulations and GZMK-expressing  $T_{EM}$  subpopulations decreased to very low levels in all inflammatory groups (Fig. 3f). Interestingly, unlike in CD4<sup>+</sup> T cells, naïve CD8<sup>+</sup> T cells did not significantly decrease compared to corresponding age-matched controls (CV/NCV groups compared with E cohort; Fig. 3f and Extended Data Fig. 4b). This result suggests that, in this context, effector CD8<sup>+</sup> T cells may arise from  $T_{EM}$  subpopulations, for example, GZMK<sup>+</sup>  $T_{EM}$  cells acquiring the GZMK<sup>+</sup>GZMB<sup>+</sup> inflammatory T ( $T_{INFLAM}$ ) phenotype and GZMB<sup>+</sup>  $T_{EMRA}$  cells acquiring the GZMK<sup>-</sup>GZMB<sup>+</sup>  $T_{INFLAM}$  phenotype.

Taken together, we find that peripheral blood CD8<sup>+</sup> T cells undergo major remodeling during both healthy aging and inflammatory contexts. During aging, there is a loss of naïve cells and an increase of  $T_{CM}$  and GZMK<sup>+</sup>  $T_{EM}$  cells. Inflammatory remodeling is characterized by a decrease in conventional  $T_{EM}$  subpopulations and an increase in inflammatory effector-like subpopulations and HLA-DR<sup>+</sup>CD38<sup>+</sup>PD-1<sup>+</sup> CD8<sup>+</sup> T cells, which are specific to individuals with COVID-19.

NK cells, B cells and myeloid cells. NK cells were split into 11 subpopulations based on the expression of CD16, CD57, CD56, GZMK and SELL (Fig. 4a-c and Extended Data Fig. 6a). There was major inflammatory-associated remodeling of NK cells (Fig. 4d,e), as seven clusters demonstrated a difference between the healthy group and at least one inflamed group: CD56+CD57-GZMK+ (enriched in CV\_moderate group), CD56-CD57-CD16- (enriched in disease groups except for CV\_severe), CD56<sup>dim</sup>CD57<sup>+</sup>CD16<sup>-</sup> (enriched in NCV group) and CD56dimCD57low (enriched in NCV and CV moderate groups; Fig. 4f and Extended Data Fig. 6b). Two clusters did not change with age but significantly decreased across all disease cohorts: CD56+CD57low and CD56+CD57+ (Fig. 4f; see Extended Data Fig. 6b for statistical evaluation between all groups). The CD56<sup>+</sup>CD57<sup>+</sup>SELL<sup>+</sup> cluster showed a similar decreasing pattern but did not reach statistical significance. Only one cluster changed significantly with age: the CD56+CD57-SELL+ cluster decreased with age (cohort E was significantly lower than cohort A; Extended Data Fig. 6b), yet it did not change with inflammation. This observation is consistent with previous reports of a decrease in CD56<sup>+</sup> NK cells with age45 (Extended Data Fig. 6c).

Our panel included a limited number of markers to resolve B cell subpopulations. B cells separated into six clusters (Fig. 4g–i) with no significant change detected in these subpopulations across age subgroups (Extended Data Fig. 6e), and there was no clear separation between samples in the MDS plot (Fig. 4k). However, the density plots indicated some inflammation-associated remodeling (Fig. 4j). Specifically, consistent with previous reports<sup>13</sup>, we observed an increase in CD27<sup>+</sup>CD38<sup>+</sup> plasmablasts in individuals with severe COVID-19 (in comparison with age-matched healthy E cohort; Fig. 4l; see Extended Data Fig. 6e for statistical evaluation between all groups). This cell subpopulation is specific to individuals with severe COVID-19 and was not significantly differ-

**Fig. 4 | Inflammatory remodeling of NK and B cells.** Cohorts: A, n = 38; B, n = 28; C, n = 20; D, n = 29; E, n = 33; NCV, n = 17; CV\_moderate, n = 18; CV\_severe, n = 18; CV\_deceased, n = 12. **a**, UMAP plot of all NK cells, colored by the cluster. **b**, Heat map of normalized gene expression for all genes used for NK cell analysis, per cluster. **c**, UMAP plots with the expression of selected markers. **d**, UMAP density plots characterizing the distribution of NK cells across conditions. **e**, MDS projection for all samples, colored by cohort. For each sample, cluster percentages were used to perform MDS. **f**, Cell proportions of each NK cell cluster across cohorts. **g**, UMAP plot of all B cells, colored by the cluster. **h**, Heat map of normalized gene expression for all genes used for B cell analysis, per cluster. **i**, UMAP plots with the expression of selected markers. **j**, UMAP density plots characterizing the distribution of B cells across conditions. **k**, MDS projection for all samples, colored by cohort. For each sample, cluster percentages were used to perform MDS. **l**, Cell proportions of each B cell cluster across cohorts. In **f** and **l**, the lower and upper hinges of all box plots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the values that are no further than 1.5 times the IQR from either the upper or the lower hinge. See Extended Data Fig. 6 for statistics related to **f** and **l**.

### RESOURCE

ent between healthy individuals and those without COVID-19. The B cell memory population, defined as CD27<sup>+</sup>CD38<sup>-</sup>SELL<sup>+</sup>, demonstrated a COVID-19-specific decrease in proportions among the B cells (statistically significant for individuals with severe COVID-19 versus participants in all age groups).

Myeloid cells demonstrated remodeling associated with infection (Extended Data Fig. 7a–d): proportions of classical monocytes and dendritic cells significantly decreased while proportions of HLA-DR<sup>low</sup> monocytes significantly increased in the disease cohorts relative to healthy controls (Extended Data Fig. 7e,f). This DR<sup>low</sup> subset



MDS1

was previously associated with an immunosuppressive monocyte phenotype<sup>46</sup>, consistent with the general features of immunosuppression reported for COVID-19 recently<sup>47</sup>.

**Protein signatures of disease linked to healthy aging.** Next, we used the SomaScan assay to analyze the proteomic signature from CV and NCV groups (WU350) and the healthy aging cohort (ABF300). SomaScan quantifies ~4,700 proteins in relative units of intensity, allowing data comparison within homogeneously collected and processed samples (Supplementary Tables 4 and 5). One caveat of our study was that samples for the cohorts were collected using different collection approaches: WU350 samples were collected in EDTA tubes, and ABF300 samples were collected in heparin tubes. While this did not affect the measurement of cellular proportions, proteomic data from the cohorts was required to be analyzed first within each cohort and then individual aging/disease signatures could be compared across cohorts.

The comparison of CV and NCV groups identified 435 upregulated proteins in individuals with COVID-19 and 464 upregulated proteins in individuals without COVID-19 (Fig. 5a). Most of these differences were driven by the severe and lethal cases of COVID-19 (Fig. 5b). Overall, the up/down COVID-19-specific signatures demonstrated a progressive increase/decrease with disease severity (Fig. 5c,d). The same pattern emerged when each COVID-19 cohort was compared to individuals without COVID-19 (Extended Data Fig. 8a-c). A relatively small number of proteins were differentially expressed between the NCV and CV\_moderate disease groups (20 CV-specific and 7 NCV-specific upregulated proteins; Fig. 5b and Extended Data Fig. 8d). Proteins upregulated in the CV group (Fig. 5c) included complement protein C9; interferon response markers MX1, ISG15 and IFIT3; ferritin subunits FTL and FTH1; heparin-binding growth factors pleiotrophin (PTN) and midkine (MDK); growth factors CLEC11A, HAMP, TINAGL1 and SFRP1; inflammation-associated soluble factors serum amyloid a1 (SAA1), fibrinogen like protein (FGL1) and granulin (GRN); soluble forms of surface receptors FOLR2 and members of CD85 family (LILRB2 and LILRA3); and two additional proteins CHST12 and DKK3 (Fig. 5d). Notably, FGL1 and LILRA3 have the potential to directly negatively impact CD8+ T cell activity by engaging with LAG3 or interfering with human leukocyte antigen (HLA) class I/II accessibility<sup>48,49</sup>. The proteins upregulated in NCV groups (Fig. 5e,f) compared to individuals with COVID-19 included AHSG (fetuin-A), KLRC4, CLEC3B, afamin (AFM) and others.

Given the different distribution of ages between the pulmonary disease cohorts, we next examined the degree to which age-related proteomic changes shape this behavior. Comparison of young (A) versus old (E) subgroups of the aging cohort revealed 241 proteins that were statistically upregulated with age and 140 downregulated proteins (Fig. 5g). Our data are consistent with the results previously published from our group and others<sup>26,50-54</sup>: proteins most upregulated with age were GDF15, SOST and ADAMTS5, as well as PTN, TAGLN, TREM2, WISP2, MYL3 and MLN, while most downregulated proteins included RET, SELL and KIT, as well as MSMP, CILP2, CTSV and CR2 (Extended Data Fig. 8e). Because we also characterized our cohorts using clinical blood tests, we compared proteomics data with the blood biochemistry analyses obtained for the same individuals from the healthy aging cohort (Fig. 1a and Supplementary Tables 6-10). A number of measured proteins strongly correlated with the clinical blood test results (Extended Data Fig. 8f): (1) creatinine kinase strongly correlated with plasma levels of SLC26A7, CKB, ACTN2, TNNI2 and MYBPC1; (2) clinical alanine aminotransferase levels correlated with plasma levels of UGDH, ALDH1A1, ASL, ALDOB, PSAT1, ACY1, FBP1 and DCXR1; (3) C-peptide and insulin levels strongly anti-correlated with IGFBP1 and ADIPOQ (as expected, insulin levels measured by clinical blood test strongly correlated with insulin levels analyzed via SomaScan profiling); (4) clinical measurements of direct high-density lipoprotein cholesterol levels positively correlated with EHMT2 protein levels and anti-correlated with WNT5A protein levels, while the latter (5) also correlated with general triglyceride levels; (6) clinical osteocalcin levels were strongly correlated with plasma levels of CHAD protein; (7) clinical thyrotropin hormone levels strongly correlated with the corresponding protein (CGA/ TSHB) levels in the proteomic data; (8) and lastly, clinically measured unsaturated iron binding capacity was strongly correlated with FTL/FTH1 and NEO1 protein levels. While this high level of concordance does not imply that SomaScan-based profiling can substitute for clinical measurements, it demonstrates the capability of unbiased profiling in characterizing the physiological state.

Gene-set enrichment analysis (GSEA) analysis demonstrated that the COVID-19 versus non-COVID-19 differentially expressed proteins strongly associated with the up/down aging signatures, consistent with the differences in the age distribution of those cohorts (Fig. 5h,i). Furthermore, we found that the COVID-19 versus non-COVID-19 signatures significantly overlapped with the up/down aging signatures (Fig. 5j,k) but not vice versa (Extended Data Fig. 8g,h), underscoring the importance of taking age into account when considering determinants of COVID-19. We found 337 unique proteins that were upregulated in COVID-19 and 421 proteins that were downregulated in individuals with COVID-19 compared to those without COVID-19 that were not age dependent. Age-associated proteins that were also significantly different in the COVID-19 versus non-COVID-19 comparison included PTN, SFRP1 and DKK3, which increased with age, and CLEC3B, which decreased with age. It is interesting to note the dissimilar age-associated behavior of two heparin-binding proteins (MDK and PTN) that were both upregulated in the COVID-19 group relative to the non-COVID-19 group

**Fig. 5 | SomaLogic plasma protein profiling demonstrates age-specific and inflammation-specific signatures in individuals with COVID-19.** Cohorts: A, n = 42; B, n = 27; C, n = 18; D, n = 29; E, n = 34; NCV, n = 27; CV\_moderate, n = 18; CV\_severe, n = 21; CV\_deceased, n = 14. **a,b**, Volcano plot for differential expression of 4,801 proteins between NCV and all CV cohorts (**a**) or CV\_moderate, CV\_severe and CV\_deceased cohorts separately (**b**). Protein names for the top ten upregulated and downregulated genes are shown. **c,e**, Box plot of average expression per sample of proteins upregulated (**c**) or downregulated (**e**) in CV cohorts compared to NCV cohort, across CV/NCV cohorts. **d,f**, Box plot with the scaled expression of selected proteins, upregulated (**d**) or downregulated (**f**) in the CV cohort compared to NCV, across CV/NCV cohorts. Genes that are differentially expressed with age are marked in red. **g**, Volcano plot for differential expression of 4,801 proteins between cohorts A and E. Protein names for the top ten upregulated and downregulated genes are shown. **h,i**, GSEA of all proteins upregulated (**h**) or downregulated (**j**) or downregulated (**k**) with age (cohorts E versus A) in proteins ranked according to differential expression between CV/NCV cohorts. **j,k**, Overlap between proteins upregulated (**j**) or downregulated (**k**) with age (cohorts E versus A) compared to proteins upregulated in COVID-19-related inflammation (CV versus NCV comparison). *P* values are one-sided and adjusted for multiple testing using the Benjamini–Hochberg method ( $P_{adj}$ ). NES, normalized enrichment scores. **I**, Box plot with the scaled expression of selected genes in cohorts A-E. Genes that are differentially expressed with age are marked in red. In **c-f** and **I**, the lower and upper hinges of all box plots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the values that are no further than 1.5 times the IQR from either the upper or the lower hinge. (Fig. 51). Consistent with our data, PTN was previously associated with aging<sup>52</sup>, while MDK does not change with age, yet serum concentrations of MDK are linked to heart injury conditions<sup>55</sup>. Another protein associated with age and COVID-19, SFRP1, a soluble mediator of WNT signaling, has also been linked to modulation of cardiac function<sup>56</sup>. Another WNT signaling modulator, DKK3, was previously linked to aging and is considered a major indicator of muscle atrophy<sup>57</sup>. A small number of proteins behaved in the opposite manner between aging and COVID-19 (11 upregulated with

CV and downregulated with age, and 7 vice versa), which included inflammatory mediators (CCL21 and SEMA4A) or apolipoproteins (APOA4 and APOE2; Extended Data Fig. 8g,h).

**COVID-19 protein profile linked to hepatocytes and muscle secretomes.** To understand the broad-level differences between individuals with COVID-19 and individuals without COVID-19, we performed pathway enrichment analysis on the differential proteins. Several pathways were upreguated or downregulated in a



#### NATURE AGING

disease-specific manner (Fig. 6a). The pathways most upregulated in individuals with COVID-19 were associated with extracellular matrix proteins (for example, WISP2 and FBLN5) and were also profoundly associated with age (Fig. 6b,c). Similarly, soluble forms of TREM2 and IGFBP2 were increased in individuals with COVID-19 and older healthy individuals. Several COVID-19-specific pathways



**Fig. 6 | Pathway enrichment analysis distinguishes COVID-19 from non-COVID-19 inflammation.** Cohorts: A, n = 42; B, n = 27; C, n = 18; D, n = 29; E, n = 34; NCV, n = 27; CV\_moderate, n = 18; CV\_severe, n = 21; CV\_deceased, n = 14. **a**, Volcano plot for GSEA for CV/NCV comparison. The top 20 upregulated and downregulated pathways, grouped by function, are shown. Pathways differentially expressed with age are marked in red. *P* values and log fold change values were calculated using the limma package (two-sided test). Significant genes were selected after correction for multiple testing using the Benjamini-Hochberg method. **b,d,f** GSEA for selected pathways upregulated with CV ( $P_{adj.} < 0.05$ ) and with age (P < 0.05), or for pathways upregulated in CV and age (**b**), in CV but not age (**d**) or downregulated with CV but not with age (**f**;  $P_{adj.} < 0.05$ ). *P* values are one-sided and adjusted for multiple testing using the Benjamini-Hochberg method ( $P_{adj.}$ ). NES are also shown. **c,e,g**, Box plots with the scaled expression of selected genes, upregulated in CV and age (**c**), in CV but not age (**e**) or downregulated with CV but not with age (**g**). Genes that were differentially expressed with age are marked with red. The lower and upper hinges of all box plots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the values that are no further than 1.5 times the IQR from either the upper or the lower hinge. ECM, extracellular matrix; ES, enrichment score.

were independent of aging signatures and included inflammatory processes (interferon, IL-6 and IL-2/stat5), complement pathways and glycosaminoglycan metabolism (Fig. 6d,e). Conversely, proteins from MAP kinase-associated pathways were downregulated in the plasma of individuals with COVID-19 relative to that of individuals without COVID-19. These proteins were mostly independent of age and included MAP2K3, BRAF, HRAS and MAP2K4. (Fig. 6f,g).

We next evaluated if cell-type-specific signatures of PBMC subpopulations are enriched in the COVID-19-specific proteome. None of the individual cell types were enriched; however, a myeloid signature (monocytes and neutrophils) was indeed upregulated in individuals with COVID-19 (Extended Data Fig. 8). To further investigate cell-type specificities, we extracted tissue-specific transcriptional signatures from the Genotype-Tissue Expression (GTEx) database (Fig. 7a; see Methods for details and Supplementary Table 11 for list of genes) and evaluated these signatures against the proteomic data ranked by the comparisons of CV versus NCV groups or by aging comparison (A versus E cohorts; Fig. 7b). Individuals with COVID-19 had a pronounced increase in liver-specific proteins accompanied by a significant decrease of muscle-specific proteins. These tissue-associated changes were unique to the COVID-19 cohort and did not vary with age. Instead, artery/aorta-specific proteins were highly upregulated with age (Fig. 7b).

Given the distinct enrichment of these tissues, we mined public scRNA-seq data for the liver<sup>58</sup> and aorta<sup>59</sup> to understand if any specific cell type is driving these signatures. When projecting 54 liver-specific genes enriched in the comparison of CV and NCV groups, we observed a very strong specificity to hepatocytes (Fig. 7c-e), indicating an important role in regulating plasma protein level alterations in COVID-19 infection. The artery/aorta-specific signature enriched in aging also demonstrated cell-type-specific enrichment in smooth muscle cells (Fig. 7f–h).

#### Discussion

Given the strong impact of age on COVID-19 pathogenesis, it is critically important to consider patient response alongside corresponding age-matched controls. In this work, we show cellular and secreted protein determinants of individuals with COVID-19 in the context of aging. To understand features that are specific to the COVID-19 immune response as opposed to the respiratory pathology immune response, we ensured that our pulmonary cohort included individuals who tested negative and positive for SARS-CoV-2. The most pronounced changes included remodeling in CD4+ and CD8+ T cell compartments shared between individuals who tested positive for COVID-19 and those who tested negative for COVID-19 and the emergence of the COVID-19-specific populations of CD8<sup>+</sup> T cells (HLA-DR+CD38+) and B cells (CD27+CD38+). The emergence of these populations was recently identified in patients with COVID-19 (refs. <sup>13,44</sup>), albeit the cohort design used in these studies could not directly establish the specificity of these subsets to COVID-19 as opposed to non-COVID-19 pathologies. We find that the TBET+EOMES+ subpopulation of CD4+ T cells (also marked with the expression of cytotoxic marker GZMB) was highly specific to moderate but not severe or lethal COVID-19 groups. Given that our cohorts included pulmonary patients with similar symptoms who tested negative for SARS-CoV-2, we directly demonstrated the specificity of these subsets and separated them from other inflammatory immune cell subpopulations. An important aspect of our study is that we considered a control cohort across multiple age groups, whereas in most studies to date, there is a significant difference between the ages of healthy cohorts and patients with COVID-19. Accordingly, our data demonstrate that the reported decrease in some immune subpopulations (for example, total CD8<sup>+</sup> T cells) is likely a reflection of the age-associated difference in the naïve CD8 T cell population rather than a specific characteristic of the COVID-19 immune response.

Proteomics analysis revealed strong age-dependent effects within the disease signatures in addition to several disease-associated markers that have not been previously reported (for example, CLEC11A and MDK). We have found divergent cell-specific and tissue-specific signatures that differed between infection status and aging. Specifically, there was major dysregulation of hepatocyte and skeletal muscle secretomes during COVID-19 infection, while healthy aging was associated with heart smooth muscle cell-associated signatures. Taken together, our data show distinct age-specific and diseasespecific alterations and provide a new insight into potential soluble mediators of the physiological impact of COVID-19.

#### Methods

Experimental model and participant details. Sample collection. The WU350 study is a prospective observational cohort study of participants with symptoms consistent with COVID-19 who presented to Barnes Jewish Hospital, St. Louis Children's Hospital or affiliated Barnes Jewish Hospital testing sites located in Saint Louis. All individuals provided written informed consent to participate in the study. Inclusion criteria required that participants were symptomatic and had a physician-ordered SARS-CoV-2 test performed during their normal clinical care. Some participants were enrolled before the return of the SARS-CoV-2 test result. Enrolled participants who tested negative for SARS-CoV-2 are included in the current paper as non-COVID-19 respiratory illness controls. Information about follow-up tests of individuals who tested negative for SARS-CoV-2 was collected to monitor for potential false-negative results. All samples were collected during evaluation for symptoms in a medical facility or within 36 h of participant admission to the hospital. Patient-reported duration of illness and other clinically relevant medical information were collected at the time of enrollment from the medical record and the participant or their legally authorized representative. Blood collected for plasma isolation was collected in BD vacutainers with EDTA. A subset of 80 WU350 samples was selected for inclusion in mass spectrometry and proteomics analysis. The selection was based on obtaining a representative subset of individuals that had the same age and sex distribution as the full WU350 cohort with BMI < 33. Comorbidities other than obesity were not considered during sample selection to maintain the typical distribution of comorbidities expected in patients with COVID-19. The study was reviewed and approved by the Washington University in St. Louis Institutional Review Board (WU350 study approval no. 202003085). The study complied with the ethical standards of the Helsinki Declaration.

The Washington University in St. Louis Institutional Review Board reviewed and approved the ABF300 study for the collection of blood samples from healthy participants (IRB approval no. 201804084). Adults who were 25 years of age and older were recruited from the St. Louis area and provided written informed consent to participate. Participants were given a screening questionnaire to establish health status. Nonobese (BMI < 30), nonsmokers, without a history of cancer, chronic inflammatory conditions (arthritis, Crohn's disease, colitis, dermatitis, fibromyalgia or lupus) or blood-borne infections were included. Participants who reported cold or flu symptoms in the previous month were excluded. Peripheral blood (approximately 100 ml) was collected in BD vacutainer tubes with sodium heparin

### **NATURE AGING**



**Fig. 7 | COVID-19 plasma protein signatures are linked to hepatocytes and skeletal muscle secretomes.** Cohorts: A, n = 42; B, n = 27; C, n = 18; D, n = 29; E, n = 34; NCV, n = 27; CV\_moderate, n = 18; CV\_severe, n = 21; CV\_deceased, n = 14. **a**, Outline of GTEx-based analysis of SomaScan data. **b**, Enrichment of GTEx-derived tissue signatures in NCV versus CV and cohort A versus E comparisons, performed using fgsea R package. *P* values are one-sided. **c**, UMAP of liver atlas cells (GSE124395) with outlined cell types (left) and with the mean expression of 54 liver-related genes from GTEx, upregulated in the CV group (right). **d**, Heat map of normalized gene expression for 54 liver-related genes from GTEx, upregulated in CV group, for each condition. **e**, Box plot with the scaled expression of selected liver-related genes, upregulated in COVID-19, across CV/NCV cohorts. **f**, UMAP of human aortic cells (GSE155468) with outlined cell types (left) and mean expression of artery/aorta-related genes from GTEx, upregulated in cohorts E versus A. **g**, Heat map of normalized gene sform GTEx, upregulated in cohorts E versus A. **g**, Heat map of normalized gene sform GTEx, upregulated in cohorts E versus A. **g**, Heat map of normalized gene expression for artery/aorta-related genes from GTEx, upregulated in cohorts E versus A. **g**, Heat map of normalized gene expression for artery/aorta-related genes from GTEx, upregulated in cohorts E versus A. **g**, Heat map of normalized gene expression for artery/aorta-related genes from GTEx, upregulated in cohorts E versus A. **g**, Heat map of normalized gene expression of selected artery/aorta-related genes from GTEx, upregulated in cohorts E versus A. for each condition. **h**, Box plot with the scaled expression of selected artery/aorta-related genes, upregulated in cohort E versus A. across A-E cohorts. In **e** and **h**, the lower and upper hinges of all box plots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the v

by venous puncture between 07:00 and 10:00 after an overnight fast. Plasma and PBMCs were isolated from this sample. An additional sample (approximately 5 ml) was collected by a venous puncture in a BD vacutainer tube with EDTA for complete blood count with differentials. Samples from healthy participants were collected from 2018 to early 2019, eliminating the possibility of concurrent or prior SARS-CoV-2 infection. Participants were stratified into age groups of 10-year

intervals for analysis and referred to as groups A (25–34 years), B (35–44 years), C (45–54 years), D (55–64 years) and E (65+ years).

*Plasma and PBMC isolation from peripheral blood.* A portion of the whole blood was centrifuged at 500g for 30 min at room temperature. The top plasma layer was carefully recovered and frozen on dry ice. Aliquots of plasma were stored at

-80 °C. The remaining whole blood was diluted in a 1:1 ratio with sterile DPBS (Sigma) with 2 mM EDTA (Cellgro). The diluted blood was overlaid on Histopaque 1077 (Sigma) and centrifuged at 500g for 30 min at room temperature. The PBMC layer at the plasma–Histopaque interface was transferred to a new tube and washed twice with cold DBPS-EDTA. Aliquots of  $1 \times 10^6$  cells were cryopreserved in CryoStor CS10 freezing medium (BioLife Solutions) and stored at -80 °C.

*CyTOF, surface and intracellular staining.* Metal-conjugated antibodies were purchased from Fluidigm when available. For all other targets, purified antibodies were obtained and conjugated using the appropriate Maxpar Antibody Labeling Kit (Fluidigm) according to the manufacturer's protocol. The Maxpar metalconjugated antibodies were stored in PBS-based antibody stabilizing solution (Candor) supplemented with 0.09% sodium azide at 4 °C. Concentrations of all antibodies were determined by titration on PBMCs before use.

Cryopreserved PBMCs were thawed and washed in CyFACS buffer (PBS, 0.1% BSA, 0.02% NaN<sub>3</sub> and 2 mM EDTA) and incubated in human TruStain FcX blocking solution for 10 min at room temperature. The surface antibody cocktail was added to the cells for 1 h at 4 °C. The cells were washed with PBS and stained in 1 ml cisplatin (2.5  $\mu$ M). Cisplatin staining was stopped by adding 3 ml CyFACS buffer.

#### Antibodies

CD19 142Nd (HIB19)	0.33x	Fluidigm	No. 3142001B
CD127 143Nd (A019D5)	1x	Fluidigm	No. 3143012B
CD69 144Nd (FN50)	1.5×	Fluidigm	No. 3144018B
CD4 145Nd (RPA-T4)	0.5×	Fluidigm	No. 3145001B
CD8 146Nd (RPA-T8)	0.33x	Fluidigm	No. 3146001B
CD11c 147Sm (Bu15)	1x	Fluidigm	No. 3147008B
CD34 148Nd (581)	1x	Fluidigm	No. 3148001B
CD45RO 149Sm (UCHL1)	1x	Fluidigm	No. 3149001B
CCR5 purified (J418F1)	7.5µg ml−1	BioLegend	No. 359102
HLA-DR 151Eu (G46-4)	0.33x	Fluidigm	No. 3151023B
EOMES 152Sm (WD1928)	0.33x	Invitrogen	No. 14-4877-82
SELL 153Eu (DREG-56)	0.1x	Fluidigm	No. 3153004B
CD45 154Sm (HI30)	0.1x	Fluidigm	No. 3154001B
CD45RA 155Gd (HI100)	0.1x	Fluidigm	No. 3155011B
PD-1 purified (EH12.2H7)	7.5µg ml⁻¹	BioLegend	No. 329902
CD27 158Gd (L128)	0.1x	Fluidigm	No. 3158010B
TBET purified (4B10)	1.65 µg ml-1	BioLegend	No. 644802
CD28 160Gd (CD28.2)	1x	Fluidigm	No. 3160003B
GZmK purified (GM26E7)	5 µg ml-1	BioLegend	No. 370502
CD57 163Dy (HCD57)	0.1x	Fluidigm	No. 3163022B
CCR7 purified (G043H7)	5µg ml⁻¹	BioLegend	No. 353202
CD16 165Ho (3G8)	1x	Fluidigm	No. 3165001B
TCR $\gamma\delta$ purified (B1)	5 µg ml-1	BioLegend	No. 331202
CD161 purified (HP-3G10)	1.65 µg ml-1	BioLegend	No. 339902
GZmB purified (GB11)	0.2 µg ml-1	Abcam	No. ab10912
CD25169Tm (2A3)	1x	Fluidigm	No. 3169003B
CD3 170Er (UCHT1)	0.25x	Fluidigm	No. 3170001B
CD38 172Yb (HIT2)	1x	Fluidigm	No. 3172007B
TCRVa7.2 purified (3C10)	5 µg ml-1	BioLegend	No. 351702
CD14 175Lu (M5E2)	1x	Fluidigm	No. 3175015B
CD56 176Yb (NCAM16.2)	0.1x	Fluidigm	No. 3176008B
CD11b 209Bi (ICRF44)	0.33x	Fluidigm	No. 3209003B

*Plasma proteomic profiling by SomaScan 5k assay.* Plasma from WU350 and ABF300 cohorts was submitted to SomaLogic for analysis on the 5k SomaScan platform (https://www.somalogic.com/wp-content/uploads/2016/08/SSM-002-Rev-3-SOMAscan-Technical-White-Paper.pdf).

*CyTOF data analysis.* Samples were run on a CyTOF 1 mass cytometer. Data were exported into Cytobank (https://www.cytobank.org/), and individual samples were manually gated to exclude normalization beads, cell debris, dead cells and select singlet cells. Next, live CD45<sup>+</sup> singlets were gated and exported for further downstream analyses with R. The samples were stained and run over 15 different days with one identical sample that was present in every run. To correct for a batch effect, we applied batch correction using the anchor sample and 95th percentile

method<sup>60</sup>. Each batch-corrected file was subsampled with flowCore package to 20,000 events to reduce the amount of data in the aggregated dataset. After, the subsampled FCS files (excluding anchor samples) were imported in R with CATALYST package in catalyst object. All markers were arcsinh normalized with a cofactor of 5. Next, we excluded doublet cells based on coexpression of CD3/CD11b, CD3/CD19, CD56/HLA-DR, TCR $\gamma$  $\delta$ /CD11b, TBET/CD11b and CD45RA/CD45RO markers. Finally, each sample was subsampled further to 7,000 events to both reduce the number of cells and accommodate the different number of cells resulting from doublet removal.

Clustering was performed with fast PhenoGraph (FastPG function from FastPG package<sup>61</sup>, run on R 3.6) using K=140. Dimensionality reduction analysis was performed with umap function (uwot package<sup>62</sup>).

To define CD4<sup>+</sup>, CD8<sup>+</sup> and other main populations, we visualized main markers and combined PhenoGraph-identified clusters that contained corresponding markers. To reanalyze CD4, CD8, NK and B cells separately, in each case we filtered cells from the population, and rerun UMAP and PhenoGraph on these cells using relevant markers. For CD4+ cells, we used CD127, CD25, CD45RA, CD45RO, EOMES, TBET AND SELL markers with K = 140 for PhenoGraph. For CD8+ cells, we used CCR5, CCR7, CD127, CD27, CD28, CD45RA, CD57, PD-1, HLA-DR, CD38, EOMES, TBET and GZMB markers with K = 50 for PhenoGraph. For B cells, we used CCR7, CD27, CD38, SELL and TBET with K = 140 for PhenoGraph. For NK cells, we used CD16, CD57, CD56, GZMK and SELL with K = 140 for PhenoGraph. PhenoGraph is known to generate many clusters, and cluster number increases with the number of cells analyzed63. Thus, we combined some of the defined clusters to get clustering that is easier to interpret. To visualize the difference between samples, we used MDS on a matrix of samples and clusters, using cluster percentages in the matrix. MDS was calculated with the cmdscale function from the stats R package. Heat maps were created with ComplexHeatmap R package4. UMAPs and box plots were created with ggplot2 R package and adapted for publication in Adobe Illustrator.

SomaScan proteomic data analysis. For proteomic expression, we used files already standardized to the external reference. For analysis, expression values were  $\log_2$  normalized. Cohorts A–E (EDTA-treated plasma) and CV/NCV (heparin-treated plasma) were analyzed separately. Only proteins with unique gene names were considered for the analysis. To find proteins differentially expressed in cohorts A–E or CV/NCV, we used the R limma package<sup>65</sup>. GSEA and enrichment of protein signatures were performed with fgsea R package<sup>66</sup>. Volcano plots were performed with the ggplot2 package. To visualize the expression of selected genes across cohorts, expression was scaled (subtracting the mean and dividing by the standard deviation) to emphasize the difference in expression. Venn diagrams were created with the R eulerr package.

Tissue enrichment analysis with GTEx database. Data for gene expression analysis in different tissues were acquired from the open database GTEx. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health (NIH), and by NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. The data used for the analyses described in this paper were obtained from the GTEx portal in January 2021. Data included RNA-seq performed with TruSeq library construction protocol (non-stranded, polyA+ selection) for 980 donors, 52 tissue subtypes, 17,382 samples and 56,200 genes. To compare samples between each other, original GTEx-acquired read counts were converted to trimmed mean of M values, and then we calculated the median value for each gene for each tissue. To curate tissue-specific gene lists, for each gene, we calculated z-scores on median tissue values across all tissues, and tissues with values higher than three sigmas were accepted as specific for that gene. We mapped 52 resulting lists of tissue-subtype-specific genes on SomaScan, and lists were 4-640 genes long. We only used tissue-subtype-specific lists that had more than 15 genes. Downstream analysis was performed according to the GSEA described above.

*Reanalysis of public data.* The PBMC dataset was downloaded from the Gene Expression Omnibus database (GEO) under GSE107011 (ref. <sup>67</sup>) with nonnormalized count values. We uploaded the dataset to Phantasus, where we filtered out genes with a mean expression of less than 3, resulting in less than 16,000 genes. We normalized data using log + 1 and quantile normalization and used limma to perform a differential expression between neutrophils and all other groups (excluding PBMCs). We have taken 400 genes enriched in neutrophils as their signature to check for the enrichment in the CV/NCV comparison from SomaLogic data.

Liver cell atlas data<sup>58</sup> (GSE124395) and aortic data<sup>59</sup> (GSE155468) were downloaded from the GEO and processed with the Seurat package. We normalized data using the 'LogNormalize' method with a scale factor of 10<sup>4</sup>, found variable genes with the FindVariableFeatures function, and scaled data with the ScaleData function. After, we run principle-component analysis, UMAP and clustering with the FindNeighbours function (using 20 PCA dimensions) and a resolution of 0.6.

Statistics and reproducibility. No statistical method was used to predetermine sample size. Seven CyTOF samples were excluded from the analysis based on the low number of live cells identified. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome

assessment. One-way analysis of variance with post hoc Tukey's test was used to compare means between multiple groups.

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

#### Data availability

The raw and processed CyTOF data generated during this study are available at Synapse under accession code syn24239844. Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding author.

Received: 2 December 2020; Accepted: 14 April 2021; Published online: 11 May 2021

#### References

- 1. Weekly epidemiological update—2 February 2021. *World Health Organization* https://www.who.int/publications/m/item/weekly-epidemiological-update—2-february-2021 (2021).
- Guan, W. et al. Clinical characteristics of coronavirus disease 2019 in China. N. Engl. J. Med. 382, 1708–1720 (2020).
- 3. Liu, J. et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2-infected patients. *EBioMedicine* **55**, 102763 (2020).
- Wang, D. et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 323, 1061–1069 (2020).
- 5. Zhao, Q. et al. Lymphopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a systemic review and meta-analysis. *Int. J. Infect. Dis.* **96**, 131–135 (2020).
- Ruan, Q., Yang, K., Wang, W., Jiang, L. & Song, J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med.* 46, 846–848 (2020).
- Costela-Ruiz, V. J., Illescas-Montes, R., Puerta-Puerta, J. M., Ruiz, C. & Melguizo-Rodríguez, L. SARS-CoV-2 infection: The role of cytokines in COVID-19 disease. *Cytokine Growth Factor Rev.* 54, 62–75 (2020).
- Lara, P. C., Macías-Verde, D. & Burgos-Burgos, J. Age-induced NLRP3 inflammasome over-activation increases lethality of SARS-CoV-2 pneumonia in elderly patients. *Aging Dis.* 11, 756–762 (2020).
- 9. Chen, X. et al. Detectable serum severe acute respiratory syndrome coronavirus 2 viral load (RNAemia) is closely correlated with drastically elevated interleukin 6 level in critically ill patients with coronavirus disease 2019. *Clin. Infect. Dis.* https://doi.org/10.1093/cid/ciaa449 (2020).
- Del Valle, D. M. et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat. Med.* 26, 1636–1643 (2020).
- Leisman, D. E. et al. Cytokine elevation in severe and critical COVID-19: a rapid systematic review, meta-analysis and comparison with other inflammatory syndromes. *Lancet Respir. Med.* 8, 1233–1244 (2020).
- 12. Cummings, M. J. et al. Epidemiology, clinical course and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet* **395**, 1763–1770 (2020).
- Mathew, D. et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science* 369, eabc8511 (2020).
- Silvin, A. et al. Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. *Cell* 182, 1401–1418 (2020).
- Zhang, J.-Y. et al. Single-cell landscape of immunological responses in patients with COVID-19. *Nat. Immunol.* 21, 1107–1118 (2020).
- 16. Palaiodimos, L. et al. Severe obesity, increasing age and male sex are independently associated with worse in-hospital outcomes, and higher in-hospital mortality, in a cohort of patients with COVID-19 in the Bronx, New York. *Metabolism* **108**, 154262 (2020).
- Blackburn, J., Yiannoutsos, C. T., Carroll, A. E., Halverson, P. K. & Menachemi, N. Infection fatality ratios for COVID-19 among noninstitutionalized persons 12 and older: results of a random-sample prevalence study. Ann. Intern. Med. https://doi.org/10.7326/M20-5352 (2020).
- Meyerowitz-Katz, G. & Merone, L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *Int. J. Infect. Dis.* 101, 138–148 (2020).
- Salje, H. et al. Estimating the burden of SARS-CoV-2 in France. Science 369, 208–211 (2020).
- Nikolich-Žugich, J. The twilight of immunity: emerging concepts in aging of the immune system. *Nat. Immunol.* 19, 10–19 (2018).
- Lucas, C. et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* 584, 463–469 (2020).
- 22. Kuri-Cervantes, L. et al. Comprehensive mapping of immune perturbations associated with severe COVID-19. *Sci. Immunol.* **5**, eabd7114 (2020).

- Milner, J. J. & Beck, M. A. The impact of obesity on the immune response to infection. *Proc. Nutr. Soc.* 71, 298–306 (2012).
- Arevalo-Rodriguez, I. et al. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. *PLoS ONE* 15, e0242958 (2020).
- Badawi, O., Liu, X., Hassan, E., Amelung, P. J. & Swami, S. Evaluation of ICU risk models adapted for use as continuous markers of severity of illness throughout the ICU stay. *Crit. Care Med.* 46, 361–367 (2018).
- 26. Tanaka, T. et al. Plasma proteomic signature of age in healthy humans. *Aging Cell* **17**, e12799 (2018).
- Kong, M., Zhang, H., Cao, X., Mao, X. & Lu, Z. Higher level of neutrophil-tolymphocyte is associated with severe COVID-19. *Epidemiol. Infect.* 148, e139 (2020).
- Liu, J. et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. J. Transl. Med. 18, 206 (2020).
- Carissimo, G. et al. Whole blood immunophenotyping uncovers immature neutrophil-to-VD2 T cell ratio as an early marker for severe COVID-19. *Nat. Commun.* 11, 5243 (2020).
- Foy, B. H. et al. Association of red blood cell distribution width with mortality risk in hospitalized adults with SARS-CoV-2 infection. *JAMA Netw. Open* 3, e2022058 (2020).
- Salvagno, G. L., Sanchis-Gomar, F., Picanza, A. & Lippi, G. Red blood cell distribution width: a simple parameter with multiple clinical applications. *Crit. Rev. Clin. Lab. Sci.* 52, 86–105 (2015).
- Di Filippo, L. et al. Hypocalcemia is highly prevalent and predicts hospitalization in patients with COVID-19. *Endocrine* https://doi.org/10.1007/ s12020-020-02383-5 (2020).
- Singh, T. & Newman, A. B. Inflammatory markers in population studies of aging. Ageing Res. Rev. 10, 319–329 (2011).
- Kairisto, V. et al. Generation of reference values for cardiac enzymes from hospital admission laboratory data. *Eur. J. Clin. Chem. Clin. Biochem.* 32, 789–796 (1994).
- Ko, G. T. C., Wai, H. P. S. & Tang, J. S. F. Effects of age on plasma glucose levels in non-diabetic Hong Kong Chinese. *Croat. Med. J.* 47, 709–713 (2006).
- Chen, J. et al. Age-related change in thyroid-stimulating hormone: a cross-sectional study in healthy euthyroid population. *Endocr. J.* 65, 1075–1082 (2018).
- Maggio, M. et al. DHEA and cognitive function in the elderly. J. Steroid Biochem. Mol. Biol. 145, 281–292 (2015).
- Mogilenko, D. A. et al. Comprehensive profiling of an aging immune system reveals clonal GZMK<sup>+</sup> CD8<sup>+</sup> T cells as conserved hallmark of inflammaging. *Immunity* 54, 99–115 (2021).
- Zhang, F. et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* 20, 928–942 (2019).
- 40. Diao, B. et al. Reduction and functional exhaustion of T cells in patients with coronavirus disease 2019 (COVID-19). *Front. Immunol.* **11**, 827 (2020).
- Lee, J. S. et al. Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Sci. Immunol.* 5, eabd1554 (2020).
- Hashimoto, K. et al. Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proc. Natl Acad. Sci. USA* 116, 24242–24251 (2019).
- Strioga, M., Pasukoniene, V. & Characiejus, D. CD8+CD28<sup>-</sup> and CD8+CD57+ T cells and their role in health and disease. *Immunology* 134, 17–32 (2011).
- Takahashi, T. et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* https://doi.org/10.1038/s41586-020-2700-3 (2020).
- 45. Müller-Durovic, B., Grählert, J., Devine, O. P., Akbar, A. N. & Hess, C. CD56-negative NK cells with impaired effector function expand in CMV and EBV co-infected healthy donors with age. *Aging* 11, 724–740 (2019).
- Mengos, A. E., Gastineau, D. A. & Gustafson, M. P. The CD14<sup>+</sup>HLA-DR<sup>lo/neg</sup> monocyte: an immunosuppressive phenotype that restrains responses to cancer immunotherapy. *Front. Immunol.* **10**, 1147 (2019).
- 47. Remy, K. E. et al. Severe immunosuppression and not a cytokine storm characterizes COVID-19 infections. *JCI Insight* 5, e140329 (2020).
- Ryu, M. et al. LILRA3 binds both classical and non-classical HLA class I molecules but with reduced affinities compared to LILRB1/LILRB2: structural evidence. *PLoS ONE* 6, e19245 (2011).
- Wang, J. et al. Fibrinogen-like protein 1 Is a major immune inhibitory ligand of LAG-3. *Cell* 176, 334–347 (2019).
- Conte, M. et al. Human aging and longevity are characterized by high levels of mitokines. J. Gerontol. A. Biol. Sci. Med. Sci. 74, 600–607 (2019).
- Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. Nat. Med. 25, 1843–1850 (2019).
- 52. Menni, C. et al. Circulating proteomic signatures of chronological age. J. Gerontol. A. Biol. Sci. Med. Sci. 70, 809-816 (2015).

- 53. Shchukina, I. et al. Enhanced epigenetic profiling of classical human monocytes reveals a specific signature of healthy aging in the DNA methylome. *Nat. Aging* 1, 124–141 (2021).
- Williams, S. A. et al. Plasma protein patterns as comprehensive indicators of health. Nat. Med. 25, 1851–1857 (2019).
- 55. Lackner, I. et al. Midkine is elevated after multiple trauma and acts directly on human cardiomyocytes by altering their functionality and metabolism. *Front. Immunol.* **10**, 1920 (2019).
- Nakamura, T. et al. Age-related increase in Wnt inhibitor causes a senescence-like phenotype in human cardiac stem cells. *Biochem. Biophys. Res. Commun.* 487, 653–659 (2017).
- 57. Yin, J. et al. Dkk3-dependent transcriptional regulation controls age-related skeletal muscle atrophy. *Nat. Commun.* 9, 1752 (2018).
- Aizarani, N. et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* 572, 199–204 (2019).
- Li, Y. et al. Single-cell transcriptome analysis reveals dynamic cell populations and differential gene expression patterns in control and aneurysmal human aortic tissue. *Circulation* 142, 1374–1388 (2020).
- 60. Schuyler, R. P. et al. Minimizing batch effects in mass cytometry data. *Front. Immunol.* **10**, 2367 (2019).
- 61. Bodenheimer, T. et al. FastPG: fast clustering of millions of single cells. Preprint at https://doi.org/10.1101/2020.06.19.159749 (2020).
- McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at https://arxiv.org/ abs/1802.03426 (2020).
- 63. Liu, X. et al. A comparison framework and guideline of clustering methods for mass cytometry data. *Genome Biol.* **20**, 297 (2019).
- Gu, Z., Eils, R. & Schlesner, M. Complex heat maps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849 (2016).
- Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47 (2015).
- Sergushichev, A. A. An algorithm for fast preranked gene-set enrichment analysis using cumulative statistic calculation. Preprint at https://doi. org/10.1101/060012 (2016).
- 67. Monaco, G. et al. RNA-seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep.* **26**, 1627–1640 (2019).

#### Acknowledgements

This research was supported by a grant from Aging Biology Foundation to the laboratory of M.N.A. This study utilized samples obtained from the Washington University School of Medicine's COVID-19 biorepository supported by the Barnes Jewish Hospital Foundation, the Siteman Cancer Center (grant P30 CA091842 from the NCI of NIH) and the Washington University Institute of Clinical and Translational Sciences (grant UL1TR002345 from the National Center for Advancing Translational Sciences of the NIH). The content is solely the responsibility of the authors and does not necessarily represent the views of the NIH.

#### Author contributions

L.A. and M.N.A. designed the study. L.A. led ABF300 healthy patient recruitment, blood collection, processing and profiling, including CyTOF and SomaLogic profiling of WU350 samples. E.E. led computational analysis of all data. D.A.M. participated in CyTOF panel design and optimization. P.T. analyzed cell-type-specific enrichments. S.B. and A.L. processed ABF300 healthy donor samples. R.P., B.G., M.W., C.W.G., C.A.G., P.A.M., C.B. and J.A.O. performed WU350 sample collection, processing and biochemical profiling.

#### **Competing interests**

The authors declare no competing interests.

#### Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43587-021-00067-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43587-021-00067-x.

Correspondence and requests for materials should be addressed to M.N.A.

**Peer review information** *Nature Aging* thanks Deborah Dunn-Walters, Adrian Hayday and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

### RESOURCE

#### **NATURE AGING**



**Extended Data Fig. 1 Demographics and clinical characteristics for CV/NCV and healthy A-E cohorts, related to Fig. 1.** Cohorts: A, n = 36; B, n = 21; C, n = 16; D, n = 24; E, n = 24. \*p < 0.05, \*\*p < 0.01, \*\*\*\*p < 0.001, \*\*\*\*p < 0.0001, ns = not significant by one-way ANOVA with post-hoc Tuckey test. In E and F, the lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the values that are no further than 1.5 × IQR from either the upper or the lower hinge. **a**, Scatterplot for CV and NCV patients. X-axis, age, Y-axis, BMI. Patients quantified by groups of younger/older than 60yo and with more/less BMI than 30. The color indicates sex. **b**, Scatterplot for CV and NCV patients. X-axis, age, Y-axis, BMI. Patients quantified by groups of younger/older than 60yo and with more/less BMI than 30. The color indicates the mortality status of the patient. **c**, Table of the demographics for CV/NCV cohorts. **d**, Table of the demographics for A-E cohorts. **e**, Distribution of clinical severity scores for ICU admitted patients. CV\_Severe, n = 17; CV\_Deceased, n = 11. **f**. Boxplot for C-reactive protein (CRP) value from clinical blood panel for CV patients cohort. CV\_Moderate, n = 13; CV\_Severe, n = 19; CV\_Deceased, n = 9. **g-i**, Heatmaps, representing significance for pairwise comparisons of the means for WBC counts (**G**), red blood cell counts and characteristics (**H**), selected clinical values (**I**) between cohorts.

# RESOURCE



Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2** | **Biochemical characteristics of healthy aging, related to Fig. 1**. In A, C, E and G, the lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the values that are no further than 1.5 × IQR from either the upper or the lower hinge. \*p < 0.05, \*\*p < 0.01, \*\*\*\*p < 0.001, \*\*\*\*p < 0.0001, ns = not significant by one-way ANOVA with post-hoc Tuckey test. **a-b**, Boxplots of clinical blood characteristics increasing with age. **b**, heatmaps representing significance for pairwise comparisons between A-E cohorts. **c-b**, Boxplots of clinical blood characteristics with increase/decrease or decrease/increase pattern with age. **f**, heatmaps representing significance for pairwise comparisons between A-E cohorts. **e-f**, Boxplots of clinical blood characteristics of clinical blood characteristics with increase/decrease or decrease/increase pattern with age. **f**, heatmaps representing significance for pairwise comparisons between A-E cohorts. **g-h**, Boxplots of clinical blood characteristics of clinical blood characteristics with age. **d**, heatmaps, representing significance for pairwise comparisons between A-E cohorts. **e-f**, Boxplots of clinical blood characteristics with increase/decrease or decrease/increase pattern with age. **f**, heatmaps representing significance for pairwise comparisons between A-E cohorts. **g-h**, Boxplots of clinical blood characteristics without a clear age-related change pattern. **h**, heatmaps representing significance for pairwise comparisons between A-E cohorts.

# RESOURCE







**Extended Data Fig. 4 | CyTOF markers and population dynamics for CD4 T cells, related to Fig. 2. a**, CD4 T cells UMAP plots for all CyTOF markers. **b**, Heatmaps, representing significance for pairwise comparisons of the CD4 T cell cluster percentages between cohorts. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*p < 0.001, ns = not significant by one-way ANOVA with post-hoc Tuckey test.

### RESOURCE

### A CD8 T cells





**Extended Data Fig. 5** | **CyTOF** markers and population dynamics for CD8 T cells, related to Fig. 4. a, CD8 T cells UMAP plots for all CyTOF markers. b, Heatmaps, representing significance for pairwise comparisons of the CD8 T cell cluster percentages between cohorts. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, s = not significant by one-way ANOVA with post-hoc Tuckey test. **c**, Cell proportion for DR+CD38+ CD8 T cells cluster across cohorts split by sex. A, n = 38; B, n = 28; C, n = 20; D, n= 29; E, n = 33; NCV, non-COVID-19, n = 17; CV\_Moderate, n = 18; CV\_Severe, n = 18; CV\_Deceased, n = 12. The lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the values that are no further than 1.5 × IQR from either the upper or the lower hinge. **d**, Distribution of CD8 T cells between GZMK and CCR5 markers, one patient. **e**, Distribution of CD8 T cells between GZMK and CD57 markers, one patient.

#### **NATURE AGING**



Extended Data Fig. 6 | See next page for caption.

**Extended Data Fig. 6 | CyTOF markers and population dynamics for NK and B cells, related to Fig. 5.** \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*p < 0.001, \*p < 0.001

#### **NATURE AGING**



**Extended Data Fig. 7** | Inflammatory remodeling of myeloid cells a, UMAP plot of all myeloid cells, colored by the cluster. **b**, UMAP plots with the expression of selected markers. **c**, MDS projection for all samples, colored by cohort. For each sample, cluster percentages were used to do MDS. **d**, UMAP density plots characterizing the distribution of NK cells across conditions. **e**, Cell proportions of each NK cell cluster across cohorts. The lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the values that are no further than  $1.5 \times IQR$  from either the upper or the lower hinge. A, n = 38; B, n = 28; C, n = 20; D, n = 29; E, n = 33; NCV, non-COVID-19, n = 17; CV\_Moderate, n = 18; CV\_Severe, n = 18; CV\_Deceased, n = 12. **f**, Heatmaps representing significance for pairwise comparisons of the B cell cluster percentages between cohorts. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001, ns = not significant by one-way ANOVA with post-hoc Tuckey test. G. UMAP plots for all CyTOF markers.

# RESOURCE



Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | SomaLogic protein plasma analysis of age- and inflammation-related features, related to Fig. 6.** Cohorts: A, n = 42; B, n = 27; C, n = 18; D, n = 29; E, n = 34; NCV, non-COVID-19, n = 27; CV\_Moderate, n = 18; CV\_Severe, n = 21; CV\_Deceased, n = 14. In A-C, E, G and H, the lower and upper hinges of all boxplots represent the 25th and 75th percentiles. Horizontal bars show the median value. Whiskers extend to the values that are no further than 1.5 × IQR from either the upper or the lower hinge. D and G, P values were calculated using the limma package (two-sided test). Significant genes were selected after correction for multiple testing using the Benjamini-Hochberg method. **a**, Boxplots of average expression per sample up- and downregulated in CV Moderate cohort (**A**), CV Severe cohort (**B**), CV Deceased cohort (**C**), compared to NCV cohort, across CV/NCV cohorts. **d**, Heatmap with the number of differentially expressed genes (p.adj < 0.05) between CV/NCV cohorts. **e**, Boxplots of z-score expression of proteins significantly up- or downregulated with age, across A-E cohorts. **f**, Selected correlations between protein supregulated in COVID-19-related inflammation (CV vs. NCV comparison). Boxplots with scaled protein expression across A-E and NCV-CV cohorts. **h**, Overlap between proteins upregulated with age (E vs. A comparison) versus proteins upregulated with acle differential inflammation (CV vs. NCV comparison). Boxplots with scaled protein expression across A-E and NCV-CV cohorts. **h**, Overlap between proteins upregulated with age (E vs. A comparison) versus proteins upregulated with ace (E vs. A comparison) versus proteins upregulated with ace (E vs. A comparison) versus proteins upregulated with ace (E vs. A comparison) versus proteins upregulated with ace (E vs. A comparison) versus proteins upregulated with ace (E vs. A comparison) versus proteins upregulated with ace (E vs. A comparison) versus proteins upregulated with ace (E vs. A comparison) versus proteins upregulated in COVI

# nature research

Corresponding author(s): Maxim N. Artyomov

Last updated by author(s): 4/1/2021

# **Reporting Summary**

Nature Research 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 Research policies, see our Editorial Policies and the Editorial Policy Checklist.

### Statistics

For	all st	atistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.
n/a	Cor	firmed
	$\boxtimes$	The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
	$\boxtimes$	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
	$\boxtimes$	The statistical test(s) used AND whether they are one- or two-sided Only common tests should be described solely by name; describe more complex techniques in the Methods section.
	$\boxtimes$	A description of all covariates tested
	$\boxtimes$	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
	$\boxtimes$	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)
	$\boxtimes$	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 <i>Give P values as exact values whenever suitable</i> .
$\boxtimes$		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
	$\boxtimes$	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
$\boxtimes$		Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
	'	Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.

icy information	about <u>availability of computer code</u>
Data collection	No software was used for data collection.
Data analysis	CyTOF data analysis. Samples were run on a CyTOF 1 mass cytometer. Data was exported into Cytobank (https://www.cytobank.org) and individual samples were manually gated to exclude normalization beads, cell debris, dead cells, and select singlet cells. Next, live CD45+ singlets were gated and exported for further downstream analyses with R (v 3.6.2 and 4.0.1). The samples were stained and run over 15 different days with one identical sample that was present in every run. To correct for a batch effect we applied batch correction using the anchor sample and 95 percentile method (Schuyler et al., 2019). Each batch-corrected file was subsampled with flowCore package (v 2.0.1) t 20,000 events to reduce the amount of data in the aggregated dataset. After, the subsampled FCS files (excluding anchor samples) were imported in R with CATALYST package (v 1.12.2) in catalyst object. All markers were arcsinh normalized with cofactor of 5. After, we excludeed doublet cells based on co-expression of CD3/CD11b, CD3/CD19, CD56/HLA-DR, TCRgd/CD11b, T-BET/CD11b, CD45RA/CD45RO markers. Finally, each sample was subsampled further to 7,000 events to both reduce the number of cells and accommodate the different number of cells resulted from doublet removal. Clustering was performed with fast Phenograph (FastPG function from FastPG package (Bodenheimer et al., 2020), run on R 3.6) using K = 140. Dimensionality reduction analysis was performed with umap function (uwot package, v 0.1.10, McInnes et al., 2020). To define CD4, CD8, and other main populations we visualized main markers and combined Phenograph-identified clusters that contained corresponding markers. To re-analyze CD4, CD8, NK and B cells separately, in each case with filtered cells from the population, and re-run UMAP and Phenograph on these cells using relevant markers. For CD4, we used CD127, CD25, CD45RA, CD45RO, EOMES, T-BET, SELL marker with K=140 for Phenograph. For CD8, we used CCR7, CD27, CD28, CD45RA, CD57, PD-1, HLA-DR, CD38, EOMES, T-BET, and GZMB markers with K=

ComplexHeatmap R package (v2.4.3, Gu et al., 2016). UMAPs and barplots were created with ggplot2 R package (v 3.3.2) and adapted for publication in Adobe Illustrator (v 22.1).

SomaScan proteomic data analysis. For proteomic expression, we got files already standardized to the external reference. For analysis, expression values were log2-normalized. Cohorts A-E (EDTA treated plasma) and CV/NCV (heparin treated plasma) were analyzed separately. Only proteins with unique gene names were considered for the analysis. To find proteins differentially expressed in A-E cohorts or CV-NCV cohorts we used the R limma package (v 3.44.3, Ritchie et al., 2015). Gene set enrichment analysis and enrichment of protein signatures was done with fgsea R package (v 1.14.0, Sergushichev, 2016). Volcano plots were done with ggplot2 package. To visualize the expression of selected genes across cohorts, expression was scaled (subtracting the mean and dividing by standard deviation) to emphasize the difference or absence of it in expression. Venn diagrams were done with R eulerr package (v 6.1.0).

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 Research guidelines for submitting code & software for further information.

#### Data

Policy information about availability of data

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable: - Accession codes, unique identifiers, or web links for publicly available datasets

- Accession codes, unique identifiers, or web links for publicity
   A list of figures that have associated raw data
- A description of any restrictions on data availability

The accession number for the raw and processed CyTOF data generated during this study is Synapse: syn24239844.

# 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.

 If sciences
 Behavioural & social sciences

 For a reference copy of the document with all sections, see <a href="mature.com/documents/nr-reporting-summary-flat.pdf">nature.com/documents/nr-reporting-summary-flat.pdf</a>

# Life sciences study design

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

Sample size	No sample size calculation was performed. Sample size was chosen based on the availability of healthy donor's samples and numbers of donors with disease were matched to numbers in healthy cohorts.
Data exclusions	Seven CyTOF samples were excluded from the analysis based on low number of live cells identified.
Replication	CyTOF data was collected over the course of 15 different days with samples from each donor cohort present on each day. Separate runs were consistent between each other without any batch effect.
Randomization	All available samples from healthy donors in ABF300 cohort were allocated into age groups. SARS-CoV-2 positive and negative WU350 samples were randomly chosen from all available samples that satisfied the BMI and age criteria.
Blinding	This study was an observational cohort study, so blinding was not applicable. No intervention was tested.

# 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/	ρ	t٢	າດ	Ч	c
111	e	ιı	IU	u	2

n/a	Involved in the study	n/a Involved in the study	
	X Antibodies	ChIP-seq	
$\boxtimes$	Eukaryotic cell lines	Flow cytometry	
$\ge$	Palaeontology and archaeology	MRI-based neuroimaging	
$\boxtimes$	Animals and other organisms		
	🔀 Human research participants		
$\boxtimes$	Clinical data		

### Antibodios

Antibodies	
Antibodies used	<ol> <li>CD19 142Nd (HIB19), Fluidigm, catalogue number: 314201B</li> <li>CD127 143Nd (A019D5), Fluidigm, catalogue number: 314401B</li> <li>CD63 144Nd (FNSO), Fluidigm, catalogue number: 314401B</li> <li>CD61 143Nd (FNA-T8), Fluidigm, catalogue number: 3146001B</li> <li>CD11 147Sm (Bu15), Fluidigm, catalogue number: 3146001B</li> <li>CD11 147Sm (Bu15), Fluidigm, catalogue number: 314000B</li> <li>CD41 418Nd (FNA-T8), Fluidigm, catalogue number: 314000B</li> <li>CD45R0 149Sm (UCH11), Fluidigm, catalogue number: 3149001B</li> <li>CCAS Purified (J418F1), Biolegend, catalogue number: 3151023B</li> <li>EOMES 152Sm (WD1928), Invitrogen, catalogue number: 14-4877-82</li> <li>SELL 153Eu (DREG-56), Fluidigm, catalogue number: 3151023B</li> <li>CD45 154Sm (H130), Fluidigm, catalogue number: 315001B</li> <li>CD45 154Sm (H130), Fluidigm, catalogue number: 316001B</li> <li>CD45 154Sm (H130), Fluidigm, catalogue number: 316001B</li> <li>CD45 154Sm (H130), Fluidigm, catalogue number: 316001B</li> <li>CD7 158Gd (L128), Fluidigm, catalogue number: 316002B</li> <li>GZmK Purified (GM26T), Biolegend, catalogue number: 316002B</li> <li>GZmK Purified (GM347), Biolegend, catalogue number: 316302B</li> <li>GZmK Purified (GM347), Biolegend, catalogue number: 316302B</li> <li>CD5 163D4 (H26347), Biolegend, catalogue number: 316302B</li> <li>CD5 165H6 (368), Fluidigm, catalogue number: 316302B</li> <li>CD16 1047H6 (H19-3610), Biolegend, catalogue number: 316302B</li> <li>CD16 1047H6 (H19-3610), Biolegend, catalogue number: 316302B</li> <li>CD25 169TM (A33), Fluidigm, catalogue number: 316302B</li> <li>CD2</li></ol>
Validation	Metal-conjugated antibodies were purchased from Fluidigm when available. Fluidigm provides the following validation statement for each antibody: Each lot of conjugated antibody is quality control tested by CyTOF <sup>®</sup> analysis of stained cells using the appropriate positive and negative cell staining and/or activation controls. For all other targets, purified antibodies were obtained and conjugated using the appropriate Maxpar Antibody Labeling Kit (Fluidigm) according the manufacturer's protocol.

### Human research participants

Policy information about studies involving human research participants

Population characteristics	Inclusion criteria for WU-350 required that subjects were symptomatic and had a physician-ordered SARS-CoV-2 test performed in the course of their normal clinical care. Some subjects were enrolled prior to the return of the SARS-CoV-2 test result. Enrolled subjects who tested negative for SARS-CoV-2 are included in the current manuscript as non-Covid-19 respiratory illness controls. All samples were collected during evaluation for symptoms in a medical facility or within 36 hours of subject admission to the hospital. Patient-reported duration of illness and other clinically relevant medical information were collected at the time of enrollment from the medical record and the subject or their legally authorized representative. Subjects were 19-90 years with approximately equal representation of male and female individuals. For ABF300 cohort, adults 25 to 80 years of age, non-obese (BMI less than 30), non-smokers without history of cancer, chronic inflammatory conditions (arthritis, Crohn's disease, colitis, dermatitis, fibromyalgia, or lupus), or blood borne infections were included. Subjects who reported cold or flu symptoms in the prior month were excluded. Samples from healthy subjects were collected in 2018 to early 2019, eliminating the possibility of concurrent or prior SARS-CoV-2 infection.
Recruitment	The Washington University 350 (WU350) study is a prospective observational cohort study of subjects with viral respiratory illness symptoms who presented to Barnes Jewish Hospital, St. Louis Children's Hospital, or affiliated Barnes Jewish Hospital testing sites located in Saint Louis, Missouri, USA. As enrollment is based on individuals who sought medical care, this study design results in an absence of mild or asymptomatic SARS-CoV-2 infections in the study cohort. For ABF300 cohort, adults 25 years of age and older were recruited from the St Louis area and provided written informed consent to participate. Subjects were given a screening questionnaire to establish health status. Self-selection bias should have minimal to no effect on outcomes as healthy individuals who chose to participate in the study should be representative of healthy individuals of their age group.
Ethics oversight	WU350 was reviewed and approved by the Washington University in St Louis Institutional Review Board (WU-350 study approval # 202003085). The study complied with the ethical standards of the Helsinki Declaration. The Washington University in St Louis Institutional Review Board reviewed and approved a ABF300 study for the collection of blood samples from healthy subjects (IRB approval #201804084).

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