Malaria Infant Paper Analysis

Paul L. Maurizio[*1]

[1]University of Chicago, Section in Genetic Medicine

*maurizio@alumni.unc.edu

5/28/2019

Contents

0.1 Load Packages

```
library(ggplot2)
library(ggfortify)
library(GGally)
library(coin)
library(xtable)
library(pscl) # zero inflated poisson regression
library(MASS) # multinomial / proportional odds regression / Box-Cox
library(ordinal) # cumulative link mixed model
library(nparLD) # nonparametric rank-based statistics for longitudinal data
library(lmerTest)
library(psych)
library(coin)
library(cowplot)
```

0.2 Load Data

```
dat <- read.csv("../data/MIS_master_data_sheet_wide.csv")
dat_long <- read.csv("../data/MIS_master_data_sheet_long.csv")
```

0.3 Clinical characteristics

```
dat_children <- subset(dat, age=="I")
dat_adults <- subset(dat, age=="A")
median(dat_children$Hb, na.rm = TRUE); IQR(dat_children$Hb, na.rm = TRUE)
## [1] 9.7
## [1] 2.3
median(dat_adults$Hb, na.rm = TRUE); IQR(dat_adults$Hb, na.rm = TRUE)
## [1] 14.1
## [1] 2.15
median(dat_children$parasites, na.rm = TRUE); IQR(dat_children$parasites, na.rm = TRUE)
## [1] 2533.5
## [1] 74277.5
median(dat_adults$parasites, na.rm = TRUE); IQR(dat_adults$parasites, na.rm = TRUE)
## [1] 576.5
## [1] 7911.5

median(dat_children$pfs25, na.rm = TRUE); IQR(dat_children$pfs25, na.rm = TRUE)
## [1] 0.465
## [1] 0.65
median(dat_adults$pfs25, na.rm = TRUE); IQR(dat_adults$pfs25, na.rm = TRUE)
## [1] 0.255
## [1] 0.1975

median(dat_children$pfs16, na.rm = TRUE); IQR(dat_children$pfs16, na.rm = TRUE)
```

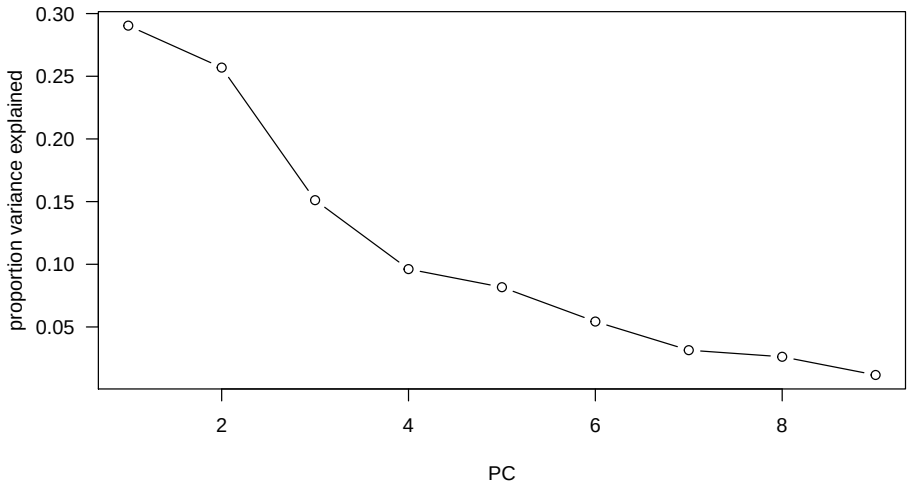
```
## [1] 16
## [1] 35.5
median(dat_adults$pfs16, na.rm = TRUE); IQR(dat_adults$pfs16, na.rm = TRUE)
## [1] 28
## [1] 37.75

median(dat_children$pfs230, na.rm = TRUE); IQR(dat_children$pfs230, na.rm = TRUE)
## [1] 25
## [1] 64.25
median(dat_adults$pfs230, na.rm = TRUE); IQR(dat_adults$pfs230, na.rm = TRUE)
## [1] 50
## [1] 21756

median(dat_children$Hb.V2, na.rm = TRUE); IQR(dat_children$Hb.V2, na.rm = TRUE)
## [1] 10.7
## [1] 2.1
median(dat_adults$Hb.V2, na.rm = TRUE); IQR(dat_adults$Hb.V2, na.rm = TRUE)
## [1] 13.5
## [1] 2.55
median(dat_children$parasites.V2, na.rm = TRUE); IQR(dat_children$parasites.V2, na.rm = TRUE)
## [1] 0
## [1] 0
median(dat_adults$parasites.V2, na.rm = TRUE); IQR(dat_adults$parasites.V2, na.rm = TRUE)
## [1] 0
## [1] 0
```

0.4 PCA on raw phenotypes

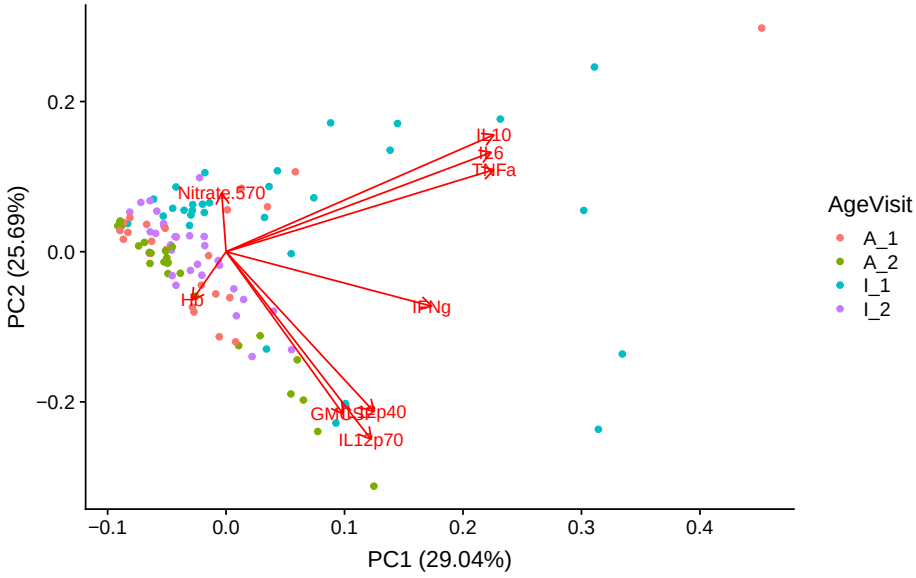
This is a PCA on a subset of the phenotypes that were collected from the majority of individuals.

```
select <- c('GMCSF','IFNg','IL10','IL12p40','IL12p70','IL6','TNFa','Nitrate.570','Hb')
dat_sub <- dat_long[,select]
dat_ref <- dat_long[complete.cases(dat_sub),]
dat_sub <- dat_sub[complete.cases(dat_sub),]
dat_ref$Visit <- as.factor(dat_ref$Visit)
dat_ref$AgeVisit <- as.factor(paste(dat_ref$Age, dat_ref$Visit, sep="_"))
pr <- prcomp(dat_sub, center=TRUE, scale.=TRUE)
summary(pr)
## Importance of components:
##                           PC1    PC2    PC3     PC4     PC5     PC6     PC7
## Standard deviation     1.6166 1.5206 1.1663 0.93028 0.85753 0.69915 0.53225
## Proportion of Variance 0.2904 0.2569 0.1511 0.09616 0.08171 0.05431 0.03148
## Cumulative Proportion  0.2904 0.5473 0.6985 0.79461 0.87631 0.93063 0.96210
##                            PC8     PC9
## Standard deviation     0.48572 0.32428
## Proportion of Variance 0.02621 0.01168
## Cumulative Proportion  0.98832 1.00000
pct.variance.explained <- as.numeric(data.frame(summary(pr)$importance)[2,])
```

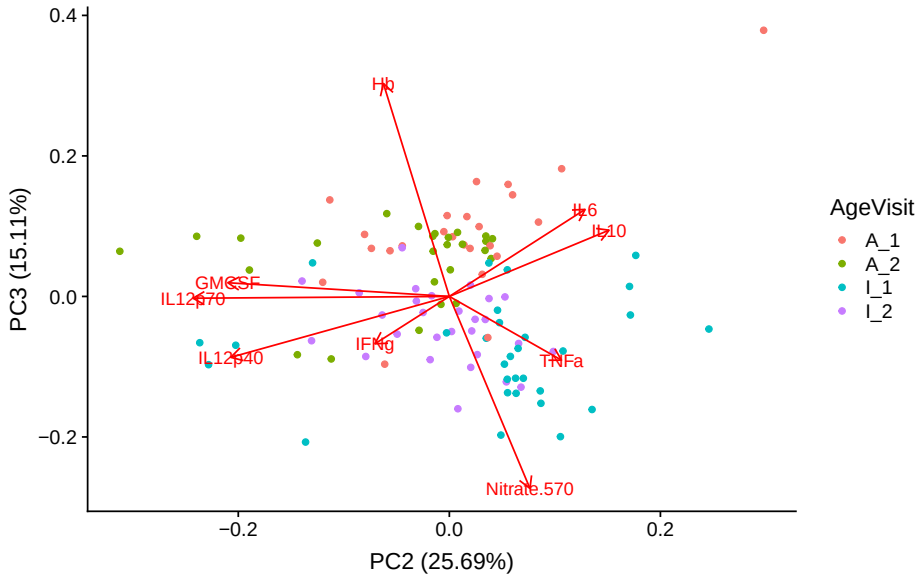
```
x <- c(1:9)
plot(pct.variance.explained ~ x, xlab="PC", ylab="proportion variance explained",
     las=1, type="b")
```




```
autoplot(pr, data = dat_ref, colour = 'AgeVisit', x=1, y=2, loadings=TRUE,
         loadings.label=TRUE)
```




```
autoplot(pr, data = dat_ref, colour = 'AgeVisit', x=2, y=3, loadings=TRUE,
         loadings.label=TRUE)
```



0.5 Parasite Load

0.5.1 Model the effects of age, considering sex, on parasite levels on V1, using the Wilcoxon test, among detectable samples only.

```
dat.sub <- subset(dat, parasites>0)
fit1 <- wilcox_test(dat.sub$parasites ~ dat.sub$age | dat.sub$sex,
                    distribution="exact"); pvalue(fit1)
## [1] 0.005679652
```

0.5.2 Model the effects of age and sex on parasite levels on V1, using zero-inflated Poisson regression.

We use the zero-inflated Poisson (ZIP) model (log link), with the binomial distribution to model the binary outcome of 0-inflation or not (probit link) (Zeileis 2008).

```
fit1 <- zeroinfl(round(parasites) ~ age * sex , data = dat, dist="poisson", link="probit")
summary(fit1)
##
## Call:
## zeroinfl(formula = round(parasites) ~ age * sex, data = dat, dist = "poisson",
##     link = "probit")
##
## Pearson residuals:
##    Min     1Q Median     3Q    Max
## -2.160 -1.967 -1.483  0.119 10.408
##
## Count model coefficients (poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)  9.987511   0.001957  5102.8   <2e-16 ***
## ageI         1.284821   0.002233   575.3   <2e-16 ***
## sexM        -1.721518   0.005692  -302.4   <2e-16 ***
## ageI:sexM    2.012143   0.005851   343.9   <2e-16 ***
##
## Zero-inflation model coefficients (binomial with probit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.8416     0.3689  -2.281   0.0225 *
## ageI          0.3528     0.4932   0.715   0.4743
## sexM          0.1671     0.5393   0.310   0.7566
## ageI:sexM    -0.6073     0.7247  -0.838   0.4021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 9
## Log-likelihood: -2.308e+06 on 8 Df
```

0.5.3 Model the effects of age on parasite levels on V2, using zero-inflated Poisson regression.

```
fit1 <- zeroinfl(round(parasites.V2) ~ age, data = dat, dist="poisson", link="probit")
summary(fit1)
##
## Call:
## zeroinfl(formula = round(parasites.V2) ~ age, data = dat, dist = "poisson",
##     link = "probit")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -0.4082 -0.4082 -0.2311 -0.1981 10.7672
##
## Count model coefficients (poisson with log link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.9890     0.1361   29.31   <2e-16 ***
## ageI          5.8160     0.1361   42.72   <2e-16 ***
##
## Zero-inflation model coefficients (binomial with probit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.7688     0.4519   3.915 9.06e-05 ***
## ageI         -0.7013     0.5386  -1.302    0.193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 16
## Log-likelihood: -9.148e+04 on 4 Df
```

0.6 Gametocytes

Check the correlation of the Pfs phenotypes.

```
p <- corr.test(log(dat[,c("pfs16", "pfs25", "pfs230")]))$p; p
##             pfs16      pfs25 pfs230
## pfs16  0.00000000 0.08389135      1
## pfs25  0.02796378 0.00000000      1
## pfs230 0.66050611 0.94653555      0
```

0.6.1 Model the effects of age and sex on pfs16, pfs25, and pfs230 levels on V1 using the Wilcoxon test

```
fit1 <- wilcox_test(dat$pfs16 ~ dat$age | dat$sex, distribution="exact"); pvalue(fit1)
## [1] 0.1095546
fit2 <- wilcox_test(dat$pfs25 ~ dat$age | dat$sex, distribution="exact"); pvalue(fit2)
## [1] 0.006847807
fit3 <- wilcox_test(dat$pfs230 ~ dat$age | dat$sex, distribution="exact"); pvalue(fit3)
## [1] 0.2377015
median(subset(dat,age=="A")$pfs25, na.rm=TRUE);
## [1] 0.255
median(subset(dat,age=="I")$pfs25, na.rm=TRUE);
## [1] 0.465
```

0.6.2 Check whether the results are similar under distributional assumptions, using lm()

```
fit1 <- lm(log(dat$pfs16) ~ age*sex, data=dat); anova(fit1)
## Analysis of Variance Table
##
## Response: log(dat$pfs16)
##           Df Sum Sq Mean Sq F value  Pr(>F)
## age        1  3.252  3.2520  3.1852 0.08014 .
## sex        1  0.585  0.5853  0.5732 0.45240
## age:sex    1  0.714  0.7135  0.6988 0.40700
## Residuals 52 53.092  1.0210
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fit2 <- lm(log(dat$pfs25) ~ age*sex, data=dat); anova(fit2)
## Analysis of Variance Table
##
## Response: log(dat$pfs25)
##           Df Sum Sq Mean Sq F value   Pr(>F)
## age        1 14.361 14.3608  8.2920 0.005769 **
## sex        1  1.841  1.8410  1.0630 0.307308
## age:sex    1  3.244  3.2442  1.8732 0.176990
## Residuals 52 90.058  1.7319
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
fit3 <- lm(log(dat$pfs230) ~ age*sex, data=dat); anova(fit3)
## Analysis of Variance Table
##
## Response: log(dat$pfs230)
##           Df  Sum Sq Mean Sq F value Pr(>F)
## age        1   32.51  32.509  1.6078 0.2106
## sex        1    0.09   0.086  0.0042 0.9483
## age:sex    1    0.88   0.877  0.0434 0.8359
## Residuals 51 1031.23  20.220
```

0.7 Antimalarial antibody

0.7.1 Model the effects of age and sex on antibody test results at V1 using multinomial ordinal logistic regression

```
dat.sub <- subset(dat, !is.na(malaria.Ab.result))
levels(dat.sub$malaria.Ab.result) <- c("neg", "grey", "pos")
with(dat.sub, table(malaria.Ab.result, sex, age))
## , , age = A
##
##                  sex
## malaria.Ab.result  F  M
##              neg   0  1
##              grey  3  2
##              pos  12 10
##
## , , age = I
##
##                  sex
## malaria.Ab.result  F  M
##              neg   1  1
##              grey  8  5
##              pos   7 10
fit1 <- polr(malaria.Ab.result ~ age*sex, data=dat.sub, method="logistic")
fit2 <- polr(malaria.Ab.result ~ age + sex, data=dat.sub, method="logistic")
fit3 <- polr(malaria.Ab.result ~ sex, data=dat.sub, method="logistic")
fit4 <- polr(malaria.Ab.result ~ 1, data=dat.sub, method="logistic")
anova(fit4,fit3,fit2,fit1)
## Likelihood ratio tests of ordinal regression models
##
## Response: malaria.Ab.result
##       Model Resid. df Resid. Dev   Test    Df  LR stat.    Pr(Chi)
## 1         1        58   94.91848
## 2       sex        57   94.68500 1 vs 2     1 0.2334787 0.62895634
## 3 age + sex        56   90.39413 2 vs 3     1 4.2908745 0.03831745
## 4 age * sex        55   89.72993 3 vs 4     1 0.6641964 0.41508237
```

0.7.2 Model the effects of age and sex on antibody test results at V2 using multinomial logistic regression.

```
dat.sub <- subset(dat, !is.na(malaria.Ab.result.V2))
levels(dat.sub$malaria.Ab.result.V2) <- c("neg", "grey", "pos")
with(dat.sub, table(malaria.Ab.result.V2, sex, age))
## , , age = A
##
##                     sex
## malaria.Ab.result.V2  F  M
##                 neg   2  0
##                 grey  3  4
##                 pos  10  9
##
## , , age = I
##
##                     sex
## malaria.Ab.result.V2  F  M
##                 neg   0  1
##                 grey  7  5
##                 pos   8  6
fit1 <- polr(malaria.Ab.result.V2 ~ age*sex, data=dat.sub, method="logistic")
fit2 <- polr(malaria.Ab.result.V2 ~ age + sex, data=dat.sub, method="logistic")
fit3 <- polr(malaria.Ab.result.V2 ~ sex, data=dat.sub, method="logistic")
fit4 <- polr(malaria.Ab.result.V2 ~ 1, data=dat.sub, method="logistic")
anova(fit4,fit3,fit2,fit1)
## Likelihood ratio tests of ordinal regression models
##
## Response: malaria.Ab.result.V2
##       Model Resid. df Resid. Dev   Test    Df    LR stat.   Pr(Chi)
## 1         1        53   91.55680
## 2       sex        52   91.55054 1 vs 2     1 0.006256122 0.9369565
## 3 age + sex        51   90.50029 2 vs 3     1 1.050246129 0.3054504
## 4 age * sex        50   90.22810 3 vs 4     1 0.272190878 0.6018659
```

0.7.3 Model the effects of age, sex, and visit, together, using a cumulative link mixed model (CLMM)/ordinal probit regression.

```
dat.sub <- subset(dat_long, !is.na(malaria.Ab.result))
levels(dat.sub$malaria.Ab.result) <- c("neg", "grey", "pos")
with(dat.sub, table(malaria.Ab.result, Sex, Age, Visit))
## , , Age = A, Visit = 1
##
##                   Sex
## malaria.Ab.result  F  M
##              neg   0  1
##              grey  3  2
##              pos  12 10
```

```
##
## , , Age = I, Visit = 1
##
##                 Sex
## malaria.Ab.result  F  M
##              neg   1  1
##              grey  8  5
##              pos   7 10
##
## , , Age = A, Visit = 2
##
##                 Sex
## malaria.Ab.result  F  M
##              neg   2  0
##              grey  3  4
##              pos  10  9
##
## , , Age = I, Visit = 2
##
##                 Sex
## malaria.Ab.result  F  M
##              neg   0  1
##              grey  7  5
##              pos   8  6
```

```
fit1 <- clmm(malaria.Ab.result ~ Age*Sex*Visit + (1|Subject_ID), data=dat.sub,
             Hess=TRUE, link="probit", nAGQ=10)
summary(fit1)
## Cumulative Link Mixed Model fitted with the adaptive Gauss-Hermite
## quadrature approximation with 10 quadrature points
##
## formula: malaria.Ab.result ~ Age * Sex * Visit + (1 | Subject_ID)
## data:    dat.sub
##
##  link   threshold nobs logLik AIC    niter     max.grad cond.H
##  probit flexible  115  -83.00 186.01 518(1557) 2.32e-05 1.5e+03
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Subject_ID (Intercept) 2.221    1.49
## Number of groups:  Subject_ID 60
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## AgeI             -3.5607     1.6383  -2.173   0.0298 *
## SexM             -1.8474     1.7176  -1.076   0.2821
## Visit            -1.2559     0.6978  -1.800   0.0719 .
## AgeI:SexM         3.5755     2.2323   1.602   0.1092
## AgeI:Visit        1.7113     0.8885   1.926   0.0541 .
## SexM:Visit        1.2400     0.9672   1.282   0.1998
## AgeI:SexM:Visit  -2.3611     1.2977  -1.819   0.0689 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##          Estimate Std. Error z value
## neg|grey   -5.283      1.563  -3.379
## grey|pos   -3.036      1.354  -2.243
```

0.8 Plasma cytokines

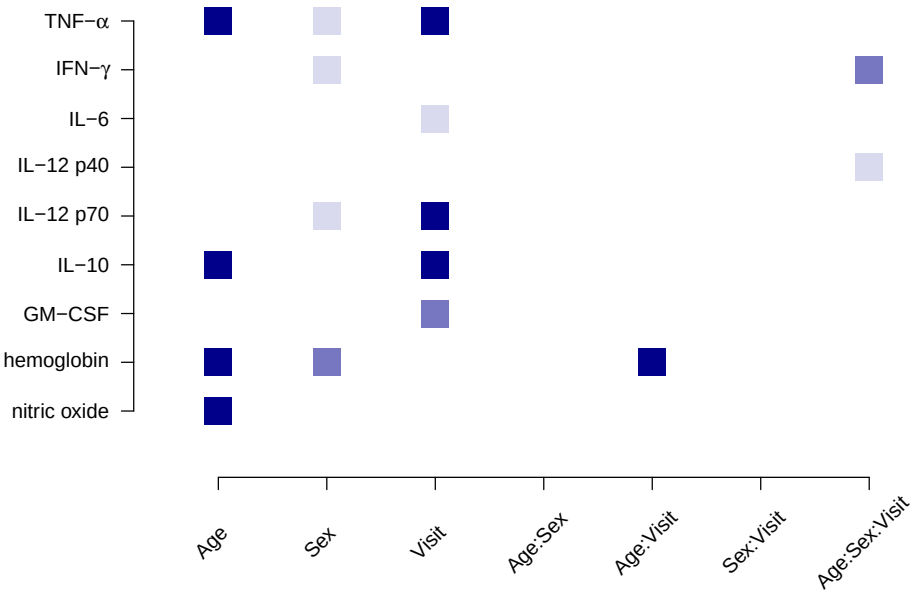
0.8.1 Model the effects of age, sex, and visit on plasma cytokine levels using nparLD (nonparametric).

```
data <- dat_long
host.secreted <- c('TNFa', 'IFNg', 'IL6', 'IL12p40', 'IL12p70', 'IL10',
                   'GMCSF', 'Hb', 'Nitrate.570')

fits <- NULL
fit.df <- as.data.frame(matrix(data=NA, nrow=length(host.secreted),ncol=7))
colnames(fit.df) <- c("Age", "Sex", "Visit", "Age:Sex",
                      "Age:Visit", "Sex:Visit", "Age:Sex:Visit")
fit.df.wholeplot <- as.data.frame(matrix(data=NA, nrow=length(host.secreted),ncol=3))
colnames(fit.df.wholeplot) <- c("Age", "Sex", "Age:Sex")

for(i in 1:length(host.secreted)){
  phen <- host.secreted[i]
  tempdata <- droplevels(subset(data, !is.na(data[phen])))
  complete.subjects <- names(summary(data$Subject_ID))[summary(data$Subject_ID)==2]
  tempdata <- droplevels(subset(data, Subject_ID %in% complete.subjects))
  y <- tempdata[,phen]

  time <- tempdata$Visit
  group1 <- tempdata$Age
  group2 <- tempdata$Sex
  subject <- tempdata$Subject_ID
  time.name <- "Visit"
  group1.name <- "Age"
  group2.name <- "Sex"
  fit <-  f2.ld.f1(y=y, time=time, group1=group1, group2=group2,
                   subject=subject, time.name=time.name,
                   group1.name=group1.name,
                   group2.name=group2.name, plot.RTE=FALSE)
  fits[[i]] <- fit
  fit.df[i,] <- fit$ANOVA.test$`p-value`
  fit.df.wholeplot[i,] <- fit$ANOVA.test.mod.Box$`p-value`
}

cat(paste0(phen, ": \n")); #print(fit$Wald.test);
```

```
rownames(fit.df) <- rownames(fit.df.wholeplot) <- host.secreted
print(fit.df)
xtable(format(fit.df, scientific = TRUE, digits=4))
print(fit.df.wholeplot)
```

0.8.2 Plot the p-values, colored by significance thresholds

```
par(oma=c(0.1,3,0.1,0.1), mar=c(5,3,0.5,0.5))
plot(1~1, xlim=c(0.5,7.5), ylim=rev(c(1,length(host.secreted)+1)), col="white",
     ylab="", xlab="",axes=FALSE)
## Warning in plot.formula(1 ~ 1, xlim = c(0.5, 7.5), ylim = rev(c(1,
## length(host.secreted) + : the formula '1 ~ 1' is treated as '1 ~ 1'
for(i in 1:7){
  for(j in 1:length(host.secreted)){
    if(fit.df[j,i] < 0.05){
      points(x=i,y=j, col=adjustcolor("darkblue", alpha.f=0.15), pch=15, cex=3)
    }
    if(fit.df[j,i] < 0.01){
      points(x=i,y=j, col=adjustcolor("darkblue", alpha.f=0.45), pch=15, cex=3)
    }
    if(fit.df[j,i] < 0.001){
      points(x=i,y=j, col="darkblue", pch=15, cex=3)
    }
  }
}
new.rownames <- c(expression(paste("TNF-",alpha)), expression(paste("IFN-", gamma)),
                  "IL-6", "IL-12 p40", "IL-12 p70", "IL-10", "GM-CSF",
                  "hemoglobin", "nitric oxide")
axis(side=2,at=c(1:length(host.secreted)),labels=new.rownames, las=2)
text(seq(1, 7, by=1)-0.1, par("usr")[3] + 0.5, labels = colnames(fit.df),
     srt = 45, pos = 1, offset=1.5, xpd = TRUE)
axis(side=1, at=c(1:7), labels=rep("",7), las=1)
```



0.9 Cell composition phenotypes

0.9.1 Model the effects of age, sex, and visit on cellular phenotypes using nparLD (nonparametric).

```
data <- dat_long
host.cellular <- c('CD33.live', 'mDC.live', 'monocytes.live', 'inflam.CD163',
                   'patrol.CD163', 'trad.CD163', 'low.traditional')
fits <- NULL
fit.df <- as.data.frame(matrix(data=NA, nrow=length(host.cellular),ncol=7))
colnames(fit.df) <- c("Age", "Sex", "Visit", "Age:Sex", "Age:Visit", "Sex:Visit",
                      "Age:Sex:Visit")
fit.df.wholeplot <- as.data.frame(matrix(data=NA, nrow=length(host.cellular),ncol=3))
colnames(fit.df.wholeplot) <- c("Age", "Sex", "Age:Sex")

for(i in 1:length(host.cellular)){
  phen <- host.cellular[i]
  tempdata <- droplevels(subset(data, !is.na(data[phen])))
  complete.subjects <- names(summary(data$Subject_ID))[summary(data$Subject_ID)==2]
  tempdata <- droplevels(subset(data, Subject_ID %in% complete.subjects))
  y <- tempdata[,phen]

  time <- tempdata$Visit
  group1 <- tempdata$Age
  group2 <- tempdata$Sex
  subject <- tempdata$Subject_ID
  time.name <- "Visit"
  group1.name <- "Age"
  group2.name <- "Sex"
```

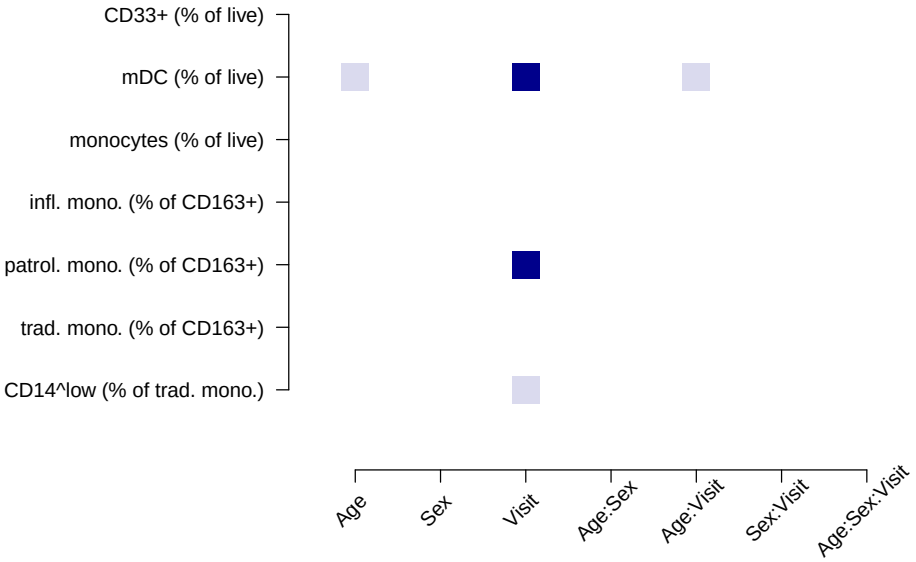
```
  fit <-  f2.ld.f1(y=y, time=time, group1=group1, group2=group2,
                   subject=subject, time.name=time.name,
                   group1.name=group1.name,
                   group2.name=group2.name, plot.RTE=FALSE)
  fits[[i]] <- fit
  fit.df[i,] <- fit$ANOVA.test$`p-value`
  fit.df.wholeplot[i,] <- fit$ANOVA.test.mod.Box$`p-value`
}

cat(paste0(phen, ": \n")); #print(fit$Wald.test);
rownames(fit.df) <- rownames(fit.df.wholeplot) <- host.cellular
```

```
#xtable(format(fit.df, scientific = TRUE, digits=4))
print(fit.df)
##                         Age       Sex        Visit   Age:Sex  Age:Visit
## CD33.live       0.12335135 0.3143957 6.351418e-02 0.8146092 0.57703951
## mDC.live        0.04665501 0.9955378 6.032446e-08 0.2024054 0.04282367
## monocytes.live  0.19027181 0.3264256 1.302573e-01 0.9151475 0.46167393
## inflam.CD163    0.12691802 0.3639364 7.734972e-01 0.9002028 0.17802588
## patrol.CD163    0.79713188 0.4550965 1.167769e-05 0.3813867 0.11039276
## trad.CD163      0.10723603 0.1738107 7.886168e-01 0.9704676 0.85103318
## low.traditional 0.43369195 0.9392206 1.647651e-02 0.2312988 0.42559641
##                 Sex:Visit Age:Sex:Visit
## CD33.live       0.3975777     0.5409413
## mDC.live        0.7180167     0.7881802
## monocytes.live  0.3036635     0.5550374
## inflam.CD163    0.4999607     0.5468366
## patrol.CD163    0.2659831     0.9464210
## trad.CD163      0.6950061     0.9854425
## low.traditional 0.3040026     0.2755606
print(fit.df.wholeplot)
##                        Age       Sex   Age:Sex
## CD33.live       0.1299105 0.3194342 0.8156053
## mDC.live        0.0520493 0.9955597 0.2082021
## monocytes.live  0.1964811 0.3313265 0.9155898
## inflam.CD163    0.1335887 0.3685465 0.9007352
## patrol.CD163    0.7981877 0.4585976 0.3855759
## trad.CD163      0.1135216 0.1799019 0.9706147
## low.traditional 0.4372928 0.9395174 0.2368036
```

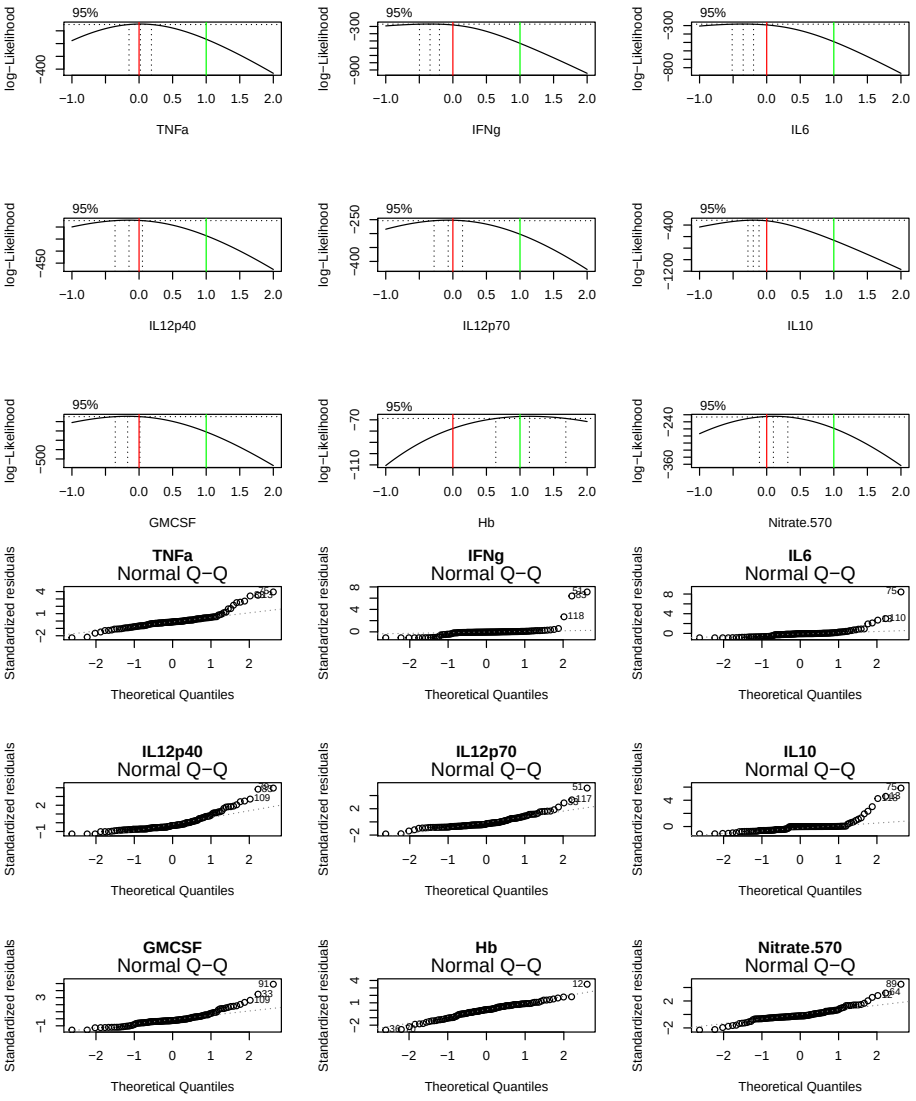
0.9.2 Plot the p-values (colored by significance thresholds)

```
par(oma=c(0.1,0.1,0.1,0.1), mar=c(5,13,0.5,0.5))
plot(1~1, xlim=c(0.5,7.5), ylim=rev(c(1,8)), col="white", ylab="", xlab="",axes=FALSE)
for(i in 1:7){
  for(j in 1:7){
    if(fit.df[j,i] < 0.05){
      points(x=i,y=j, col=adjustcolor("darkblue", alpha.f=0.15), pch=15, cex=3)
    }
```

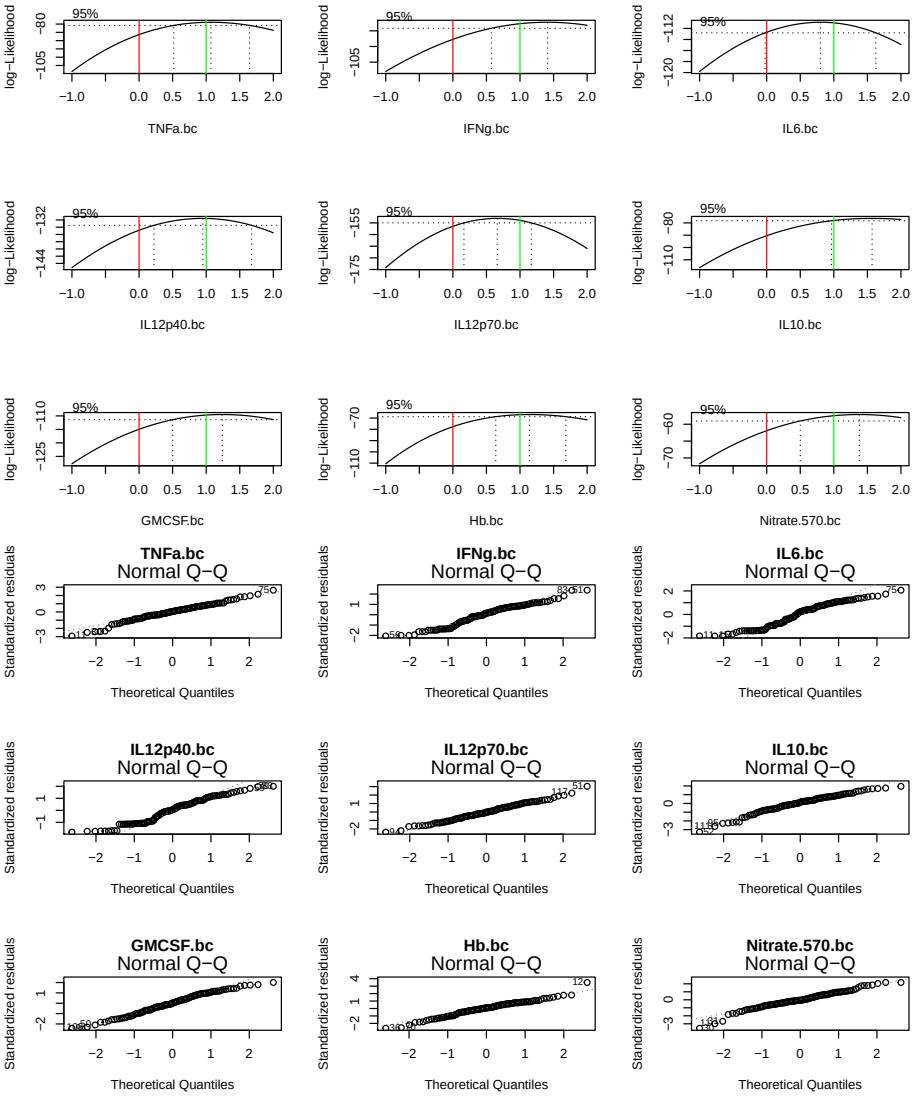
```
      if(fit.df[j,i] < 0.01){
        points(x=i,y=j, col=adjustcolor("darkblue", alpha.f=0.45), pch=15, cex=3)
      }
      if(fit.df[j,i] < 0.001){
        points(x=i,y=j, col="darkblue", pch=15, cex=3)
      }
    }
  }
  new.rownames <- c("CD33+ (% of live)",
                    "mDC (% of live)",
                    "monocytes (% of live)",
                    "infl. mono. (% of CD163+)",
                    "patrol. mono. (% of CD163+)",
                    "trad. mono. (% of CD163+)",
                    "CD14^low (% of trad. mono.)")
  axis(side=2,at=c(1:7),labels=new.rownames, las=2)
  text(seq(1, 7, by=1)-0.1, par("usr")[3], labels = colnames(fit.df),
       srt = 45, pos = 1, offset=1.5, xpd = TRUE)
  axis(side=1, at=c(1:7), labels=rep("",7), las=1)
```



In order to use a parametric (linear mixed model) with our data (lmer) we need to deal with heteroskedastic residuals. We can find a power transform that helps normalize them using Box-Cox analysis (Box and Cox, 1964).


95%
log−Likelihood
TNFa
IFNg
IL6
IL12p40
IL12p70
IL10
GMCSF
Hb
Nitrate.570
Normal Q−Q
Standardized residuals
Theoretical Quantiles

0.9.3 Based on the Box-Cox analysis, choose a sensible transform proximal to the lambda value (+/- sqrt, +/- 1/3 root, log, no transform, etc.): Use log (natural) for lambda ~ 0, and no transform for lambda ~ 1.



0.9.4 Use `lmer`:

```
dat_long$Visit <- dat_long$Visit - 1
fits1 <- fits2 <- summaries <- list()
row_names <- c("(Intercept)", "AgeI", "SexM", "Visit", "AgeI:SexM",
               "AgeI:Visit", "SexM:Visit", "AgeI:SexM:Visit")
summary_table <- data.frame(row.names = row_names)
#for(i in c(1:10,18:20)){
for(i in 1:length(bcphens.t)){
  # Subject-level random intercepts will absorb all the age-specific variation,
```

```
  # so we leave them out and instead estimate global age-specific effects,
  # and only model the within-subject (visit) slopes
  expr1 <- paste0(bcphens.t[i] , "~ Sex*Visit + (0+Visit|Subject_ID)")
  expr2 <- paste0(bcphens.t[i], "~ Age*Sex*Visit + (0+Visit|Subject_ID)")
  fit1 <- lmer(expr1, data=dat_long, na.action=na.exclude)
  fit2 <- lmer(expr2, data=dat_long, na.action=na.exclude)
  fits1[[i]] <- fit1
  names(fits1)[i] <- bcphens.t[i]
  fits2[[i]] <- fit2
  names(fits2)[i] <- bcphens.t[i]
  cat("\n##-------------------------")
  cat(paste0(as.character(bcphens.t[i]), " : "))
  cat("-------------------------##\n")
  cat("\n##------------")
  cat("SUMMARY")
  cat("--------------##\n")
  print(summary(fit2))
  cat("\n##------------")
  cat("ANOVA")
  cat("--------------##\n")
  print(anova(fit2,fit1))
  cat("\n##------------")
  cat("RANOVA")
  cat("--------------##\n")
  print(ranova(fit2))
  summaries[[i]] <- as.data.frame(summary(fit2)[[10]][,5])
  summary_table <- cbind(summary_table, p=summaries[[i]])
}

colnames(summary_table) <- bcphens
summary_table <- t(summary_table)
xtable(format(summary_table, scientific = TRUE, digits=4))
dat_long$Visit <- dat_long$Visit + 1
```

0.10 Cytokine ratios

0.10.1 Model the effects of age, sex, and visit on blood analyte ratios using nparLD; we omit IL12p40, NO and Hb, resulting in 15 proportions tested.

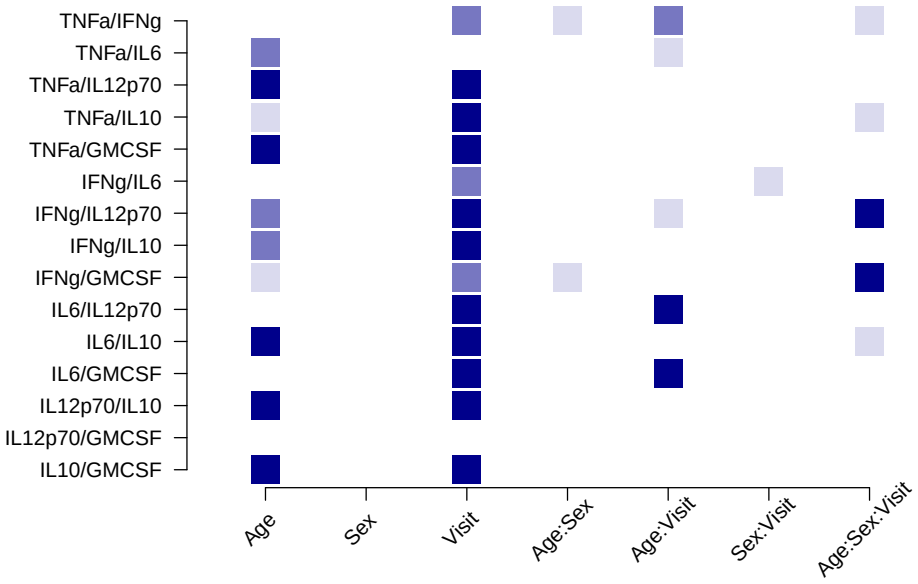
```
ratiotest <- c('TNFa','IFNg','IL6','IL12p70','IL10','GMCSF')
dat.ratios <- dat_long[,c("Subject_ID", "Sample", "age.years", "Age", "Sex", "Visit")]
ratio.combos <- t(combn(ratiotest,2))
ratio.colnames <- paste(ratio.combos[,1], ratio.combos[,2], sep="/")
for(i in 1:length(ratio.colnames)){
  dat.ratios[,ratio.colnames[i]] <- dat_long[,ratio.combos[i,1]]/dat_long[,ratio.combos[i,2]]
}

fits <- NULL
```

```
fit.df <- as.data.frame(matrix(data=NA, nrow=length(ratio.colnames),ncol=7))
colnames(fit.df) <- c("Age", "Sex", "Visit", "Age:Sex", "Age:Visit", "Sex:Visit",
                      "Age:Sex:Visit")
fit.df.wholeplot <- as.data.frame(matrix(data=NA, nrow=length(ratio.colnames),ncol=3))
colnames(fit.df.wholeplot) <- c("Age", "Sex", "Age:Sex")

for(i in 1:length(ratio.colnames)){
  phen <- ratio.colnames[i]
  cat(paste0(phen, ": \n")); #print(fit$Wald.test);
  #tempdata <- droplevels(subset(dat.ratios, !is.na(data[phen])))
  complete.subjects <- names(summary(data$Subject_ID))[summary(data$Subject_ID)==2]
  tempdata <- droplevels(subset(dat.ratios, Subject_ID %in% complete.subjects))
  y <- tempdata[,phen]

  time <- tempdata$Visit
  group1 <- tempdata$Age
  group2 <- tempdata$Sex
  subject <- tempdata$Subject_ID
  time.name <- "Visit"
  group1.name <- "Age"
  group2.name <- "Sex"
  fit <-  f2.ld.f1(y=y, time=time, group1=group1, group2=group2,
                   subject=subject, time.name=time.name,
                   group1.name=group1.name,
                   group2.name=group2.name, plot.RTE=FALSE)
  fits[[i]] <- fit
  fit.df[i,] <- fit$ANOVA.test$`p-value`
  fit.df.wholeplot[i,] <- fit$ANOVA.test.mod.Box$`p-value`
}

rownames(fit.df) <- rownames(fit.df.wholeplot) <- ratio.colnames
```

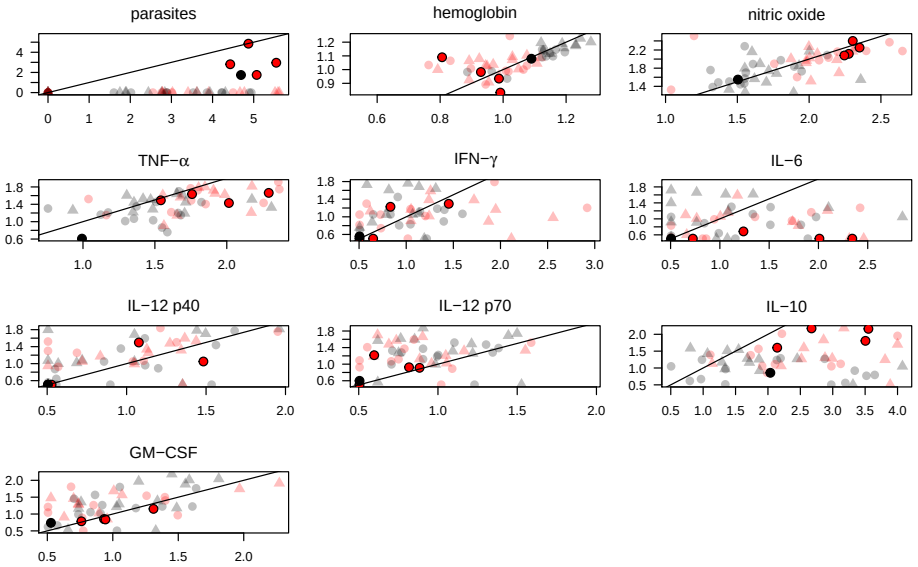
```
#xtable(format(fit.df, scientific = TRUE, digits=4))
print(fit.df)
##                          Age       Sex        Visit    Age:Sex    Age:Visit
## TNFa/IFNg     9.120635e-02 0.5852038 2.679570e-03 0.02673352 0.0082616320
## TNFa/IL6      2.856784e-03 0.6442682 6.790523e-01 0.09455301 0.0269344575
## TNFa/IL12p70  5.860592e-08 0.4937122 3.807258e-18 0.32700283 0.4422558598
## TNFa/IL10     2.706203e-02 0.6957773 4.555445e-16 0.83749812 0.2539669683
## TNFa/GMCSF    1.906851e-05 0.4097557 8.119534e-13 0.84663218 0.5673586057
## IFNg/IL6      3.077146e-01 0.9417818 1.781863e-03 0.93540234 0.3535413331
## IFNg/IL12p70  1.128398e-03 0.7761798 5.501030e-06 0.12440072 0.0219111074
## IFNg/IL10     1.034297e-03 0.6960740 2.843759e-19 0.65692822 0.6256864901
## IFNg/GMCSF    2.532751e-02 0.3089323 1.092911e-03 0.03405905 0.3338340221
## IL6/IL12p70   2.524049e-01 0.8248457 1.499572e-11 0.64024467 0.0001385155
## IL6/IL10      5.282103e-05 0.3498184 3.055106e-19 0.87884185 0.5373833237
## IL6/GMCSF     3.188063e-01 0.5405496 1.492876e-10 0.27020448 0.0008994290
## IL12p70/IL10  7.762802e-06 0.5190226 4.753808e-22 0.85362176 0.5150001829
## IL12p70/GMCSF 6.257536e-01 0.9158885 9.158735e-02 0.24463452 0.4147820477
## IL10/GMCSF    2.796421e-05 0.4986923 1.241835e-22 0.76941913 0.6190910080
##                Sex:Visit Age:Sex:Visit
```

```
## TNFa/IFNg     0.58400643  0.0169148255
## TNFa/IL6      0.13702180  0.6000227758
## TNFa/IL12p70  0.37742929  0.4588496621
## TNFa/IL10     0.90633123  0.0483083382
## TNFa/GMCSF    0.99001948  0.2217603615
## IFNg/IL6      0.01892378  0.3071071312
## IFNg/IL12p70  0.43237932  0.0008848825
## IFNg/IL10     0.81496355  0.5466323136
## IFNg/GMCSF    0.91417711  0.0009115823
## IL6/IL12p70   0.31854725  0.1506788411
## IL6/IL10      0.16435034  0.0125232641
## IL6/GMCSF     0.07861986  0.3733004096
## IL12p70/IL10  0.73584373  0.1018831981
## IL12p70/GMCSF 0.85554122  0.3112755688
## IL10/GMCSF    0.88302614  0.0748478475
print(fit.df.wholeplot)
##                         Age       Sex    Age:Sex
## TNFa/IFNg     9.756590e-02 0.5876821 0.03141559
## TNFa/IL6      4.571634e-03 0.6464594 0.10138971
## TNFa/IL12p70  1.759777e-06 0.4969169 0.33178815
## TNFa/IL10     3.242638e-02 0.6977063 0.83846500
## TNFa/GMCSF    1.207824e-04 0.4148172 0.84764572
## IFNg/IL6      3.126348e-01 0.9420737 0.93572656
## IFNg/IL12p70  2.037338e-03 0.7773628 0.13075292
## IFNg/IL10     1.945056e-03 0.6978248 0.65895016
## IFNg/GMCSF    3.075933e-02 0.3148248 0.04008269
## IL6/IL12p70   2.581658e-01 0.8257988 0.64239146
## IL6/IL10      1.848243e-04 0.3543693 0.87946316
## IL6/GMCSF     3.243046e-01 0.5437265 0.27625071
## IL12p70/IL10  5.083825e-05 0.5222403 0.85443645
## IL12p70/GMCSF 6.283891e-01 0.9164109 0.25145816
## IL10/GMCSF    1.250330e-04 0.5020743 0.77073520
```

0.10.2 Plot the p-values in a grid.



0.11 Treatment failure, recrudescence or reinfection



0.12 Continuous age-associated effects within group

0.12.1 Model the phenotypes vs. age for adults and young children at V1.

```
phens <- c('GMCSF', 'IFNg', 'IL10', 'IL12p40', 'IL12p70',
           'IL6', 'TNFa', 'Nitrate.570',
```

```
                'pfs25', 'pfs16', 'pfs230', 'Hb', 'malaria.Ab', 'parasites')

phens <- phens[!(phens=="malaria.Ab")]
dat.I <- subset(dat, age="I")
dat.A <- subset(dat, age="A")

cat("##-----------------##\n")
cat("CHILDREN \n")
cat("##-----------------##\n")
for(i in 1:length(phens)){
  cat("##-----------------")
  cat(phens[i])
  cat("-----------------## \n")
  form <- paste0(phens[i], "~ age.years*sex")
  try(fit1 <- lm(form, data=dat.I))
  try(print(summary(fit1)))
  fit1 <- NULL
}

for(i in 1:length(phens)){
  cat("##-----------------")
  cat(phens[i])
  cat("-----------------## \n")
  form <- paste0(phens[i], "~ age.years*sex")
  try(fit1 <- lm(form, data=dat.A))
  try(print(summary(fit1)))
  fit1 <- NULL
}
```

0.12.2 Model the phenotypes vs. age for adults and young children at V2.

```
phens.V2 <- paste0(phens, ".V2")
phens.V2 <- phens.V2[!(phens.V2 %in% c("pfs25.V2", "pfs16.V2", "pfs230.V2",
                                       "malaria.Ab.V2"))]

cat("##-----------------##\n")
cat("CHILDREN \n")
cat("##-----------------##\n")
for(i in 1:length(phens.V2)){
  cat("##-----------------")
  cat(phens.V2[i])
  cat("-----------------## \n")
  form <- paste0(phens.V2[i], "~ age.years*sex")
  try(fit1 <- lm(form, data=dat.I))
  try(print(summary(fit1)))
  fit1 <- NULL
}

cat("##-----------------##\n")
```

```
cat("ADULTS \n")
cat("##-----------------##\n")
for(i in 1:length(phens.V2)){
  cat("##-----------------")
  cat(phens.V2[i])
  cat("-----------------## \n")
  form <- paste0(phens.V2[i], "~ age.years*sex")
  try(fit1 <- lm(form, data=dat.A))
  try(print(summary(fit1)))
  fit1 <- NULL
}
```

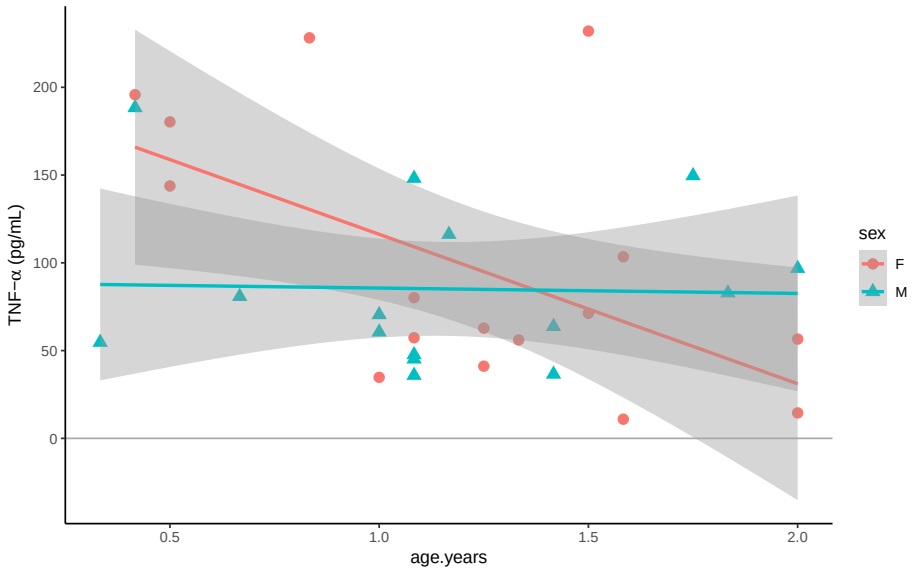
0.12.3 Model the log2FC of phenotypes vs. age for adults and young children.

```
dat$TNFa.FC <- log2(dat$TNFa.V2) - log2(dat$TNFa)
dat$IFNg.FC <- log2(dat$IFNg.V2) - log2(dat$IFNg)
dat$IL6.FC <- log2(dat$IL6.V2) - log2(dat$IL6)
dat$IL12p40.FC <- log2(dat$IL12p40.V2) - log2(dat$IL12p40)
dat$IL12p70.FC <- log2(dat$IL12p70.V2) - log2(dat$IL12p70)
dat$IL10.FC <- log2(dat$IL10.V2) - log2(dat$IL10)
dat$GMCSF.FC <- log2(dat$GMCSF.V2) - log2(dat$GMCSF)
dat$Hb.FC <- log2(dat$Hb.V2) - log2(dat$Hb)
dat$Nitrate.570.FC <- log2(dat$Nitrate.570.V2) - log2(dat$Nitrate.570)
phens <- c("TNFa", "IFNg", "IL6", "IL12p40", "IL12p70", "IL10", "GMCSF",
           "Hb", "Nitrate.570")
phens.FC <- paste0(phens, ".FC")
dat.I <- subset(dat, age="I")
dat.A <- subset(dat, age="A")

cat("##-----------------##\n")
cat("CHILDREN \n")
cat("##-----------------##\n")
for(i in 1:length(phens.FC)){
  cat("##-----------------")
  cat(phens.FC[i])
  cat("-----------------## \n")
  form <- paste0(phens.FC[i], "~ age.years*sex")
  try(fit1 <- lm(form, data=dat.I))
  try(print(summary(fit1)))
  fit1 <- NULL
}

cat("##-----------------##\n")
cat("ADULTS \n")
cat("##-----------------##\n")
for(i in 1:length(phens.FC)){
  cat("##-----------------")
  cat(phens.FC[i])
  cat("-----------------## \n")
```

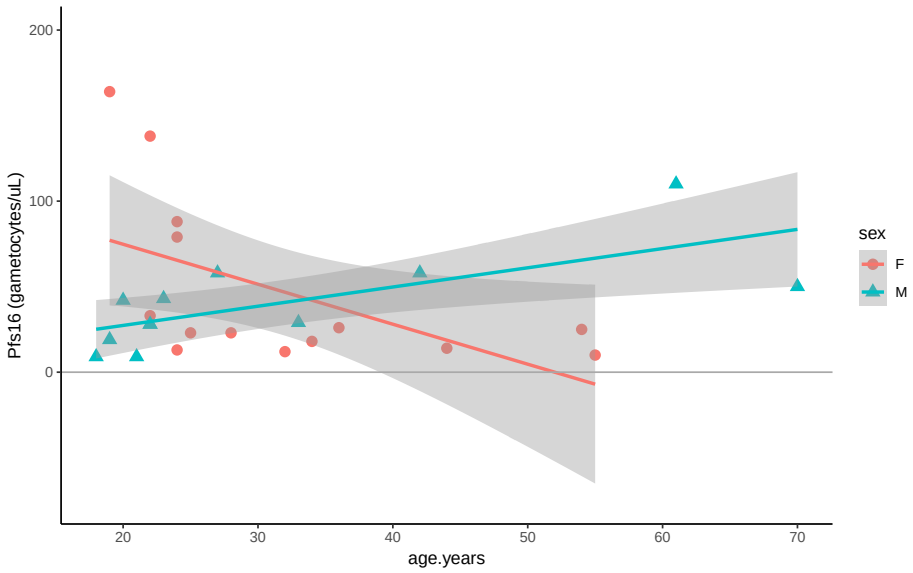
```
  form <- paste0(phens.FC[i], "~ age.years*sex")
  try(fit1 <- lm(form, data=dat.A))
  try(print(summary(fit1)))
  fit1 <- NULL
}
```

0.12.4 Plot the phenotypes vs. age for adults and young children - V1.

```
dat.I <- droplevels(subset(dat, age=="I"))
dat.A <- droplevels(subset(dat, age=="A"))

p1 <- ggplot(dat.I, aes_string(x="age.years", y="TNFa", colour="sex", shape="sex"))
p2 <- p1 + geom_point(aes(pch=sex), size=3) + geom_smooth(method=lm) +
  theme_classic() +
  geom_hline(yintercept=0, color = "darkgrey") + ylab(expression(paste("TNF-",alpha," (pg/mL)")))
plot(p2)
```

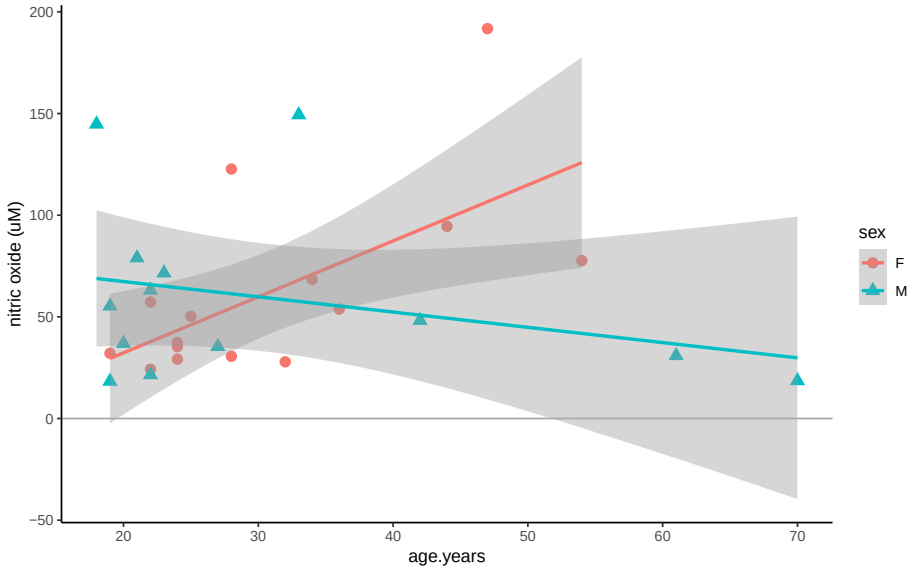



```
p1 <- ggplot(dat.A, aes_string(x="age.years", y="pfs16", colour="sex", shape="sex"))
p2 <- p1 + geom_point(aes(pch=sex), size=3) + geom_smooth(method=lm) + ylim(-75,200) +
  theme_classic() + geom_hline(yintercept=0, color = "darkgrey") + ylab("Pfs16 (gametocytes/uL)")
plot(p2)
```



0.12.5 Plot the phenotypes vs. age for adults and young children - V2.

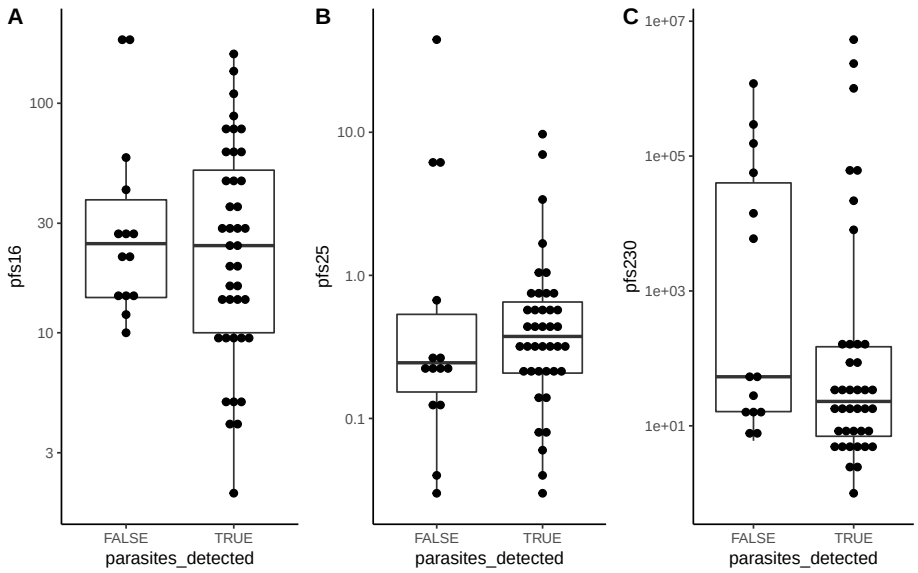
```
p1 <- ggplot(dat.A, aes_string(x="age.years", y="Nitrate.570.V2", colour="sex",
                               shape="sex"))
p2 <- p1 + geom_point(aes(pch=sex), size=3) + geom_smooth(method=lm) + theme_classic() +
  geom_hline(yintercept=0, color = "darkgrey") + ylab("nitric oxide (uM)")
plot(p2)
```



0.13 Comparison of parasites and gameocytes at V1

Compare estimated levels of gametocytemia in samples with parasitemia = 0 and parasitemia > 0.

```
dat_long$parasites_detected <- as.factor(ifelse(dat_long$parasites > 0, TRUE,
                                                ifelse(dat_long$parasites < 1, FALSE, NA)))

# subset on data for which parasite counts were obtained

dat_sub <- droplevels(subset(dat_long, !is.na(parasites_detected)))
dat_sub$AgeSex <- as.factor(paste(dat_sub$Age, dat_sub$Sex, sep="_"))

p1 <- ggplot(dat_sub, aes(x=parasites_detected, y=pfs16)) + scale_y_log10() +
  geom_boxplot(outlier.shape=NA) + theme_classic() +
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5)
p2 <- ggplot(dat_sub, aes(x=parasites_detected, y=pfs25)) + scale_y_log10() +
  geom_boxplot(outlier.shape=NA) + theme_classic() +
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5)
p3 <- ggplot(dat_sub, aes(x=parasites_detected, y=pfs230)) + scale_y_log10() +
  geom_boxplot(outlier.shape=NA) + theme_classic() +
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5)

plot_grid(p1, p2, p3, ncol = 3, labels=LETTERS[1:3])
```




```
fit1 <- wilcox_test(dat_sub$pfs25 ~ dat_sub$parasites_detected | dat_sub$AgeSex,
                    distribution="exact"); pvalue(fit1)
## [1] 0.3270372
fit2 <- wilcox_test(dat_sub$pfs16 ~ dat_sub$parasites_detected | dat_sub$AgeSex,
                    distribution="exact"); pvalue(fit2)
## [1] 0.7073396
fit3 <- wilcox_test(dat_sub$pfs230 ~ dat_sub$parasites_detected | dat_sub$AgeSex,
```

```
                        distribution="exact"); pvalue(fit3)
## [1] 0.125032
```

0.14 Effect of parasite load at V1 on transformed phenotypes examined previously

```
dat_long$scaled.parasites <- scale(dat_long$parasites, center=TRUE, scale=TRUE)

dat_long_V1 <- droplevels(subset(dat_long, Visit==1))

fit <- lm(GMCSF.bc ~ Age*Sex*scaled.parasites, data=dat_long_V1); anova(fit)[["Pr(>F)"]][3]
## [1] 0.046483

fit <- lm(IFNg.bc ~ Age*Sex*scaled.parasites, data=dat_long_V1); anova(fit)[["Pr(>F)"]][3]
## [1] 0.09511465

fit <- lm(IL10.bc ~ Age*Sex*scaled.parasites, data=dat_long_V1); anova(fit)[["Pr(>F)"]][3]
## [1] 0.03842163

fit <- lm(IL12p40.bc ~ Age*Sex*scaled.parasites, data=dat_long_V1); anova(fit)[["Pr(>F)"]][3]
## [1] 0.9186304

fit <- lm(IL12p70.bc ~ Age*Sex*scaled.parasites, data=dat_long_V1); anova(fit)[["Pr(>F)"]][3]
## [1] 0.7692469

fit <- lm(IL6.bc ~ Age*Sex*scaled.parasites, data=dat_long_V1); anova(fit)[["Pr(>F)"]][3]
## [1] 0.05399335

fit <- lm(TNFa.bc ~ Age*Sex*scaled.parasites, data=dat_long_V1); anova(fit)[["Pr(>F)"]][3]
## [1] 0.5099813

fit <- lm(Hb.bc ~ Age*Sex*scaled.parasites, data=dat_long_V1); anova(fit)[["Pr(>F)"]][3]
## [1] 0.2476384

fit <- lm(Nitrate.570.bc ~ Age*Sex*scaled.parasites, data=dat_long_V1); anova(fit)[["Pr(>F)"]][3]
## [1] 0.874238
```