medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Mild SARS-CoV-2 infection modifies DNA methylation of peripheral blood mononuclear cells from COVID-19 convalescents

Johanna Huoman, PhD[1]*, Shumaila Sayyab, PhD[1]*, Eirini Apostolou, PhD[2], Lovisa Karlsson, MSc[1], Lucas Porcile[1], Muhammad Rizwan, MSc[2], Sumit Sharma, PhD[3], Jyotirmoy Das, PhD[1], Anders Rosén[2], PhD and Maria Lerm, PhD[1]**

*Shared first authorship

**Corresponding author

[1]Division of Inflammation and Infection, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

[2]Division of Cell Biology, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

[3]Division of Molecular Medicine and Virology, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

Corresponding author:

Maria Lerm, PhD, Professor

Department of Biomedical and Clinical Sciences

Division of Inflammation and Infection, Lab1, floor 12

Faculty of Medicine and Health Sciences

SE-581 85 Linköping, Sweden

Cell phone: +46-73-2707786

Visit us at: https://liu.se/en/employee/marle69

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

ABSTRACT

Background: Coronaviruses such as SARS-CoV-2 may circumvent host defence mechanisms by hijacking host proteins, possibly by altering DNA methylation patterns in host cells. While most epigenetic studies have been performed in severely ill COVID-19 patients, studies on individuals who have recovered from mild-to-moderate disease remain scarce. The aim of this study was to assess epigenome-wide DNA methylation patterns in COVID-19 convalescents compared to uninfected controls from before and after the pandemic outbreak began.

Methods: DNA was extracted from peripheral blood mononuclear cells originating from uninfected controls before (Pre20, n=5) and after (Con, n=18) 2020, COVID-19 convalescents (CC19, n=14) and symptom-free individuals with a SARS-CoV-2-specific T cell response (SFT, n=6), as well as from Pre20 (n=4) samples stimulated *in vitro* with SARS-CoV-2. Subsequently, epigenome-wide DNA methylation analyses were performed using the Illumina MethylationEPIC 850K array, and statistical and bioinformatic analyses comprised differential DNA methylation, pathway over-representation and module identification network analyses.

Results: DNA methylation patterns of COVID-19 convalescents were altered as compared to uninfected controls, with similar results observed in *in vitro* stimulations of PBMC with SARS-CoV-2. Differentially methylated genes from the *in vivo* comparison constituted the foundation for the identification of a possibly SARS-CoV-2-induced module, containing 66 genes of which six could also be identified in corresponding analyses of the *in vitro* data (TP53, INS, HSPA4, SP1, ESR1 and FAS). Pathway over-representation analyses revealed involvement of Wnt, cadherin and apoptosis signalling pathways amongst others. Furthermore, numerous interactions were found between the obtained differentially methylated genes from both settings and the network analyses when overlaying the data unto the SARS-CoV-2 interactome.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Conclusions: Epigenome-wide DNA methylation patterns of individuals that have recovered from mild-to-moderate COVID-19 are different from those of non-infected controls. The observed alterations during both *in vivo* and *in vitro* exposure to SARS-CoV-2 showed involvement in interactions and pathways that are highly relevant to COVID-19. The present study provides indications that DNA methylation is one of several epigenetic mechanisms that is altered upon SARS-CoV-2 infection. Further studies on the mechanistic underpinnings should determine whether the observed effects are reflecting host-protective antiviral defence or targeted viral hijacking to evade host defence.

Keywords (max. 10 words): COVID-19, DNA methylation, interactome, *in vitro* stimulation, mild-to-moderate, module identification, network analysis, PBMC, SARS-CoV-2

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

BACKGROUND

Severe acute respiratory syndrome (SARS) caused by coronaviruses is not new to the world, but at the emergence of the SARS coronavirus 2 (SARS-CoV-2) in December 2019, the global community was largely unprepared. Despite the outbreak of SARS-CoV-1 in 2003, very limited understanding of coronavirus biology and no vaccine portfolio was available at the time of the SARS-CoV-2 outbreak in 2019. To understand SARS-CoV-2 biology, the underlying mechanisms of how the virus interacts with its host needs to be scrutinized and this knowledge is crucial for the development of effective treatment and prevention of coronavirus disease 19 (COVID-19), the disease caused by the virus.

DNA methylation (DNAm) is the most stable epigenetic modification, as it ensures heritability in the cell division process, but is at the same time highly dynamic in response to environmental stimuli (1). The malleability and flexibility of the DNA methylome decreases with increasing age (2), and environmental factors such as smoking and nutrition may alter DNAm patterning in various cell types, including different immune cells (3). Epigenetic changes in *i.a.* immune cell populations have been reported both in immune-related diseases and allergies (4) as well as infectious diseases (5-7). In line with this, we have observed that immune cells of asymptomatic, tuberculosis-exposed individuals carry a lasting DNAm biosignature (8-10) that is linked to protection against mycobacterial infection (8). However, epigenetic alterations can also be induced by pathogens for their own benefit (11-14).

A majority (40-80%) of individuals infected with SARS-CoV-2 show no or mild symptoms of COVID-19 and proceed into convalescence thereafter, while a smaller, but non-negligible, proportion of individuals show severe or life-threatening manifestations (15, 16). However, thus far, no studies have addressed whether and how the epigenome is altered in subjects with a recent mild-to-moderate SARS-CoV-2 infection. In this study, we set out to examine epigenome-wide DNAm patterns in convalescent COVID-19 (CC19) subjects, after a mild-to-moderate disease course. Understanding how convalescent COVID-19 individuals mount an epigenetically encoded defence strategy against new viruses such as SARS-CoV-2, for

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

which no pre-existent immunity was present, may reveal how a functional defence towards SARS-CoV-2 is mounted, and guide development of novel diagnostic and preventive measures. Indeed, we could show that a number of genes that interact with SARS-CoV-2 interacting proteins were epigenetically modulated in these individuals, suggesting that appropriate host defence may be initiated on a cellular level by altered DNAm patterning in virus-exploited host proteins.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

RESULTS

COVID-19 convalescents display altered DNAm patterns compared to non-infected controls

As we were interested in studying DNAm as a defence mechanism in COVID-19, we compared epigenome-wide DNAm patterning in peripheral blood mononuclear cells (PBMC) from non-infected controls (Con, n=19), COVID-19 convalescents who had recovered from mild or

Figure 1.

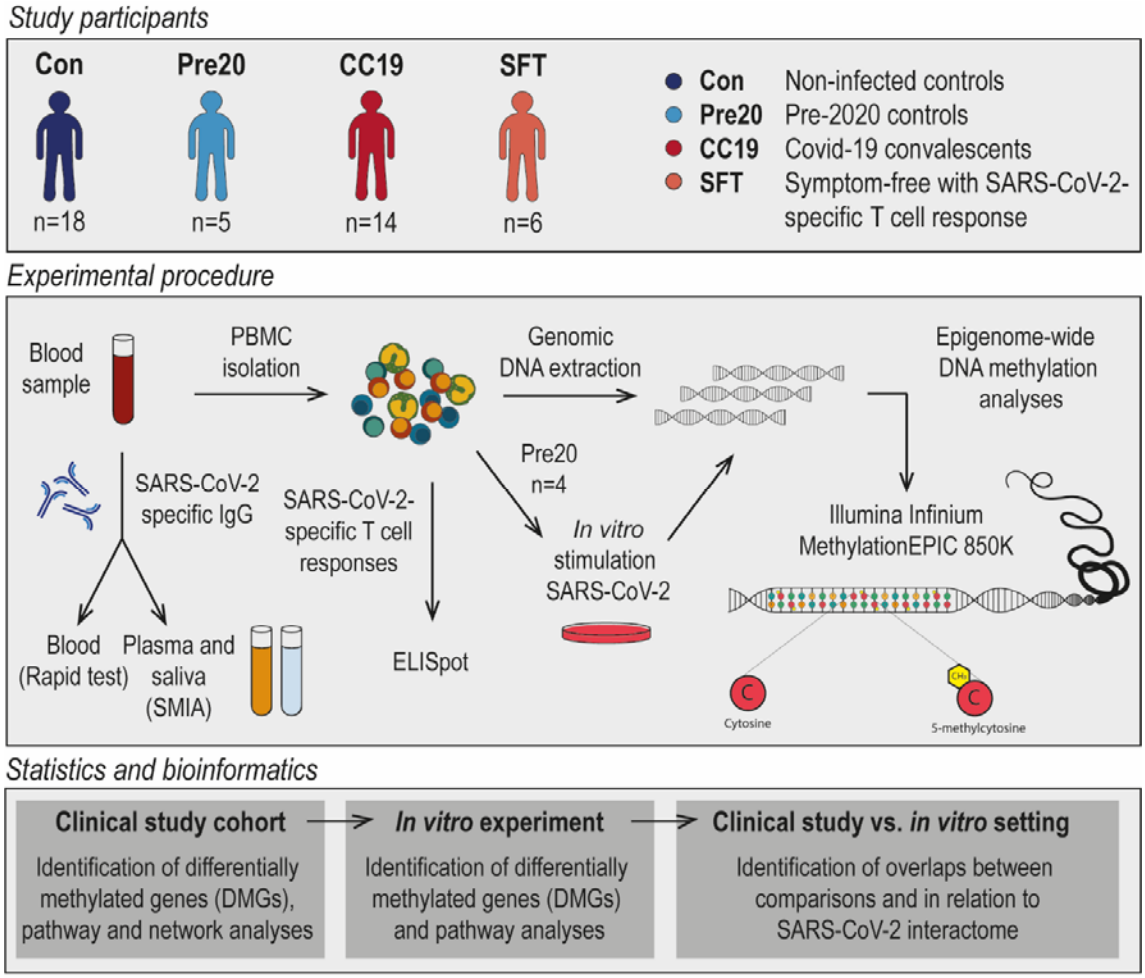


Figure 1. Outline of included participants, experimental procedures as well as statistical and bioinformatic approaches utilised in the present study. CC19 – convalescent COVID-19, Con – non-infected control, DMG – differentially methylated gene, Pre20 – Pre-2020 non-infected control,

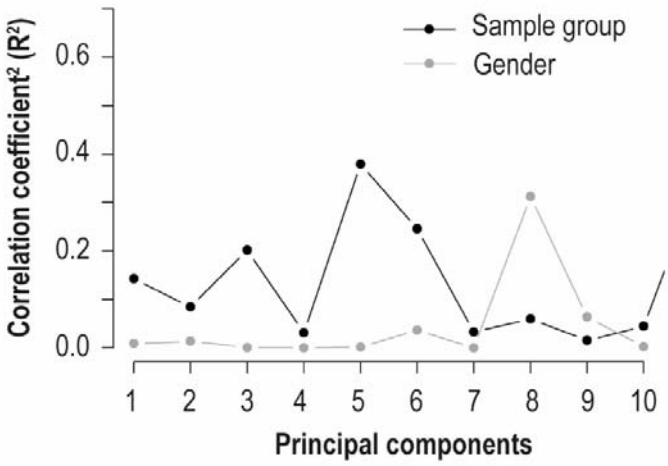
medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

SFT – symptom-free individuals with SARS-CoV-2-specific T cell response, SMIA – suspension multiplex immunoassay.

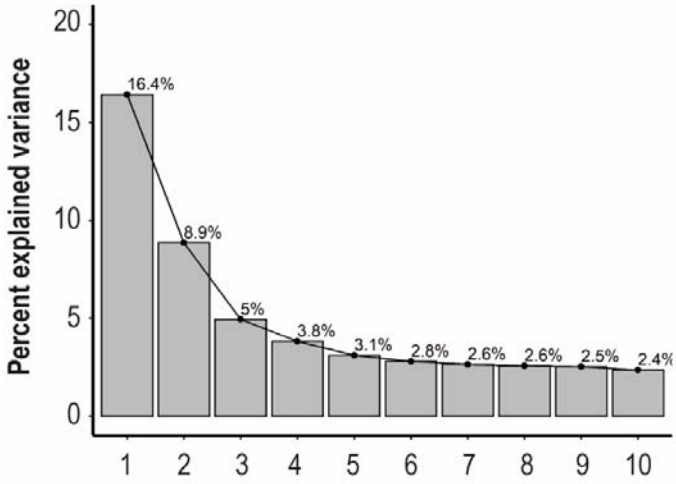
moderate symptoms (CC19, n=14), donor blood collected before the pandemic (Pre20, n=5) and from asymptomatic individuals presenting with SARS-CoV-2-specific T cell responses (SFT, n=6, Figure 1). Comparisons of demographic variables revealed no significant differences between any of the groups (Table S1). To examine any inherent differences in the DNA methylome between the different sample groups, principal component analyses (PCA) were performed. Three principal components (PC) were identified as both contributing to the variation within the DNAm data and correlating with the sample groups (Figure 2A-B). A three-dimensional illustration of these three most contributing components revealed that the CC19 subjects are distinct from the Con, Pre20 and SFT subjects, whose centroids cluster more

Figure 2.

A.



B.



medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

C.

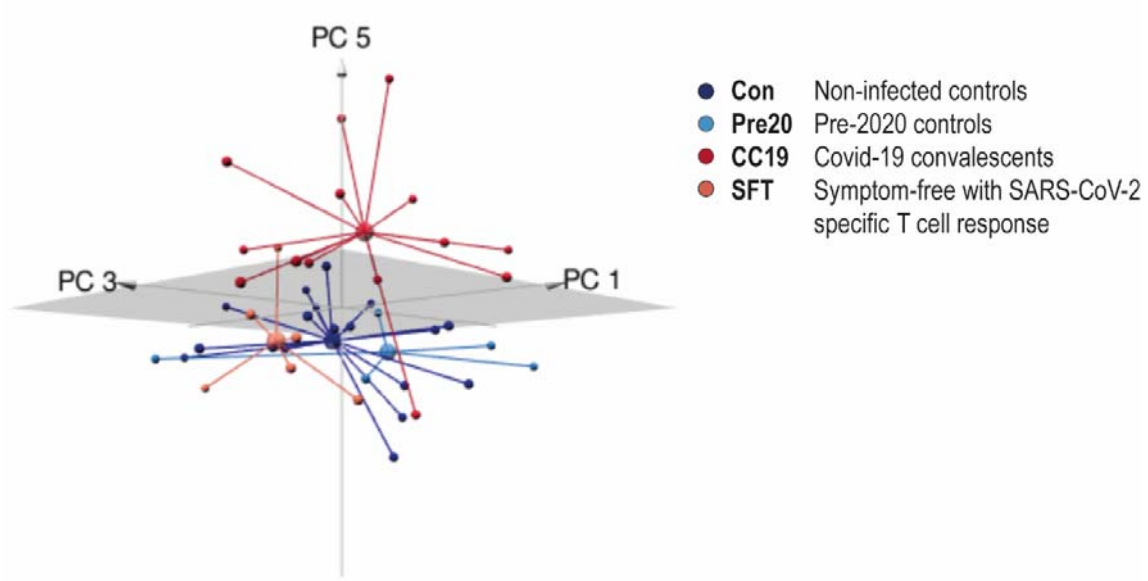
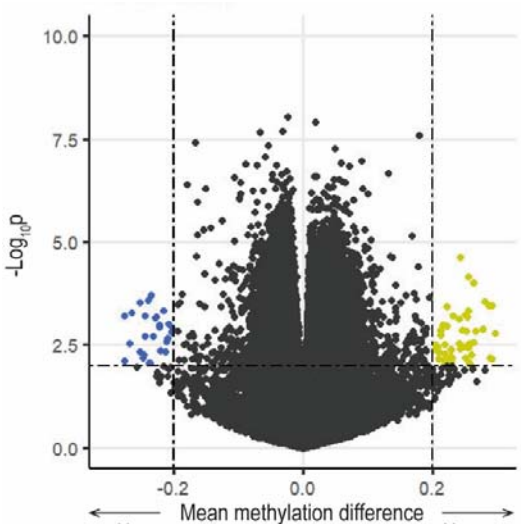


Figure 2. PCA analysis of PBMC DNA methylomes. Upon filtering and normalisation, the DNAm data were subjected to PCA analysis. A. shows a correlation plot of the PCA-derived eigenvalues and the DNAm group data projected as Con/Pre20/CC19/SFT and male/female. In B. a scree-plot shows degree to which the identified components contribute to the variation observed within the DNAm data. C. shows a 3D-PCA plot of principal component (PC)1, PC3 and PC5, where the group means are illustrated as centroids.

Figure 3.

A.



medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

B.

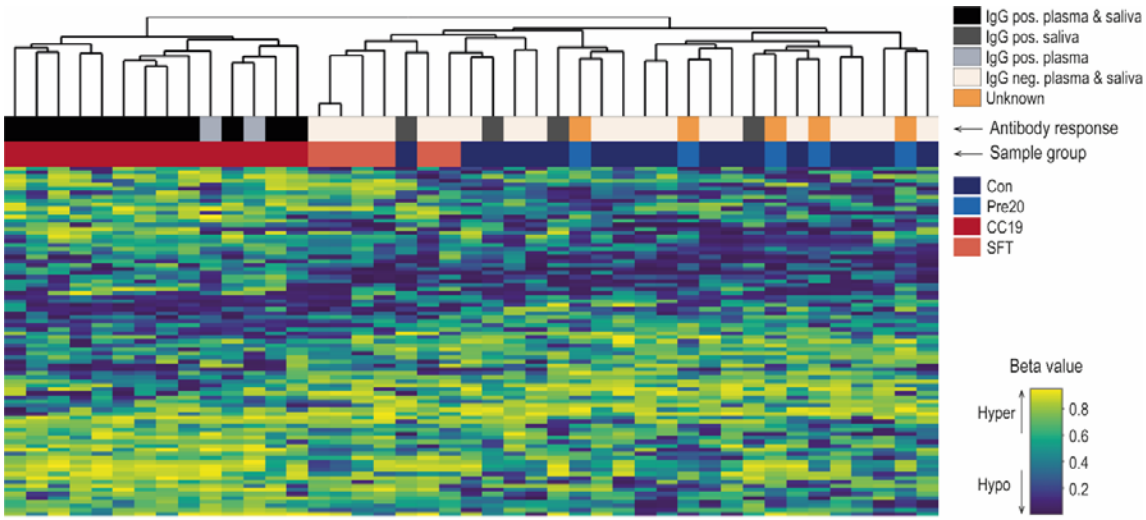


Figure 3. Identification of differentially methylated CpGs in CC19 subjects *vs.* uninfected controls. DMCs were identified comparing CC19s to Cons and Pre20s by computing a linear model on the DNAm data. A. illustrates a volcano plot of the CC19 *vs.* Con + Pre20 DNAm data. The dash-dotted horizontal line represents a nominal p-value cut-off of 0.01, and the vertical lines represent a cut-off in mean methylation difference (MMD) in CC19 *vs*. Con + Pre20 of > ± 0.2. B. shows a heatmap representing an unsupervised hierarchical clustering analysis of individual β values of the 87 identified DMCs in A. The individuals' antibody status is indicated as a grey-scale (unknown in anonymous Pre20 blood donors, orange).

closely together (Figure 2C, Figure S1). The observed methylome differences prompted us to identify differentially methylated CpGs (DMCs), which we defined as CpG sites with a nominal p-value of <0.01 along with a mean methylation difference (MMD) of >0.2. We found 87 DMCs, 30 of which were hypomethylated and 57 DMCs were hypermethylated when comparing the DNA methylomes of CC19s to the merged groups of Cons and Pre20s (Figure 3A, Table S2a). This identified DMC signature could furthermore accurately distinguish the CC19s from Cons, Pre20s and SFTs (Figure 3B), suggesting that a past SARS-CoV-2 infection may have resulted in modulation of the epigenome that persists at least a couple of

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

months after the virus is eliminated from the body. Interestingly, a majority of CC19s showed positive SARS-CoV-2-specific IgG responses both in the circulation and in saliva (Figure 3B). The individuals who were positive for SARS-CoV-2-specific T cells or antibodies in saliva, while being negative for antibodies in plasma, aligned with the controls in the PCA and unsupervised clustering analyses (Figure 3A-B).

Differentially methylated genes of COVID-19 convalescents identify a putatively SARS-CoV-2-induced module

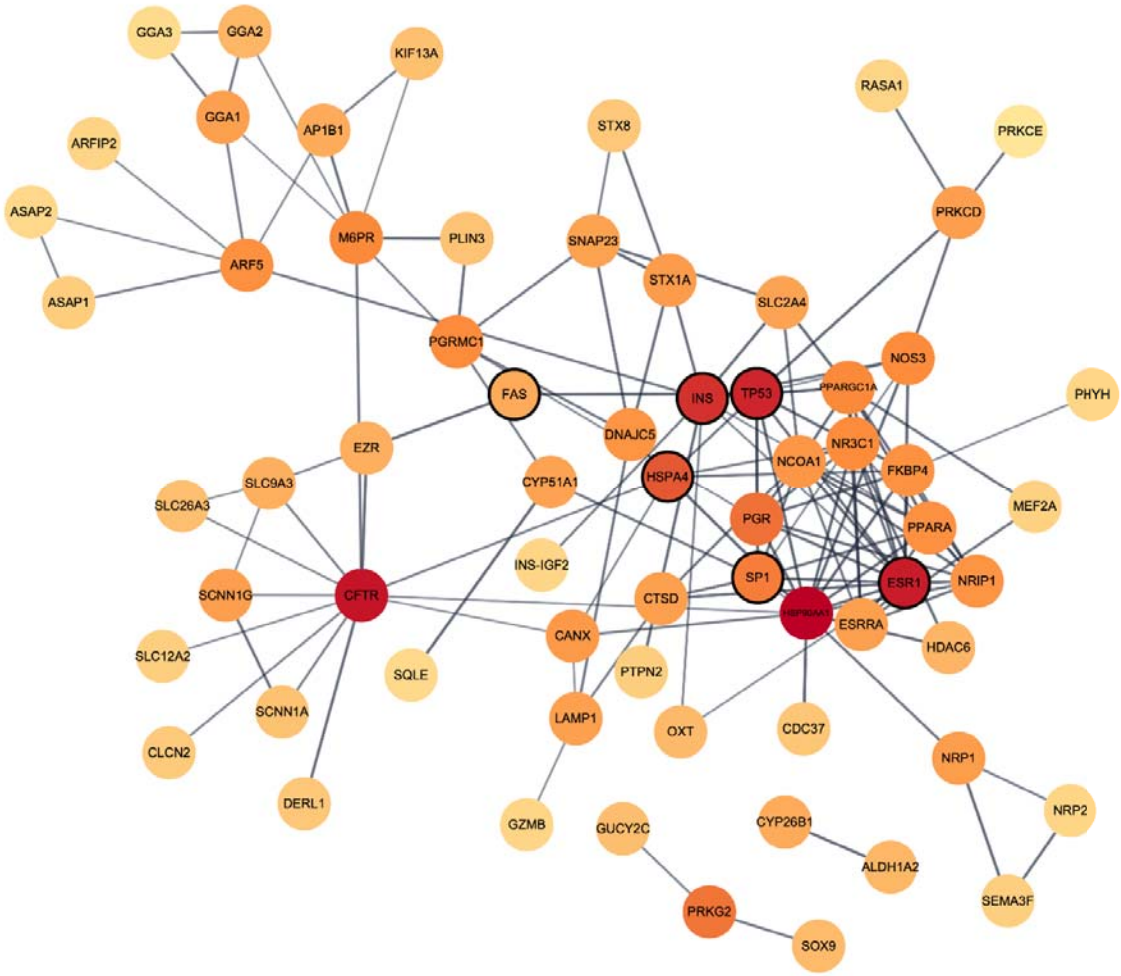
To further explore the biological impact of SARS-CoV-2 exposure in the CC19 subjects, the identified DMCs were annotated to their respective differentially methylated genes (DMG), resulting in 54 unique genes, of which 18 genes were hypomethylated, 35 hypermethylated, and one gene featured a mixed methylation pattern (Table S2b). Subsequent pathway over-representation analyses using the identified DMGs from the CC19 to the combined Con and Pre20 subject comparison revealed involvement in two significantly over-represented pathways (Wnt and integrin signalling pathways, Table S3).

As a means to elaborate on the wider interaction context in which the DMGs act with other proteins, the DMGs (n=54) were used as seed genes in the identification of SARS-CoV-2-induced modules in network analyses. The resulting module consisted of 66 genes from the protein-protein interaction (PPI) network, with 139 interactions, which is significantly more interactions than the expected (34 interactions) for a network of that size (Figure 4A, Table S4). Six of these genes were present in at least two module identification methods (*INS*, *HSPA4*, *SP1*, *ESR1*, *TP53*, *FAS*), and they were all located in the centre of the module. The four genes with the highest combined centrality scores were *HSP90AA1*, *TP53*, *INS* and *CFTR*. Pathway over-representation analyses of the 66 module genes revealed involvement in pathways such as apoptosis signalling, muscarinic acetylcholine receptor 1 and 3 signalling and gonadotropin-releasing hormone pathway (Figure 4B).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Figure 4.

A.



B.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Figure 4. Network illustration and analysis of significantly differentiated methylated genes from the *in vivo* setting. A. shows the network module constructed by means of the graph clustering algorithm MCODE with the 54 DMGs as input. Nodes (n=66) represent genes and connecting lines represent high-confidence protein-protein interactions within the network (STRING combined score > 0.7). Combined ranked scores of centrality quantification of degree, betweenness and closeness is visualised as a colour (light orange to dark red) continuum, with dark red nodes constituting the most central parts of the network. Nodes that were also found both when utilising two other module identifying methods (DIAMOnD and WGCNA) and when performing the same analyses on the *in vitro* data set using MCODE are enclosed with a black line. B. displays results from pathway over-representation analyses of the 66 identified network genes in the protein-protein interaction network using PANTHER. Pathways with an FDR-corrected p-value < 0.05 were considered significant.

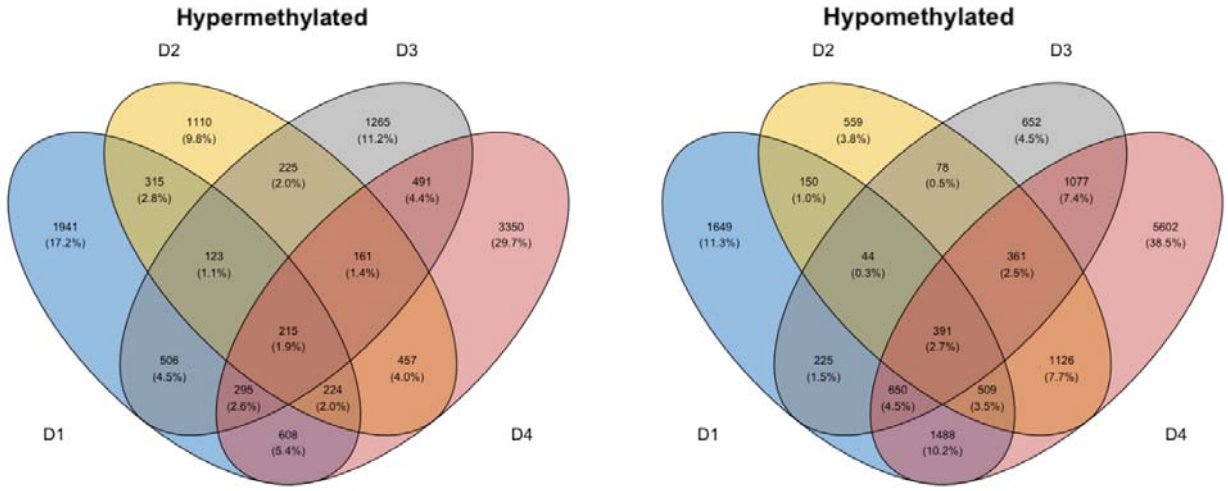
PBMCs stimulated with SARS-CoV-2 *in vitro* reveal differential methylation in multiple pathways important for the viral life cycle

In the present study, we only had access to self-reported time-after-onset of COVID-19 symptoms (Table S5), thus making the immediate effects of SARS-CoV-2 exposure on the epigenome impossible to analyse. Moreover, as the virus-induced DNAm patterns in the CC19's may fade over time, we set out to examine the possible role of SARS-CoV-2-induced DNAm patterns in host defence in an *in vitro* setting. To this end, we exposed PBMCs collected from blood donors in 2019 to SARS-CoV-2 at a low multiplicity of infection for 48h to mimic immediate *in vivo* exposure to the virus (Figure S2). Exploring the intraindividual DNAm differences between stimulated and unstimulated cells, a set of DMCs (n=3693) were identified to be shared between all four individuals, of which 1523 were hypermethylated (Table S6a) and 2170 were hypomethylated (Table S6b). These DMCs mapped to in total 606 DMGs (542 unique genes, Table S6c), consisting of 215 hypermethylated and 391 hypomethylated genes (Figure 4A), which were significantly over-represented in a number of pathways including several glutamate receptor pathways, muscarinic acetylcholine receptor 1 and 3 signalling pathway, as well as the Wnt and cadherin signalling pathways (Figure 4B).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Figure 4.

A.



B.

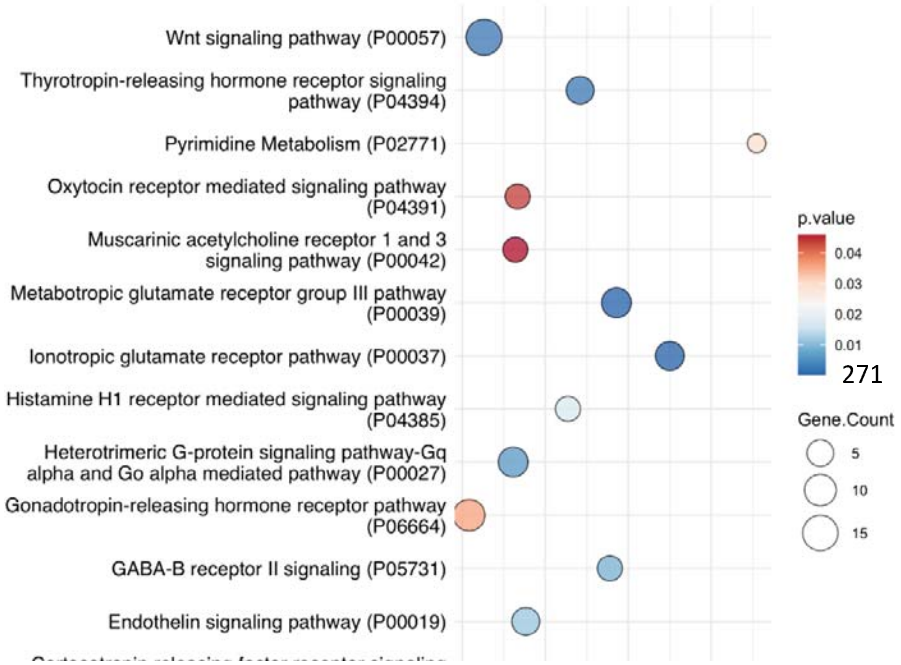


Figure 4.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Differential DNAm analyses of PBMCs stimulated *in vitro* with SARS-CoV-2. A. Venn diagrams depicting the overlap of DMCs from the SARS-CoV-2 *in vitro* stimulated PBMCs. Intraindividual comparisons of differential DNAm were performed in treated *vs.* untreated PBMCs from four different blood donors (D1-D4) collected before the start of the COVID-19 pandemic (2014-2019). DMCs were defined as a fold change in M-value >|2|. These DMCs were further mapped to their corresponding annotated genes (DMGs, n=542). B shows results from pathway over-representation analyses in PANTHER based on the 542 DMGs originating from the SARS-CoV-2 *in vitro* stimulated PBMCs compared to non-stimulated PBMCs. Pathways with a nominal p-value < 0.05 were considered significant.

Comparisons between *in vivo* and *in vitro* setting as well as network analyses reveal overlaps to SARS-CoV-2 interactome

As similar pathways were revealed in the findings from the clinical study and the SARS-CoV-2 stimulations, we wanted to explore further similarities in DNAm between the *in vivo* and *in vitro* settings. Analyses of the overlap of shared DMGs identified in the two comparisons revealed eight overlapping DMGs (*OR12D3*, *PCSK6*, *INPP5A*, *RAD51B*, *CDH4*, *PHACTR3*, *CDH13*, *SFTA2*), of which one (*PCSK6*) was identified as directly interacting with SARS-CoV-2. Additionally, to understand the biological context of the genes identified in the *in vitro* comparison, we performed network analyses in the same manner as for the *in vivo* comparison. These analyses found a module consisting of six genes (*TP53*, *INS*, *HSPA4*, *SP1*, *ESR1* and *FAS*), which were among the previously identified module genes from the *in vivo* setting and also were identical to those that had been identified by more than two module identification methods (Figure 4A). Furthermore, explorations of the overlap between identified genes in the differential DNAm analyses and network module analyses to the genes from the SARS-CoV-2 interactome identified numerous interactions in the *in vivo* (n=11/54), *in vitro* (n=100/542) and network module setting (n=33/66) (Figure S3).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

DISCUSSION

The epigenetic events triggered during a mild COVID-19 infection are largely unknown, despite the fact that these individuals make up a majority of all SARS-CoV-2-infected individuals. In this study, we observed changes in the DNA methylome of PBMCs from CC19s compared to non-infected individuals. A number of recent studies have studied DNAm patterns in severely ill patients with COVID-19, mainly reflecting the acute phase of the immune response. For instance, genes involved in antiviral responses driven by interferons were shown to be transcriptionally inhibited by hypermethylation, in severely ill COVID-19 patients compared to controls, while genes originating from inflammatory responses were granted transcriptional accessibility through general hypomethylation (17). Other studies reported on DNAm patterns in whole blood of COVID-19 sufferers, comparing hospitalised severely ill individuals to mildly ill and healthy individuals (18), pre-pandemic controls (19) and asymptomatic individuals (20), yet again showing mainly engagement of several antiviral immunity-related pathways. Whether the changes we found are reflecting an antiviral defence mechanism or reflect viral manipulation of the host epigenome warrants further studies.

The main finding of our study was that a number of genes in the networks deriving from both the DNA methylomes of mildly ill COVID-19 subjects and the *in vitro* stimulated PBMCs methylomes were shared and consistently found by several module identification methods. This may indicate the importance of these genes as hubs for protein-protein interactions in the course of SARS-CoV-2 infection and recovery. One of these genes was tumor protein 53 (TP53), an evolutionarily conserved protein that is one of the most well-studied hub genes in

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

cell signalling due to its central role in cancer (21) and that interacts with a variety of viral proteins from different classes (https://thebiogrid.org)(22). The ability of mutual inhibition and downregulation has been shown for TP53 and one of the previously identified SARS coronaviruses – SARS-CoV (23). Furthermore, TP53 has in several other studies been identified as a hub gene, in whole blood from COVID-19 patients (24), and interacting with ACE2 in SARS-CoV-2-infected human induced pluripotent stem cell-derived cardiomyocytes (25). In line with findings from our study, transcriptomic analyses of PBMCs from a small group of patients infected with SARS-CoV-2 revealed involvement of apoptosis and p53 signalling pathways (26). The relevance of this is supported by studies of the SARS-CoV-2 interactome, where TP53 was identified as a central player in apoptosis-mediated pathways (27). In terms of apoptosis, the module gene Fas cell surface death receptor, FAS, is also highly relevant. Higher expression of FAS on CD4+ T cells have been shown to correlate with lower cell counts in Covid-19 patients (28). Along the same lines, elevated circulating levels of the soluble form of FAS have recently been suggested to be causally contributing to the severity of Covid-19, and may in turn originate from genetic splice variants (29). Furthermore, apoptosis of T cells in PBMCs induced by FAS was reported to be increased in Covid-19 patients (30), which along with the involvement of TP53 could explain the lymphopenia frequently observed in COVID-19 subjects.

Interestingly, reports on differentially expressed genes overlapping between acute respiratory distress syndrome and venous thromboembolism datasets identified both TP53 and HSP90AA1, one of our other identified central genes, among the top ranked hub genes in their networks (31). HSP90AA1 was furthermore shown to be upregulated in bronchial cells of patients with mild COVID-19 disease, as compared to those with a severe disease course (32), suggesting that this gene may be of particular importance in the mounting of a protective antiviral response. Another heat shock protein in the network derived from our *in vivo* and *in vitro* data, HSPA4, directly interacts with the SARS-CoV-2 M and N proteins and also three of the virus' non-structural proteins (https://thebiogrid.org)(22). In fact, this heat

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

shock protein, a member of the HSP70 family, also binds proteins of the Human Herpes Virus 4 and HIV. Although our study does not provide any evidence for a protective role, HSP70 family members have been discussed as antiviral defence components (33, 34). In addition, HSP70 members have been suggested as drug targets in viral infections, including SARS-CoV-2 (35). Another interesting hub gene was *CFTR*, for which there is evidence for correlations of the inactivating delta F508 polymorphism, which is protective against chloride ion secreting diarrhoeas, with prevalence and mortality in Covid-19 (36). This is particularly interesting as SARS-CoV-2-induced diarrhoea has been suggested to involve Ca2+ activated chloride channels (37). Similarly, it has been hypothesised that transport of chloride ions over CFTR may be implicated in Covid-19-induced lung oedema (38). Furthermore, the muscarinic acetylcholine receptor 1 and 3 signalling pathway was present in over-representation analyses of genes from the network analyses as well as the *in vitro* stimulations. In post-viral fatigue patients, including post-SARS-CoV and myalgic encephalomyelitis/chronic fatigue syndrome patients, this signalling pathway is dysfunctional due to the development of anti-muscarinic receptor autoantibodies (39, 40). Although this was not investigated in our study, this could suggest that these pathways found may be implicated in the development of for instance post-acute COVID-19 syndrome, as the effects we observe may have persisted for months after the initial exposure to the virus. Altogether, the network centrality of the hub genes that we derived from the *in vivo* and *in vitro* data suggests that they may be of particular importance in the interaction with epigenetically modulated genes upon SARS-CoV-2 infection. Nevertheless, further studies are needed to elucidate the mechanistic role of these genes during infection and recovery from COVID-19.

Although an obvious limitation of the study is the lack of validation of the DNAm findings on a transcriptional level, it serves as a pilot study that generates hypotheses for further studies within the field. Hence, whether the observed DNAm patterns are indeed associated or even causally linked to host protective or host detrimental immune responses still needs to be addressed in future studies. With more well-designed, larger, consecutive sample materials,

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

possibly also in closer proximity to the time of infection with SARS-CoV-2, it will be possible to study the role of DNAm alterations in anti-viral defence and in viral manipulation of the same.

An advantage of the investigation of epigenetic modifications in in mild to moderately ill patients, is that we may be able to discern DNAm differences that otherwise would have been masked due to an overriding inflammatory response. These subtle changes may not only be relevant to how a less severe immune response is mounted towards SARS-CoV-2, but also in the case of long-COVID-19. The presentation of longstanding symptoms could be caused by detrimentally changed DNAm patterns, originally triggered as a short-term anti-viral response. This should be explored in detail in further studies since the risk is that these short-term responses may permanently alter and erroneously manifest in the DNA methylome.

CONCLUSIONS

In conclusion, we found epigenome-wide differences in DNAm patterns of individuals that had recovered from a mild-to-moderate disease course of COVID-19 compared to non-infected controls. The DNAm changes observed during *in vivo* and *in vitro* exposure to SARS-CoV-2 were translated to pathways of central relevance to COVID-19 through network analyses. The study suggests that DNAm is one of several epigenetic mechanisms that are altered upon SARS-CoV-2 infection. However, whether the effects are reflecting targeted viral hijacking to evade host defence or host-protective antiviral defence mechanisms remains to be determined.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

METHODS

Study population

In this study, participants were enrolled between May 29th and July 10th 2020 during the first wave of the SARS-CoV-2 pandemic in Linköping, Sweden. Individuals who had recovered from and individuals who had not experienced COVID-19 were recruited after announcements with leaflets. Exclusion criteria were the existence of current active SARS-CoV-2 infection and/or other infectious disease symptoms, as well as being younger than 18 years. The study participants voluntarily entered the study in a consecutive manner. The study was conducted on blood and saliva samples from in total 38 individuals from three different groups; non-infected controls (Con, n=18), COVID-19 convalescents (CC19, n=14) and symptom-free individuals with SARS-CoV-2-specific T cell responses (SFT, n=6). Additionally, blood samples from anonymous healthy blood donors from the blood bank at Linköping University Hospital before 2020 were included as a separate group in the analyses (pre20, n=5), collected between 2014-2019 prior to the outbreak of the pandemic. CC19 participants presented with either mild or asymptomatic initial infection, and none was admitted to hospital. Cons were defined as neither having any positive circulating IgG-antibody or T cell responses to SARS-CoV-2, while CC19s were defined by the presence of SARS-CoV-2-specific IgG antibodies in plasma using suspension multiplex immunoassay

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

(SMIA), some of which were positive for IgG in saliva, rapid test and in T cell responses as well. From the included individuals, the following information was retrieved using health questionnaires: self-reported COVID-19 symptoms (if applicable, one or several of the following: fever, headache, shortness of breath, loss of smell/taste, cough, fatigue, muscle pain, nausea, sinusitis/congestion), date of self-reported symptoms, weeks between symptoms and sampling, age, sex, smoking, weight, height, comorbidities as well as medications. The blood and saliva from the study participants was processed in a Biosafety level-2 facility. For samples from the natural exposure cohort, all participants provided written informed consent, and the present study was approved by the Regional Ethics Committee for Human Research in Linköping (Dnr. 2019-0618). Regarding the anonymous blood samples used for *in vitro* experiments, informed consent was given by the healthy donors at the time of blood donation and the use of the donated blood for research purposes was guaranteed as per the guidelines of Regional Ethics Committee for Human Research in Linköping and the Helsinki Declaration.

PBMC and plasma isolation from whole blood

Peripheral blood was collected in three 10 ml EDTA tubes (BD Vacutainer, 10331254, Fisher Scientific, Sweden). Up to 20 ml of whole blood was used for PBMC isolation after Ficoll-Paque Plus gradient centrifugation (GE17-1440-03, GE Healthcare Life Sciences, Sigma-Aldrich, Sweden) with SepMateTM tubes (85450, StemCell technologies, France) according to the manufacturer's protocol. Cells were frozen in 10% DMSO (10103483, Fischer Scientific, Sweden) in fetal bovine serum (FBS) (10270106, Gibco, Fischer Scientific, Sweden) and kept at -150°C until analysis. After thawing, the cells were washed twice in cell culture medium (RPMI medium 1640, 31870-025, 10% fetal bovine serum, 1% penicillin/streptomycin, 15140, 1% L-glutamine, 25030081, all from Gibco, Fischer Scientific, Sweden) further on termed as complete culture medium, prior to further processing. Up to 10

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

ml of whole blood was used for plasma separation by centrifugation (2000g for 15min, 4°C) and aliquots were stored at -80°C till further analysis.

Measurements of SARS-CoV-2-specific T cell responses using ELISpot

Peptides for the spike (S) protein of SARS-CoV-2 were obtained from Mabtech (3629-1, Sweden) and were reconstituted with di-methyl-sulphoxide (DMSO) at a concentration of 200 µg/ml according to the manufacturer's instructions. The SARS-CoV-2 S1 scanning pool contains 166 peptides consisting of 15-mers, overlapping with 11 amino acids, covering the S1 domain of the spike S1 protein (amino acid 13-685). The peptides were combined into one pool. IFN-γ ELISpot Plus kit was purchased from Mabtech (3420-4HST-10, Sweden). Briefly, the pre-coated wells were plated with unfractionated PBMCs at counts of 300 000 cells/well, and the cells were cultured with peptides for the S protein of SARS-CoV-2 at a final concentration of 2 µg/ml (diluted in complete culture medium) for 20 to 22 hrs in a 37°C, 5% CO_2 incubator. Cells cultured with medium alone were used as negative controls. Stimulation with anti-CD3 antibody at a concentration of 1 µg/ml was used as a positive control for each subject. Anti-CD28 antibody (3608-1-50, Mabtech, Sweden) was included at a final concentration of 0.1 µg/ml as a co-stimulator. All experiments were conducted in duplicates and results represent the mean of the duplicates. The plates were then processed according to the manufacturer's protocol. Estimation of specific T cell numbers was expressed as spot-forming cells per $1x10^6$ PBMCs (SFC). SFC were counted using an automated reading system (BioSys Bioreader 5000 Pro-F beta, Bio-sys GmbH, Germany) and assessed with the Bioreader 5000 analyser. A stimulation index was calculated by dividing the SFC elicited by a SARS-CoV-2 stimulus by the SFC present in the negative control wells. An increment value was calculated by subtracting the SFC from the negative control wells from the SFC of the stimulated wells. A stimulus was considered to be positive when the stimulation index was >2, and the increment value was >10.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Saliva samples

Prior to saliva collection, participants were required to rinse their mouth with water and confirmed they did not show documented oral disease or injury, that they had fasted, refrained from smoking, chewing a gum, taking oral medication, tooth brushing for a minimum of 1 hour before sampling and that no dental work had been performed within 24 hours prior to sample collection. Donors were asked to provide a 5 ml sample of saliva in a 50 ml sterile conical tube by passive drooling.

All saliva samples were stored/transported on ice upon receipt of the laboratory for processing to preserve sample integrity. Samples were centrifuged (2500g for 20 minutes at 4°C) to pellet cells and insoluble matter. The supernatant was collected and samples were complemented with cOmplete™ protease (#11836170001, Sigma) and PierceTM phosphatase inhibitor cocktails (#88667, Thermo Scientific), aliquoted and frozen/stored at -80°C on the same day. On the day of the assay, samples were thawed and micro-centrifuged (2500g for 10 minutes at 4°C) prior to analysis.

Antibody responses in plasma and saliva using Suspension Multiplex Immunoassay (SMIA)

MagPlex-C microspheres (Luminex Corp., Austin, TX, USA) were used for the coupling of antigens according to the manufacturer's protocol as previously described (41). Briefly, 200 µl of the stock microsphere solution (1.25×10^7 beads/ml) were coupled by adding 10 µg of recombinant SARS-CoV-2 Spike protein RBD His-Tag (#40592-V08B, SinoBiological Inc., USA). After the coupling, beads were incubated in phosphate buffered saline (PBS: 0.15 M sodium chloride, 10 mM sodium phosphate, pH 7.4) containing 0.05% (v/v) Tween 20 (PBS-T) for 15 min on a rocking shaker at RT. The beads were then washed with 0.5 ml StabilGuard solution (SurModics, Eden Prairie, MN, USA, #SG01-1000) using a magnetic separator (Milliplex® MAG handheld magnetic separation block for 96-well plates, Millipore

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Corp. Missouri, USA. Cat. #40-285) and resuspended in 400 µl of StabilGuard solution. The coupled beads were stored at 4°C in the dark until further use.

For plasma samples, 50 µl of plasma diluted 1:1000, and for saliva samples 50 µl of sample diluted 1:2 in PBS-T containing and 1% (v/v) BSA (Sigma-Aldrich Sweden AB, Stockholm, Sweden, #Sigma-Aldrich-SRE0036) (PBS-T + 1% BSA) was added per well of a flat bottom, 96-well µClear non-binding microtiter plate (Greiner Bio-One GmbH, Frickenhausen, Germany, #Greiner-655906). Fifty microliters of a vortexed and sonicated antigen-coupled bead mixture suspended in PBS-T + 1% BSA (~50 beads/µl) was then added to each well. The plate was incubated in the dark at 600 rpm for 1h at RT. The wells were then washed twice with 100 µl of PBS using a magnetic plate separator. The beads were resuspended in 100 µl of 1 µg/ml goat anti-human IgG-PE labelled antibody (Southern BioTech,, Birmingham, AL, USA. Cat. #2040-09) in PBS-T + 1% BSA and incubated for 30 min at RT in the dark with rotation at 600 rpm. The beads were subsequently washed twice with PBS, resuspended in 100 µl of PBS and analysed in a FlexMap 3D® instrument (Luminex Corporation, Austin, TX, USA) according to the manufacturer's instructions. A minimum of 100 events for each bead number was set to read and the median value was obtained for the analysis of the data. All sample analyses were repeated three times. A naked, non-antigen-coupled bead was included as a blank along with PBS-T + 1% BSA as a negative control.

Rapid test for SARS-CoV-2-specific IgG antibodies

SARS-CoV-2-specific antibody levels were validated using the Wondfo SARS-CoV-2 antibody test (lateral flow method) (Cat. # W195, Guangzhou, China) for rapid antibody testing. 10 µl of blood was added to the sample well and 80 µl of buffer solution in the buffer well, provided in the box by the manufacturer. The results were recorded as positive or negative based on band appearance according to the instructions by the manufacturer.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

In vitro stimulation with SARS-CoV-2

PBMC samples from four healthy blood donors, frozen in 2019 in -150 °C in foetal bovine serum (FBS) with 10% DMSO, were thawed and added to 10 ml of Gibco Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific, Waltham, US) containing 1% L-glutamine (Cat no: 25030-024, Gibco, Waltham, Massachusetts, USA), 1% penicillin-streptomycin (Cat no: 15140148 Gibco) and 10% normal human serum (NHS) (pooled from 5 donors) filtered through a 40 µm strainer and pre-heated to 37 °C. The cells were washed two times by centrifugation at 330g for 10 min. The pellet was resuspended in 1.5 ml medium and 2 million per donor were seeded in six-well plates and incubated for 16-24 h. The cell culture media were collected, and centrifugated at 330g for 5 min to pellet the non-adherent cells.

For *in vitro* infection experiments, SARS-CoV-2 virus previously isolated in a Biosafety level 3 lab according to local safety regulations from the nasopharyngeal aspirate of a COVID-19 patient (early April 2020) was used (42). The isolated virus was passaged five times in Vero E6 cells and for cell infection experiments, freeze-thawed medium supernatants of 4-5 days infected cells or mock supernatants were used. Virus titers were determined using immunoperoxidase assay. In brief, two-day old confluent cells (in a 96-well plate) were first washed with Dulbecco's Modified Eagle's Medium (DMEM) (Gibco, Code: 13345364) containing 100 µg/ml gentamicin, and 100 µl of 10-fold serially diluted SARS-CoV-2 virus lysate was added in quadruplicate. SARS-CoV-2 or mock Vero cell supernatant was added to the PBMC cultures corresponding to a multiplicity of infection of 0.01. 2 hours post infection the cells were washed twice with DMEM and 100 µl of fresh DMEM (containing 2% FBS and 100 µg/ml gentamicin) was added, and the plate was incubated for 8 hours at 37°C in presence of 5% CO_2. After incubation, the supernatant was discarded, and the cells were fixed for 2 hours with 4% formaldehyde. Next, Triton-X (1:500 in phosphate buffered saline, PBS) was added for 15 min, washed once with PBS and incubated for 2 hours at 37°C with PBS containing 3% BSA. Next, the cells were incubated with mouse-anti-dsRNA antibody

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

(Scions, Code: J2 at 1:100 dilution) for 1.5 h followed by detection using horseradish peroxidase–conjugated goat anti-mouse IgG (heavy plus light chain) (Catalog: 1706516, Bio-Rad Laboratories, Hercules, CA, USA) (1:1000) for 1 h. The plates were washed five times with PBS between every incubation, all incubations were done at room temperature and the antibody dilutions were made in PBS containing 1% BSA. Finally, the SARS CoV-2 infected Vero E6 cells were identified using 3-aminoethylcarbazole (AEC) substrate. The spots representing virus-infected cells were counted under the light microscope and the virus lysate was titrated to be 5×10^6 per ml.

Cells were monitored in the IncuCyte S3 live cell analysis system (Sartorious, Göttingen, Germany) to allow quantification of cell death in SARS-CoV-2 infected wells versus controls. After 48h incubation the cell culture media was collected from each well and centrifugated at 330g for 5 min to collect the non-adherent cells. Lysis buffer (RLT from the AllPrep® DNA/RNA Mini Kit, Qiagen, Hilden, Germany) was added to the wells to lyse adherent cells and the mixture was then added to the pelleted non-adherent cells in order to collect DNA (according to the manufacturer's instructions) from the entire PBMC fraction.

Epigenome-wide DNA methylation analyses

DNA extraction and quantification

For the performance of epigenome-wide DNA methylation analyses, DNA was extracted from the above isolated PBMCs (approximately $2x10^6$ cells) using the AllPrep® DNA/RNA Mini Kit (Cat no: 80204, Qiagen, Hilden, Germany) according to the manufacturer´s instructions. Concentrations of extracted DNA were measured using the Qubit® 4.0 Fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, U.S), using dsDNA High Sensitivity (HS) Assay Kit and RNA HS Assay Kit. The measurement was performed according to the manufacturer's instructions.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Illumina MethylationEPIC 850K array

DNA samples were sent to the Bioinformatics and Expression analysis Core facility, Karolinska Institutet, Stockholm, Sweden, where the samples first went through bisulphite conversion on site, followed by the performance of the Illumina Infinium MethylationEPIC 850K array. 200 ng of DNA from each sample was analysed.

Statistics

Descriptive analyses on demographic variables

Initial descriptive analyses of demographic variables were performed on the available information about age, gender, smoking and BMI (kg/m^2). Continuous variables were compared using an unpaired two-tailed t-test and categorical variables were examined using the Pearson $\chi 2$ test or Fisher's exact test (if the number of observations was smaller than five), see Table S1.

DNA methylation analyses

The resulting raw IDAT-files from the MethylationEPIC array analyses were processed in R programming environment (version 4.0.2). The analyses were identically performed for the clinical *in vivo* cohort and the *in vitro* experiment, unless stated otherwise.

Pre-processing and quality control *in vivo*

The resulting raw IDAT-files containing the raw DNA methylation profiles for each cell type were analysed in R (version 4.0.2) using the minfi package(43) (version 1.36.0) and the data were pre-processed in several steps. The following filters were applied: i) removal of probes with detection p-values above 0.01, ii) removal of non-CpG probes, iii) removal of multi-hit probes, iv) removal of all probes in X and Y chromosomes. We removed the sex chromosomes from our data set, as female X-inactivation skews the distribution of beta values (Figure S4). Of the initial 865 918 probes, 841 524 probes remained upon filtering.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

After filtering, quality control was performed, and normalisation of the data was done with subset-quantile within array (SWAN) normalisation method (44). The β-values and M-values of the samples were calculated against each probe per sample. The quality of the data was assessed before and after the normalisation (Figure S5). Thereafter, we performed singular value decomposition (SVD) analyses using the ChAMP package (45) (version 2.19.3) to identify underlying components of variation within the filtered and normalised data set (Figure S6). Significant components consisted of slide, batch and sample groups that contributed to variation within the data set. Corrections were performed for the identified components using ComBat from the SVA package (46) (version 3.38.0). As PBMCs consist of multiple nucleated cell types in peripheral blood, we utilised the Houseman method to infer cell type proportions within the samples (47). No differences could be determined in cell type proportions between any of the individuals or between sample groups (Table S7), motivating our choice of not correcting for these cell type proportions.

Differential DNA methylation analysis *in vivo*

As we were interested in studying CpGs that were differentially methylated between CC19s and non-infected controls from both before and after the start of the COVID-19 pandemic, we performed differential DNA methylation analyses, using the limma package (version 3.46.0). A linear model was fitted to the filtered, normalised and SVD-corrected DNA methylation data. Identified sources of variation that were still present upon SVD correction provided the basis for the inclusion of these variables as co-variates in the models, in this case gender and BMI (Figure S6). For each investigated probe, moderated t-statistics, log2 Fold Change (logFC) and p-values were computed. The logFC values represent the average beta methylation difference (from hereon referred to as mean methylation difference, MMD) between the CC19s *vs.* non-infected controls (Cons + Pre20). Differentially methylated CpGs (DMCs) were defined as CpG sites having a nominal p-value of less than 0.01 along with an MMD of > 0.2. As a means to ascertain the quality of the identified DMCs, genomic inflation

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

and pertaining bias were estimated using the BACON package (version 1.18.0). As the estimated genomic inflation for the comparison was close to 1 (genomic inflation: 1.20, bias: 0.01, Figure S7), this suggested that no major genomic inflation was present in the comparisons, and no correction for this was deemed necessary. The distribution of the DMCs among all investigated DNA methylation sites were illustrated by creating volcano plots (EnhancedVolcano, version 1.8.0). Thereafter, the DMCs were mapped to their corresponding DMGs. DMGs contained at least one DMC, and were considered hyper- or hypomethylated if all DMCs within the gene were hyper- or hypomethylated, respectively. If both hyper- and hypomethylated genes were present in the same gene, the gene was considered having a mixed methylation pattern.

Pre-processing and quality control *in vitro*

The resulting raw IDAT-files containing the raw DNA methylation profiles for each cell type were analysed in R (version 4.0.2) using the minfi package (43) (version 1.36.0) and the data were pre-processed in several steps. The following filters were applied: i) removal of probes with detection p-values above 0.01, ii) removal of non-CpG probes, iii) removal of multi-hit probes, iv) removal of all probes in X and Y chromosomes. In this dataset, we did not have any information on demographic variables, as the samples derived from anonymous donors. However, we still removed the sex chromosomes from our data set, as female X-inactivation skews the distribution of beta values. Of the initial 861 728 probes, 837 694 probes remained upon filtering. After filtering, quality control was performed, and normalisation of the data was done with subset-quantile within array (SWAN) normalsation method (44). The Houseman method was utilised to infer cell type proportions within the samples (47), yet again revealing no differences could be determined in cell type proportions between any of the individuals (Table S7), motivating our choice of not correcting for these cell type proportions. The β-values and M-values of the samples were calculated against each probe per sample. The quality of the data was assessed before and after the normalisation (Figure S8). SVA

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

package (version 3.40) was applied to correct the batch effect. Cell deconvolution was performed using FlowSorted.Blood.EPIC package (version 1.11).

Differential DNA methylation *in vitro*

To evaluate the difference between the MOCK and INFECTION, the fold change was calculated using the cut-off obtained from the density plot (M-value >|2|; Figure S9) for each CpG site. Only those CpGs with higher values than the cut-off, were selected for further analysis. Venn analysis was performed among the samples using the ggVennDiagram (version 1.1) package in R (version 4.0.3) and bioconductor (version 3.12).

Pathway over-representation analyses

To make biological sense of the putatively SARS-CoV-2-induced DNA methylation differences, we performed PANTHER pathway over-representation test analyses using the PANTHER database (version 16.0). The Fisher's exact test was used for generation of nominal p-values (significance level set to p-value of < 0.05), in case false discovery rate correction was too stringent. The significantly enriched pathways were displayed in dot plots generated in R using ggplot2 package (version 3.3.3).

Network analyses

A network analysis was conducted to generate further and wider biological insight about the DMGs generated in the *in vivo* setting. An input object was constructed using the pre-2020 (Pre20, n=5) and post-2020 (Con, n=18) non-exposed controls and COVID-19 convalescents (CC19, n=14), as a two-column data frame containing gene annotation and P-value of the significant DMGs (n=54). The graph clustering algorithm MCODE (48) was used to identify molecular complexes and create a large disease module, which was then fitted to a protein-protein interaction network, and both were analysed and rendered in Cytoscape (version 3.8.0). High confidence interactions with a STRINGdb confidence value >0.7 were displayed

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

in the network. Centrality measurements of degree, betweenness and closeness were used to expose the most central nodes in the network. Finally, a functional enrichment of the genes present within the module was carried out using StringDB (49). In addition, the inference of modules was performed with two other methods from the MODifieR package (DIAMOnD and WGCNA)(50) to study whether it was possible to condense the module genes to fewer genes of particular interest within the network, for both the *in vivo* and the *in vitro* setting.

Overlap to SARS-CoV-2 interactome

A publicly available protein-protein interaction (PPI) network of SARS-COV-2 and human genes curated by BioGRID (version 4.4.197) was downloaded from the Network Data Exchange in Cytoscape (version 3.8.0). The DMGs from the *in vivo* and *in vitro* setting alongside the gene list from the module generated by MCODE were overlapped onto the PPI network to visualise their respective distributions.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

LIST OF ABBREVIATIONS

CC19 – convalescent COVID-19 individuals

Con – Control (uninfected pandemic)

COVID-19 – Coronavirus disease -19

DMC – differentially methylated CpG site

DMG – differentially methylated gene

DNAm – DNA methylation

MMD – mean methylation difference

PBMC – peripheral blood mononuclear cell

PC – principal component

PCA – principal component analysis

PPI – protein-protein interaction

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Pre20 – Control (uninfected, pre-pandemic 2020)

SARS – severe acute respiratory syndrome

SARS-CoV-2 – severe acute respiratory syndrome corona virus 2

SFT – symptom-free individuals with T cell response

SMIA – suspension multiplex immunoassay

REFERENCES

1. Schmidl C, Delacher M, Huehn J, Feuerer M. Epigenetic mechanisms regulating T-cell responses. The Journal of allergy and clinical immunology. 2018;142(3):728-43.
2. Mueller AL, McNamara MS, Sinclair DA. Why does COVID-19 disproportionately affect older people? Aging. 2020;12(10):9959-81.
3. Martin EM, Fry RC. Environmental Influences on the Epigenome: Exposure-Associated DNA Methylation in Human Populations. Annual Review of Public Health. 2018;39(1):1-25.
4. Zhang L, Lu Q, Chang C. Epigenetics in Allergy and Autoimmunity. Advances in Experimental Medicine and Biology. 2020;1253:3-55.
5. Jin J, Xu H, Wu R, Niu J, Li S. Aberrant DNA methylation profile of hepatitis B virus infection. Journal of Medical Virology. 2019;91(1):81-92.
6. Zhang X, Justice AC, Hu Y, Wang Z, Zhao H, Wang G, et al. Epigenome-wide differential DNA methylation between HIV-infected and uninfected individuals. Epigenetics. 2016;11(10):750-60.
7. Oriol-Tordera B, Berdasco M, Llano A, Mothe B, Gálvez C, Martinez-Picado J, et al. Methylation regulation of Antiviral host factors, Interferon Stimulated Genes (ISGs) and T-cell responses associated with natural HIV control. PLoS pathogens. 2020;16(8).
8. Das J, Idh N, Pehrson I, Paues J, Lerm M. A DNA methylome biosignature in alveolar macrophages from TB-exposed individuals predicts exposure to mycobacteria. MedRxiv : the preprint server for health sciences. 2021.
9. Karlsson L, Das J, Nilsson M, Tyrén A, Pehrson I, Idh N, et al. A Differential DNA Methylome Signature of Pulmonary Immune Cells from Individuals Converting to Latent Tuberculosis Infection. MedRxiv : the preprint server for health sciences. 2021.
10. Pehrson I, Das J, Idh N, Karlsson L, Rylander H, Hård af Segerstad H, et al. DNA methylomes derived from alveolar macrophages display distinct patterns in latent tuberculosis - implication for interferon gamma release assay status determination. 2021.
11. Menachery VD, Schäfer A, Burnum-Johnson KE, Mitchell HD, Eisfeld AJ, Walters KB, et al. MERS-CoV and H5N1 influenza virus antagonize antigen presentation by altering the epigenetic landscape. Proceedings of the National Academy of Sciences. 2018;115(5).

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

12. Schäfer A, Baric R. Epigenetic Landscape during Coronavirus Infection. Pathogens. 2017;6(1):8.
13. Sen R, Garbati M, Bryant K, Lu Y. Epigenetic mechanisms influencing COVID-191. Genome. 2021;99(999):1-14.
14. Chlamydas S, Papavassiliou AG, Piperi C. Epigenetic mechanisms regulating COVID-19 infection. Epigenetics. 2020:1-8.
15. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China. JAMA. 2020;323(13):1239-42.
16. Li J, Huang DQ, Zou B, Yang H, Hui W, Rui F, et al. Epidemiology of COVID-19: A systematic review and meta-analysis of clinical characteristics, risk factors, and outcomes. Journal of Medical Virology. 2021;93(3):1449-58.
17. Corley MJ, Pang A, Dody K, Mudd PA, Patterson BK, Seethamraju H, et al. Genome-wide DNA methylation profiling of peripheral blood reveals an epigenetic signature associated with severe COVID-19. Journal of Leukocyte Biology. 2021.
18. Zhou S, Zhang J, Xu J, Zhang F, Li P, He Y, et al. An epigenome-wide DNA methylation study of patients with COVID-19. Annals of human genetics. 2021.
19. Balnis J, Madrid A, Hogan KJ, Drake LA, Chieng HC, Tiwari A, et al. Blood DNA methylation and COVID-19 outcomes. Clinical epigenetics. 2021;13(1):118.
20. Castro de Moura M, Davalos V, Planas-Serra L, Alvarez-Errico D, Arribas C, Ruiz M, et al. Epigenome-wide association study of COVID-19 severity with respiratory failure. EBioMedicine. 2021;66:103339.
21. Levine AJ. p53: 800 million years of evolution and 40 years of discovery. Nature reviews Cancer. 2020;20(8):471-80.
22. Oughtred R, Rust J, Chang C, Breitkreutz B-JJ, Stark C, Willems A, et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein science : a publication of the Protein Society. 2021;30(1):187-200.
23. Ma-Lauer Y, Carbajo-Lozoya J, Hein MY, Müller MA, Deng W, Lei J, et al. p53 down-regulates SARS coronavirus replication and is targeted by the SARS-unique domain and PLpro via E3 ubiquitin ligase RCHY1. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(35):201.
24. Vastrad B, Vastrad C, Tengli A. Identification of potential mRNA panels for severe acute respiratory syndrome coronavirus 2 (COVID-19) diagnosis and treatment using microarray dataset and bioinformatics methods. 3 Biotech. 2020;10(10):422.
25. Wicik Z, Eyileten C, Jakubik D, Simões SNN, Martins DC, Pavão R, et al. ACE2 Interaction Networks in COVID-19: A Physiological Framework for Prediction of Outcome in Patients with Cardiovascular Risk Factors. Journal of clinical medicine. 2020;9(11).
26. Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerging microbes & infections. 2020;9(1):761-70.
27. Tiwari R, Mishra AR, Gupta A, Nayak D. Structural similarity-based prediction of host factors associated with SARS-CoV-2 infection and pathogenesis. Journal of biomolecular structure & dynamics. 2021:1-12.
28. Bellesi S, Metafuni E, Hohaus S, Maiolo E, Marchionni F, D'Innocenzo S, et al. Increased CD95 (Fas) and PD-1 expression in peripheral blood T lymphocytes in COVID-19 patients. British journal of haematology. 2020;191(2):207-11.
29. Klaric L, Gisby JS, Papadaki A, Muckian MD, Macdonald-Dunlop E, Zhao JH, et al. Mendelian randomisation identifies alternative splicing of the FAS death receptor as a mediator of severe COVID-19. medRxiv : the preprint server for health sciences. 2021.
30. Zhu L, Yang P, Zhao Y, Zhuang Z, Wang Z, Song R, et al. Single-Cell Sequencing of Peripheral Mononuclear Cells Reveals Distinct Immune Response Landscapes of COVID-19 and Influenza Patients. Immunity. 2020;53(3):685-696000.
31. Mishra A, Chanchal S, Ashraf MZ. Host-Viral Interactions Revealed among Shared Transcriptomics Signatures of ARDS and Thrombosis: A Clue into COVID-19 Pathogenesis. TH open : companion journal to thrombosis and haemostasis. 2020;4(4).
32. Ma D, Liu S, Hu L, He Q, Shi W, Yan D, et al. Single-cell RNA sequencing identify SDCBP in ACE2-positive bronchial epithelial cells negatively correlates with COVID-19 severity. Journal of cellular and molecular medicine. 2021.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

33. Kim MY, Shu Y, Carsillo T, Zhang J, Yu L, Peterson C, et al. hsp70 and a novel axis of type I interferon-dependent antiviral immunity in the measles virus-infected brain. Journal of virology. 2013;87(2):998-1009.
34. Kim MY, Ma Y, Zhang Y, Li J, Shu Y, Oglesbee M. hsp70-dependent antiviral immunity against cytopathic neuronal infection by vesicular stomatitis virus. Journal of virology. 2013;87(19):10668-78.
35. Tampere M, Pettke A, Salata C, Wallner O, Koolmeister T, Cazares-Körner A, et al. Novel Broad-Spectrum Antiviral Inhibitors Targeting Host Factors Essential for Replication of Pathogenic RNA Viruses. Viruses. 2020;12(12).
36. Delanghe JR, De Buyzere ML, Speeckaert MM. Genetic Polymorphisms in the Host and COVID-19 Infection. Advances in experimental medicine and biology. 2021;1318:109-18.
37. Donowitz M, Tse C-MM, Dokladny K, Rawat M, Horwitz I, Ye C, et al. SARS-COV-2 induced Diarrhea is inflammatory, Ca 2+ Dependent and involves activation of calcium activated Cl channels. bioRxiv : the preprint server for biology. 2021.
38. Abdel Hameid R, Cormet-Boyaka E, Kuebler WM, Uddin M, Berdiev BK. SARS-CoV-2 may hijack GPCR signaling pathways to dysregulate lung ion and fluid transport. American journal of physiology Lung cellular and molecular physiology. 2021;320(3).
39. Blomberg J, Gottfries C-GG, Elfaitouri A, Rizwan M, Rosén A. Infection Elicited Autoimmunity and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: An Explanatory Model. Frontiers in immunology. 2018;9:229.
40. Skiba MA, Kruse AC. Autoantibodies as Endogenous Modulators of GPCR Signaling. Trends in pharmacological sciences. 2021;42(3):135-50.
41. Rizwan M, Rönnberg B, Cistjakovs M, Lundkvist Å, Pipkorn R, Blomberg J. Serology in the Digital Age: Using Long Synthetic Peptides Created from Nucleic Acid Sequences as Antigens in Microarrays. Microarrays. 2016;5(3):22.
42. Nissen K, Hagbom M, Krambrich J, Akaberi D, Sharma S, Ling J, et al. Presymptomatic viral shedding and infective ability of SARS-CoV-2; a case report. Heliyon. 2021;7(2).
43. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics (Oxford, England). 2014;30(10):1363-9.
44. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome biology. 2012;13(6).
45. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics (Oxford, England). 2014;30(3):428-30.
46. Leek JT, Johnson EW, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882-3.
47. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012;13:86.
48. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC bioinformatics. 2003;4:2.
49. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research. 2019;47(D1).
50. de Weerd HA, Badam TVS, Martínez-Enguita D, Åkesson J, Muthas D, Gustafsson M, et al. MODifieR: an ensemble R package for inference of disease modules from transcriptomics networks. Bioinformatics (Oxford, England). 2020;36(12):3918-9.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

DECLARATIONS

Ethics approval and consent to participate:

All participants provided informed consent prior to inclusion in the study. Ethical permission for this study has been granted by the Regional Ethics Committee for Human Research in Linköping (Dnr. 2019-0618).

Consent for application

Not applicable

Availability of data and materials

The datasets used and/or analysed in the presented work will be available upon publication due to a pending patent, for reference: GeneExpression Omnibus (GEO-ID: GSE178962).

Competing interests

None of the authors declare any competing interests.

medRxiv preprint doi: https://doi.org/10.1101/2021.07.05.21260014; this version posted July 6, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Funding

This study was supported by grants from the Swedish Heart and Lung Foundation (M.L.: 20200319), the Swedish Research Council (A.R.: Covid-19/biobank 210202#1) and the Open Medicine foundation (A.R.: OMF190626).

Author's contributions

M.L. and A.R. designed the study. A.R., E.A. and M.R. were responsible for sample collection and performed the ELIspot and SMIA analyses. M.L., L.K and S.Sh. performed the *in vitro* SARS-CoV-2 stimulation of PBMCs. L.K. prepared all PBMC samples for epigenome-wide DNA methylation analyses. J.H. and J.D. provided guidance and expertise on the performance of the statistical and bioinformatic analyses. S.Sa. has lead and performed the majority of the statistical and bioinformatic analyses, with support from J.D., L.K., L.P. and J.H. The findings were presented by J.H., S.Sa., L.K., J.D. and L.P. All authors interpreted and discussed the results. J.H. drafted the manuscript. All authors contributed to and approved the final draft for publication.

Acknowledgements

We would like to thank the Bioinformatics and Expression analysis Core facility at Karolinska Institutet for their fruitful collaboration and the performance of the Illumina Infinium MethylationEPIC 850K arrays described in this paper.

Author's information (optional)

Not applicable