# DATA SUPPLEMENT

**SUPPLEMENTAL METHODS**

**Primary Exome Data Analysis**

Exome base call data was converted to FASTQ using bcl2fastq v2.20.0.422. Reads were aligned using BWA-MEM v0.7.1553 to the reference GRCh37 (hs37d5.fa), separate read groups were assigned for all reads from one lane, and duplicates were masked using Samblaster v0.1.2454. Standard QC was performed using FastQC55. Variants were then called using GATK HaplotypeCaller v3.7.22

**Secondary Exome Data Analysis: HLA Typing**

Exome data was analysed using OptiType v1.3.556.

**Secondary Exome Data Analysis: KIR2DS4f Typing**

Unlike HLA, no ready-made bioinformatics tool is available for accurately haplotyping the genes in the KIR family, only KPI57, which requires whole-genome sequencing data. We thus limited the analysis of the functional variant of KIR2DS458 being present (KIR2DS4f+) in the data as follows: The non-functional variant is the major allele of KIR2DS4 that is present in the human reference GRCh37. KIR2DS4f positive samples thus show an insertion of CCCGGAGCTCCTATGACATGTA in exon 4 of KIR2DS4 (dbSNP rs138504928). Therefore, we screened the variant-calling results for this variant being present in the data and used the genotype called by GATK HaplotypeCaller23 for subsequent analysis (0/0 = KIR2DS4f negative, 0/1 = KIR2DS4f het. positive, 1/1 = KIR2DS4f hom. positive).

**SUPPLEMENTAL FIGURES**

**Supplemental Figure 1.** Comparison of frequencies in data sets, DS 1 - 3. Blue dots indicate alleles which are found in both data sets; grey dots correspond to alleles found in only one data set.

Chart, scatter chart

Description automatically generated

**Supplemental Figure 2.** Comparison of frequencies in data sets, DS 1 - 3 with the population wide frequencies. Blue dots indicate alleles which are found in both data sets; grey dots correspond to alleles found in only one data set.

Chart, scatter chart

Description automatically generated

**Supplemental Figure 3.** First, we used a t-test to test whether any of the first four principal components (PCs) was significantly associated with either intensive care unit (ICU) or intubation. This was not the case (p < 0.05). Next, we removed the outliers based on the IQR method as implemented in the boxplot.stats function in R and re-calculated the PCA. Then, we repeated the meta-analysis using PC1 and PC2 as covariates.

Principal component analysis was based on SNP genotyping data. Each dot is a single patient. Left panel: Colors indicate patients who carry the HLA-C\*04:01 allele. Right panel for data set 1 (DS1)only: Colors indicate self-reported geographical origin. Top row shows SNP typing results for data set 1 (DS1), bottom row shows SNP typing results for data set 2 (DS2). Symbol shapes indicate samples which were removed from subsequent analysis.

Chart, scatter chart

Description automatically generated

**Supplemental Figure 4.**  Principal component analysis based on SNP genotyping data after removing the outliers (2nd round). Each dot is a single patient. Left panel: Colors indicate patients who carry the HLA-C\*04:01 allele. Right panel (for DS1 only): Colors indicate self-reported geographical origin. Top row shows SNP typing results for data set 1 (DS1), bottom row shows SNP typing results for data set 2 (DS2).

Chart, scatter chart

Description automatically generated

**Supplemental Figure 5.** Comparison of the meta-log-OR calculated in the original meta-analysis (x-axis) and the meta-analysis including only individuals with sequenced exomes, with outliers removed from data set 1 (DS1, 8 outliers, 42 remaining) and data set (DS2, 10 outliers, 93 remaining). Colors indicate whether the effects have the same direction in both analyses (green) or whether the directions are opposite (red).

Diagram

Description automatically generated

**Supplemental Figure 6.** Population stratification in data set 1 (DS1): Association between alleles and intubation / ICU status stratified by reported country of origin in DS1. Top row shows absolute numbers, bottom row shows frequencies.

Chart, bar chart

Description automatically generated

**Supplemental Figure 7.** Association between alleles and intubation / ICU status stratified by reported ethnicity in data set, DS3. Top row shows absolute numbers, bottom row shows frequencies.

DS3 was exceptional among the analysed data sets as there were several large ethnicity groups present, including Hispanic, Caucasian, African American and other. Given that the allele HLA-C\*04:01 is present in these groups at different frequencies, and the incidence and severity of COVID-19 differs between these ethnicities, we sought to demonstrate that the observed association between COVID-19 severity and HLA-C\*04:01 in DS3 is not a result of coincidence. While we have included ethnicity as a covariate in the association test, we have also inspected the association separately for each group. Due to small sample sizes, none of the associations is significant in groups other than ethnicity Caucasian.

Chart, bar chart

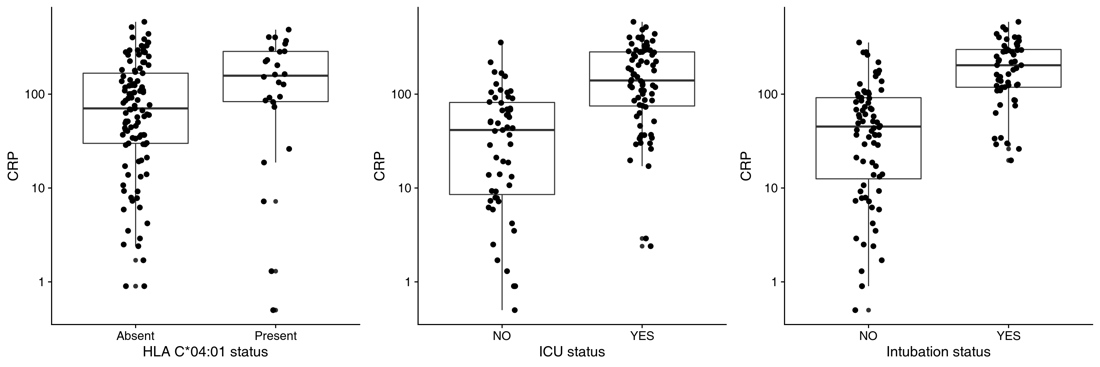
Description automatically generated

**Supplemental Figure 8**.

Middle line denotes the median; box defines the  
interquartile range; whiskers show the range excluding the outliers  
defined as data points more than 1.5 times the interquartile range away  
from the box.

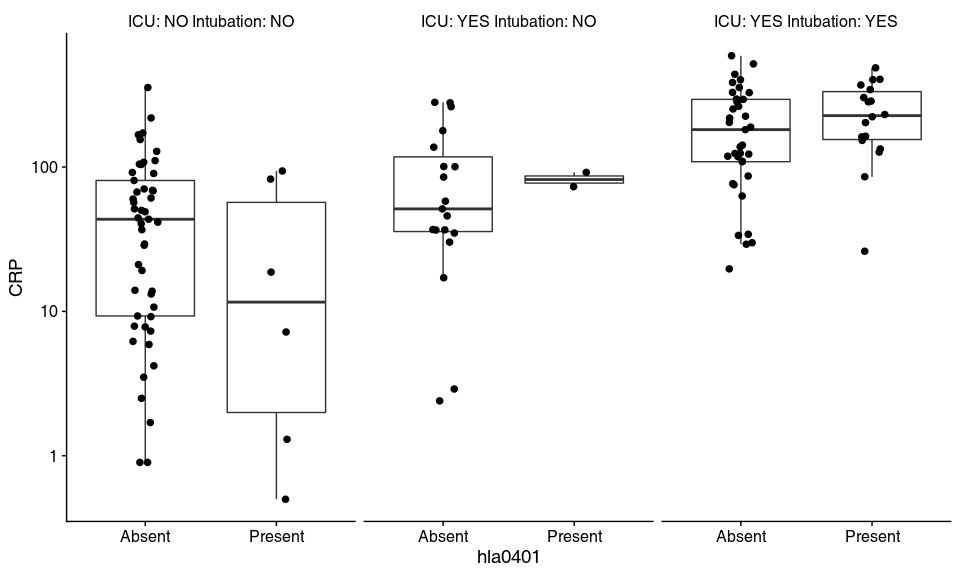
**a)**

Association of CRP levels with the presence of the HLA-C\*04:01 allele (left), intensive care unit (ICU) status (middle) and intubation status (right) in DS1. CRP was significantly associated with each of the three outcome variables (Wilcoxon rank sum test).



**b)**

Association between the presence of the HLA C\*04:01 allele stratified by ICU / intubation status.



**SUPPLEMENTAL TABLES**

**Supplemental Table 1.** Summary statistics for different cohorts. For numeric variables (BMI and age), the minimum, first quartile, median, third quartile and maximum are given.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subgroup | Number | BMI | BMI (NA) | Gender | Age (years) | ICU | ICU (NA) | Intubation | Intubation (NA) | Data set |
| Germany | 135 | 20.0 [ 23.3 < 27.4> 30.9] 356.4 | 54 | M: 90, F: 45 | 21 [48 <60> 71] 90 | YES: 77, NO: 58 | 0 | NO: 79, YES: 56 | 0 | DS1 |
| Granada | 18 | All values missing | 18 | M: 11, F: 7 | 30 [47 <56> 66] 91 | Only one value: NO | 0 | Only one value: NO | 0 | DS2 |
| Sevilla | 115 | All values missing | 115 | F: 71, M: 44 | 21 [42 <56> 65] 97 | NO: 111, YES: 4 | 0 | NO: 113, YES: 2 | 0 | DS2 |
| Switzerland | 20 | 22.1 [25.8 <28.0> 31.4] 45.0 | 0 | M: 16, F: 4 | 31 [51 <60> 75] 86 | NO: 11, YES: 9 | 0 | NO: 12, YES: 8 | 0 | DS2 |
| US | 147 | 16.2 [ 26.0 < 29.5> 33.7] 105.0 | 1 | M: 95, F: 52 | 21 [48 <60> 72] 87 | YES: 85, NO: 62 | 0 | NO: 86, YES: 61 | 0 | DS3 |
| Total | 435 | 16.2 [ 24.5 < 28.8> 32.4] 356.4 | 188 | M: 256, F: 179 | 21 [46 <58> 71] 97 | NO: 260, YES: 175 | 0 | NO: 308, YES: 127 | 0 | – |

**Supplemental Table 2.** Summary of number of alleles at different loci for data sets, DS 1-3 as well as number of alleles remaining after filtering for frequency > 0.05 and number of carriers > 5 in all data sets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Locus | DS1 | DS2 | DS3 | Filtered |
| DMA | – | – | 3 | 0 |
| DMB | – | – | 4 | 0 |
| DOA | – | – | 3 | 0 |
| DOB | – | – | 6 | 0 |
| DPA1 | 8 | 6 | 12 | 2 |
| DPB1 | 23 | 26 | 33 | 4 |
| DQA1 | 15 | 15 | 20 | 6 |
| DQB1 | 20 | 19 | 29 | 5 |
| DRA | – | – | 3 | 0 |
| DRB1 | 36 | 32 | 43 | 5 |
| DRB3 | 7 | 6 | 6 | 2 |
| DRB4 | 3 | 3 | – | 0 |
| DRB5 | 5 | 3 | 5 | 1 |
| E | – | – | 7 | 0 |
| F | – | – | 5 | 0 |
| H | – | – | 13 | 0 |
| HLA\_A | 30 | 27 | 43 | 4 |
| HLA\_B | 48 | 43 | 81 | 2 |
| HLA\_C | 30 | 30 | 44 | 4 |
| J | – | – | 2 | 0 |
| K | – | – | 4 | 0 |
| L | – | – | 3 | 0 |
| Total | 225 | 210 | 369 | 35 |

**Supplemental Table 3.** Top five results for the association of human leukocyte antigen (HLA) alleles with admission to the intensive care unit (ICU) and intubation with BMI as additional covariate. There was a significant association (adjusted p-value, p Adjusted) of HLA-C\*04:01 with intubation. OR 1-3, odds ratios (with 95% CI) in data sets 1-3; OR Meta-analysis, odds ratio from meta-analysis; RR, risk ratio with 95% CI; p-value, p-value from meta-analysis; p Adjusted, p-value corrected for multiple testing using Bonferroni correction.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| trait | Allele | OR  DS1 | OR  DS2 | OR  DS3 | OR Meta-analysis | RR | p-value | P Adjusted |
| icu | HLA\_C 04:01 | 3.6 [0.9 - 13.6] | 0.9 [0.2 - 3.6] | 3.1 [1.1 - 8.5] | 2.3 [1.2 - 4.8] | 1.2 [0.9 – 1.6] | 0.011 | 0.73 |
| ,, | DQB1 06:02 | 1.4 [0.4 - 4.6] | 7.6 [1.1 - 50.2] | 2.9 [0.7 - 11.9] | 2.5 [1.1 - 5.5] | 1.4 [1.0 – 1.9] | 0.03 | 1 |
| ,, | DQA1 01:02 | 1.1 [0.4 - 3.2] | 5.2 [1.5 - 18.4] | 0.6 [0.3 - 1.4] | 1.1 [0.7 - 2.0] | 1.1 [0.8 – 1.5] | 0.035 | 1 |
| ,, | HLA\_C 07:02 | 1.2 [0.4 - 3.3] | 2.3 [0.5 - 10.1] | 0.4 [0.2 - 1.0] | 0.8 [0.4 - 1.4] | 1.1 [0.8 – 1.5] | 0.041 | 1 |
| ,, | HLA\_B 07:02 | 0.7 [0.2 - 2.0] | 1.9 [0.4 - 8.5] | 0.3 [0.1 - 1.1] | 0.7 [0.3 - 1.3] | 0.8 [0.5 – 1.2] | 0.046 | 1 |
| intub | HLA\_C 04:01 | 8.5 [2.1 - 35.2] | 1.5 [0.3 - 6.9] | 3.4 [1.3 - 8.9] | 3.6 [1.8 - 7.2] | 1.5 [1.1 – 2.2] | 0.00036 | 0.025 |
| ,, | HLA\_B 07:02 | 0.4 [0.1 - 1.2] | 1.5 [0.3 - 8.6] | 0.3 [0.1 - 1.3] | 0.5 [0.2 - 1.0] | 0.6 [0.3 – 1.1] | 0.022 | 1 |
| ,, | DQA1 01:01 | 4.8 [1.4 - 17.1] | 0.0 [0.0 - Inf] | 1.6 [0.5 - 5.0] | 2.6 [1.1 - 6.1] | 1.1 [0.7 – 1.7] | 0.024 | 1 |
| ,, | DRB1 13:01 | 0.5 [0.0 - 5.8] | 1.6 [0.2 - 17.9] | 3.1 [1.1 - 9.3] | 2.2 [0.9 - 5.5] | 1.3 [0.8 – 2.1] | 0.035 | 1 |
| ,, | DQB1 03:02 | 1.5 [0.4 - 5.6] | 0.7 [0.1 - 6.3] | 2.7 [1.1 - 7.1] | 2.0 [1.0 - 4.1] | 1.7 [1.2 – 2.4] | 0.039 | 1 |

**Supplemental Table 4.** Top five results for the association of human leukocyte antigen (HLA) alleles with admission to the intensive care unit (ICU) and intubation after removing the Sevilla cohort. There was a significant association (adjusted p-value, p Adjusted) of HLA-C\*04:01 with intubation. OR 1-3, odds ratios (with 95% CI) in data sets 1-3; OR Meta-analysis, odds ratio from meta-analysis; RR, risk ratio with 95% CI; p-value, p-value from meta-analysis; p Adjusted, p-value corrected for multiple testing using Bonferroni correction.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| trait | Allele | OR  DS1 | OR  DS2 | OR  DS3 | OR Meta-analysis | RR | p-value | P Adjusted |
| icu | HLA\_C 04:01 | 3.4 [1.2 - 9.8] | 1.0 [0.1 - 8.0] | 3.2 [1.1 - 8.8] | 2.9 [1.4 - 5.8] | 1.3 [1.1 – 1.7] | 0.0028 | 0.12 |
| ,, | HLA\_C 07:01 | 0.4 [0.2 - 1.0] | 6.8 [0.6 - 80.1] | 1.2 [0.4 - 3.6] | 0.7 [0.4 - 1.3] | 0.9 [0.7 – 1.1] | 0.03 | 1 |
| ,, | DQA1 01:02 | 0.8 [0.3 - 1.7] | 6.6 [1.0 - 42.1] | 0.6 [0.3 - 1.3] | 0.8 [0.5 - 1.4] | 0.9 [0.7 – 1.2] | 0.058 | 1 |
| ,, | HLA\_C 07:02 | 1.0 [0.4 - 2.2] | 2.4 [0.2 - 23.8] | 0.4 [0.2 - 0.9] | 0.7 [0.4 - 1.2] | 0.9 [0.7 – 1.1] | 0.085 | 1 |
| ,, | DQA1 01:03 | 2.1 [0.7 - 5.7] | 1.0 [0.1 - 7.4] | 1.7 [0.7 - 4.0] | 1.7 [0.9 - 3.2] | 1.2 [1.0 – 1.5] | 0.091 | 1 |
| intub | HLA\_C 04:01 | 5.4 [1.9 - 15.1] | 1.5 [0.2 - 12.4] | 3.3 [1.3 - 8.5] | 3.8 [1.9 - 7.3] | 1.7 [1.3 – 2.2] | 8.3e-05 | 0.0035 |
| ,, | DQA1 01:02 | 0.7 [0.3 - 1.5] | 8.9 [1.3 - 58.7] | 0.5 [0.2 - 1.2] | 0.7 [0.4 - 1.3] | 0.9 [0.7 – 1.2] | 0.019 | 0.8 |
| ,, | HLA\_A 03:01 | 0.7 [0.3 - 1.8] | 2.8 [0.3 - 25.9] | 4.3 [1.2 - 14.9] | 1.4 [0.7 - 2.9] | 1.1 [0.8 – 1.6] | 0.035 | 1 |
| ,, | DQA1 01:01 | 2.7 [1.0 - 7.1] | 0.0 [0.0 - Inf] | 1.6 [0.5 - 4.7] | 2.1 [1.0 - 4.4] | 1.2 [0.9 – 1.7] | 0.039 | 1 |
| ,, | HLA\_C 07:01 | 0.6 [0.3 - 1.3] | 9.5 [0.8 - 118.3] | 1.8 [0.6 - 5.7] | 1.0 [0.5 - 1.9] | 0.9 [0.6 – 1.3] | 0.042 | 1 |

**Supplemental Table 5.** Results of the association test between SNP-derived PCs and ICU status, intubation status and presence of the HLA C\*04:01 allele. Table shows p-values from a two-sided t-test with Welch correction, corrected for multiple testing using Bonferroni correction.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **PC1** | **PC2** | **PC3** | **PC4** |
| **ICU** | 1 | 1 | 1 | 1 |
| **Intubation** | 1 | 1 | 1 | 0.52 |
| **HLA-C 04:01** | 1 | 1 | 1 | 1 |

**Supplemental Table 6.** Associations for viral load were tested in DS1 only. As response variables, viral load as peak viral load was chosen. Associations with p.value < 0.05 are shown below.

|  |  |  |
| --- | --- | --- |
| Allele | p.value | p\_adj |
| HLA\_A 01:01 | 0.0062 | 0.73 |
| HLA\_C 03:03 | 0.016 | 1 |
| HLA\_B 15:01 | 0.018 | 1 |
| HLA\_A 01:01 | 0.022 | 1 |
| DPB1 04:02 | 0.024 | 1 |
| DPB1 01:01 | 0.025 | 1 |
| HLA\_A 11:01 | 0.03 | 1 |
| DRB5 02:02 | 0.042 | 1 |

**Supplemental Table 7.** Top 25 HLA alleles with the lowest number of binding peptides. Tightly, number of tigthly binding peptides (binding affinity < 50 mM); Loosely, number of loosely binding peptides (binding affinity < 500 mM).

|  |  |  |
| --- | --- | --- |
| Allele | Tigthly | Loosely |
| HLA\_B 46:01 | 0 | 3 |
| HLA\_B 52:01 | 0 | 4 |
| HLA\_C 01:02 | 0 | 4 |
| HLA\_B 14:03 | 0 | 5 |
| HLA\_B 27:03 | 0 | 5 |
| HLA\_B 82:01 | 0 | 5 |
| HLA\_C 04:01 | 0 | 6 |
| HLA\_B 58:02 | 0 | 7 |
| HLA\_B 15:10 | 0 | 9 |
| HLA\_B 51:08 | 0 | 11 |
| HLA\_B 37:01 | 0 | 27 |
| HLA\_B 13:02 | 0 | 28 |
| HLA\_B 44:05 | 0 | 30 |
| HLA\_B 78:01 | 0 | 37 |
| HLA\_A 74:01 | 0 | 79 |
| HLA\_B 14:01 | 1 | 14 |
| HLA\_B 14:02 | 1 | 14 |
| HLA\_B 48:01 | 1 | 19 |
| HLA\_A 25:01 | 1 | 20 |
| HLA\_B 40:12 | 1 | 27 |
| HLA\_B 48:03 | 1 | 27 |
| HLA\_B 44:10 | 1 | 32 |
| HLA\_B 51:01 | 1 | 34 |
| HLA\_B 38:02 | 1 | 36 |
| HLA\_B 27:02 | 1 | 39 |

**Supplemental Table 8.** Association between HLA C\*04:01 status and diverse clinical variables. For categorical variables (sex, diabetes, hypertension, death outcome), Fisher’s exact test was used. For continuous variables (age, weight, BMI, CRP), Wilcoxon’s test was used. Values are p-values, not corrected for multiple testing.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Type | DS1 | DS2 | DS3 |
| age | Continuous | 0.68 | 0.77 | 0.97 |
| weight | Continuous | 0.44 | 0.11 | NA |
| BMI | Continuous | 0.93 | 0.31 | 0.58 |
| CRP | Continuous | 0.021 | 0.62 | NA |
| sex | Categorical | 0.49 | 0.58 | 1 |
| Diabetes | Categorical | 0.34 | 1 | 0.34 |
| Hypertension | Categorical | 0.51 | 0.79 | NA |
| Death | Categorical | 0.74 | 0.56 | 1 |