

Tobacco exposure associated with oral microbiome anaerobiosis in the New York City Health and Nutrition Examination Study (NYC HANES) II

Francesco Beghini^{1*}

¹Laboratory of Computational Metagenomics - Centre for Integrative Biology - University of Trento - Italy

*francesco.beghini@unitn.it (mailto:francesco.beghini@unitn.it)

5 May 2023

Contents

- 1 Table 1: Demographics & Descriptive Statistics
- 2 Microbial composition (8 most abundant genera)
- 3 Alpha diversity
- 4 Serum cotinine vs Smoking status
- 5 Beta Diversity
 - 5.1 Cigarette smokers vs Never smokers
 - 5.2 Cigarette smokers + Never smokers + Former smokers
 - 5.3 Cigarette smokers + Never smokers + Former smokers + Secondhand smokers
 - 5.4 All smoking statuses
 - 5.5 Discrete serum blood Cotinine levels
- 6 Differential analysis
 - 6.1 DESeq2
 - 6.2 edgeR
 - 6.3 Cigarette smoker vs Never smoker: crude phylum level
- 7 Analysis on biosis of bacteria
 - 7.1 Odds ratio smokers vs never smokers
 - 7.2 Cigarette smokers vs Never smokers
 - 7.3 Secondhand smokers

7.4 Continuous cotinine

7.5 Hookah smokers vs Never smokers

7.6 E-cigarette smokers vs Never smokers

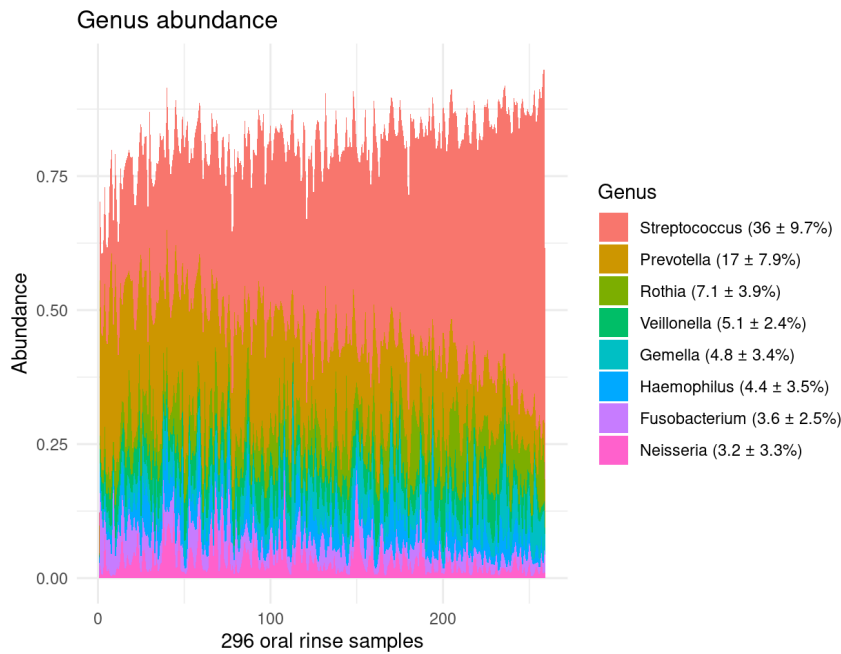
7.7 Cigar/cigarillos smokers vs Never smokers

1 Table 1: Demographics & Descriptive Statistics

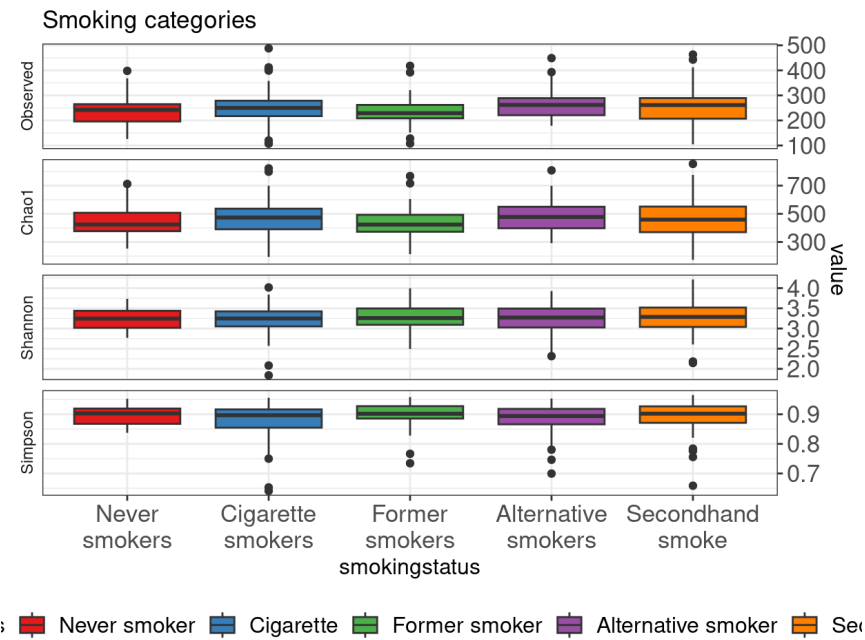
	Never smoker	Cigarette	Former smoker	Alternative smoker	Secondhand
n	43	86	43	49	38
GENDER = Female (%)	28 (65.1)	45 (52.3)	25 (58.1)	22 (44.9)	22 (57.9)
RACE (%)					
Non-Hispanic White	13 (30.2)	24 (27.9)	25 (58.1)	19 (38.8)	10 (26.3)
Non-Hispanic Black	13 (30.2)	33 (38.4)	4 (9.3)	9 (18.4)	11 (28.9)
Hispanic	10 (23.3)	19 (22.1)	10 (23.3)	12 (24.5)	14 (36.8)
Asian	3 (7.0)	9 (10.5)	3 (7.0)	3 (6.1)	2 (5.3)
Other	4 (9.3)	1 (1.2)	1 (2.3)	6 (12.2)	1 (2.6)
EDU4CAT (%)					
College graduate or more	16 (37.2)	18 (20.9)	21 (48.8)	17 (34.7)	9 (23.7)
Less than High school diploma	8 (18.6)	24 (27.9)	4 (9.3)	10 (20.4)	14 (36.8)
High school graduate/GED	7 (16.3)	24 (27.9)	8 (18.6)	11 (22.4)	10 (26.3)
Some College or associate's degree	12 (27.9)	20 (23.3)	10 (23.3)	11 (22.4)	5 (13.2)
SPAGE (mean (SD))	45.42 (16.50)	45.85 (13.07)	55.47 (18.00)	35.59 (16.44)	37.76 (14.70)
AGEGRP5C (%)					
20-29	7 (16.3)	10 (11.6)	3 (7.0)	26 (53.1)	14 (36.8)
30-39	11 (25.6)	17 (19.8)	7 (16.3)	8 (16.3)	11 (28.9)
40-49	10 (23.3)	25 (29.1)	7 (16.3)	4 (8.2)	3 (7.9)
50-59	6 (14.0)	19 (22.1)	8 (18.6)	7 (14.3)	6 (15.8)
60 AND OVER	9 (20.9)	15 (17.4)	18 (41.9)	4 (8.2)	4 (10.5)

	Never smoker	Cigarette	Former smoker	Alternative smoker	Secondhand
DBTS_NEW (%)					
Yes	5 (11.6)	5 (5.8)	7 (16.3)	3 (6.1)	2 (5.3)
No	38 (88.4)	81 (94.2)	36 (83.7)	38 (77.6)	36 (94.7)
NA	0 (0.0)	0 (0.0)	0 (0.0)	8 (16.3)	0 (0.0)
SR_ACTIVE (%)					
Very active	15 (34.9)	27 (31.4)	11 (25.6)	16 (32.7)	17 (44.7)
Somewhat active	20 (46.5)	37 (43.0)	20 (46.5)	25 (51.0)	15 (39.5)
Not very active/not active at all	8 (18.6)	22 (25.6)	12 (27.9)	8 (16.3)	6 (15.8)
INC25KMOD (%)					
Less Than \$20,000	5 (11.6)	31 (36.0)	8 (18.6)	20 (40.8)	14 (36.8)
\$20,000-\$49,999	15 (34.9)	20 (23.3)	9 (20.9)	14 (28.6)	9 (23.7)
\$50,000-\$74,999	6 (14.0)	11 (12.8)	3 (7.0)	6 (12.2)	4 (10.5)
\$75,000-\$99,999	8 (18.6)	4 (4.7)	6 (14.0)	4 (8.2)	2 (5.3)
\$100,000 or More	6 (14.0)	11 (12.8)	13 (30.2)	2 (4.1)	6 (15.8)
NA	3 (7.0)	9 (10.5)	4 (9.3)	3 (6.1)	3 (7.9)
COTININE (median [IQR])	0.04 [0.04, 0.04]	271.49 [189.99, 360.99]	0.04 [0.04, 0.04]	10.54 [0.28, 55.36]	3.01 [1.39, 5.48]
OHQ_3 (%)					
Yes	4 (9.3)	9 (10.5)	5 (11.6)	4 (8.2)	4 (10.5)
No	39 (90.7)	76 (88.4)	38 (88.4)	45 (91.8)	34 (89.5)
NA	0 (0.0)	1 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)

2 Microbial composition (8 most abundant genera)



3 Alpha diversity



```
aovSMOKINGSTATUS <- aov(Shannon ~ smokingstatus, alphadiv)
summary(aovSMOKINGSTATUS)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## smokingstatus  4   0.09  0.02162   0.168  0.954
## Residuals    254  32.62  0.12841
```

```
aovSMOKINGSTATUS <- aov(Observed ~ smokingstatus, alphadiv)
summary(aovSMOKINGSTATUS)
```

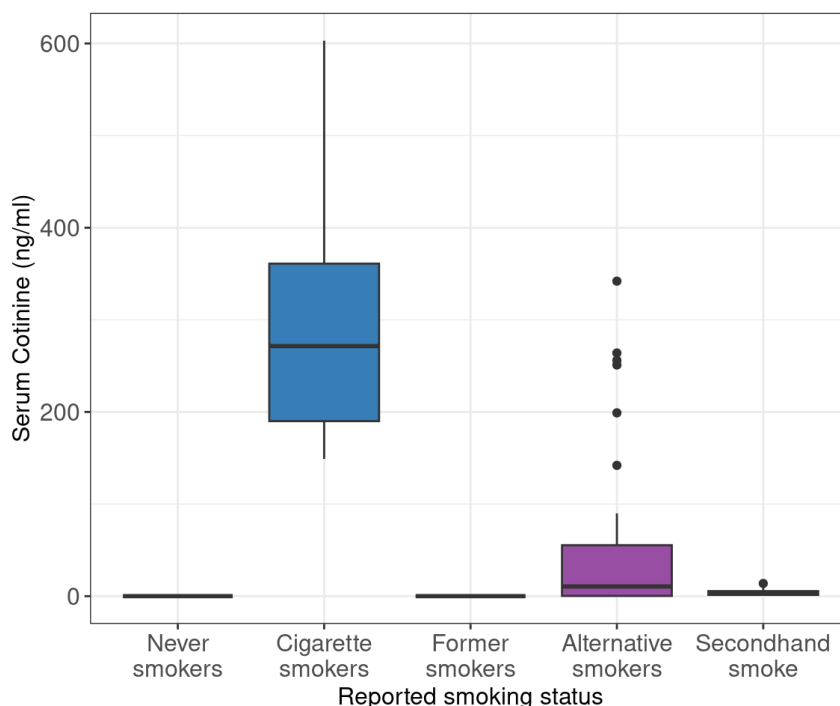
```
##              Df Sum Sq Mean Sq F value Pr(>F)
## smokingstatus  4   34597     8649    2.12 0.0788 .
## Residuals    254 1036314     4080
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

aovSMOKINGSTATUS <- aov(Chao1 ~ smokingstatus, alphadiv)
summary(aovSMOKINGSTATUS)

##              Df Sum Sq Mean Sq F value Pr(>F)
## smokingstatus  4   74507    18627    1.33 0.259
## Residuals    254 3556471    14002
```

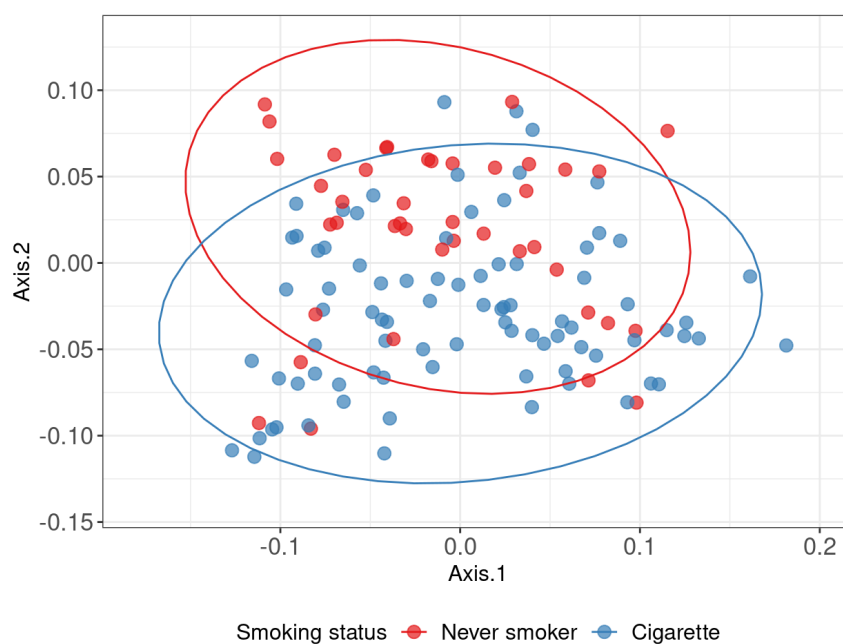
4 Serum cotinine vs Smoking status

```
ggplot(metadata, aes(smokingstatus, COTININE, fill = smokingstatus))
  stat_boxplot() +
  scale_fill_manual(values = scale_palette) +
  theme_bw() +
  scale_x_discrete(labels = c("Never\nsmokers", "Cigarette\nsmokers",
                              "Former\nsmokers", "Alternative\nsmokers",
                              "Secondhand\nsmoke")) +
  xlab('Reported smoking status') +
  ylab('Serum Cotinine (ng/ml)') +
  guides(fill = FALSE) +
  theme(strip.background = element_rect(colour = "transparent", fill
    strip.text.y = element_text(angle = 180),
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 12),
    legend.text = element_text(size=12),
    legend.title = element_text(size=12),
    legend.position = 'bottom')
```

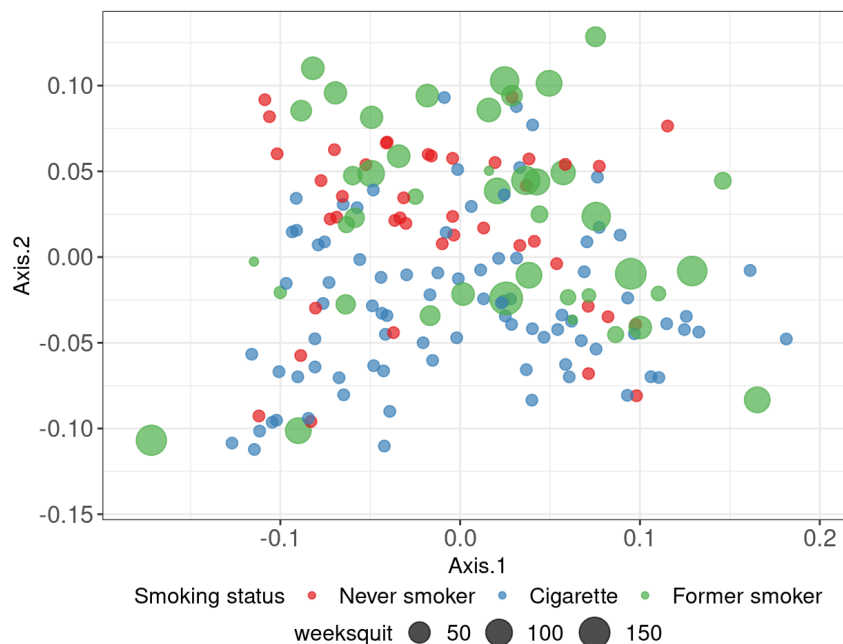


5 Beta Diversity

5.1 Cigarette smokers vs Never smokers

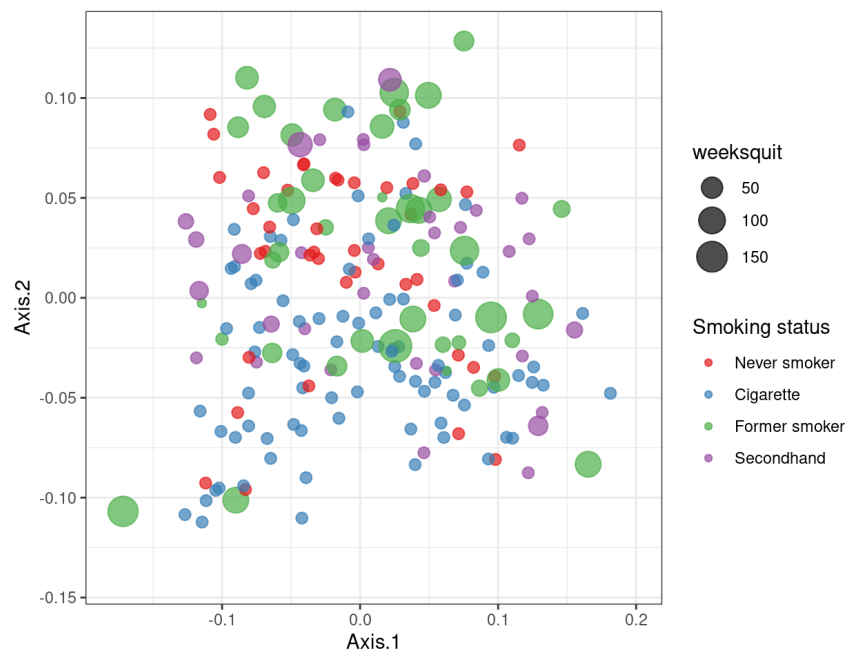


5.2 Cigarette smokers + Never smokers + Former smokers

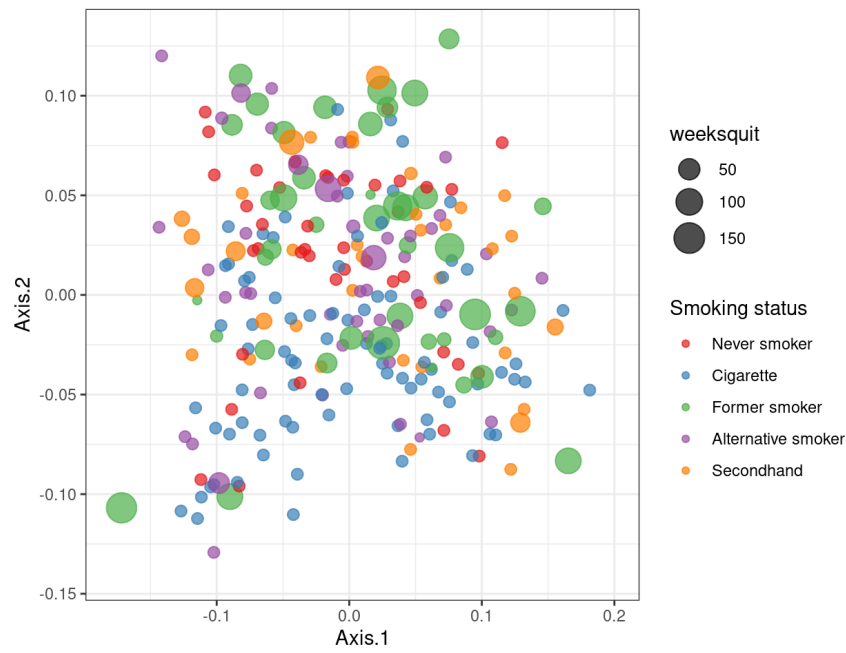


5.3 Cigarette smokers + Never smokers + Former smokers +

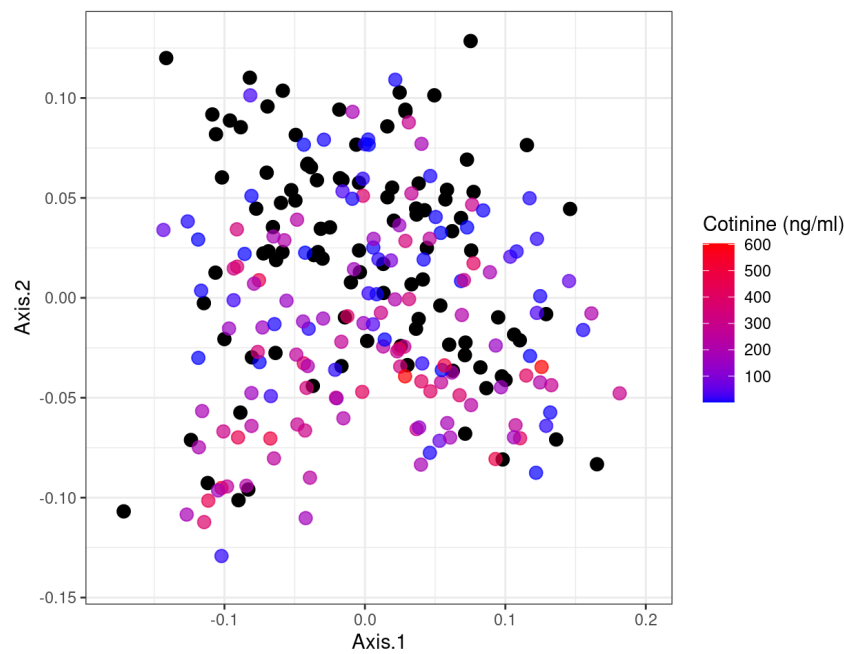
Secondhand smokers



5.4 All smoking statuses



5.5 Discrete serum blood Cotinine levels



5.5.1 Test group with PERMANOVA (adonis, vegan package)

contrast	r2	pvalue
Never smoker vs Cigarette	0.0505876	0.001
Cigarette vs Former smoker	0.0424802	0.001
Cigarette vs Secondhand	0.0349217	0.004
Cigarette vs Alternative smoker	0.0290550	0.008
Never smoker vs Secondhand	0.0230414	0.104
Never smoker vs Former smoker	0.0186827	0.160
Former smoker vs Alternative smoker	0.0143434	0.279
Former smoker vs Secondhand	0.0129641	0.327
Alternative smoker vs Secondhand	0.0091159	0.477
Never smoker vs Alternative smoker	0.0089463	0.510


```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## vegan::adonis2(formula = distwu_2cat ~ COTININE, data = metadata_2
##           Df SumOfSqs      R2      F Pr(>F)
## COTININE   1  0.09755 0.04867 6.4971 0.002 **
## Residual 127  1.90674 0.95133
## Total    128  2.00429 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5.5.1.1 Three categories

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## vegan::adonis2(formula = distwu_3cat ~ smokingstatus + RACE + GEND
##           Df SumOfSqs      R2      F Pr(>F)
## smokingstatus  2  0.13485 0.05649 5.1964 0.001 ***
## RACE           4  0.07660 0.03209 1.4759 0.110
## GENDER         1  0.02369 0.00993 1.8259 0.094 .
## AGEGRP4C       1  0.03184 0.01334 2.4535 0.052 .
## SR_ACTIVE      2  0.02715 0.01138 1.0464 0.354
## EDU3CAT        2  0.03124 0.01309 1.2037 0.256
## DBTS_NEW       1  0.01153 0.00483 0.8884 0.469
## Residual      158  2.05013 0.85886
## Total         171  2.38703 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5.5.1.2 All categories

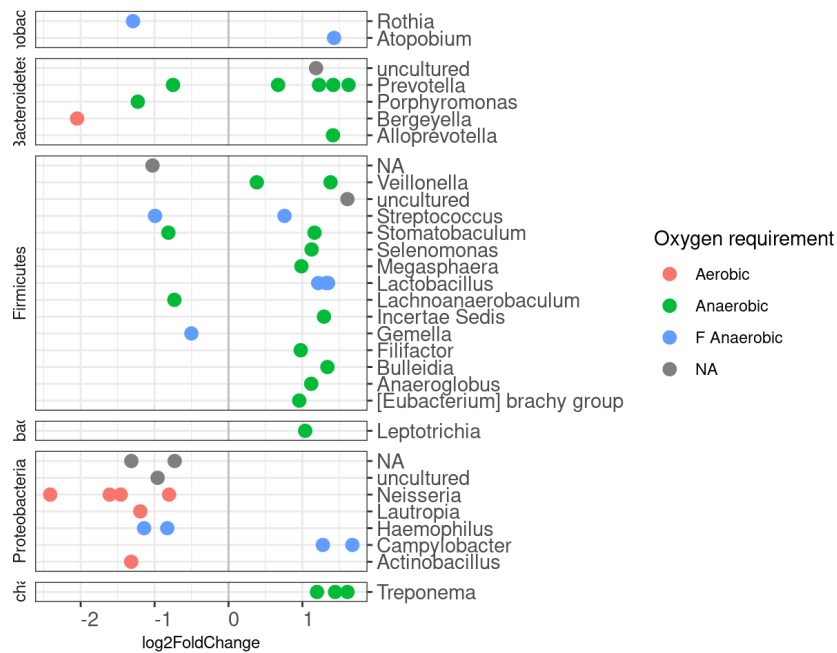
```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## vegan::adonis2(formula = distwu ~ smokingstatus, data = metadata)
##           Df SumOfSqs      R2      F Pr(>F)
## smokingstatus  4  0.1768 0.04838 3.2285 0.001 ***
## Residual      254  3.4775 0.95162
## Total         258  3.6543 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6 Differential analysis

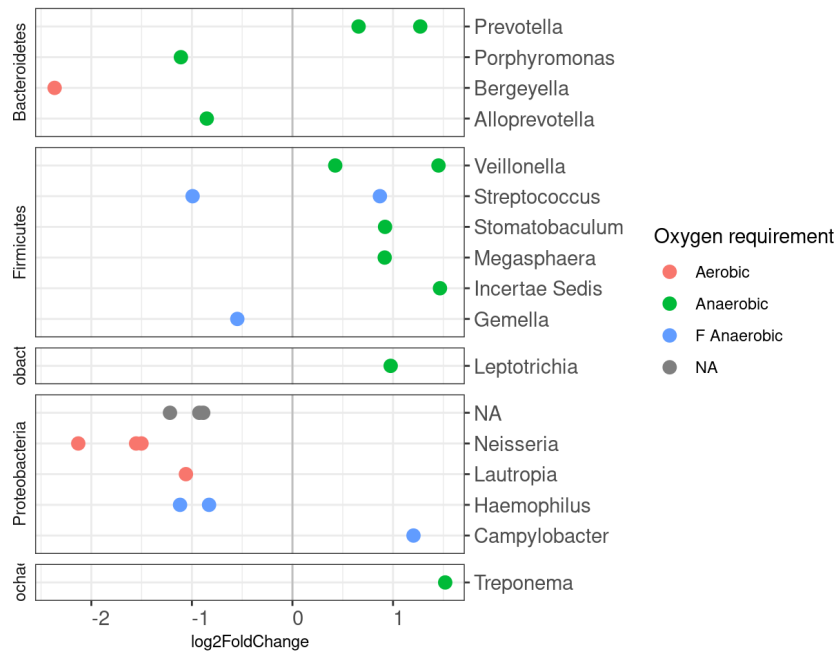
6.1 DESeq2

DESeq2 is presented as sensitivity analysis. Main results are calculated by edgeR.

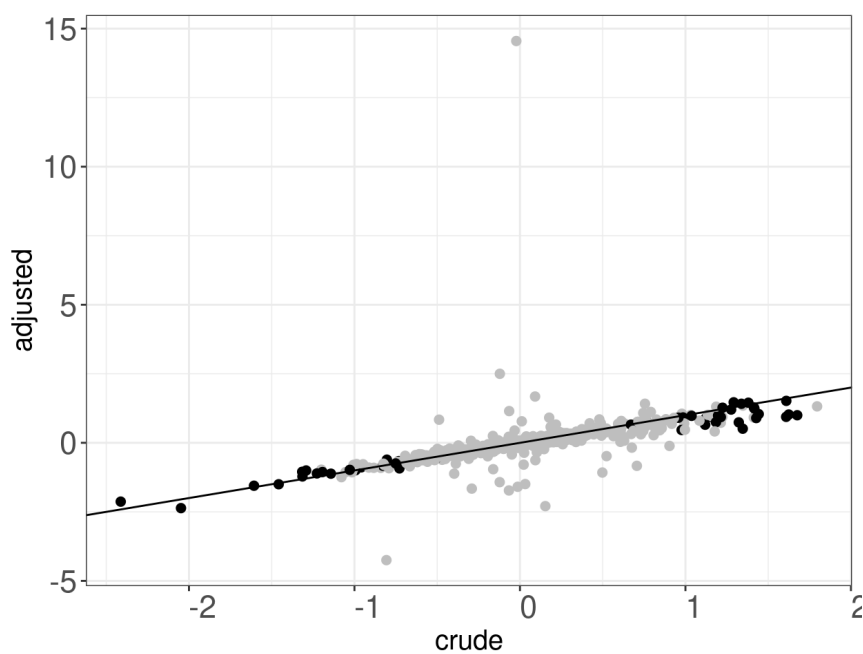
6.1.1 Current smokers vs. never smokers crude



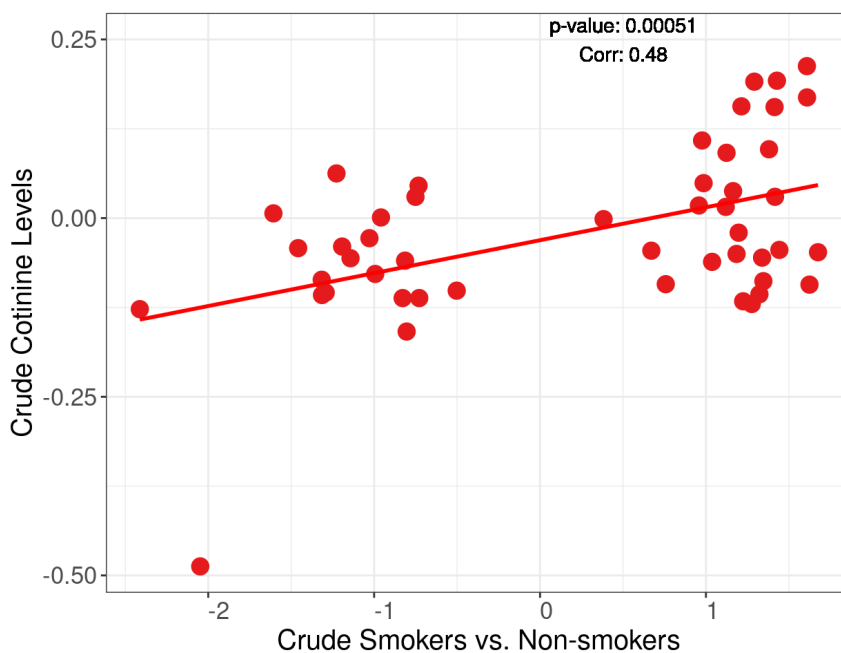
6.1.2 Current smokers vs. never smokers adjusting for confounding



6.1.3 Plot crude vs. adjusted from DESeq2

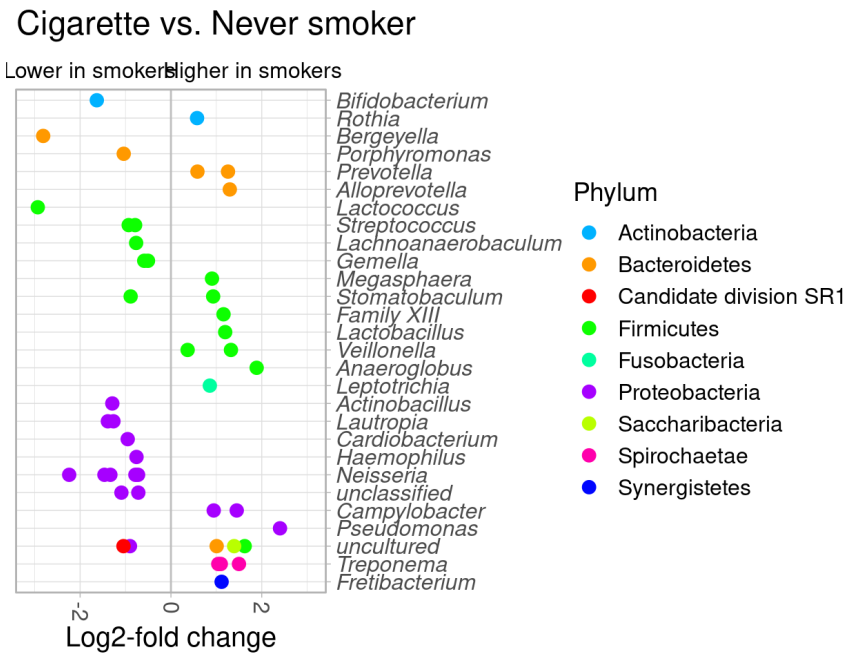


6.1.4 Secondhand vs Cigarette/Never

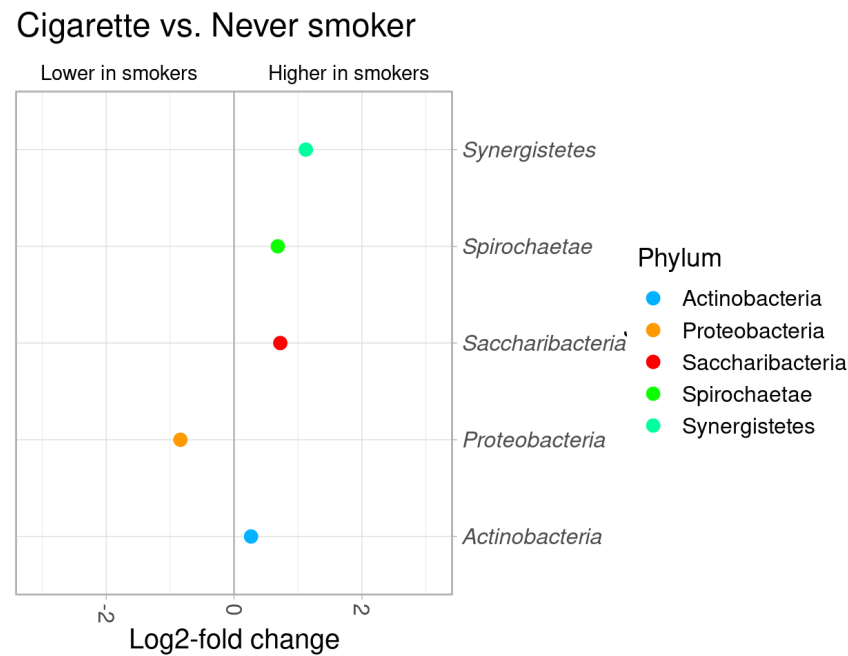


6.2 edgeR

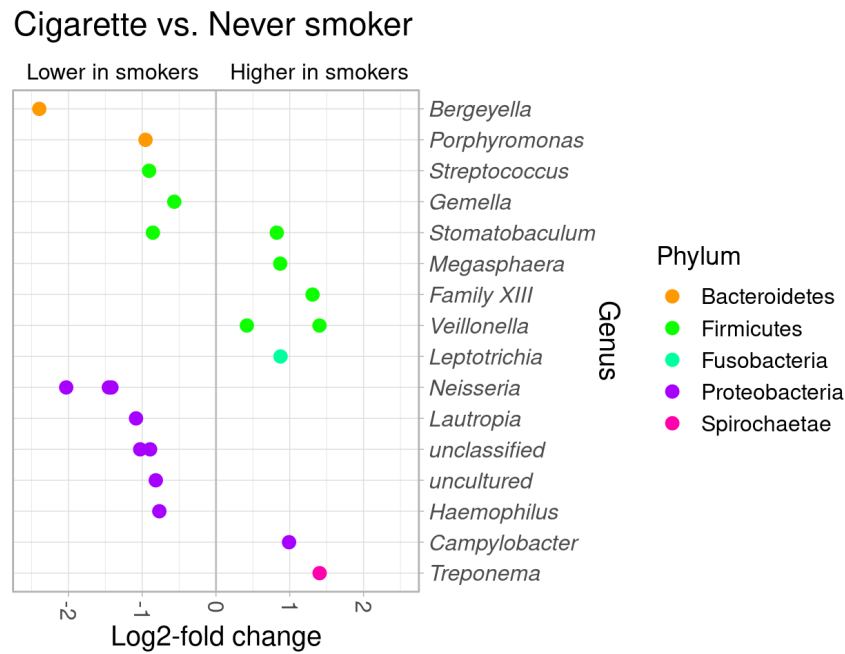
6.2.1 Cigarette smoker vs Never smoker: crude



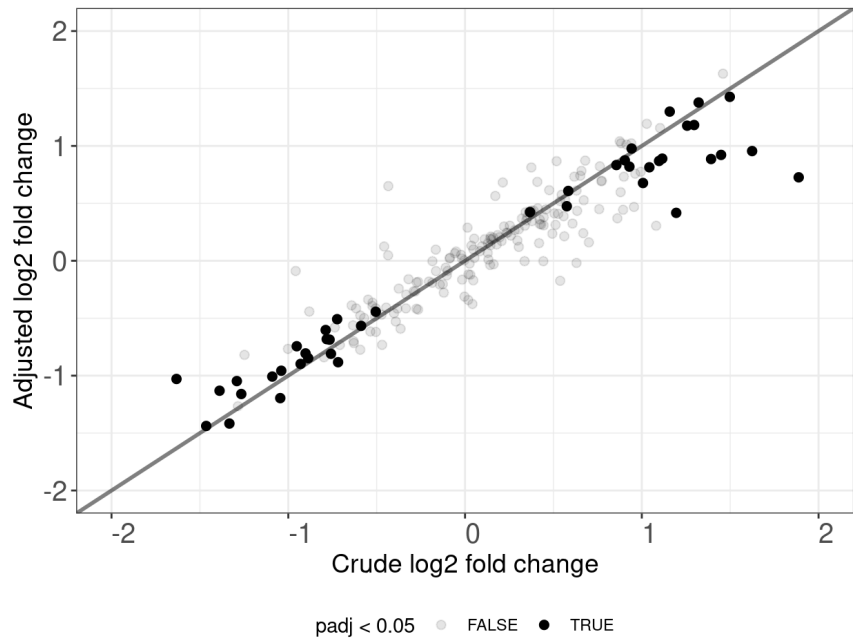
6.3 Cigarette smoker vs Never smoker: crude phylum level



6.3.1 Cigarette smoker vs Never smoker: Adjusted




6.3.2 Plot crude vs adjusted beta coefficients



6.3.3 Secondhand vs Cigarette/Never: only significant OTUs

A function for merging smoking and cotinine edgeR results:

```
mergeResults <- function(obj.smoking, obj.cotinine) {  
  intersect.filters <-  
    intersect(rownames(obj.smoking), rownames(obj.cotinine))  
  intersect.filters <- sort(intersect.filters)  
  obj.smoking <- obj.smoking[intersect.filters, ]  
  obj.cotinine <- obj.cotinine[intersect.filters, ]  
  stopifnot(all.equal(rownames(obj.smoking), rownames(obj.cotinine)))  
  ##  
  res <-  
    data.frame(  
      logFC_smoking = obj.smoking$table$logFC,  
      logFC_cotinine = obj.cotinine$table$logFC,  
      FDR_smoking = obj.smoking$table$FDR,  
      FDR_cotinine = obj.cotinine$table$FDR,  
      OTU = rownames(obj.smoking$table)  
    )  
  res <- cbind(res, obj.smoking[, c("Phylum", "Class", "Order", "Fami  
  return(res)  
}
```



And a plotting function:

```

plotSmokingcotinineComparison <-
  function(obj.smoking,
           obj.cotinine,
           labtext = "Crude",
           FDRcutoff = 1) {
    dataset <- mergeResults(obj.smoking, obj.cotinine)
    dataset <- dataset[dataset$FDR_smoking < FDRcutoff,]
    corr.result <- with(dataset,
                        cor.test(logFC_smoking, logFC_cotinine))
    p <- ggplot(dataset) +
      theme_bw() +
      xlab(paste(labtext, "Smokers vs. Non-smokers")) +
      ylab(paste(labtext, "Cotinine Levels (ng/ml)")) +
      geom_text(aes(1,-0.3, label = sprintf(
        "p-value: %s\ncor: %s \nFDR<0.05: %s",
        signif(corr.result$p.value, 2),
        signif(corr.result$estimate, 2),
        sum(dataset$FDR_smoking < 0.05)
      )), size = 5) +
      geom_point(aes(logFC_smoking, logFC_cotinine, color = FDR_smoking),
                 size = 3) +
      scale_colour_manual(values = setNames(c('black', 'grey'), c(TRUE, FALSE)) +
                          FALSE) +
      geom_smooth(
        aes(logFC_smoking, logFC_cotinine),
        method = lm,
        se = FALSE,
        color = "red"
      ) +
      theme(
        axis.text = element_text(
          hjust = 0,
          vjust = 0.5,
          size = 13
        ),
        strip.text = element_text(size = 13),
        axis.title = element_text(size = 15)
      )
    return(p)
  }

```

Two thresholds for serum cotinine:

```

NYC_HANES.secondhand.original <- subset_samples(NYC_HANES, smokingstatus == "secondhand")
NYC_HANES.secondhand.strict <- subset_samples(NYC_HANES.secondhand.original,
sort(sample_data(NYC_HANES.secondhand.original)$COTININE)

```

```

## [1] 1.0022 1.0394 1.0418 1.0715 1.0915 1.2015 1.2282 1.27
## [10] 1.3822 1.4094 1.5315 1.6918 2.0094 2.5994 2.6759 2.86
## [19] 2.9720 3.0459 3.1082 3.1722 3.4715 3.8394 4.3418 4.52
## [28] 4.7522 5.7220 6.3118 7.1315 7.7382 9.1118 9.6415 10.29
## [37] 10.7915 13.7894

```

```
summary(sample_data(NYC_HANES.secondhand.original)$COTININE > 10)
```

```
##      Mode    FALSE    TRUE
## logical      34      4
```

Number of OTUs with FDR < 0.05 and total that passed non-specific screens in both analysis:

```
summary(edger.smoker_secondhand_crude$FDR_smoking < 0.05)
```

```
##      Mode    FALSE    TRUE
## logical      93      28
```

*# note that the different total number for strict cotinine cutoff is
due to non-specific filtering in edgeR*

```
summary(edger.smoker_secondhand_crude.strict$FDR_smoking < 0.05)
```

```
##      Mode    FALSE    TRUE
## logical      87      27
```

```
corarray <- array(dim=c(2, 2, 2), dimnames=list(secondhand=c("cot14",
                                                             counfounding=c("crude
                                                             FDRcutoff=c("no", "ye
pararray <- array(dim=c(2, 2, 2), dimnames=list(secondhand=c("cot14", '
                                                             counfounding=c("crude
                                                             FDRcutoff=c("no", "ye
```

Correlation between crude smoking and cotinine coefficients, all OTUs:

```
corr <- with(edger.smoker_secondhand_crude,
             cor.test(logFC_smoking, logFC_cotinine))
pararray["cot14", "crude", "no"] <- corr$p.value
corarray["cot14", "crude", "no"] <- corr$estimate
```

Strict definition of second-hand smoke:

```
corr <- with(edger.smoker_secondhand_crude.strict,
             cor.test(logFC_smoking, logFC_cotinine))
pararray["cot10", "crude", "no"] <- corr$p.value
corarray["cot10", "crude", "no"] <- corr$estimate
```

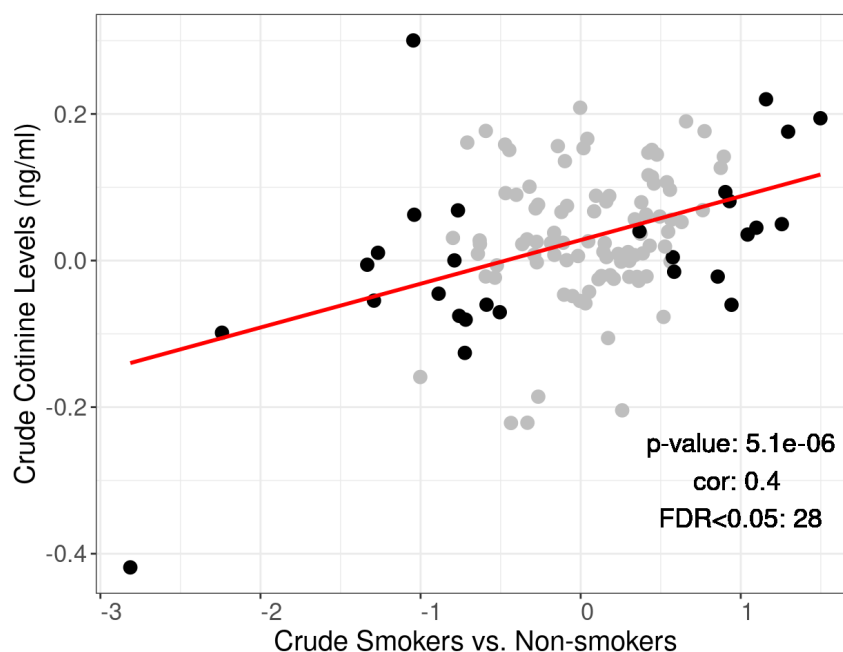
Correlation between crude smoking and cotinine coefficients, FDR < 0.05 in smoking:

```
corr <- with(filter(edger.smoker_secondhand_crude, FDR_smoking < 0.05),
             cor.test(logFC_smoking, logFC_cotinine))
corarray["cot10", "crude", "yes"] <- corr$p.value
pararray["cot10", "crude", "yes"] <- corr$estimate
```

All OTUs, black for FDR < 0.05 on smoking coefficients, grey for FDR >= 0.05

```
p <-
  plotSmokingcotinineComparison(dasmoking.crude,
                                dacot.crude,
                                labtext = "Crude",
                                FDR = 1)
```

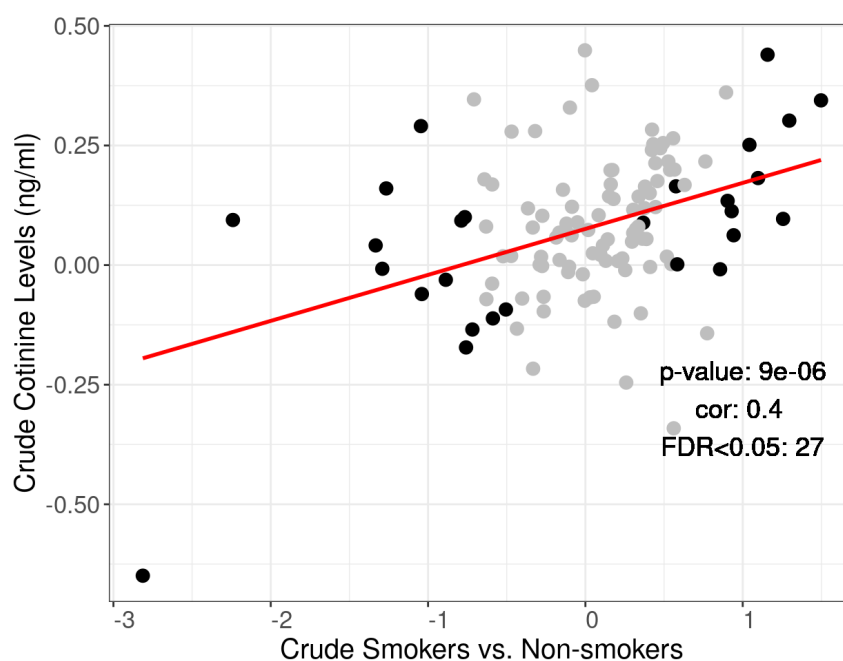
p



Repeat, using the strict definition of second-hand smokers (cotinine < 10)

```
p <-
  plotSmokingcotinineComparison(dasmoking.crude,
                                dacot.crude.strict,
                                labtext = "Crude",
                                FDR = 1)
```

p



6.3.4 Adjusted Secondhand vs Adjusted Cigarette/Never: only significant OTUs

Number of differentially abundant OTUs, and total

```
nrow(edger.smoker_secondhand_adjusted)
```

```
## [1] 121
```

```
summary(edger.smoker_secondhand_adjusted$logFC_smoking < 0.05)

##      Mode    FALSE     TRUE
## logical      63      58

summary(edger.smoker_secondhand_adjusted$logFC_cotinine < 0.05)

##      Mode    FALSE     TRUE
## logical      38      83
```

Correlation between adjusted smoking and cotinine coefficients, all OTUs:

```
corr <- with(edger.smoker_secondhand_adjusted,
             cor.test(logFC_smoking, logFC_cotinine))
parray["cot14", "adjusted", "no"] <- corr$p.value
corarray["cot14", "adjusted", "no"] <- corr$estimate
```

As above, using strict serum cotinine cutoff (<10)

```
corr <- with(edger.smoker_secondhand_adjusted.strict,
             cor.test(logFC_smoking, logFC_cotinine))
parray["cot10", "adjusted", "no"] <- corr$p.value
corarray["cot10", "adjusted", "no"] <- corr$estimate
```

Correlation between adjusted smoking and cotinine coefficients, FDR < 0.05 in smoking:

```
with(filter(edger.smoker_secondhand_adjusted, FDR_smoking < 0.05),
     cor.test(logFC_smoking, logFC_cotinine))

##
## Pearson's product-moment correlation
##
## data: logFC_smoking and logFC_cotinine
## t = 3.8131, df = 17, p-value = 0.001391
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3249855 0.8660843
## sample estimates:
##      cor
## 0.6789709

parray["cot14", "adjusted", "yes"] <- corr$p.value
corarray["cot14", "adjusted", "yes"] <- corr$estimate
```

Sensitivity analysis:

```
round(corarray, 2)

## , , FDRcutoff = no
##
##      counfounding
## secondhand crude adjusted
##      cot14  0.4    0.28
##      cot10  0.4    0.34
##
## , , FDRcutoff = yes
##
##      counfounding
## secondhand crude adjusted
##      cot14  NA    0.34
##      cot10  0     NA

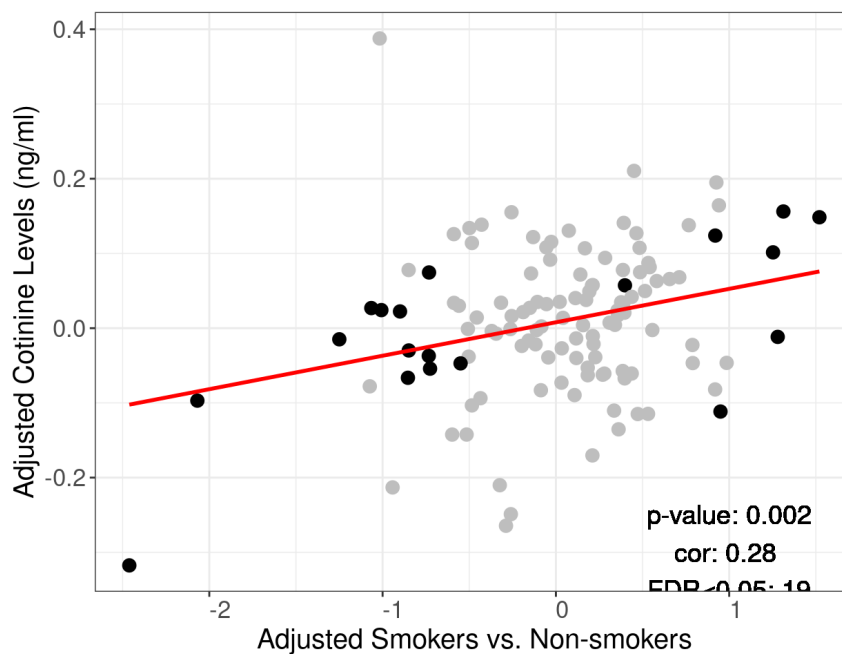
signif(parray, 1)
```

```
## , , FDRcutoff = no
##
##      confounding
## secondhand crude adjusted
##      cot14 5e-06    2e-03
##      cot10 9e-06    3e-04
##
```

```
## , , FDRcutoff = yes
##
##      confounding
## secondhand crude adjusted
##      cot14    NA    3e-04
##      cot10 0.6    NA
```

```
p <-
  plotSmokingcotinineComparison(
    obj.smoking = dasmoking.adjusted,
    obj.cotinine = dacot.adjusted,
    labtext = "Adjusted",
    FDRcutoff = 1
  )
```

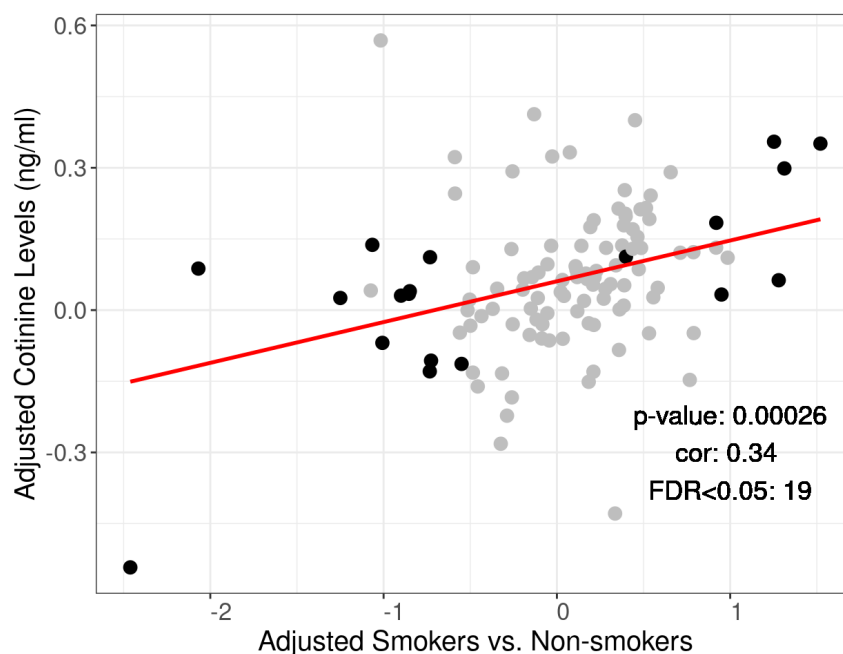
```
p
```



Repeat, using the strict definition of second-hand smokers (cotinine < 10)

```
p <-
  plotSmokingcotinineComparison(
    obj.smoking = dasmoking.adjusted,
    obj.cotinine = dacot.adjusted.strict,
    labtext = "Adjusted",
    FDRcutoff = 1
  )
```

```
p
```

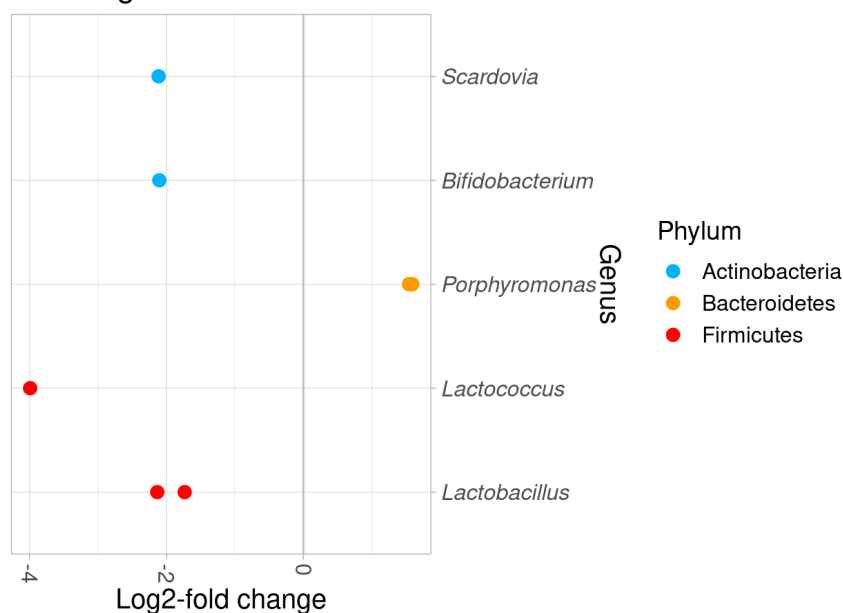


6.3.5 Write supplemental file

```
stopifnot(all.equal(rownames(edger.smoker_secondhand_adjusted),
                        rownames(edger.smoker_secondhand_crude)))
crude <- edger.smoker_secondhand_crude[, 1:4]
adjusted <- edger.smoker_secondhand_adjusted[, 1:4]
colnames(crude) <- paste0(colnames(crude), "_crude")
colnames(adjusted) <- paste0(colnames(adjusted), "_adjusted")
sfile <- cbind(edger.smoker_secondhand_crude[, -1:-4], crude, adjusted)
readr::write_excel_csv(sfile, path="SupplementalFile1.csv")
```

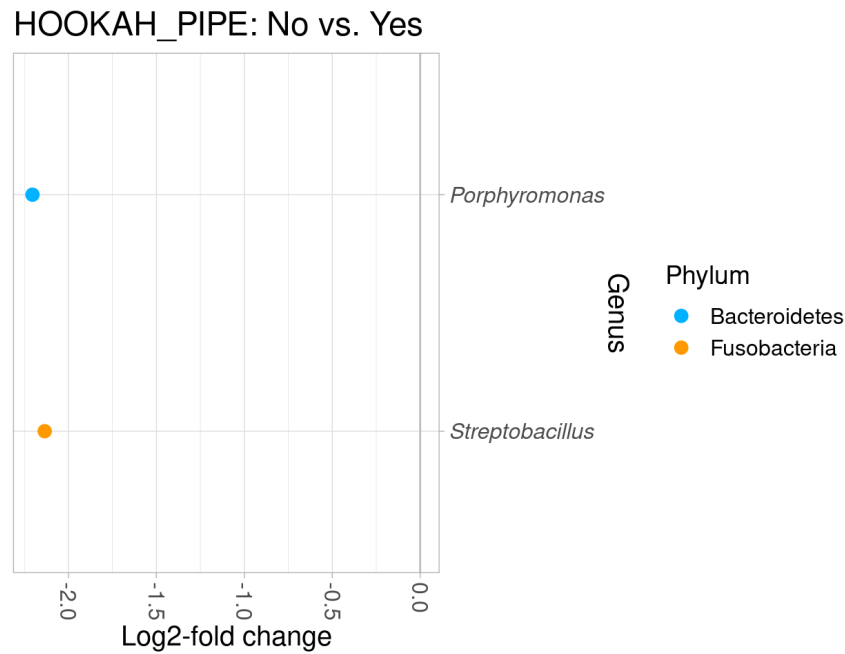
6.3.6 Alternative smokers vs Never smoker: crude

smokingstatus: Alternative smoker vs. Never smoker



6.3.6.1 E-cigarette vs never smokers

6.3.6.2 Hookah pipe vs never smokers



6.3.6.3 Cigars and cigarillos vs never smokers


[1] "No results"

7 Analysis on biosis of bacteria


7.1 Odds ratio smokers vs never smokers

```
## $tab
##
##           Enriched in smokers      p0 Depleted in smc
## OTU is aerobic                23 0.0362776
## OTU is not aerobic            611 0.9637224
##
##           p1 oddsratio      lower      upper
## OTU is aerobic      0.1074219 1.0000000      NA      NA
## OTU is not aerobic  0.8925781 0.3127808 0.1894164 0.5164911 2.682
##
## $measure
## [1] "wald"
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "fisher.exact"
```

```
## $tab
##
##           Enriched in smokers           p0 Depleted in s
## OTU is anaerobic           407 0.6419558
## OTU is not anaerobic       227 0.3580442
##
##           p1 oddsratio      lower      upper
## OTU is anaerobic  0.5195312  1.000000      NA      NA
## OTU is not anaerobic 0.4804688  1.658143  1.307573  2.102704  3.107
##
## $measure
## [1] "wald"
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "fisher.exact"
```



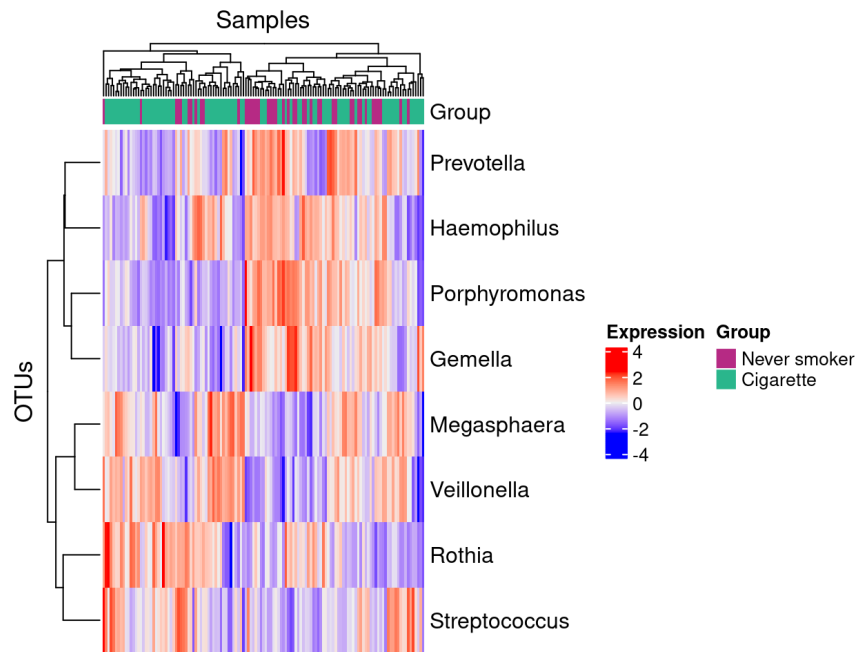
```
## $tab
##
##           Enriched in smokers           p0 Depleted in ir
## OTU is F Anaerobic           204 0.3217666
## OTU is not F Anaerobic       430 0.6782334
##
##           p1 oddsratio      lower      upper
## OTU is F Anaerobic  0.3730469  1.0000000      NA      NA
## OTU is not F Anaerobic 0.6269531  0.7973213  0.624298  1.018298  0.6
##
## $measure
## [1] "wald"
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "fisher.exact"
```



7.2 Cigarette smokers vs Never smokers

7.2.1 GSEA

```
## DataFrame with 2 rows and 4 columns
##      GENE.SET      ES      NES      PVAL
## <character> <numeric> <numeric> <numeric>
## 1      fana    -0.322   -0.738    0.769
## 2      anae     0.282    0.545    0.957
```



7.2.2 ORA

```
## 100 permutations completed
## 200 permutations completed
## 300 permutations completed
## 400 permutations completed
## 500 permutations completed
## 600 permutations completed
## 700 permutations completed
## 800 permutations completed
## 900 permutations completed
## 1000 permutations completed

## DataFrame with 2 rows and 4 columns
##      GENE.SET  GLOB.STAT  NGLGB.STAT    PVAL
##    <character> <numeric>  <numeric> <numeric>
## 1      fana         3      0.300    0.186
## 2      anae         5      0.208    0.531
```

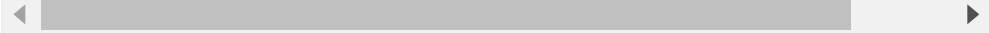
7.3 Secondhand smokers

7.3.1 GSEA

```
## DataFrame with 2 rows and 4 columns
##      GENE.SET    ES    NES    PVAL
##    <character> <numeric> <numeric> <numeric>
## 1      anae    0.77    1.30    0.164
## 2      fana   -0.50   -1.11    0.360
```



```
##          logFC      AveExpr      t      P.Value  adj.P.Val
## anae  9.312388e-05  0.01061843  2.635005  0.009409142  0.02519808 .
## fana -1.031180e-04 -0.02637825 -2.421534  0.016798721  0.02519808 .
## aero  9.331210e-05  0.10299623  1.397472  0.164590432  0.16459043 -1
```



7.5 Hookah smokers vs Never smokers

7.5.1 GSEA

```
## DataFrame with 2 rows and 4 columns
##      GENE.SET      ES      NES      PVAL
## <character> <numeric> <numeric> <numeric>
## 1      anae      0.586      1.15      0.299
## 2      fana     -0.532     -1.17      0.301
```

7.5.2 ORA

```
## DataFrame with 2 rows and 4 columns
##      GENE.SET  NR.GENES  NR.SIG.GENES      PVAL
## <character> <numeric>   <numeric> <numeric>
## 1      anae       23         1      0.697
## 2      fana       10         0      1.000
```

7.6 E-cigarette smokers vs Never smokers

7.6.1 GSEA

```
## DataFrame with 2 rows and 4 columns
##      GENE.SET      ES      NES      PVAL
## <character> <numeric> <numeric> <numeric>
## 1      anae      0.451      1.020      0.466
## 2      fana      0.329      0.832      0.683
```

7.6.2 ORA

```
## DataFrame with 2 rows and 4 columns
##      GENE.SET  NR.GENES  NR.SIG.GENES      PVAL
## <character> <numeric>   <numeric> <numeric>
## 1      anae       29         0         1
## 2      fana       12         0         1
```

7.7 Cigar/cigarillos smokers vs Never smokers

7.7.1 GSEA

```
## DataFrame with 2 rows and 4 columns
##      GENE.SET      ES      NES      PVAL
## <character> <numeric> <numeric> <numeric>
## 1      anae     -0.487     -0.983      0.508
## 2      fana      0.372      0.813      0.663
```

7.7.2 ORA

```
## DataFrame with 2 rows and 4 columns
##      GENE.SET  NR.GENES NR.SIG.GENES      PVAL
##  <character> <numeric>  <numeric> <numeric>
## 1      fana       10         1    0.323
## 2      anae       21         0    1.000
sessionInfo()
```

```

## R version 4.3.0 (2023-04-21)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Pop!_OS 22.04 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.10.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## time zone: America/New_York
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    ba
##
## other attached packages:
##  [1] edgeR_3.42.2      limma_3.56.0      ggplot2_3.4
##  [4] magrittr_2.0.3    dplyr_1.1.2       nychanesmic
##  [7] lsr_0.5.2         phyloseq_1.44.0   Biobase_2.6
## [10] BiocGenerics_0.46.0 BiocStyle_2.28.0
##
## loaded via a namespace (and not attached):
##  [1] splines_4.3.0      bitops_1.0-7
##  [3] tibble_3.2.1       R.oo_1.25.0
##  [5] graph_1.78.0       XML_3.99-0.14
##  [7] lifecycle_1.0.3    doParallel_1.0.17
##  [9] lattice_0.21-8     vroom_1.6.3
## [11] MASS_7.3-59        survey_4.1-1
## [13] sass_0.4.5         rmarkdown_2.21
## [15] jquerylib_0.1.4    yaml_2.3.7
## [17] DBI_1.1.3          RColorBrewer_1.1-3
## [19] ade4_1.7-22        zlibbioc_1.46.0
## [21] Rtsne_0.16         GenomicRanges_1.52.0
## [23] purrr_1.0.1        R.utils_2.12.2
## [25] RCurl_1.98-1.12    circlize_0.4.15
## [27] labelled_2.11.0     GenomeInfoDbData_1.2.10
## [29] IRanges_2.34.0     S4Vectors_0.38.1
## [31] irlba_2.3.5.1      GSVA_1.48.0
## [33] vegan_2.6-4         microbiome_1.22.0
## [35] annotate_1.78.0     DelayedMatrixStats_1.22.0
## [37] permute_0.9-7       codetools_0.2-19
## [39] DelayedArray_0.26.1 tidyselect_1.2.0
## [41] shape_1.4.6         farver_2.1.1
## [43] ScaledMatrix_1.8.0  epitools_0.5-10.1
## [45] matrixStats_0.63.0 stats4_4.3.0
## [47] jsonlite_1.8.4      GetoptLong_1.0.5
## [49] multtest_2.56.0     e1071_1.7-13
## [51] survival_3.5-5      iterators_1.0.14
## [53] foreach_1.5.2       tools_4.3.0
## [55] Rcpp_1.0.10         glue_1.6.2

```

```

## [57] xfun_0.39 mgcv_1.8-42
## [59] DESeq2_1.40.1 MatrixGenerics_1.12.0
## [61] GenomeInfoDb_1.36.0 HDF5Array_1.28.1
## [63] withr_2.5.0 BiocManager_1.30.20
## [65] fastmap_1.1.1 mitools_2.4
## [67] rhdf5filters_1.12.1 fansi_1.0.4
## [69] SparseM_1.81 rsvd_1.0.5
## [71] digest_0.6.31 R6_2.5.1
## [73] colorspace_2.1-0 Cairo_1.6-0
## [75] RSQLite_2.3.1 R.methodsS3_1.8.2
## [77] utf8_1.2.3 tidyr_1.3.0
## [79] generics_0.1.3 data.table_1.14.8
## [81] class_7.3-21 httr_1.4.5
## [83] S4Arrays_1.0.1 pkgconfig_2.0.3
## [85] gtable_0.3.3 blob_1.2.4
## [87] ComplexHeatmap_2.16.0 SingleCellExperiment_1.22.0
## [89] XVector_0.40.0 htmltools_0.5.5
## [91] bookdown_0.33 biomformat_1.28.0
## [93] GSEABase_1.62.0 clue_0.3-64
## [95] scales_1.2.1 png_0.1-8
## [97] knitr_1.42 rstudioapi_0.14
## [99] tzdb_0.3.0 reshape2_1.4.4
## [101] rjson_0.2.21 nlme_3.1-162
## [103] proxy_0.4-27 cachem_1.0.8
## [105] zoo_1.8-12 rhdf5_2.44.0
## [107] GlobalOptions_0.1.2 stringr_1.5.0
## [109] safe_3.40.0 parallel_4.3.0
## [111] AnnotationDbi_1.62.1 pillar_1.9.0
## [113] grid_4.3.0 vctrs_0.6.2
## [115] BiocSingular_1.16.0 beachmat_2.16.0
## [117] xtable_1.8-4 cluster_2.1.4
## [119] Rgraphviz_2.44.0 evaluate_0.20
## [121] KEGGgraph_1.60.0 readr_2.1.4
## [123] cli_3.6.1 locfit_1.5-9.7
## [125] compiler_4.3.0 rlang_1.1.1
## [127] crayon_1.5.2 tableone_0.13.2
## [129] labeling_0.4.2 plyr_1.8.8
## [131] forcats_1.0.0 stringi_1.7.12
## [133] BiocParallel_1.34.0 munsell_0.5.0
## [135] Biostrings_2.68.0 Matrix_1.5-4
## [137] sas7bdat_0.7 hms_1.1.3
## [139] sparseMatrixStats_1.12.0 bit64_4.0.5
## [141] Rhdf5lib_1.22.0 KEGGREST_1.40.0
## [143] statmod_1.5.0 SummarizedExperiment_1.30.1
## [145] highr_0.10 haven_2.5.2
## [147] igraph_1.4.2 memoise_2.0.1
## [149] bslib_0.4.2 bit_4.0.5
## [151] EnrichmentBrowser_2.30.0 ape_5.7-1

```