ADVANZ4 Exploratory Report

INTRODUCTION

This report contains the analysis of microbiome data including Alpha-Diversity (Rarefaction and lmm) and Beta-Diversity (Ordination, Hierarchical Clustering and detection of differentially abundant taxa) analysis.

Report parameters

The report was generated using the following parameters:

- Diversity catalog (**diversity_slot**): igc
- Taxonomic classification (**taxa_slot**): metaphlan
- Taxonomic level (**tax_level**): Species
- Metadata variables (**metadata_vars**): group, risk_group, center, gender, ethnic_group, CD4diff_48, CD8diff_48, CD4after_48, CD8after_48, CD4, CD8, CD8_CD38_DR, CRP, IL6, TNFa, sCD14, time_point
- Group variable (**group_var**): group

Metadata description

The following table shows a summary of selected metadata variables with descriptive statistics.

Characteristic	**N**	**Overall**, N = 271[1]	**DTG**, N = 144[1]	**RTVr**, N = 127[1]	**p-value**[2]
risk_group	259				0.12
hts		109 (42%)	50 (36%)	59 (48%)	
msm		145 (56%)	85 (62%)	60 (49%)	
pwid		5 (1.9%)	2 (1.5%)	3 (2.5%)	
Unknown		12	7	5	
center	271				
bellvitge		51 (19%)	35 (24%)	16 (13%)	
clinic		120 (44%)	57 (40%)	63 (50%)	
hgtp		17 (6.3%)	9 (6.2%)	8 (6.3%)	
mataro		9 (3.3%)	0 (0%)	9 (7.1%)	
sant_pau		23 (8.5%)	14 (9.7%)	9 (7.1%)	
vall_hebron		51 (19%)	29 (20%)	22 (17%)	
Unknown		0	0	0	
gender	271				0.9
female		35 (13%)	19 (13%)	16 (13%)	
male		236 (87%)	125 (87%)	111 (87%)	
Unknown		0	0	0	
ethnic_group	267				0.005
asian		4 (1.5%)	4 (2.8%)	0 (0%)	
black		18 (6.7%)	7 (4.9%)	11 (8.9%)	
caucassian		106 (40%)	66 (46%)	40 (33%)	
hispanic		105 (39%)	56 (39%)	49 (40%)	
other		34 (13%)	11 (7.6%)	23 (19%)	
Unknown		4	0	4	
CD4diff_48	205				0.006
<50		7 (3.4%)	0 (0%)	7 (7.0%)	
>50		198 (97%)	105 (100%)	93 (93%)	
Unknown		66	39	27	
CD8diff_48	205				0.024
<50		69 (34%)	43 (41%)	26 (26%)	
>50		136 (66%)	62 (59%)	74 (74%)	
Unknown		66	39	27	
CD4after_48	205				<0.001

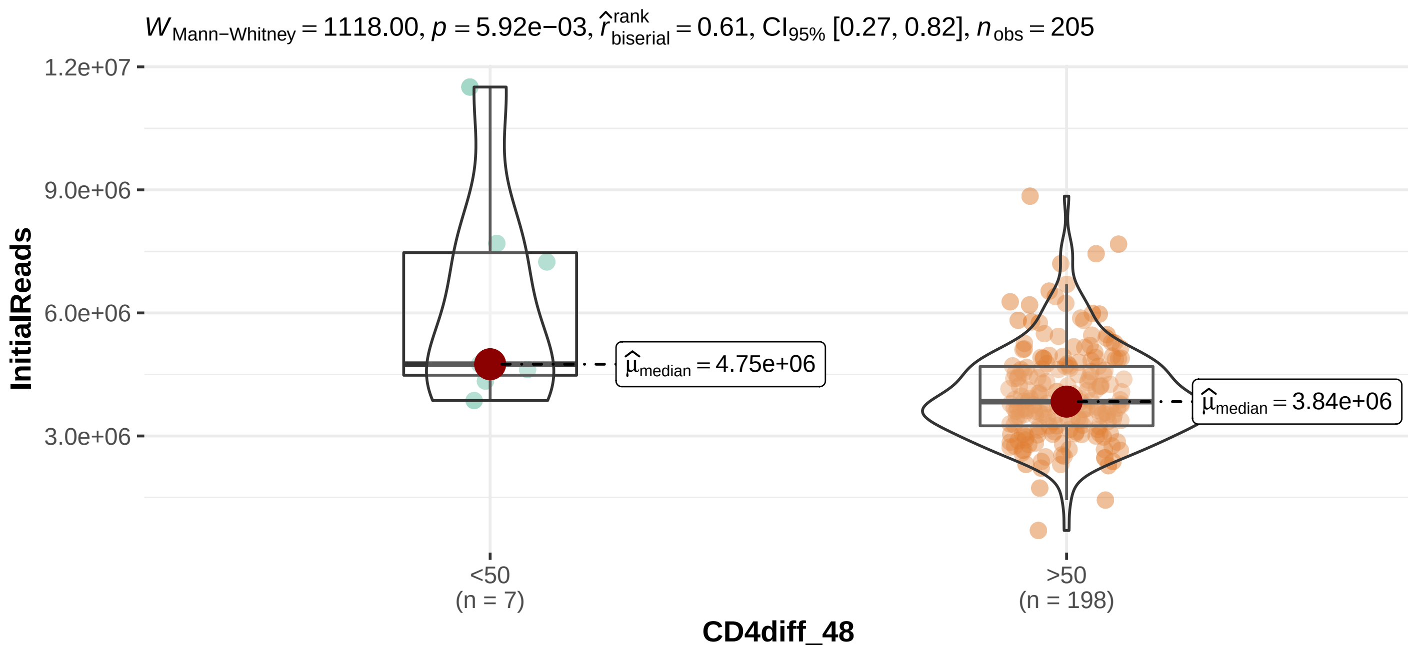
high		15 (7.3%)	4 (3.8%)	11 (11%)	
low		90 (44%)	37 (35%)	53 (53%)	
mid		100 (49%)	64 (61%)	36 (36%)	
Unknown		66	39	27	
CD8after_48	205				0.048
high		171 (83%)	93 (89%)	78 (78%)	
low		3 (1.5%)	0 (0%)	3 (3.0%)	
mid		31 (15%)	12 (11%)	19 (19%)	
Unknown		66	39	27	
CD4	201	113 (40, 237)	140 (60, 245)	89 (30, 216)	0.049
Unknown		70	38	32	
CD8	201	697 (462, 1,132)	725 (504, 1,168)	688 (414, 1,053)	0.3
Unknown		70	38	32	
CD8_CD38_DR	151	31 (19, 48)	31 (17, 48)	31 (20, 47)	0.8
Unknown		120	67	53	
CRP	133	0.18 (0.09, 0.51)	0.15 (0.08, 0.42)	0.20 (0.10, 0.59)	0.14
Unknown		138	77	61	
IL6	128	8 (2, 20)	8 (2, 23)	7 (2, 17)	0.7
Unknown		143	77	66	
TNFa	152	12 (9, 17)	12 (8, 19)	13 (9, 16)	0.8
Unknown		119	66	53	
sCD14	151	2,204 (1,720, 2,986)	1,976 (1,620, 2,836)	2,250 (1,885, 3,139)	0.046
Unknown		120	67	53	
time_point	271				0.9
0		81 (30%)	40 (28%)	41 (32%)	
24		51 (19%)	29 (20%)	22 (17%)	
48		71 (26%)	38 (26%)	33 (26%)	
96		68 (25%)	37 (26%)	31 (24%)	
Unknown		0	0	0	

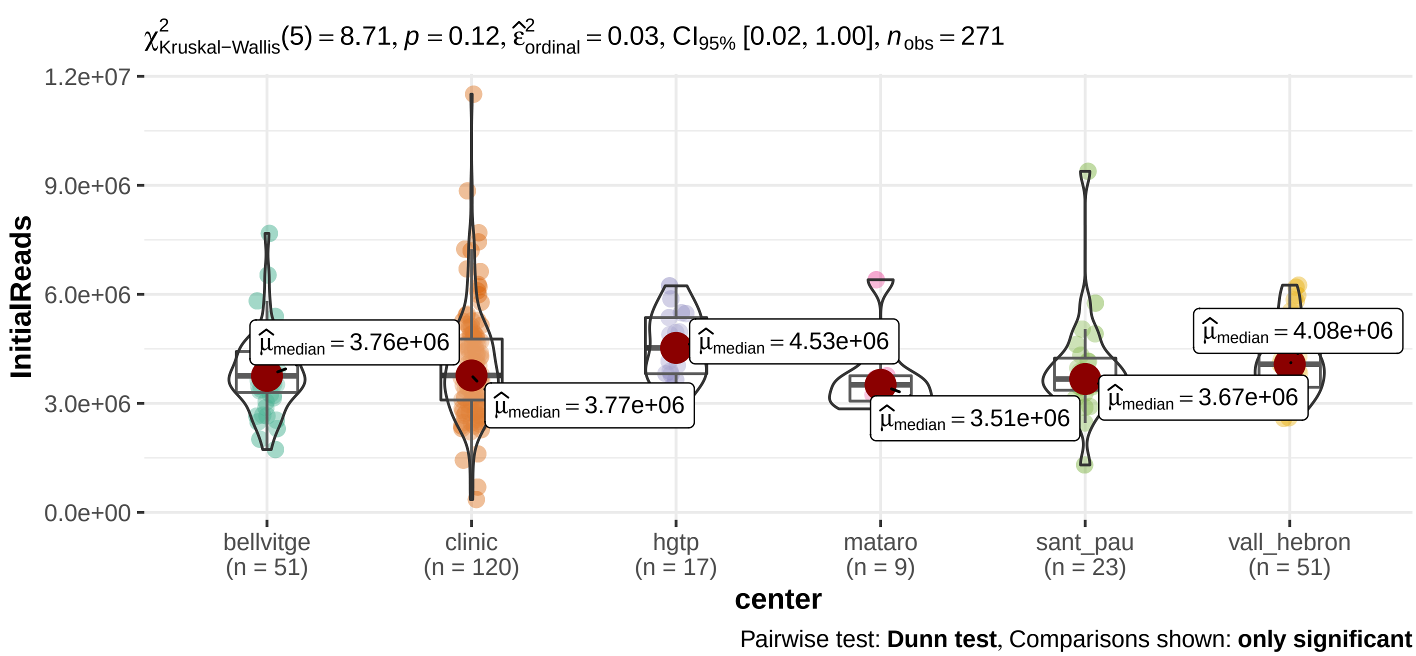
[1]n (%); Median (IQR)
[2]Fisher's exact test; Pearson's Chi-squared test; Wilcoxon rank sum test

QUALITY CONTROL

The objective of this section is to check if there is any association between the total number of sequencing reads and the different levels of each metadata variable. Showing all significant associations. If the number of significant plots is less than 2, the best non-significant associations will be displayed.

Categorical variables:

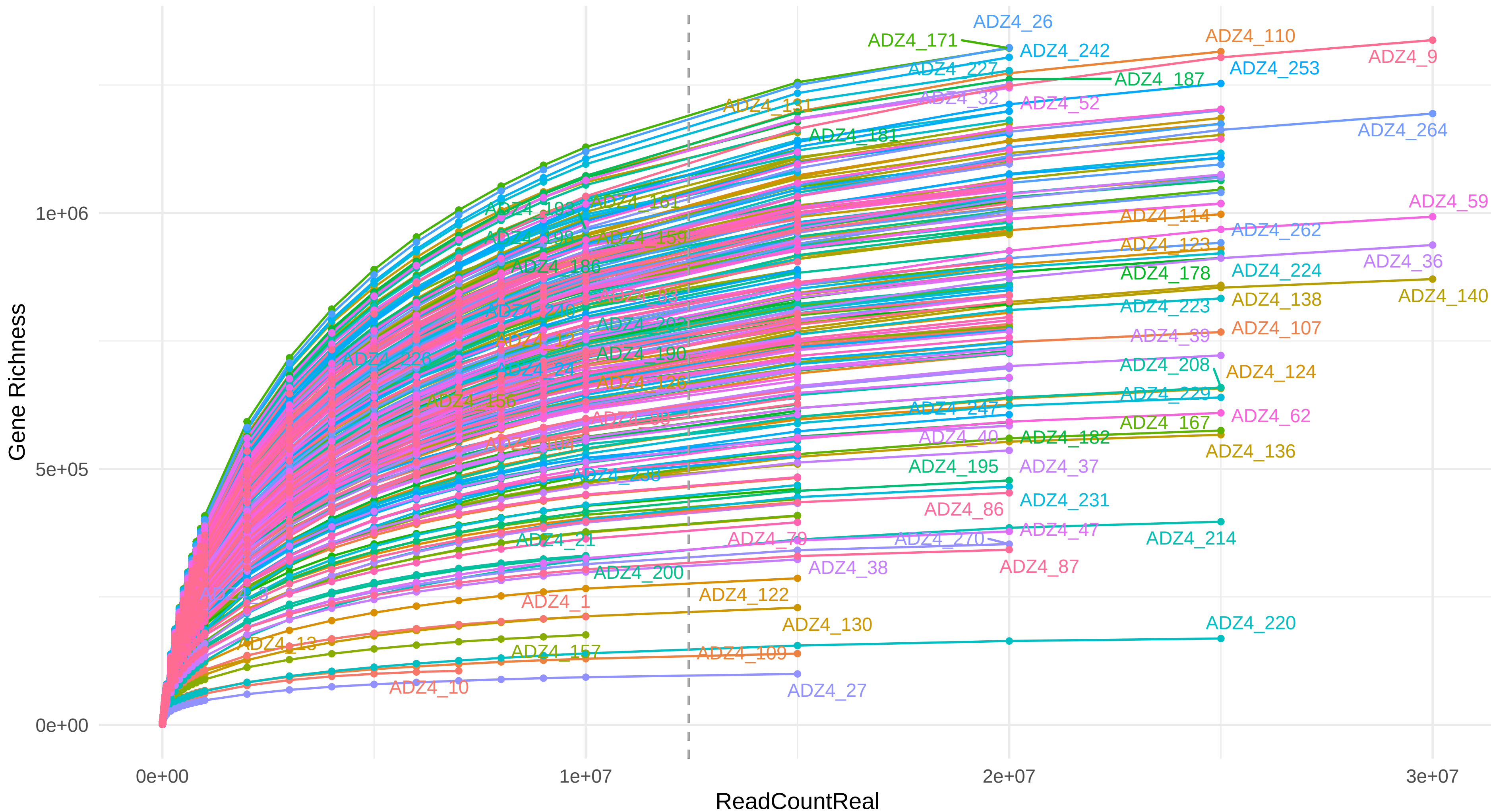




ALPHA-DIVERSITY BASED ANALYSIS

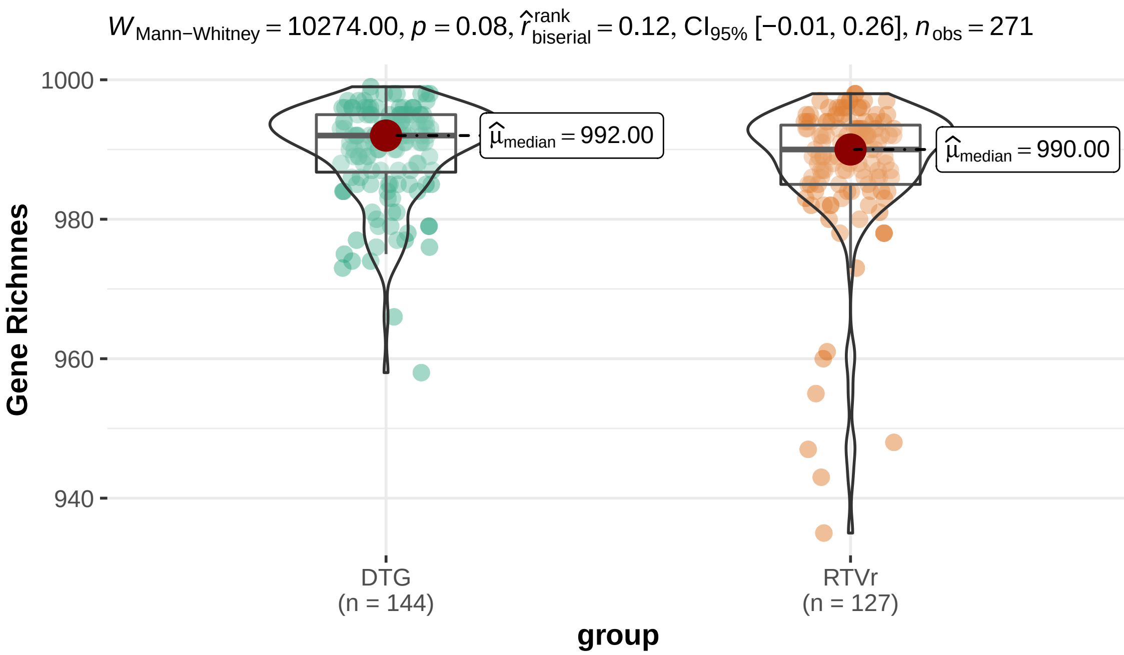
Rarefaction analysis

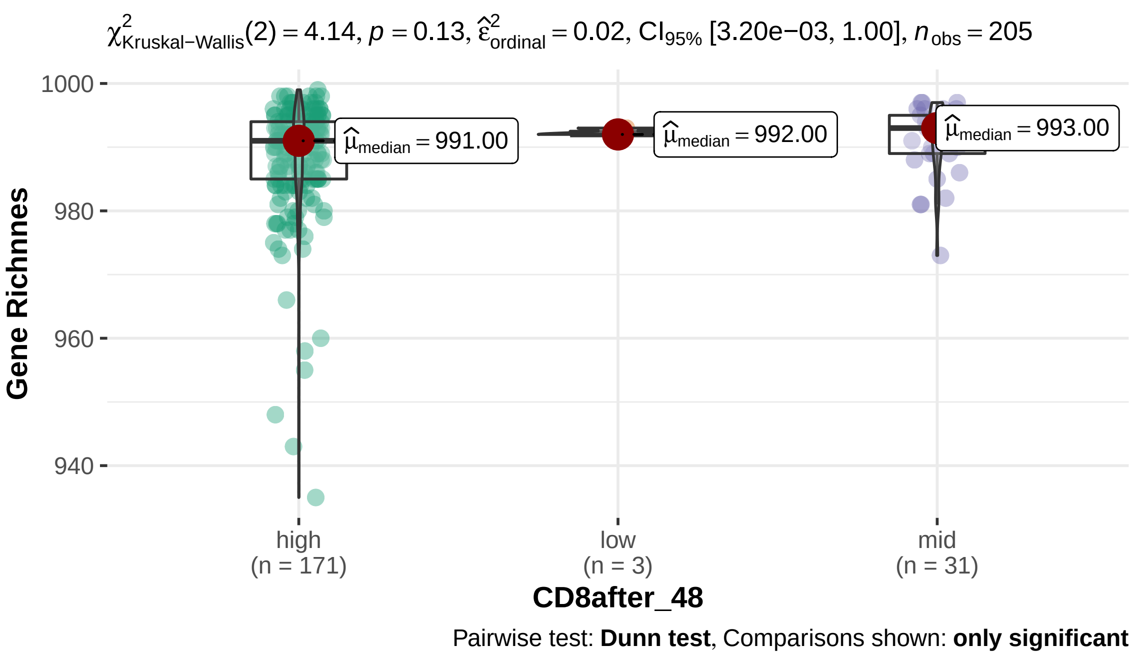
Representation of the increase in the Gene Richness in relation to the total number of mapped reads for each of the samples. Dashed line represents the quantile at a probability of 2%.



Gene Richness by Categorical Variables

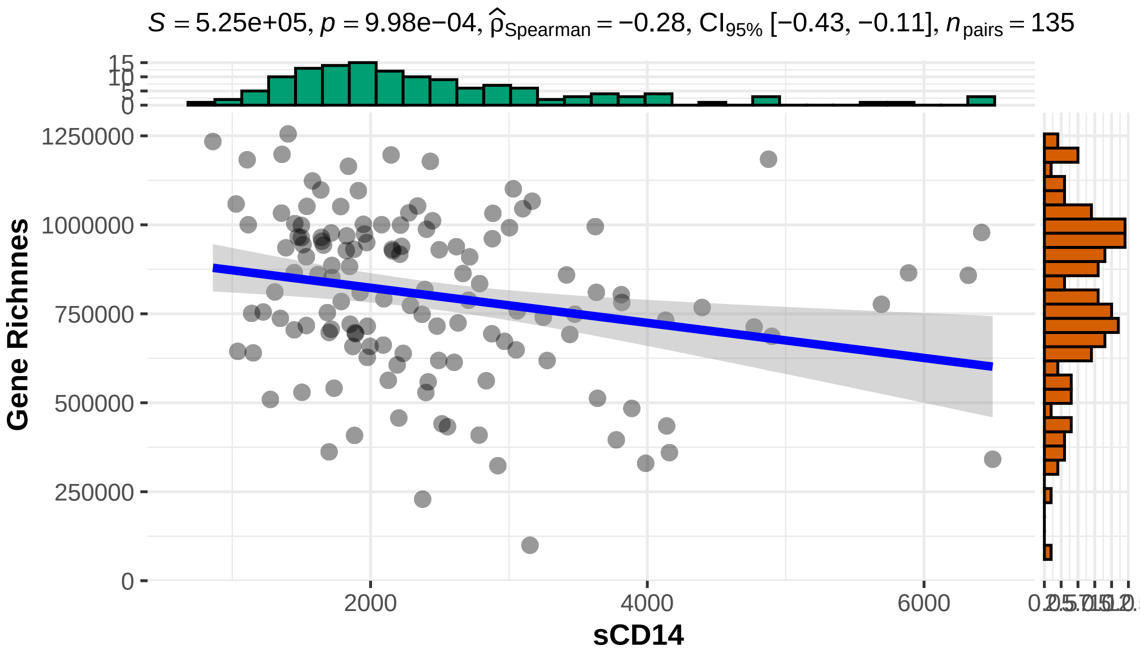
This section shows the genetic wealth for each of the different levels of each of the categorical variables. Plots were produced using the ggstatsplot package. The upper text presents information on inferential statistics and the bottom one provides information about Bayesian hypothesis-testing and estimation. Showing all significant associations. If the number of significant plots is less than 2, the best non-significant associations will be displayed.

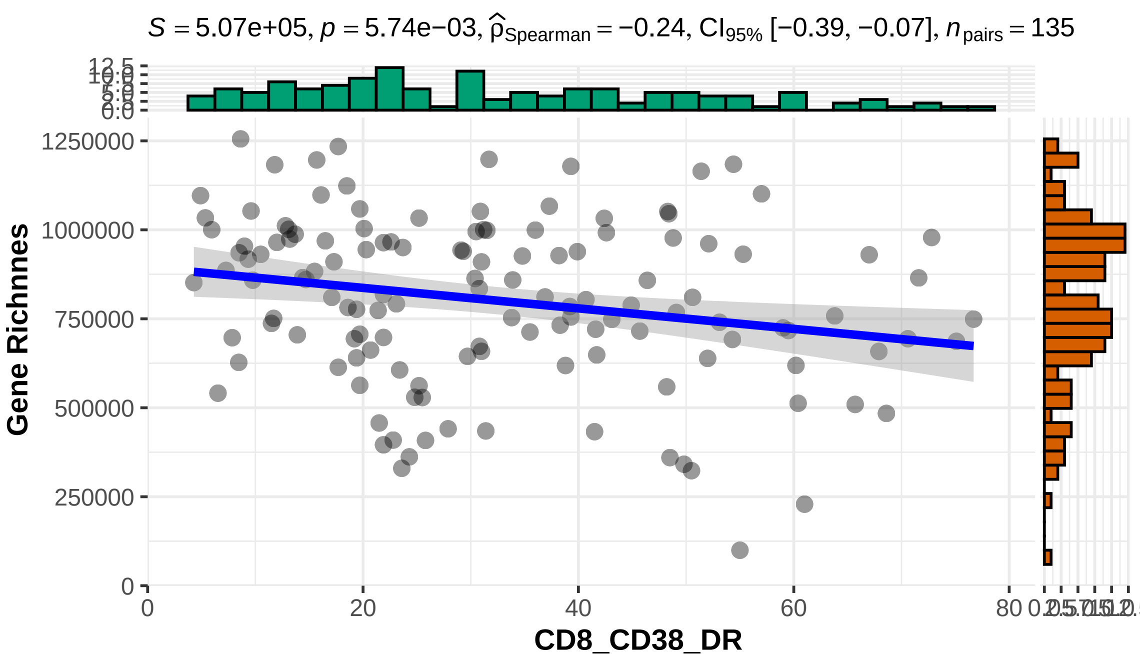


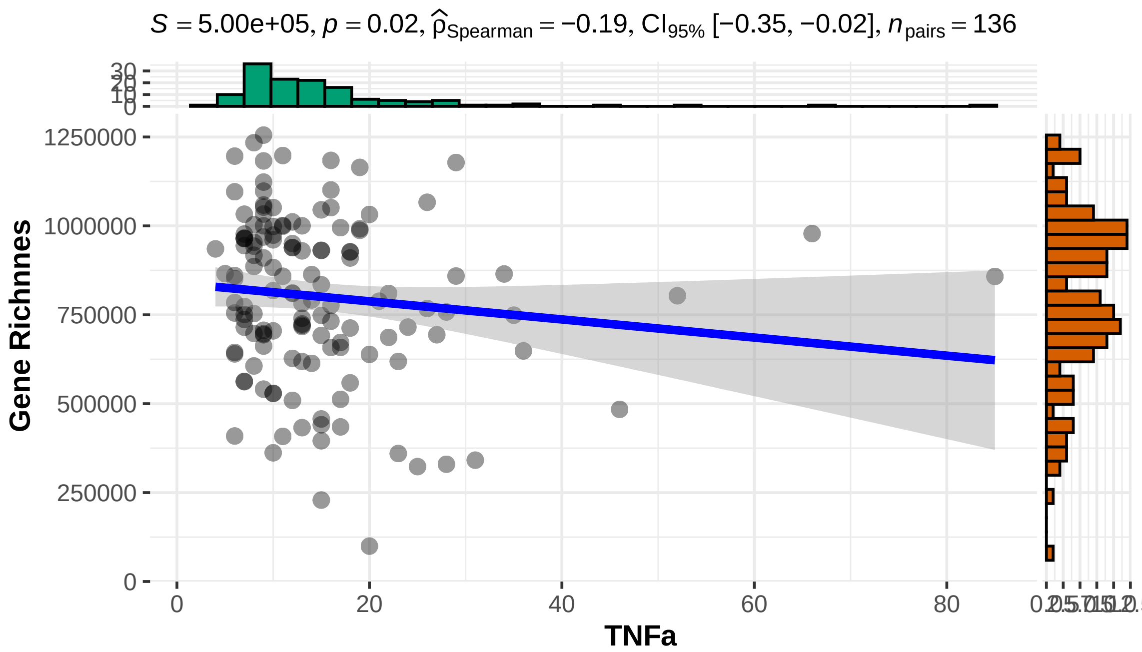


Gene Richness by Numeric Variables

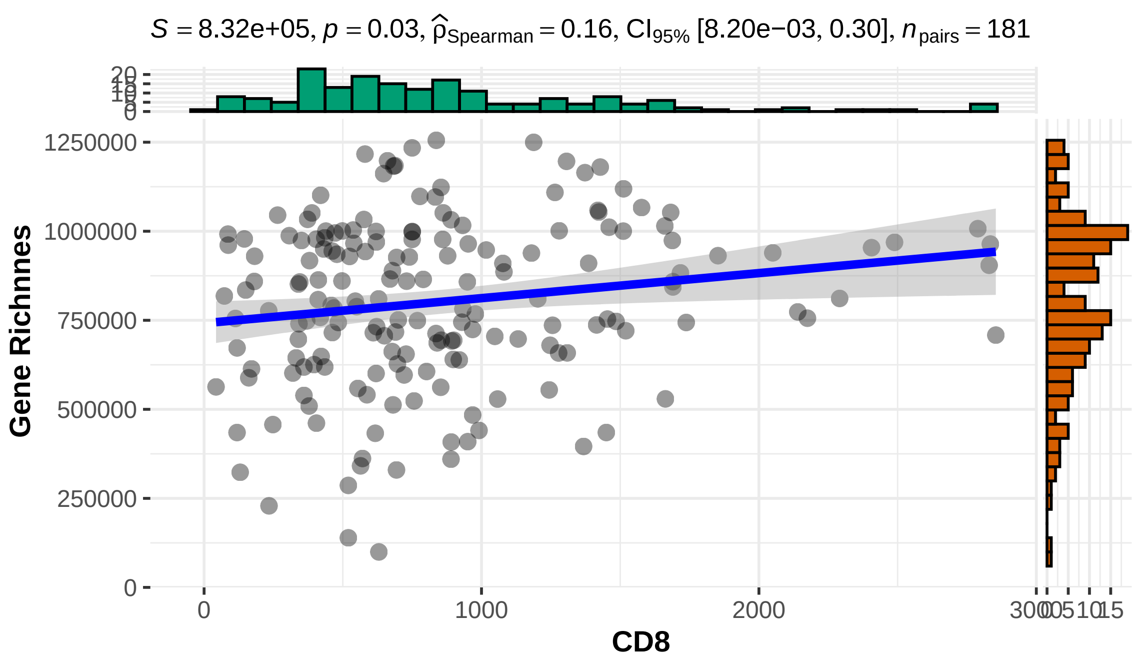
This section shows the genetic wealth for each of the different levels of each of the numeric variables. Plots were produced using the ggstatsplot package. The upper text presents information on inferential statistics and the bottom one provides information about Bayesian hypothesis-testing and estimation. Showing all significant associations. If the number of significant plots is less than 2, the best non-significant associations will be displayed.







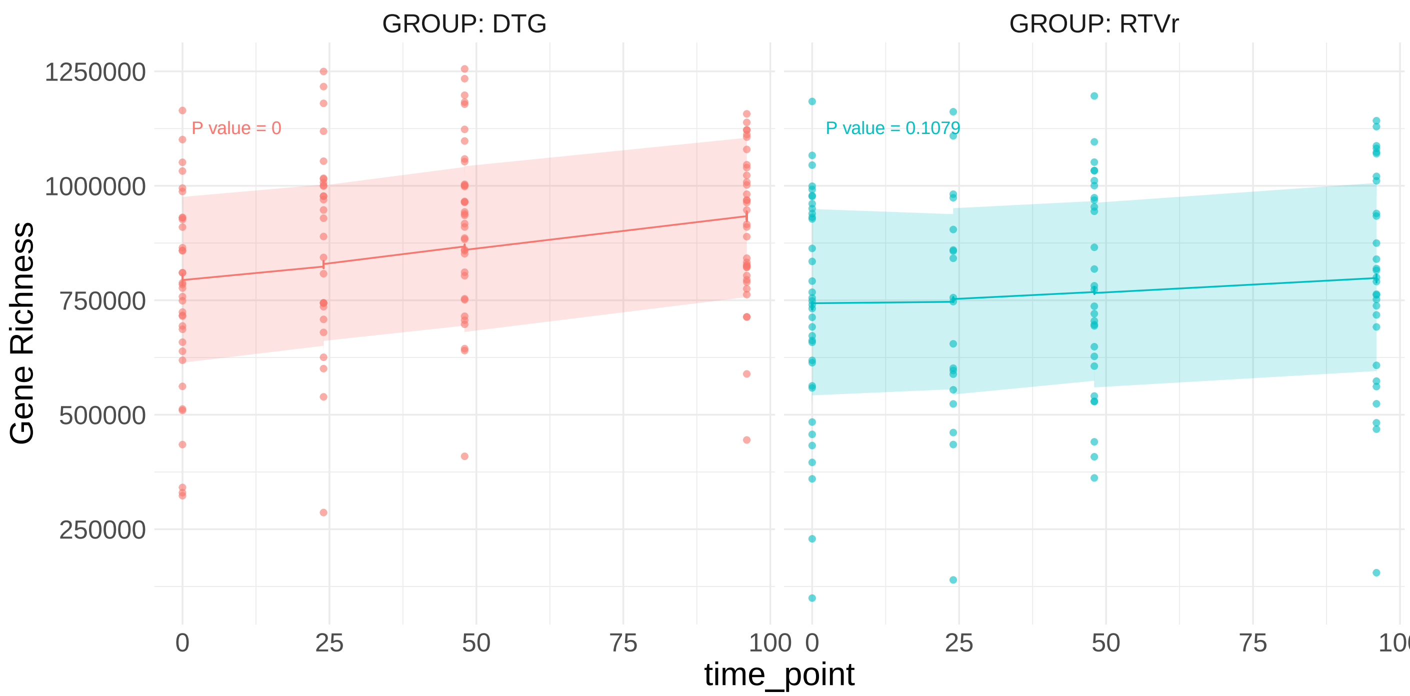
$S = 5.00e+05, p = 0.02, \widehat{\rho}_{Spearman} = -0.19, CI_{95\%} [-0.35, -0.02], n_{pairs} = 136$
Gene Richnnes
TNFa



$S = 8.32e+05, p = 0.03, \widehat{\rho}_{Spearman} = 0.16, CI_{95\%} [8.20e-03, 0.30], n_{pairs} = 181$
Gene Richnnes
CD8

Gene Richness by Longitudinal Variable

In this section, Gene Richness was correlated with longitudinal metadata variables using liner mixed models.

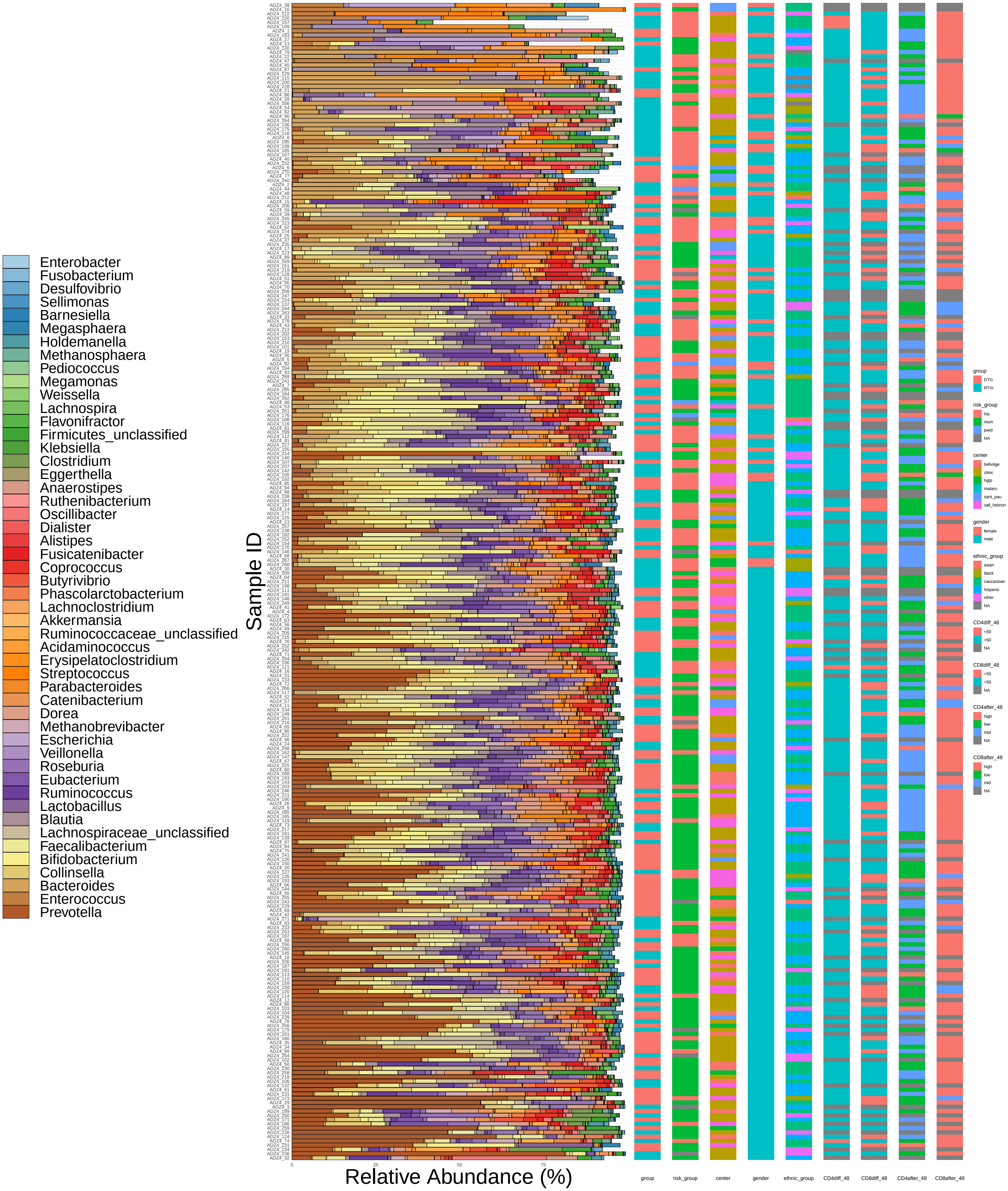


BETA-DIVERSITY BASED ANALYSIS

Taxa description based on Non-Metric Multidimendional Scaling (NMDS) ordering

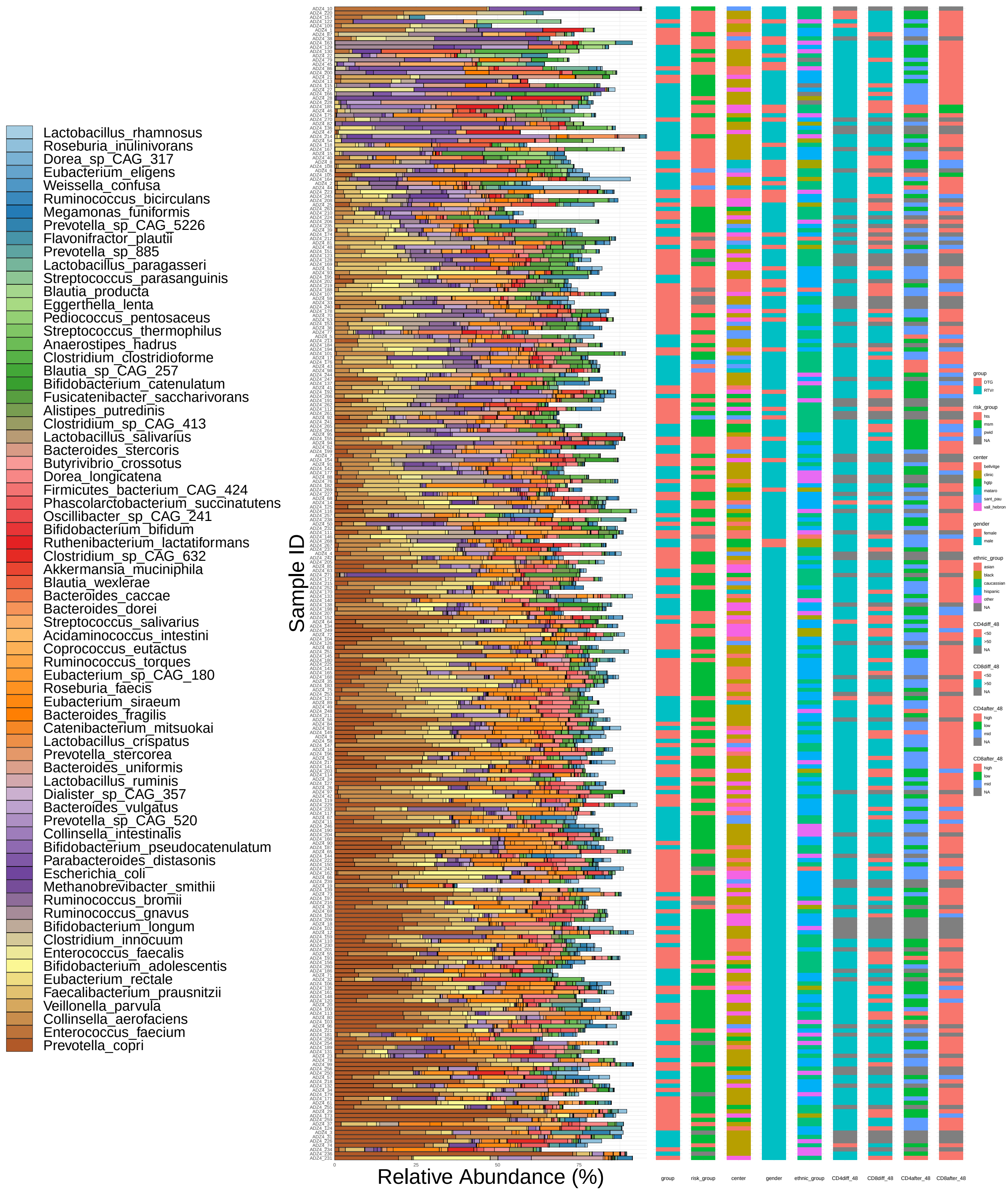
Bar plot showing the relative abundance of different taxa in each sample. Sample axis order was determined using Non-metric Multidimensional Scaling (NMDS). The Bray method was used for distance calculations on Shotgun data, for 16s data Wunifrac distance was used. In the upper part of the graph are shown the distribution of the values of the variables selected from the metadata.

```
#> $Genus
```



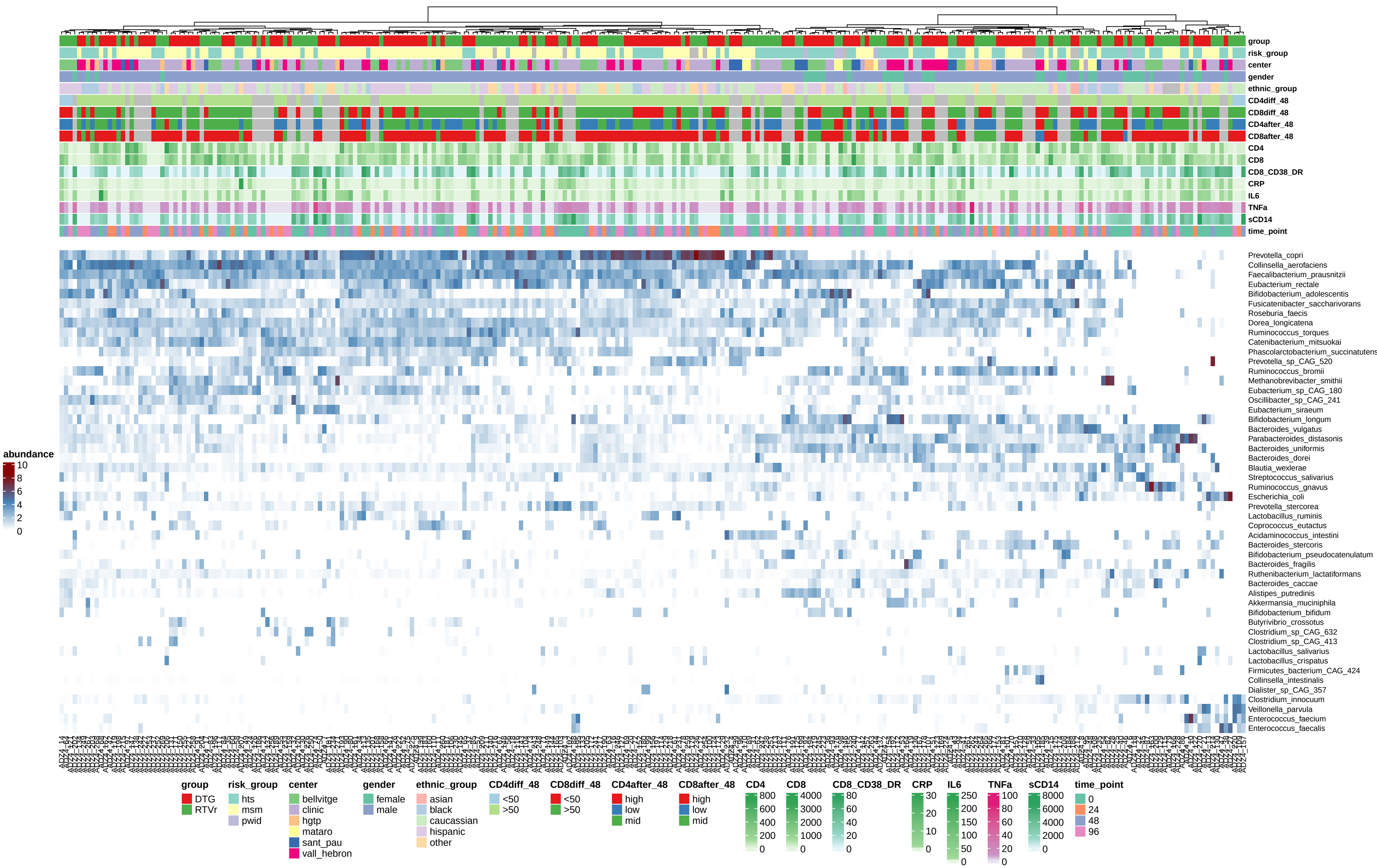
Enterobacter
Fusobacterium
Desulfovibrio
Sellimonas
Barnesiella
Megasphaera
Holdemanella
Methanosphaera
Pediococcus
Megamonas
Weissella
Lachnospira
Flavonifractor
Firmicutes_unclassified
Klebsiella
Clostridium
Eggerthella
Anaerostipes
Ruthenibacterium
Oscillibacter
Dialister
Alistipes
Fusicatenibacter
Coprococcus
Butyrivibrio
Phascolarctobacterium
Lachnoclostridium
Akkermansia
Ruminococcaceae_unclassified
Acidaminococcus
Erysipelatoclostridium
Streptococcus
Parabacteroides
Catenibacterium
Dorea
Methanobrevibacter
Escherichia
Veillonella
Roseburia
Eubacterium
Ruminococcus
Lactobacillus
Blautia
Lachnospiraceae_unclassified
Faecalibacterium
Bifidobacterium
Collinsella
Bacteroides
Enterococcus
Prevotella
Sample ID
Relative Abundance (%)
group
risk_group
center
gender
ethnic_group
CD4diff_48
CD8diff_48
CD4after_48
CD8after_48


```
#>
#> $Species
```


Lactobacillus_rhamnosus
Roseburia_inulinivorans
Dorea_sp_CAG_317
Eubacterium_eligens
Weissella_confusa
Ruminococcus_bicirculans
Megamonas_funiformis
Prevotella_sp_CAG_5226
Flavonifractor_plautii
Prevotella_sp_885
Lactobacillus_paragasseri
Streptococcus_parasanguinis
Blautia_producta
Eggerthella_lenta
Pediococcus_pentosaceus
Streptococcus_thermophilus
Anaerostipes_hadrus
Clostridium_clostridioforme
Blautia_sp_CAG_257
Bifidobacterium_catenulatum
Fusicatenibacter_saccharivorans
Alistipes_putredinis
Clostridium_sp_CAG_413
Lactobacillus_salivarius
Bacteroides_stercoris
Butyrivibrio_crossotus
Dorea_longicatena
Firmicutes_bacterium_CAG_424
Phascolarctobacterium_succinatutens
Oscillibacter_sp_CAG_241
Bifidobacterium_bifidum
Ruthenibacterium_lactatiformans
Clostridium_sp_CAG_632
Akkermansia_muciniphila
Blautia_wexlerae
Bacteroides_caccae
Bacteroides_dorei
Streptococcus_salivarius
Acidaminococcus_intestini
Coprococcus_eutactus
Ruminococcus_torques
Eubacterium_sp_CAG_180
Roseburia_faecis
Eubacterium_siraeum
Bacteroides_fragilis
Catenibacterium_mitsuokai
Lactobacillus_crispatus
Prevotella_stercorea
Bacteroides_uniformis
Lactobacillus_ruminis
Dialister_sp_CAG_357
Bacteroides_vulgatus
Prevotella_sp_CAG_520
Collinsella_intestinalis
Bifidobacterium_pseudocatenulatum
Parabacteroides_distasonis
Escherichia_coli
Methanobrevibacter_smithii
Ruminococcus_bromii
Ruminococcus_gnavus
Bifidobacterium_longum
Clostridium_innocuum
Enterococcus_faecalis
Bifidobacterium_adolescentis
Eubacterium_rectale
Faecalibacterium_prausnitzii
Veillonella_parvula
Collinsella_aerofaciens
Enterococcus_faecium
Prevotella_copri
Sample ID
Relative Abundance (%)
0
25
50
75
group
risk_group
center
gender
ethnic_group
CD4diff_48
CD8diff_48
CD4after_48
CD8after_48
group
DTG
RTVr
risk_group
hts
msm
pwid
NA
center
bellvitge
clinic
hgtp
mataro
sant_pau
vall_hebron
gender
female
male
ethnic_group
asian
black
caucassian
hispanic
other
NA
CD4diff_48
<50
>50
NA
CD8diff_48
<50
>50
NA
CD4after_48
high
low
mid
NA
CD8after_48
high
low
mid
NA

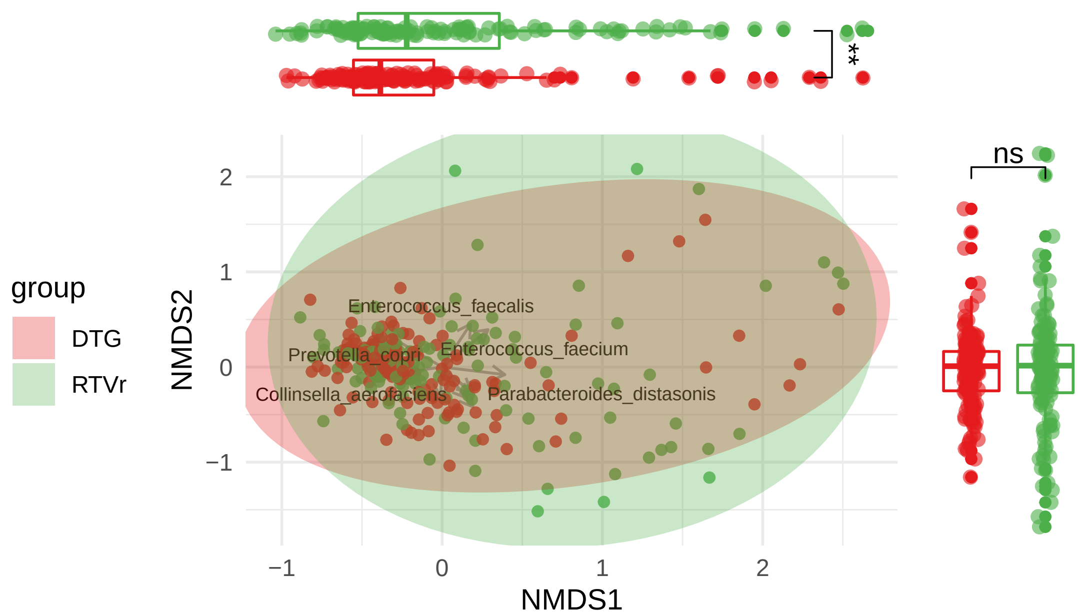
Hierarchical Clustering Analysis

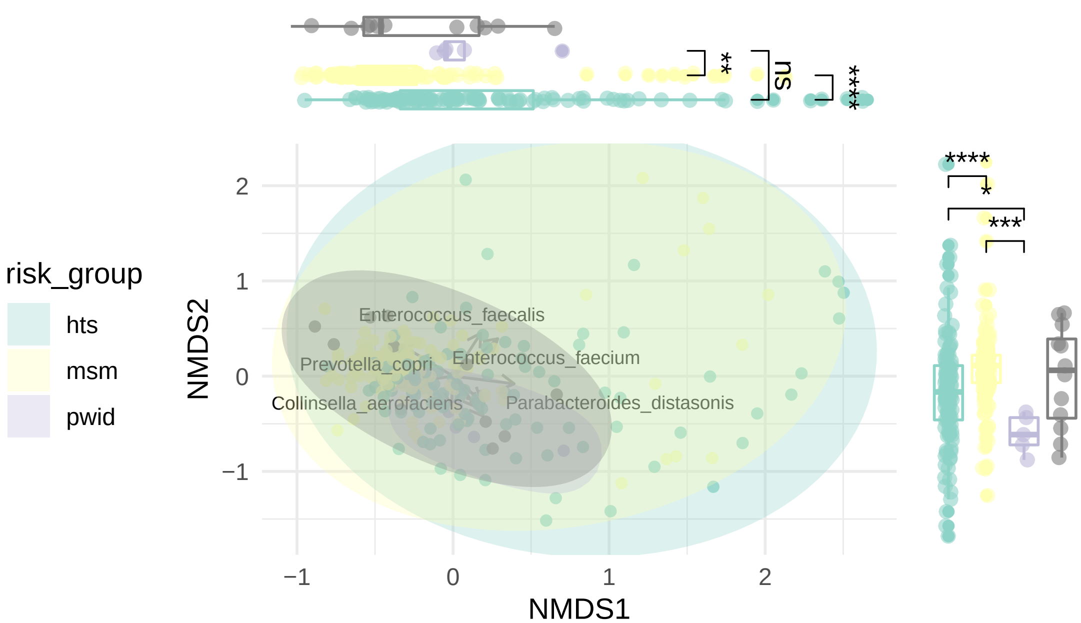
Heatmap showing the abundance of different taxas in each sample. Sample order was determined using ward.D2 hierarchical clustering. The categorical and numeric variables present in the mre object were used for sample annotation.

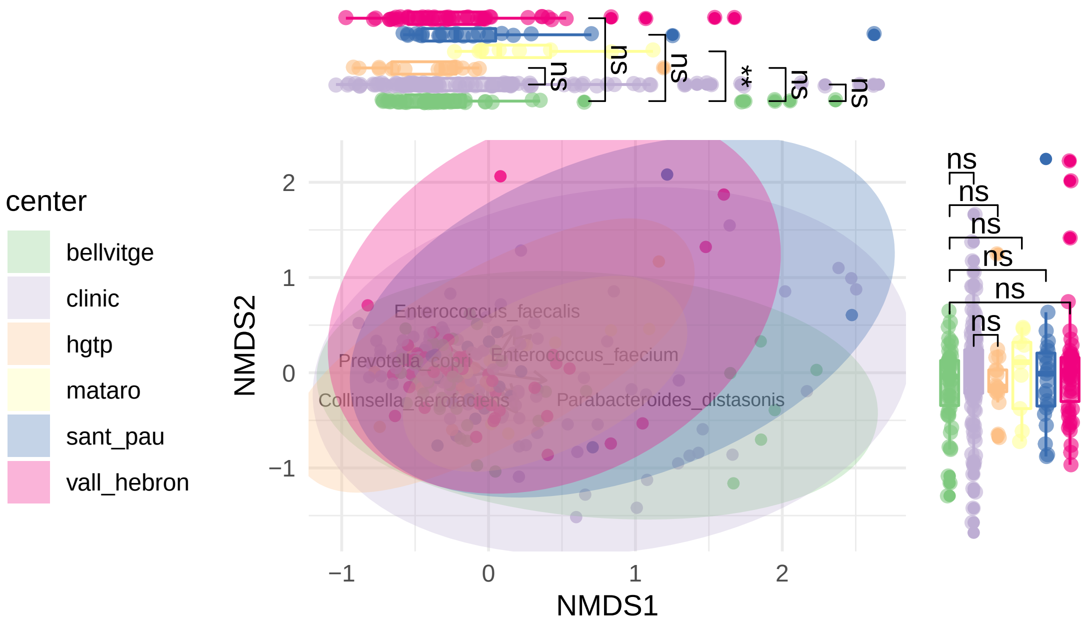


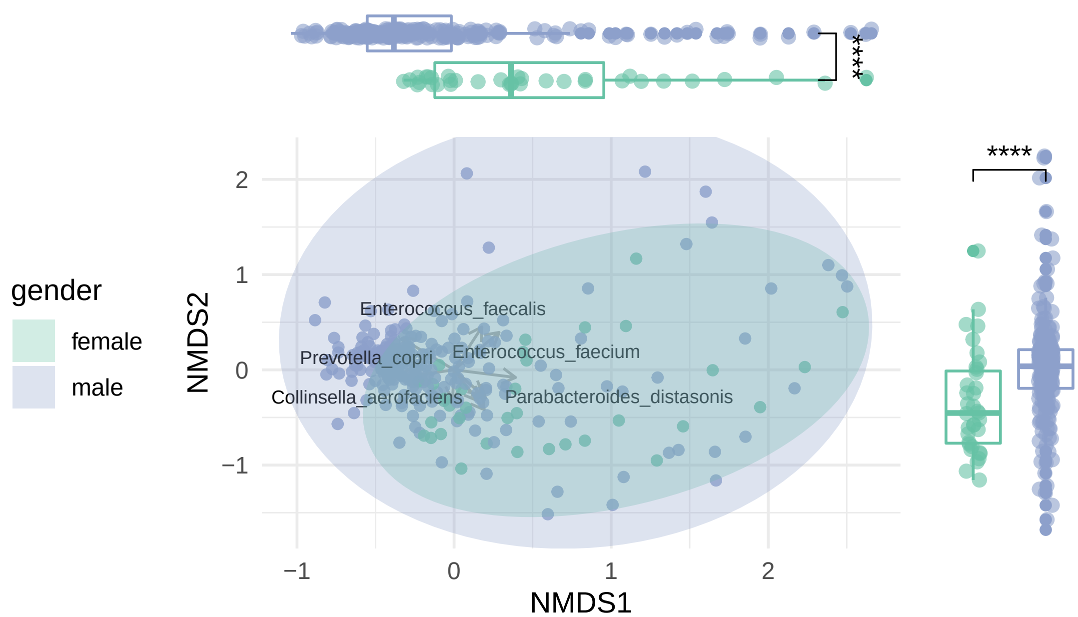
Ordination Analysis (Non-metric multidimensional scaling) by Catecorical Variables.

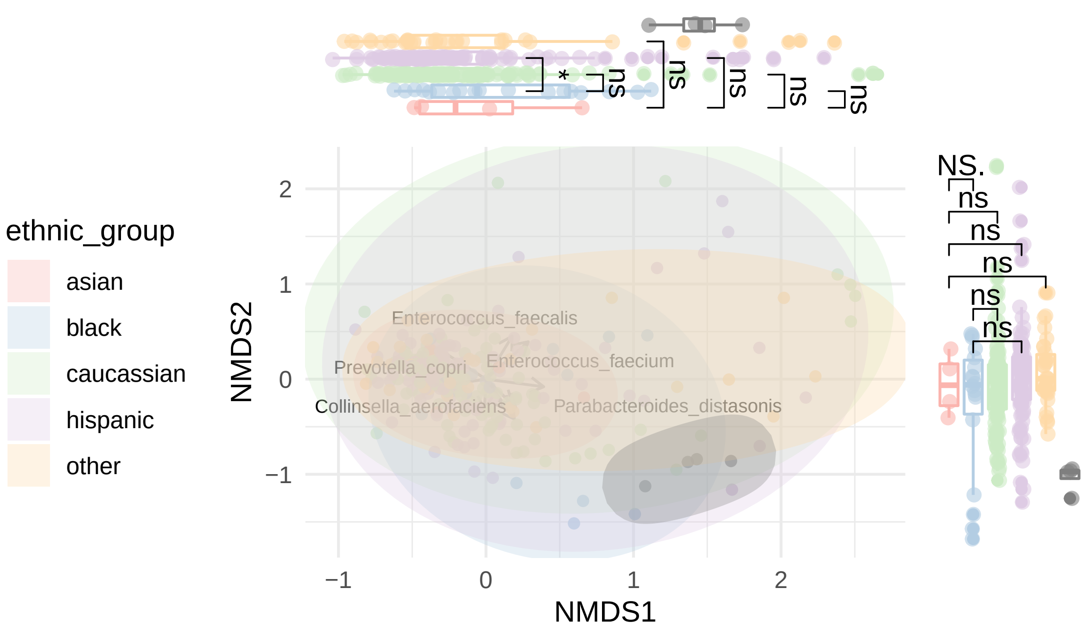
Non-metric multidimensional scaling plot of categorical metadata variables and microbial community compositions. One plot for each categorical metadata. NMDS analysis within the vegan package of R software package based on dissimilarities calculated using the Bray-Curtis (Shotgun data) or WUnifrac (16s data) index of bacterial communities composition for the relative abundance of each OTU in relation to the categorical variables.

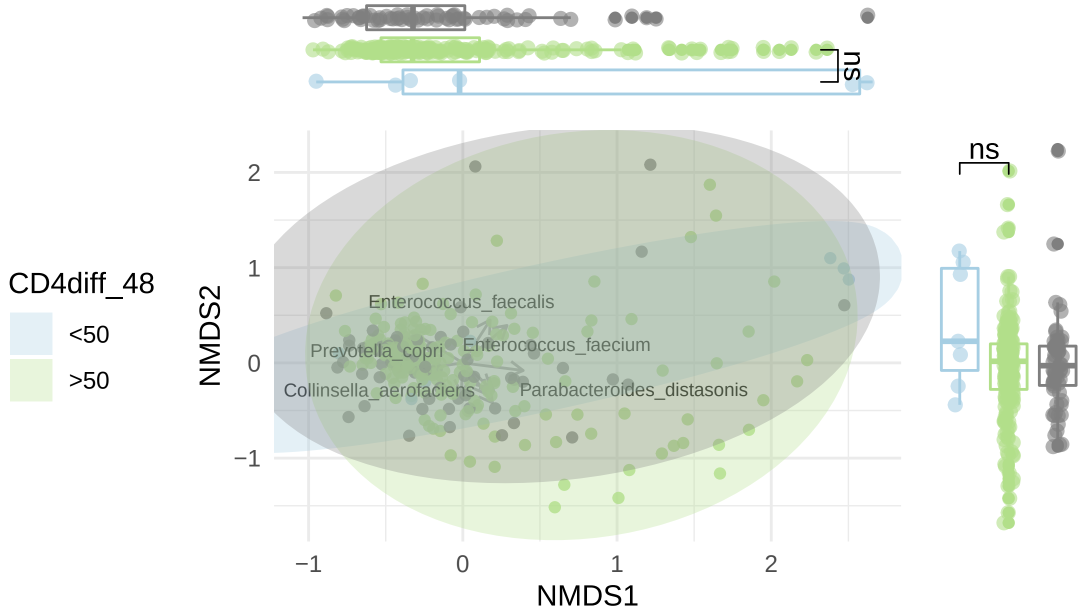


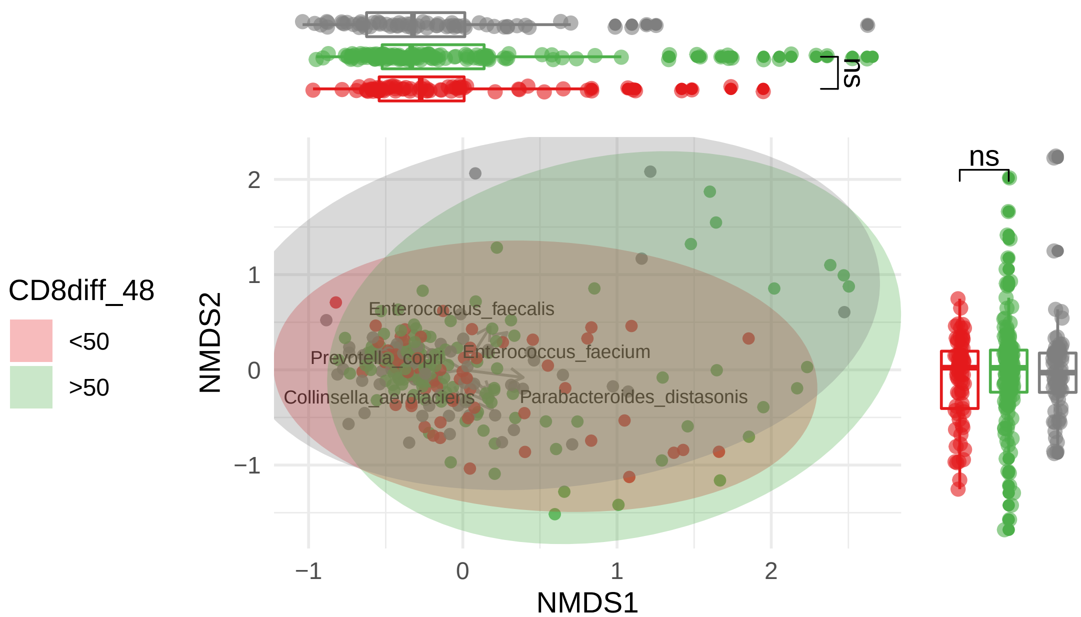


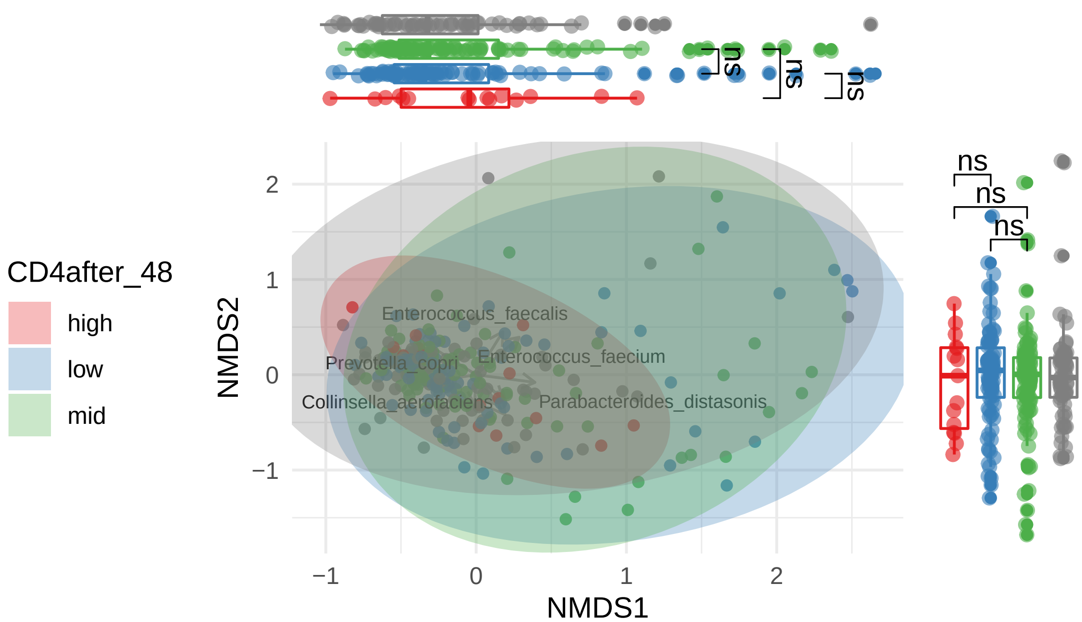

center
bellvitge
clinic
hgtp
mataro
sant_pau
vall_hebron
NMDS1
NMDS2
Enterococcus_faecalis
Enterococcus_faecium
Prevotella_copri
Collinsella_aerofaciens
Parabacteroides_distasonis
ns
**

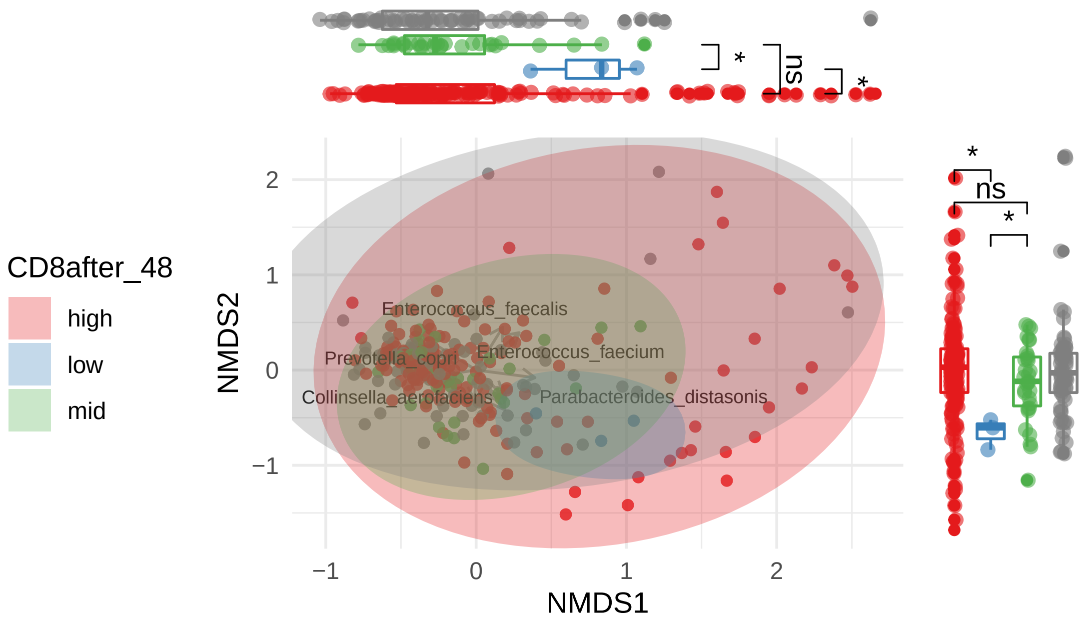

gender
female
male
NMDS1
NMDS2
Enterococcus_faecalis
Enterococcus_faecium
Prevotella_copri
Collinsella_aerofaciens
Parabacteroides_distasonis
****

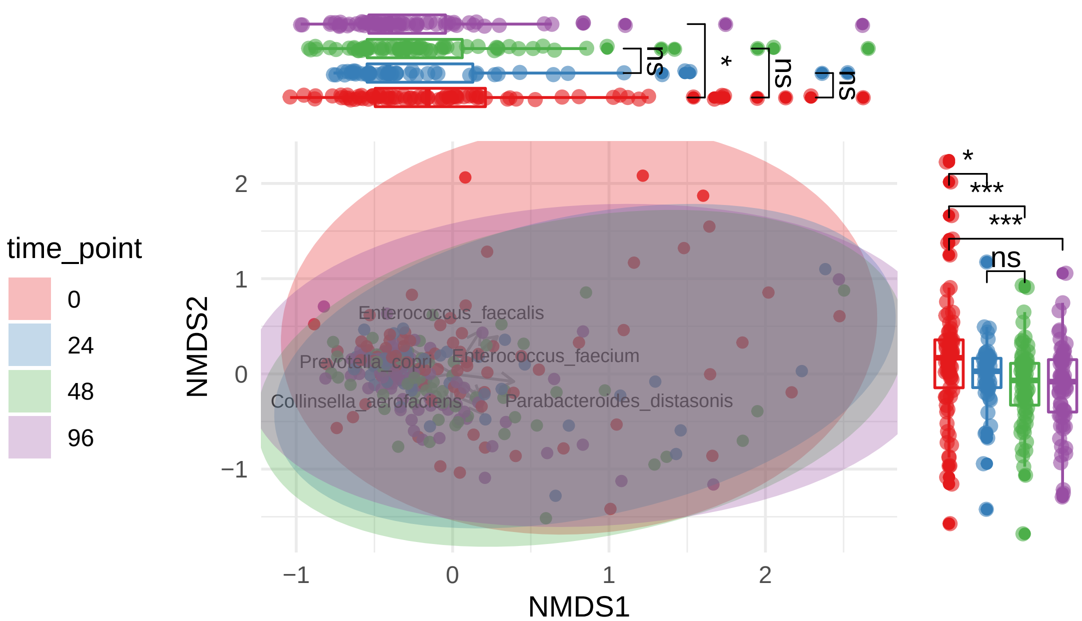

ethnic_group
asian
black
caucassian
hispanic
other
NMDS1
NMDS2
Enterococcus_faecalis
Enterococcus_faecium
Prevotella_copri
Collinsella_aerofaciens
Parabacteroides_distasonis
NS.
ns
*


CD4diff_48
<50
>50
NMDS2
NMDS1
Enterococcus_faecalis
Prevotella_copri
Enterococcus_faecium
Collinsella_aerofaciens
Parabacteroides_distasonis
ns
ns


CD8diff_48
<50
>50
NMDS2
NMDS1
Enterococcus_faecalis
Prevotella_copri
Enterococcus_faecium
Collinsella_aerofaciens
Parabacteroides_distasonis
ns
ns


CD4after_48
high
low
mid
NMDS2
NMDS1
Enterococcus_faecalis
Prevotella_copri
Enterococcus_faecium
Collinsella_aerofaciens
Parabacteroides_distasonis
ns
ns
ns


CD8after_48
high
low
mid
NMDS2
NMDS1
Enterococcus_faecalis
Enterococcus_faecium
Prevotella_copri
Collinsella_aerofaciens
Parabacteroides_distasonis
ns
*

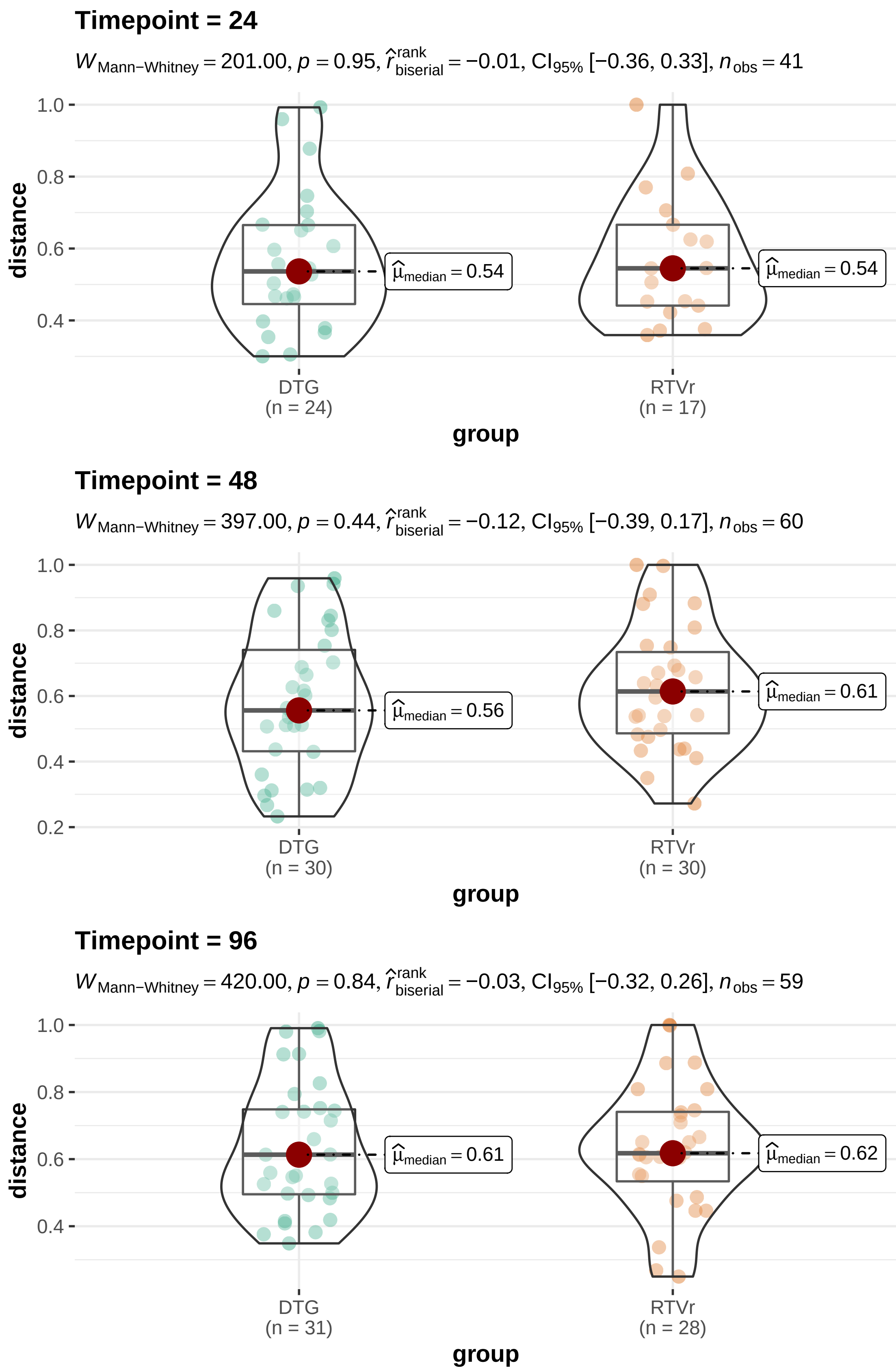

time_point
0
24
48
96
NMDS2
NMDS1
Enterococcus_faecalis
Enterococcus_faecium
Prevotella_copri
Collinsella_aerofaciens
Parabacteroides_distasonis
ns
*
***

The statistics of the marginal boxplots were calculated using the ANOVA test. The Permutational Multivariate Analysis of Variance Using Distance Matrices (PERMANOVA) was computed using the `vegan::adonis()` function. The bottom table shows the results of the PERMANOVA for each categorical variable.

id	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr..F.
group	1	0.9565124	0.9565124	3.409143	0.012514788	0.001
risk_group	2	2.8339413	1.4169706	5.165222	0.038788070	0.001
center	5	2.2744560	0.4548912	1.625573	0.029758458	0.002
gender	1	2.7318254	2.7318254	9.971147	0.035742574	0.001
ethnic_group	4	1.8604200	0.4651050	1.684405	0.025071370	0.001
CD4diff_48	1	0.7980095	0.7980095	2.833267	0.013764863	0.004
CD8diff_48	1	0.3653768	0.3653768	1.287498	0.006302382	0.170
CD4after_48	2	0.7747815	0.3873907	1.368068	0.013364202	0.078
CD8after_48	2	1.5293299	0.7646650	2.736507	0.026379405	0.002
time_point	3	1.2042928	0.4014309	1.424795	0.015756690	0.026
record_id	94	46.4281638	0.4939166	2.897412	0.607455395	0.001
cluster	1	8.2019180	8.2019180	32.337087	0.107312005	0.001

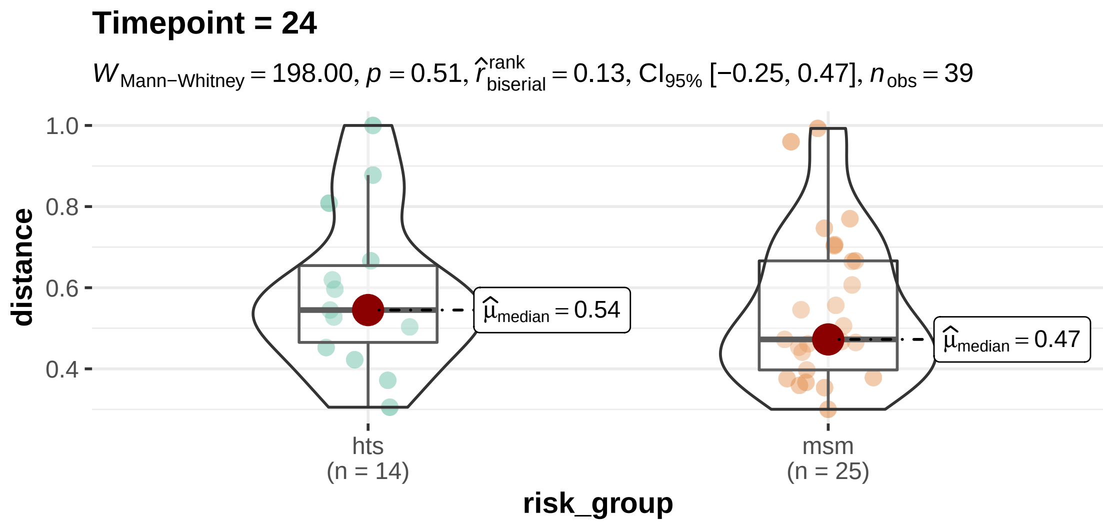
Changes in Beta diversity along longitudinal variable by each categorical variable

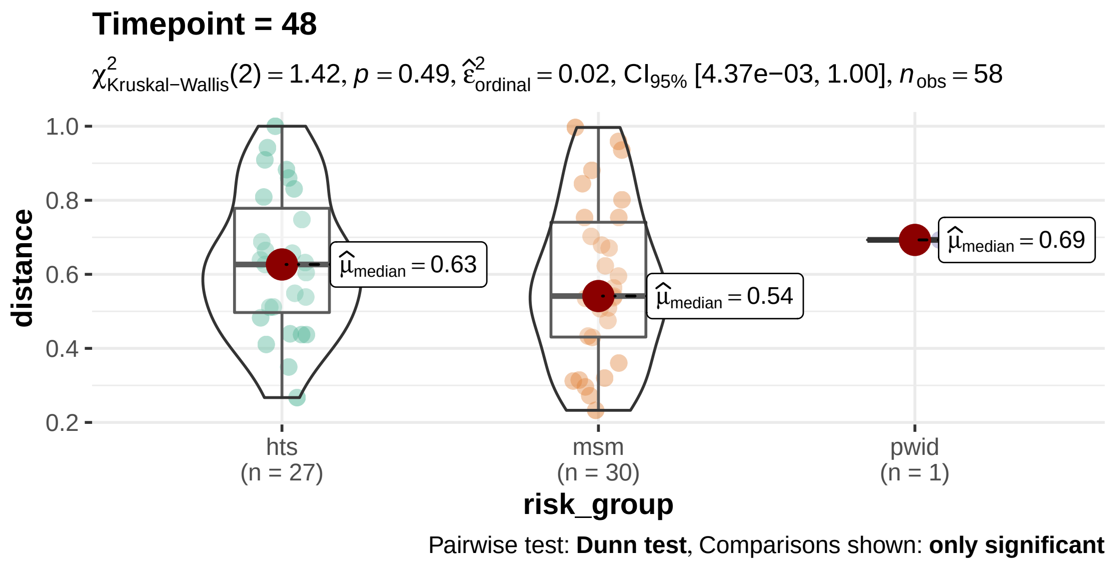
```
#> $group
```

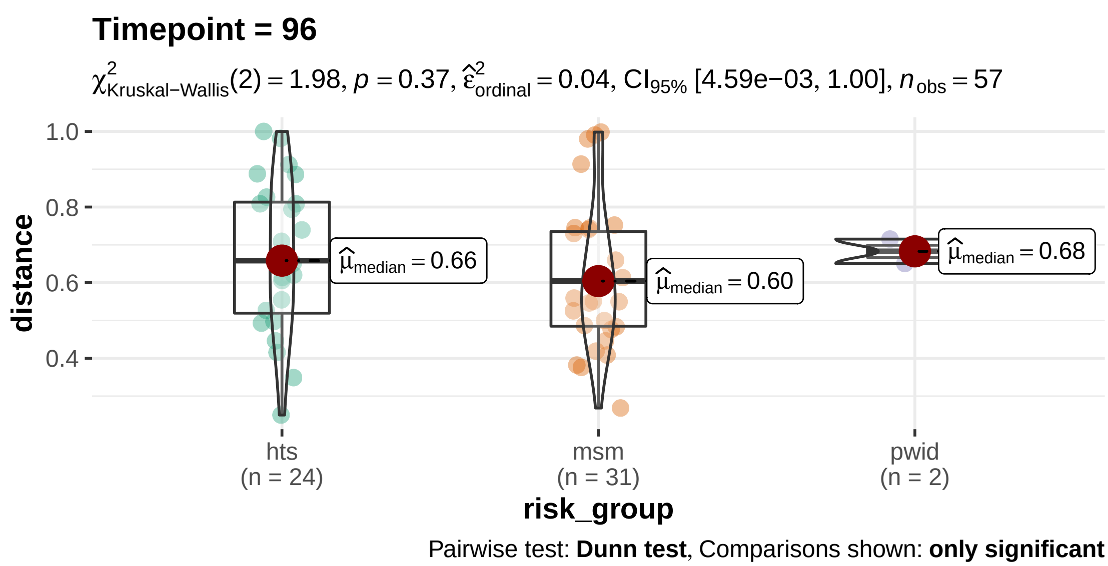



```
#>
```

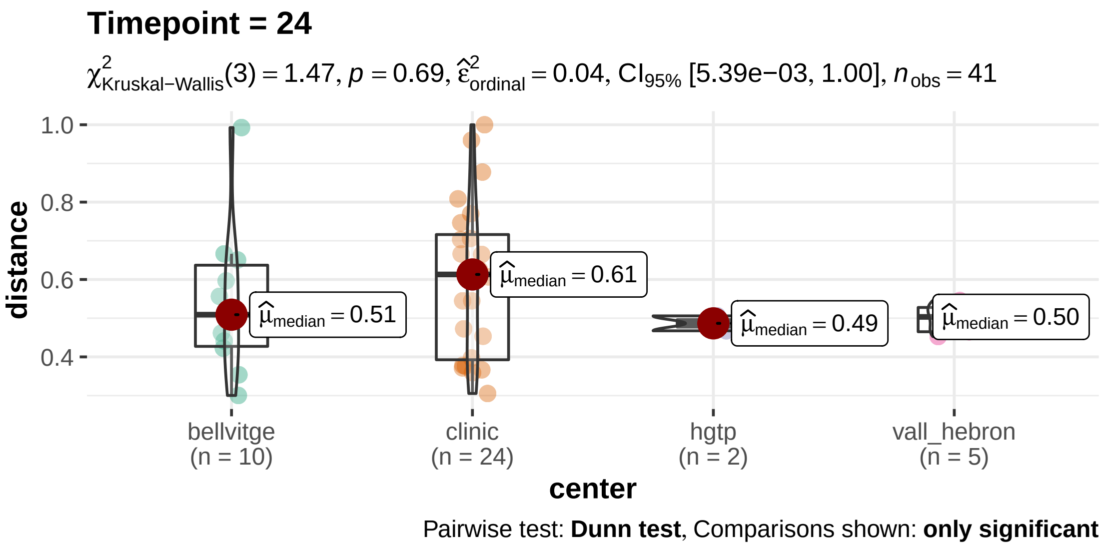
#> $risk_group

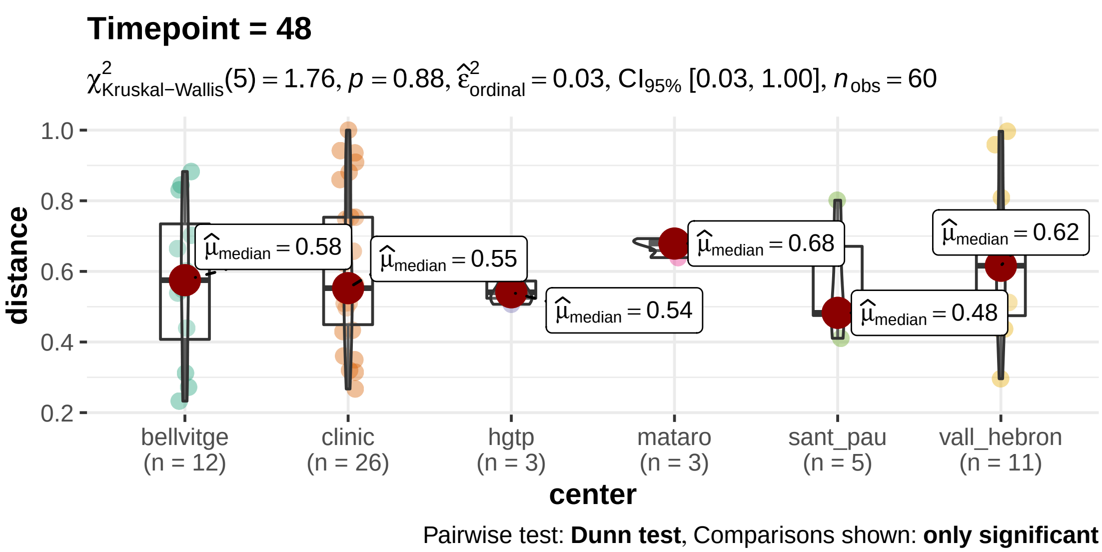


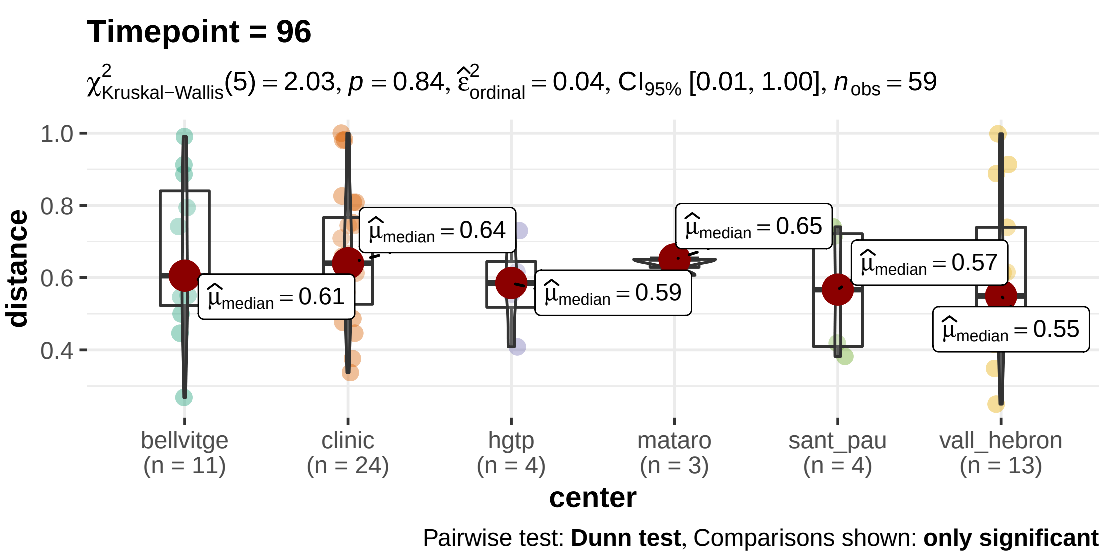




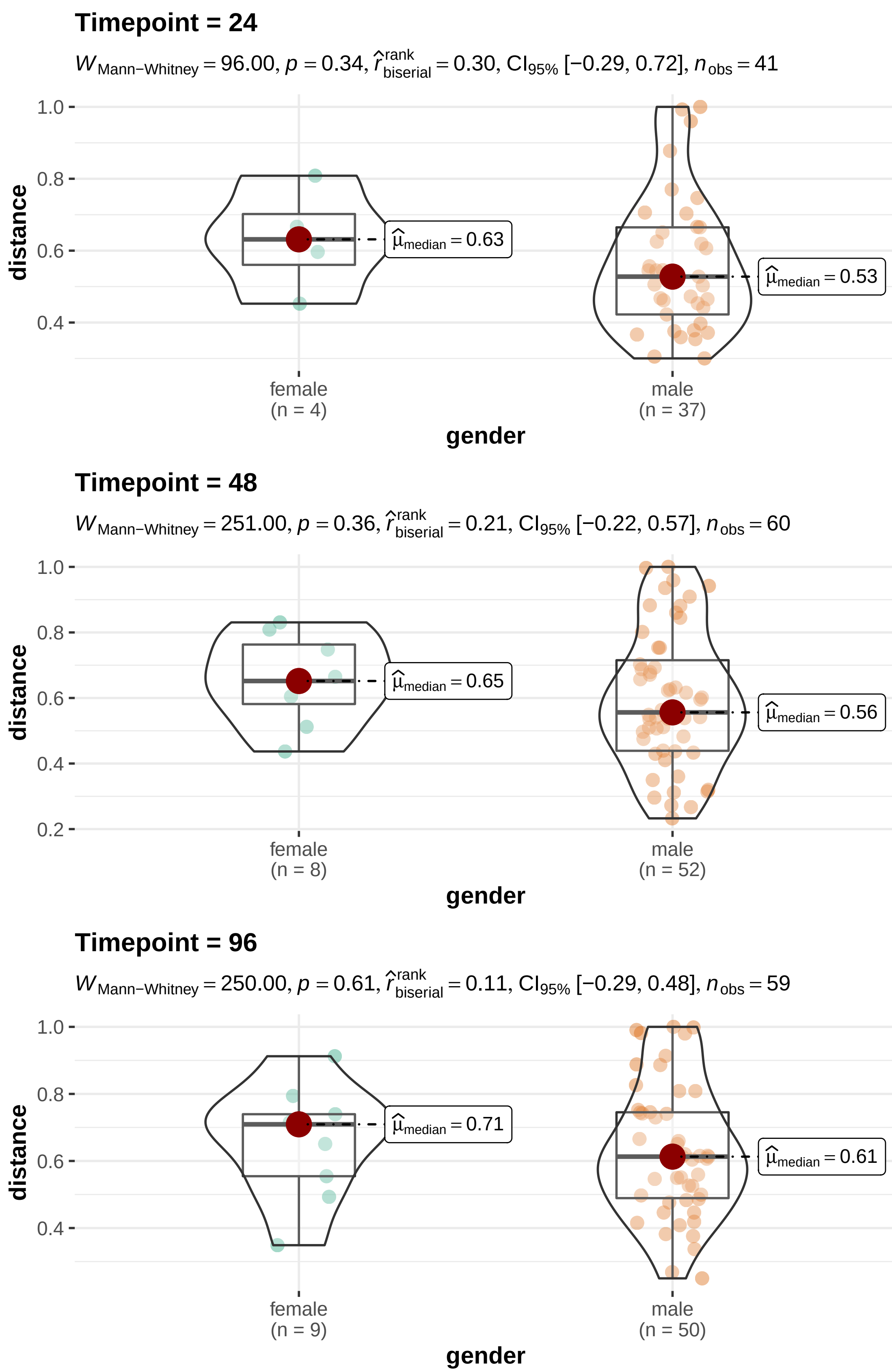
#>
#> $center


Timepoint = 24
$\chi^2_{\text{Kruskal-Wallis}}(3) = 1.47, p = 0.69, \hat{\varepsilon}^2_{\text{ordinal}} = 0.04, \text{CI}_{95\%}$ [5.39e−03, 1.00], $n_{\text{obs}} = 41$
distance
1.0
0.8
0.6
0.4
$\hat{\mu}_{\text{median}} = 0.51$
$\hat{\mu}_{\text{median}} = 0.61$
$\hat{\mu}_{\text{median}} = 0.49$
$\hat{\mu}_{\text{median}} = 0.50$
bellvitge (n = 10)
clinic (n = 24)
hgtp (n = 2)
vall_hebron (n = 5)
center
Pairwise test: **Dunn test**, Comparisons shown: **only significant**

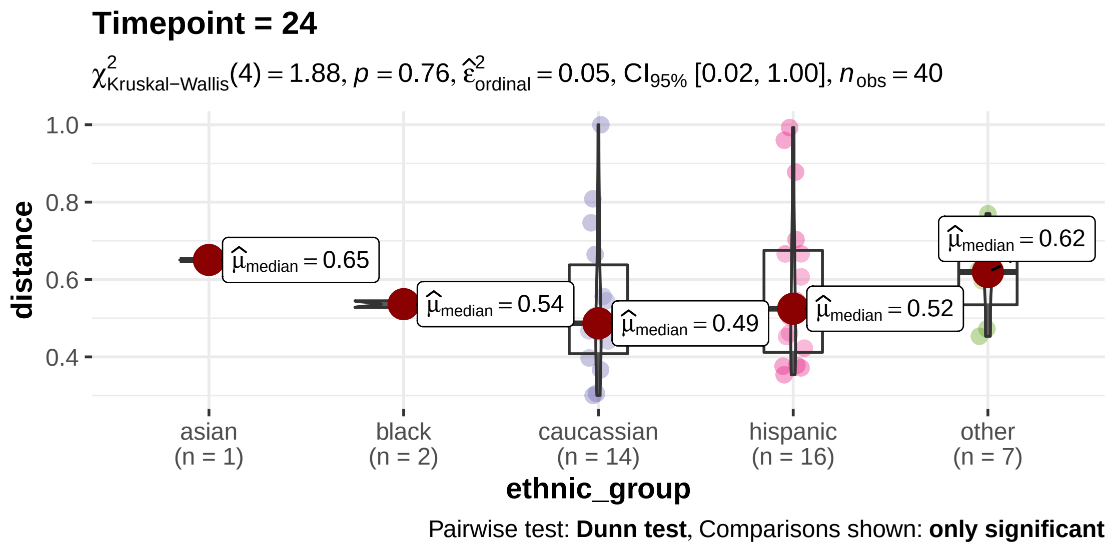

Timepoint = 48
$\chi^2_{\text{Kruskal-Wallis}}(5) = 1.76, p = 0.88, \hat{\varepsilon}^2_{\text{ordinal}} = 0.03, \text{CI}_{95\%}$ [0.03, 1.00], $n_{\text{obs}} = 60$
distance
1.0
0.8
0.6
0.4
0.2
$\hat{\mu}_{\text{median}} = 0.58$
$\hat{\mu}_{\text{median}} = 0.55$
$\hat{\mu}_{\text{median}} = 0.54$
$\hat{\mu}_{\text{median}} = 0.68$
$\hat{\mu}_{\text{median}} = 0.48$
$\hat{\mu}_{\text{median}} = 0.62$
bellvitge (n = 12)
clinic (n = 26)
hgtp (n = 3)
mataro (n = 3)
sant_pau (n = 5)
vall_hebron (n = 11)
center
Pairwise test: **Dunn test**, Comparisons shown: **only significant**

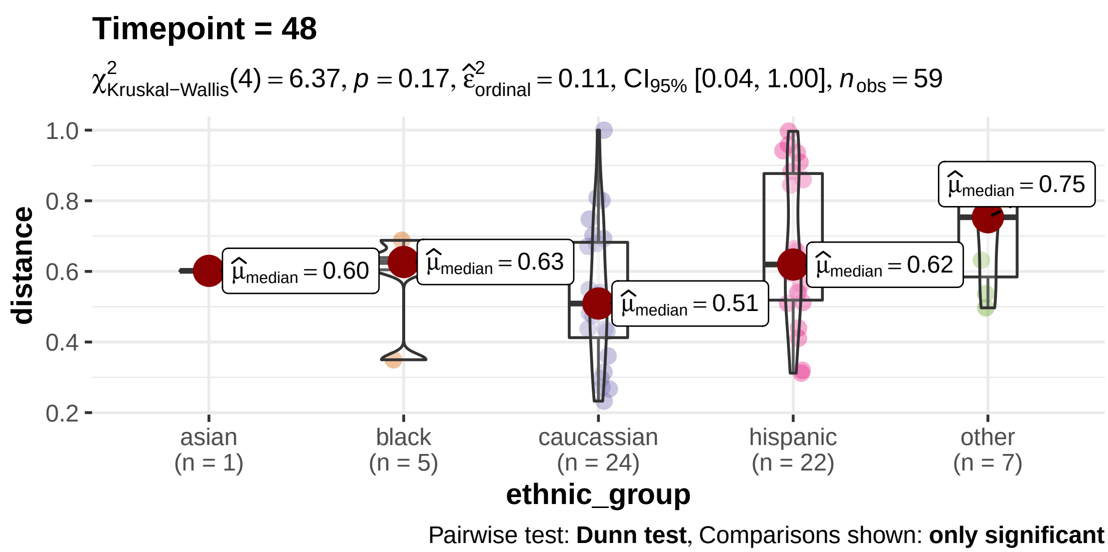

Timepoint = 96
$\chi^2_{\text{Kruskal-Wallis}}(5) = 2.03, p = 0.84, \hat{\varepsilon}^2_{\text{ordinal}} = 0.04, \text{CI}_{95\%}$ [0.01, 1.00], $n_{\text{obs}} = 59$
distance
1.0
0.8
0.6
0.4
$\hat{\mu}_{\text{median}} = 0.61$
$\hat{\mu}_{\text{median}} = 0.64$
$\hat{\mu}_{\text{median}} = 0.59$
$\hat{\mu}_{\text{median}} = 0.65$
$\hat{\mu}_{\text{median}} = 0.57$
$\hat{\mu}_{\text{median}} = 0.55$
bellvitge (n = 11)
clinic (n = 24)
hgtp (n = 4)
mataro (n = 3)
sant_pau (n = 4)
vall_hebron (n = 13)
center
Pairwise test: **Dunn test**, Comparisons shown: **only significant**

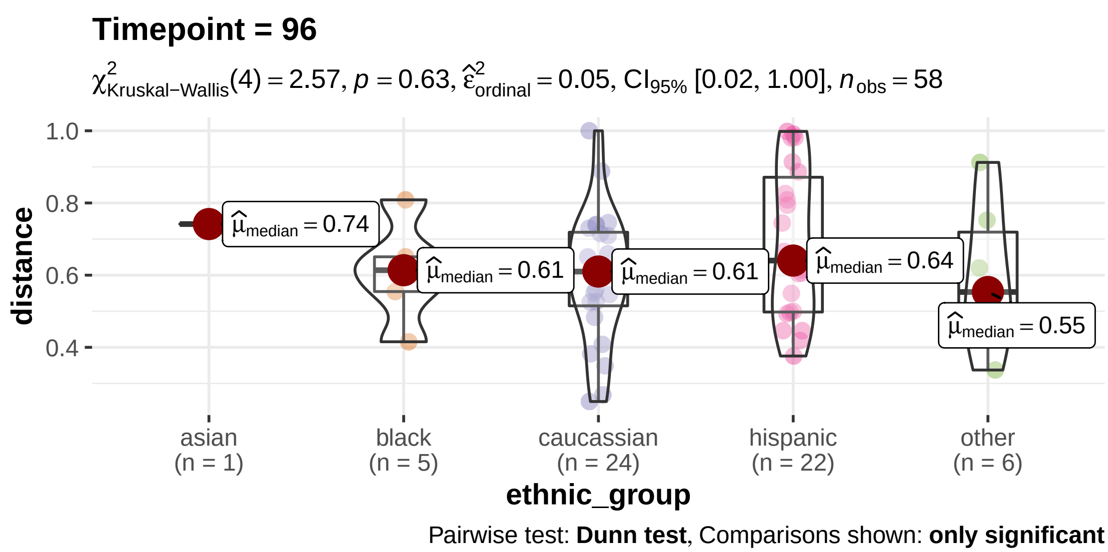
```
#>
#> $gender
```



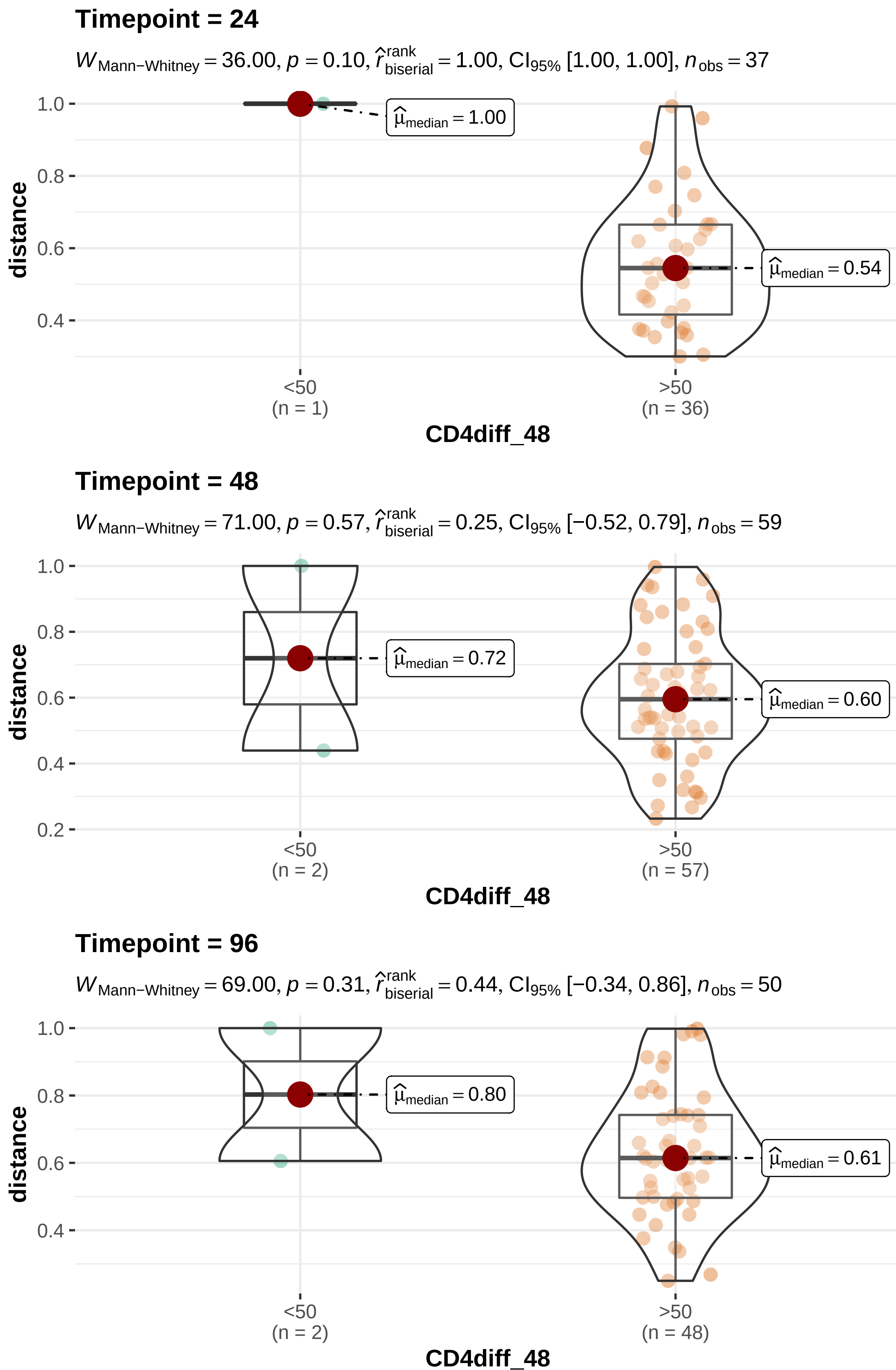

```
#>
#> $ethnic_group
```


Timepoint = 24
χ²Kruskal−Wallis(4) = 1.88, p = 0.76, ε̂²ordinal = 0.05, CI95% [0.02, 1.00], nobs = 40
distance
1.0
0.8
0.6
0.4
μ̂median = 0.65
μ̂median = 0.54
μ̂median = 0.49
μ̂median = 0.52
μ̂median = 0.62
asian
(n = 1)
black
(n = 2)
caucassian
(n = 14)
hispanic
(n = 16)
other
(n = 7)
ethnic_group
Pairwise test: Dunn test, Comparisons shown: only significant

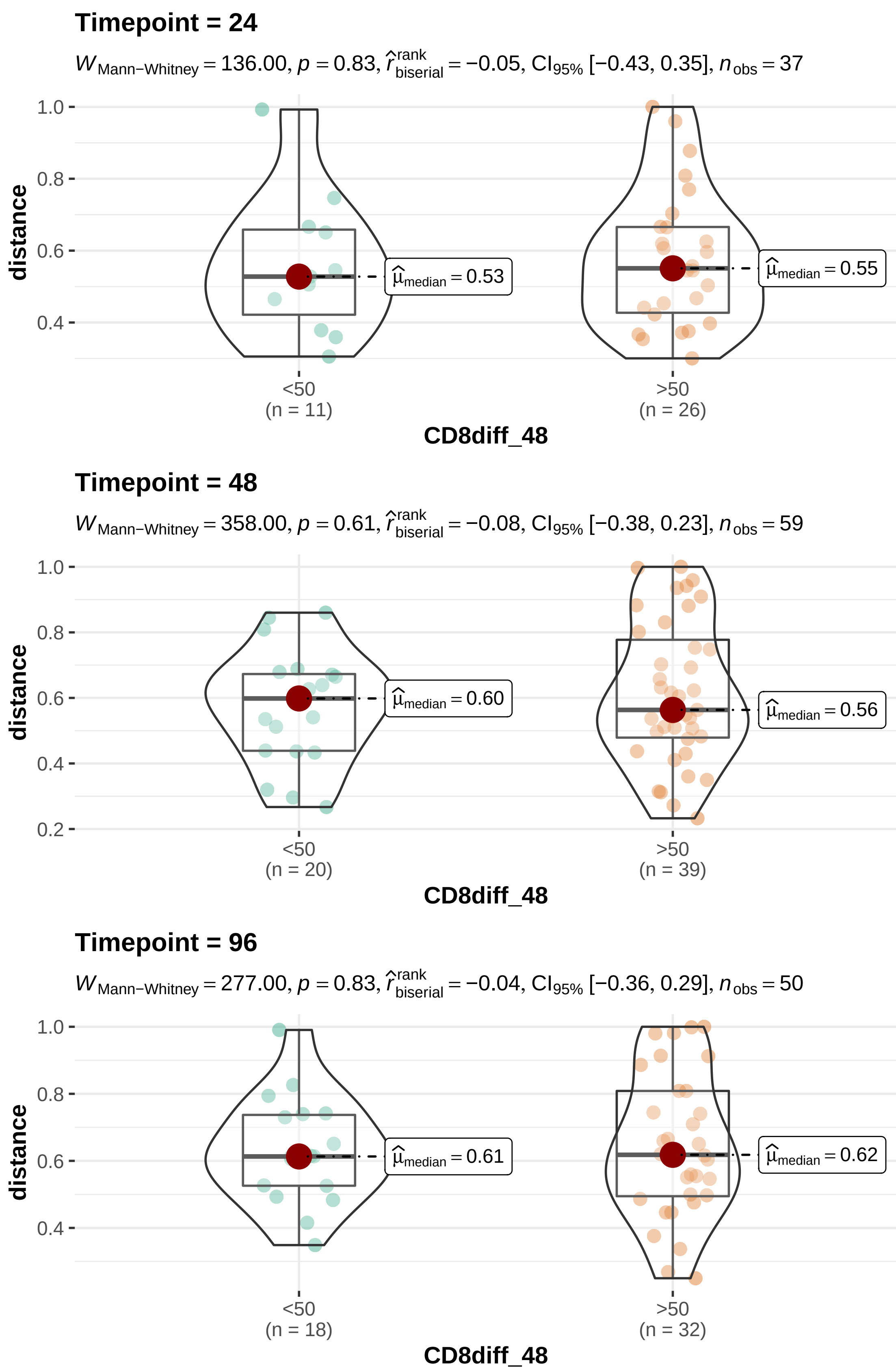

Timepoint = 48
χ²Kruskal−Wallis(4) = 6.37, p = 0.17, ε̂²ordinal = 0.11, CI95% [0.04, 1.00], nobs = 59
distance
1.0
0.8
0.6
0.4
0.2
μ̂median = 0.60
μ̂median = 0.63
μ̂median = 0.51
μ̂median = 0.62
μ̂median = 0.75
asian
(n = 1)
black
(n = 5)
caucassian
(n = 24)
hispanic
(n = 22)
other
(n = 7)
ethnic_group
Pairwise test: Dunn test, Comparisons shown: only significant


Timepoint = 96
χ²Kruskal−Wallis(4) = 2.57, p = 0.63, ε̂²ordinal = 0.05, CI95% [0.02, 1.00], nobs = 58
distance
1.0
0.8
0.6
0.4
μ̂median = 0.74
μ̂median = 0.61
μ̂median = 0.61
μ̂median = 0.64
μ̂median = 0.55
asian
(n = 1)
black
(n = 5)
caucassian
(n = 24)
hispanic
(n = 22)
other
(n = 6)
ethnic_group
Pairwise test: Dunn test, Comparisons shown: only significant

```
#>
#> $CD4diff_48
```


Timepoint = 24
W Mann−Whitney = 36.00, p = 0.10, r̂ rank biserial = 1.00, CI95% [1.00, 1.00], n obs = 37
μ̂ median = 1.00
μ̂ median = 0.54
distance
1.0
0.8
0.6
0.4
<50
(n = 1)
>50
(n = 36)
CD4diff_48
Timepoint = 48
W Mann−Whitney = 71.00, p = 0.57, r̂ rank biserial = 0.25, CI95% [−0.52, 0.79], n obs = 59
μ̂ median = 0.72
μ̂ median = 0.60
distance
1.0
0.8
0.6
0.4
0.2
<50
(n = 2)
>50
(n = 57)
CD4diff_48
Timepoint = 96
W Mann−Whitney = 69.00, p = 0.31, r̂ rank biserial = 0.44, CI95% [−0.34, 0.86], n obs = 50
μ̂ median = 0.80
μ̂ median = 0.61
distance
1.0
0.8
0.6
0.4
<50
(n = 2)
>50
(n = 48)
CD4diff_48

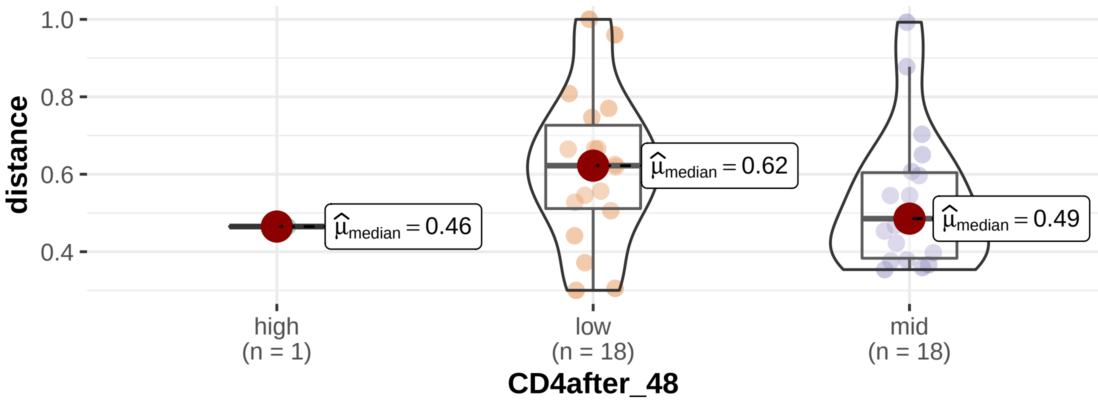
```
#>
#> $CD8diff_48
```


Timepoint = 24
W Mann−Whitney = 136.00, p = 0.83, r̂ rank biserial = −0.05, CI95% [−0.43, 0.35], n obs = 37
μ̂ median = 0.53
μ̂ median = 0.55
distance
1.0
0.8
0.6
0.4
<50
(n = 11)
>50
(n = 26)
CD8diff_48
Timepoint = 48
W Mann−Whitney = 358.00, p = 0.61, r̂ rank biserial = −0.08, CI95% [−0.38, 0.23], n obs = 59
μ̂ median = 0.60
μ̂ median = 0.56
distance
1.0
0.8
0.6
0.4
0.2
<50
(n = 20)
>50
(n = 39)
CD8diff_48
Timepoint = 96
W Mann−Whitney = 277.00, p = 0.83, r̂ rank biserial = −0.04, CI95% [−0.36, 0.29], n obs = 50
μ̂ median = 0.61
μ̂ median = 0.62
distance
1.0
0.8
0.6
0.4
<50
(n = 18)
>50
(n = 32)
CD8diff_48

```
#>
#> $CD4after_48
```

Timepoint = 24

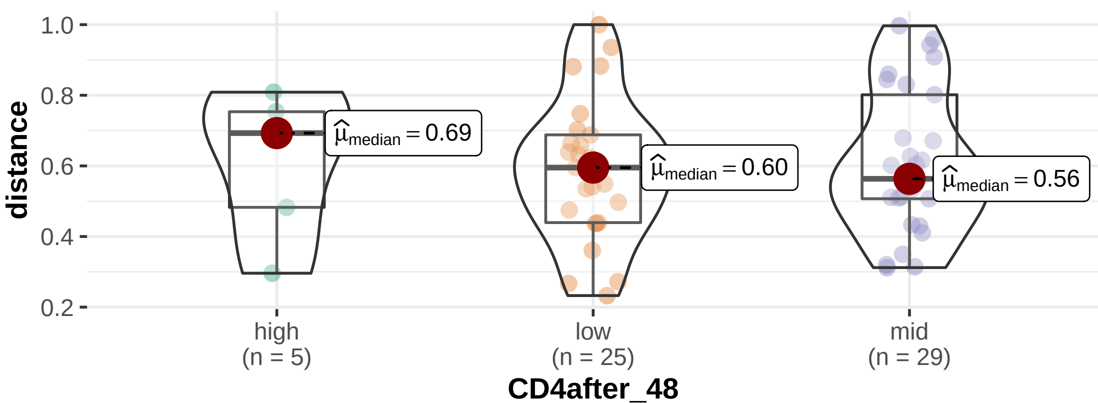
$\chi^2_{\text{Kruskal-Wallis}}(2) = 2.69$, $p = 0.26$, $\widehat{\varepsilon}^2_{\text{ordinal}} = 0.07$, $\text{CI}_{95\%}$ [6.93e−03, 1.00], $n_{\text{obs}} = 37$



Pairwise test: **Dunn test**, Comparisons shown: **only significant**

Timepoint = 48

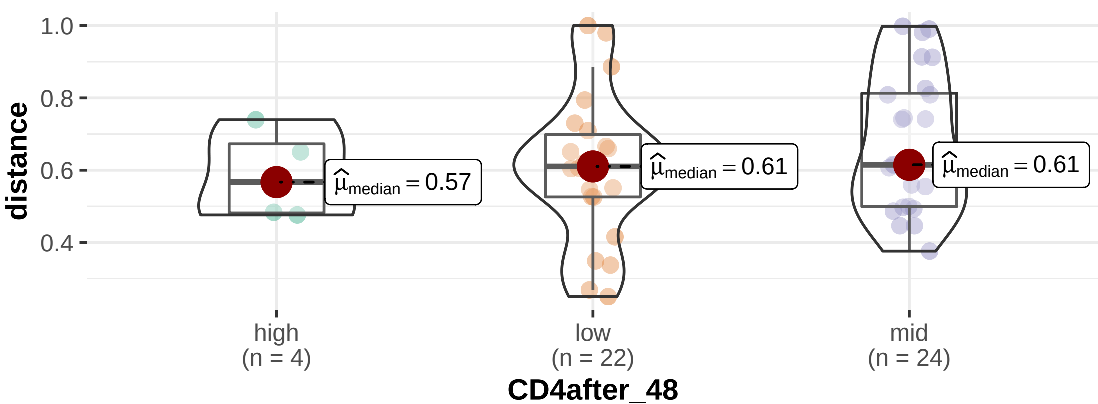
$\chi^2_{\text{Kruskal-Wallis}}(2) = 0.10$, $p = 0.95$, $\widehat{\varepsilon}^2_{\text{ordinal}} = 1.71\text{e}{-03}$, $\text{CI}_{95\%}$ [1.52e−03, 1.00], $n_{\text{obs}} = 59$



Pairwise test: **Dunn test**, Comparisons shown: **only significant**

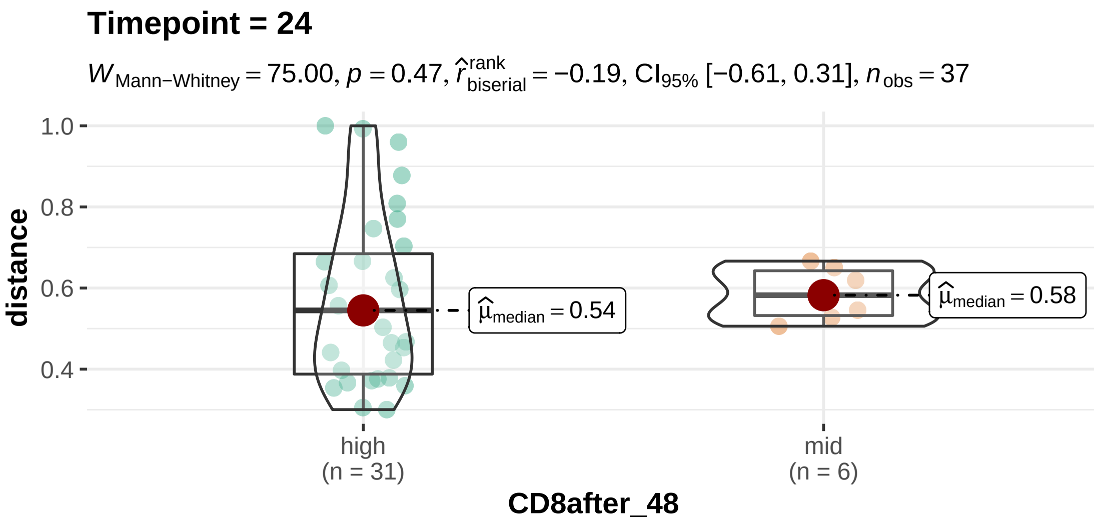
Timepoint = 96

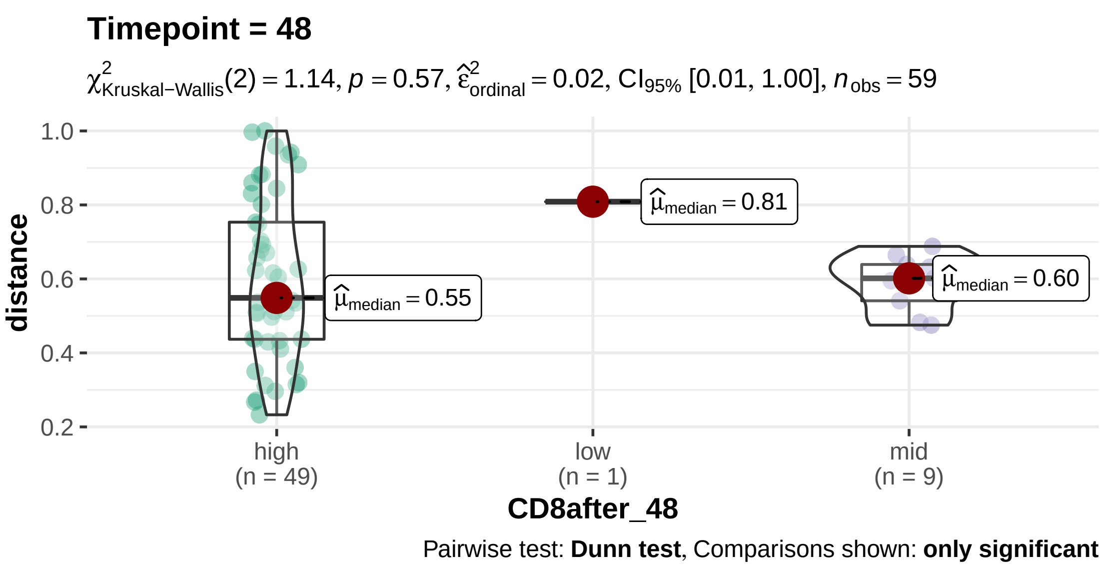
$\chi^2_{\text{Kruskal-Wallis}}(2) = 1.62$, $p = 0.45$, $\widehat{\varepsilon}^2_{\text{ordinal}} = 0.03$, $\text{CI}_{95\%}$ [4.56e−03, 1.00], $n_{\text{obs}} = 50$

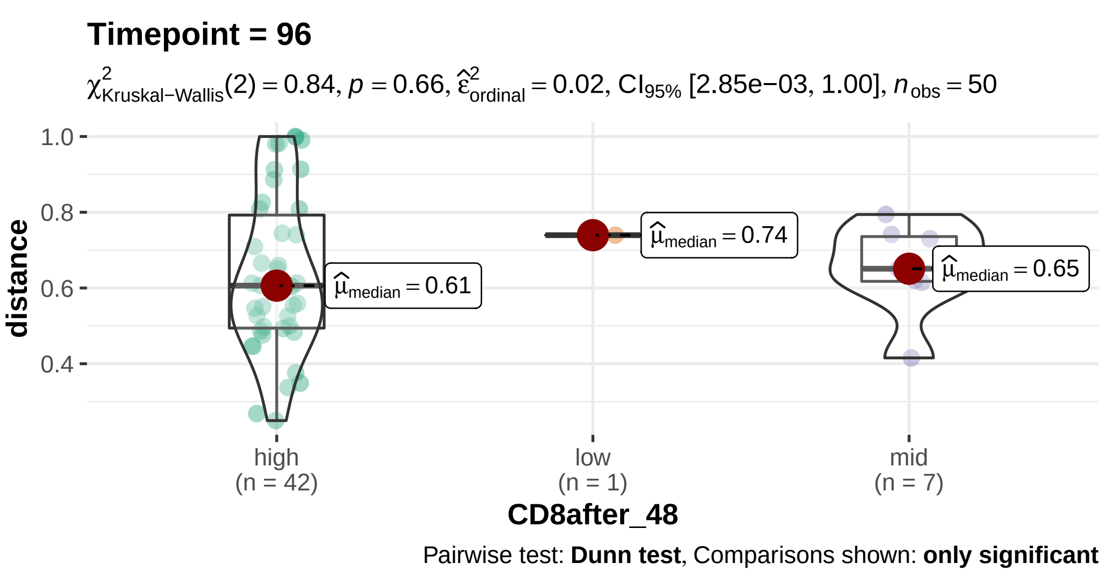


Pairwise test: **Dunn test**, Comparisons shown: **only significant**

```
#>
#> $CD8after_48
```

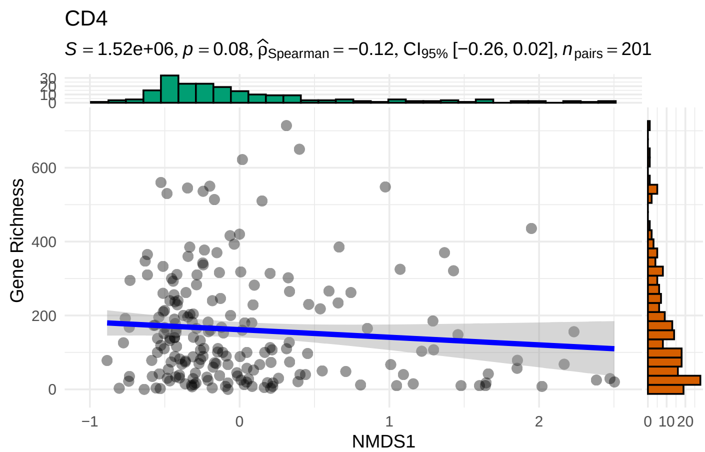

Timepoint = 24
$W_{Mann-Whitney} = 75.00$, $p = 0.47$, $\hat{r}_{biserial}^{rank} = -0.19$, $CI_{95\%}$ [−0.61, 0.31], $n_{obs} = 37$
distance
1.0
0.8
0.6
0.4
$\hat{\mu}_{median} = 0.54$
$\hat{\mu}_{median} = 0.58$
high
(n = 31)
mid
(n = 6)
CD8after_48

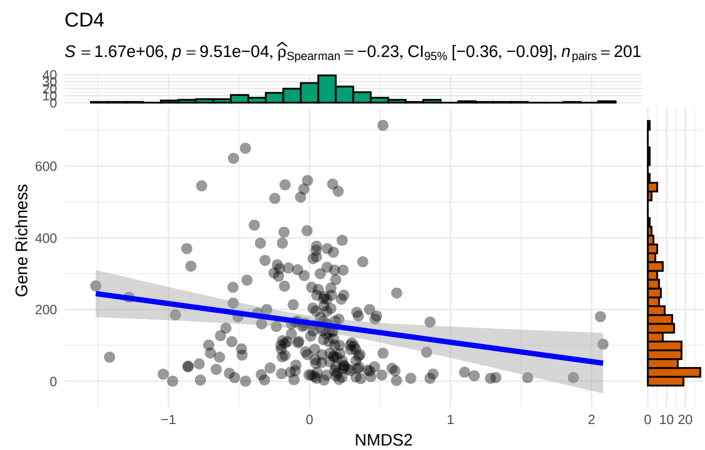

Timepoint = 48
$\chi^2_{Kruskal-Wallis}(2) = 1.14$, $p = 0.57$, $\hat{\varepsilon}^2_{ordinal} = 0.02$, $CI_{95\%}$ [0.01, 1.00], $n_{obs} = 59$
distance
1.0
0.8
0.6
0.4
0.2
$\hat{\mu}_{median} = 0.55$
$\hat{\mu}_{median} = 0.81$
$\hat{\mu}_{median} = 0.60$
high
(n = 49)
low
(n = 1)
mid
(n = 9)
CD8after_48
Pairwise test: **Dunn test**, Comparisons shown: **only significant**

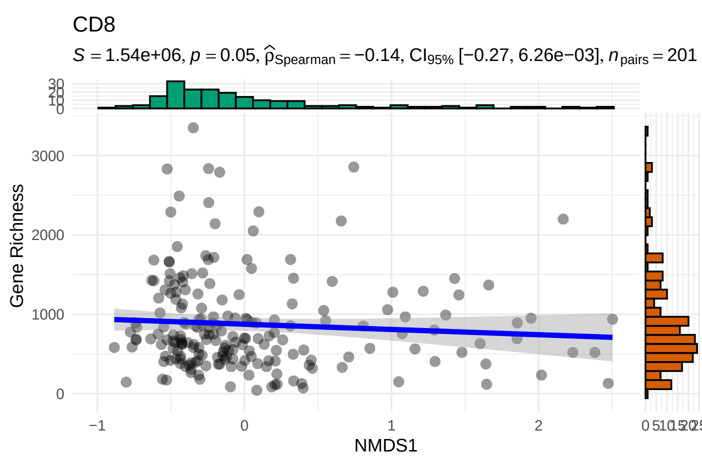

Timepoint = 96
$\chi^2_{Kruskal-Wallis}(2) = 0.84$, $p = 0.66$, $\hat{\varepsilon}^2_{ordinal} = 0.02$, $CI_{95\%}$ [2.85e−03, 1.00], $n_{obs} = 50$
distance
1.0
0.8
0.6
0.4
$\hat{\mu}_{median} = 0.61$
$\hat{\mu}_{median} = 0.74$
$\hat{\mu}_{median} = 0.65$
high
(n = 42)
low
(n = 1)
mid
(n = 7)
CD8after_48
Pairwise test: **Dunn test**, Comparisons shown: **only significant**

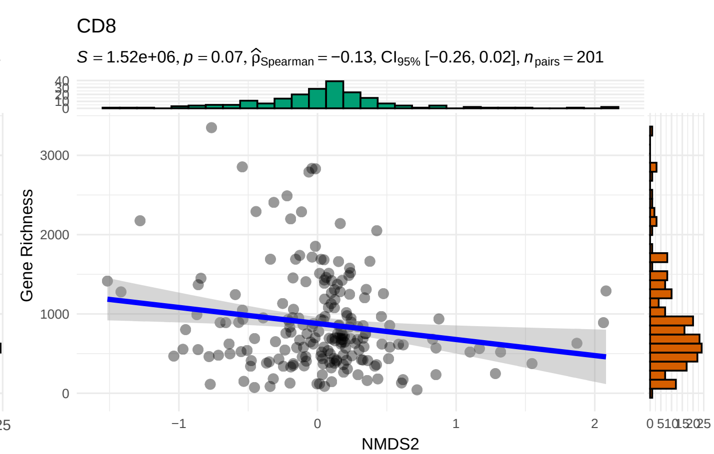
Ordination Analysis (Non-metric multidimensional scaling) by Numeric Variables.

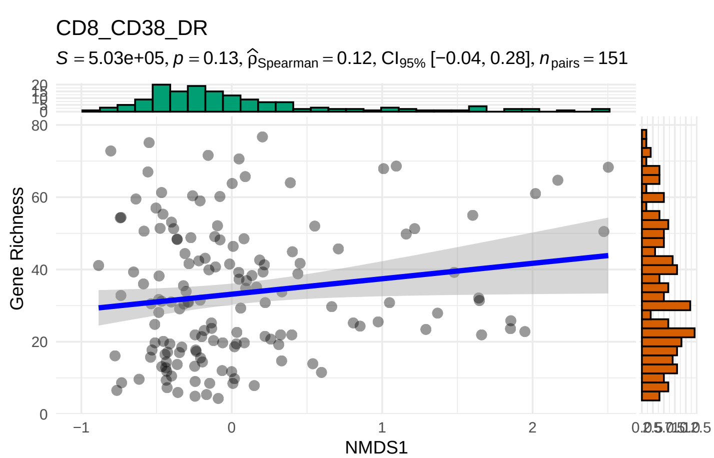
Correlation analysis between NMDS components (NMDS1 and NMDS2) and numeric variables. Plots were produced using the ggstatsplot package. The upper text presents information on inferential statistics and the bottom one provides information about Bayesian hypothesis-testing and estimation.

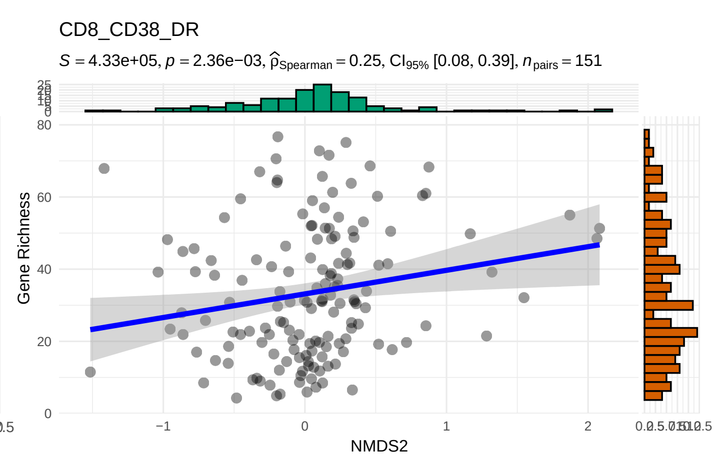


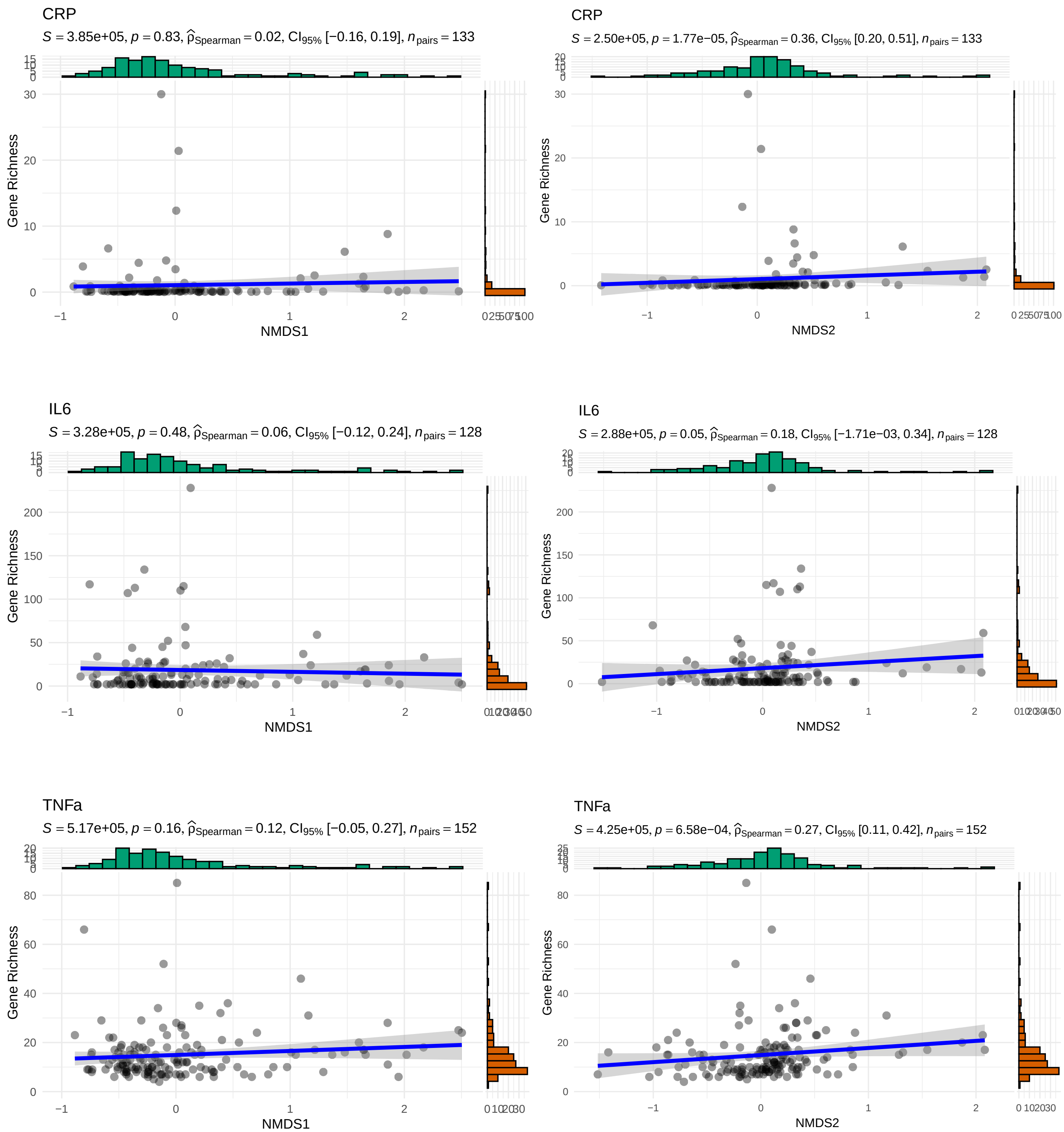




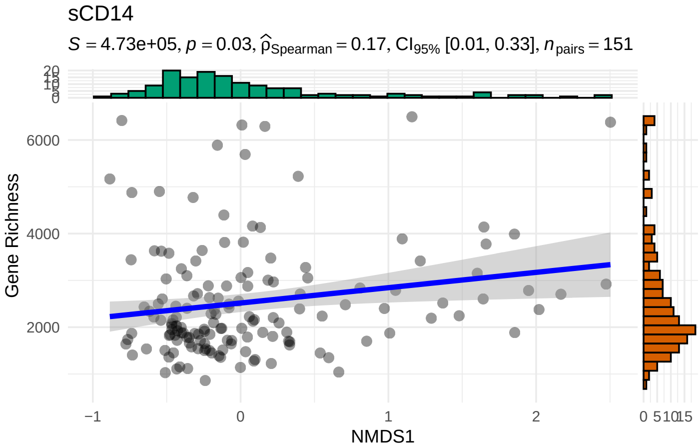




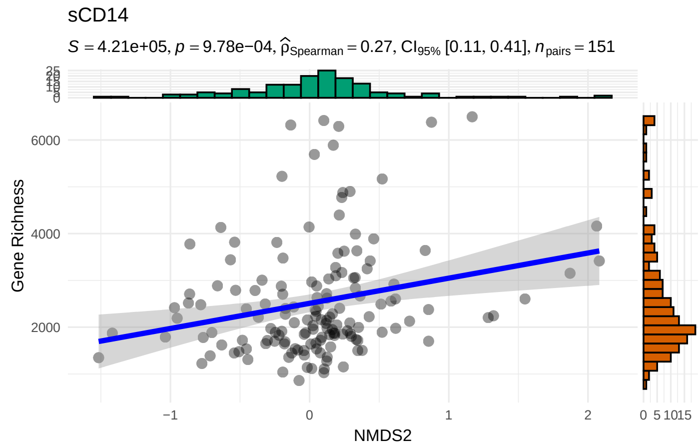




CRP
$S = 3.85\text{e}{+}05$, $p = 0.83$, $\widehat{\rho}_{\text{Spearman}} = 0.02$, $\text{CI}_{95\%}$ [−0.16, 0.19], $n_{\text{pairs}} = 133$
Gene Richness
NMDS1
CRP
$S = 2.50\text{e}{+}05$, $p = 1.77\text{e}{-}05$, $\widehat{\rho}_{\text{Spearman}} = 0.36$, $\text{CI}_{95\%}$ [0.20, 0.51], $n_{\text{pairs}} = 133$
Gene Richness
NMDS2
IL6
$S = 3.28\text{e}{+}05$, $p = 0.48$, $\widehat{\rho}_{\text{Spearman}} = 0.06$, $\text{CI}_{95\%}$ [−0.12, 0.24], $n_{\text{pairs}} = 128$
Gene Richness
NMDS1
IL6
$S = 2.88\text{e}{+}05$, $p = 0.05$, $\widehat{\rho}_{\text{Spearman}} = 0.18$, $\text{CI}_{95\%}$ [−1.71e−03, 0.34], $n_{\text{pairs}} = 128$
Gene Richness
NMDS2
TNFa
$S = 5.17\text{e}{+}05$, $p = 0.16$, $\widehat{\rho}_{\text{Spearman}} = 0.12$, $\text{CI}_{95\%}$ [−0.05, 0.27], $n_{\text{pairs}} = 152$
Gene Richness
NMDS1
TNFa
$S = 4.25\text{e}{+}05$, $p = 6.58\text{e}{-}04$, $\widehat{\rho}_{\text{Spearman}} = 0.27$, $\text{CI}_{95\%}$ [0.11, 0.42], $n_{\text{pairs}} = 152$
Gene Richness
NMDS2



sCD14
$S = 4.73e+05$, $p = 0.03$, $\widehat{\rho}_{Spearman} = 0.17$, $CI_{95\%}$ [0.01, 0.33], $n_{pairs} = 151$
Gene Richness
NMDS1

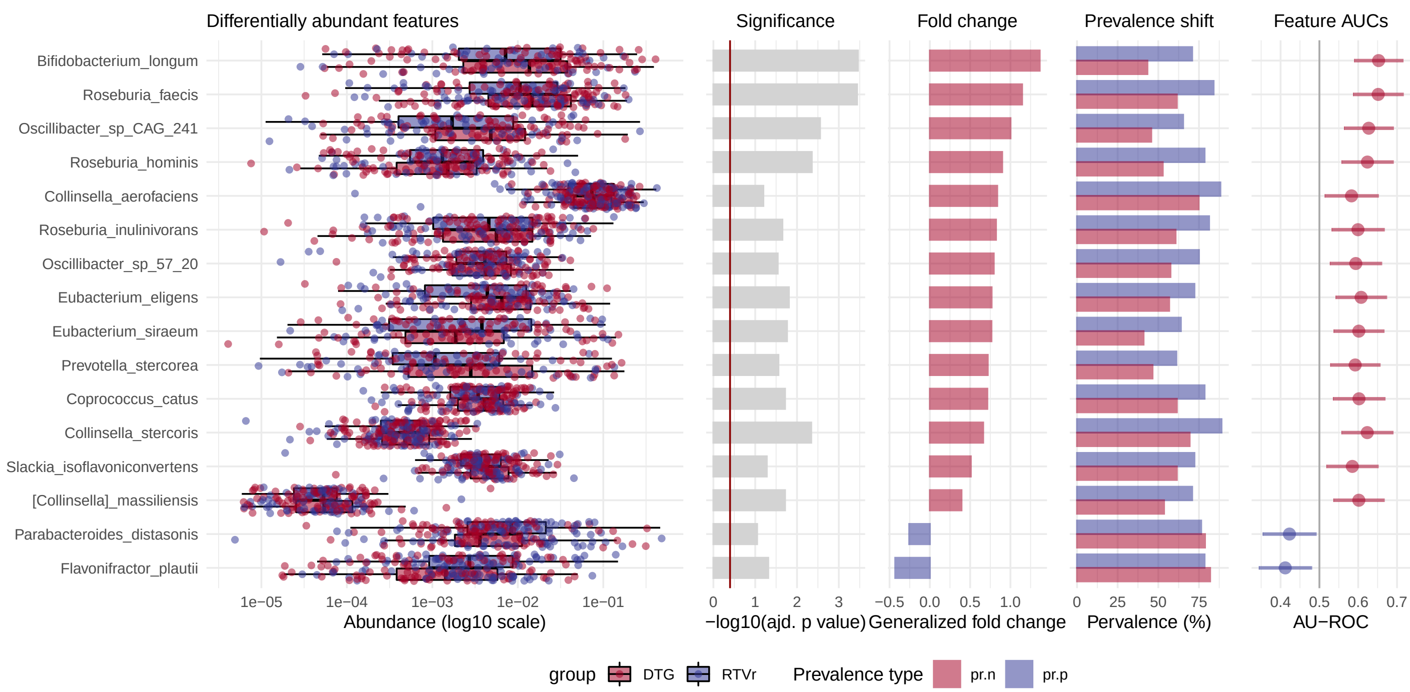


sCD14
$S = 4.21e+05$, $p = 9.78e-04$, $\widehat{\rho}_{Spearman} = 0.27$, $CI_{95\%}$ [0.11, 0.41], $n_{pairs} = 151$
Gene Richness
NMDS2

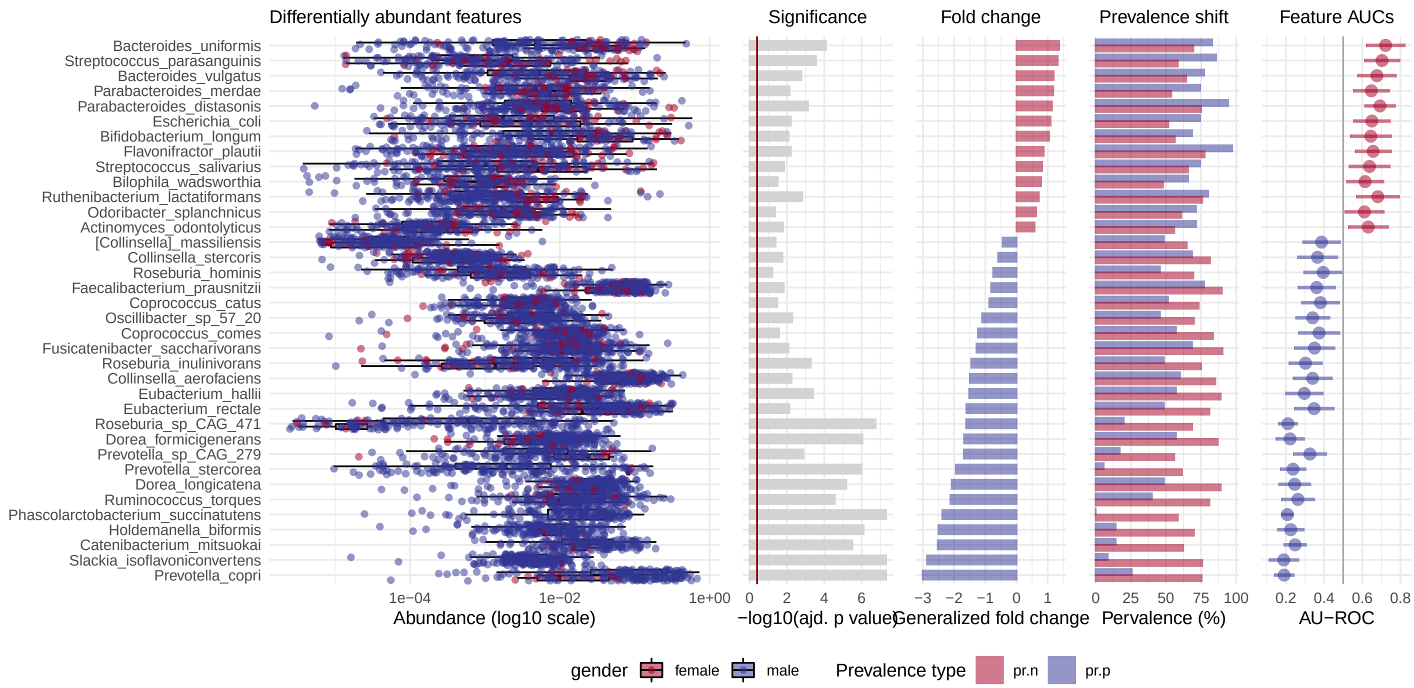
Statistical Inference of Associations between Microbial Communities And host phenoTypes (SIAMCAT).

Detection of changes in community composition that are associated with metadata variables using LASSO logistic regression modeling. For this purpose, the abundance matrix was relativized and expressed in times ones. Feature selection was performed with "prevalence" method removing those features with low prevalence across samples (relative abundance cutoff value set as 0.5). Feature normalization was performed using centred log-ratio ("log.clr") transformation. Finally, for visualization purposes, only those associations with an fdr < 0.05 (Default value) were considered.

```
#> $group
```




```
#>
#> $gender
```



Differentially abundant features
Significance
Fold change
Prevalence shift
Feature AUCs
Bacteroides_uniformis
Streptococcus_parasanguinis
Bacteroides_vulgatus
Parabacteroides_merdae
Parabacteroides_distasonis
Escherichia_coli
Bifidobacterium_longum
Flavonifractor_plautii
Streptococcus_salivarius
Bilophila_wadsworthia
Ruthenibacterium_lactatiformans
Odoribacter_splanchnicus
Actinomyces_odontolyticus
[Collinsella]_massiliensis
Collinsella_stercoris
Roseburia_hominis
Faecalibacterium_prausnitzii
Coprococcus_catus
Oscillibacter_sp_57_20
Coprococcus_comes
Fusicatenibacter_saccharivorans
Roseburia_inulinivorans
Collinsella_aerofaciens
Eubacterium_hallii
Eubacterium_rectale
Roseburia_sp_CAG_471
Dorea_formicigenerans
Prevotella_sp_CAG_279
Prevotella_stercorea
Dorea_longicatena
Ruminococcus_torques
Phascolarctobacterium_succinatutens
Holdemanella_biformis
Catenibacterium_mitsuokai
Slackia_isoflavoniconvertens
Prevotella_copri
1e−04
1e−02
1e+00
Abundance (log10 scale)
0 2 4 6
−log10(ajd. p value)
−3 −2 −1 0 1
Generalized fold change
0 25 50 75 100
Pervalence (%)
0.2 0.4 0.6 0.8
AU−ROC
gender
female
male
Prevalence type
pr.n
pr.p