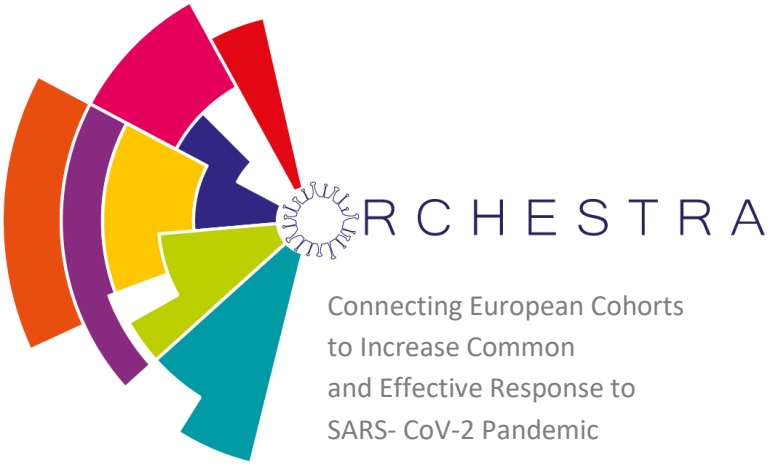

WP6 D6.12

Report on dataset of genome-wide methylation patterns

Helmholtz Zentrum München (HMGU)

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Project Classification

Project Acronym:	ORCHESTRA
Project Title:	Connecting European Cohorts to Increase Common and Effective Response to SARS- CoV-2 Pandemic
Coordinator:	UNIVR
Grant Agreement Number:	101016167
Funding Scheme:	Horizon 2020
Start:	1st December 2020
Duration:	48 months
Website:	www.orchestra-cohort.eu
Email:	info@orchestra.eu

Document Classification

WP No:	WP6
Deliverable No:	D6.12
Title:	Report on dataset of genome-wide methylation patterns
Lead Beneficiary:	HMGU
Other Involved Beneficiaries:	all WP6 members
Nature:	Report
Dissemination Level:	Public
Due Delivery Date:	March 2024 (M40)
Submission Date:	May, 2024 (M42)
Justification of delay:	Delay in data acquisition
Status:	Final
Version:	1.0
Author(s):	Pamela Matías-García, Melanie Waldenberger, Annette Peters

History of Changes

Version	Date	Created/Modified by
0.1	04.04.2024	Pamela Matías-García, Melanie Waldenberger
0.2	29.04.2024	Pamela Matías-García, Melanie Waldenberger
1.0	14.05.2024	Pamela Matías-García, Melanie Waldenberger

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Table of contents

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Summary of partners and cohorts

PARTNER	COUNTRY	COHORT	PRINCIPAL INVESTIGATOR
SAS	Spain	SAS	Jesús RODRIGUEZ BAÑO
LIH	Luxembourg	CON-VINCE	Rejko KRÜGER
HMGU	Germany	NAKO	Annette PETERS
UMCG	Netherlands	LIFELINES COVID-19	Lude FRANKE
Tirschenreuther Study (External partner)	Germany	TiKoCo-19	Iris HEID
NAPKON (External partner)	Germany	NAPKON	Gabriele ANTON

List of Figures

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

List of Tables

Executive summary

Genome-wide methylation profiling on patient and population-based cohorts, as described in Task 6.7, has been performed to assess acquired epigenomic regulatory features through COVID-19 infection and/or vaccination. The Illumina EPIC chip technology has been used to measure DNA methylation on more than 850 000 CpG sites in 1488 samples from selected cohorts from WP3 and WP4. The main objective of this document is to provide an overview of the data obtained from profiling DNA-methylation in COVID-19 positive and negative individuals from population- and hospital-based cohorts within ORCHESTRA, as well as to summarize the data acquisition and quality control processing conducted on these datasets.

Background

1. Participating cohorts

This deliverable includes analysis of samples collected through a cross-sectional case-control study set up for Task 3.3, which aims to investigate whether infection with SARS-CoV-2 has a persistent impact on the DNA methylation (DNAm) profile of individuals up to 4 months after infection and explore if these molecular changes are associated with the severity of disease. For this purpose, patient and population-based cohorts with individuals with varying degrees of disease severity and available blood and/or genomic DNA samples for DNAm profiling were identified: NAKO (HMGU) (Peters et al., 2022), CON-VINCE (LIH) (Snoeck et al., 2020), Lifelines COVID-19 (UMCG) (Mc Intyre et al., 2021), Tirschenreuther Study (Wagner et al., 2021) and NAPKON (Schons et al., 2022). Participants from these cohorts were selected if they met the following inclusion criteria:

i) answered a baseline survey,
ii) provided a whole blood sample during this survey,
iii) provided informed consent for (epi)genetic analyses,
iv) had (self-reported) data on COVID-19 testing within the last 4 months prior to the survey / blood draw,
v) had not received a COVID-19 vaccine, and
vi) were 18 years or older.

Cases were identified as individuals having positive serology results or positive COVID-19 test (PCR, rapid test, self-test), and controls were identified as individuals having negative serology results and negative self-reported COVID-19 test. Cases and controls were frequency-matched based on sex and age.

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

DNA methylation levels from individuals from fragile populations were assessed to explore whether there are epigenetic differences post-vaccination and if these epigenetic differences are associated with vaccination responsiveness, differences in blood cell composition and/or disease conditions. For this purpose, 143 whole blood samples collected from the SAS WP4 cohorts (haematological patients, oncological patients, solid organ transplant recipients, rheumatological patients) for longitudinal methylation profiling of pre- and post-vaccination samples of fragile populations have been processed and their derived data has undergone quality control.

2. DNAm profiling

Whole blood or DNA samples from the selected cohorts were shipped to HMGU, with details for each cohort described in the following subsections. Samples meeting basic quality measures were used for DNA-methylation (DNAm) analyses. For each sample from all included cohorts, 750 ng of genomic DNA was used for bisulfite conversion with the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA, USA).

Subsequent methylation analysis was performed on an Illumina (San Diego, CA, USA) iScan platform using the Infinium MethylationEPIC BeadChip according to standard protocols provided by Illumina. GenomeStudio software version 2011.1 with Methylation Module version 1.9.0 was used for initial quality control of assay performance and for generation of methylation data export files.

a. SAS

There were 90 individuals measured at several time points (at most three, each one occurring after one of the following stages of Covid-19 vaccination: 1st dose, 2nd dose, FU M3). Thus, SAS methylation data for a sample size of 223 samples was generated in one batch. Whole blood samples (n=223) were shipped to HMGU for genomic DNA extraction, which was then used as a starting point for the bisulfite conversion.

b. CONVINCE

Whole blood samples from 100 individuals were shipped to HMGU and genomic DNA extracted; genomic DNA from 100 samples was used as a starting point for the bisulfite conversion.

c. NAKO / NAPKON

Buffy coat samples from 214 NAKO participants were processed to extract genomic DNA; aliquots of genomic DNA were available for 341 NAKO participants and 242 NAPKON participants. Genomic DNA from 797 samples in total was used as a starting point for the bisulfite conversion.

d. LifeLines

Genomic DNA from 168 samples was shipped to HMGU and used as a starting point for the bisulfite conversion.

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

e. Tirschenreuther Study

Genomic DNA from 200 samples was shipped to HMGU and used as a starting point for the bisulfite conversion.

3. Data pre-processing and Quality Control (QC)

Further quality control and preprocessing of the data were performed in R v4.1.3 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/), with the package minfi v1.40.0 (Aryee et al., 2014) and following primarily the CPACOR pipeline (Lehne et al., 2015). Raw intensities were read into R (command read.metharray) and background corrected (bgcorrect.illumina). Of note, the cut-off for the detection rate in each study is determined individually, considering the results from the sex prediction and median intensity quality control steps. The goal is to optimize the sample size while maintaining data quality, which is why the acceptable range for this threshold typically falls between 80% and 95%.

a. SAS

The detection rate threshold, considering the autosomes only, was set to 95% and all samples passed that threshold (Figure 1).

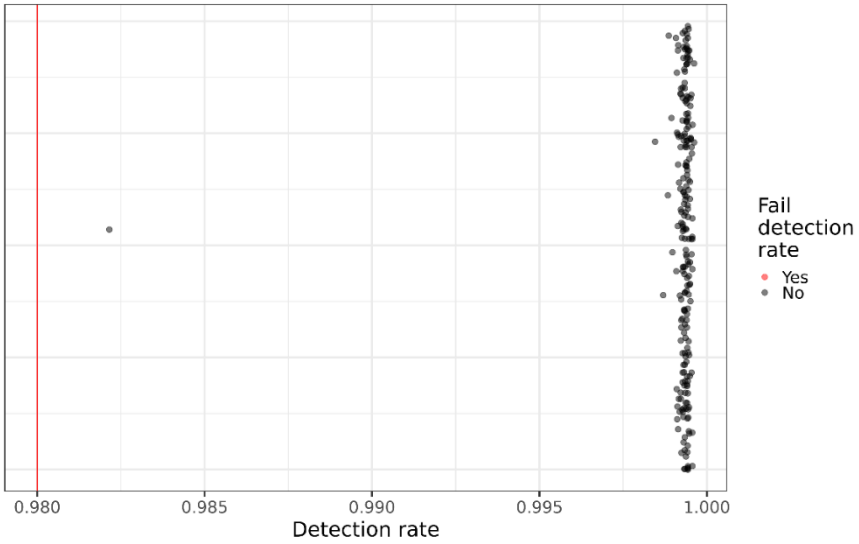

Figure 1. Detection rate p-value in SAS

A total of 57450 probes were removed (some overlapping multiple categories): cross-reactive probes as given in published lists (N=44 493, (McCartney et al., 2016; Pidsley et al., 2016)); probes with SNPs with minor allele frequency >5% at the CG position (N=11370) or the single base extension (N=5597) as given by minfi; and 3751 with >5% missing values (autosomes only) (Figure 2). A total of 808 409 probes remained for analysis.

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

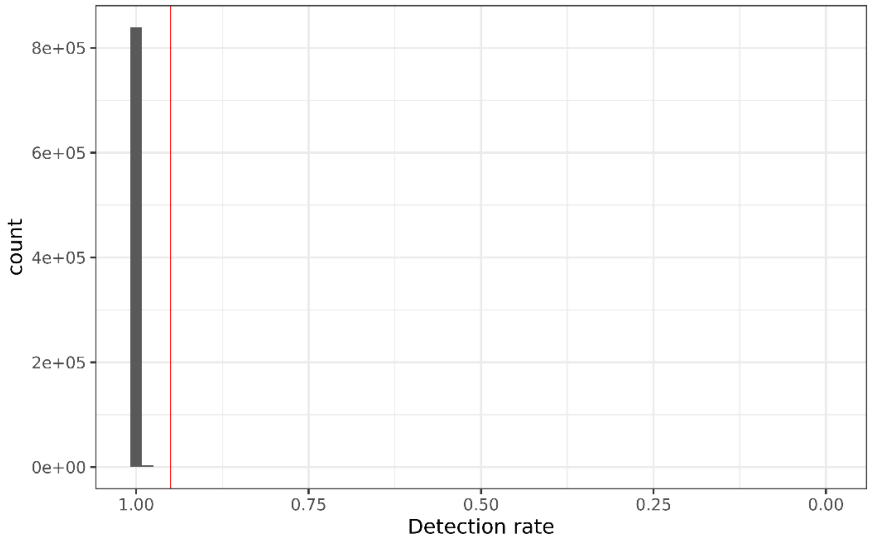

Figure 2. CpG call rate filter in SAS

No sample was removed prior to normalization, that is, all samples passed the sex prediction (Figure 3) and median intensity (Figure 4) quality control steps (implemented in minfi, commands getSex and getQC respectively), and had less than 5% missing values on the autosomes.

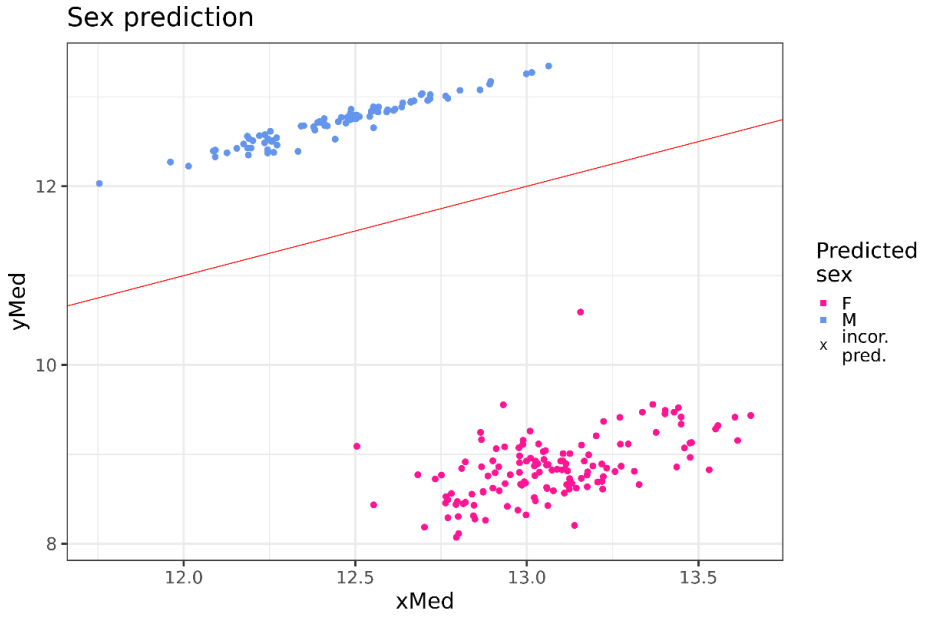

Figure 3. Sex prediction QC in SAS

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

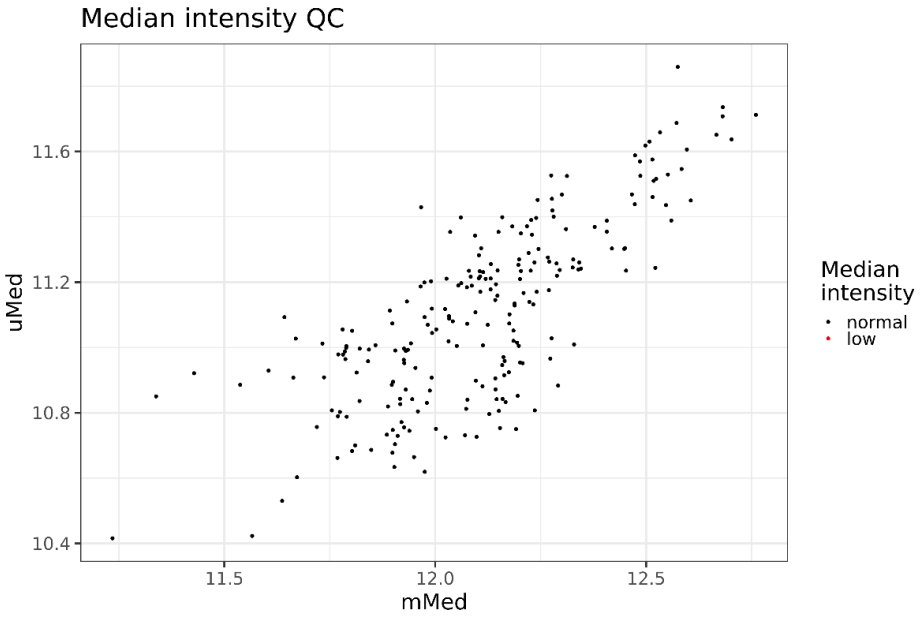

Figure 4. Median intensity QC step in SAS

b. CONVINCE

The detection rate threshold, considering the autosomes only, was set to 85% and all samples passed that threshold (Figure 5).

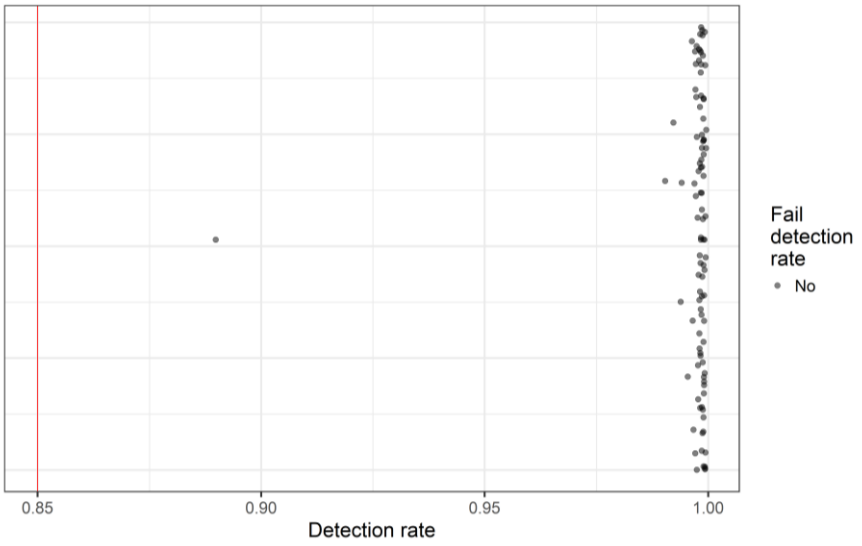

Figure 5. Detection rate p-value in CONVINCE

A total of 60 186 probes were removed (some overlapping multiple categories): cross-reactive probes as given in published lists (N=44 493, (McCartney et al., 2016; Pidsley et al., 2016)); probes with SNPs with minor allele frequency >5% at the CG position (N=11 370) or the single base extension (N=5597) as given by minfi; and 6228 with >5% missing values (autosomes only) (Figure 6). A total of 805 673 probes remained for analysis.

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

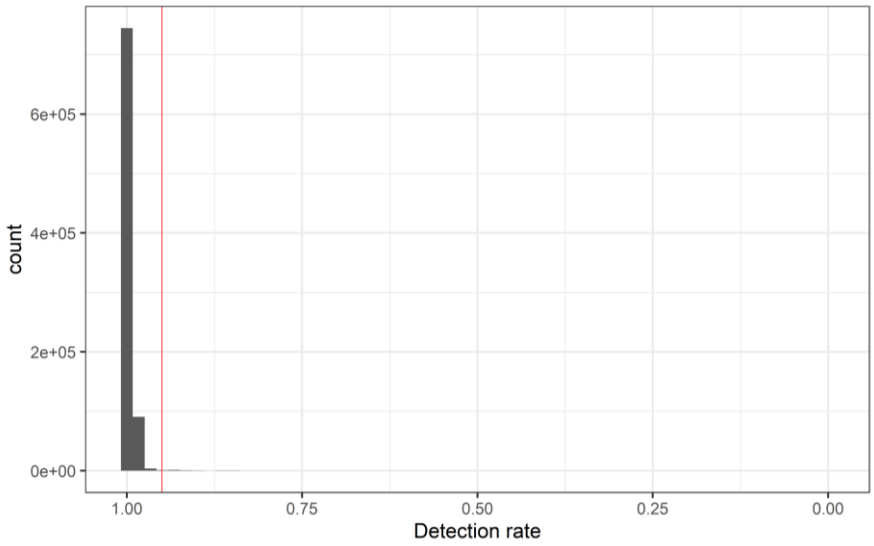

Figure 6. CpG call rate filter in CONVINCE

No sample was removed prior to normalization, that is, all samples passed the sex prediction (Figure 7) and median intensity (Figure 8) quality control steps (implemented in minfi, commands getSex and getQC respectively), and had less than 15% missing values on the autosomes. For additional details on the median intensity QC steps and one flagged sample, see the sample and probe summary in section Data Normalization - CONVINCE.

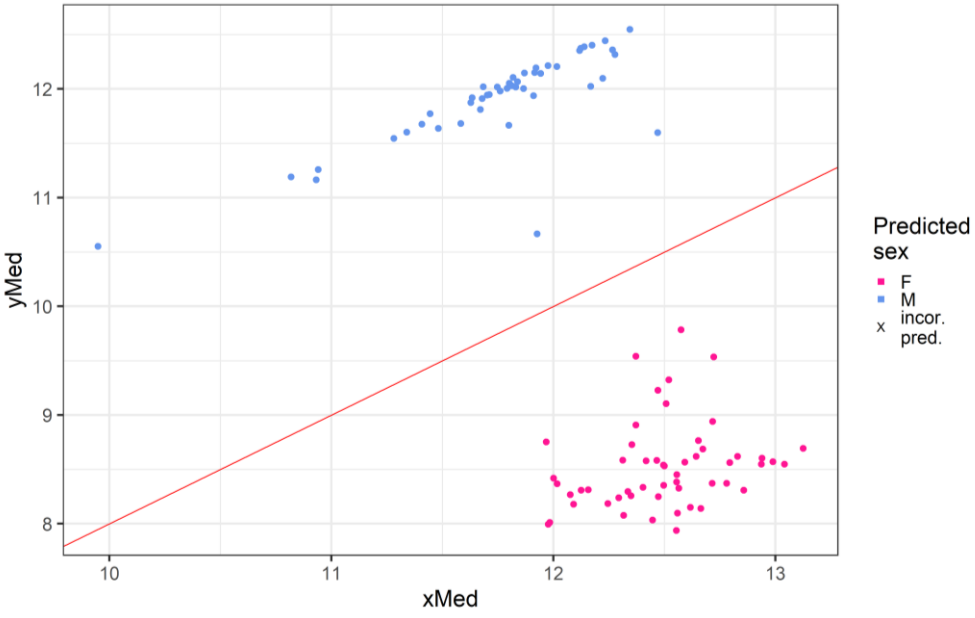

Figure 7. Sex prediction QC in CONVINCE

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

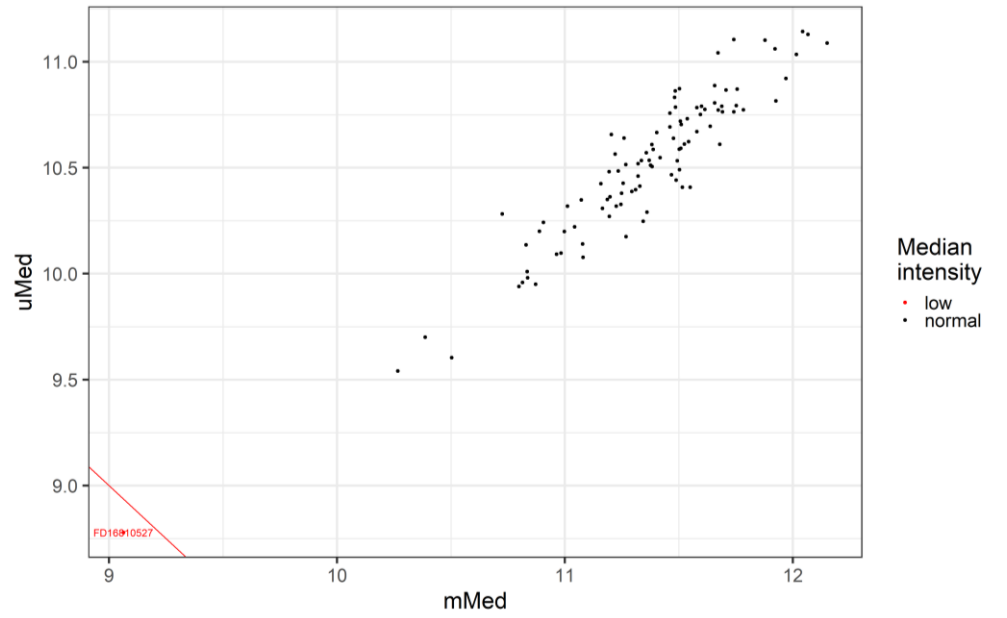

Figure 8. Median intensity QC step in CONVINCE

c. NAKO / NAPKON

The detection rate threshold, considering the autosomes only, was set to 90% and 12 samples fell below that threshold (Figure 9).

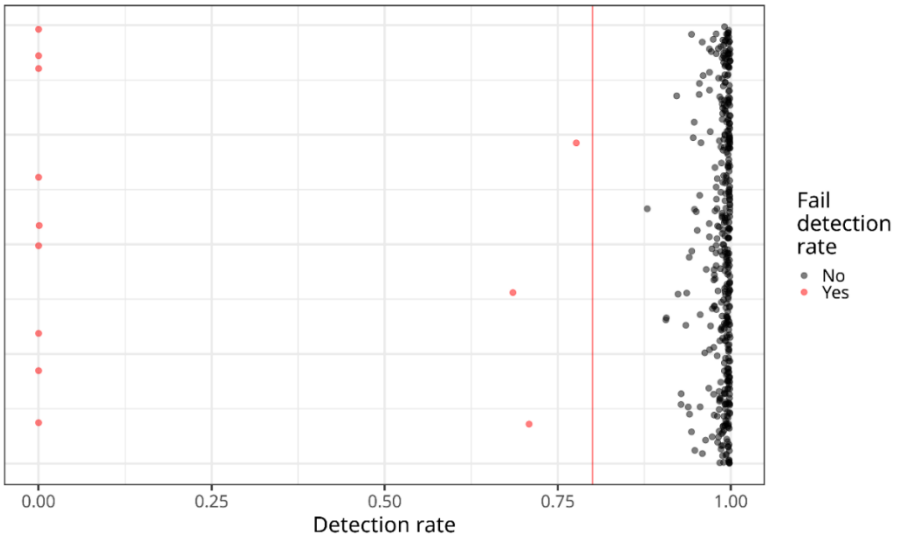

Figure 9. Detection rate p-value in NAKO + NAPKON

A total of 60 186 probes were removed (some overlapping multiple categories): cross-reactive probes as given in published lists (N=44 493, (McCartney et al., 2016; Pidsley et al., 2016)); probes with SNPs with minor allele frequency >5% at the CG position (N=11 370) or the single base extension (N=5597) as given by minfi; and 22 602 with >15% missing values (autosomes only) (Figure 10). A total of 789 299 probes remained for analysis.

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

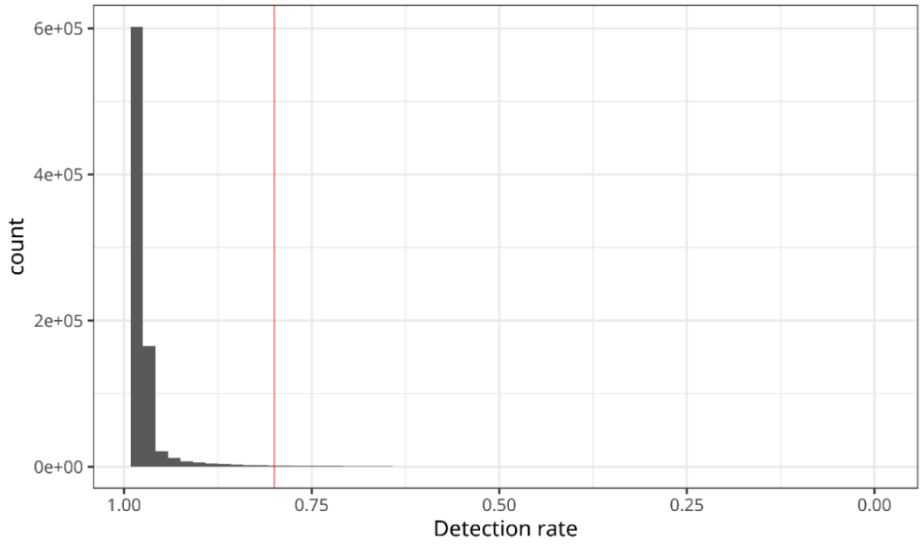

Figure 10. CpG call rate filter in NAKO + NAPKON

12 samples were removed prior to normalization: four for sex prediction failure (identified with crosses in Figure 11) and 12 for not passing median intensity quality control (Figure 12, red dots), as implemented in minfi (commands getSex and getQC respectively). The detection rate threshold (testing the autosomes only) was set to 80%, so that samples with greater than 20% missing values on the autosomes were removed (N=12, which all overlap with those failing the median intensity QC step).

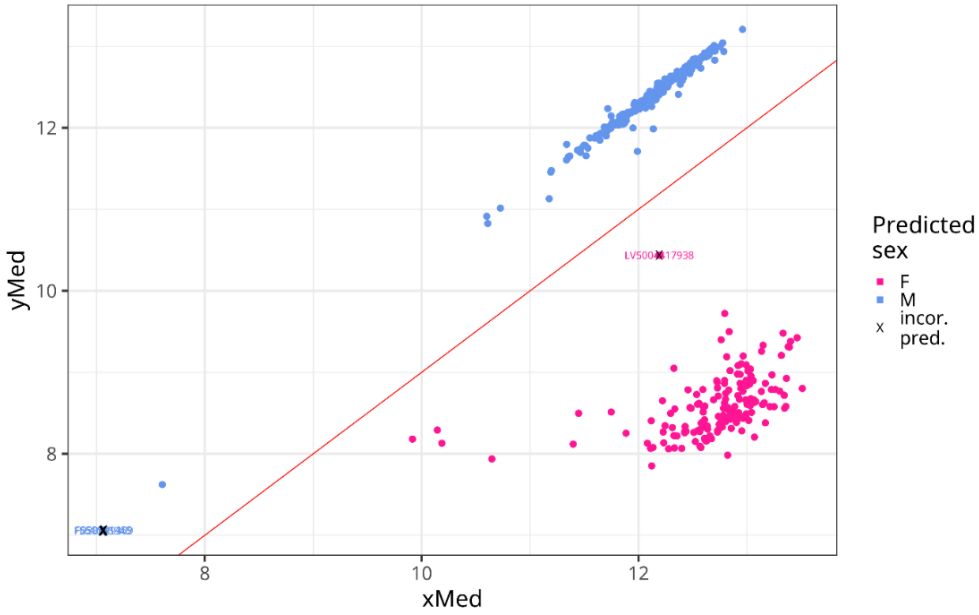

Figure 11. Sex prediction QC in NAKO + NAPKON

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

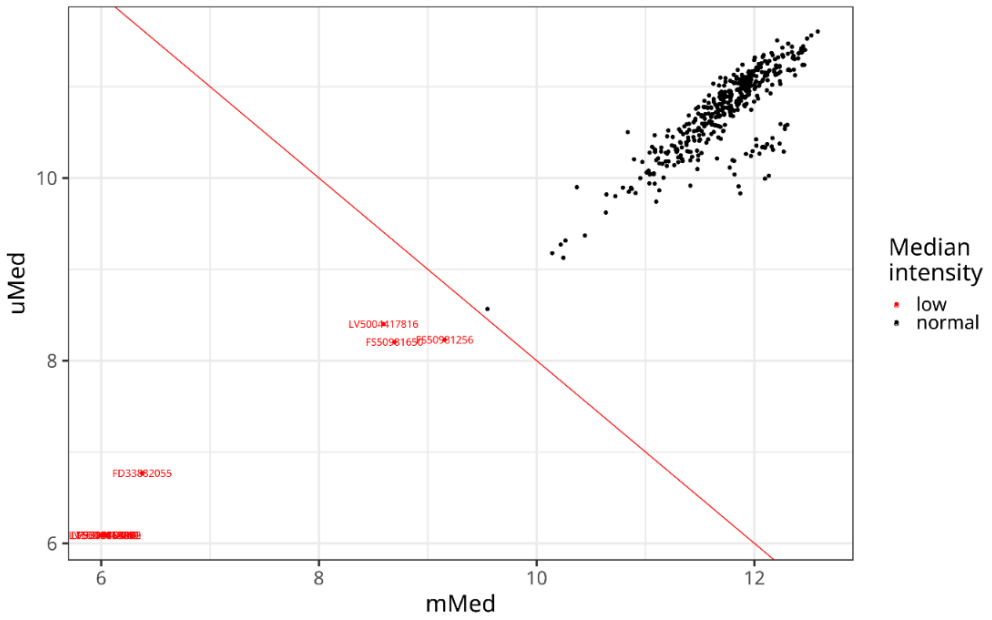

Figure 12. Median intensity QC step in NAKO + NAPKON

d. LifeLines

The detection rate threshold, considering the autosomes only, was set to 80% and 5 samples fell below that threshold (Figure 13).

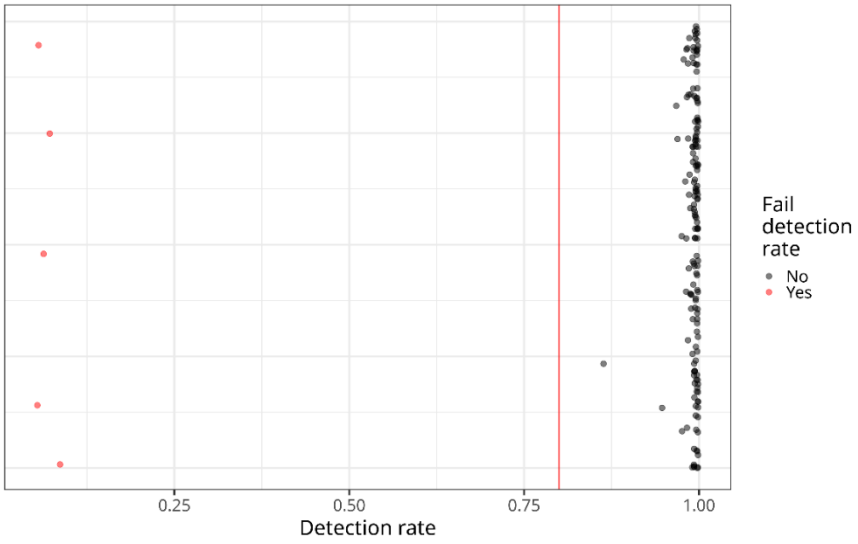

Figure 13. Detection rate p-value in Lifelines COVID-19

A total of 60 186 probes were removed (some overlapping multiple categories): cross-reactive probes as given in published lists (N=44 493, (McCartney et al., 2016; Pidsley et al., 2016)); probes with SNPs with minor allele frequency >5% at the CG position (N=11370) or the single base extension (N=5597) as given by minfi; and 18283 with >10% missing values (autosomes only) (Figure 14).

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

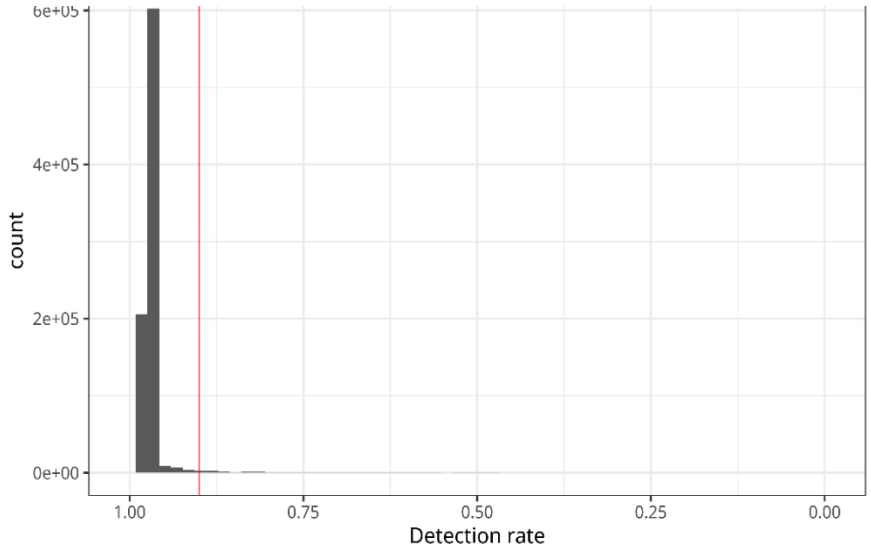

Figure 14. CpG call rate filter in Lifelines COVID-19

Five samples were removed prior to normalization: two for sex prediction failure (identified with crosses in Figure 15) and five for not passing median intensity quality control (Figure 16, red dots), as implemented in minfi (commands getSex and getQC respectively). The detection rate threshold (testing the autosomes only) was set to 80%, so that samples with greater than 20% missing values on the autosomes were removed (N=5, which all overlap with those failing the median intensity QC step).

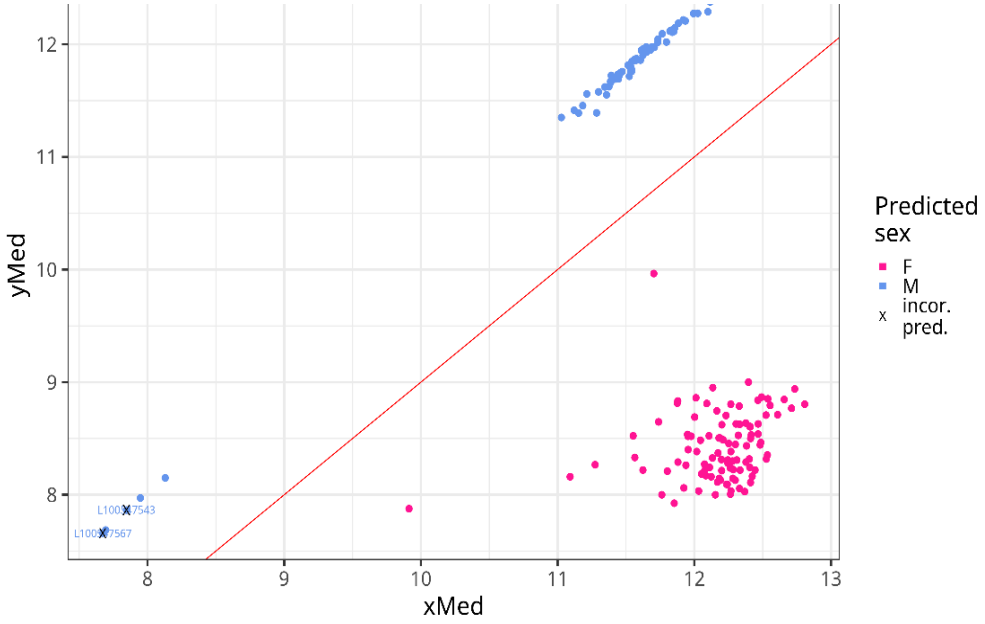

Figure 15. Sex prediction QC in Lifelines COVID-19

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

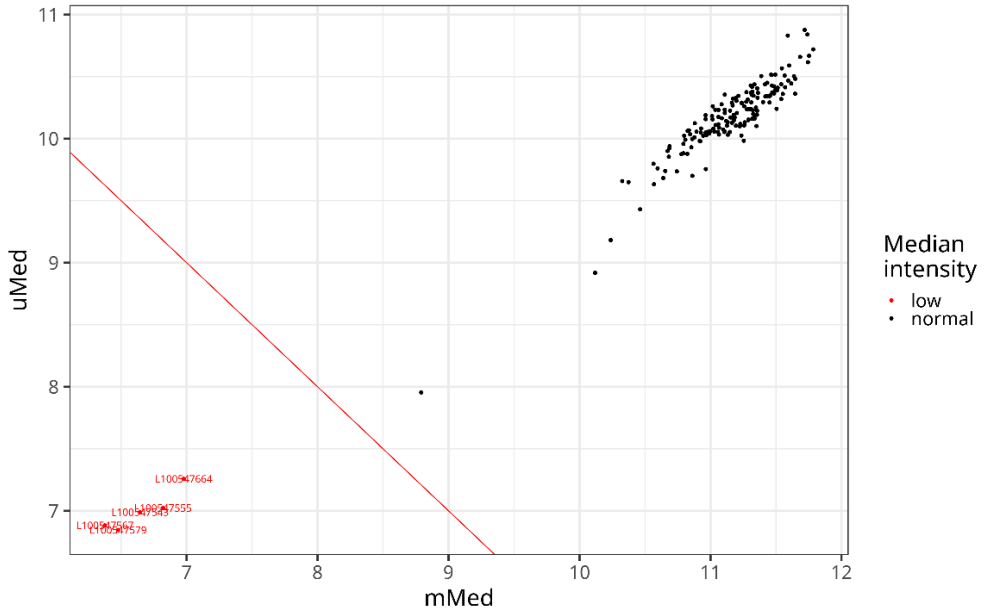

Figure 16. Median intensity QC step in Lifelines COVID-19

e. Tirschenreuther Study

The detection rate threshold, considering the autosomes only, was set to 95% and all samples passed that threshold (Figure 17).

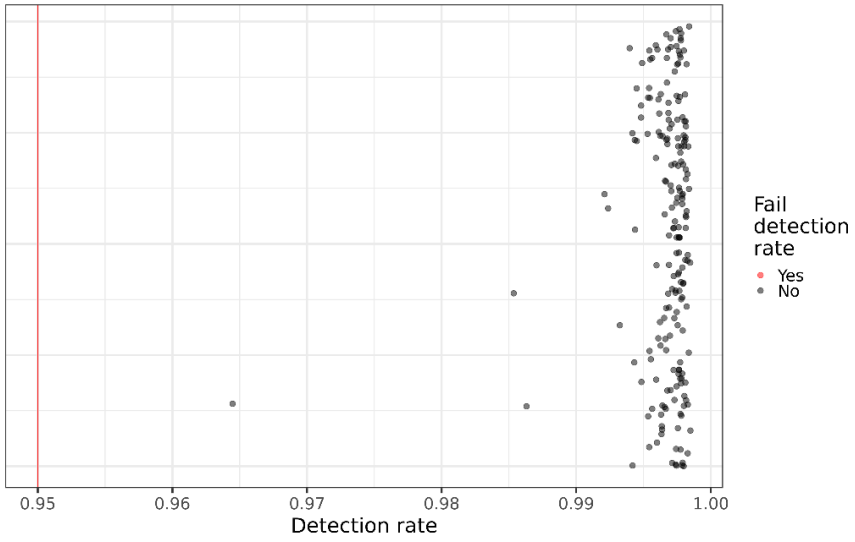

Figure 17. Detection rate p-value in TiKoCo-19

A total of 61 824 probes were removed (some overlapping multiple categories): cross-reactive probes as given in published lists (N=44 493, (McCartney et al., 2016; Pidsley et al., 2016)); probes with SNPs with minor allele frequency >5% at the CG position (N=11370) or the single base extension (N=5597) as given by minfi; and 8781 with >5% missing values (autosomes only) (Figure 18).

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

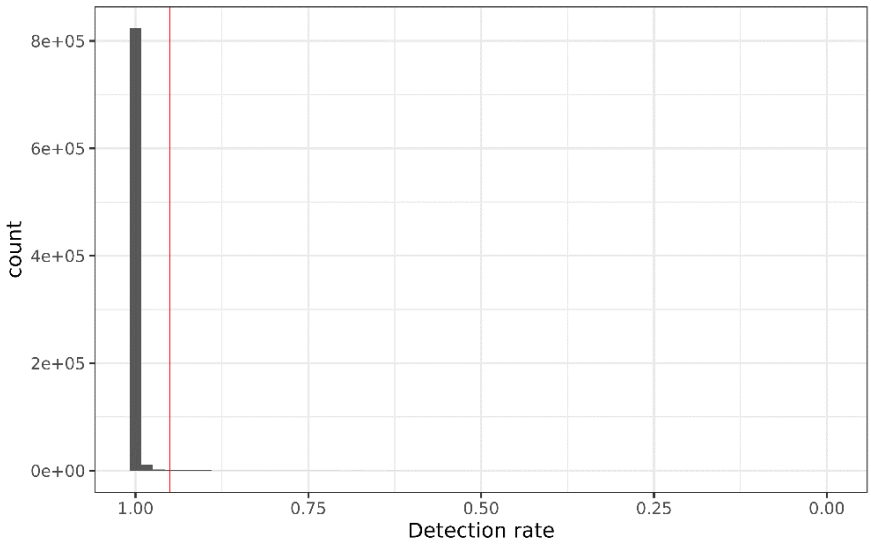

Figure 18. CpG call rate filter in TiKoCo-19

No sample was removed prior to normalization, that is, all samples passed the sex prediction (Figure 19) and median intensity (Figure 20) quality control steps (implemented in minfi, commands getSex and getQC respectively), and had less than 5% missing values on the autosomes.

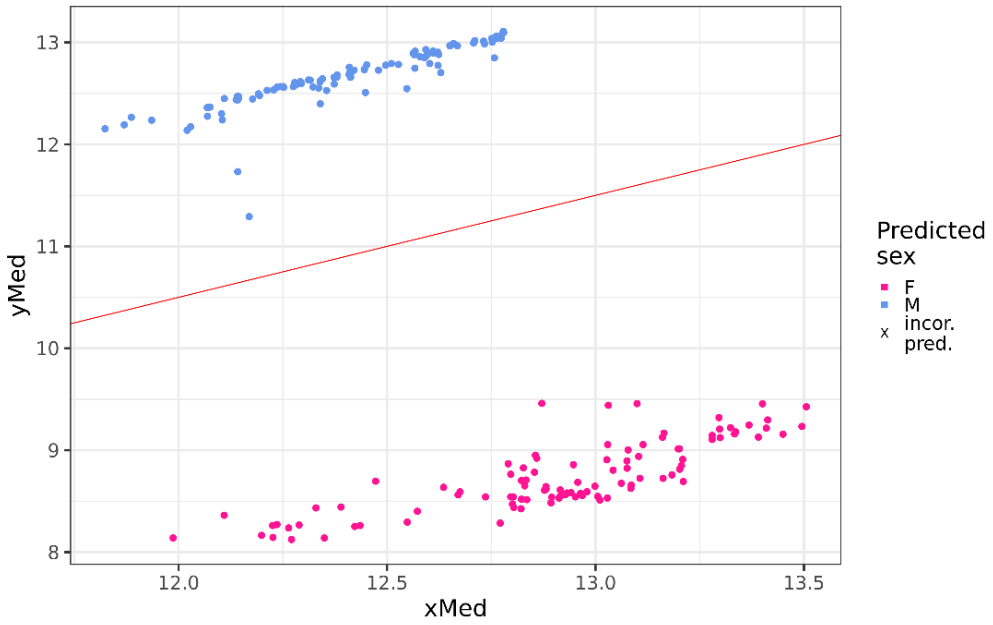

Figure 19. Sex prediction QC step in TiKoCo-19

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

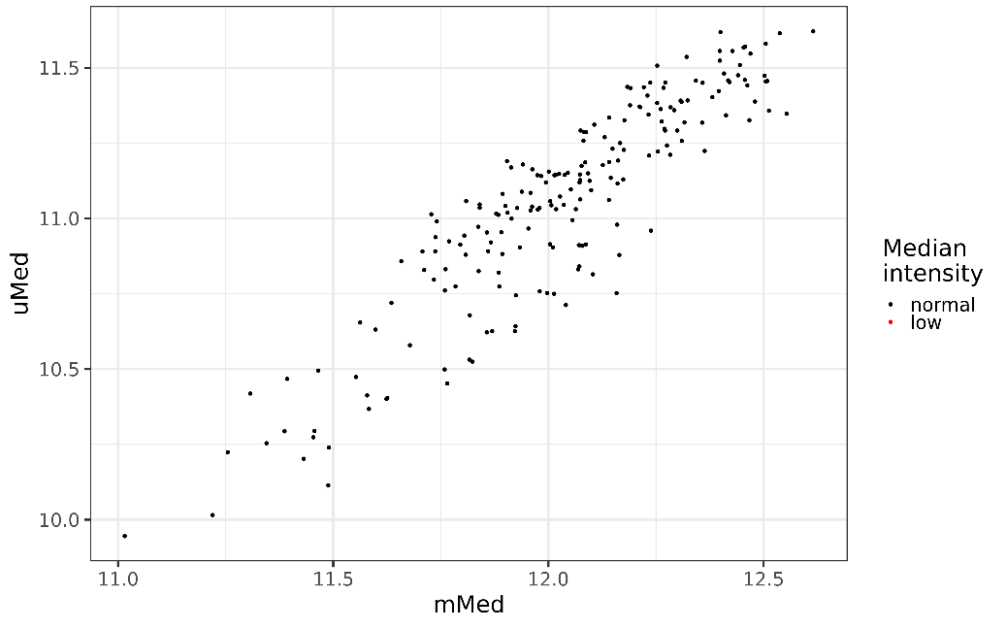

Figure 20. Median intensity QC step in TiKoCo-19

4. Data Normalization

Quantile normalization (QN, R package limma v3.50.3 (Ritchie et al., 2015)) was then performed separately on the signal intensities divided into the 6 probe types: type II red, type II green, type I green unmethylated, type I green methylated, type I red unmethylated, type I red methylated (Lehne et al., 2015). For the autosomes, QN was performed for all samples together; for the X and Y chromosomes, men and women were processed separately. The transformed intensities were then used to generate methylation beta values, a measure from 0 to 1 indicating the percentage of cells methylated at a given locus (Figure 21, Figure 22, Figure 23, Figure 24). Probes from the X chromosome (N=17743, following quality control) and the Y chromosome (N=379) are to be excluded from further analysis.

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

a. SAS

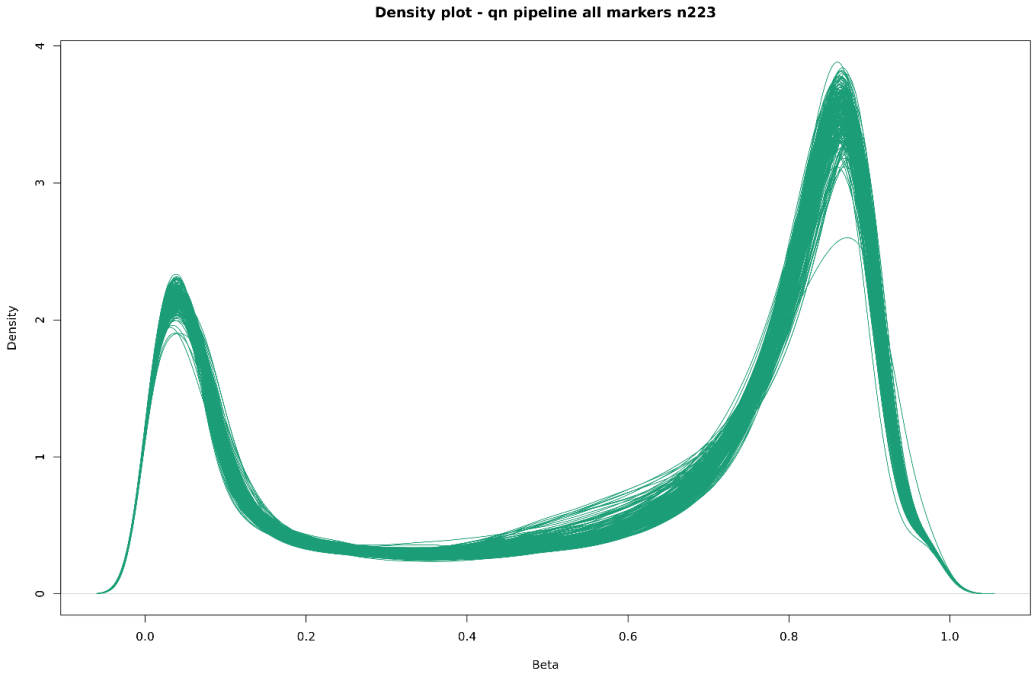

Figure 21. Density plot of SAS normalized DNAm data

Probe count summary:

The array contains 865 859 probes (846 232 on the autosomes, 19 090 on the X chromosome, 537 on the Y chromosome). 44 493 were cross-reactive probes, and 11 370 and 5597 had SNPs in the CG position and single base extensions respectively, and 3751 failed the detection p-value filter, a total of 60 000. However, many probes overlapped multiple categories: a total of 57 450 were removed. This leaves a total of 808 409 probes: 790 287 from the autosomes, 17 743 from the X chromosome, 379 from the Y chromosome.

Notes on probe count: the original EPIC array had 866 895 probes, of which 59 are rs SNP probes for quality control. A “Product Quality Notice” (Tracking Number: PQN0223) issued by Illumina on April 19, 2017 indicated that 977 probes were removed due to underperformance, hence the total of 865 859.

Sample count summary:

223 samples were measured, and all passed quality control according to the steps described above.

ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167

b. CONVINCE

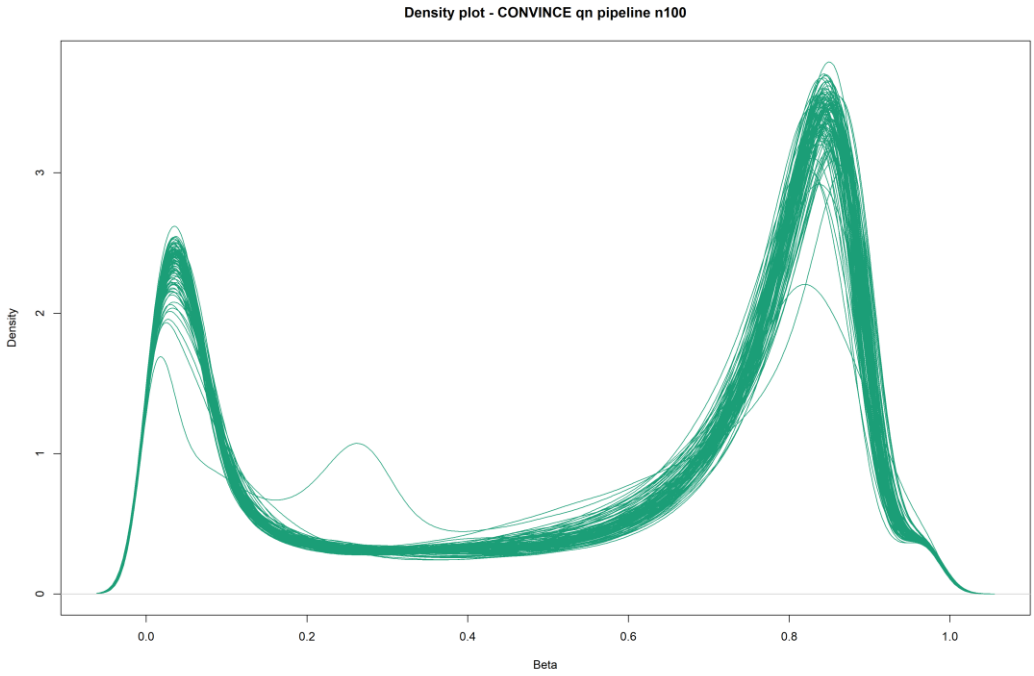

Figure 22. Density plot of CONVINCE normalized DNAm data

Probe count summary:

The array contains 865 859 probes (846 232 on the autosomes, 19 090 on the X chromosome, 537 on the Y chromosome). 44 493 were cross-reactive probes, and 11 370 and 5597 had SNPs in the CG position and single base extensions respectively, and 6228 failed the detection p-value filter, a total of 62 477. However, many probes overlapped multiple categories: a total of 60 186 were removed. This leaves a total of 805 673 probes: 787 551 from the autosomes, 17743 from the X chromosome, 379 from the Y chromosome.

Notes on probe count: the original EPIC array had 866 895 probes, of which 59 are rs SNP probes for quality control. A “Product Quality Notice” (Tracking Number: PQN0223) issued by Illumina on April 19, 2017 indicated that 977 probes were removed due to underperformance, hence the total of 865 859.

Sample count summary:

100 samples were measured, and all passed quality control according to the cut-offs described above.

Notes on sample count: One sample might be problematic for future analyses. Its median intensity is 8.920736 and its sample detection rate is 0.8898399. The later explains the chosen sample detection rate cut-off of 85%, all other samples of CONVINCE have a detection rate greater than 95%. Furthermore, the CpGs density for that sample (after normalization) shows 3 peaks, which clearly differs from all other samples of CONVINCE.

ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167

c. NAKO / NAPKON

Probe count summary:

The array contains 865 859 probes (846 232 on the autosomes, 19 090 on the X chromosome, 537 on the Y chromosome). 44 493 were cross-reactive probes, and 11 370 and 5597 had SNPs in the CG position and single base extensions respectively, and 18 283 failed the detection p-value filter, a total of 79 743. However, many probes overlapped multiple categories: a total of 87 760 were removed. This leaves a total of 789 299 probes.

Notes on probe count: the original EPIC array had 866 895 probes, of which 59 are rs SNP probes for quality control. A “Product Quality Notice” (Tracking Number: PQN0223) issued by Illumina on April 19, 2017 indicated that 977 probes were removed due to underperformance, hence the total of 865 859.

Sample count summary:

797 individuals were measured. Four were removed due to sex mismatch, 10 were removed due to failing quality control on the raw intensities and 10 failed the detection p-value filter (4 overlap with sex mismatch filter and complete overlap with intensity filter). A list with more information on the samples failing QC has been provided in “NAKO_NAPKON_samples_failing_qc.csv”.

d. LifeLines

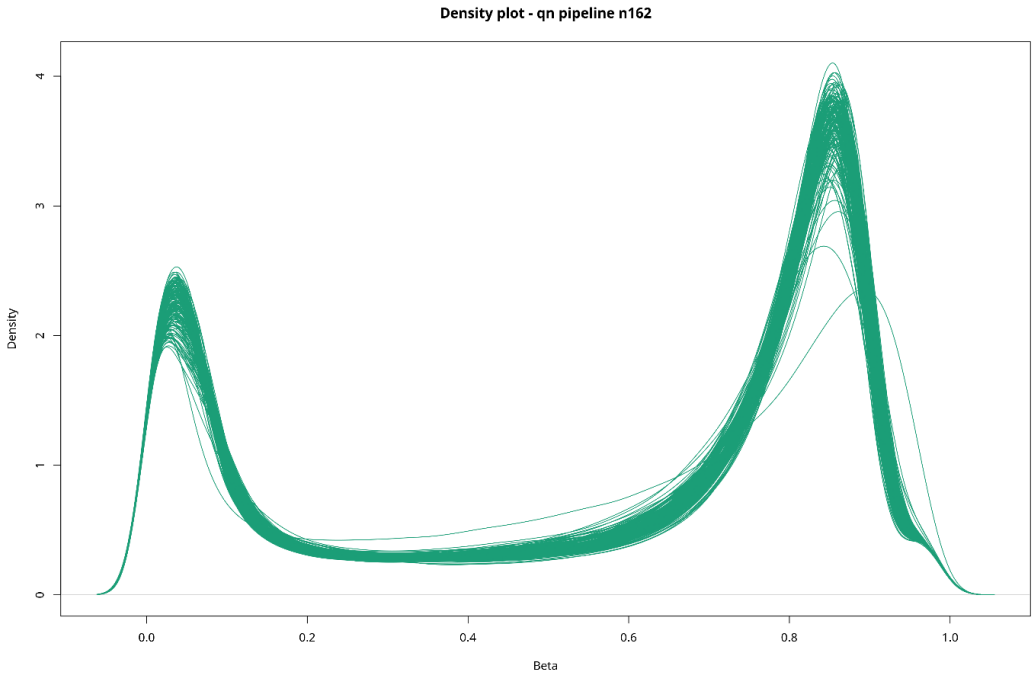

Figure 23. Density plot of Lifelines COVID-19 normalized DNAm data

Probe count summary:

The array contains 865 859 probes (846 232 on the autosomes, 19 090 on the X chromosome, 537 on the Y chromosome). 44 493 were cross-reactive probes, and 11 370 and 5597 had SNPs in the CG position and single base extensions respectively, and 18 283 failed the detection p-value filter, a total of 79 743. However, many probes overlapped

ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167

multiple categories: a total of 71 386 were removed. This leaves a total of 794 473 probes: 776 351 from the autosomes, 17 743 from the X chromosome, 379 from the Y chromosome.

Notes on probe count: the original EPIC array had 866 895 probes, of which 59 are rs SNP probes for quality control. A “Product Quality Notice” (Tracking Number: PQN0223) issued by Illumina on April 19, 2017 indicated that 977 probes were removed due to underperformance, hence the total of 865 859.

Sample count summary:

167 individuals were measured. Two were removed due to sex mismatch, 5 were removed due to failing quality control on the raw intensities and 5 failed the detection p-value filter (2 overlap with sex mismatch filter and complete overlap with intensity filter), leaving 162 individuals passing quality control. A list with more information on the samples failing QC has been provided in “LifeLines_samples_failing_qc.csv”.

Notes on sample count: One sample might be problematic for future analyses; this sample has been identified in the file “LifeLines_samples_failing_qc.csv”. The CpGs density for that sample (after normalization) shows two peaks shifted in comparison to all other samples from LifeLines. Further investigation of this sample is needed to determine whether the differences in its DNAm profile might be due to certain blood diseases and/or other chronic conditions affecting blood counts.

e. Tirschenreuther Study

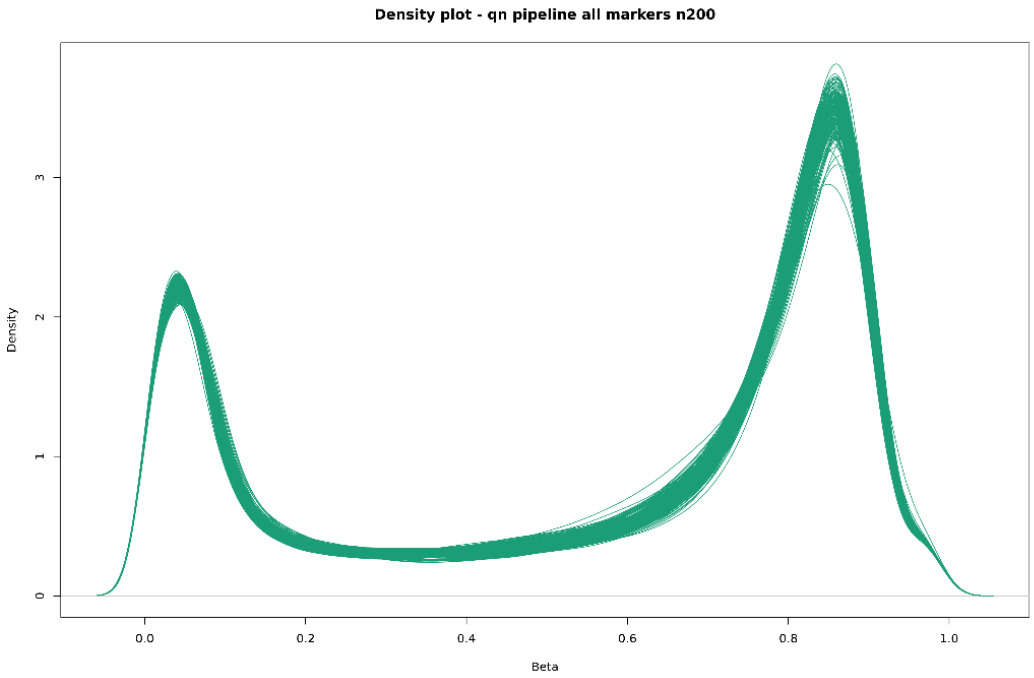

Figure 24. Density plot of TiKoCo-19 QN normalized DNAm data

Probe count summary:

The array contains 865 859 probes (846 232 on the autosomes, 19 090 on the X chromosome, 537 on the Y chromosome). 44 493 were cross-reactive probes, and 11 370 and 5 597 had SNPs in the CG position and single base extensions respectively, and 8781

ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167

failed the detection p-value filter, a total of 64 953. However, many probes overlapped multiple categories: a total of 61824 were removed. This leaves a total of 804 035 probes: 785913 from the autosomes, 17743 from the X chromosome, 379 from the Y chromosome.

Notes on probe count: the original EPIC array had 866 895 probes, of which 59 are rs SNP probes for quality control. A “Product Quality Notice” (Tracking Number: PQN0223) issued by Illumina on April 19, 2017 indicated that 977 probes were removed due to underperformance, hence the total of 865 859.

Sample count summary:

200 samples were measured, and all passed quality control according to the cut-offs described above.

5. Estimation of white blood cell type proportions

As methylation levels in blood can be strongly influenced by leukocyte composition, the white blood cell type proportions (WBC %) were calculated as per the method of Houseman et al (Houseman et al., 2012) using the command estimateCellCounts (minfi v1.46.0) on the raw intensities and using the default parameters. Two estimates were performed: a) using the default types ("CD8T", "CD4T", "NK", "Bcell", "Mono", ”Gran”); b) using "CD8T", "CD4T", "NK", "Bcell", "Mono", "Eos", "Neu".

To further explore the white blood cell deconvolution from bulk blood DNAm samples, an alternative method to obtain additional estimates of white blood cell counts using the R package FlowSorted.BloodExtended.EPIC and the function estimateCellCounts2() on the raw intensities and default parameters was done. This produces estimates for 12 cell types ("Bas", "Bmem", "Bnv", "CD4mem", "CD4nv", "CD8mem", "CD8nv", "Eos", "Mono", "Neu", "NK", "Treg") (Salas et al., 2022).

6. Results

Results from the other cohorts introduced in the participating cohorts section will be presented within the scope of WP3.

a. SAS: Population characteristics

Blood samples for 90 individuals from the fragile patient population at SAS were available for DNAm profiling. This cohort consisted of HIV positive subjects (HIV+), oncological and haematological patients, solid organ transplant recipients (SOT), hemodialysis patients and rheumatological patients.

Patients in this study were enrolled at 1st or 2nd dose within Task 4.5. In the subset of samples presented in these analyses, each individual had up to three samples taken, each one from one of the following timepoint of COVID-19 vaccination: 1st dose, 2nd dose, 3- and 6-month follow-up visits. Of note, in the first two time-points blood sampling was performed before receiving the vaccine, thus the sample from the visit to obtain the 1st vaccine was used as baseline and the sample from the visit for the 2nd vaccination considered to assess the effects of the first vaccine. A total of 221 samples were analyzed (Figure 25).

ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167

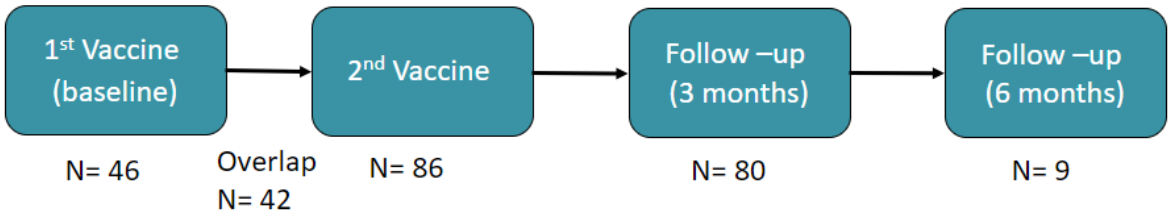

Figure 25. Overview of SAS study

b. Epigenome-wide association (EWAS) on vaccination

Statistical analysis, including the fitting of the Generalized Estimating Equations (GEE) model, was conducted using the statistical software R (R Core Team, 2021) for data analysis and visualization.

The association between DNA-methylation (DNAm) and vaccination was examined using generalized estimating equations (GEE) models with the function *geeglm* from the package "geepack" (version 1.3.10). DNAm levels were defined as the dependent variable and vaccination as the predictor variable. One model was run per CpG site: in this EWAS a total of 808 409 models were run. FDR correction was used to control the rate of false positives among all the significant results and to identify significant CpGs for downstream analyses, but the more conservative Bonferroni correction was also included as reference. Given both the exploratory and longitudinal nature of these analyses, associations with p FDR<0.05 were considered as statistically significant.

In model 1, the association between DNAm and vaccination was adjusted for age, sex and disease condition, using the patient ID to consider repeated DNAm measures per individual as shown next:

Model1: DNAm at CpG ~ vaccination + age + sex + condition code

Epigenome-wide results of this EWAS are displayed in Figure 26. The X-axis represents the genomic position of the epigenetic markers tested in the study arranged along the chromosomes in order of their physical position. The Y-axis represents the statistical significance of the associations between DNA methylation levels and vaccination status. Raw p-values obtained from the linear regression were transformed into -log10 scale for better visualization. As it can be seen in the Manhattan plot below, cg05447102 in chromosome 13 shows the strongest association with having received a first vaccine in the studied population (dot in red).

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

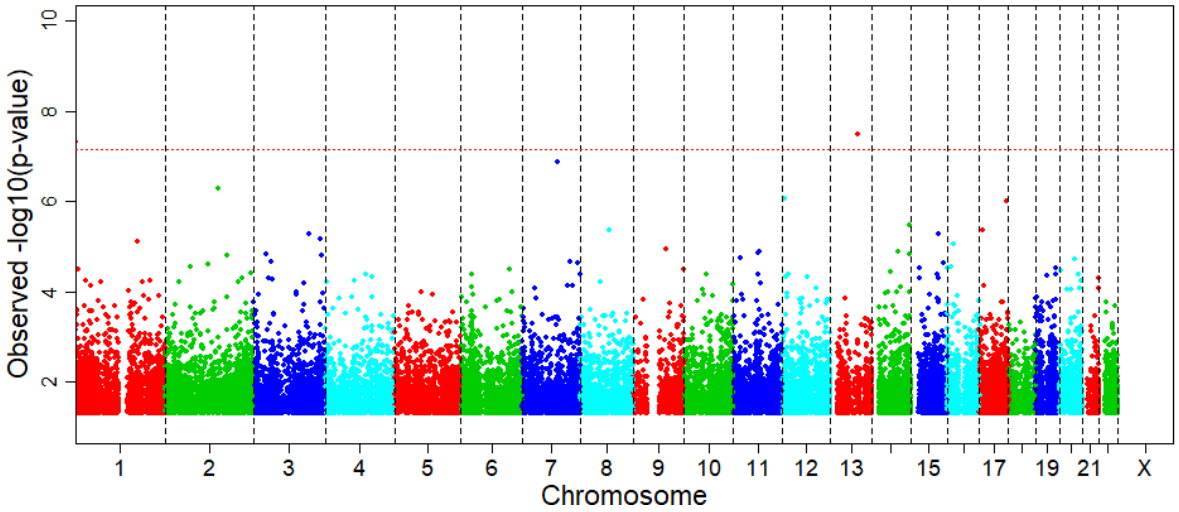

Figure 26. Manhattan plot of associations between genome-wide CpG probes and 1st vaccination adjusted for age, sex and patient condition.

Figure caption: The horizontal red line denotes the significance threshold based on the Bonferroni correction; points above this threshold are considered statistically significant. Each data point on the plot represents an epigenetic marker tested in the study, where they are colored based on the chromosome. The higher a point on the y-axis, the stronger the association with vaccination status.

After correction for multiple testing of the epigenome-wide analyses, three CpGs were statistically significant at FDR < 0.05 (Table 1). Higher methylation levels in two of these CpGs and lower methylation at cg11148246 was associated with receiving the first vaccine, as reflected in the positive and negative beta coefficients respectively.

Table 1. Significant CpGs from model adjusting for age, sex and patient condition

CpG	Beta	SE	p val	p FDR	Chr	Gene
cg05447102	0.013	0.002	3.16E-08	0.016	13	NA
cg23931819	0.008	0.001	4.62E-08	0.016	1	*PUSL1*
cg11148246	-0.012	0.002	1.31E-07	0.031	7	*DYNC1I1*

CpG: CpG probe, Beta: beta coefficient from linear regression, SE: standard error from beta coefficient, p val: p value, p FDR: FDR-corrected p value, Chr: chromosome, Gene: gene annotated to CpG

A second model additionally adjusted for whole blood counts, as shown below:

Model 2: DNAm in CpG ~ vaccination + age + sex + condition code + estimation of white blood cell type

Epigenome-wide results of this EWAS are displayed in Figure 27. Eleven CpGs were significantly associated with having received a first vaccination (FDR<0.05, Figure caption: The horizontal red line denotes the significance threshold based on the Bonferroni correction; points above this threshold are considered statistically significant. Each data point on the plot represents an epigenetic marker tested in the study, where they are colored based on the chromosome. The higher a point on the y-axis, the stronger the association with vaccination status.

Table 2). Of these 11 CpGs, two are also significant at the more conservative Bonferroni threshold, as shown in Figure 27.

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

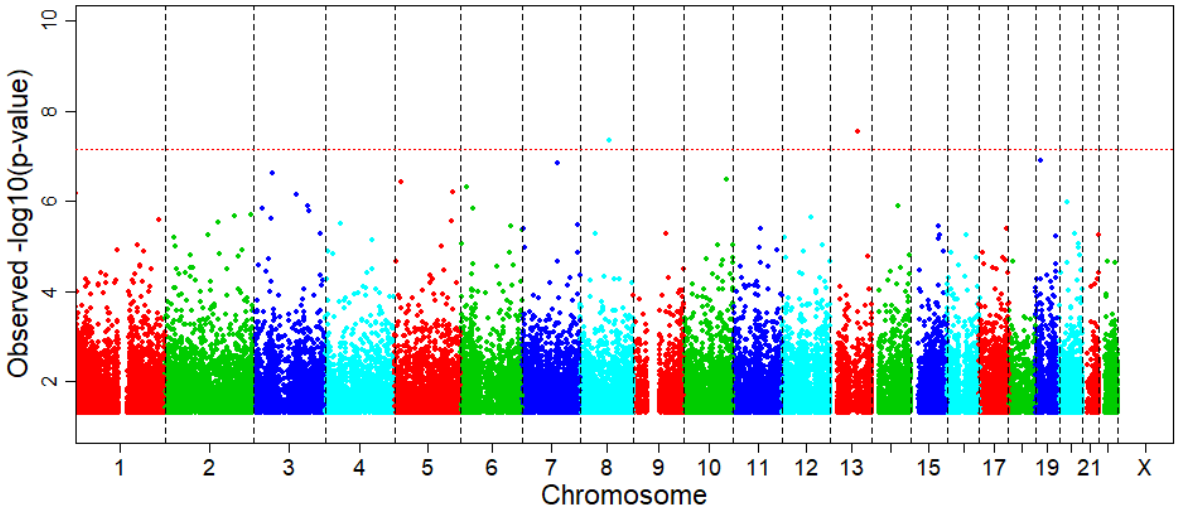

Figure 27. Manhattan plot of associations between genome-wide CpG probes and 1st vaccination adjusted for age, sex, patient condition and white blood cell proportions

Figure caption: The horizontal red line denotes the significance threshold based on the Bonferroni correction; points above this threshold are considered statistically significant. Each data point on the plot represents an epigenetic marker tested in the study, where they are colored based on the chromosome. The higher a point on the y-axis, the stronger the association with vaccination status.

Table 2. Significant CpGs from model adjusting for age, sex, patient condition and white blood cell proportions

CpG	Beta	SE	p val	p FDR	Chr	Gene
cg05447102	0.014	0.002	2.73E-08	0.016	13	-
cg06724127	-0.017	0.003	4.45E-08	0.016	8	-
cg17676631	0.023	0.004	1.23E-07	0.025	19	-
cg11148246	-0.013	0.002	1.41E-07	0.025	7	*DYNC1I1*
cg11575340	-0.014	0.003	2.29E-07	0.032	3	-
cg09031776	0.017	0.003	3.23E-07	0.037	10	-
cg22187952	0.027	0.005	3.75E-07	0.037	5	*FAM134B*
cg06611753	-0.026	0.005	4.86E-07	0.042	6	-
cg11648049	0.012	0.002	6.19E-07	0.045	5	-
cg23931819	0.007	0.001	6.45E-07	0.045	1	*PUSL1*
cg06630764	0.020	0.004	7.28E-07	0.046	3	*STXBP5L*

CpG: CpG probe, Beta: beta coefficient from linear regression, SE: standard error from beta coefficient, p val: p value, p FDR: FDR-corrected p value, Chr: chromosome, Gene: gene annotated to CpG

Gene ontology and KEGG pathway enrichment analyses were performed using the goMeth function from the missMethyl R package (v.1.32.1) (Maksimovic et al., 2021), which accounts for biases derived from the use of array-based DNAm measurements. The list of genes used as background included all genes covered in the EPIC array. No GO terms or pathways were found to be enriched at FDR<0.05.

These analyses show three CpG sites to be consistently associated with having received a first vaccination. Higher methylation levels at cg05447102 (in chromosome 13, not annotated to any gene) and cg23931819 (annotated to *PUSL1*) were associated with receiving the first vaccine in both the basic model and the model additionally adjusting for differences in white

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

blood cell proportions. Similarly, lower methylation levels in cg11148246 (annotated to *DYNC1I1*) were consistently associated with vaccination in both statistical models.

Interestingly, higher DNAm levels in cg05447102 and cg23931819 have been reported as associated with rheumathoid arthritis and age (Liu et al., 2013; Mulder et al., 2021). Contrary to the positive direction of effect observed in our analyses, smoking seems to be associated with lower DNAm levels in cg23931819 (Joehanes et al., 2016; Sikdar et al., 2019).

A few of the additional CpG sites identified in the fully adjusted model, namely cg11148246, and cg17676631, have also been reported as positively associated with age (Mulder et al., 2021). Likewise, a recent epigenome-wide analyses of chronic low-grade inflammation across diverse population cohorts reported suggestive associations between CRP and changes in DNAm levels of cg22187952, cg11648049, cg06630764 and cg23931819, as well as CpGs in *STXBP5L* (although not specifically cg06630764) (Hillary et al., 2024). Of note, no EWAS results were found in the literature for cg11575340 and cg06611753.

Taken together, the results here presented seem to suggest an interplay between vaccination, DNA methylation and inflammation in mixed fragile populations. Further research is needed to investigate if the initial vaccine-related changes persist over time and to determine if these changes are consistent across various conditions in the population under study. Subsequent studies should validate our analyses and explore the molecular mechanisms that may be responsible for these associations.

References

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, *30*(10), 1363-1369. https://doi.org/10.1093/bioinformatics/btu049

Hillary, R. F., Ng, H. K., McCartney, D. L., Elliott, H. R., Walker, R. M., Campbell, A., Huang, F., Direk, K., Welsh, P., Sattar, N., Corley, J., Hayward, C., McIntosh, A. M., Sudlow, C., Evans, K. L., Cox, S. R., Chambers, J. C., Loh, M., Relton, C. L., . . . Suderman, M. (2024). Blood-based epigenome-wide analyses of chronic low-grade inflammation across diverse population cohorts. *Cell Genom*, *4*(5), 100544. https://doi.org/10.1016/j.xgen.2024.100544

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., Wiencke, J. K., & Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, *13*, 86. https://doi.org/10.1186/1471-2105-13-86

Joehanes, R., Just, A. C., Marioni, R. E., Pilling, L. C., Reynolds, L. M., Mandaviya, P. R., Guan, W., Xu, T., Elks, C. E., Aslibekyan, S., Moreno-Macias, H., Smith, J. A., Brody, J. A., Dhingra, R., Yousefi, P., Pankow, J. S., Kunze, S., Shah, S. H., McRae, A. F., . . . London, S. J. (2016). Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet*, *9*(5), 436-447. https://doi.org/10.1161/circgenetics.116.001506

Lehne, B., Drong, A. W., Loh, M., Zhang, W., Scott, W. R., Tan, S. T., Afzal, U., Scott, J., Jarvelin, M. R., Elliott, P., McCarthy, M. I., Kooner, J. S., & Chambers, J. C. (2015). A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol*, *16*(1), 37. https://doi.org/10.1186/s13059-015-0600-x

Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., Shchetynsky, K., Scheynius, A., Kere, J., Alfredsson, L., Klareskog, L., Ekström, T. J., & Feinberg, A. P. (2013). Epigenome-

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol*, *31*(2), 142-147. https://doi.org/10.1038/nbt.2487

Maksimovic, J., Oshlack, A., & Phipson, B. (2021). Gene set enrichment analysis for genome-wide DNA methylation data. *Genome Biology*, *22*(1), 173. https://doi.org/10.1186/s13059-021-02388-x

Mc Intyre, K., Lanting, P., Deelen, P., Wiersma, H. H., Vonk, J. M., Ori, A. P. S., Jankipersadsing, S. A., Warmerdam, R., van Blokland, I., Boulogne, F., Dijkema, M. X. L., Herkert, J. C., Claringbould, A., Bakker, O., Lopera Maya, E. A., Bültmann, U., Zhernakova, A., Reijneveld, S. A., Zijlstra, E., . . . Franke, L. (2021). Lifelines COVID-19 cohort: investigating COVID-19 infection and its health and societal impacts in a Dutch population-based cohort. *BMJ Open*, *11*(3), e044474. https://doi.org/10.1136/bmjopen-2020-044474

McCartney, D. L., Walker, R. M., Morris, S. W., McIntosh, A. M., Porteous, D. J., & Evans, K. L. (2016). Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom Data*, *9*, 22-24. https://doi.org/10.1016/j.gdata.2016.05.012

Mulder, R. H., Neumann, A., Cecil, C. A. M., Walton, E., Houtepen, L. C., Simpkin, A. J., Rijlaarsdam, J., Heijmans, B. T., Gaunt, T. R., Felix, J. F., Jaddoe, V. W. V., Bakermans-Kranenburg, M. J., Tiemeier, H., Relton, C. L., van, I. M. H., & Suderman, M. (2021). Epigenome-wide change and variation in DNA methylation in childhood: trajectories from birth to late adolescence. *Hum Mol Genet*, *30*(1), 119-134. https://doi.org/10.1093/hmg/ddaa280

Peters, A., Peters, A., Greiser, K. H., Göttlicher, S., Ahrens, W., Albrecht, M., Bamberg, F., Bärnighausen, T., Becher, H., Berger, K., Beule, A., Boeing, H., Bohn, B., Bohnert, K., Braun, B., Brenner, H., Bülow, R., Castell, S., Damms-Machado, A., . . . Zschocke, J. (2022). Framework and baseline examination of the German National Cohort (NAKO). *Eur J Epidemiol*, *37*(10), 1107-1124. https://doi.org/10.1007/s10654-022-00890-5

Pidsley, R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., Molloy, P., Van Djik, S., Muhlhausler, B., Stirzaker, C., & Clark, S. J. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol*, *17*(1), 208. https://doi.org/10.1186/s13059-016-1066-1

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Salas, L. A., Zhang, Z., Koestler, D. C., Butler, R. A., Hansen, H. M., Molinaro, A. M., Wiencke, J. K., Kelsey, K. T., & Christensen, B. C. (2022). Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. *Nat Commun*, *13*(1), 761. https://doi.org/10.1038/s41467-021-27864-7

Schons, M., Pilgram, L., Reese, J. P., Stecher, M., Anton, G., Appel, K. S., Bahmer, T., Bartschke, A., Bellinghausen, C., Bernemann, I., Brechtel, M., Brinkmann, F., Brünn, C., Dhillon, C., Fiessler, C., Geisler, R., Hamelmann, E., Hansch, S., Hanses, F., . . . Vehreschild, J. J. (2022). The German National Pandemic Cohort Network (NAPKON): rationale, study design and baseline characteristics. *Eur J Epidemiol*, *37*(8), 849-870. https://doi.org/10.1007/s10654-022-00896-z

Sikdar, S., Joehanes, R., Joubert, B. R., Xu, C. J., Vives-Usano, M., Rezwan, F. I., Felix, J. F., Ward, J. M., Guan, W., Richmond, R. C., Brody, J. A., Küpers, L. K., Baïz, N., Håberg, S. E., Smith, J. A., Reese, S. E., Aslibekyan, S., Hoyo, C., Dhingra, R., . . . London, S. J. (2019). Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. *Epigenomics*, *11*(13), 1487-1500. https://doi.org/10.2217/epi-2019-0066

Snoeck, C. J., Vaillant, M., Abdelrahman, T., Satagopam, V. P., Turner, J. D., Beaumont, K., Gomes, C. P. C., Fritz, J. V., Schröder, V. E., Kaysen, A., Pavelka, L., Stute, L., Meyers, G. R., Pauly, L., Hansen, M., Pauly, C., Aguayo, G. A., Perquin, M., Hanff,

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

A.-M., . . . group, o. b. o. t. C.-V. s. (2020). Prevalence of SARS-CoV-2 infection in the Luxembourgish population – the CON-VINCE study. *medRxiv*, 2020.2005.2011.20092916. https://doi.org/10.1101/2020.05.11.20092916

Wagner, R., Peterhoff, D., Beileke, S., Günther, F., Berr, M., Einhauser, S., Schütz, A., Niller, H. H., Steininger, P., Knöll, A., Tenbusch, M., Maier, C., Korn, K., Stark, K. J., Gessner, A., Burkhardt, R., Kabesch, M., Schedl, H., Küchenhoff, H., . . . Überla, K. (2021). Estimates and Determinants of SARS-Cov-2 Seroprevalence and Infection Fatality Ratio Using Latent Class Analysis: The Population-Based Tirschenreuth Study in the Hardest-Hit German County in Spring 2020. *Viruses*, *13*(6). https://doi.org/10.3390/v13061118

Acknowledgements

The present study was carried out in the context of the ORCHESTRA Project (www.orchestra-cohort.eu). ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101016167. The views expressed in this abstract are the sole responsibility of the authors and the Commission is not responsible for any use that may be made of the information it contains.

We would like to extend our appreciation to the members of the ORCHESTRA Advisory Panel: Prof. Marc Lallemant, Prof. Leonard Leibovici, Prof. Jaques Scheres, Prof. Ranjeeta Thomas, Prof. Maria Vehreschild and Dr. Catherine Cohet (EMA) and Dr. Anastasia Pharris (ECDC) as well as to the Ethics Board: Prof. GreetIeven, Prof. Nikolaus Forgó and Prof. Satinder Aneja for their support and guidance throughout.

Methylation analysis and initial quality control was performed at the Core Facility Genotyping (Helmholtz Munich) by Nadine Lindemann under the supervision of Dr. Melanie Waldenberger. Further methylation data quality control and preprocessing was performed by Dr. Thomas Delerue, Research Unit Molecular Epidemiology, Institute of Epidemiology, Helmholtz Munich.

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167