Enhancer deregulation in *TET2* Mutant Clonal Hematopoiesis is associated with increased COVID-19 related inflammation severity and mortality

Moritz Binder[1,2†], MD MPH, Terra L. Lasho[1†], PhD, Wazim Mohammed Ismail[2,3†], PhD, Nana A. Ben-Crentsil[4†], MBChB MS, Jenna A. Fernandez[1†], PhD, Minsuk Kim[5], PhD, Susan M. Geyer[6], PhD, Amelia Mazzone[2,3], PhD, Christy M. Finke[1], BS, Abhishek A. Mangaonkar[1], MD, Jeong-Heon Lee[2], PhD, Kwan Hyun Kim[2], PhD, Vernadette A. Simon[7], MS, Fariborz Rakhshan Rohakthar[7], MS, Amik Munankarmy[3], MD, Susan M. Schwager[1], Jonathan J. Harington[4], Melissa R. Snyder[3], PhD, Nathalie M. Droin[8], PhD, Eric Solary[8], MD, Keith D. Robertson[2,9], PhD, Eric D. Wieben[10], PhD, Eric Padron[4‡], MD, Nicholas Chia[5‡], PhD, Alexandre Gaspar-Maia[2,3‡], PhD, and Mrinal M. Patnaik[1,2‡], MD

[†] Co-first authors | [‡] Co-corresponding authors

[1] Division of Hematology, Department of Internal Medicine, Mayo Clinic, Rochester, MN 55905, United States
[2] Epigenomics Program, Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905, United States
[3] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, United States
[4] H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, United States
[5] Microbiome Program, Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905, United States
[6] Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN 55905, United States
[7] Medical Genome Facility, Genome Analysis Core, Mayo Clinic, Rochester, MN 55905, United States
[8] INSERM U1287, Gustave Roussy Cancer Center, 94805 Villejuif, France
[9] Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, United States
[10] Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN 55905, United States

Correspondence to:

Mrinal M. Patnaik, MD
Associate Professor of Medicine
Mayo Clinic College of Medicine and Science

Consultant
Division of Hematology
Department of Internal Medicine
Mayo Clinic
200 First Street SW
Rochester, MN 55905

Email: patnaik.mrinal@mayo.edu
Phone: 507-293-7814

Alexandre Gaspar-Maia, PhD
Assistant Professor of Laboratory Medicine and Pathology
Mayo Clinic College of Medicine and Science

Senior Associate Consultant
Experimental Pathology and Laboratory Medicine
Department of Laboratory Medicine and Pathology
Mayo Clinic
200 First Street SW
Rochester, MN 55905

Email: maia.alexandre@mayo.edu
Phone: 507-293-2464

Nicholas Chia, PhD
Associate Professor of Surgery
Mayo Clinic College of Medicine and Science

Senior Associate Consultant
Department of Quantitative Health Sciences
Mayo Clinic
200 First Street SW
Rochester, MN 55905

Email: chia.nicholas@mayo.edu
Phone: 507-538-0614

Eric Padron, MD
Associate Professor of Oncologic Sciences
H. Lee Moffitt Cancer Center and Research Institute

Associate Member
Department of Malignant Hematology
H. Lee Moffitt Cancer Center
12902 USF Magnolia Drive
Tampa, FL 33612

Email: epadron@health.usf.edu
Phone: 866-663-3488

Abstract: 180 | **Word Count Excluding Methods:** 5361 | **Tables:** 1 | **Figures:** 6
Supplementary Figures: 6 | **Supplementary Tables:** 8

ABSTRACT

Coronavirus disease 2019 (COVID-19) is associated with significant morbidity and mortality, albeit with considerable heterogeneity among affected individuals. It remains unclear which factors determine severity of illness and long-term outcomes. Emerging evidence points towards an important role of preexisting host factors, such as a deregulated inflammatory response at the time of infection. Here, we demonstrate the negative impact of clonal hematopoiesis, a prevalent clonal disorder of ageing individuals, on COVID-19-related cytokine release severity and mortality. We show that mutations in the gene coding for the methylcytosine dioxygenase *TET2* promotes amplification of classical and intermediate monocyte subsets. Using single cell multiomic sequencing approaches, we identify cell-specific gene expression changes associated with the loss of TET2 and significant epigenomic deregulation affecting enhancer accessibility of a subset of transcription factors involved in monocyte differentiation. We further identify *EGR1* down-regulation secondary to *TET2*-mediated hypermethylation, resulting in overexpression of *MALAT1,* a lncRNA that plays a role in hematopoietic stem cell differentiation and monocyte lineage commitment. Together, these data provide a mechanistic insight to the poor prognostic value of clonal hematopoiesis in patients infected with Sars-COV2.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the pathogen responsible for coronavirus disease 2019 (COVID-19), has resulted in an ongoing pandemic associated with significant morbidity, mortality, and long-term sequelae.[1-3] While infections can vary from asymptomatic carrier states to severe cytokine release syndrome (CRS), acute respiratory distress (ARDS) and associated multi organ dysfunction syndrome (MODS), reasons for clinical heterogeneity have partially been investigated, with several viral (strain type) and host factors (age, comorbidities, immune status, ACE2 receptor polymorphisms, among others) congruently being involved.[4]

Severe COVID-19 infections have been associated with a relative increase in classical monocytes ($CD14^{high}$ $CD16^{low}$), due to the migration and pulmonary macrophage differentiation of inflammatory nonclassical monocytes ($CD14^{low}$ $CD16^{high}$, usually <4%), along with significantly elevated levels of calprotectin (S100A8 / S100A9) in affected individuals.[5] Classical monocytes physiologically are primed for phagocytosis, innate immune responses and monocyte migration, while non-classical monocytes are involved in complement and Fc gamma-mediated phagocytosis and adhesion.[6] Expanded classical monocyte fractions (>94%) have also consistently been documented in chronic myelomonocytic leukemia (CMML), a myeloid neoplasm characterized by peripheral blood monocytosis, with a somatic gene signature composed of recurrent mutations involving *TET2* (60%), *ASXL1* (40%) and *SRSF2* (50%).[7-9] Clonal hematopoiesis (CH) is defined by the acquisition and subsequent expansion of leukemia-associated driver mutations in hematopoietic stem and progenitor cells (HSPC) in apparently healthy individuals.[10-12] CH has been associated with a low, but increased risk of hematological neoplasms and an increased all-cause mortality, largely due to cardiovascular disease and inflammation associated endothelial dysfunction.[11-14]

In a large cohort of 515 patients with COVID-19, CH was associated with severe COVID-19 outcomes, including increased mortality.[15] However, majority of these patients had underlying visceral malignancies, with several getting chemo- / immunotherapy, and except for a non-significant trend with *PPM1D* mutations, there were no clear mutational associations detected.[15] Among common CH mutations, *TET2* mutations are most likely to be associated with aging, with a monocyte bias in hematopoiesis and are strongly associated with macrophage and endothelial dysfunction.[10,12,16] In addition its role in regulating DNA methylation and hydroxymethylation, TET2 selectively mediates active repression of *IL6* transcription during inflammation resolution in innate myeloid cells.[17] In the context of HIV-1 infection, TET2 is degraded by viral protein R (Vpr), to sustain IL6 expression and enhance viral replication and disease progression.[18] Additionally, decreased *TET2* expression also favors transcription of the *MIF* (macrophage migration inhibitory factor) gene and increased production of its corresponding cytokine.[19] Together, these data point to a role of *TET2* in the regulation of immune and inflammatory responses.

In a cohort of 227 patients, we demonstrate an age- and comorbidity-independent adverse impact of *TET2* CH on COVID-19 outcomes. As expected, we observe an increase in CpG site DNA methylation, mostly in distal enhancer-like elements, in patients with *TET2* mutant CH compared to other CH genotypes and wildtype cases. Single-cell proteogenomics indicates that, in the context of *TET2* CH, the deregulated repartition of circulating monocyte subsets is amplified, and single-cell transcriptomics demonstrates an increased expression of S100A8 and S100A9 alarmin genes in monocytes, neutrophils, and dendritic cells of affected patients. Having observed that patients with *TET2* mutant CH and COVID-19 specifically overexpress *MALAT1*, a regulatory long noncoding RNA (lncRNA), we decipher a pathway that involves the epigenetic down-regulation of the transcription factor EGR1[20] as being responsible for enhanced *MALAT1* expression. These

data highlight the role of *TET2* mutant CH in enhancing immune cell dysregulation in the context of COVID-19 infection, which may account for increased viral disease severity.

RESULTS

Concomitant clonal hematopoiesis in patients with COVID-19 is associated with increased severity of cytokine release and increased age- and comorbidity-independent mortality

A cohort of 227 patients with COVID-19 was included in the study, median age 60 years (range 19-99), of which 52 (24.6%) patients had evidence of CH (Tab. 1 and Fig. 1A). Apart from the fact that patients with both COVID-19 and CH were older (median age for CH with COVID-19 69.5 years versus 57 years for CH without COVID-19; $p < 0.0001$), there were no significant differences in sex ($p = 0.5$), race ($p = 0.2$), comorbidities ($p = 0.12$), hospitalization rates ($p = 0.21$), oxygen requirements ($p = 0.11$), or incidence of CRS ($p=0.37$). There were no differences in baseline blood indices (MCV and RDW) between the two groups, in hemoglobin levels ($p = 0.099$), absolute neutrophil counts (ANC, $p = 0.15$), absolute monocyte counts (AMC; $p = 0.86$), or platelet counts ($p = 0.41$), respectively (Supp. Table 1-2). Apart from elevated MCP-1 (monocyte chemoattractant protein-1) levels obtained at COVID-19 diagnosis in the COVID-19 CH cohort compared to the COVID-19 cohort without CH ($p = 0.045$), there were no other significant differences in clinically measured cytokines / chemokines, including IL1b, IL6, GM-CSF and TNF-alpha, or inflammatory surrogates like C-reactive protein ($p = 0.087$) and serum ferritin ($p = 0.62$) (Supp. Table 3).

Seventy-three CH mutations were seen in 52 patients, with 13 (25%) having 2 CH mutations, 1 (2%) having 3 CH mutations and 2 (4%) having 4 CH mutations, respectively (Fig. 1A). While none of these patients underwent bone marrow (BM) biopsies, they did not have an underlying diagnosed hematological neoplasm at the time of COVID-19 detection and none of these patients demonstrated disease evolution at last follow-up (median 15 months). The most common CH mutations encountered included *DNMT3A* (n=25, 34%), *TET2* (n=15, 20.5%), *SF3B1* (n=6, 8%), *ASXL1* (n=5, 6.8%), *TP53* (n=4, 5.4%) and *PPM1D* (n=3, 4%), respectively (Fig. 1A-B). None of the *TP53* and *PPM1D* mutated patients had prior exposures to cytotoxic chemotherapy or ionizing radiation therapy. Four patients had 2 *DNMT3A* mutations, 2 patients had 2 *TET2* mutations, and one patient had both a *TET2* and a *DNMT3A* mutation. There was only one patient with the *DNMT3A* R882 hot spot mutation, commonly seen in AML and MDS, with a variant allele fraction (VAF) of 2%.[21]

Patients with CH and COVID-19 had higher grades of CRS in comparison to those without CH as documented by CTCAE v5.0 criteria ($p = 0.008$; Fig. 1C) and by the WHO COVID-19 severity criteria ($p = 0.023$, Fig. 1D).[22,23] There were no significant differences between the two groups with regards to incidence rates of acute lung injury ($p = 0.8$), ARDS ($p = 0.49$), acute kidney injury ($p = 0.5$), MODS ($p = 0.09$) and venous thromboembolism ($p = 0.76$). Both groups received similar therapeutic interventions including remdesivir, IL-6 directed monoclonal antibody therapies (tocilizumab and siltuximab), corticosteroids, and access to clinical trials ($p = 0.45$, Supp. Table 4-6). At last follow-up (median 15 months), 14 deaths (6%) have been documented, 8 (16%) in COVID-19 patients with CH and 6 (7%) in COVID-19 patients without CH. On a univariate and multivariate survival analysis that included several clinical and laboratory variables, the presence of CH negatively impacted overall survival (OS) in patients with COVID-19 ($p = 0.005$, Fig. 1E). The one-month estimated survival rate was 97.6% for COVID-19 patients without CH and 81.8% in COVID-19 patients with CH (median OS not reached versus 13 months). With this analysis, we demonstrate an association of higher grades of CRS with CH in COVID-19, along with an increased age-independent mortality.

The number of patients with individual CH mutations was too small to assess the impact of individual mutations on outcomes. Given the demonstrated role of *TET2* in the regulation of immune and inflammatory responses, we decided to further explore how somatic *TET2* mutations could impact COVID-19 outcomes, using other CH mutations and patients without CH as comparator groups.

TET2 mutant CH in the context of COVID-19 is associated with enhanced DNA methylation at CpG residues in contrast to patients with DNMT3A mutant CH and COVID-19

We first assessed DNA methylation status using the Illumina Infinium Methylation EPIC array on peripheral blood mononuclear cells (PBMC) from the COVID-19 and CH cohorts, respectively (Fig. 2A). *TET2* mutations in particular have been associated with DNA hypermethylation at key enhancer sites in granulocytes and mononuclear cells in patients with CH, with these elements known to be regulators of leukocyte function, inflammation and adaptive immune responses.[24] We included 9 patients with CH and COVID-19 (*TET2* - 4*, DNMT3A* - 3*, TET2* / *DNMTA* co-mutant - 1, *ASXL1* - 1), 4 patients with COVID-19 and no CH and 5 CH patients without COVID-19 (*TET2* - 3, *DNMT3A* - 2), respectively (Fig. 2A). Even though there were no global changes in DNA methylation between the three groups ($p = 0.37$, Supp. Fig. 2A), *TET2* mutant patients with COVID-19 demonstrated increased methylation at CpG sites ($\beta > 0.75$), a finding also seen in the *DNMT3A* / *TET2* co-mutated patient with COVID-19 (Fig. 2B, Supp. Fig. 2B). Differential methylation analysis also revealed an increased number of hypermethylated sites in *TET2* mutant patients with COVID-19 when compared to *DNMT3A* mutant patients with COVID-19, a finding also seen when comparing all COVID-19 patients with CH versus COVID-19 patients without CH (Fig. 2C). *TET2* mutant CH patients had 656 hypomethylated sites and 4,701 hypermethylated sites, of which the majority of the hypermethylated sites met classification criteria for distal enhancer-like elements when using the ENCODE cis-regulatory element definitions (70% of hypermethylated CpGs and 59% of hypomethylated CpGs overlap with distal enhancer-like regions) (Fig. 2D-E).[25] Pathway analysis revealed that these hypermethylated sites are in or near genes involved in a number of diseases and functions related to inflammation and immune response, especially in leukocyte function (Fig. 2F). Notably, by analyzing the enrichment of such differentially methylated regions against published genomic ChIP-seq databases of all myeloid cell types, we were able to determine potential transcription factor binding sites that could be disrupted by the observed changes in methylation. Potential candidate transcription factors that can be deregulated by *TET2* disruption that were identified included SREPB2, RXR, SMC1, CDK8, CREBBP and EGR1 (Fig. 2G). This is consistent with the fact that *TET2* activity in myeloid cells is associated with the recruitment of transcription factors involved in cell survival, proliferation, and differentiation.[26]

Single cell proteogenomics reveal that TET2 mutations bias hematopoiesis towards predominant classical and intermediate monocyte fraction monopoiesis

We then carried out comprehensive proteogenomic assessments at single-cell resolution on PBMC, on 10 patients with COVID-19 and CH (*TET2* mutant - 6, *DNMT3A* mutant - 4, *ASXL1* mutant - 1, with 1 patient having both *DNMT3A* and *TET2* mutations; Fig. 2A), 8 patients with COVID-19 and no CH, and 6 age-matched CH controls without COVID-19 (*TET2* mutant - 5, *DNMT3A* mutant - 1). In the COVID-19 CH cohort 1 patient had dual *TET2* mutations and 2 patients had dual *DNMT3A* mutations, while in the CH only cohort, 3 patients had dual *TET2* mutations. All *TET2* mutations were truncating with a median VAF of 5.5% in the COVID-19 CH cohort (range, 2-14%) and a median VAF of 25% (range 13-40%) in the CH only cohort (although

median VAF was higher in the COVID-19 CH cohort, none of the patients had unexplained cytopenias, or met criteria for a *bona fide* myeloid neoplasm). Similarly, the median *DNMT3A* VAF was 7% (range 4-21%) in the COVID-19 CH cohort, while there was only one *DNMT3A* mutation (VAF - 18%) in the CH only cohort. The single *ASXL1* mutation seen was a truncating mutation in the COVID-19 CH cohort with a VAF of 19%.

In total, we included 74,694 single cells in the final analysis, after rigorous quality control and exclusion of cells with allele drop out, as previously described (Fig. 3B and Methods section).[27] Of these 74,694 sequenced cells, 7262 (9.7%) had detectable CH mutations, of which 3156 (43%) were *DNMT3A* mutant, 1834 (25%) were *TET2* mutant and 460 (6.3%) were *ASXL1* mutant (Fig. 3C-F). In comparison to *TET2* and *ASXL1* mutant CH, where CH mutations were largely present in classical and intermediate monocytes, in *DNMT3A* mutant CH, the mutations were more commonly seen in $CD4^+$ T-lymphocytes (Fig. 3C-F), a lineage bias that has previously been described.[28] We then compared the single-cell immunophenotypes in *TET2* mutant versus wildtype cells in 6 patients with adequate number of cells for comparison and showed that *TET2* mutations were more likely to be seen in classical and intermediate monocyte fractions, compared to the wildtype cases (Fig. 3G). The only exception was patient #35, a 90-year-old male with *TET2* (VAF-2%) and *CBL* (VAF-30%) co-mutant CH with COVID-19, who had a baseline expansion in his classical and intermediate monocytes. In addition to the classical and intermediate monocyte fractions, *TET2* mutations were also enriched in circulating dendritic cells and plasmacytoid dendritic cells (Fig. 3B and Supp. Fig. 3C), which are known to be clonal in CMML.[29,30]

We carried out a similar analysis in 3 patients with sole *DNMT3A* mutant CH and did not see a similar trend for enrichment in classical and intermediate monocytes (Supp. Fig. 3D), while one *DNMT3A* and *TET2* co-mutant CH patient (patient #44) did have monocyte compartment enrichment (probably due to the *TET2* mutation). We also globally assessed surface immunophenotypes in all the COVID-19 CH *TET2* mutant versus wildtype cells, *TET2* mutant versus wildtype CH cells without COVID-19 and the CH negative COVID-19 cells and demonstrate an enrichment in classical monocytes in the *TET2* mutant versus wildtype cells in both *TET2* mutant cohorts (Supp. Fig. 3E-G). Commensurate with the median *TET2* VAF, there were 220 (1%) *TET2* mutant cells identified out of 21,755 sequenced cells in the *TET2* mutant CH COVID-19 cohort and 1305 (10%) *TET2* mutant cells identified out of 12,871 sequenced cells in the *TET2* CH no COVID-19 cohort. Since we only had single-cell data on one patient with *ASXL1* mutant CH, we were not able to complete an accurate assessment with regards to monocyte portioning.

Based on these findings we conclude that classical and intermediate monocytes and circulating dendritic cells are enriched in *TET2* mutations in *TET2* mutant CH. Given the inflammatory nature of these cells, and the inflammatory associations of mutant *TET2*, this may represent an important mechanism in priming the innate and adaptive immune systems towards a state of pervasive inflammation in COVID-19.

Single-cell RNA-seq validates the classical and intermediate monocyte enrichment associated with TET2 mutations and demonstrates overexpression of S100A8 and S100A9 alarmin genes, along with MALAT1, in TET2 mutant CH with COVID-19 versus TET2 mutant CH without COVID-19 and other genotypes

To explore changes induced by SARS-CoV-2 infection between patients with CH and patients without CH, we assessed PBMC from a cohort of 34 patients, which included 11 patients with CH and COVID-19 (*TET2* - 6*, DNTM3A* - 3*, ASXL1* - 1, *EZH2* - 1), 15 patients with COVID-19 and no CH and 8 CH patients without COVID-19 (*TET2* - 5, *DNMT3A* - 1, *TET2/DNMT3A* co-mutant - 1,

ASXL1 - 1), respectively (Fig. 4A, Supp. Fig. 4A). By pooling 102,072 cells from all 34 patients and subjecting them to dimensionality reduction using PCA and UMAP, followed by cell type identification using SingleR (see Methods section), we identified the typical populations of B-cells, NK-cells, $CD4^+$ T-cells, $CD8^+$ T-cells, CD14 monocytes, CD16 monocytes, and an intermediate monocytic population that included gene signatures from both CD16 and CD14 monocytes, among others (Fig. 4B-C). Analysis of the proportion of monocytes per subtype and per genotype allowed us to determine a significant enrichment of classical and intermediate monocytes in the COVID-19^+ / *TET2*^{MT} patients specifically, when compared with COVID-19^+ / CH^- and COVID-19^+ / CH^+ for mutations other than *TET2*, which will from now onwards will be referred to as COVID-19^+ / CH^+ / *TET2*^{WT} (Fig. 4D-E, Supp. Fig. 4C-D). Interestingly, monocyte *HLA-DRA* expression was reduced in COVID-19 patients even when comparing *TET2* mutants with and without COVID-19 (Supp. Fig. 4E) in all three monocytic subtypes, in line with previous observations demonstrating the accumulation of *HLA-DRA* low monocytes in severe COVID-19.[5] Importantly, we were able to detect a decrease in the percentage of cells expressing *TET2* in the COVID-19^+ / *TET2*^{MT} cohort, specifically in myeloid lineages, in corroboration with our mutational analysis, suggesting a disruption of *TET2* expression in the *TET2*^{MT} patients (Fig. 4F).

To further investigate a potential association of *TET2* mutations with COVID-19 severity, we performed differential expression analysis within each cell type, comparing COVID-19^+ / CH^- with COVID-19^+ / CH^+ mutant individually and by grouping all the CH mutations (Fig. 5A, Supp. Fig. 5A, Supp. Table 7). An overall increase in expression of *S100A8* and *S100A9* alarmin genes was observed in B-cells and dendritic cells (Supp. Fig. 5B), a potential biomarker of severity that has been documented in plasma samples from patients with COVID-19.[5] Even though we were able to determine differences in the other CH mutations with COVID-19, given that we had identified *TET2* mediated classical and intermediate monocyte fraction biased hematopoiesis and the fact that classical monocytes are known to be inflammatory, we decided to pursue transcriptional changes associated with *TET2* mutations more specifically.

Epigenetic down-regulation of EGR1 is responsible for transcriptional deregulation of TET2 mutated cells in COVID19 patients

Interestingly, despite having a pronounced increase in number of monocytes, *TET2* mutant patients had a striking down-regulation of genes among those lineages (*MARCKS, FTH1*, and *LGALS2)* including *TET2* itself and other transcription factors known to define macrophage differentiation such as *EGR1*, suggesting the enrichment for classical and intermediate monocytes could be in part due to an inability to differentiate to tissue resident macrophages.[20] Pathway analysis of differentially expressed genes indicated signatures of inflammatory response, such as activation of granulocytes, rheumatoid arthritis, and inflammation of joints (Fig. 5B).

To better understand novel mechanisms associated with the inflammatory response, we decided to focus on two non-coding RNAs, *MALAT1* and *NEAT1*, both highly over-expressed in COVID-19^+ / *TET2*^{MT} when compared with COVID-19^+ / CH^- and when compared to COVID-19^- / *TET2*^{MT} as well (Fig. 5C-E, Supp. Fig. 5C). *MALAT1* was overexpressed predominantly in monocytes (all three subsets) and dendritic cells in COVID-19 patients with *TET2* mutations, but was also overexpressed in COVID-19 patients with CH in comparison to healthy individuals, with data being obtained from previously published data sets.[31] *NEAT1* also demonstrated a similar pattern, but the difference between *TET2* mutants and COVID-19 patients without CH was not statistically significant (Fig. 5D-E). We also looked at the specific *TET2*^{MT} versus CH^- differentially expressed gene signatures in classical monocytes that were concordant between COVID-19^+ cohort and the COVID-19^- cohort (where both *MALAT1* and *NEAT1* are included) and discordant (Fig. 5C and

Supp. Fig. 5C). This signature revealed genes that were in the same direction (up-regulated or down-regulated) in comparison to healthy individuals with CH (patients in a non-inflammatory setting), and COVID-19^{+} patients with and without *TET2* mutations. Besides the two non-coding RNAs *MALAT1* and *NEAT1*, there was an enrichment in up-regulation of several mitochondrial genes, an indication of cellular stress.[32]

Given the potential effect of *TET2* mutations on chromatin as an epigenetic regulator and our inability to test methylation status at the single-cell level, we decided to perform single-cell profiling of both expression (scRNA-seq) and open chromatin (scATAC-seq) from the same PBMC using the 10X Genomics Multiome platform. From a cohort of 18 patients, which included 6 patients with CH and COVID-19 (*TET2* - 3*, DNTM3A* - 2*, ASXL1* - 1), 6 patients with COVID-19 and no CH and 6 CH patients without COVID-19 (*TET2* - 4, *DNMT3A* - 1, *TET2* / *DNMT3A* co-mutant - 1), respectively (Supp. Fig. 6A), we pooled 43,160 cells and we performed dimensionality reduction using PCA and UMAP followed by cell type identification by mapping the expression data from Multiome onto the scRNA-seq data and transferring labels using Azimuth (Supp. Fig. 6A-B). From each cell type, we can gather an expression signature and open chromatin profile (cut sites). Using this technology, we were able to validate the monocytic enrichment (Supp. Fig. 6C) and the overexpression of proinflammatory non-coding RNAs *MALAT1* and *NEAT1* in the classical monocytes in the COVID-19^{+} / *TET2*^{MT} cohort compared to the COVID-19^{+} / CH^{-} cohort (Supp. Fig. 6D).

Since chromatin accessibility is reflective of the active enhancer and promoter structure and it is strongly associated with the methylation status, we aimed to provide mechanistic insights on the deregulation of the epigenetic landscape in *TET2* mutant CH. Notably, analysis of differentially accessible peaks between COVID-19^{+} / CH^{-} and COVID-19^{+} / *TET2*^{MT} shows that there is decreased accessibility in the main myeloid lineages, including classical, non-classical and intermediate monocytes, whereas in some lineages there is an increase in open chromatin accessibility, such as in T-cells (Supp. Fig. 6E). Similarly, the number of open chromatin regions (cut sites per cell) are reduced in monocytes (Fig. 6A), a strong indicator that *TET2*^{MT} could be decreasing chromatin accessibility in these cells secondary to enhancer and promoter methylation. In fact, co-accessibility analysis of cut sites within each cell type identifies potential chromatin interactions such as interactions between enhancers and promoters, quantified as connections per peak, which revealed a significant loss of accessibility in the COVID-19^{+} / *TET2*^{MT} cohort. This loss was more prominent when analyzing connections that have CpGs associated with them, another strong indicator of the specific effect of *TET2* mutations associated with methylation patterns (Fig. 6B, Supp. Fig. 6F).

To understand which transcription factors could be mostly affected by *TET2*-driven disruption, we first performed analysis of differentially accessible motifs in classical monocytes and validated several of those by performing differential enrichment analysis directly from ChIP-seq datasets found in the literature.[33] Once again, the major differences between COVID-19^{+} / CH^{-} cohort and the COVID-19^{+} / *TET2*^{MT} were found in classical monocytes and predominantly affected transcription factors highly associated with myeloid differentiation, such EGR1, CEBPB and SPI1 (Fig. 6C-D). Expression of some of these transcription factors was also analyzed in the same cell types to ensure that the chromatin accessibility analysis could also reflect changes in expression levels and whether the expression changes could also be myeloid lineage specific (Fig. 6E). EGR1, CEBPB and SPI1 were all found to be mostly expressed in monocytes and showed down-regulation in the *TET2*^{MT}. Interestingly, EGR1 was not identified in the motif analysis but only in the ChIP-seq enrichment analysis, and TET2 binding was found at EGR1 binding sites in monocytes (Fig. 6F). One explanation for this may be the targeting behavior of EGR1 that is (1) associated with known motifs when acting as an activator and (2) associated with co-factors of

the NuRD (nucleosome modeling and deacetylase) complex member CHD4 when acting as a repressor. Its activating effect seems to be associated with genes that are involved in macrophage differentiation, while its repressive effect targets inflammatory genes. In this context, we hypothesized that *EGR1* loss of expression was positively affecting the expression of *MALAT1* and *NEAT1* because of its potential role as a repressor of these proinflammatory genes.

To functionally validate the role of *TET2* mutations in EGR1 function, we decided to overlap the chromatin accessibility signal of the most represented cell types and corresponding expression, and we were able to identify several unique regulatory regions that were associated with *MALAT1* and *NEAT1*, while also being sites for CpG DNA methylation (Fig. 6G). Finally, we tested if *EGR1* could act as a repressor of *MALAT1* and *NEAT1* by downregulating *EGR1* through shRNA in two cell line models and were able to validate that the downregulation of *EGR1* led to up-regulation of *MALAT1*, but not *NEAT1* (Fig. 6H, Supp. Fig. 6G, see Methods section).

DISCUSSION

While several host and viral factors impacting COVID-19 severity have been previously described, in this study, we demonstrate the importance of CH involving epigenetic regulators, in particular *TET2* mutations, as an important age-independent host factor enhancing COVID-19 related CRS severity and mortality.[1-3,15,34] TET2 is a dioxygenase that iteratively demethylates DNA in a ferrous, alpha ketoglutarate, and ascorbic acid dependent manner and is recruited to enhancer elements by transcription factors, where it facilitates DNA demethylation and transcription.[26,35] Somatic *TET2* mutations in HSPC have been described in age related CH and form the basis for several hematological neoplasms including CMML and acute myeloid leukemia (AML).[28,36-38] *TET2* mutations in HSPC bias hematopoiesis towards monocytosis, with biallelic *TET2* mutations or *TET2* and *SRSF2* co-mutations giving rise to clonal monocytosis, setting the stage for evolution to CMML, a myeloid neoplasm characterized by sustained, classical fraction monopoiesis.[8,39] *TET2* mutations are inherently proinflammatory, with *TET2* CH being associated with higher levels of IL1β / NLRP3 and IL6 and with murine models showing the need for an intact tet2 to non-catalytically recruit hdac2 to the Il6 promoter site to repress Il6 signaling.[17,40,41]

While the association of age-related CH and severe COVID-19 has been hypothesized, we for the first time provide biological insights into this association, in an unselected population of COVID-19 patients with age-related CH, without underlying hematological or visceral neoplasms.[15] The frequency of CH in the COVID-19 cohort was 24.6%, with 29% having more than one CH mutation and with CH prevalence increasing with age. The most frequent CH mutations seen were *DNMT3A* (34%) and *TET2* (20.5%), followed by *SF3B1* (8%) and *ASXL1* (6.8%), consistent with prior observations.[10,11] Patients with CH and COVID-19 had higher grades of CRS in comparison to those without CH, with CH having an age-independent adverse impact on mortality. There was no further stratification of this effect based on CH mutational VAF or the number of CH mutations. Due to the smaller sample size, we were not able to assess the impact of individual mutations on morbidity and mortality. None of the patients included in this cohort, at diagnosis of COVID-19, or at last follow-up, had evidence of a hematological neoplasm. Except for elevated MCP-1 levels in the COVID-19+/CH+ cohort, at COVID-19 diagnosis, there were no significant differences in other clinically measured cytokines between the two groups. MCP-1, also called CCL2, is a key chemokine that regulates the migration of monocytes and macrophages in response to inflammation.[42]

Methylation studies using the Illumina Methylation EPIC array demonstrated increased DNA methylation in *TET2* mutant COVID-19 cases, occurring predominantly at the distal enhancer like elements, in comparison to CH involving other genes (*DNMT3A*) and the non-CH COVID-19

cohort. This is in line with prior observations assessing DNA methylation on whole blood samples in patients with CH, CH associated cytopenias and AML, where *TET2* mutations were associated with hypermethylated enhancers (ELF2, ETS1, SPI1 and C/EBP family), impacting the regulation of leukocyte function and inflammation.[24]

We then assessed the distribution of *TET2* mutant CH in COVID-19^{+} / CH^{+} cohort and in a COVID-19^{-} / CH^{+} cohort using single-cell proteogenomics and show that in *TET2*^{MT} CH, the *TET2* mutations were enriched in classical and intermediate fraction monocytes. Classical monocytes are proinflammatory and express high levels of chemokine receptors, such as CCR1, CCR2, CCR5, CXCR1, and CXCR2, highlighting their potential to migrate to inflammatory sites and have an enhanced ability to secrete proinflammatory cytokines/chemokines, such as IL-6, IL-8, MCP-1/CCL2, CCL3, and CCL5.[6,43] Classical monocytes can also differentiate into monocyte-derived inflammatory macrophages and dendritic cells and play an important role in tissue-based inflammation.[44] Intermediate monocytes are known to express the highest levels of antigen presentation-related molecules and have the capacity to produce TNF-α, IL-1β, IL-6, and CCL3 upon TLR stimulation and are known to be increased in patients with systemic infections. Enhanced classical monopoiesis is also a hallmark of CMML, a myeloid neoplasm associated with sustained peripheral blood monocytosis and autoimmunity (M01>94%).[8,45] In addition, the *TET2*^{MT} CH cohort was noted to have increased levels of circulating dendritic cells and HSPC, compared to the *TET2*^{WT} patients with COVID-19. This enhanced classical and intermediate fraction monopoiesis was not seen in *DNTM3A* mutant CH.

We then carried out single-cell RNA-seq and using SingleR references validated our single-cell proteogenomic assessments, demonstrating enrichment of *TET2* mutations in classical and intermediate fraction monocytes in *TET2*^{MT} patients with and without COVID-19, compared to *TET2*^{WT} patients. In addition, we show that *TET2*^{MT} CH patients with COVID-19 also had higher fractions of circulating dendritic cells, plasmacytoid dendritic cells, basophils and HSPC, compared to the *TET2*^{WT} patients with COVID-19. These findings were not seen in *DNMT3A*^{MT} CH patients or in controls. These data are in line with prior single-cell RNA-seq data in COVID-19 patients, where authors demonstrated a relative increase in the classical monocyte subset thought to be due to increased non-classical monocyte migration into pulmonary alveolar and parenchymal spaces, a phenomenon largely facilitated by elevated calprotectin levels.[5] Commensurate with their observations, we also demonstrate that the classical monocyte subsets seen in COVID-19 have a low expression of HLA-DR, a finding that is thought to correlate with calprotectin and IL-6 overproduction.[46] While our study was not designed to assess the mechanisms of enhanced classical monopoiesis, we believe that enhanced nonclassical monocyte transmigration is not the only mechanism behind this observation and that in the context of COVID-19, calprotectin bursts can drive emergency myelopoiesis and monopoiesis.[47] Our single-cell RNA-seq data demonstrated an increase in expression of lncRNA's *MALAT1* and *NEAT1* in *TET2*^{MT} CH patients with COVID-19, in comparison to *TET2*^{MT} CH without COVID-19 and *TET2*^{WT} controls, along with a marked increase in expression of several mitochondrial genes, usually a sign of cell stress, likely induced by the replicating SARS-CoV-2 virus.

MALAT1 is a highly conserved, large sized, nuclear lncRNA that has been associated with tumor growth and progression in the context of several visceral malignancies. In the hematopoietic system, *MALAT1* is believed to regulate HSPC proliferation and is transcriptionally regulated by *TP53*.[48] Importantly, *MALAT1* skews macrophage differentiation into the proinflammatory M1 macrophages, which are activated by INF-γ and/or TLR ligands.[49] Monocytes are postmitotic cells that can infiltrate tissues secreting inflammatory cytokines, but usually undergo an elaborate differentiation program generating macrophages that can phagocytose and activate the adaptive immune system by effective antigen presentation. The monocyte-macrophage development axis

is regulated by a conserved pool of transcription factors including PU.1, CEBPRA/B and AP1. Recently, the role of EGR1 and its association with the NuRD complex, repressing inflammatory enhancers in developing macrophages, blunting inflammatory responses has been elucidated.[20] In addition, TET2 has been shown to physically interact with EGR1 at the promoter regions of MIF in healthy monocytes, with TET2 disruption resulting in overproduction of MIF.[19] In our study, using a combination of single-cell RNA-seq and scATAC-seq (in the same nucleus), we demonstrate that the overexpression of *MALAT1* in *TET2*^{MT} CH is specific to the context of COVID-19 and correlates with decreased expression of *EGR1*, in comparison to *TET2*^{MT} CH without COVID-19 and COVID-19 without CH. Importantly, we demonstrate a decrease in chromatin accessibility largely in classical and intermediate monocytes in *TET2*^{MT} CH patients with COVID-19, by demonstrating decreased cut site counts and decreased number of connections per peak. We demonstrate that loss of *EGR1* expression positively impacts *MALAT1* expression in the context of inflammatory COVID-19 stimuli and that several of the EGR1 regulatory sites are enriched in CpG dinucleotides that are hypermethylated in *TET2*^{MT} CH with COVID-19, compared to *TET2*^{MT} CH without COVID-19 and *TET2*^{WT} controls. Finally, using representative cell lines, we demonstrate that EGR1 can be a potential context specific repressor of *MALAT1* by down-regulating EGR1 through shRNA, which in turn leads to up-regulation of *MALAT1*.

In summary, our results validate CH as a risk factor for severe COVID-19. With regards to *TET2* mutant CH, our study identifies a pathway that involves the epigenetic down-regulation of *EGR1* through hypermethylation of its enhancer. In turn, *MALAT1*, a key lncRNA that likely plays a role in HSPC differentiation and monocyte lineage commitment, is overexpressed likely resulting in classical and intermediate monopoiesis with an HLA-DR low suppressive phenotype, contributing to COVID-19 severity. *TET2* mutant CH thus joins the list of host biomarkers predicting COVID-19 related CRS severity and mortality.

ACKNOWLEDGMENTS

The authors thank the Genome Analysis Core and the Biospecimens Accessioning and Processing (Mayo Clinic) for technical support, and Alessandro Gardini for help with access to their published datasets. This work was supported by the Mayo Clinic Center for Individualized Medicine.

AUTHOR CONTRIBUTIONS

M.P., A.G.M., N.C., conceived and designed the study with the help of M.B., T.L., W.M.I. and E.P. A.M., N.N.B-C., C.M.F., K.H.K., V.A.S., F.R.R., A.M., M.R.S. and J-H.L. performed experiments. W.M.I., T.L., M.B., J.F. M.K., S.M.G., A.A.M., S.M.S. analyzed the data. M.P., A.G.M. and M.B., wrote the manuscript with the help of W.M.I, K.R., E.P., N.M.D. and N.C. All authors critically revised and approved the final version of the manuscript.

COMPETING INTERESTS

MP has received research funding from Kura Oncology and StemLine Pharmaceuticals.

.

MATERIALS AND METHODS

Authorizations, patient cohorts, cell collection and sorting

This study was carried out at the Mayo Clinic in Rochester, Minnesota, after approval from the Mayo Clinic Institutional Review Board (IRB- 15-003786). In all cases, diagnosis was according to the 2016 iteration of the WHO classification of myeloid malignancies.[50] PB and BM samples were collected in EDTA tubes after informed consent.

Target Capture Chip Assay for Bulk Sequencing

Sample DNA was extracted from peripheral blood mononuclear cells isolated by gradient centrifugation and re-suspended in a concentration of 500 ng in 50µl of low TE buffer. Paired-end indexed libraries were prepared using the Sureselect XT Low Input Library prep protocol on the Agilent Bravo liquid handler following the manufacturer's protocol (NewEngland Biolabs, Ipswitch, MA, and Agilent Technologies, Ankeny, IA). Briefly, 200ng of target DNA was fragmented using the Covaris LE220 plus sonicator. The settings of duty factor 30%, peak incident power (PIP) 450, cycles per burst 200, time 180 seconds generated double-stranded DNA fragments with blunt or sticky ends with a fragment size mode of between 150-200bp. The ends were repaired using the Sureselect End-Repair-A-Tailing enzyme mix. Adapter ligated DNA fragments were size selected to enrich for 200 bp inserts (~320 bp total library size) using AMPURE XP bead purification. The size selected adapter-modified fragments were enriched, and specific indexes were added by 12 cycles of PCR using universal index primers. The concentration and size distribution of the libraries was determined on an Agilent Bioanalyzer DNA 1000 chip.
The Custom Capture hybrid-target enrichment probes were designed using Agilent SureSelect design software (Agilent Technologies, Santa Clara, CA). The targeted gene panel was comprised of 62,962 single probes with size 1.805Mbp, and covered the coding regions, UTRs, and overlapping intron/exon regions for 205 genes described and / or enriched for CH mutations (see Appendix: Target Gene Panel). The custom capture was carried out using the Agilent Bravo liquid handler following Agilent's SureSelect XT Low. Purified capture products were then amplified using the SureSelect Post-Capture primer mix for 14 cycles. Libraries were validated and quantified on the Agilent Bioanalyzer. Samples were sequenced by 150 paired end reads, 21 samples to a Flow Cell, on an Illumina NovaSeq 6000 SP with an expected depth of ≥1,000X coverage (Illumina, SanDiego, CA).

Secondary bioinformatics analysis included quality assessment and alignment to the hg19 build reference genome using Novoalign (Novocraft Technologies, Malaysia), followed by GATK based single nucleotide and small insertion / deletion variant calling, structural variation discovery, and annotation. The quality of sequencing chemistry was evaluated using FastQC. After alignment, PCR duplication rates and percent reads mapped on target were used to assess the quality of the sample preparations. Realignment and recalibration steps were implemented in the GATK.[51] Somatic single nucleotide variations (SNVs) were then genotyped using SomaticSniper, whereas insertions and deletions were called by GATK Somatic Indel Detector.[51,52] Each variant in coding regions was functionally annotated by SnpEff, SAVANT, ClinVar, dbNSFP, OMIM, and the Human Gene Annotation Database to predict biological effects.[53-58] Each variant was also annotated with allele frequency from the Exoma Aggregation Consortium.[59] Variants of significant interest were visually inspected using IGV.[60] The total list of all variants was internally compared for internal

duplicates indicative of false positives, and variants of concern are additionally hand annotated for identification using Alamut Software. Interpretation for relevant alterations included absence in international normal variant allele databases (GnomAD, ExAC), deleterious effect on protein function by multiple phenotype prediction models, somatic and functional annotation in literature, consequence of variant (nonsense, truncating, etc.) and location proximal to important domains.

Single-Cell Proteogenomics

Single-cell DNAseq + proteogenomics was performed using the Mission Bio Tapestri platform according to the manufacturer's specifications. Briefly, a cocktail of pre-titered oligo-linked antibodies targeting 42 unique cell-surface markers and 3 antibodies for isotype control, TotalSeq™ D Human Heme Oncology Kit (BioLegend®, San Diego, CA [https://www.biolegend.com/en-us/totalseq/single-cell-dna]) was utilized to capture the cell immunophenotype, along with targeted myeloid scDNAseq genotyping panel targeted against 45 genes with 312 amplicons covering relevant genes dysregulated in myeloid disorders (designed and manufactured by Mission Bio, San Francisco, CA [https://designer.missionbio.com/catalogpanels/Myeloid]). See Appendices: Myeloid Panel Coverage and Myeloid Panel Design for specific genomic coordinates. Cryopreserved PBMC patient samples were gently thawed, washed, and run through a Dead Cell Removal Kit (Miltenyi, Auburn, CA). The resulting viable cell fraction was counted, blocked with Human TruStain FcX™ (BioLegend®) and re-suspended at a concentration of 25,000 cells/µL. This fraction was incubated with BioLegend TotalSeq Kit Cocktail described, washed, and filtered through a Flowmi cell strainer (MilliporeSigma, St. Louis, MO), and quantified using a Countess II cell counter. The cells were then diluted to a concentration of 4,000 cells per µL in Cell Buffer. Next, 35 µL of cell suspension was loaded onto a microfluidics cartridge and cells were encapsulated on the Tapestri instrument followed by cell lysis with protease digestion followed by heat inactivation using a thermal cycler. The cell lysate was reintroduced into the cartridge and processed such that each cell possessed a unique molecular barcode. Amplification of the targeted DNA regions was performed by incubating the barcoded DNA emulsions in a thermocycler as follows: 98 °C for 6 min (4 °C per s); ten cycles of 95°C for 30s, 72°C for 10s, 61°C for 9 min, 72°C for 20s (1 °C per s); ten cycles of 95°C for 30s, 72°C for 10s, 48°C for 9 min, 72°C for 20s (1°C per s); and 72°C for 6 min (4°C per s). Emulsions were broken, DNA digested and purified with 0.72X AMPure XP beads (Beckman Coulter). The beads were pelleted and washed with 80% ethanol and the generated by amplifying DNA libraries with Mission Bio V2 index primers in the thermocycler using the following program: 95°C for 3 min; 10 cycles of 98°C for 20s for DNA libraries and 16-20 cycles for protein libraries, 62°C for 20s, 72°C for 45s; 72°C for 2 min. Final libraries were purified with 0.69X AMPure XP beads. All libraries were sized and quantified using an Agilent Bioanalyzer and pooled for sequencing on an Illumina NovaSeq 6000 SP with 2 x 150bp multiplexed runs. FASTQ files generated by sequencers were processed using the Tapestri Pipeline V2 which handles adapter trimming, sequence alignment (BWA), barcode correction, cell finding and variant calling (using GATK 4.1.7 haplotype caller). Generated loom and H5 files were then processed with Tapestri Insights v2.2 (Mission Bio) and/or the Python-based Mosaic package (GitHub). Tapestri Insights analysis used default filter criteria (for example, genotype quality ≥30 and reads per cell per target ≥10) or whitelisting of known variants and annotation-based information (including ClinVar and DANN). Only cells with complete genotype information for all variants (previously detected in bulk sample) were included for downstream processing.

Cell type identification for the proteogenomics platform

A rigorous quality control was performed on antibody-derived tag (ADT) count profiles of the Tapestri single-cell proteogenomic data. We first filtered out cells with less than 200 or more than

50,000 total ADT counts, and then removed cells with less than 10 unique measured ADTs. After ADT-based filtering, one sample was excluded from the analysis as there were less than 500 cells remaining for the sample. ADT counts for each surface protein were then scaled using counts per 10 million with a pseudocount of +1 and normalized using the centered-log ratio (CLR) transformation. Subsequently, following the step II of the "dsb" (denoised and scaled by background) protocol normalized ADT profiles of antibodies for isotype control were used to remove cell-to-cell technical noise.[61] In result, we obtained a normalized and denoised ADT count matrix representing 42 different surface protein expression profiles of 74,694 single cells from 24 patients. This full matrix was used to generate UMAP plots following dimensionality reduction using PCA.

To identify cell types for cells profiled by the Tapestri platform, we compared ADT count profiles of the Tapestri data with that of a reference CITE-seq data, i.e., Azimuth reference.[62] Among the 42 cell-surface markers targeted by our Tapestri analysis, 36 were also targeted by the reference CITE-seq analysis. In the CITE-seq analysis, expression of 8 out of the 36 shared markers were assayed using two different antibodies targeting the same antigen (CD3, CD4, CD11b, CD38, CD44, CD45, CD56, and CD138). For these antigens, we used geometric means of ADT counts for both antibodies. CD5, CD7, CD10, CD33, CD62L, and FceRIa were the markers that were targeted by the Tapestri platform but not by the reference CITE-seq analysis. ADT count matrices of the 36 shared markers for the Tapestri and CITE-Seq analyses were scaled, normalized, and denoised as described above (for denoising CITE-seq data, 3 IgG antibodies were selected as isotype controls). Subsequently, for each cell from the Tapestri data, 20 nearest neighbor cells from the CITE-seq data were identified using Euclidian distance, and cell type was called based on their identities using majority vote.

To determine the distribution of cell types annotated by Tapestri, all individual cells (identified by unique oligo barcode by patient) from all cases were pre-defined and labeled as either non-*TET2*^{WT} or *TET2*^{MUT} via Tapestri Insights 2.2, based on bulk detection and Tapestri. The normalized level of relative surface protein expression for markers: CD3, CD19, CD14, CD56, CD11c, HLA-DR, and CD64 was then used to delineate the main cell types. Because CD3+ and CD19+ cell types had extremely clear bimodal cutoffs, these cells (CD3+ = T cells and CD19+ = B-cells) were then used, along with UMAP sourced by protein expression proximity, to determine the relative cutoff expression for other surface markers and definition of cell type. We then distributed each cell into a specific group type: COVID-19^{+} / CH^{-}, COVID-19^{+} / CH^{+}, or COVID-19^{-} / CH^{+} for determination of the distribution of cell types comprising *TET2*^{MT} versus *TET2*^{WT} for each group.

Single-cell profiling of PBMCs (single-cell RNA-seq and 10X Genomics Multiome)

Frozen PBMCs were thawed in a 37°C water bath for 3-5 min until no ice was visible. Cells were washed twice with 1 mL PBS + 0.04% BSA and pelleted (300×g for 5 min at 4°C). Dead cells were removed according to the Demonstrated Protocol: Removal of Dead Cells from Single Cell Suspensions for Single Cell RNA Sequencing (10X Genomics). Using the MACS Dead Cell Removal Kit (Miltenyi Biotec), the pellet was resuspended in 100 µL Dead Cell Removal MicroBeads and incubated for 15 minutes at room temperature. After incubation, the cell suspension was diluted with 1X Binding Buffer and applied to a MS column. The dead cells were retained on the column and the live cells passed through the column and were collected. After dead cell removal, the samples were washed twice with 1 mL PBS + 0.04% BSA and the cell concentration was determined using a Cellometer K2 cell counter (Nexcelom Biosciences). The cells were then aliquoted for scRNA-seq and Multiome (see Methods section).

Single-cell RNA-seq

The cells were first counted and measured for viability using either the Vi-Cell XR Cell Viability Analyzer (Beckman-Coulter, Brea, CA) or a basic hemocytometer and light microscope. The barcoded Gel Beads were thawed from -80°C and the cDNA master mix was prepared according to the manufacturer's instruction for Chromium Next GEM Single Cell 3' Library and Gel Bead Kit (10x Genomics, Pleasanton, CA). Based on the desired number of cells to be captured for each sample, a volume of live cells was mixed with the cDNA master mix. A per sample concentration of 400,000 cells per milliliter or better is required for the standard targeted cell recovery of approximately 4000 cells. The stock concentration requirements would not change for higher cell recovery numbers. The cell suspension and master mix, thawed Gel Beads and partitioning oil were added to a Chromium Single Cell G chip. The filled chip was loaded into the Chromium Controller, where each sample was processed and the individual cells within the sample were captured into uniquely labeled GEMs (Gel Beads-In-Emulsion). The GEMs were collected from the chip and taken to the bench for reverse transcription, GEM dissolution, and cDNA clean-up. The resulting cDNA contained a pool of uniquely barcoded molecules. A portion of the cleaned and measured pooled cDNA continued to library construction, where standard Illumina sequencing primers and a 10x Genomics unique i7 Sample index were added to each cDNA pool. All cDNA pools and resulting libraries were measured using Qubit High Sensitivity assays (Thermo Fisher Scientific, Waltham, MA) and Agilent Bioanalyzer High Sensitivity chips (Agilent, Santa Clara, CA).

Libraries are sequenced at between 40,000 and 50,000 fragment reads per cell following Illumina's standard protocol using the Illumina cBot and HiSeq 3000/4000 PE Cluster Kit (Illumina, San Diego, CA). The flow cells are sequenced as 100 X 2 paired end reads on an Illumina HiSeq 4000 HD using HiSeq 3000 / 4000 sequencing kit and HCS v3.4.0.38 collection software. Base-calling is performed using Illumina's RTA version 2.7.7.

Single-cell Multiome

After approximately 4,000 cells were aliquoted for scRNA-seq, the remainder were used for single-cell Multiome ATAC + Gene Expression (10X Genomics). Nuclei were isolated according to the Demonstrated Protocol: Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression (10x Genomics, CG000365 Rev A). Briefly, cells were added to a 2.0 mL low binding tube and centrifuged (300×g for 5 min at 4°C) using a swinging bucket rotor. The supernatant was removed, and the cell pellet was resuspended in 100 µL of chilled 10x Genomics Lysis Buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.1 % NP-40 Substitute, 0.01% digitonin, 1% BSA, 1 mM DTT, 1 U/µL RNase inhibitor 40 U/mL) by pipette-mixing 10 times. Cells were incubated on ice for 3 min, followed by dilution with 1 mL of chilled Wash Buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 1% BSA, 1 mM DTT, 1 U/mL RNase inhibitor 40 U/mL). Nuclei were then centrifuged (500×g for 3 min at 4°C), and the supernatant was slowly removed. The nuclei were washed one additional time with 1 mL Wash Buffer. Nuclei were resuspended in chilled diluted nuclei buffer (1X Nuclei Buffer (10x Genomics), 1 mM DTT, 1 U/mL RNase inhibitor 40 U/mL); the concentration was determined using a Cellometer K2 cell counter (Nexcelom Biosciences) and the samples were adjusted to a concentration appropriate for our targeted nuclei recovery. The single-cell ATAC library construction and gene expression library construction were carried out as described in the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide (CG000338 Rev A). ATAC and GEX libraries were sequenced separately on an HiSeq 4000 (Illumina) before demultiplexing, alignment to the reference genome, and post-alignment quality control.

DNA Methylation

Genomic DNA was isolated and checked for quality by standard protocols. 1 µg genomic DNA then underwent bisulfite treatment using the TrueMethyl oxBS Module (Tecan Genomics, Männedorf, Switzerland) according to the manufacturer's specifications. Briefly, DNA was purified using magnetic beads, denatured, and underwent bisulfite conversion followed by desulfonation and purification. The TrueMethyl converted DNA samples were then eluted in 10 µL and then processed through the Illumina Infinium MethylationEPIC BeadChip array (Illumina, San Diego, CA) protocol. In brief, 7 µL of converted DNA was denatured with 1 µL of 0.4N sodium hydroxide prior to whole genome amplification on the MSA4 plate. All other steps were followed as per manufacturer's guidelines.

Quality control of Infinium MethylationEPIC BeadChips was performed via the Genome Studio Methylation Module (Illumina). Subset-quantile Within Array Normalization (SWAN) was performed on the Infinium MethylationEPIC BeadChip IDAT files via the R package "minfi".[63,64] The resultant β-values are used to identify changes in DNA methylation (Δβ) between groups. Unless otherwise noted, a change in absolute methylation level of 10% ($\Delta\beta > |0.1|$) and a *p* value of < 0.01 were considered significant.

CpG site relation within candidate *cis*-regulatory elements (cCREs) was annotated using Bedtools v2.27.1 to the ENCODE Registry of candidate cis-Regulatory Elements (cCREs) in the human genome.[25] To evaluate if the differentially methylated CpGs were located in transcription factor binding sites, ReMAP was used to assess the enrichment of transcription factors binding at these differentially methylated regions.[33] This tool overlaps the differentially methylated regions with a catalog of transcription factor binding profiles collected from public datasets of ChIP-seq experiments. A myeloid specific library was made by selecting all peaks from macrophages, monocytes, and dendritic cells.

Statistical analysis

Distribution of continuous variables was statistically compared using Mann-Whitney or Kruskal-Wallis tests, while nominal or categorical variables were compared using the Chi-Square or Fischer's exact test. Time to event analyses used the method of Kaplan-Meier for univariate comparisons using the log-rank test. OS was calculated from the date of diagnosis to date of death or last follow-up, while AML-free survival (LFS) was calculated from date of diagnosis to date of AML transformation or death. Poisson regression models were fit to compare the numbers of different cell types across groups ($COVID^{+/-}$ and $CHIP^{+/-}$). These Poisson models incorporated an offset term to reflect the total number of cells for a given patient (to be able to compare cell types across consistent intervals in the setting of varying total cell counts: $\ln(Y) = \ln(N) + \beta_0 + \beta_1 X_1 + \epsilon$).

Single-cell data analysis

Single-cell RNA-seq: Sequenced reads from the droplet libraries were processed using 10x Genomics Cell Ranger v6.0.1.[65] The reads were aligned to the pre-built human reference transcriptome GRCh38 - v2020-A (July 7, 2020) provided by 10X Genomics. Read trimming, alignment, UMI counting, and cell calling were performed by Cell Ranger. Doublet prediction on scRNA-seq data was done using Scrublet v0.2.1 with default parameters.[66] Downstream processing was done using Seurat v4.0.4.[67] Count matrices from all samples were combined and batch-corrected using Seurat v4 integration method. Count matrix from each sample was log-normalized, scaled to mean 0 and variance 1, and dimensionality reduced using PCA on the top

2000 variable genes across all samples. The reciprocal PCA and reference-based integration options were applied in the anchor finding step due to large data size. Six patient samples, one from each sex and each condition (COVID-19$^+$ / CH$^-$, COVID-19$^-$ / CH$^+$, COVID-19$^+$ / CH$^+$), were chosen as references, and PCs 1-50 were used for the reference-based integration. The integrated data included a total of 114,453 cells from 35 patient samples. Cells with more than 50% of reads mapped to mitochondrial genes, those with less than 200 unique genes detected and those that were predicted as doublets by Scrublet were removed. QC statistics from Cell Ranger are reported in Supp. Table 8. After QC filtering the number of cells was reduced to 105,169. Genes that were not expressed in at least 3 cells were also removed from downstream analysis. Uniform manifold approximation and projection (UMAP) was made using the top 50 PCs obtained by running PCA on the integrated (batch-corrected) gene expression matrix. Cell type identification was done with SingleR v1.10.0 using immune data from celldex v1.6.0 as reference.[68,69]

Single-cell Multiome: Sequenced reads from the gene expression (GEX) and DNA accessibility (ATAC) droplet libraries of the Multiome assay were processed using 10x Genomics Cell Ranger ARC v2.0.0. The reads were aligned to the pre-built human reference genome GRCh38 - v2020-A-2.0.0 (May 3, 2021) provided by 10X Genomics. Read trimming, alignment, duplicate marking (ATAC), UMI counting (GEX), peak calling (ATAC) and joint cell calling were performed by Cell Ranger. Downstream processing was done using Seurat v4.0.4 and Signac v1.4.0.[70] GEX and ATAC count matrices were integrated (batch-corrected) independently using Seurat. Sample GEX count matrices were integrated following the same steps as used for the scRNA-seq data. Default options (CCA and pairwise anchor-finding) were used in the integration anchor finding step, since the data size was smaller than our scRNA-seq data. To merge the ATAC data from all samples, a common peak set was created by merging peaks from all samples using the reduce function from the R package GenomicRanges. Peaks that were smaller than 20 base pairs or larger than 10000 base pairs after merging were removed. The count matrix for each sample with the new common peak set was then recalculated using Signac. The count matrices were normalized using Term Frequency - Inverse Document Frequency (TF-IDF) and dimensionality reduced using singular value decomposition (SVD) using only peaks with non-zero counts in at least 20 cells, these two steps together known as latent semantic indexing (LSI) generating LSI components. The samples were then integrated using Seurat V4 integration with the reciprocal LSI (rLSI) method used on LSI components 2 to 50 (since the first LSI component correlates with sequencing depth) in the pairwise anchor finding step. The UMAP was calculated using the integrated LSI components 2 to 50. Seurat's weighted nearest neighbor (WNN) algorithm was used on principal components 1 to 50 (GEX) and integrated LSI components 2 to 50 (ATAC) together to obtain a combined UMAP projection of both GEX and ATAC counterparts of the complete scMultiome dataset. The integrated data included 57,817 cells from 18 patient samples. Cells with more than 50% of reads mapped to mitochondrial genes, those with less than 200 unique genes detected (GEX), those with less than 200 unique peaks detected (ATAC) and those with transcription start site (TSS) enrichment score (as calculated by Signac) less than 1 were removed. After QC filtering the number of cells was reduced to 43,160. Cell type identification was done by using Azimuth algorithm to map the scMultiome GEX data to the scRNA-seq data since they were generated from the same cohort. The labels were then transferred from the scRNA-seq data to the single-cell Multiome data.

Differential gene expression analysis: Differential gene expression testing was done on the log-normalized counts using Seurat's FindMarkers function with default parameters unless specified otherwise. The statistical test applied was Wilcoxon Rank Sum test with p-values adjusted using Bonferroni correction based on the total number of genes in the dataset. Statistically significant differentially expressed genes were selected by keeping only genes that fall below the adjusted

p-value threshold of 0.05. Differential gene expression testing comparing any two conditions (e.g., COVID-19$^+$ / *TET2*^{MT} versus COVID-19$^+$ / CH$^-$) was always done for each cell type independently (although shown together in the volcano plots for efficient visualization), unless specified otherwise. To make sure that the results were not driven by a single patient sample we applied a leave-one-out approach on all tests where we removed cells from one sample at a time redoing the tests and keeping only the genes that passed the adjusted p-value threshold in all tests. Pathway analysis was done using the Ingenuity Pathway Analysis platform on differentially expressed genes.

Differential DNA accessibility and motif analysis: The differentially accessible peaks were identified by comparing the TF-IDF normalized cut-site counts of any two pair of cell populations using Seurat's FindMarkers function. Here, the method used was the logistic regression framework along with testing for the number of fragments in peaks as a latent variable. Motif enrichments in peaks were estimated using ChromVAR enrichment scores on the JASPAR2020 motif matrix set.[71,72] Enrichment of cut-sites in monocyte and macrophage specific ChIP-seq peaks of select DNA binding proteins were also estimated using ChromVAR by providing ChIP-seq peaks from ReMap2022 as input.[33] Co-accessibility scores between pairs of peaks were calculated using Cicero v1.3.5.[73]

Cell culture and shRNA Knockdown of EGR1

All cell lines were purchased from the American Type Culture collection (ATCC). Human Erythroleukemia (HEL) and human TF-1 cells were cultured in RPMI-1640 medium (Sigma Aldrich #R8758) supplemented with 10% Fetal Bovine Serum (Sigma Aldrich #F0926) at 37°C, 5% CO_2. Growth medium for TF-1 was also supplemented with 2ng/mL of GM-CSF (R&D systems #7954-GM-010). Mycoplasma PCR testing of these cells was performed every four weeks. Cells were grown for no more than 20 passages in total for any experiment.

Viral particles were made from shRNA lentiviral vector targeting human EGR1 (Sigma-Aldrich MISSION shRNA Target Clone ID TRCN000027385, target sequence CATCTCTCTGAACAACGAGAA) using the two-vector packaging system. Non-target PLKO.1 shRNA viral particles were also made with the same system using the TRC non-target control lentiviral vector (Addgene # 10879). HEL and TF-1 cells were infected with EGR1 shRNA and non-target shRNA viral particles using spinfection and infected cells were selected with puromycin (2 µg/mL for TF-1 and 4 µg/mL for HEL). Knockdown of EGR1 in selected cells was confirmed by western blot.

Western blot analysis

HEL and TF-1 cells were lysed with RIPA buffer with Halt protease and phosphatase inhibitors (Life Technologies, #78440) on ice for 30 minutes followed by centrifugation at 14,000 rpm for 15 minutes. Lysed samples were normalized by protein concentration determined using the bicinchoninic assay (BCA) (Fisher Scientific #23227). Normalized samples were heated in SDS (Sodium dodecyl sulfate) loading buffer then loaded into 10% bis-tris gels. Transfer to PVDF membrane was done overnight. Membranes were then incubated with primary rabbit anti EGR1 antibody (Cell Signaling Technology, #4153S) at 4°C overnight. Vinculin (Sigma-Aldrich #V9131) and beta actin (Thermo Fisher Scientific #4333762F) were used as loading controls.

RT-qPCR

Total RNA was extracted using the Quick-RNA miniprep kit (Zymo Research #R1055). 1µg of total RNA for each sample was reverse transcribed to cDNA using the High-Capacity RNA-to-cDNA™ Kit (Thermo Fisher Scientific #4387406). Each PCR reaction was carried out in three technical replicates and four biological replicates. Primers used for RT-qPCR are as follows: human NEAT1 FWD: TGGCTAGCTCAGGGCTTCAG human NEAT1 REV: TCTCCTTGCCAAGCTTCCTTC, human 18S FWD: GCTTAATTTGACTCAACACGGGA, human 18S REV: AGCTATCAATCTGTCAATCCTGTC, human MALAT1 FWD: GACGGAGGTTGAGATGAAGC, human MALAT1 REV: ATTCGGGGCTCTGTAGTCCT

Availability of data and materials

The bulk sequencing, scDNA-seq + proteogenomics, DNA Methylation, scRNA-seq and Multiome (GEX + ATAC) datasets generated and analyzed in this study have been deposited into the NCBI Gene Expression Omnibus (GEO) data base (https://www.ncbi.nlm.nih.gov/geo/) with accession number GSE210435 (reviewers access token: cfyhqccwzbitfij). Code and scripts used for analysis are made available in the GitHub repository https://github.com/LabFunEpi/CC_multiomics.

REFERENCES

1. Goyal P, Choi JJ, Pinheiro LC, et al. Clinical Characteristics of Covid-19 in New York City. N Engl J Med 2020;382:2372-4.
2. Moore JB, June CH. Cytokine release syndrome in severe COVID-19. Science 2020;368:473-4.
3. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA 2020;323:1775-6.
4. Sah P, Fitzpatrick MC, Zimmer CF, et al. Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. Proc Natl Acad Sci U S A 2021;118.
5. Silvin A, Chapuis N, Dunsmore G, et al. Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19. Cell 2020;182:1401-18 e18.
6. Kapellos TS, Bonaguro L, Gemund I, et al. Human Monocyte Subsets and Phenotypes in Major Chronic Inflammatory Diseases. Front Immunol 2019;10:2035.
7. Carr RM, Vorobyev D, Lasho T, et al. RAS mutations drive proliferative chronic myelomonocytic leukemia via a KMT2A-PLK1 axis. Nat Commun 2021;12:2901.
8. Selimoglu-Buet D, Badaoui B, Benayoun E, et al. Accumulation of classical monocytes defines a subgroup of MDS that frequently evolves into CMML. Blood 2017;130:832-5.
9. Selimoglu-Buet D, Riviere J, Ghamlouch H, et al. A miR-150/TET3 pathway regulates the generation of mouse and human non-classical monocyte subset. Nat Commun 2018;9:5455.
10. Jaiswal S, Ebert BL. Clonal hematopoiesis in human aging and disease. Science 2019;366.
11. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 2014;371:2488-98.
12. Jaiswal S, Natarajan P, Silver AJ, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. N Engl J Med 2017;377:111-21.
13. Kusne Y, Lasho T, Mangaonkar A, et al. Remarkable stability in clonal hematopoiesis involving leukemia-driver genes in patients without underlying myeloid neoplasms. Am J Hematol 2021;96:E392-E6.
14. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nat Commun 2016;7:12484.
15. Bolton KL, Koh Y, Foote MB, et al. Clonal hematopoiesis is associated with risk of severe Covid-19. Nat Commun 2021;12:5975.
16. Gondek LP, DeZern AE. Assessing clonal haematopoiesis: clinical burdens and benefits of diagnosing myelodysplastic syndrome precursor states. Lancet Haematol 2020;7:e73-e81.
17. Zhang Q, Zhao K, Shen Q, et al. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. Nature 2015;525:389-93.
18. Lv L, Wang Q, Xu Y, et al. Vpr Targets TET2 for Degradation by CRL4(VprBP) E3 Ligase to Sustain IL-6 Expression and Enhance HIV-1 Replication. Mol Cell 2018;70:961-70 e5.
19. Pronier E, Imanci A, Selimoglu-Buet D, et al. Macrophage migration inhibitory factor is overproduced through EGR1 in TET2(low) resting monocytes. Commun Biol 2022;5:110.
20. Trizzino M, Zucco A, Deliard S, et al. EGR1 is a gatekeeper of inflammatory enhancers in human macrophages. Sci Adv 2021;7.
21. Patnaik MM, Barraco D, Lasho TL, et al. DNMT3A mutations are associated with inferior overall and leukemia-free survival in chronic myelomonocytic leukemia. Am J Hematol 2017;92:56-61.
22. Common Terminology Criteria for Adverse Events (CTCAE) v5.0. 2017. athttps://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm.)
23. Clinical management of COVID-19: Living guideline. 2020. at https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2022-1.)
24. Tulstrup M, Soerensen M, Hansen JW, et al. TET2 mutations are associated with hypermethylation at key regulatory enhancers in normal and malignant hematopoiesis. Nat Commun 2021;12:6061.

25. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-74.

26. Ko M, Rao A. TET2: epigenetic safeguard for HSC. Blood 2011;118:4501-3.

27. Miles LA, Bowman RL, Merlinsky TR, et al. Single-cell mutation analysis of clonal evolution in myeloid malignancies. Nature 2020;587:477-82.

28. Buscarlet M, Provost S, Zada YF, et al. Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. Blood 2018;132:277-80.

29. Lucas N, Duchmann M, Rameau P, et al. Biology and prognostic impact of clonal plasmacytoid dendritic cells in chronic myelomonocytic leukemia. Leukemia 2019;33:2466-80.

30. Mangaonkar AA, Reichard KK, Binder M, et al. Bone marrow dendritic cell aggregates associate with systemic immune dysregulation in chronic myelomonocytic leukemia. Blood Adv 2020;4:5425-30.

31. Stephenson E, Reynolds G, Botting RA, et al. Single-cell multi-omics analysis of the immune response in COVID-19. Nat Med 2021;27:904-16.

32. Zhao Q, Wang J, Levichkin IV, Stasinopoulos S, Ryan MT, Hoogenraad NJ. A mitochondrial specific stress response in mammalian cells. EMBO J 2002;21:4411-9.

33. Hammal F, de Langen P, Bergon A, Lopez F, Ballester B. ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. Nucleic Acids Res 2022;50:D316-D25.

34. Melenotte C, Silvin A, Goubet AG, et al. Immune responses during COVID-19 infection. Oncoimmunology 2020;9:1807836.

35. Ko M, An J, Bandukwala HS, et al. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature 2013;497:122-6.

36. Abdel-Wahab O, Mullally A, Hedvat C, et al. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood 2009;114:144-7.

37. Busque L, Patel JP, Figueroa ME, et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. Nat Genet 2012;44:1179-81.

38. Coltro G, Mangaonkar AA, Lasho TL, et al. Clinical, molecular, and prognostic correlates of number, type, and functional localization of TET2 mutations in chronic myelomonocytic leukemia (CMML)-a study of 1084 patients. Leukemia 2020;34:1407-21.

39. Itzykson R, Kosmider O, Renneville A, et al. Clonal architecture of chronic myelomonocytic leukemias. Blood 2013;121:2186-98.

40. Cook EK, Izukawa T, Young S, et al. Comorbid and inflammatory characteristics of genetic subtypes of clonal hematopoiesis. Blood Adv 2019;3:2482-6.

41. Sano S, Oshima K, Wang Y, et al. Tet2-Mediated Clonal Hematopoiesis Accelerates Heart Failure Through a Mechanism Involving the IL-1beta/NLRP3 Inflammasome. J Am Coll Cardiol 2018;71:875-86.

42. Deshmane SL, Kremlev S, Amini S, Sawaya BE. Monocyte chemoattractant protein-1 (MCP-1): an overview. J Interferon Cytokine Res 2009;29:313-26.

43. Wong KL, Tai JJ, Wong WC, et al. Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets. Blood 2011;118:e16-31.

44. Menezes S, Melandri D, Anselmi G, et al. The Heterogeneity of Ly6C(hi) Monocytes Controls Their Differentiation into iNOS(+) Macrophages or Monocyte-Derived Dendritic Cells. Immunity 2016;45:1205-18.

45. Patnaik MM, Timm MM, Vallapureddy R, et al. Flow cytometry based monocyte subset analysis accurately distinguishes chronic myelomonocytic leukemia from myeloproliferative neoplasms with associated monocytosis. Blood Cancer J 2017;7:e584.

46. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, et al. Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. Cell Host Microbe 2020;27:992-1000 e3.

47. Chen G, Wu D, Guo W, et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. J Clin Invest 2020;130:2620-9.

48. Ma XY, Wang JH, Wang JL, Ma CX, Wang XC, Liu FS. Malat1 as an evolutionarily conserved lncRNA, plays a positive role in regulating proliferation and maintaining undifferentiated status of early-stage hematopoietic cells. BMC Genomics 2015;16:676.

49. Cui H, Banerjee S, Guo S, et al. Long noncoding RNA Malat1 regulates differential activation of macrophages and response to lung injury. JCI Insight 2019;4.

50. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016;127:2391-405.

51. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.

52. Larson DE, Harris CC, Chen K, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics 2012;28:311-7.

53. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012;6:80-92.

54. Fiume M, Williams V, Brook A, Brudno M. Savant: genome browser for high-throughput sequencing data. Bioinformatics 2010;26:1938-44.

55. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 2014;42:D980-5.

56. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Hum Mutat 2011;32:894-9.

57. Stenson PD, Ball EV, Mort M, et al. Human Gene Mutation Database (HGMD): 2003 update. Hum Mutat 2003;21:577-81.

58. Tryggvadottir L, Vidarsdottir L, Thorgeirsson T, et al. Prostate cancer progression and survival in BRCA2 mutation carriers. J Natl Cancer Inst 2007;99:929-35.

59. Robinson D, Van Allen EM, Wu YM, et al. Integrative clinical genomics of advanced prostate cancer. Cell 2015;161:1215-28.

60. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24-6.

61. Mule MP, Martins AJ, Tsang JS. Normalizing and denoising protein expression data from droplet-based single cell profiling. Nat Commun 2022;13:2099.

62. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. Cell 2019;177:1888-902 e21.

63. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363-9.

64. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol 2012;13:R44.

65. Zheng GX, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun 2017;8:14049.

66. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst 2019;8:281-91 e9.

67. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. Cell 2021;184:3573-87 e29.

68. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 2019;20:163-72.

69. Monaco G, Lee B, Xu W, et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. Cell Rep 2019;26:1627-40 e7.

70. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. Nat Methods 2021;18:1333-41.

71. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res 2022;50:D165-D73.

72. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 2017;14:975-8.

73. Pliner HA, Packer JS, McFaline-Figueroa JL, et al. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. Mol Cell 2018;71:858-71 e8.

TABLE 1

Table 1 Clinical and laboratory characteristics of 227 patients with COVID-19 stratified by clonal hematopoiesis status.

Characteristic	**All patients** (n=227)	**No CH** (n=171)	**CH** (n=56)	**p-value**
Age at CH evaluation (years)				
Median (range)	60 (19 – 99)	57 (19 – 93)	69.5 (42 – 99)	<0.0001
Mean ± sd	58.3 ±17.8	54.3 ±17.0	70.5 ±17.8	
<75 years		154	33	<0.0001
75+ years		17	23	
<40 years		37	0	
40-74 years		117	33	<0.0001
75+ years		17	23	
<60 years		93	16	
60-74 years		61	17	<0.0001
75+ years		17	23	
Sex				
Female	100	78	22	0.50
Male	127	93	34	
Race/ethnicity group				
NHW	162	119	43	
NH Black/African Am.	23	16	7	
NH Asian	7	7	0	0.20
Hispanic or Latino	19	14	5	
Unknown/missing/refused	16	15	1	
NHW	162	119	43	
NHB	23	16	7	
Hispanic or Latino	19	14	5	0.08
Other race / unknown	23	22	1	
Comorbidities				
Diabetes	7	4	10	
Obesity	10	10	3	
Hypertension	34	24	0	
Coronary heart disease	1	1	10	0.12
COPD	9	5	0	
More than one	79	58	4	
None	53	45	21	
Missing	34	24	8	
No comorbidities reported	53	45	8	
One comorbidity reported	61	44	17	0.32
2+ comorbidities reported	79	58	21	
Missing	34	24	10	

CH = Clonal Hematopoiesis | NHW = Non-Hispanic White | NHB = Non-Hispanic Black | p-value reflects the comparison of the measures or factors between those with versus without CH. For continuous measures, a Wilcoxon rank sum test was used, for categorical measures, a chi-square or Fisher's exact test was used, depending on numbers in each subset

FIGURES

Figure 1

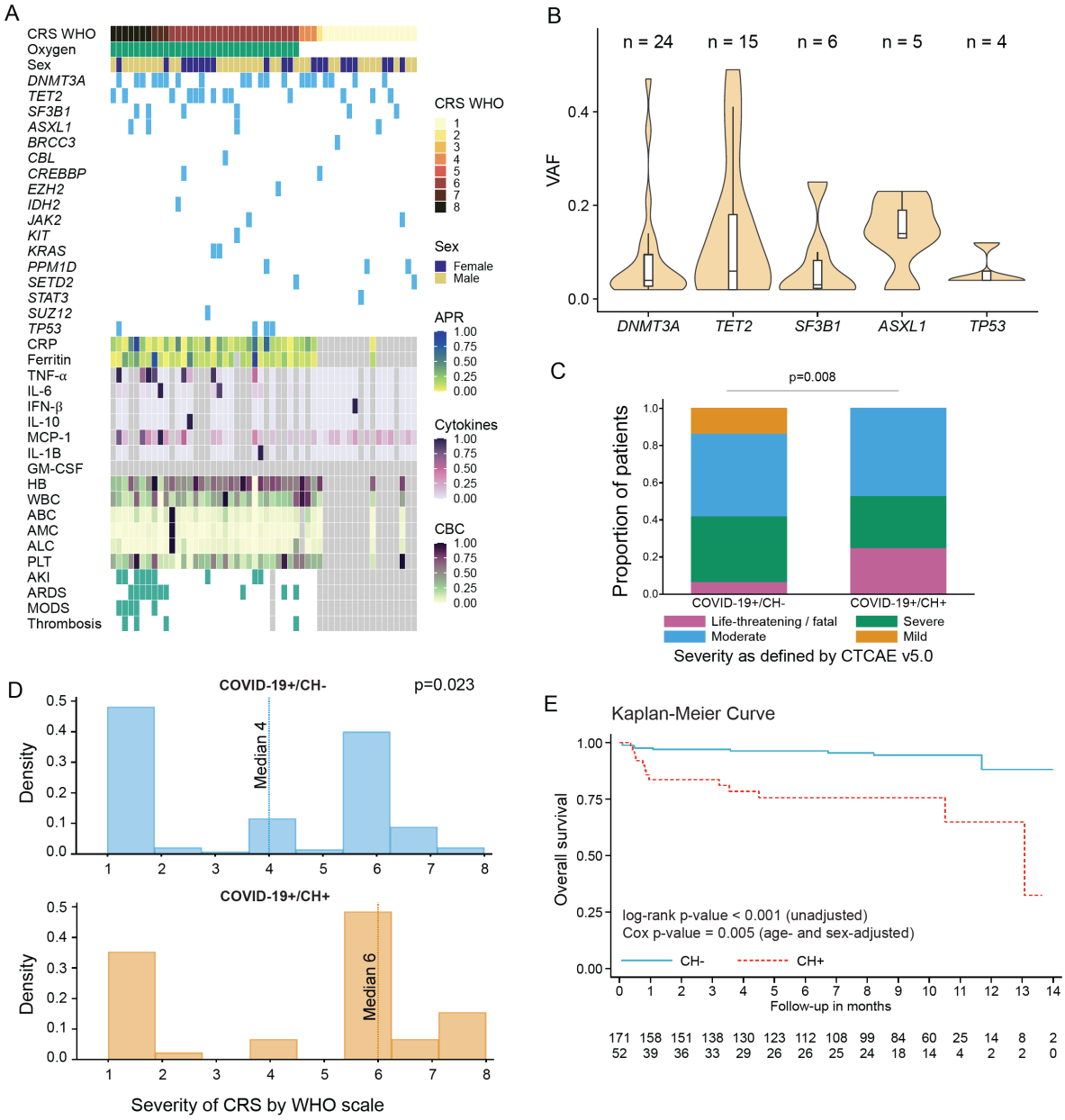


Figure 1. Clonal hematopoiesis is associated with increased inflammatory morbidity and mortality in COVID-19 infection. A, Heatmap showing the patient information (each column represents a single patient) for all CH patients included in the study (stratified by CH status). Rows with gene names (italicized) show presence (color) or absence (white) of a given mutation. APR, Cytokines and CBC are min-max scaled within their category. CRS, cytokine release syndrome; APR, acute phase reactants; CBC, complete blood count; CRP, C-reactive protein; HGB, hemoglobin; WBC, white blood cell count; ANC, absolute neutrophil count; AMC, absolute monocyte count, ALC, absolute lymphocyte count; PLT, platelet count; AKI, acute kidney injury; ARDS, acute respiratory distress syndrome; MODS, multiple organ dysfunction syndrome. B, Violin plots showing the distribution of variant allele frequencies of the five most common CH mutations among the COVID-19$^+$ / CH$^+$ patients included in this study. C, Proportion of COVID-19$^+$ patients with CRS separated by CH$^+$ and CH$^-$. CRS grade is stratified by mild, moderate, severe, and life-threatening as defined by CTCAE v5.0 criteria. Statistical significance in the difference was tested using Fisher's Exact Test. D, Comparison of CRS severity by WHO scale for COVID-19$^+$ / CH$^-$ and COVID$^-$19$^+$ / CH$^-$ cohorts. E, Survival of 223 COVID-19$^+$ patients stratified by CH status. One month OS estimates: CH$^-$ 97.6% versus CH$^+$ 83.7%. One year OS estimates: CH$^-$ 88.1% versus CH$^+$ 64.8%. CH$^+$ association with worse overall survival remained consistent after adjusting for age and sex at the time of COVID-19 diagnosis (HR 3.48, 95% 1.45 - 8.36, $p = 0.005$). There was no evidence for a violation of the proportional hazards assumption ($p = 0.204$; using a hypothesis test based on scaled Schoenfeld residuals).

Figure 2

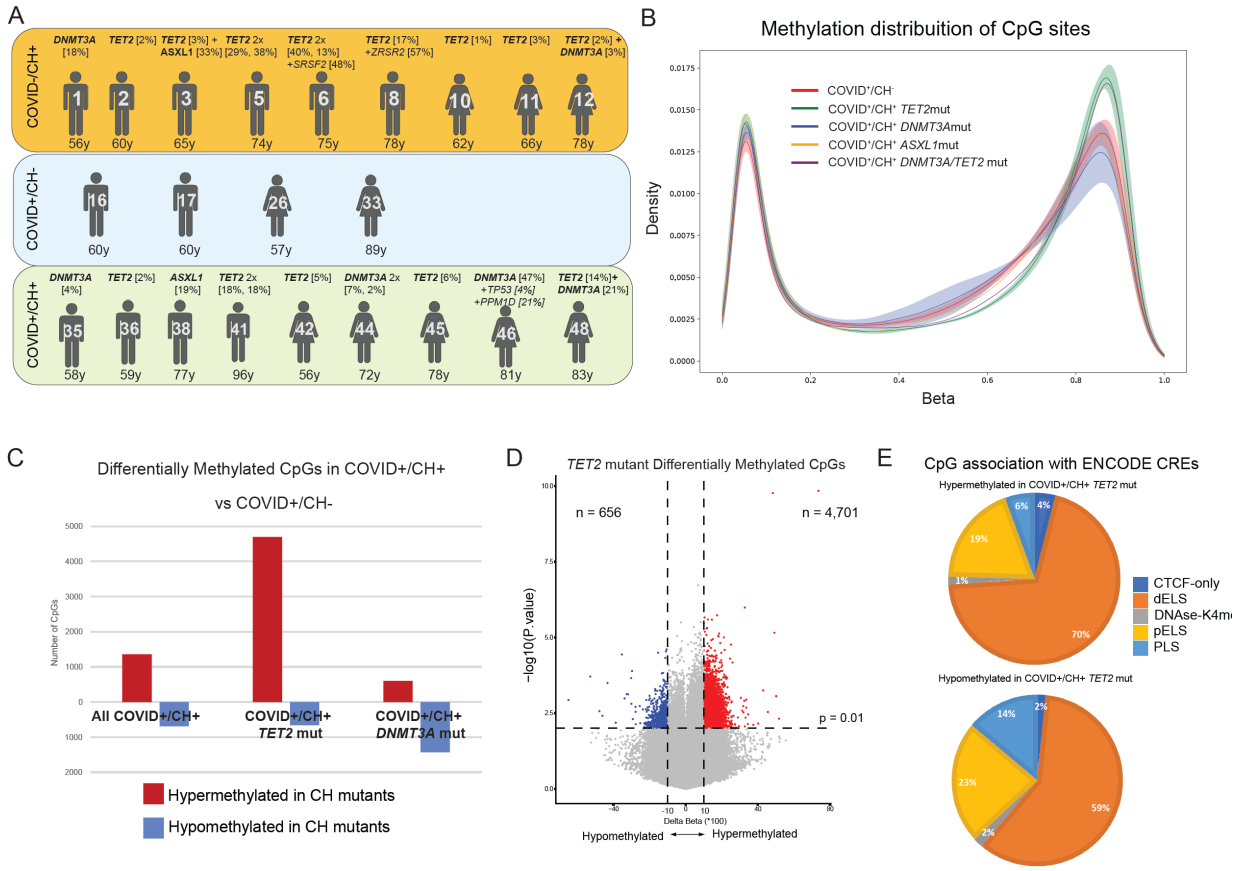

A
COVID-/CH+
DNMT3A [18%]
TET2 [2%]
TET2 [3%] + ASXL1 [33%]
TET2 2x [29%, 38%]
TET2 2x [40%, 13%] +SRSF2 [48%]
TET2 [17%] +ZRSR2 [57%]
TET2 [1%]
TET2 [3%]
TET2 [2%] + DNMT3A [3%]
1 2 3 5 6 8 10 11 12
56y 60y 65y 74y 75y 78y 62y 66y 78y
COVID+/CH-
16 17 26 33
60y 60y 57y 89y
COVID+/CH+
DNMT3A [4%]
TET2 [2%]
ASXL1 [19%]
TET2 2x [18%, 18%]
TET2 [5%]
DNMT3A 2x [7%, 2%]
TET2 [6%]
DNMT3A [47%] +TP53 [4%] +PPM1D [21%]
TET2 [14%]+ DNMT3A [21%]
35 36 38 41 42 44 45 46 48
58y 59y 77y 96y 56y 72y 78y 81y 83y
B
Methylation distribution of CpG sites
COVID+/CH-
COVID+/CH+ TET2mut
COVID+/CH+ DNMT3Amut
COVID+/CH+ ASXL1mut
COVID+/CH+ DNMT3A/TET2 mut
Density
Beta
C
Differentially Methylated CpGs in COVID+/CH+ vs COVID+/CH-
Number of CpGs
All COVID+/CH+
COVID+/CH+ TET2 mut
COVID+/CH+ DNMT3A mut
Hypermethylated in CH mutants
Hypomethylated in CH mutants
D
TET2 mutant Differentially Methylated CpGs
n = 656
n = 4,701
p = 0.01
−log10(P.value)
Delta Beta (*100)
Hypomethylated
Hypermethylated
E
CpG association with ENCODE CREs
Hypermethylated in COVID+/CH+ TET2 mut
Hypomethylated in COVID+/CH+ TET2 mut
CTCF-only
dELS
DNAse-K4me
pELS
PLS
F
Quantity of leukocytes
Cell death of immune cells
Cell death of phagocytes
Apoptosis of phagocytes
Leukocyte migration
Cell movement of phagocytes
Quantity of mast cells
Psoriatic arthritis
Apoptosis of leukocytes
Morphology of leukocytes
Quantity of mononuclear leukocytes
Cell viability of monocyte-derived dendritic cells
Morphology of phagocytes
Cell movement of leukocytes
Quantity of T lymphocytes
Cell viability of leukocytes
-log10(p-value)
G
Myeloid − ReMapEnrich
nb.overlaps
q.significance
SREBP2:monocyte
RXR:macrophage
SMC1A:monocyte
SMC1:monocyte
CDK8:monocyte
CREBBP:monocyte
EGR1:macrophage
STAT3:monocyte
SPI1:DC
CREBBP:monocyte
CEBPB:monocyte
CREBBP:monocyte
SPI1:DC
STAT3:monocyte
SPI1:primary−monocyte,DC
CHD4:macrophage
CREBBP:monocyte
INTS13:monocyte
CEBPB:monocyte
IRF1:CD14,monocyte
mapped.peaks.ratio

Figure 2. Enhancer hypermethylation driven by *TET2* mutations. A, Overview of patient cohorts - COVID-19$^-$ / CH$^+$ (patients carrying CH mutations), COVID-19$^+$ / CH$^-$ (patients diagnosed with COVID-19), COVID-19$^+$ / CH$^+$ (patients carrying CH mutations and diagnosed with COVID-19). CH mutations are shown above, and patient age is shown below. B, Density plot demonstrating the differences in β-value distribution between the COVID-19$^+$ / CH$^-$ and COVID-19$^+$ / CH$^+$ patient cohorts. The COVID-19$^+$ / CH$^+$ cohort is grouped by the mutation type. C, Number of differentially methylated CpG sites comparing the COVID-19$^+$ / CH$^+$ cohort to the COVID-19$^+$ / CH$^-$ cohort, the COVID-19$^+$ / CH$^+$ TET2 mutant cohort to the COVID-19$^+$ / CH$^-$ cohort, and the COVID-19$^+$ / CH$^+$ DNMT3A mutant cohort to the COVID-19$^+$ / CH$^-$ cohort. Bar graph shows the number of CpG sites that are hypo- and hyper-methylated in each. D, Volcano plot of results from *TET2*-specific analyses. Horizontal dashed line indicates $p = 0.01$. The numbers on the upper corners are the hyper- and hypo-methylated CpGs with methylation difference >10%, shown by the vertical dashed lines. Red and blue dots represent hyper- and hypomethylated CpG sites, respectively, below the significance threshold. E, Pie Chart showing the proportion of *TET2*-specific differentially methylated CpGs which intersect with ENCODE candidate Cis-Regulatory Elements (cCREs). F, Ingenuity pathway analysis diseases and functions annotation of differentially methylated CpG sites comparing the COVID-19$^+$ / CH$^+$ TET2 mutant cohort to the COVID-19$^+$ / CH$^-$ cohort. G, Dot plot showing transcription factor binding enrichment in the TET2-specific differentially methylated CpGs. Myeloid cell specific DNA binding sites were obtained from ReMap 2022. The x-axis represents the percentage of transcription factor binding sites overlapping the query CpG list. The bubble size represents the number of CpGs overlapping with transcription factor binding sites. The color represents enrichment significance as measured by -log10 (q-value), which is obtained after adjusting the p-value under the Poisson distribution using randomized query regions.

Figure 3

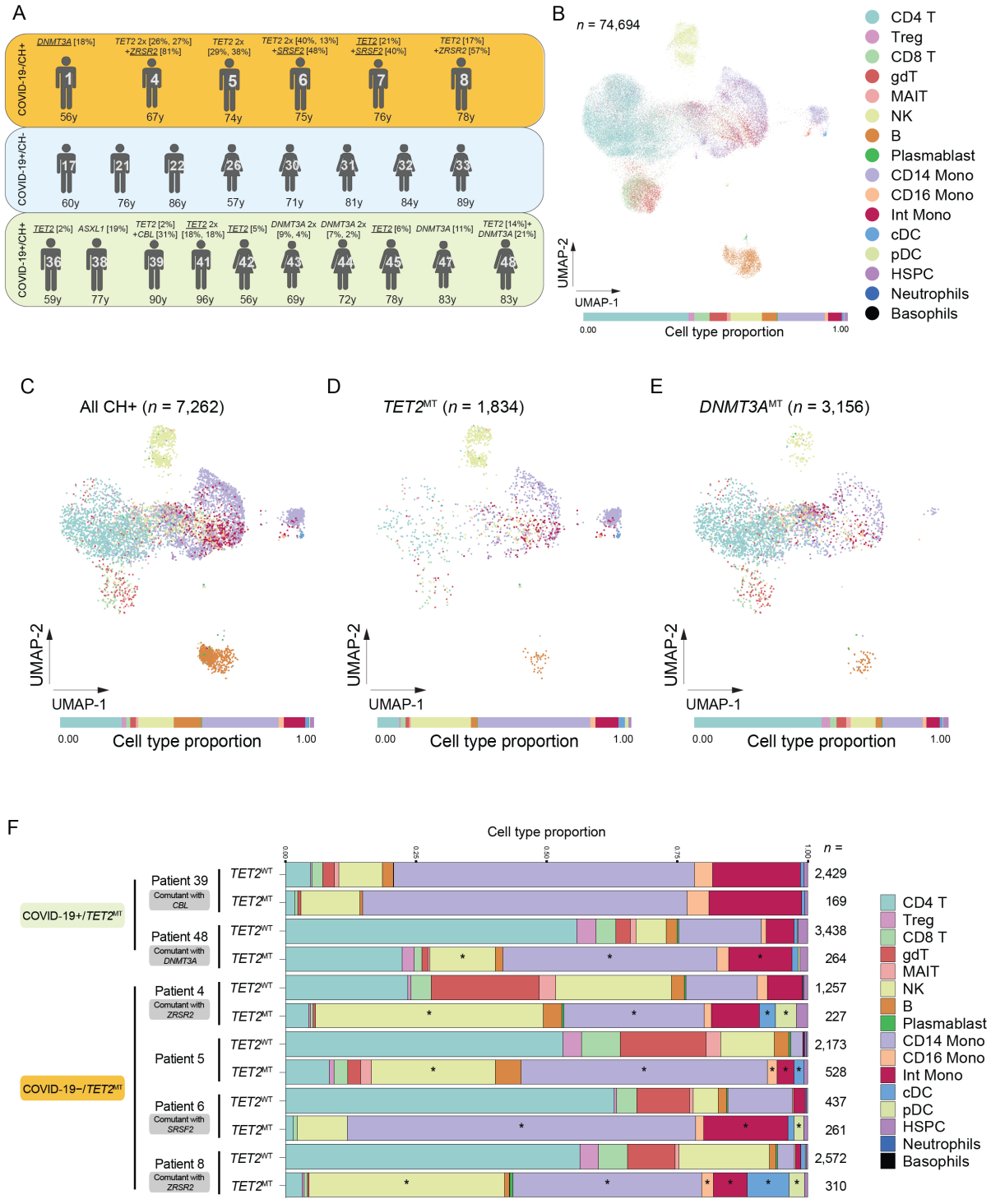

A
COVID-19-/CH+
DNMT3A [18%]
TET2 2x [26%, 27%] +ZRSR2 [81%]
TET2 2x [29%, 38%]
TET2 2x [40%, 13%] +SRSF2 [48%]
TET2 [21%] +SRSF2 [40%]
TET2 [17%] +ZRSR2 [57%]
1 56y
4 67y
5 74y
6 75y
7 76y
8 78y
COVID-19+/CH-
17 60y
21 76y
22 86y
26 57y
30 71y
31 81y
32 84y
33 89y
COVID-19+/CH+
TET2 [2%]
ASXL1 [19%]
TET2 [2%] +CBL [31%]
TET2 2x [18%, 18%]
TET2 [5%]
DNMT3A 2x [9%, 4%]
DNMT3A 2x [7%, 2%]
TET2 [6%]
DNMT3A [11%]
TET2 [14%]+ DNMT3A [21%]
36 59y
38 77y
39 90y
41 96y
42 56y
43 69y
44 72y
45 78y
47 83y
48 83y
B
n = 74,694
UMAP-2
UMAP-1
0.00
Cell type proportion
1.00
CD4 T
Treg
CD8 T
gdT
MAIT
NK
B
Plasmablast
CD14 Mono
CD16 Mono
Int Mono
cDC
pDC
HSPC
Neutrophils
Basophils
C
All CH+ (n = 7,262)
D
TET2MT (n = 1,834)
E
DNMT3AMT (n = 3,156)
F
Cell type proportion
n =
COVID-19+/TET2MT
Patient 39
Comutant with CBL
TET2WT 2,429
TET2MT 169
Patient 48
Comutant with DNMT3A
TET2WT 3,438
TET2MT 264
COVID-19−/TET2MT
Patient 4
Comutant with ZRSR2
TET2WT 1,257
TET2MT 227
Patient 5
TET2WT 2,173
TET2MT 528
Patient 6
Comutant with SRSF2
TET2WT 437
TET2MT 261
Patient 8
Comutant with ZRSR2
TET2WT 2,572
TET2MT 310

Figure 3. Identification of cell types carrying CH mutations by combined single-cell surface protein and genotype analysis. A, Overview of the patient cohorts, i.e., COVID-19^{+} / CH^{+} (patients carrying CH mutations), COVID-19^{+} / CH^{-} (patients diagnosed with COVID-19), and COVID-19^{+} / CH^{+} (patients carrying CH mutations and diagnosed with COVID-19). Variant allele frequencies (VAFs) determined by bulk sequencing are shown in the parentheses next to gene names. Gene name is underlined if the CH mutation identified by bulk sequencing is not recovered by the single-cell platform (Mission Bio Tapestri). Patient age is shown at the bottom. B, UMAP plot based on single-cell surface protein measurements by the MissionBio Tapestri platform, showing 74,694 cells from all 24 patient samples colored by cell types annotated by reference mapping. C−E, UMAP plots showing cells carrying mutations in any of CH genes (C), *TET2* (D), and *DNMT3A* (E). F, Proportions of cell types in *TET2*^{WT} versus *TET2*^{MT} (*TET2* mutated) cells for each patient with *TET2* mutation. Treg, regulatory T-cells; gdT, gamma delta T-cells; MAIT, Mucosal-associated invariant T-cells; NK, natural killer cells; Mono, monocytes; Int, intermediate; cDC, classical dendritic cells; pDC, plasmacytoid dendritic cells; HSPC, hematopoietic stem and progenitor cells.

Figure 4

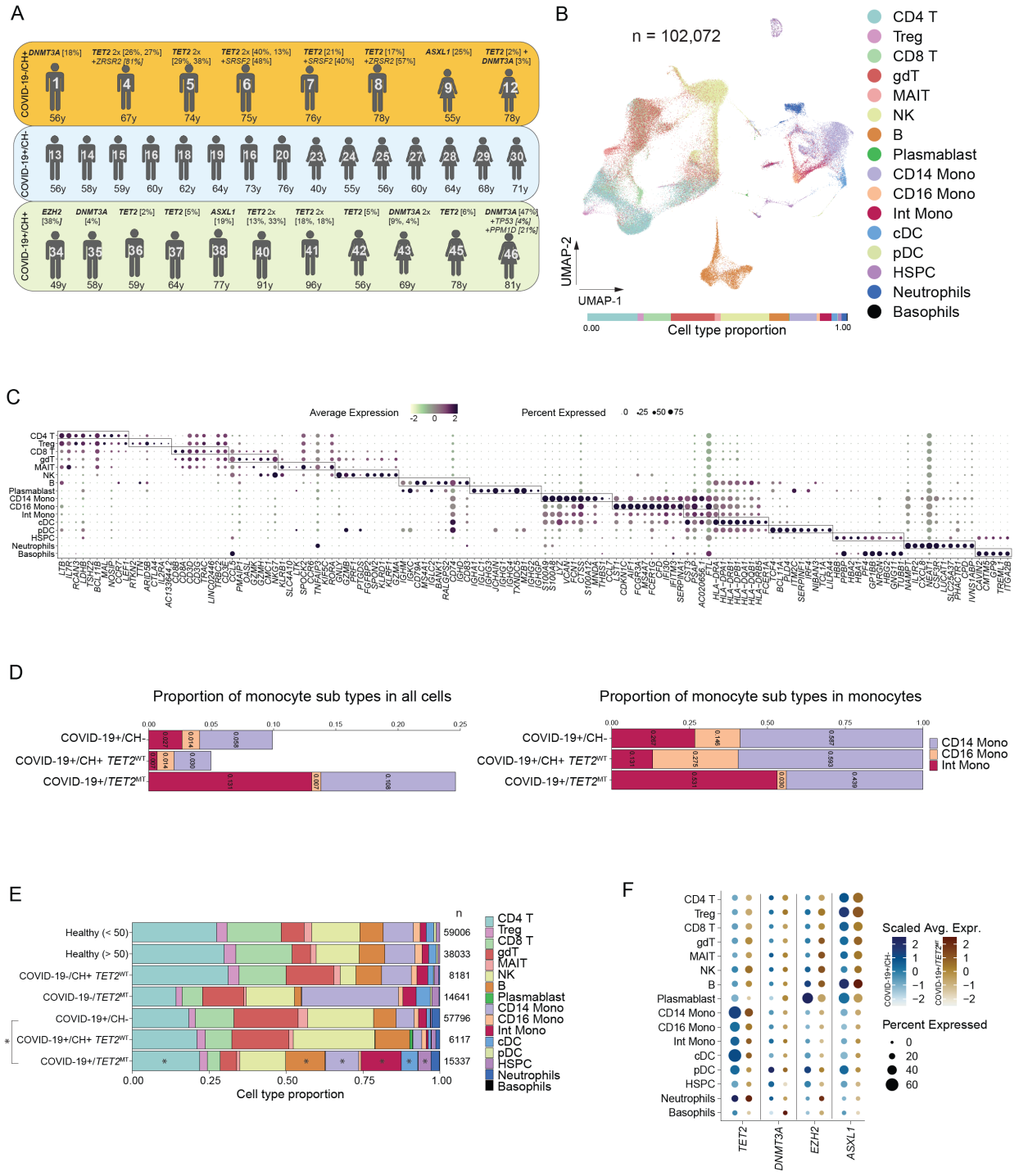

A
COVID-19-/CH+
COVID-19+/CH-
COVID-19+/CH+
B
n = 102,072
UMAP-2
UMAP-1
Cell type proportion
CD4 T
Treg
CD8 T
gdT
MAIT
NK
B
Plasmablast
CD14 Mono
CD16 Mono
Int Mono
cDC
pDC
HSPC
Neutrophils
Basophils
C
Average Expression
Percent Expressed
D
Proportion of monocyte sub types in all cells
Proportion of monocyte sub types in monocytes
COVID-19+/CH-
COVID-19+/CH+ TET2WT
COVID-19+/TET2MT
CD14 Mono
CD16 Mono
Int Mono
E
n
Healthy (< 50)
Healthy (> 50)
COVID-19-/CH+ TET2WT
COVID-19-/TET2MT
COVID-19+/CH-
COVID-19+/CH+ TET2WT
COVID-19+/TET2MT
59006
38033
8181
14641
57796
6117
15337
Cell type proportion
F
Scaled Avg. Expr.
Percent Expressed
TET2
DNMT3A
EZH2
ASXL1

Figure 4. Identification of biomarkers of COVID-19 severity associated with CH mutations. A, Overview of the patient cohorts - COVID-19$^+$ / CH$^+$ (patients carrying CH mutations), COVID-19$^+$ / CH$^-$ (patients diagnosed with COVID-19), COVID-19$^+$ / CH$^+$ (patients carrying CH mutations and diagnosed with COVID-19). Patient age is shown at the bottom. B, UMAP projection showing 105,169 cells from all 35 patient samples colored by cell types identified using SingleR. Treg, regulatory T-cells; gdT, gamma delta T-cells; MAIT, Mucosal-associated invariant T cells; NK, natural killer cells; Mono, monocytes; Int, intermediate; cDC, classical dendritic cells; pDC, plasmacytoid dendritic cells; HSPC, hematopoietic stem and progenitor cells. C, Dot plot showing the log normalized expression of top marker genes differentially expressed in each cell type compared to all other cells. D, Proportion of average number of monocyte subtypes (averaged by sample) showing significant monocyte expansion (see Methods section for hypothesis test), specifically classical monocyte expansion in COVID-19$^+$ / *TET2*^{MT} samples. E, Proportion of each cell type stratified by 7 conditions. Data for the healthy cohort was obtained from previously published work.[31] The healthy cohort is further stratified by age: Below 50 and above 50 years. The COVID-19$^-$ / CH$^+$ cohort and COVID-19$^+$ / CH$^+$ cohort are stratified into two groups: Patients with *TET2* mutations (COVID-19$^+$ / *TET2*^{MT}) and patients with any CH mutation other than *TET2* (COVID-19$^+$ / CH$^+$ / *TET2*^{WT}). Statistical significance of the change in cell-type proportions between the two groups COVID-19$^+$ / CH$^-$ and COVID-19$^+$ / *TET2*^{MT} was tested using a Poisson regression model (see Methods section). Cell types with significant increase in proportions in COVID-19$^+$ / *TET2*^{MT} ($p < 0.05$) are marked with an asterisk. F, Dot plot showing log-normalized gene expression of CH genes across different cell types, in COVID-19$^+$ / CH$^-$ and COVID-19$^+$ / *TET2*^{MT} patients.

Figure 5

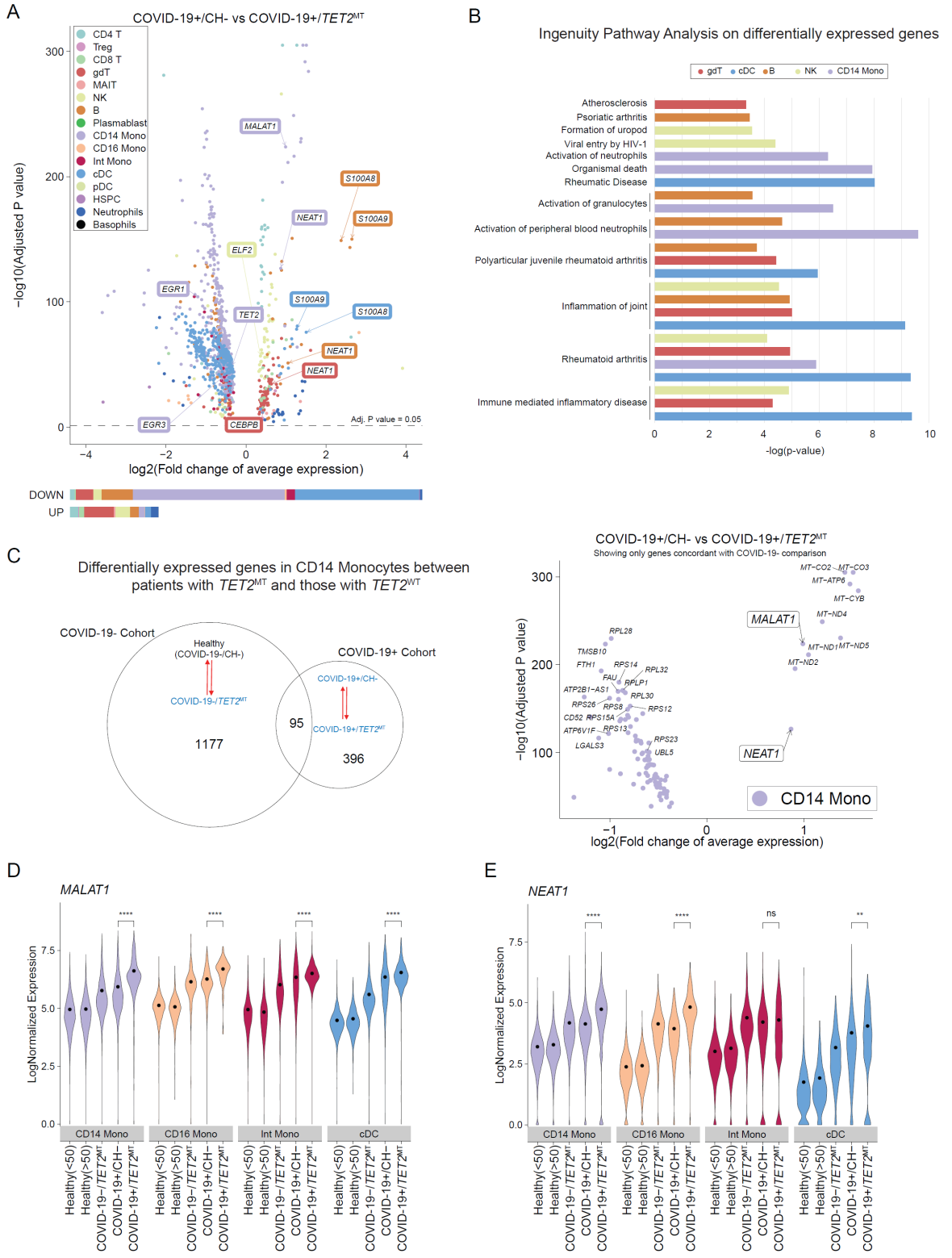

A
COVID-19+/CH- vs COVID-19+/TET2MT
CD4 T
Treg
CD8 T
gdT
MAIT
NK
B
Plasmablast
CD14 Mono
CD16 Mono
Int Mono
cDC
pDC
HSPC
Neutrophils
Basophils
MALAT1
S100A8
NEAT1
S100A9
ELF2
EGR1
TET2
CEBPB
EGR3
Adj. P value = 0.05
−log10(Adjusted P value)
log2(Fold change of average expression)
DOWN
UP
B
Ingenuity Pathway Analysis on differentially expressed genes
gdT
cDC
B
NK
CD14 Mono
Atherosclerosis
Psoriatic arthritis
Formation of uropod
Viral entry by HIV-1
Activation of neutrophils
Organismal death
Rheumatic Disease
Activation of granulocytes
Activation of peripheral blood neutrophils
Polyarticular juvenile rheumatoid arthritis
Inflammation of joint
Rheumatoid arthritis
Immune mediated inflammatory disease
-log(p-value)
C
Differentially expressed genes in CD14 Monocytes between patients with TET2MT and those with TET2WT
COVID-19- Cohort
Healthy (COVID-19-/CH-)
COVID-19-/TET2MT
1177
COVID-19+ Cohort
COVID-19+/CH-
COVID-19+/TET2MT
95
396
COVID-19+/CH- vs COVID-19+/TET2MT
Showing only genes concordant with COVID-19- comparison
MT-CO2
MT-CO3
MT-ATP6
MT-CYB
MT-ND4
MT-ND1
MT-ND5
MT-ND2
MALAT1
NEAT1
RPL28
TMSB10
FTH1
RPS14
RPL32
FAU
RPLP1
ATP2B1-AS1
RPL30
RPS26
RPS8
RPS12
CD52
RPS15A
ATP6V1F
RPS13
LGALS3
RPS23
UBL5
CD14 Mono
−log10(Adjusted P value)
log2(Fold change of average expression)
D
MALAT1
LogNormalized Expression
CD14 Mono
CD16 Mono
Int Mono
cDC
Healthy(<50)
Healthy(>50)
COVID-19−/TET2MT
COVID-19+/CH−
COVID-19+/TET2MT
E
NEAT1
ns
LogNormalized Expression

Figure 5. Expression analysis identifies key regulators of inflammation in COVID-19^{+} CH. A, Volcano plot showing significantly differentially expressed genes (adjusted $p < 0.05$, Wilcoxon rank sum test) comparing cells from COVID-19^{+} / *TET2*^{MT} cohort with those from COVID-19^{+} / CH^{-} cohort. Cells from each cell type were tested independently. Bars below the volcano plot show the proportion of genes per cell type that are down- and up-regulated in COVID-19^{+} / *TET2*^{MT} cohort. B, Ingenuity pathway analysis of differentially expressed genes (comparing COVID-19^{+} / *TET2*^{MT} versus COVID-19^{+} / CH^{-}) enriched in each cell type. C, Venn diagram (left) comparing the lists of differentially expressed genes in CD14 monocytes between *TET2*^{MT} versus non-mutant patients, in the COVID-19^{+} cohort and the COVID-19^{-} cohort. Volcano plot (right) showing significantly differentially expressed genes in CD14 monocytes comparing COVID-19^{+} / *TET2*^{MT} versus COVID-19^{+} / CH^{-} showing only genes that are concordant with the COVID-19^{-} comparison (genes that show the same trend: Up- or down-regulated in both comparisons, comparing COVID-19^{-} / *TET2*^{MT} cohort versus healthy cohort). D-E, Violin plots showing expression of *MALAT1* and *NEAT1*. Black dot shows mean expression. ns: $p > 0.05$, *: $p <= 0.05$, **: $p <= 0.01$, ***: $p <= 0.001$, ****: $p <= 0.0001$ (Wilcoxon rank sum test).

Figure 6

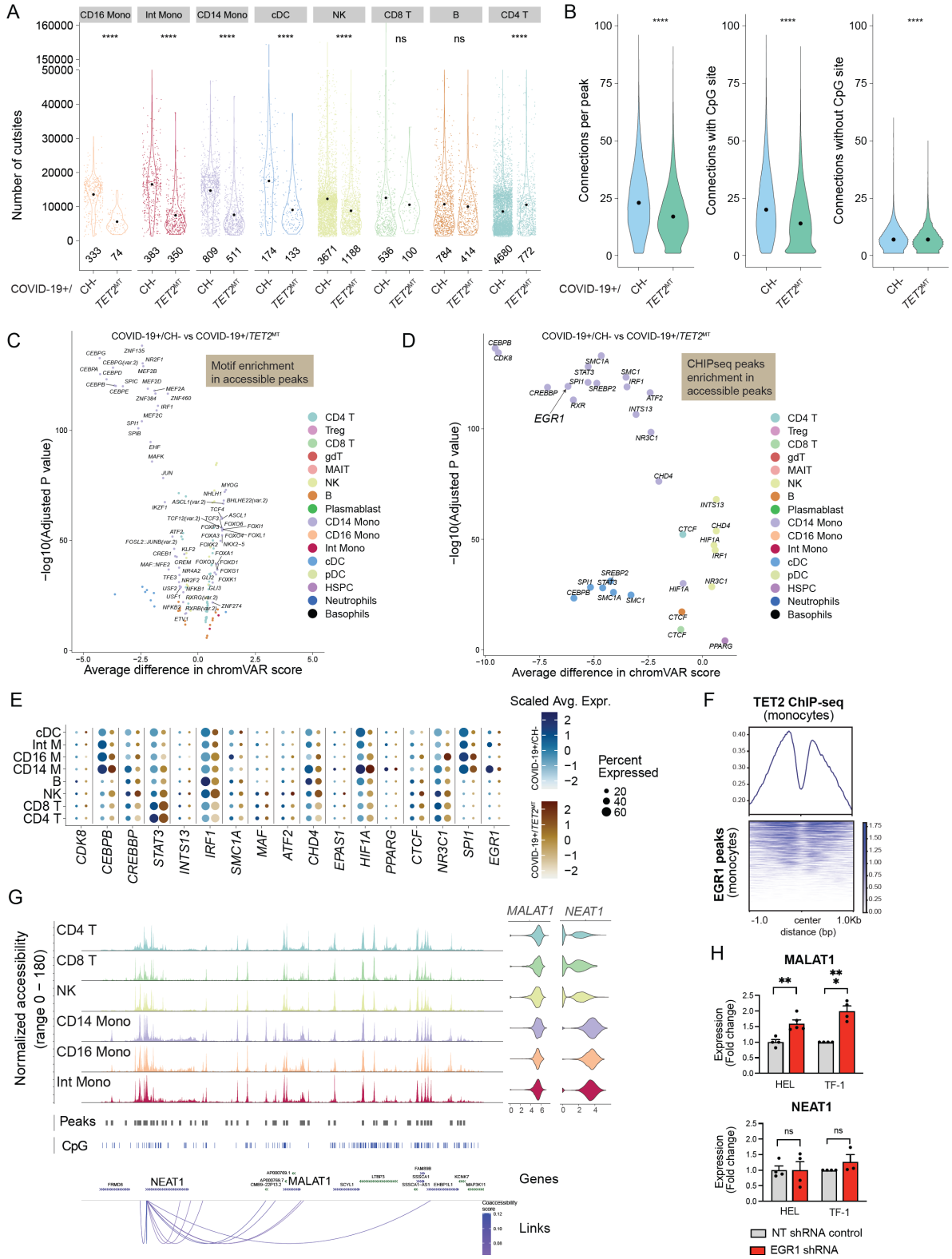
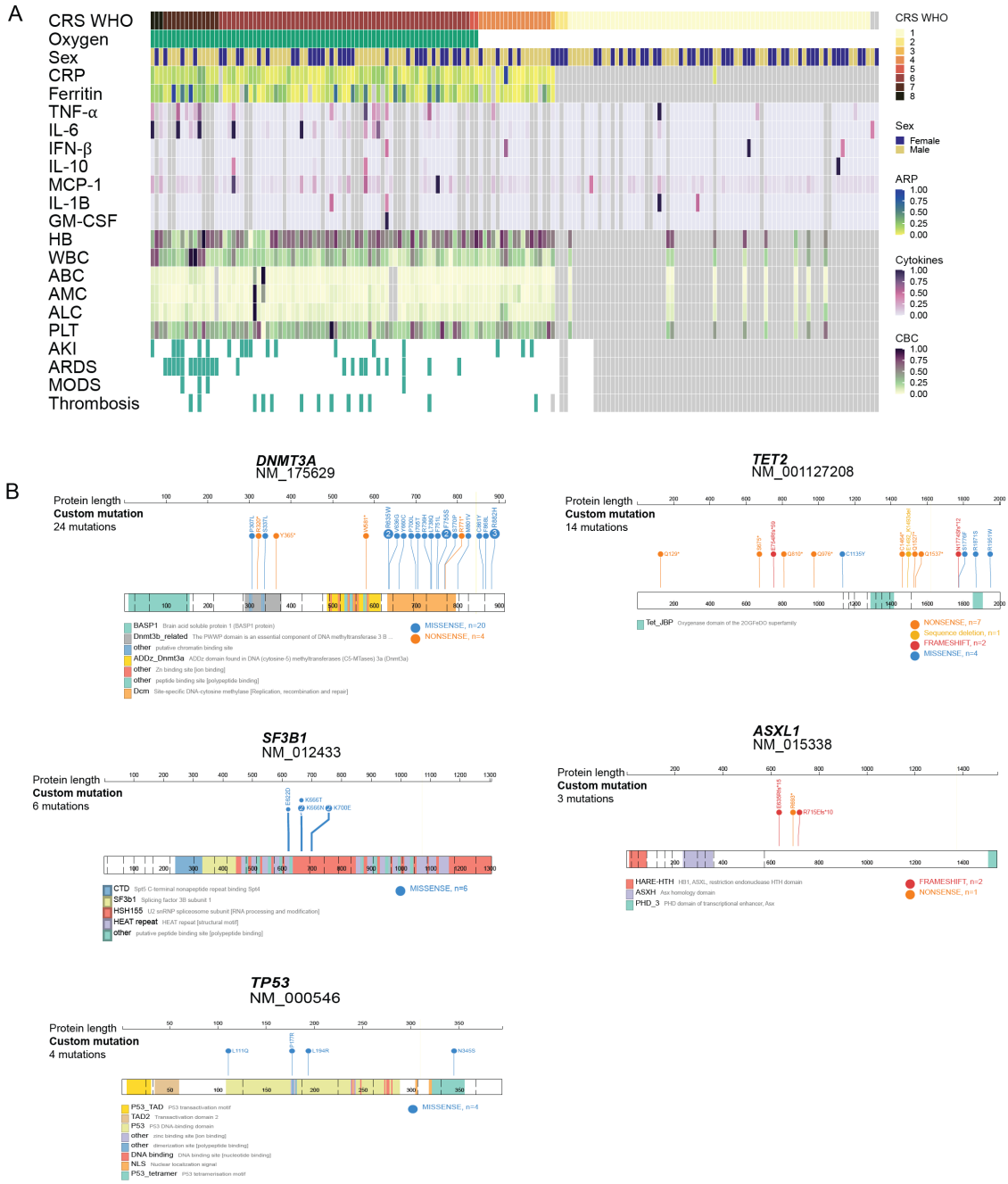

A
CD16 Mono
Int Mono
CD14 Mono
cDC
NK
CD8 T
B
CD4 T
Number of cutsites
COVID-19+/
CH-
TET2MT
B
Connections per peak
Connections with CpG site
Connections without CpG site
C
COVID-19+/CH- vs COVID-19+/TET2MT
Motif enrichment in accessible peaks
−log10(Adjusted P value)
Average difference in chromVAR score
D
CHIPseq peaks enrichment in accessible peaks
EGR1
CD4 T
Treg
CD8 T
gdT
MAIT
NK
B
Plasmablast
CD14 Mono
CD16 Mono
Int Mono
cDC
pDC
HSPC
Neutrophils
Basophils
E
Scaled Avg. Expr.
Percent Expressed
cDC
Int M
CD16 M
CD14 M
B
NK
CD8 T
CD4 T
CDK8
CEBPB
CREBBP
STAT3
INTS13
IRF1
SMC1A
MAF
ATF2
CHD4
EPAS1
HIF1A
PPARG
CTCF
NR3C1
SPI1
EGR1
F
TET2 ChIP-seq (monocytes)
EGR1 peaks (monocytes)
distance (bp)
G
Normalized accessibility (range 0 – 180)
CD4 T
CD8 T
NK
CD14 Mono
CD16 Mono
Int Mono
MALAT1
NEAT1
Peaks
CpG
Genes
Links
H
MALAT1
NEAT1
Expression (Fold change)
HEL
TF-1
NT shRNA control
EGR1 shRNA

Figure 6. Deregulation of the myeloid regulatory network driven by TET2 mutations. A, Violin plots showing the total number of cut sites (sum of TF-IDF normalized cut site counts) in each cell type comparing COVID-19$^+$ / CH$^-$ and COVID-19$^+$ / *TET2*^{MT} cohorts. Black dots show the mean value. ns: p > 0.05, *: p <= 0.05, ****: p <= 0.0001 (Wilcoxon rank sum test). B, Violin plots showing the number of co-accessibility connections per peak predicted using Cicero (co-access score > 0) comparing CD14 monocytes from COVID-19$^+$ / CH$^-$ and COVID-19$^+$ / *TET2*^{MT} cohorts, showing all peaks (left), peaks with CpG site (middle), and peaks without CpG site (right). Peaks with CpG sites were identified by overlapping peaks with CpG sites (EPIC array) using bedtools intersect. Black dots show the mean value. ****: p <= 0.0001 (Wilcoxon rank sum test). C, Volcano plot showing transcription factor motif enrichment scores (calculated using ChromVAR) significantly different between COVID-19$^+$ / CH$^-$ and COVID-19$^+$ / *TET2*^{MT} cohorts (adjusted p value < 0.05, Wilcoxon rank sum test). Cells from each cell type were tested independently. D, Volcano plot showing binding sites enrichment scores (ChromVAR) of DNA binding proteins significantly different between COVID-19$^+$ / CH$^-$ and COVID-19$^+$ / *TET2*^{MT} cohorts (adjusted p value < 0.05, Wilcoxon rank sum test). Cells from each cell type were tested independently. Curated ChIP-seq peaks of DNA binding proteins assayed in myeloid cells, obtained from ReMap 2022, were used as the DNA binding sites to calculate enrichment. E, Dot plot showing log-normalized gene expression of the transcription factor genes across different cell types, in COVID-19$^+$ / CH$^-$ and COVID-19$^+$ / *TET2*^{MT} patients. F, Average signal profile (top) and heatmap (bottom) of ChIP-seq signal scores of TET2 in isolated monocytes around EGR1 binding sites (EGR1-gained peaks during monocyte differentiation from CD34+ hematopoietic stem cells). G, Coverage plot showing the chromatin accessibility, gene expression (violin plots), co-accessibility connections (links) and CpG sites (EPIC array) around the *MALAT1* and *NEAT1* gene locus. H, qPCR for detection of MALAT1 and NEAT1 using RNA extracted from HEL and TF-1 cells with shRNA construct against EGR1 (and non-targeting, NT, as control). Data are mean with SE n=4 (p>0.05, t-test).

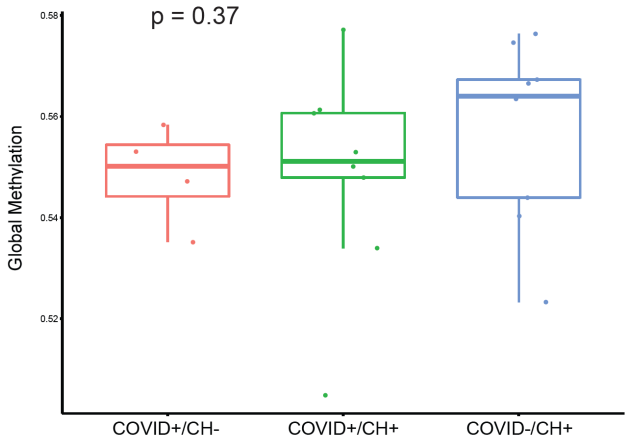
Supplementary Figure 1



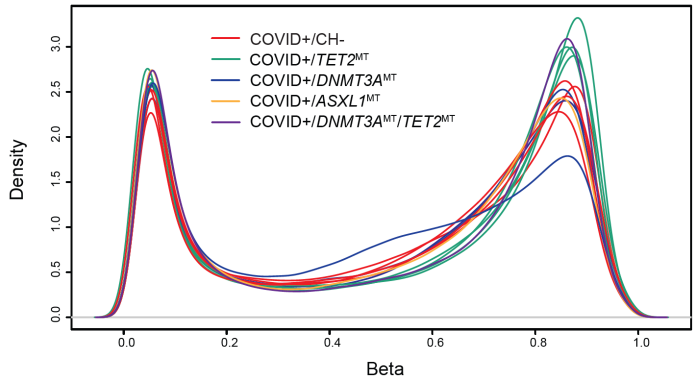
Supplementary Figure 1. A, Heatmap showing the patient information (each column represents a single patient) for all COVID-19 patients included in the study. APR, Cytokines and CBC are min-max scaled within their category. CRS, cytokine release syndrome; APR, acute phase reactants; CBC, complete blood count; CRP, C-reactive protein; HGB, hemoglobin; WBC, white blood cell count; ANC, absolute neutrophil count; AMC, absolute monocyte count, ALC, absolute lymphocyte count; PLT, platelet count; AKI, acute kidney injury; ARDS, acute respiratory distress syndrome; MODS, multiple organ dysfunction syndrome. B, Lollipop plots showing type and location along the protein sequence of the CH mutations for *DNMT3A*, *TET2*, *SF3B1*, *ASXL1*, and *TP53*. The number of recurrently detected alterations is indicated by the text within each disc, as well as by disc size. Colors indicate the type of mutation: blue, missense; orange, nonsense; red, frameshift; yellow: sequence deletion.

Supplementary Figure 2

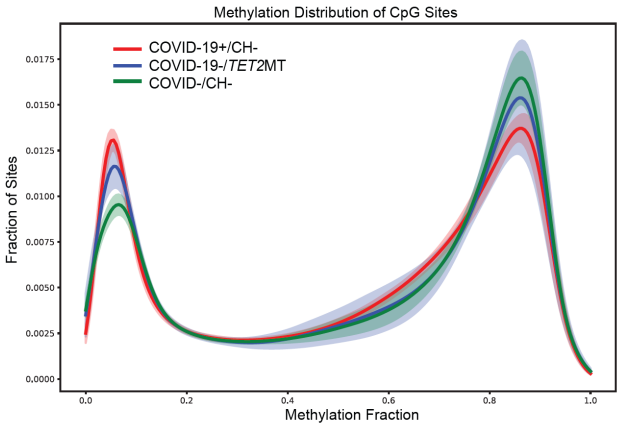
A



B

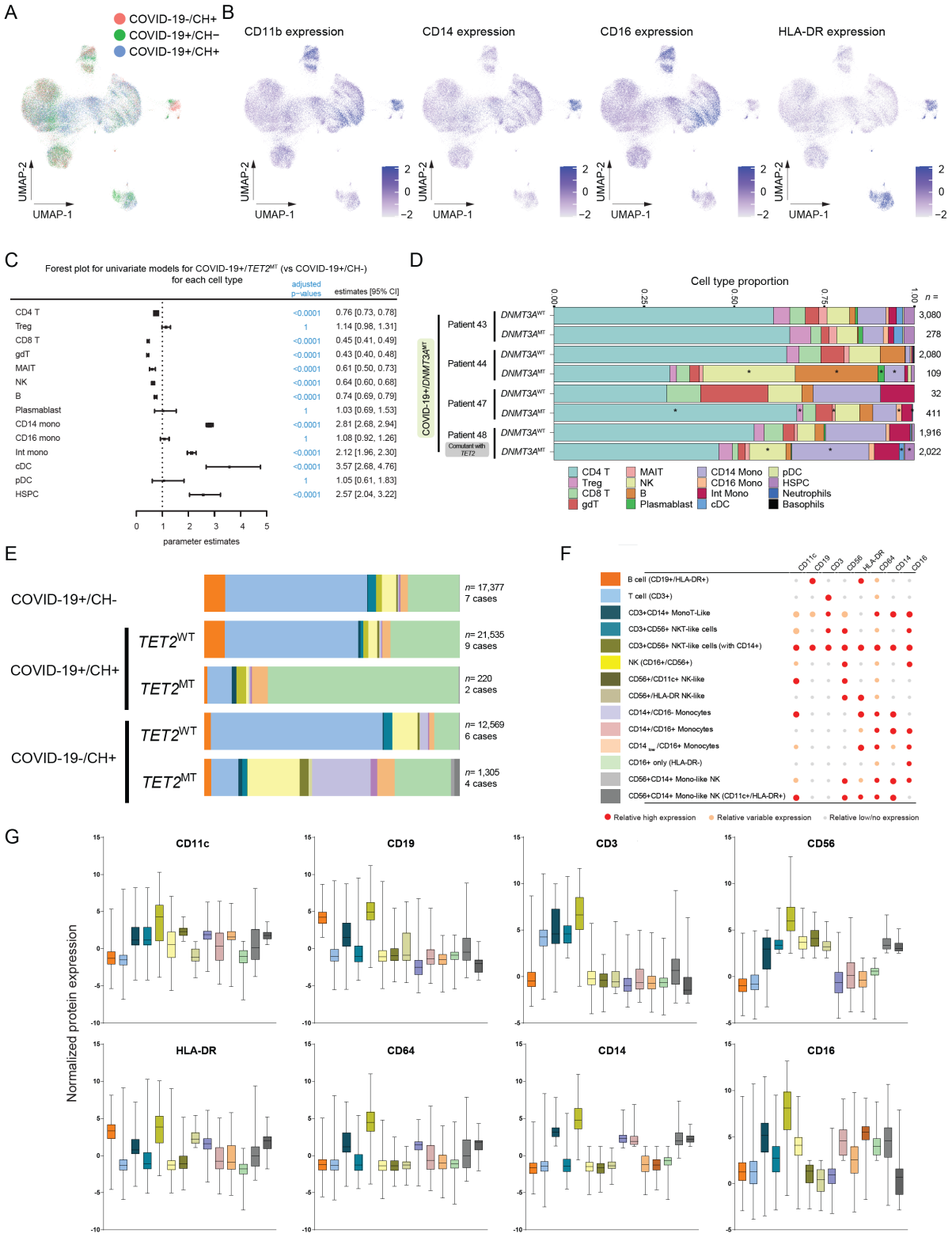


C



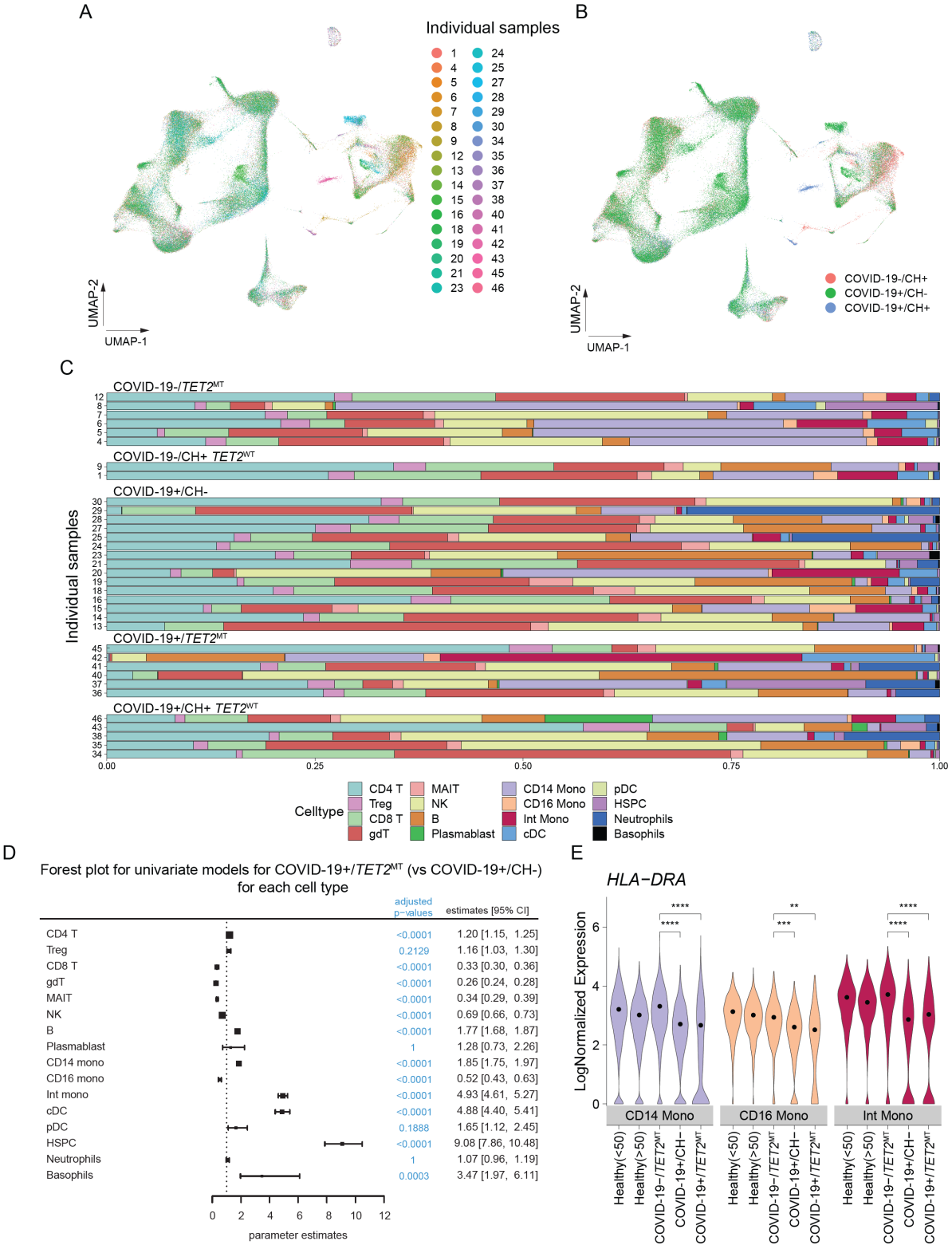
Supplementary Figure 2. A, Global average methylation levels (β-values) at 866,238 CpG sites in COVID-19^{+} / CH^{-} (n = 4), COVID-19^{+} / CH^{+} (n = 9), and COVID-19^{-} / CH^{+} (n = 9). P-value obtained using a Kruskal-Wallis test. B, Density plot demonstrating the differences in individual sample β-value distribution between the COVID-19^{+} / CH^{-} and COVID-19^{+} / CH^{+} patient cohorts. The COVID-19^{+} / CH^{+} cohort is grouped by the mutation type. C, Density plot demonstrating the differences in β-value distribution between the COVID-19^{+} / CH^{-}, COVID^{-} / CH^{-}, and COVID-19^{-} / CH^{+} / *TET2*^{MT} patient cohorts.

Supplementary Figure 3



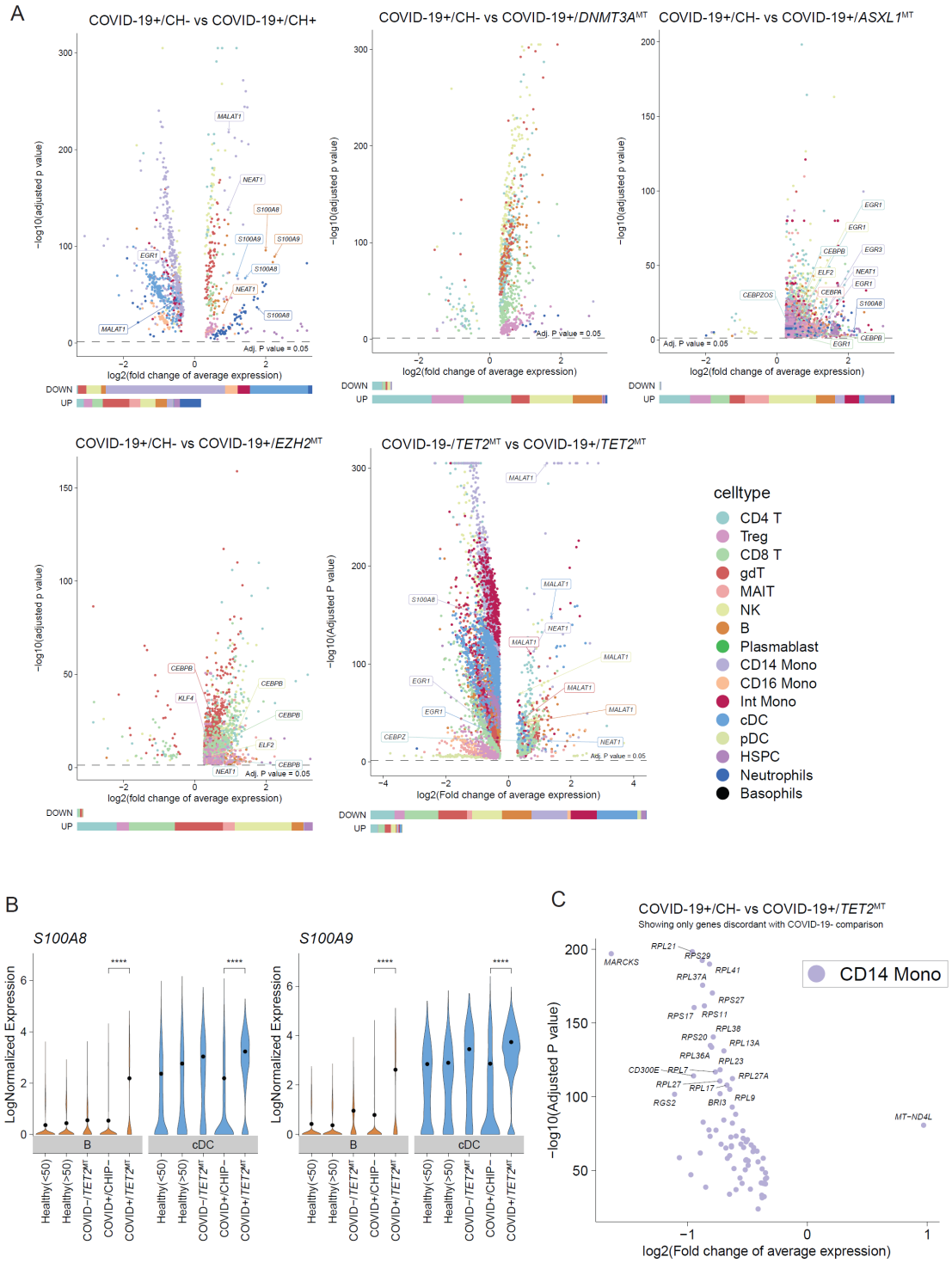
Supplementary Figure 3. A, UMAP plot based on single-cell surface protein measurements by the Mission Bio Tapestri platform, showing cells colored by patient cohorts, i.e., COVID-19^{-} / CH^{+} (patients carrying CH mutations), COVID-19^{+} / CH^{-} (patients diagnosed with COVID-19), and COVID-19^{+} / CH^{+} (patients carrying CH mutations and diagnosed with COVID-19). B, UMAP plots showing expression levels of cell surface markers expressed on typical blood monocytes. C, Forest plot for univariate models for COVID-19^{+} / *TET2*^{MT} cohort (versus COVID-19^{+} / CH^{-} cohort) for each cell type. D, Proportions of cell types in *DNMT3A*^{WT} versus *DNMT3A*^{MT} (*DNMT3A* mutated) cells for each patient with *DNMT3A* mutation. Treg, regulatory T-cells; gdT, gamma delta T-cells; MAIT, Mucosal-associated invariant T-cells; NK, natural killer cells; Mono, monocytes; Int, intermediate; cDC, classical dendritic cells; pDC, plasmacytoid dendritic cells; HSPC, hematopoietic stem and progenitor cells. E, Representative fraction of cell types in *TET2*^{WT} (non-*TET2* mutated) versus *TET2*^{MT} (*TET2* mutated) cells for each patient group. The number of cases involved in each group, as well as the total number of cells analyzed for each, is listed on the right. Cell types were called based on normalized relative surface protein expression. F, Cell types with correlative relative levels of surface marker expression. G, Box plots showing distribution of expression levels for surface markers defining cell types.

Supplementary Figure 4



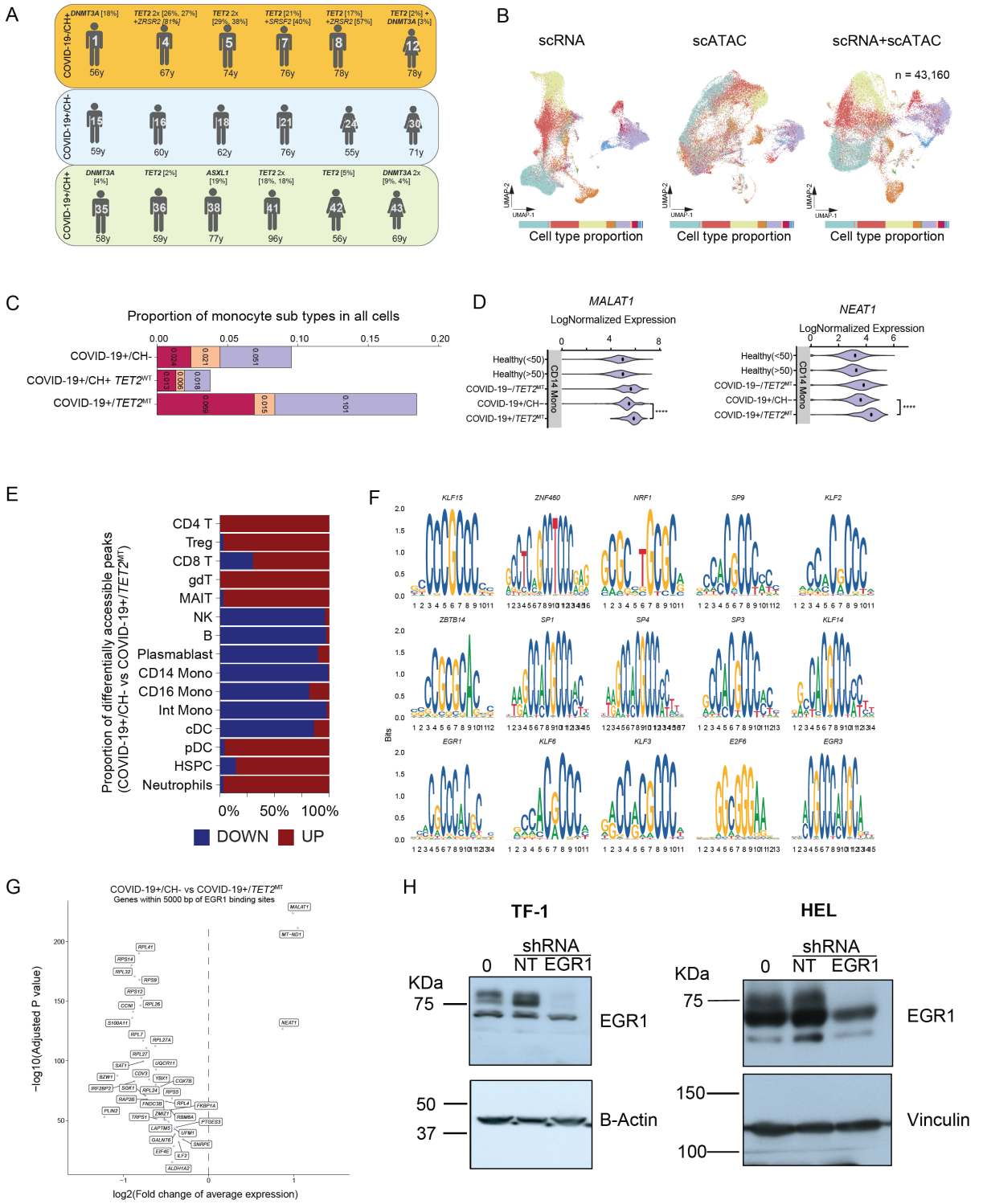
Supplementary Figure 4. A,B, UMAP projection showing cells colored by patient sample and by patient cohorts - COVID-19^{-} / CH^{+} (patients carrying CH mutations), COVID-19^{+} / CH^{-} (patients diagnosed with COVID-19), COVID-19^{+} / CH^{+} (patients carrying CH mutations and diagnosed with COVID-19). C, Proportion of each cell type shown for each individual patient sample. D, Forest plot for univariate models for COVID-19^{+} / *TET2*^{MT} cohort (versus COVID-19^{+} / CH^{-} cohort) for each cell type. E, Violin plots showing expression of *HLA-DRA*. Black dot shows mean expression. ns: $p > 0.05$, *: $p <= 0.05$, **: $p <= 0.01$, ***: $p <= 0.001$, ****: $p <= 0.0001$ (Wilcoxon rank sum test).

Supplementary Figure 5



Supplementary Figure 5. A, Volcano plots showing significantly differentially expressed genes (adjusted $p < 0.05$, Wilcoxon rank sum test) comparing cells from COVID-19^{+} / CH^{+} cohort grouped by each CH mutation they carry, with that from COVID-19^{+} / CH^{-} cohort. Cells from each cell type were tested independently. Bars below the volcano plot show the proportion of genes per cell type that are down- and up-regulated in COVID-19^{+} / CH^{+} cohort. B, Violin plots showing expression of *S100A8* and *S100A9*. Black dot shows mean expression. ns: $p > 0.05$, *: $p <= 0.05$, **: $p <= 0.01$, ***: $p <= 0.001$, ****: $p <= 0.0001$ (Wilcoxon rank sum test). C, Volcano plot showing significantly differentially expressed genes in CD14 monocytes comparing COVID-19^{+} / *TET2*^{MT} versus COVID-19^{+} / CH^{-}, that show the opposing trend when comparing COVID-19^{-} / *TET2*^{MT} cohort versus healthy cohort: Up-regulated genes in one comparison are down-regulated in the other, and vice-versa.

Supplementary Figure 6



Supplementary Figure 6. A, Overview of the patient cohorts - COVID-19^{-} / CH^{+} (patients carrying CH mutations), COVID-19^{+} / CH^{-} (patients diagnosed with COVID-19), COVID-19^{+} / CH^{+} (patients carrying CH mutations and diagnosed with COVID-19) whose samples were assayed using single-cell Multiome analysis. Patient age is shown at the bottom. B, UMAP projections of GEX (left), ATAC (middle) and GEX+ATAC combined using Seurat's weighted nearest neighbor algorithm (right), showing 43,160 cells from all 18 patient samples colored by cell types identified using Azimuth algorithm mapping the cells to our scRNA-seq data considered as reference. Treg, regulatory T-cells; gdT, gamma delta T-cells; MAIT, Mucosal-associated invariant T-cells; NK, natural killer cells; Mono, monocytes; Int, intermediate; cDC, classical dendritic cells; pDC, plasmacytoid dendritic cells; HSPC, hematopoietic stem and progenitor cells. C, Proportion of average number of monocyte subtypes (averaged by sample) showing monocyte expansion in COVID-19^{+} / *TET2*^{MT} samples, similar to the observation in scRNA-seq analysis. D, Violin plots showing expression of *MALAT1* and *NEAT1*. Black dot shows mean expression. ns: $p > 0.05$, *: $p <= 0.05$, **: $p <= 0.01$, ***: $p <= 0.001$, ****: $p <= 0.0001$ (Wilcoxon rank sum test). E, Bar plot showing the proportions of differentially accessible peaks ($p < 0.05$, Wilcoxon rank sum test) that are up- or down-regulated in each cell type, in COVID-19^{+} / *TET2*^{MT} patients compared to COVID-19^{+} / CH^{-} patients. F, Motif plots of the top 15 motifs that lost accessibility in COVID-19^{+} / CH^{-} cohort CD14 monocytes compared to COVID-19^{+} / *TET2*^{MT} cohort CD14 monocytes. G, Volcano plot showing significantly differentially expressed genes (adjusted $p < 0.05$, Wilcoxon rank sum test), that are within 5000 base pairs from EGR1 binding sites (EGR1-gained peaks during monocyte differentiation from CD34+ hematopoietic stem cells), comparing CD14 Monocytes from COVID-19^{+} / *TET2*^{MT} cohort with those from COVID-19^{+} / CH^{-} cohort. H, Immunoblots from HEL and TF-1 cell lysates with parental, non-targeting and EGR1 knockdown probed for EGR1. Vinculin and B-Actin were used as loading controls.