CCR1 regulatory variants linked to pulmonary macrophage recruitment in severe COVID-19

Authors: Bernard Stikker[1], Grégoire Stik[2,3], Rudi Hendriks[1]* and Ralph Stadhouders[1,4]*

[1] Department of Pulmonary Medicine, Erasmus MC, Rotterdam, the Netherlands
[2] Centre for Genomic Regulation (CRG) and Institute of Science and Technology (BIST), Barcelona, Spain
[3] Universitat Pompeu Fabra (UPF), Barcelona, Spain
[4] Department of Cell Biology, Erasmus MC, Rotterdam, the Netherlands
*Corresponding authors (r.hendriks@erasmusmc.nl, r.stadhouders@erasmusmc.nl)

Abstract

Genome-wide association studies have identified 3p21.31 as the main risk locus for severe symptoms and hospitalization in COVID-19 patients. To elucidate the mechanistic basis of this genetic association, we performed a comprehensive epigenomic dissection of the 3p21.31 locus. Our analyses pinpoint activating variants in regulatory regions of the chemokine receptor-encoding *CCR1* gene as potentially pathogenic by enhancing infiltration of monocytes and macrophages into the lungs of patients with severe COVID-19.

Main text

Coronavirus disease 2019 (COVID-19) is a potentially life-threatening respiratory disorder caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1]. Clinical manifestations of SARS-CoV-2 infection range from no or mild symptoms to respiratory failure. Life-threatening disease is often associated with an excessive inflammatory response to SARS-CoV-2, involving elevated systemic cytokine levels and profound organ infiltration by monocytes and macrophages[2]. Besides clinical characteristics such as age and various comorbidities[3], genetic differences may play a role in predisposing individuals to progress towards severe disease. Genome-wide association studies (GWASs) have recently identified genetic variants associated with hospitalized COVID-19[4,5]. The strongest genetic association was consistently detected on chromosome 3 at 3p21.31, harbouring multiple genes. However, it is currently unclear how variants and genes in this locus affect COVID-19 disease pathophysiology.

Here, we integrated COVID-19 GWAS findings with publically available data from large-scale epigenome profiling, 3D chromatin interaction mapping, transcriptomics and expression quantitative trait loci (eQTL) studies. Version 4 COVID-19 GWAS meta-analyses (provided by the COVID-19 Host Genetics Initiative[6]; https://www.covid19hg.org/) confirmed the striking association between the 3p21.31 locus and severe COVID-19, both when comparing hospitalized COVID-19 patients with healthy control subjects (**Fig.1a**) or with non-hospitalized patients (**Supplementary Fig.1a**). We focused on the former comparison (8,638 hospitalized COVID-19 patients vs. 1,736,547 control subjects) to maximize the number of associated SNPs for downstream analysis. Regional association plots generated using the Functional Mapping and Annotation (FUMA) platform[7] revealed a region of 743 kb in high linkage-disequilibrium (LD; $r^2 > 0.8$) with independent significant ($P < 5e-8$) GWAS SNPs (**Fig.1b**). This genomic region includes 12 known protein-coding genes, including a cluster of chemokine receptor genes highly relevant for anti-viral immunity[8].

Figure 1

Figure 1. The 3p21.31 severe COVID-19 risk locus is highly active in monocytes and macrophages that express *CCR1*. (**a**) Manhattan plot showing the strong association of the 3p21.31 locus with hospitalized COVID-19 (based on v4 GWAS meta-analysis release, https://www.covid19hg.org/). (**b**) FUMA region plot of the 3p21.31 locus highlighting all variants in high linkage-disequilibrium ($r^2 > 0.8$, $P < 0.05$) with independent significant ($P < 5e-8$) GWAS SNPs. Bar graph denotes SNP distribution. (**c-d**) Normalized gene expression levels (transcripts per million; TPM) of 3p21.31 candidate genes across 54 tissues (GTEx database, panel c) or various immune cell subsets from peripheral blood (DICE database, panel d). (**e**) UCSC genome browser view of H3K27Ac ChIP-Seq tracks for the indicated cell or tissue types (BLUEPRINT/ENCODE databases; fibro. = fibroblast, epith. = epithelial, Mph = macrophage, mem. B = memory B cell). Genes and FUMA SNPs are shown above, shading indicates location of the most significant GWAS SNPs ('top SNPs'; orange) or monocyte/macrophage-specific active chromatin regions ('ACR1-3'; in yellow). (**f**) Circos plots showing significant chromatin interactions with the *CCR1* promoter (green dashed line) in inflammatory macrophages (M1 Mph) or CD8+ T cells measured by promoter-capture Hi-C (BLUEPRINT database; freq.: frequency). ACRs are indicated in orange. (**g**) Schematic indicating active chromatin hub formation involving the ACRs (enhancer; Enh.), CTCF binding sites and indicated *CCR* genes in monocytes/macrophages. (**h**) Hi-C interaction matrices (5 kb resolution, smoothened) for *in vitro* generated ('induced') macrophages before (left) and after (right) auxin-inducible CTCF degradation[15], resulting in weaker interactions (indicated by numbers) between indicated CCR genes and/or ACR1 (color code as in panel g). CTCF ChIP-Seq peaks are indicated below. (**i**) Gene expression changes of indicated genes in macrophages after CTCF degradation.

Approximately 96% of all COVID-19-associated SNPs fall in non-coding regions (**Fig.1b**), suggesting they may affect COVID-19 pathogenesis by altering local gene regulation. Gene expression analysis using GTEx v8 data[9] indicated largely tissue-invariant expression for several 5' genes (i.e. *LIMD1*, *LZTFL1*, *FYCO1*, *SACM1L*), whereas the seven chemokine receptor genes showed highly tissue-specific activity (**Fig.1c**). *CCR1* in particular was highly expressed in tissues containing hematopoietic cells (i.e. whole blood and spleen), but also in the lung. More specifically, mining ENCODE[10], BLUEPRINT[11] and DICE[12] databases revealed strong transcriptional activity of *CCR1*/*CCR2*/*CCR5* and nearby chromatin specifically in monocytes and macrophages as indicated by high levels of transcription, H3K27Ac and chromatin accessibility (**Fig.1d-e**, **Supplementary Fig.1b**). Regulatory activity was concentrated in three active chromatin regions (ACRs) (**Fig.1e**). Promoter-capture Hi-C data of hematopoietic cells[13] revealed extensive monocyte/macrophage-specific chromatin interactions between the three ACRs, as exemplified by *CCR1* and *CCR5* promoter interaction profiles in inflammatory macrophages and T cells (**Fig.1f**, **Supplementary Fig.1c-d**). Notably, the 49 kb region with the strongest statistical association in GWAS (labelled 'top SNPs') showed few signs of regulatory activity in the relevant tissue and cell types we examined (**Fig.1e**, **Supplementary Fig.1b**). Moreover, this region did not engage in chromatin interactions with nearby promoters (**Supplementary Fig.1e**).

Together, this analysis reveals that the 3p21.31 COVID-19 risk locus is predominantly active in monocytes and macrophages. Activity is centered on *CCR1* and its surroundings, which are organized in a 3D chromatin hub involving the other active *CCR* genes (i.e. *CCR2*, *CCR5*) and putative enhancer elements (**Fig.1g**) - a conformation often used for complex tissue-specific gene regulation[14]. In support, all three ACRs were bound by the architectural CCCTC-binding factor CTCF[14] and acute CTCF depletion in *in vitro* generated macrophages[15] resulted in chromatin hub disruption and *CCR1*/*CCR2*/*CCR5* - but not *CCR3* - downregulation (**Fig.1h-i**, **Supplementary Fig.1f-g**). Importantly, CCR1 as

well as CCR2 are critical mediators of monocyte/macrophage polarization and tissue infiltration[16], which are pathogenic hallmarks of severe COVID-19[2].

Figure 2. Variant prioritization and eQTL analysis identify _CCR1_ regulatory variants linked to increased _CCR1_ expression. (**a**) CADD and RegulomeDB scores for FUMA SNPs across the 3p21.31 locus. Red dashed lines indicate selected thresholds. SNPs passing the threshold are indicated in orange; those within monocyte H3K27Ac+/DNAse+ regulatory regions are in red. (**b**) UCSC genome browser view of H3K27Ac ChIP-Seq and DNAse-seq signals in the indicated cell types centered on the four candidate causal variants near _CCR1_. _CCR1_ promoter region shown in panel d is indicated by a green rectangle. (**c**) Zoom-in view of the _CCR1_ promoter (top) showing H3K27Ac, chromatin accessibility (DNAse-Seq) and binding of AP1 (i.e. ATF2, BATF, CREM) or IRF4 transcription factors in the indicated cells (ENCODE database) overlapping two candidate causal SNPs. Binding motifs modified by the variants are indicated (bottom). Rs3181080 optimizes an IRF-AP1 composite motif (A>T), while rs34919616 disrupts a BCL6 motif (G>A). (**d**) eQTL analysis of rs3181080/rs34919616 for _CCR1_ and _CCR3_. Dots indicate independent tissue eQTLs (P<0.005) detected per gene, normalized effect sizes show association with lower (-) or higher (+) expression. (**e**) Schematic indicating how 3p21.31 variants can increase risk of severe disease upon SARS-CoV2 infection. See text for details.

To further prioritize variants with potential biological significance we used RegulomeDB and CADD SNP annotations retrieved by FUMA[7]. Stringent filters for both scores were combined with localization within a putative monocyte regulatory region (H3K27Ac+ and DNAse+), yielding four unique candidate causal SNPs (**Fig.2a**). These variants mostly clustered near *CCR1* within ACR2 and altered putative transcription factor binding motifs (**Fig.2b**, **Supplementary Fig.2a**). Two SNPs within the *CCR1* promoter - including the rs3181080 variant that uniquely emerged as a top hit in both RegulomeDB and CADD annotations (**Fig.2a**) - affected binding motifs of known regulators of the macrophage inflammatory expression program (**Fig.2c**). Rs3181080 (beta=0.284, GWAS P=1.8e-16) optimizes a composite Interferon Regulatory Factor (IRF)-Activator Protein 1 (AP1) motif, which is used for cooperative binding of IRF and AP1 family transcription factors that promote monocyte/macrophage activation[17]. In line with *CCR1* activation by IRF/AP1 factors, binding of AP1 proteins and IRF4 to rs3181080 was detected in *CCR1*-expressing GM12878 lymphoblastoid cells (**Fig.2c**, **Supplementary Fig.2b**). Previous experiments in mouse macrophages[18] confirmed IRF binding to the *Ccr1* promoter (**Supplementary Fig.2c**). The second *CCR1* promoter variant, rs34919616 (beta=0.283, GWAS P=1.04e-16), disrupts a critical nucleotide in a motif for BCL6, a suppressor of inflammatory gene expression in macrophages[19,20]. To directly link these *CCR1* promoter variants to altered gene expression, we performed eQTL analysis using the GTEx v8 database[9]. Strikingly, the presence of rs3181080 and rs34919616, which are in very high LD (r^2=1), strongly correlates with higher *CCR1* expression in multiple tissues (n=7, P<0.005), including lung and lymphoblastoid cells (**Fig.2d**). In contrast, negative associations were detected with *CCR3* (**Fig.2d**).

Taken together, these data show that COVID-19-associated 3p21.31 variants likely induce increased *CCR1* expression in monocytes and macrophages by modulating transcription factor binding to the *CCR1* promoter. Our findings offer a mechanistic explanation for the genetic association of 3p21.31 risk alleles with severe COVID-19. Elevated CCR1 expression may enhance lung infiltration by monocytes and macrophages upon viral infection[16], provoking a rapid and deleterious hyperinflammation as observed in COVID-19 patients suffering from severe disease[2,21] (**Fig.2e**). In support of this notion, single cell RNA-sequencing revealed increased levels of *CCR1* and its ligands *CCL2/CCL3* in pulmonary macrophages from critical COVID-19 patients[22,23]. Additionally, CCL2 plasma levels showed the highest predictive value for mortality in a COVID-19 patient cohort[21]. These findings are in line with excessive pulmonary influx of CCR1+ monocytes and subsequent differentiation into inflammatory tissue macrophages as a hallmark of severe COVID-19 (**Fig.2e**)[22].

In summary, we provide evidence that common genetic variants may increase susceptibility to develop severe COVID-19 by modulating chemokine receptor expression in monocytes and macrophages, resulting in aggravated inflammatory responses and more severe disease. These data contribute to our understanding of COVID-19 disease heterogeneity and support therapeutic targeting of CCR1 activity in hospitalized COVID-19 patients.

References

1. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
2. Merad, M. & Martin, J. C. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. *Nat. Rev. Immunol.* **20**, 355–362 (2020).
3. Zheng, Z. *et al.* Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *J. Infect.* **81**, e16–e25 (2020).
4. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
5. Zeberg, H. & Pääbo, S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* **587**, 610–612 (2020).
6. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* **28**, 715–718 (2020).
7. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, (2017).
8. Hughes, C. E. & Nibbs, R. J. B. A guide to chemokines and their receptors. *FEBS J.* **285**, 2944–2971 (2018).
9. Aguet, F. *et al.* Cell type–specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020).
10. Abascal, F. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
11. Fernández, J. M. *et al.* The BLUEPRINT Data Analysis Portal. *Cell Syst.* **3**, 491-495.e5 (2016).
12. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715.e16 (2018).
13. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384.e19 (2016).
14. Ribeiro De Almeida, C., Stadhouders, R., Thongjuea, S., Soler, E. & Hendriks, R. W. DNA-binding factor CTCF and long-range gene interactions in V(D)J recombination and oncogene activation. *Blood* **119**, 6209–6218 (2012).
15. Stik, G. *et al.* CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. *Nat. Genet.* **52**, 655–661 (2020).
16. Mantovani, A. *et al.* The chemokine system in diverse forms of macrophage activation and polarization. *Trends Immunol.* **25**, 677–686 (2004).
17. Molawi, K. & Sieweke, M. H. Transcriptional Control of Macrophage Identity, Self-Renewal, and Function. in *Adv. Immunol.* **120**, 269–300 (2013).
18. Mancino, A. *et al.* A dual cis-regulatory code links IRF8 to constitutive and inducible gene expression in macrophages. *Genes Dev.* **29**, 394–408 (2015).
19. Xu, F. *et al.* Bcl6 Sets a Threshold for Antiviral Signaling by Restraining IRF7 Transcriptional Program. *Sci. Rep.* **6**, 1–13 (2016).
20. Toney, L. M. *et al.* BCL-6 regulates chemokine gene transcription in macrophages. *Nat. Immunol.* **1**, 214–220 (2000).
21. Lucas, C. *et al.* Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* **584**, 463 (2020).
22. Chua, R. L. *et al.* COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* **38**, 970–979 (2020).
23. Trump, S. *et al.* Hypertension delays viral clearance and exacerbates airway hyperinflammation in patients with COVID-19. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-00796-1.
24. Abascal, F. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
25. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).

Acknowledgements

We thank staff of the Erasmus MC Pulmonary Medicine department for fruitful discussions. We are indebted to the COVID-19 Host Genetics Initiative for immediately making their GWAS analyses publically available, and to the ENCODE, BLUEPRINT and GTEx consortia for providing the scientific community with invaluable genomics data. R.W.H. and B.S. are supported by Dutch Lung Foundation grant 4.1.18.226. R.S. is supported by an

Erasmus MC Fellowship and a Dutch Lung Foundation Junior Investigator grant (4.2.19.041JO).

Supplementary Figures

Supplementary Figure 1

Supplementary Figure 1. Transcriptional activity and chromatin interactions at the 3p21.31 severe COVID-19 risk locus. (**a**) Manhattan plot showing the association of the 3p21.31 locus with hospitalized COVID-19 (based on v4 GWAS meta-analysis release, https://www.covid19hg.org/). (**b**) UCSC genome browser overview of RNA-Seq, DNAse-Seq and H3K27Ac ChIP-Seq tracks for the indicated cell or tissue types (ENCODE database; BEC: bronchial epithelial cell, fibro.: fibroblast, epith.: epithelial). Shading indicates location of the most significant GWAS SNPs ('top SNPs'; orange) or monocyte/macrophage-specific active chromatin regions ('ACR1-3'; in yellow). (**c**) Circos plots showing significant chromatin interactions with the *CCR1* promoter (green dashed line) in monocytes, CD4+ T cells and total B cells measured by promoter-capture Hi-C (BLUEPRINT database; freq.: frequency). ACRs are indicated in orange. (**d**) Significant chromatin interactions with the *CCR5* promoter (green dashed line) in inflammatory macrophages (M1 Mph) or CD4+ T cells. (**e**) Lack of chromatin interactions emanating from the location of the most significant GWAS SNPs (indicated in panel b by orange shading; 'top SNPs'). The circus plot for inflammatory macrophages (M1 Mph) is depicted, but identical results were obtained for the indicated cell types. (**f**) Normalized gene expression levels (transcripts per million; TPM) for indicated 3p21.31 genes in *in vitro* generated ('induced') macrophages[15]. (**g**) UCSC genome browser view of H3K27Ac and CTCF ChIP-Seq signal for the indicated cell types.

Supplementary Figure 2

Supplementary Figure 2. Transcription factor binding at 3p21.31 COVID-19 risk variants. (**a**) Transcription factor binding motifs overlapping with four candidate causal SNPs. Single nucleotide substitutions are indicated above the motif logo. (**b**) UCSC genome browser view of the *CCR1* gene (located inside an intron of *CCR3*) showing H3K27Ac ChIP-Seq and RNA-Seq data tracks for monocytes and GM12878 lymphoblastoid cells (ENCODE database). (**c**) UCSC genome browser view of the mouse *Ccr1* gene. Locations of ChIP-Seq peaks for H3K27Ac and IRF8 detected in unstimulated or lipopolysaccharide (LPS) stimulated bone marrow-derived macrophages are indicated by colored rectangles (data obtained from ref[18]).

Methods

GWAS data retrieval
Version 4 COVID-19 GWAS meta-analysis data was retrieved from The COVID-19 Host Genetics Initiative at https://www.covid19hg.org/. GWAS data (GRCh37/hg38 genome build) was obtained from two studies: B1_ALL (Hospitalized COVID-19 vs. non-hospitalized COVID-19; 2430 cases versus 8478 controls) and B2_ALL (Hospitalized COVID-19 vs. population; 8638 cases versus 1736547 controls). GWAS summary statistics files were used to generate input files for FUMA using standard data table processing functions in Rstudio v.1.3.1093.

Functional Mapping and Annotation (FUMA) analysis
FUMA[7] was performed for both B1_ALL and B2_ALL GWASs using default settings, with exception of the r^2 used to define independent significant SNPs, which was set to ≥0.8. Manhattan and regional plots were generated by FUMA's SNP2GENE function; heatmaps visualizing tissue-specific gene expression data (GTEx v8) were obtained through FUMA's GENE2FUNCTION function. The Combined Annotation Dependent Depletion (CADD) and RegulomeDB scores for all significant SNPs were obtained from FUMA. As thresholds to identify candidate causal variants, we used CADD score >14 and RegulomeDB score <3. SNPs were further filtered based on their combined overlap with H3K27Ac ChIP-Seq (active chromatin mark) and DNAse-seq (chromatin accessibility for regulatory factor recruitment) peaks (data obtained from ENCODE[24]).

Epigenome and transcriptome analysis
ChIP-Seq, RNA-Seq and DNAse-Seq epigenomic data used were retrieved from public ENCODE[24], BLUEPRINT[11] and DICE[12] databases. ENCODE and BLUEPRINT data were visualized in the UCSC Genome Browser (https://genome.ucsc.edu). Significant FUMA SNPs were converted to GRCh38/hg38 using UCSC LiftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver) to allow aligning variants to the epigenomic profiles. Circos plots visualizing promoter-capture HiC data from the BLUEPRINT consortium were generated using https://www.chicp.org/chicp/, with a threshold normalized interaction value of 7. Transcription factor binding motifs were obtained using HOMER[25]. Immune cell-specific gene expression data was obtained using the DICE online platform (https://dice-database.org/) and visualized as averaged values (from n=81-91 individuals) using Morpheus (https://software.broadinstitute.org/morpheus/). RNA-Seq, ChIP-Seq and in-situ Hi-C from *in vitro* generated ('induced') macrophages (prior to and after auxin-inducible CTCF degradation) were obtained from GSE140528 and analyzed as previously described[15].

Expression quantitative trait locus (eQTL) analysis
eQTL analysis was performed using data obtained from the GTEx v8 database at https://gtexportal.org/home/, searching by variant for statistically significant (P<0.005) tissue eQTLs. The number of tissue eQTLs for *CCR1*/*CCR3* associated with

rs3181080/rs34919616 and their normalized effect sizes (NES values, denoting the direction of the eQTL effect) were determined using the multi-tissue eQTL plot function.