Loss of Y in leukocytes as a risk factor for critical COVID-19 in men

Bożena Bruhn-Olszewska^{1,#}, Hanna Davies^{1,#}, Daniil Sarkisyan^{1,#}, Ulana Juhas², Edyta Rychlicka-Buniowska², Magdalena Wójcik², Monika Horbacz², Marcin Jąkowski², Paweł Olszewski², Jakub O. Westholm³, Agata Smiałowska³, Karol Wierzbą⁴, Åsa Torinsson Nalwai⁵, Niklas Jern⁵, Lars-Magnus Andersson^{6,11}, Josef D. Järhult⁷, Natalia Filipowicz², Eva Tiensuu Janson⁸, Sten Rubertsson⁹, Miklós Lipcsey^{9,10}, Magnus Gisslén^{6,11}, Michael Hultström^{9,12*}, Robert Frithiof^{9*}, Jan P. Dumanski^{1,2,*,&}

¹ Department of Immunology, Genetics and Pathology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

² 3P-Medicine Laboratory, Medical University of Gdańsk, Dębinki 7, 80-211 Gdańsk, Poland.

³ National Bioinformatics Infrastructure Sweden, Department of Biochemistry and Biophysics, Stockholm University, Science for Life Laboratory, Stockholm, Sweden.

⁴ Department of Internal Medicine, Connective Tissue Diseases and Geriatrics, Medical University of Gdańsk, Gdańsk, Poland

⁵ Department of Laboratory Medicine, Institute of Biomedicine and Biobank Core Facility, Sahlgrenska Academy at the University of Gothenburg, Gothenburg Sweden.

⁶ Department of Infectious Diseases, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

⁷ Zoonosis Science Center, Department of Medical Sciences, SE-75185, Uppsala, Sweden.

⁸ Department of Medical Sciences, Endocrine Oncology Unit, Uppsala University, Uppsala, Sweden.

⁹ Department of Surgical Sciences, Anesthesiology and Intensive Care, Uppsala University, Uppsala, Sweden.

¹⁰ Hedenstierna laboratory, Department of Surgical Sciences, Uppsala University, Sweden.

¹¹ Department of Infectious Diseases, Region Västra Götaland, Sahlgrenska University Hospital, Gothenburg, Sweden.

¹² Integrative physiology, Department of Medical Cell Biology, Uppsala University, Uppsala, Sweden.

- shared first authors with equal contributions

* - shared senior authors

& - corresponding author

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Abstract:

COVID-19 shows an unexplained, strong male bias for severity and mortality. Loss of Y (LOY) in myeloid cells is a risk factor candidate in COVID-19 because of associations with many age-related diseases and its effect on transcription of immune genes. We report the highest levels of LOY in cells that are crucial for the development of severe COVID-19 phenotype, such as low-density neutrophils, granulocytes, and monocytes reaching 46%, 32%, and 29%, respectively, from men with critical COVID-19 (n=139). LOY in sorted subpopulations of leukocytes correlated with increased thrombocyte count, thromboembolic events, invasive mechanical ventilation and a history of vessel disease. In recovered patients, LOY decreased in whole blood and peripheral blood mononuclear cells. Moreover, sc-RNA-seq analysis of CD14⁺ monocytes from 30 COVID-19 males and 34 controls revealed pervasive transcriptional downregulation in LOY-cells, notably affecting HLA class I and II genes important for antigen presentation. The data support a link between LOY and emergency myelopoiesis as well as the role of LOY in modulation of COVID-19 severity. Our results might also be relevant for other viral infections showing similar male bias.

Leukocytes from aging males frequently show mosaic loss of chromosome Y (LOY) [1, 2]. It is detectable in whole blood DNA from >40% of the men above the age of 70 years [3] and reaches 57% in the analysis of 93-year-old men, indicating an increasing frequency with age [4]. Recent single-cell analyses of peripheral blood mononuclear cells (PBMCs) from 29 aging men (median age 80 years, range 64-94 years) identified cells with LOY in every studied subject, making this the most common post-zygotic mutation [5]. In serial analysis of male blood samples LOY is a dynamic process [6]. LOY is most common in leukocytes, but has also been shown in other tissues although with considerably lower frequencies [4, 7]. Risk factors for LOY include age, smoking and germline predisposition [1-3, 8]. LOY in whole blood has been associated with increased risk for all-cause mortality, hematological and non-hematological cancers and other age-related disorders such as Alzheimer's disease, diabetes and cardiovascular events, among others [1, 2, 7, 9-14]. Thus, carriers of LOY have an increased risk for diseases both inside and outside of the hematopoietic system and the mechanism(s) behind these associations largely remain to be elucidated. Recent studies suggest that LOY could have a direct physiological role through LOY-associated transcriptional effects (LATE) on global gene expression in a pleiotropic manner as well as being a biomarker of genomic instability in somatic tissues [5]. Moreover, dysregulation of immune genes, including pathways related to viral life-cycle, was pronounced in LOY-cells [5, 15].

Coronavirus disease 2019 (COVID-19) is caused by the new SARS-CoV-2 virus [16]. Critically affected patients present with bilateral pneumonia, in many cases progressing to acute respiratory distress syndrome with a strong decrease in pulmonary function requiring treatment with non-invasive or invasive mechanical ventilation [17]. Results suggest that neutrophil activation is a hallmark of COVID-19 together with lymphocytopenia and that patients with the critical disease show substantial elevation of circulating neutrophils [18-20]. In COVID-19 patients, neutrophils adopt a so-called low-density phenotype, when they are activated and prone to spontaneous release of neutrophil extracellular traps (NETs) in capillaries. Intra-vascular aggregations of NETs lead to occlusion of the affected vessels, disturbed microcirculation and organ damage. Furthermore, hyper-coagulation and thromboembolic events are also contributing to the mortality from COVID-19 [21, 22]. These symptoms are possibly a consequence of dysfunction in myeloid cell lineages and growing evidence corroborate this hypothesis [23, 24]. Another feature of the disease is a strong and unexplained bias of males who are critically ill and die from COVID-19. Of 7874 patients treated and registered since March 2020 at Intensive Care Units (ICU) in Sweden, 70.1% were men (Swedish Intensive Care Database [SIR], data from September 9, 2021). Other studies of COVID-19 patients confirm this strong male bias [25, 26] and this aspect of COVID-19 pathogenesis is understudied [27]. Thus, there might exist a so far unidentified male-specific risk factor, predisposing men for a life-threatening disease course of COVID-19. We hypothesized that LOY in cells from myeloid lineage might be linked to the development of the critical disorder and predominantly studied here critically ill COVID-19 males.

Myeloid lineage cells from COVID-19 patients show the highest levels of LOY

We initially reanalyzed the publicly available single-cell RNA sequencing dataset [18], adding LOY-status for all sequenced cells as described [5] (**Fig. 1** and **Fig. S1**). All critically ill COVID-19 patients (WHO grade ≥ 5) described in Schulte-Schrepping et al, 2020 were included; i.e. 6 male patients with a total of 9 samples (three patients were sampled twice) taken during ICU stay. In this re-analysis, we used two previously published cell type annotations [18, 28] with concordant results. Neutrophils, immature neutrophils (also called low-density neutrophils, LDNs) and other myeloid cells clearly showed the highest levels of LOY.

In order to extend the above initial analysis, we collected ~16 ml blood from 139 critically ill male COVID-19 patients, treated at two ICUs at University Hospitals (Uppsala and Gothenburg) in Sweden. The criterion for inclusion in our study as a critical COVID-19 patient was treatment at ICU (WHO grade 6 or higher). The median age for this cohort was 65 years, ranging 19-86 years. We also collected history of smoking habits, comorbidity data, as well as numerous clinical and biochemical parameters recorded during ICU-treatment (**Table S1**). We further designed a new protocol for cell-sorting using FACS, allowing analysis of LOY in LDNs, granulocytes, monocytes and PBMCs as well as in whole blood for all patients. We also analyzed 38 healthy controls, which were used for cell-sorting, followed by LOY analysis (n=17), as well for single-cell transcriptome analysis (n=34). The median age of controls was 71 years, ranging 61-91 years. Moreover, we studied four male subjects with a milder course of COVID-19, i.e. hospitalized patients with confirmed SARS-CoV-2 infection but with no signs of severe disease at the time of sampling.

The number of LDNs varied most among all cell types that were analyzed. Using FACS, we counted LDNs in all critically ill COVID-19 patients. Their median number (per 10,000 FACS events, unadjusted for age and smoking) was 60, ranging 1-458. The corresponding range of numbers for LDNs in milder COVID-19 patients (n = 4) and healthy controls (n=17) were 1-73 and 4-70 per 10,000 FACS events, respectively. **Figure 2** shows medians, 80% and 95% highest density intervals (HDI) for LDN cell counts, adjusted for age, age² and smoking. The LDN cell counts in critically ill patients were significantly higher compared to healthy controls (median 86.8, 95%-HDI 61.8–119.1, $p=1e-9$) and to milder COVID-19 patients (median 63.1, 95%-HDI 2.6–107.2, $p=.0089$).

The analysis of LOY status for all samples in this study was performed using droplet-digital PCR (ddPCR) as described [6]. The advantage of using ddPCR was a low input of DNA (~50 ng) required to perform the LOY measurement, which was particularly important for analysis of LDNs from COVID-19 patients having low counts of these cells. Another advantage was a higher sensitivity in the lower range of measurements allowing to score LOY with 5% cut-off for cells with the mutation (**Fig. 3**). We were able to successfully detect LOY in DNA from LDNs sorted on FACS starting at about 10,000 cells. It should be stressed, however, that the

number of successful measurements of LOY in LDNs was considerably lower than for other sources of DNA, due to the low number of LDNs from some COVID-19 patients. LOY percentages for all ICU-patients in five different cell populations (LDNs, granulocytes, monocytes, PBMCs and whole blood) were compared (**Fig. 3**). Adjusting for age, age² and smoking the %LOY in LDNs was significantly higher than in monocytes (median 1.83%, 95%-HDI 0.94%–2.72%, MVT adjusted $p=5e-4$), in PBMCs (median 1.48%, 95%-HDI 0.58%–2.41%, $p=0.013$) and in blood (median 1.38%, 95%-HDI 0.5%–2.27%, $p=0.018$), and %LOY in granulocytes was significantly higher than in monocytes (median 0.75%, 95%-HDI 0.22%–1.26%, $p=0.037$). **Figures 3B** and **S2** show unadjusted values, also pointing to the highest level of LOY in LDNs compared with other cell types among critically ill patients. The range of cells with LOY also varied in different cell populations. The highest percentages of cells with LOY per subject in the five cell types were as follows: granulocytes, 85.5%; monocytes, 85.4%; whole blood, 81.7%; PBMCs, 51.5%; and LDNs, 36%.

LOY decreases in COVID-19 patient samples at the recovery stage

We further collected follow-up blood samples from 17 critically ill patients who recovered from the disease. The median number of days when the follow-up specimens were taken (from the date of first sampling) was 119 days, ranging 93-143. These samples were processed in the same way as described above for all critically ill patients. **Figure 4A** shows medians, 80%- and 95%-HDI for estimates of LDN cell counts, adjusted for age, age² and smoking, demonstrating a radical decrease in the follow-up specimens (median 76.4, 95%-HDI 55.5-104, $p=6e-11$). Paired comparisons of cell numbers from LDNs and monocytes in 16 patients, where both ICU and recovery samples were available, also showed statistically significant decrease in the follow-up samples for LDNs ($p=0.0066$) and monocytes ($p=0.013$) using non-parametric paired Mann-Whitney-Wilcoxon test (**Fig. 4B**).

We also performed paired LOY analysis in the five sorted populations of cells in 16 patients, where both ICU and recovery samples were available. Because the number of sorted LDNs from recovery samples were too low for DNA isolation (below 10,000 cells isolated via FACS from 16 ml of blood), we studied LOY only in the four other cell types; i.e. whole blood, PBMCs, granulocytes and monocytes. The %LOY values in DNA from whole blood ($p=0.29$) and PBMCs ($p=0.25$) decreased significantly in samples taken at the recovery stage using non-parametric paired Mann-Whitney-Wilcoxon test (**Fig. 4B**). In conclusion, the results from comparisons between ICU and recovery samples suggest the overall decrease of cell numbers that are mostly associated with the critical course of COVID-19 (LDNs and monocytes) as well as the overall decrease in the load of cells with LOY. The latter result underscores the dynamic character of aberrant cell clones with LOY that are present in COVID-19 patients.

LOY correlate to clinical variables and co-morbidity in COVID-19 patients

Previous epidemiological studies showed an association of LOY with blood cell differentiation; that is thrombocyte- and erythrocyte cell counts were positively and negatively associated with LOY, respectively [29, 30]. In particular, binding sites of FLI1, a fate-determining factor promoting hematopoietic stem cell differentiation into thrombocytes rather than erythrocytes, showed a significant heritability enrichment of LOY GWAS signals using ChIP-seq data [29]. This effect of LOY might be important for the clinical outcome of critically ill male COVID-19 patients that display LOY, since increased counts of thrombocytes, hyper-coagulation and thromboembolic events are contributing to the high mortality in COVID-19 [21, 22]. We have therefore analyzed in our cohort of critically ill patients the correlation of LOY with the cell counts of thrombocytes and erythrocytes. For the patients treated at ICU in Uppsala, the cell counts for thrombocytes and erythrocytes as well as hemoglobin measurements were collected at the day of admission to ICU as well as the minimum- and the maximum values for these cells during the entire length of ICU stay. We estimated robust correlations of %LOY in five studied cell populations versus the above-mentioned clinical variables (**Fig. 5A** and **Table S1**). After adjusting for age, age² and smoking status, the %LOY in PBMCs was significantly negatively correlated with maximum erythrocyte count ($R = -0.25$, 95%-HDI -0.45 – 0.04) and with hemoglobin measurement, both at day 1 ($R = -0.21$, 95%-HDI -0.4 – 0.007) and maximum values (-0.23 , 95%-HDI -0.41 – 0.02). In contrast, maximum thrombocyte counts were significantly positively correlated with %LOY in monocytes ($R = 0.023$, 95%-HDI 0.02 – 0.42), granulocytes ($R = 0.023$, 95%-HDI 0.02 – 0.43), and whole blood ($R = 0.24$, 95%-HDI 0.04 – 0.43). Increased counts of thrombocytes in critically ill patients that have higher levels of LOY may predispose them for thromboembolic complications during ICU stay. Thus, we tested this variable against %LOY in blood and sorted subsets of leukocytes. Indeed, %LOY in LDNs (median 4.9%, 95%-HDI 0.29 – 9.6 , $p = 0.041$) and monocytes (median 2.7%, 95%-HDI 0.54 – 4.8 , $p = 0.013$) positively correlated with clinically confirmed thromboembolic events during ICU stay (**Fig. 5B**).

The necessity to apply invasive mechanical ventilation (intubation and treatment in respirator) for some COVID-19 patients is one of the most severe consequences of this infection. We studied the possible association of this variable with %LOY, adjusting for age, age² and smoking. We grouped our critically ill patients (**Table S1**) into those who did and did not receive this form of treatment. The %LOY in blood was significantly higher in patients treated with invasive mechanical ventilation (median 1.15%, 95%-HDI 0.04% – 2.3% , $p = 0.048$) and showed a tendency in the same direction also in monocytes (median 0.99%, 80%-HDI 0.17% – 1.77%) (**Fig. 5C**). This result suggests that %LOY might have a predictive value for the identification of patients that are at high risk for treatment with invasive mechanical ventilation and is another indication that LOY may contribute to severity of COVID-19.

We also tested a number of other clinical parameters (**Table S1**) that were collected for our critical COVID-19 cohort for correlation with %LOY, among them information about pre-existing comorbidities, prior to ICU admission. The co-morbidity, named “Vessel Disease (VD)”, was defined as the presence in the medical history of patients of any of the following conditions: coronary heart disease, stroke, diagnoses related to central vessels (aorta aneurysm, aorta stenosis, etc.) and diseases of peripheral vessels (intermittent claudication, skin ulcers of legs, etc). **Figure 5D** shows that, adjusted for age, age² and smoking, the %LOY was significantly higher in VD patients in blood (median 4.0%, 95%-HDI 2.0%-5.8%, $p=4e-5$), in granulocytes (median 2.8%, 95%-HDI 0.52%-4.8%, $p=0.013$) and in PBMCs (median 2.0%, 95%-HDI 0.64%-3.26%, $p=0.0033$). Thus, we show that %LOY in blood is associated with common comorbidities present among COVID-19 patients treated at ICUs.

sc-RNA-seq associates LOY and COVID-19 with a predominant down-regulation of genes in monocytes

We applied here single-cell transcriptomics using 10X Genomics platform for PBMCs from 30 males with critical COVID-19 and 34 healthy male controls to compare differentially expressed genes (DEGs) between cells with LOY and in normal state (abbreviated as non-LOY) in classical CD14⁺ monocytes. They were selected for the differential gene expression analysis because dysfunction of monocytes has been linked to pathogenesis of COVID-19 [18]. Furthermore, monocytes are abundant in the circulation of controls and expand in numbers in the COVID-19 patients. Yet another reason is our above described results showing that monocytes are frequently affected by LOY in COVID-19 patients and %LOY in monocytes displayed interesting correlations with clinical parameters. Cell types were annotated using Azimuth workflow as described [31, 32]. Cells were further classified as LOY-cells when they had no detectable expression from genes located in male specific region of chromosome Y (MSY) [5]. We performed three pairwise comparisons between gene expression levels at the baseline and three other groups. The baseline was the data from 22,676 non-LOY CD14⁺ monocytes from the 34 controls. This was compared with 3,741 LOY cells in the 34 controls (LOY effect), 41,486 non-LOY cells from 30 COVID-19 patients (COVID-19 effect) and with 6,015 LOY cells from COVID-19 patients (combined effect). We excluded MSY genes from these comparisons because their expression was absent in the single-cell data; i.e. this was the definition of LOY for single-cells [5]. The predominant pattern observed for top 500 DEGs was down-regulation of gene expression (**Fig. S3**, **Fig. S4** and **Fig. 6**). The detailed results from the top 100 DEGs, including statistical significance for either down- or up-regulation in the three comparisons is illustrated in **Fig S4**. Similar analysis for top 500 DEGs is summarized in **Table S2**.

Furthermore, as is apparent from **Figs. 6** and **S4**, there was a dominant profile among down-regulated genes, with a consecutive and seemingly additive decrease of expression in each of the three pairwise comparisons.

Genes showing a concordant down-regulation (average log fold-change <-0.1 in each of the three comparisons) were labeled as having a "down" pattern. There were 21 “down” pattern genes among the top 100 DEGs using this threshold. Similarly, genes showing a concordant up-regulation (average log fold-change >0.1 in each of the three comparisons) were labeled as having an "up" pattern. There were only six “up” pattern genes in the top 100 DEGs (**Fig. S4, Table S2**). The largest category of down-regulated genes was related to ribosome biogenesis/assembly, protein translation and protein targeting to membrane. Among the top 100 DEGs, there were 27 genes for ribosomal proteins and translation elongation or initiation factors. The second largest category of down-regulated genes among the top 100 (or top 500) DEGs are connected to antigen processing and presentation, predominantly via HLA class II genes and to a lesser degree through HLA class I system.

We further performed pathway enrichment analysis in Gene Ontology (biological process and molecular function) as well as in WikiPathway using 21 strictly defined “down” pattern genes from the top 100 DEGs (**Table S2**). As expected, top-enriched pathways were: translation (GO:0006412); peptide biosynthetic process (GO:0043043); MHC class II receptor activity (GO:0032395); CD4 receptor binding (GO:0042609); RNA binding (GO:0003723); MHC class II protein binding (GO:0042289); cytoplasmic ribosomal proteins (WP477); and pathogenesis of SARS-CoV-2 (WP4884). Essentially identical results were obtained using 58 strictly defined “down” pattern genes selected from the top 500 DEGs. Furthermore, there was no difference in results obtained when down-regulated MSY genes (due to LOY) were added into the enrichment analysis (**Table S2**). The corresponding analysis of up-regulated genes in Gene Ontology: Biological Processes (using six strictly defined “up” pattern genes from the top 100 DEGs, alternatively using 38 “up” pattern genes from the top 500 DEGs) revealed the following three most enriched pathways: neutrophil degranulation (GO:0043312); neutrophil activation (GO:0002283); neutrophil mediated immunity (GO:0002446) (**Table S2**).

Discussion

The COVID-19 pandemic is the most serious new global disease of this century. Understanding the male bias in severity and mortality of COVID-19 is a central but poorly explored aspect of COVID-19 pathogenesis, which might be important for future strategies aiming at better prevention and treatment of critically ill patients [27]. We suggest here that LOY might be a factor underlying the higher severity and mortality of this infection in men. This would be another example of an age-related disorder associated with LOY, but the first acute, often life-threatening viral infection, representing a step forward in the understanding of the role of LOY in the susceptibility to disease, especially in relation to the dysfunction of the immune system. Male predominance for critical COVID-19 is consistent with a similar bias for prior SARS and Middle East

respiratory syndrome epidemics (caused by SARS-CoV and MERS-CoV viruses, respectively) [33-35]. Moreover, male-female differences have also been reported for seasonal influenza A virus infections in Australia and Japan [36, 37]. Consequently, the importance of our results might not be limited to COVID-19 and this should be studied further in COVID-19 as well as other common viral infections.

We have recently reported [5], and show it here again in the context of COVID-19 that LOY can be scored on single-cell RNA-seq data, allowing analysis of the impact of this large mutation on transcription in selected leukocyte subpopulations. Using this approach, we re-analyzed published sc-RNA-seq data [18] showing that LDNs (or immature neutrophils), which play a key role in the pathogenesis of COVID-19, exhibit high levels of LOY in critically ill COVID-19 males (**Fig. 1**). It should be mentioned here that sc-RNA-seq has frequently been used in studies of COVID-19 and there exist numerous additional publicly available sc-RNA-seq datasets, which will facilitate follow-up and extension of our results in additional cohorts. Our results from sc-RNA-seq were confirmed at the DNA level, as we demonstrated that LDNs from critically ill patients displayed LOY, representing 46% of the studied COVID-19 patients (**Fig. 3**). Comparing the literature and considering the young age of patients in our cohort (median age 65 years), this is a very high number of subjects showing LOY in bulk DNA. This observation is also in line with other results suggesting that the myeloid lineage is the most affected by LOY [5, 38].

Epidemiological studies have associated LOY with myeloid cell differentiation [29, 30]. A high neutrophil to lymphocyte ratio is also a recognized marker of COVID-19 severity [39]. Considering the high percentage of LOY cells among neutrophils of COVID-19 patients, it is possible that clonal expansion and functional disturbance of these cells might be a consequence of LOY. We also demonstrated a decreased level of LOY in blood and PBMCs in recovering patients, suggesting that LOY could be a biomarker of severe disease. Neutrophils are the cell type with the second highest daily turnover (after erythrocytes) and are intensively produced during infections [40]. Therefore, neutrophils are most prone to show clonal hematopoiesis of indeterminate potential (CHIP), which is associated with age-related accumulation of various post-zygotic mutations, among them LOY. In line with the above, we have recently shown that LOY in monocytes is accompanied by pathogenic somatic mutations in genes associated with DNA methylation, transcription regulation and DNA repair [38].

Our sc-RNA-seq results show that many members of HLA class II genes and *CD74*, as well as one crucial component of HLA class I system (beta-2 micro-globulin, *B2M*) are among the top down-regulated genes (**Figs. 6 and S4, Table S2**) in CD14⁺ monocytes of COVID-19 patients. It is noteworthy that *CD74* was recently reported as factor that can block the entry of the SARS-CoV-2 and other viruses [41]. We have also recently shown that LOY is associated with dysregulation of autosomal genes and one example was *LAG3*, which is a component of HLA class II system [5]. Furthermore, previous work on immuno-phenotyping of

leukocytes in critically ill COVID-19 patients showed down-regulation of some HLA class II proteins [42, 43]. Thus, our results together with published data suggest that LOY can negatively affect the immune system and could be responsible for dysregulation of numerous immune genes [5, 15].

We report correlations of LOY to clinical variables in critically ill patients, such as increased thrombocyte counts, which is in line with previous reports from non-covid cohorts [29, 30]. Thromboembolic complications also correlated with increased levels of LOY in LDNs and monocytes, which may contribute to the severity of COVID-19 disease in males with LOY in these cell fractions, eventually leading to organ failure, in particular affecting lungs. A recent paper [44] shows that thrombocytes from COVID-19 patients secrete increased levels of S100A8/A9 proteins forming heterodimer protein calprotectin, inducing the expression of coagulation-associated genes in vascular endothelial cells. The *S100A8* gene is the top upregulated gene in monocytes with LOY from COVID-19 patients (**Fig. S4** and **Table S2**) and this might represent another link between LOY and increased severity of COVID-19. In conclusion, LOY might have a predictive value for identification of patients that are at high risk for severe course of COVID-19 and this should be studied in larger cohorts. While the detailed mechanism of how LOY is involved in clonal expansion and disturbed function of myeloid cells remains largely unknown, our results support a link between LOY and emergency myelopoiesis and the role of LOY in modulation of COVID-19 disease severity, which ought to stimulate further research.

ACKNOWLEDGEMENTS

We thank all the anonymous COVID-19 patients and volunteer healthy controls for acceptance to participate in the study, sample contribution and information provided in the medical questionnaire. We thank Maria Lindström, Anna Beskow, Susanne Aldén and the lab team at the Uppsala Biobank for their support with preparation of COVID-19 samples from patients. Uppsala Clinical Research Center (UCR), part of Uppsala University and Uppsala University Hospital, are also acknowledged in this respect. We thank the Uppsala Intensive Care COVID-19 research group; Tomas Luther, Anders Larsson, Katja Hanslin, Anna Gradin, Sara Galien, Jacob Rosén, Sara Bülow Anderberg, Labolina Spång, Erik Danielsson, Philip Karlsson and Amanda Svensson for the help with patient recruitment and sample collection. We acknowledge the Biobank West, a part of Sahlgrenska University Hospital and Gothenburg University for support with preparation of COVID-19 samples from patients recruited in Gothenburg. We thank Dr. Jakub Mieczkowski for bioinformatic advice. This study was supported by grants from the Science for Life Laboratory/KAW via National COVID-19 research, Swedish Heart-Lung Foundation, Swedish Cancer Society, the Swedish Research Council, Hjärnfonden, Alzheimerfonden, Konung Gustav V:s och Drottning Victorias Frimurarestiftelse, and the Foundation for Polish Science under the International Research Agendas Programme (grant number MAB/2018 /6; co-financed by the European Union under the European Regional Development Fund) to J.P.D.; SciLifeLab/KAW national COVID-19 research program project grant to MH (KAW 2020.0182 and KAW 2020.0241), the Swedish Heart Lung Foundation to M.H. (20190639, 20190637 and 20210089) and the Swedish Research Council [2014-02569, 2014-07606, and 2016-02606 (the latter to J.D.J.)] as well as The Swedish Kidney Foundation (F2020-0054) to R.F. This study was also supported by grants from the SciLifeLab National COVID-19 Research Program, financed by the Knut and Alice Wallenberg Foundation [2020.0182 and 2020.0241, the Swedish Research Council (#2021-06545)] and the Swedish State Support for Clinical Research (ALFGBG-717531) to M.G. J.O.W was supported by KAW Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab. Single cell RNA sequencing was performed at SNP&SEQ facilities at Uppsala University and SciLifeLab. SNP&SEQ is part of the National Genomics Infrastructure (NGI) Sweden supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Support by NBIS (National Bioinformatics Infrastructure Sweden) is gratefully acknowledged. Data handling and computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) partially funded by the Swedish Research Council (#2018-05973).

Materials and Methods

Study design, sample collection and clinical variables

Male patients with COVID-19 were enrolled in Sweden, at Uppsala University Hospital (as part of the PronMed-study) and Sahlgrenska University Hospital in Gothenburg with confirmed positive SARS-CoV-2 nasopharyngeal test verified by RT-PCR. The COVID-19 cohort included 139 critically ill patients admitted to ICU, 4 milder patients (not requiring corticosteroid treatment, ICU-care, or high-flow oxygen at time of sampling) and 17 after recovery 3-6 months after discharge. Samples from COVID-19 patients were collected between July 2020 and August 2021 in Uppsala and Gothenburg, Sweden. Control samples were collected from 38 healthy individuals, in Gdansk, Poland. Informed consent was obtained from all the participants, or from next of kin for ICU-admitted patients, who were not able to give consent themselves. The research was approved by the local research ethics committee in Uppsala and Gothenburg, Sweden Dnr. 2017/043 with amendments 2019/00169, 2020-01623, 2020-20719 and 2021-01469; Dnr. 2020-01771 (COVID-19 cohorts) and Dnr. 2021-02205 (COVID-19 recovery samples). The Bioethics Committee at Medical University of Gdansk Dnr. NKBBN/564/2018 approved sample collection for control cohort in Poland. The study was conducted according to the guidelines of the Declaration of Helsinki. The full set of clinical variables that was used for statistical tests is shown in **Table S1**.

Sample preparation

16 ml of blood was collected into BD Vacutainer® CPT™ Mononuclear Cell Preparation Tubes (BD). 1 ml of whole blood was saved for further DNA analysis. The tube was spun down within 4 hours of collection; PBMC fraction was washed with PBS and cryopreserved in FBS (Fetal Bovine Serum) for long time storage at -150°C. Granulocytes together with red blood cells were recovered from the fraction below the gel of the CPT tube. The red blood cells were then lysed twice with 40 ml of 1x RBC lysis buffer (PharmLyse, BD) for 15 min at room temperature. The granulocytes were washed with PBS and cryopreserved in FBS (Fetal Bovine Serum) for long time storage in -150°C. Cryopreserved PBMCs were thawed and washed with PBS according to the 10X Genomics protocol for the handling of frozen cells. Cell number and viability was determined with Trypan blue and Countess II FL automated cell counter (Thermo Fisher). Approximately 200 000 cells were saved for DNA extraction.

Preparation of cells for FACS

For FACS analysis, the PBMCs were mixed with FITC labeled CD66b clone G10F5 (BD), APC labeled CD15 clone HI98 (BD), PE-CF594 labeled CD14 clone MφP9 (BD) and incubated for 20 min at 4°C. After

incubation, PBMCs were washed with PBS and cell pellets were resuspended in 1 ml of PBS containing 3 mM EDTA. Before sorting, cells were filtered through 20 µm CellTrics strainer to remove aggregates. The target cell populations were isolated using FACS Aria III (Becton Dickinson). Data were acquired and analyzed using BD FACSDiva™ Software (Becton Dickinson). Live cells were sorted based on their FSC and SSC, a singlets gate based on FSC-H and FSC-A was used to remove doublet cells. Neutrophils were identified as CD15+ and CD66b+; monocytes were defined based on their size and as CD14+. Cells were sorted to achieve a purity of at least 96%. After sorting, cells were centrifuged and cell pellets were stored in -80°C freezer.

DNA extraction

For samples with >50 000 cells the DNA was extracted using an in-house protocol previously described [5]. Cells were pelleted by centrifugation at 4000 rpm for 10 minutes and lysed with buffer containing 10 mM EDTA, 10 mM Tris-HCL (pH 7.9), 50 mM NaCl, and 1% N-Lauroylsarcosine sodium salt (Sigma) with 10 mg/ml proteinase K (Sigma). The DNA was then precipitated and resuspended in Low-TE. Samples with cell number between 10 000 to 50 000 cells were handled with an optimized cell lysis protocol. Lysis buffer without detergent was added to the cell pellet, together with 10 mg/ml proteinase K (Sigma) and incubated for 2 h in 50 °C. The cell lysate was then incubated in 95°C for 10 min to inactivate proteinase K. DNA from whole blood and granulocyte samples were extracted using QIAmp DNA Blood Midi kit (Qiagen).

Analysis of the level of Loss of chromosome Y (LOY) using ddPCR.

The LOY status was determined using ddPCR as described previously [6]. The DNA samples were pre-digested with HindIII (Thermo Fischer). Subsequently 50 ng of the digested DNA was used in the analysis. The digested DNA was mixed with PCR supermix for probes without dUTP (BioRad) together with primers and probes for the AMELX/AMELY TaqMan-assay, number C_990000001_10 (Thermo Fisher). Quantification of the relative number of chromosomes X and Y in the sample was obtained by targeting the 6 bp sequence difference present between the AMELX and AMELY genes. Droplets were generated using the automated droplet generator (Bio-Rad), PCR was conducted using the T100 thermal cycler (Bio-Rad) and QX200 Droplet reader was used for the fluorescent measurements of droplets. The data was analyzed using the QuantaSoft software (Bio-Rad).

Preparation of PBMCs for single cell RNA sequencing (sc-RNA-seq)

For sc-RNA-seq, samples from 30 individuals with critical COVID-19 and 34 controls were collected as described above. PBMCs were dissolved in a buffer containing PBS with 0.04% BSA at a concentration of

10⁶ cells/ml. The PBMCs were loaded on Chromium Next GEM Single Cell 3' Reagent kit v3.1 (10x Genomics) for sequence library preparations according to manufacturer's instructions. The single cell libraries were then sequenced using the NovaSeq 6000 and v1.5 sequencing chemistry (Illumina Inc.). The single cell library preparation and sequencing were performed at the Science for Life technology platform SNP&SEQ, Uppsala University, Sweden. CellRanger v6.0.1 (10x Genomics) was used to process sc-RNA-seq. To generate a digital gene expression (DGE) matrix for each sample, we mapped their reads to GRCh38 reference genome and recorded the number of UMIs for each gene in each cell. We excluded one control sample having more than 10 times the average number of reads and one critical covid sample having less than 40% fraction reads in cells.

sc-RNA-seq data analysis of 10x Chromium data

sc-RNA-seq UMI count matrices were imported to R 4.1.0 and gene expression data analysis was performed using the R/Seurat package 4.0.4 [32]. We excluded cells based on the following quality criteria: more than 25% mitochondrial reads, less than 50 expressed genes or more than 7,500 expressed genes. We further excluded genes that were expressed in less than five cells. In addition, mitochondrial genes and MALAT1 have been excluded from further analysis. We removed doublets following R/DoubletFinder 2.0.3 best practice workflow [45].

We normalized count data using SCTransform function in Seurat, integrated data by embedding it into PBMC reference and annotated cell types using Azimuth workflow [32]. For two-dimensional data visualization we performed UMAP based on the first 50 dimensions of the supervised PCA data reduction. Cells classified as LOY-cells had no detectable expression from chromosome Y, according to Dumanski et al. 2021 [5].

Differential expression tests

Differential expression (DE) tests were performed in CD14⁺ monocytes using FindMarker functions in Seurat with a hurdle model tailored to sc-RNA-seq data in MAST. Three pairwise comparisons were performed: group of cells without LOY from controls (baseline) vs. each of LOY-cells from controls (LOY), cells without LOY from critical COVID-19 patients (COV) and LOY-cells from critical covid patients (BOTH) groups. Genes with >0.1 log-fold changes, expressed in ≥20% of cells in ≥1 tested groups, and Bonferroni-corrected p-values <0.05 in at least one pairwise comparison were regarded as significantly differentially expressed genes (DEGs). Top 500 DEGs ordered by adjusted p-values (breaking ties by log-fold changes) were summarized in **Table S2**.

GO enrichment analysis

Genes showing concordant down-regulation (average log FC < 0 for all three comparisons) were named as showing “*down*” pattern. Among the top 100 significant DEGs, 21 DEGs showing a *down* pattern were taken as a gene set for enrichment test by R/enrichR package vs. GO Biological Process 2021 (GO:BP2021) and vs. GO Molecular Function 2021 (GO:MF2021). Since LOY-cells by definition have no gene expression from chromosome Y, only more conservative results involving genes not located on chromosome Y were reported. We noted that adding genes from chromosome Y into analysis or considering top 500 genes result in the same pathways showing significant enrichment, (see **Table S2** for details).

Statistical analysis

Processing and statistical analysis was performed after completion of the experiments by the statisticians (D.S., J.O.W., A.S. and M.J.), who were not involved in execution of experiments. Therefore, the results of intermediate statistical analyses could not affect acquisition of experimental data. Data was analyzed using Bayesian regression models via full Bayesian framework by calling Stan 2.21 [46] from R 4.1 [47] using the brms 2.16 [48] interface. Predictors and outcomes were centered and scaled. To reduce the influence of outliers, Student’s t response distribution family with identity link function was used to model percent LOY outcomes and hurdle negative binomial distribution family with log link function was used to model FACS events counts. Models had no intercepts with indexing approach to predictors [49]. According to Stan recommendations [50] weakly informative priors were used for group-level effects, residual SD and group-level SD. P-values, adjusted using the multivariate t distribution with the same covariance structure as the estimates, were produced by frequentist summary in emmeans 1.6 [51]. Medians of the posterior distribution and 80%- and 95%-HDI were plotted. The contrast between groups was defined as significant if 95%-HDI did not include zero and adjusted P-value was ≤ 0.05 . Moreover, FACS-sorting events count data and percent LOY data were analysed by Wilcoxon matched pairs signed-rank test. The significance level was set to $p \leq 0.05$.

Data and Code Availability

Data supporting the findings of this study are available within the article, its Supplementary material or upon request. sc-RNA-seq data generated during this study are deposited at the European Genome-phenome Archive (EGA) under access number EGAS-nnn, which is hosted by the EBI and the CRG. Code is available on <https://www.github.com/>

FIGURES & FIGURE LEGENDS

Figure 1

Fig. 1. Distribution of cell types and cells with loss of chromosome Y (LOY) in PBMCs from critically ill patients with COVID-19. The dataset comes from Schulte-Schrepping et al. 2020 (PMID: 32810438) and we used cell type annotation reported therein. Nine PBMCs samples from six male patients with WHO grade ≥ 5 were used for calculations. **A)** Proportion of cell types in PBMCs; each data point represents a single sample. **B)** Proportion of cells with LOY per cell type and per sample; each data point represents a single sample. **C)** UMAP projection of sc-RNA-seq profiles, selected cell types are colored. **D)** UMAP visualization of sc-RNA-seq profiles colored according to LOY-status. Cells classified as LOY-cells had no detectable expression from chromosome Y [5].

Figure 2

Fig. 2. Differences in LDNs counts between controls and critically ill patients (treated at ICU). FACS images showing the gating strategy employed to identify low-density neutrophils (LDNs, also called immature neutrophils) for representative control (**A**) and ICU patient (**B**). PBMCs were gated for size and granularity first, LDNs were identified as CD15+ (APC) and CD66b+ (FITC) later. **C**) Comparison of the numbers of LDNs in PBMCs between critically ill patients (ICU, n=135), milder patients (n=4) and healthy controls (n=17). Points show medians; thick and thin horizontal bars show 80%- and 95%-HDI. P-values are shown after MVT correction.

Figure 3

A

B

% of patients with at least 5% of cells with LOY unadjusted for age and smoking

	Median age (range)	Blood	PBMCs	Gran	Mono	LDNs
ICU-treated male covid-19 patients (n LOY/n max)	65 (19-86)	32.61 (45/138)	26.62 (37/139)	31.82 (42/132)	29.55 (39/132)	46.07 (41/89)

Fig. 3. Comparison of %LOY across five cell populations for critically ill patients. A) Comparison of loss of chromosome Y percentages (%LOY) in monocytes, PBMCs, whole blood, granulocytes and low-density neutrophils (LDNs) within ICU-patients. Points show medians; thick and thin horizontal bars show 80%- and 95%-HDI. P-values are shown after MVT correction. **B)** Unadjusted percent of patients with LOY.

Figure 4

Fig. 4. Leukocyte counts and LOY status for patients during ICU-treatment and during recovery. **A)** Comparison of the numbers of low-density neutrophils (LDNs) at intensive care unit (ICU, $n=135$) and 3-6 months after the discharge from ICU (recovery, $n=17$) in PBMCs. Points denote medians; thick and thin horizontal bars show 80%- and 95%-HDI, respectively. **B)** Paired comparison of loss of chromosome Y percentages (%LOY) in whole blood and PBMC, and numbers of LDNs and monocytes cells in PBMCs in 16 patients, where both ICU and recovery samples were available. Each dotted line connects the ICU and recovery data within the same patient. Boxplots show median, IQR as hinges and largest values no further than $1.5 \times \text{IQR}$ away as whiskers. Unadjusted p -values from paired two-sided Mann-Whitney-Wilcoxon non-parametric tests are shown.

Figure 5

Fig. 5. Clinical correlations. **A)** Correlation between %LOY vs thrombocyte, erythrocyte and hemoglobin (Hb) measurements. Robust correlation coefficients between %LOY in five types of cells and cell counts are shown; “.day1”, “.min” and “.max” denote levels at admission to ICU, as well as minimal and maximal measurements during ICU treatment, respectively. **B)** Comparison of %LOY in LDNs and monocytes with (n=20) and without (n=115) thromboembolic events at ICU. **C)** Comparison of %LOY in blood and monocytes at ICU, with (n=80) and without (n=56) invasive mechanical ventilation (IMV). **D)** Comparison of %LOY in four types of cells at ICU, with (n=26) and without (n=112) vessel disease (VD). Points show medians; thick and thin horizontal bars show 80%- and 95%-HDI, respectively. Asterisk denote that 95%-HDI is separated from zero and thus the correlation coefficient is statistically significant. Unadjusted p-values are shown.

Fig. 6. LOY- and covid-associated pattern of down-regulation of HLA-related gene expression in CD14⁺ monocytes. Violin-plots and median expression levels are shown after normalization by *sctransform* from Seurat: cells without LOY from controls (green; “baseline”; 22,676 cells), LOY-cells from controls (orange; 3,741 cells), cells without LOY from critical COVID-19 patients (blue; 41,486 cells) and LOY-cells from critical COVID-19 patients (magenta; 6,015 cells).

References

1. Forsberg LA, Rasi C, Malmqvist N, Davies H, Pasupulati S, Pakalapati G, Sandgren J, de Stahl TD, Zaghlool A, Giedraitis V, Lannfelt L, Score J, Cross NC, Absher D, Janson ET, Lindgren CM, Morris AP, Ingelsson E, Lind L, Dumanski JP: Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer, *Nature Genetics* 2014, 46:624-628; (<http://www.ncbi.nlm.nih.gov/pubmed/24777449>)
2. Dumanski JP, Lambert JC, Rasi C, Giedraitis V, Davies H, Grenier-Boley B, Lindgren CM, Campion D, Dufouil C, European Alzheimer's Disease Initiative I, Pasquier F, Amouyel P, Lannfelt L, Ingelsson M, Kilander L, Lind L, Forsberg LA: Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease, *Am J Hum Genet* 2016, 98:1208-1219; (<http://www.ncbi.nlm.nih.gov/pubmed/27231129>)
3. Thompson D, Genovese G, Halvardson J, Ulirsch J, Wright D, Terao C, Davidsson O, Day F, Sulem P, Jiang Y, Danielsson M, Davies H, Dennis J, Dunlop M, Easton D, Fisher V, Zink F, Houlston R, Ingelsson M, Kar S, Kerrison N, Kristjansson R, Li R, Loveday C, Mattisson J, McCarroll S, Murakami Y, Murray A, Olszewski P, Rychlicka-Buniowska E, Scott R, Thorsteinsdottir U, Tomlinson I, Torabi Moghadam B, Turnbull C, Wareham N, Gudbjartsson D, INTEGRAL-ILCCO, The Breast Cancer Association Consortium, CIMBA, The Endometrial Cancer Association Consortium, The Ovarian Cancer Association Consortium, The PRACTICAL Consortium, The Kidney Cancer GWAS Meta-Analysis Project, eQTLGen Consortium, BIOS Consortium, 23andMe Research Team, Kamatani Y, Finucane H, Hoffmann E, Jackson S, Stefansson K, Auton A, Ong K, Machiela M, Loh P-R, Dumanski J, Chanock S, Forsberg L, Perry J: Genetic predisposition to mosaic Y chromosome loss in blood, *Nature* 2019, 575:652-657; (<https://www.ncbi.nlm.nih.gov/pubmed/31748747>)

4. Forsberg L, Halvardson J, Rychlicka E, Danielsson M, Torabi Moghadam B, Mattisson J, Rasi C, Davies H, Lind L, Giedraitis V, Lannfelt L, Kilander L, Ingelsson M, Dumanski J: Mosaic loss of chromosome Y (LOY) in leukocytes matters, *Nature Genetics* 2019, 51:4-7; (<https://www.ncbi.nlm.nih.gov/pubmed/30374072>)
5. Dumanski J, Halvardson J, Davies H, Rychlicka-Buniowska E, Mattisson J, Torabi Moghadam B, Nagy N, Węglarczyk K, Bukowska-Strakova K, Danielsson M, Olszewski P, Piotrowski A, Oerton E, Ambicka A, Przewoźnik M, Bełch L, Grodzicki T, Chłosta P, Imreh S, Giedraitis V, Kilander L, Nordlund J, Ameur A, Gyllensten U, Johansson A, Józkowicz A, Siedlar M, Klich-Rączka A, Jaszczyński J, Enroth S, Baran J, Ingelsson M, Perry J, Ryś J, Forsberg L: Immune cells lacking Y chromosome show dysregulation of autosomal gene expression, *Cell Mol Life Sci* 2021, 78:4019–4033; (<https://www.ncbi.nlm.nih.gov/pubmed/33837451>)
6. Danielsson M, Halvardson J, Davies H, Torabi Moghadam B, Mattisson J, Rychlicka-Buniowska E, Heintz J, Lannfelt L, Giedraitis V, Ingelsson M, Dumanski J, Forsberg L: Longitudinal changes in the frequency of mosaic chromosome Y loss in peripheral blood cells of aging men varies profoundly between individuals, *Eur J Hum Genet* 2020, 28:349-357; (<https://www.ncbi.nlm.nih.gov/pubmed/31654039>)
7. Haitjema S, Kofink D, van Setten J, van der Laan S, Schoneveld A, Eales J, Tomaszewski M, de Jager S, Pasterkamp G, Asselbergs F, den Ruijter H: Loss of Y Chromosome in Blood Is Associated with Major Cardiovascular Events during Follow-up in Men after Carotid Endarterectomy, *Circulation: Cardiovascular Genetics* 2017, 10:e001544; (<https://www.ncbi.nlm.nih.gov/pubmed/28768751>)
8. Dumanski JP, Rasi C, Lonn M, Davies H, Ingelsson M, Giedraitis V, Lannfelt L, Magnusson PK, Lindgren CM, Morris AP, Cesarini D, Johannesson M, Tiensuu Janson E, Lind L, Pedersen NL, Ingelsson E, Forsberg LA: Smoking is associated with mosaic loss of chromosome Y, *Science* 2015, 347:81-83; (<http://www.ncbi.nlm.nih.gov/pubmed/25477213>)
9. Ganster C, Kampfe D, Jung K, Braulke F, Shirneshan K, Machherndl-Spandl S, Suessner S, Bramlage CP, Legler TJ, Koziolok MJ, Haase D, Schanz J: New data shed light on Y-loss-related pathogenesis in myelodysplastic syndromes, *Genes Chromosomes Cancer* 2015, 54:717-724; (<http://www.ncbi.nlm.nih.gov/pubmed/26394808>)
10. Loftfield E, Zhou W, Graubard BI, Yeager M, Chanock SJ, Freedman ND, Machiela MJ: Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank, *Sci Rep* 2018, 8:12316; (<https://www.ncbi.nlm.nih.gov/pubmed/30120341>)
11. Loftfield E, Zhou W, Yeager M, Chanock SJ, Freedman ND, Machiela MJ: Mosaic Y Loss Is Moderately Associated with Solid Tumor Risk, *Cancer Res* 2019, 79:461-466; (<https://www.ncbi.nlm.nih.gov/pubmed/30510122>)
12. Persani L, Bonomi M, Lleo A, Pasini S, Civardi F, Bianchi I, Campi I, Finelli P, Miozzo M, Castronovo C, Sirchia S, Gershwin ME, Invernizzi P: Increased loss of the Y chromosome in peripheral blood cells in male patients with autoimmune thyroiditis, *Journal of autoimmunity* 2012, 38:193-196; (<http://www.ncbi.nlm.nih.gov/pubmed/22196921>)
13. Lleo A, Oertelt-Prigione S, Bianchi I, Calzari L, Finelli P, Miozzo M, Lazzari R, Floreani A, Donato F, Colombo M, Gershwin ME, Podda M, Invernizzi P: Y chromosome loss in male patients with primary biliary cirrhosis, *Journal of autoimmunity* 2013, 41:87-91; (<http://www.ncbi.nlm.nih.gov/pubmed/23375847>)
14. Grassmann F, Kiel C, den Hollander A, Weeks D, Lotery A, Cipriani V, Weber B, International Age-related Macular Degeneration Genomics Consortium (IAMDC): Y chromosome mosaicism is associated with age-related macular degeneration, *European Journal of Human Genetics* 2019, 27:36-41; (<https://www.ncbi.nlm.nih.gov/pubmed/30158665>)
15. Mattisson J, Danielsson M, Hammond M, Davies H, Gallant CJ, Nordlund J, Raine A, Edén M, Kilander L, Ingelsson M, Dumanski JP, Halvardson J, Forsberg LA: Leukocytes with chromosome Y loss have reduced abundance of the cell surface immunoprotein CD99, *Scientific Reports* 2021, 11:15160; (<https://doi.org/10.1038/s41598-021-94588-5>)
16. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus I, Research T: A Novel Coronavirus from Patients with Pneumonia in China, 2019, *N Engl J Med* 2020, 382:727-733; (<https://www.ncbi.nlm.nih.gov/pubmed/31978945>)

17. Hu B, Guo H, Zhou P, Shi ZL: Characteristics of SARS-CoV-2 and COVID-19, *Nat Rev Microbiol* 2021, 19:141-154; (<https://www.ncbi.nlm.nih.gov/pubmed/33024307>)
18. Schulte-Schrepping J, Reusch N, Paclik D, Bassler K, Schlickeiser S, Zhang B, Kramer B, Krammer T, Brumhard S, Bonaguro L, De Domenico E, Wendisch D, Grasshoff M, Kapellos TS, Beckstette M, Pecht T, Saglam A, Dietrich O, Mei HE, Schulz AR, Conrad C, Kunkel D, Vafadarnejad E, Xu CJ, Horne A, Herbert M, Drews A, Thibeault C, Pfeiffer M, Hippenstiel S, Hocke A, Muller-Redetzky H, Heim KM, Machleidt F, Uhrig A, Bosquillon de Jarcy L, Jurgens L, Stegemann M, Glosenkamp CR, Volk HD, Goffinet C, Landthaler M, Wyler E, Georg P, Schneider M, Dang-Heine C, Neuwinger N, Kappert K, Tauber R, Corman V, Raabe J, Kaiser KM, Vinh MT, Rieke G, Meisel C, Ulas T, Becker M, Geffers R, Witznath M, Drosten C, Suttorp N, von Kalle C, Kurth F, Handler K, Schultze JL, Aschenbrenner AC, Li Y, Nattermann J, Sawitzki B, Saliba AE, Sander LE, Deutsche C-OI: Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment, *Cell* 2020, 182:1419-1440 e1423; (<https://www.ncbi.nlm.nih.gov/pubmed/32810438>)
19. Ren X, Wen W, Fan X, Hou W, Su B, Cai P, Li J, Liu Y, Tang F, Zhang F, Yang Y, He J, Ma W, He J, Wang P, Cao Q, Chen F, Chen Y, Cheng X, Deng G, Deng X, Ding W, Feng Y, Gan R, Guo C, Guo W, He S, Jiang C, Liang J, Li YM, Lin J, Ling Y, Liu H, Liu J, Liu N, Liu SQ, Luo M, Ma Q, Song Q, Sun W, Wang G, Wang F, Wang Y, Wen X, Wu Q, Xu G, Xie X, Xiong X, Xing X, Xu H, Yin C, Yu D, Yu K, Yuan J, Zhang B, Zhang P, Zhang T, Zhao J, Zhao P, Zhou J, Zhou W, Zhong S, Zhong X, Zhang S, Zhu L, Zhu P, Zou B, Zou J, Zuo Z, Bai F, Huang X, Zhou P, Jiang Q, Huang Z, Bei JX, Wei L, Bian XW, Liu X, Cheng T, Li X, Zhao P, Wang FS, Wang H, Su B, Zhang Z, Qu K, Wang X, Chen J, Jin R, Zhang Z: COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas, *Cell* 2021, 184:1895-1913 e1819; (<https://www.ncbi.nlm.nih.gov/pubmed/33657410>)
20. Huckriede J, Anderberg SB, Morales A, de Vries F, Hultström M, Bergqvist A, Ortiz-Pérez JT, Sels JW, Wichapong K, Lipcsey M, van de Poll M, Larsson A, Luther T, Reutelingsperger C, de Frutos PG, Frithiof R, Nicolaes GAF: Evolution of NETosis markers and DAMPs have prognostic value in critically ill COVID-19 patients, *Sci Rep* 2021, 11:15701;
21. Spiezia L, Boscolo A, Poletto F, Cerruti L, Tiberio I, Campello E, Navalesi P, Simioni P: COVID-19-Related Severe Hypercoagulability in Patients Admitted to Intensive Care Unit for Acute Respiratory Failure, *Thromb Haemost* 2020, 120:998-1000; (<https://www.ncbi.nlm.nih.gov/pubmed/32316063>)
22. Kollias A, Kyriakoulis KG, Dimakakos E, Poulakou G, Stergiou GS, Syrigos K: Thromboembolic risk and anticoagulant therapy in COVID-19 patients: emerging evidence and call for action, *Br J Haematol* 2020, 189:846-847; (<https://www.ncbi.nlm.nih.gov/pubmed/32304577>)
23. Zuo Y, Zuo M, Yalavarthi S, Gockman K, Madison JA, Shi H, Woodard W, Lezak SP, Lugogo NL, Knight JS, Kanthi Y: Neutrophil extracellular traps and thrombosis in COVID-19, *J Thromb Thrombolysis* 2021, 51:446-453; (<https://www.ncbi.nlm.nih.gov/pubmed/33151461>)
24. Busch MH, Timmermans S, Nagy M, Visser M, Huckriede J, Aendekerk JP, de Vries F, Potjewijd J, Jallah B, Ysermans R, Oude Lashof AML, Breedveld PH, van de Poll MCG, van de Horst ICC, van Bussel BCT, Theunissen R, Spronk HMH, Damoiseaux J, Ten Cate H, Nicolaes GAF, Reutelingsperger CP, van Paassen P: Neutrophils and Contact Activation of Coagulation as Potential Drivers of COVID-19, *Circulation* 2020, 142:1787-1790; (<https://www.ncbi.nlm.nih.gov/pubmed/32946302>)
25. Peckham H, de Grijter NM, Raine C, Radziszewska A, Ciurtin C, Wedderburn LR, Rosser EC, Webb K, Deakin CT: Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission, *Nat Commun* 2020, 11:6317; (<https://www.ncbi.nlm.nih.gov/pubmed/33298944>)
26. Gebhard C, Regitz-Zagrosek V, Neuhauser HK, Morgan R, Klein SL: Impact of sex and gender on COVID-19 outcomes in Europe, *Biol Sex Differ* 2020, 11:29; (<https://www.ncbi.nlm.nih.gov/pubmed/32450906>)
27. Scully EP, Haverfield J, Ursin RL, Tannenbaum C, Klein SL: Considering how biological sex impacts immune responses and COVID-19 outcomes, *Nat Rev Immunol* 2020, 20:442-447; (<https://www.ncbi.nlm.nih.gov/pubmed/32528136>)
28. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carre C, Burdin N, Visan L, Ceccarelli M, Poidinger M, Zippelius A, Pedro de Magalhaes J, Larbi A: RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types, *Cell Rep* 2019, 26:1627-1640 e1627; (<https://www.ncbi.nlm.nih.gov/pubmed/30726743>)

29. Terao C, Momozawa Y, Ishigaki K, Kawakami E, Akiyama M, Loh PR, Genovese G, Sugishita H, Ohta T, Hirata M, Perry JRB, Matsuda K, Murakami Y, Kubo M, Kamatani Y: GWAS of mosaic loss of chromosome Y highlights genetic effects on blood cell differentiation, *Nat Commun* 2019, 10:4719; (<https://www.ncbi.nlm.nih.gov/pubmed/31624269>)
30. Lin SH, Loftfield E, Sampson JN, Zhou W, Yeager M, Freedman ND, Chanock SJ, Machiela MJ: Mosaic chromosome Y loss is associated with alterations in blood cell counts in UK Biobank men, *Sci Rep* 2020, 10:3655; (<https://www.ncbi.nlm.nih.gov/pubmed/32108144>)
31. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R: Comprehensive Integration of Single-Cell Data, *Cell* 2019, 177:1888-1902 e1821; (<https://www.ncbi.nlm.nih.gov/pubmed/31178118>)
32. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R: Integrated analysis of multimodal single-cell data, *Cell* 2021, 184:3573-3587 e3529; (<https://www.ncbi.nlm.nih.gov/pubmed/34062119>)
33. Karlberg J, Chong DS, Lai WY: Do men have a higher case fatality rate of severe acute respiratory syndrome than women do?, *Am J Epidemiol* 2004, 159:229-231; (<https://www.ncbi.nlm.nih.gov/pubmed/14742282>)
34. Leong HN, Earnest A, Lim HH, Chin CF, Tan C, Puhaindran ME, Tan A, Chen MI, Leo YS: SARS in Singapore--predictors of disease severity, *Ann Acad Med Singap* 2006, 35:326-331; (<https://www.ncbi.nlm.nih.gov/pubmed/16829999>)
35. Alghamdi IG, Hussain, II, Almalki SS, Alghamdi MS, Alghamdi MM, El-Sheemy MA: The pattern of Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive epidemiological analysis of data from the Saudi Ministry of Health, *Int J Gen Med* 2014, 7:417-423; (<https://www.ncbi.nlm.nih.gov/pubmed/25187734>)
36. Wong KC, Luscombe GM, Hawke C: Influenza infections in Australia 2009-2015: is there a combined effect of age and sex on susceptibility to virus subtypes?, *BMC Infect Dis* 2019, 19:42; (<https://www.ncbi.nlm.nih.gov/pubmed/30630435>)
37. Eshima N, Tokumaru O, Hara S, Bacal K, Korematsu S, Tabata M, Karukaya S, Yasui Y, Okabe N, Matsuishi T: Sex- and age-related differences in morbidity rates of 2009 pandemic influenza A H1N1 virus of swine origin in Japan, *PLoS One* 2011, 6:e19409; (<https://www.ncbi.nlm.nih.gov/pubmed/21559366>)
38. Ljungstrom V, Mattisson J, Halvardson J, Pandzic T, Davies H, Rychlicka-Buniowska E, Danielsson M, Lacaze P, Cavelier L, Dumanski JP, Baliakas P, Forsberg LA: Loss of Y and clonal hematopoiesis in blood-two sides of the same coin?, *Leukemia* 2021, (<https://www.ncbi.nlm.nih.gov/pubmed/34725452>)
39. Carissimo G, Xu W, Kwok I, Abdad MY, Chan YH, Fong SW, Puan KJ, Lee CY, Yeo NK, Amrun SN, Chee RS, How W, Chan S, Fan BE, Andiappan AK, Lee B, Rotzschke O, Young BE, Leo YS, Lye DC, Renia L, Ng LG, Larbi A, Ng LF: Whole blood immunophenotyping uncovers immature neutrophil-to-VD2 T-cell ratio as an early marker for severe COVID-19, *Nat Commun* 2020, 11:5243; (<https://www.ncbi.nlm.nih.gov/pubmed/33067472>)
40. Sender R, Milo R: The distribution of cellular turnover in the human body, *Nature medicine* 2021, 27:45-48; (<https://www.ncbi.nlm.nih.gov/pubmed/33432173>)
41. Bruchez A, Sha K, Johnson J, Chen L, Stefani C, McConnell H, Gaucherand L, Prins R, Matreyek KA, Hume AJ, Muhlberger E, Schmidt EV, Olinger GG, Stuart LM, Lacy-Hulbert A: MHC class II transactivator CIITA induces cell resistance to Ebola virus and SARS-like coronaviruses, *Science* 2020, 370:241-247; (<https://www.ncbi.nlm.nih.gov/pubmed/32855215>)
42. Mudd PA, Crawford JC, Turner JS, Souquette A, Reynolds D, Bender D, Bosanquet JP, Anand NJ, Striker DA, Martin RS, Boon ACM, House SL, Remy KE, Hotchkiss RS, Presti RM, O'Halloran JA, Powderly WG, Thomas PG, Ellebedy AH: Distinct inflammatory profiles distinguish COVID-19 from influenza with limited contributions from cytokine storm, *Sci Adv* 2020, 6:(<https://www.ncbi.nlm.nih.gov/pubmed/33187979>)
43. Wilk AJ, Rustagi A, Zhao NQ, Roque J, Martinez-Colon GJ, McKechnie JL, Ivison GT, Ranganath T, Vergara R, Hollis T, Simpson LJ, Grant P, Subramanian A, Rogers AJ, Blish CA: A single-cell atlas of

- the peripheral immune response in patients with severe COVID-19, *Nature medicine* 2020, 26:1070-1076; (<https://www.ncbi.nlm.nih.gov/pubmed/32514174>)
44. Barrett TJ, Cornwell M, Myndzar K, Rolling CC, Xia Y, Drenkova K, Biebuyck A, Fields AT, Tawil M, Luttrell-Williams E, Yuriditsky E, Smith G, Cotzia P, Neal MD, Kornblith LZ, Pittaluga S, Rapkiewicz AV, Burgess HM, Mohr I, Stapleford KA, Voora D, Ruggles K, Hochman J, Berger JS: Platelets amplify endotheliopathy in COVID-19, *Sci Adv* 2021, 7:eabh2434; (<https://www.ncbi.nlm.nih.gov/pubmed/34516880>)
 45. McGinnis CS, Murrow LM, Gartner ZJ: DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors, *Cell Syst* 2019, 8:329-337 e324; (<https://www.ncbi.nlm.nih.gov/pubmed/30954475>)
 46. Stan Development Team: RStan: the R interface to Stan. R package version 2.21.2. Edited by 2020, p. (<http://mc-stan.org/>.)
 47. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna; Austria. Edited by 2021, p. (<https://www.R-project.org/>)
 48. Bürkner P-C: brms: An R Package for Bayesian Multilevel Models Using Stan, *Journal of Statistical Software* 2017, 80:1 - 28; (<https://www.jstatsoft.org/index.php/jss/article/view/v080i01>)
 49. McElreath R: Statistical Rethinking. A Bayesian Course with Examples in R and STAN. Edited by Chapman and Hall/CRC, 2020,
 50. Gelman A: Prior Choice Recommendations. In: Stan-dev/stan. Ed. GitHub. Edited by 2019, p. (<https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>)
 51. Lenth RV: emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.6.3. Edited by 2021, p. (<https://CRAN.R-project.org/package=emmeans>)
 52. McDavid A, Finak G, Yajima M: MAST: Model-based Analysis of Single Cell Transcriptomics. R package version 1.18.0. Edited by 2021, p. (<https://github.com/RGLab/MAST/>)