Explore for RNA shedding correlates

Sanmi

6/24/2021

To explore relationship between demographic variables and shedding

* Inputs: +sgene\_drop.csv  
  +paired data (mask+nomask) for seronegative subjects:

Variables explored: bmi, age, days-past-onset

# Import data

#import sgene  
sgene\_drop<-read.csv("../working\_files/paired/paired\_sgene\_drop.csv")  
#Import paired files to plot  
pcr\_dat<-read.csv("../working\_files/paired/paired\_pred\_negsero\_rpr.csv") #replicate per row

# Count replicates per sample type

pcr\_dat%>%   
# distinct(sample\_id,.keep\_all = T)%>%  
 group\_by(sample\_type)%>%  
 summarise(subjects=length(unique(study\_id)),  
 n\_replicates=n())

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 10 x 3  
## sample\_type subjects n\_replicates  
## <fct> <int> <int>  
## 1 G-II Coarse Aerosol 47 142  
## 2 G-II Coarse Aerosol mask 47 142  
## 3 G-II Cone Swab 41 124  
## 4 G-II Cone Swab mask 41 124  
## 5 G-II Fine Aerosol 47 142  
## 6 G-II Fine Aerosol mask 47 142  
## 7 IcePac 11 32  
## 8 Midturbinate Swab 47 142  
## 9 Phone Swab 47 142  
## 10 Saliva 47 143

# Count samples provided by SGTF cases

sgene\_drop%>%  
 distinct(study\_id,sample\_id,.keep\_all = T)%>%  
 group\_by(sample\_type)%>%  
 summarise(n\_samples=n())

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 10 x 2  
## sample\_type n\_samples  
## <fct> <int>  
## 1 G-II Coarse Aerosol 6  
## 2 G-II Coarse Aerosol mask 6  
## 3 G-II Cone Swab 6  
## 4 G-II Cone Swab mask 6  
## 5 G-II Fine Aerosol 6  
## 6 G-II Fine Aerosol mask 6  
## 7 IcePac 1  
## 8 Midturbinate Swab 6  
## 9 Phone Swab 6  
## 10 Saliva 6

# Process data for plot

#steps  
#replace -1 with NA for the non-detects  
#log10(x+1),x=rna copies this will keeo non-detects as zero  
#replace NA with 41 for the non-detects  
  
pcr\_dat<-pcr\_dat%>%  
 group\_by(sample\_id)%>%  
 mutate(av\_ngene=log10(av\_quant+1))%>% #added 1 to the copy numbers to avoid infinity error when estimating log 0  
 distinct(sample\_id,.keep\_all = T)%>% #keep one replicate per sample  
 ungroup()  
  
  
#Sample type X Subjects X Samples  
pcr\_dat%>%   
 group\_by(sample\_type)%>%  
 summarise(subjects=length(unique(study\_id)),  
 samples=n())

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 10 x 3  
## sample\_type subjects samples  
## <fct> <int> <int>  
## 1 G-II Coarse Aerosol 47 71  
## 2 G-II Coarse Aerosol mask 47 71  
## 3 G-II Cone Swab 41 62  
## 4 G-II Cone Swab mask 41 62  
## 5 G-II Fine Aerosol 47 71  
## 6 G-II Fine Aerosol mask 47 71  
## 7 IcePac 11 16  
## 8 Midturbinate Swab 47 71  
## 9 Phone Swab 47 71  
## 10 Saliva 47 71

# Contruct datframe for the correlation plots

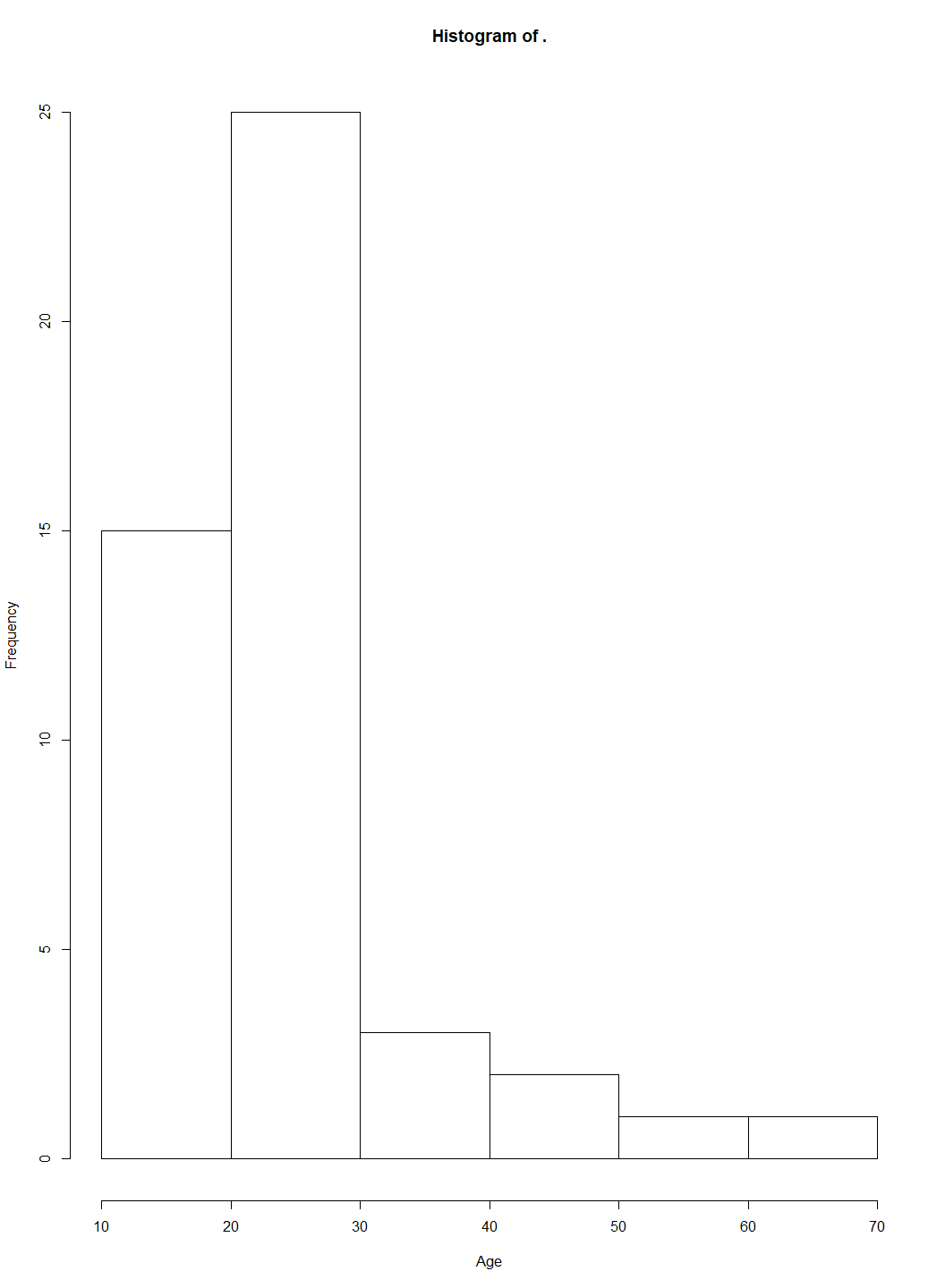
#remove phone swabs for the correlation plots  
plot\_cor<-pcr\_dat%>%  
 filter(sample\_type!="Phone Swab") #remove phone swabs  
  
plot\_cor$S.gene.dropout <- factor(plot\_cor$S.gene.dropout, levels=c(T,F), labels=c("True","False")) #adjust levels for proper display  
  
plot\_cor<-plot\_cor%>%  
 select(study\_id,sample\_date,S.gene.dropout,av\_ngene,sample\_type,ct\_ngene\_qpcr,ct\_sgene\_qpcr,age\_entry,days\_past\_onset,bmi)  
  
#quick count post average per sample  
plot\_cor%>%  
 group\_by(sample\_type)%>%  
 summarise(subjects=length(unique(study\_id)),  
 samples=n())

## `summarise()` ungrouping output (override with `.groups` argument)

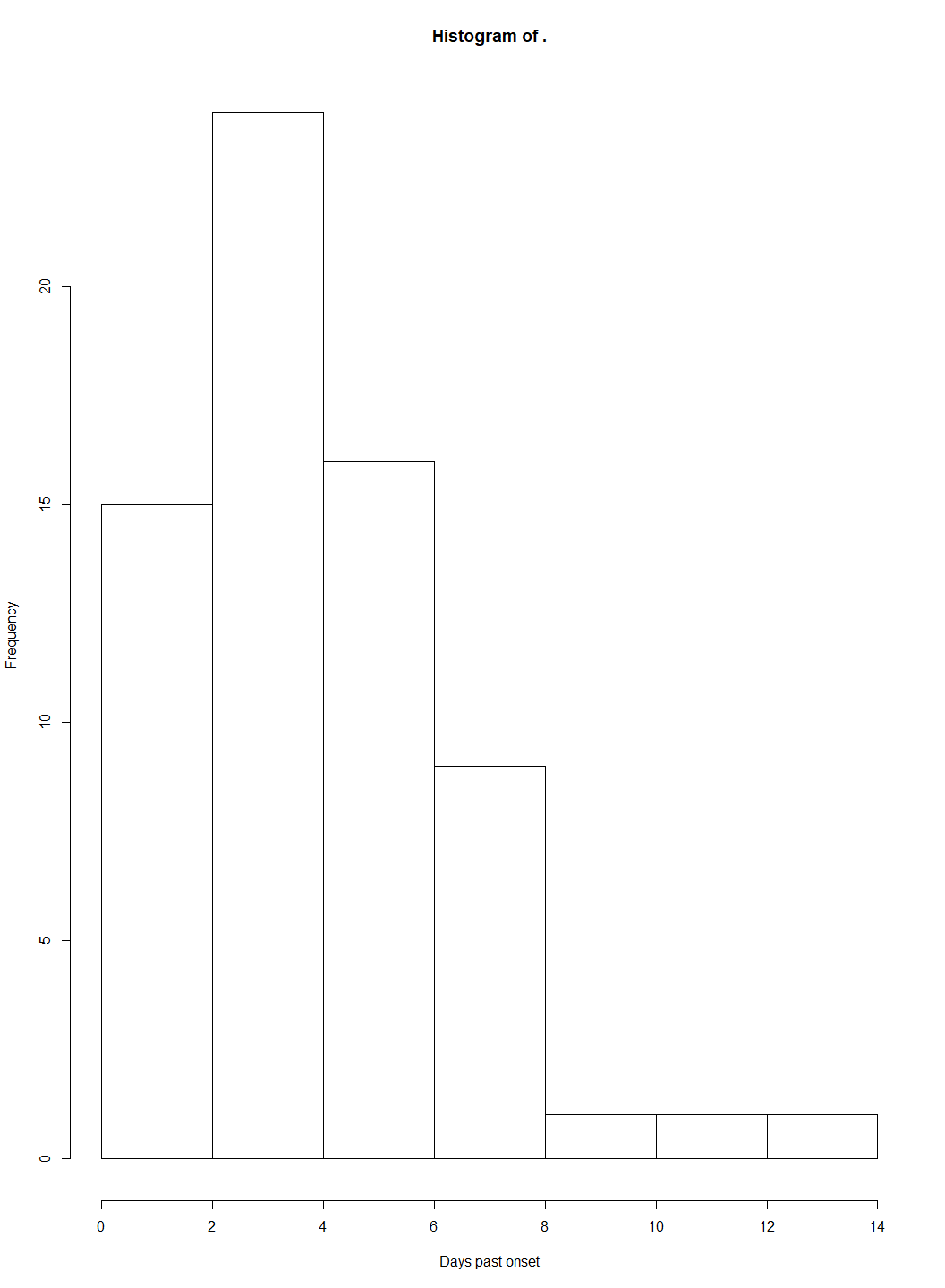
## # A tibble: 9 x 3  
## sample\_type subjects samples  
## <fct> <int> <int>  
## 1 G-II Coarse Aerosol 47 71  
## 2 G-II Coarse Aerosol mask 47 71  
## 3 G-II Cone Swab 41 62  
## 4 G-II Cone Swab mask 41 62  
## 5 G-II Fine Aerosol 47 71  
## 6 G-II Fine Aerosol mask 47 71  
## 7 IcePac 11 16  
## 8 Midturbinate Swab 47 71  
## 9 Saliva 47 71

# Distribution age,bmi,days-post-onset

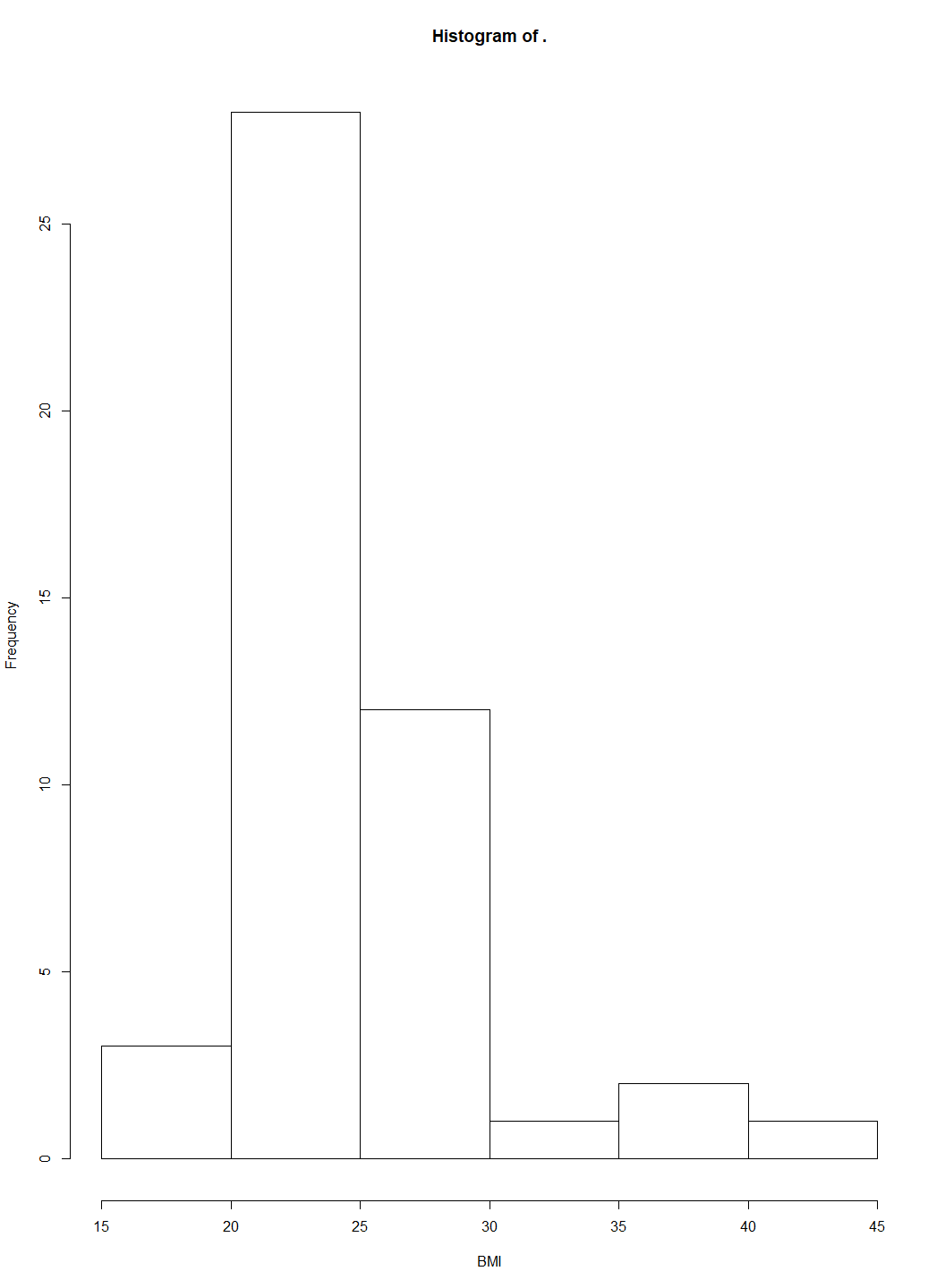
#Age  
plot\_cor%>%  
 distinct(study\_id,.keep\_all = T)%>%  
 pull(age\_entry)%>%  
 hist(xlab="Age")



#Days past onset  
plot\_cor%>%  
 distinct(study\_id,sample\_date,.keep\_all = T)%>%  
 pull(days\_past\_onset)%>%  
 hist(xlab="Days past onset")



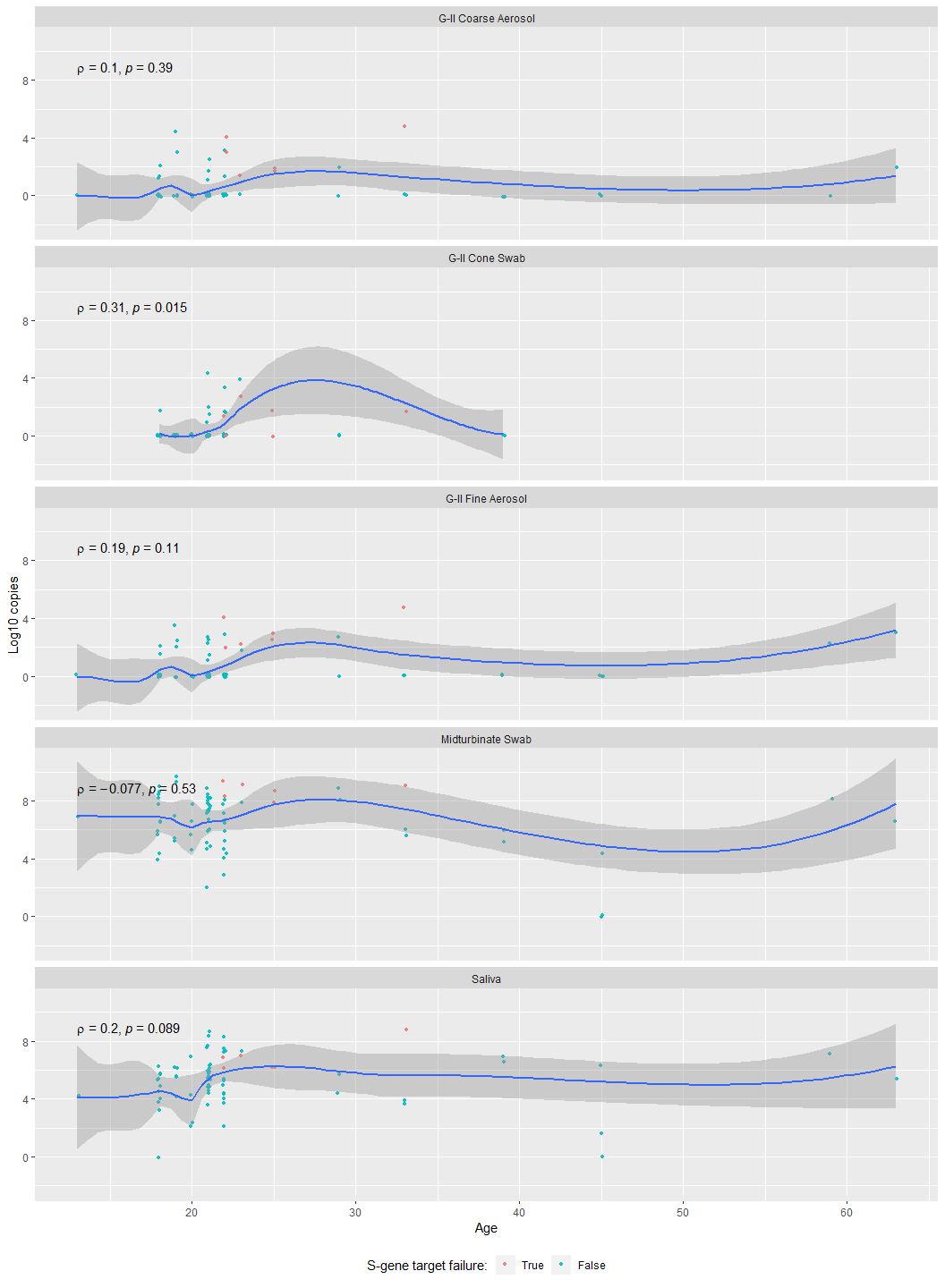
#Body mass index  
plot\_cor%>%  
 distinct(study\_id,.keep\_all = T)%>%  
 pull(bmi)%>%  
 hist(xlab="BMI")



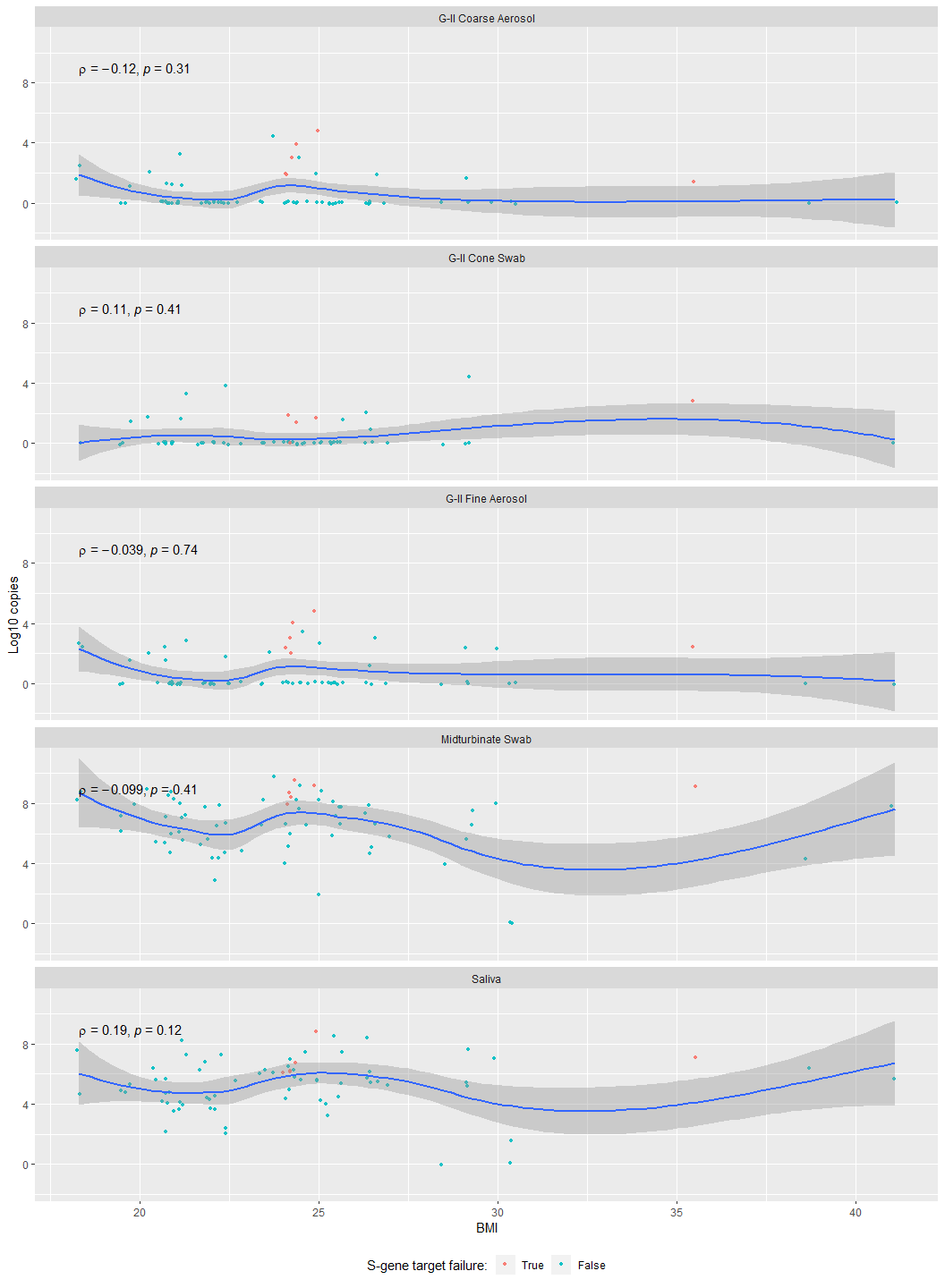
# Cor plots

## Age

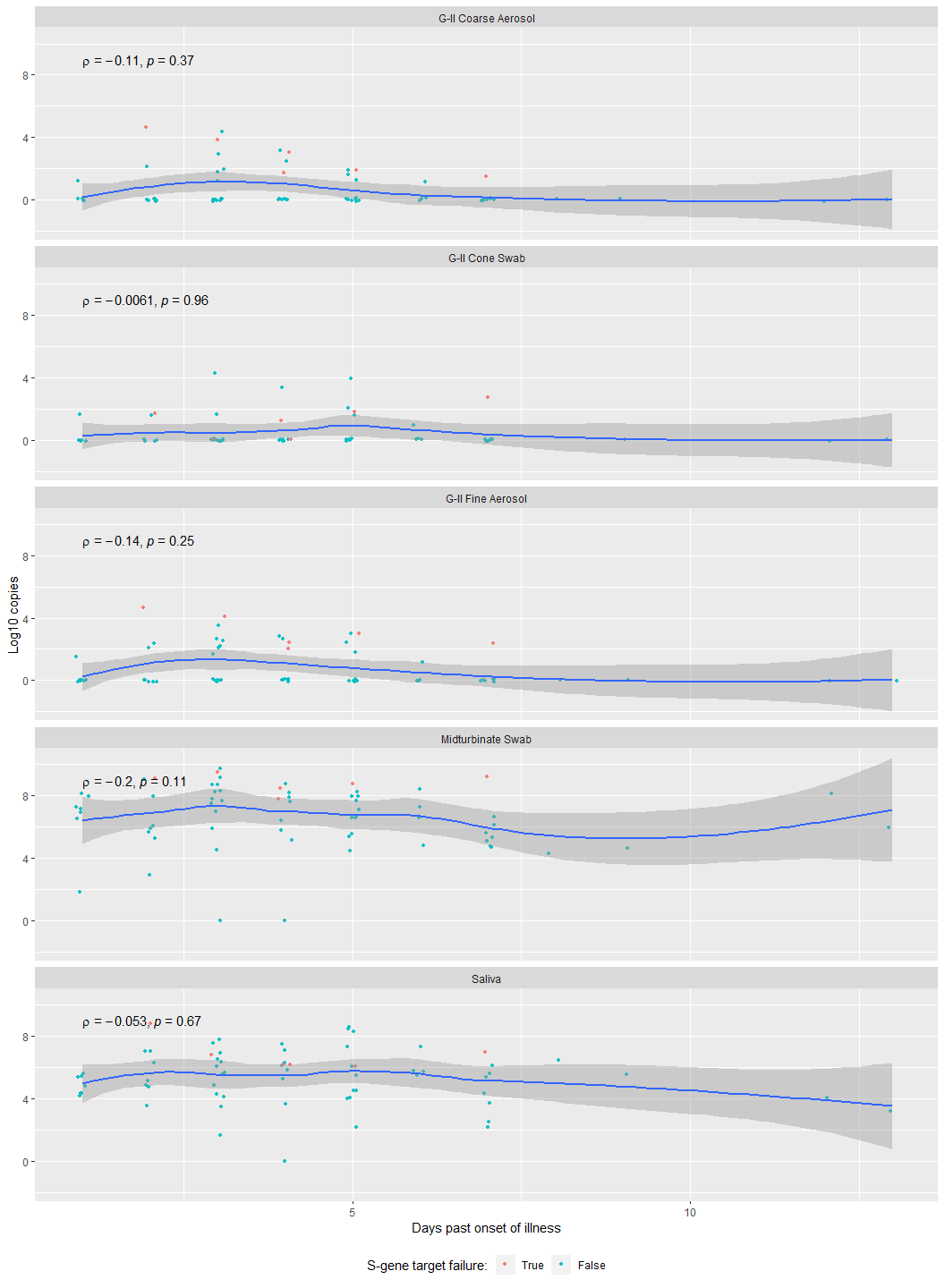
plot\_cor%>%  
 filter(sample\_type%in%c("G-II Cone Swab","G-II Coarse Aerosol","G-II Fine Aerosol","Midturbinate Swab","Saliva"))%>%  
 rename(`S-gene target failure:`=S.gene.dropout)%>%  
 ggplot(aes(x=age\_entry,y=av\_ngene))+  
 geom\_point(aes(color=`S-gene target failure:`),size=1,alpha=0.9,position=position\_jitter(h=.1,w=.1))+  
 geom\_smooth(method = "loess")+  
 stat\_cor(method = "spearman",cor.coef.name = "rho")+  
 facet\_wrap(. ~ sample\_type,ncol = 1)+  
 ylab("Log10 copies")+  
 xlab("Age")+  
 theme(legend.position = "bottom")

 ## BMI

#BMI  
plot\_cor%>%  
 filter(sample\_type%in%c("G-II Cone Swab","G-II Coarse Aerosol","G-II Fine Aerosol","Midturbinate Swab","Saliva"))%>%  
 rename(`S-gene target failure:`=S.gene.dropout)%>%  
 ggplot(aes(x=bmi,y=av\_ngene))+  
 geom\_point(aes(color=`S-gene target failure:`),size=1,alpha=0.9,position=position\_jitter(h=.1,w=.1))+  
 geom\_smooth(method = "loess")+  
 stat\_cor(method = "spearman",cor.coef.name = "rho")+  
 facet\_wrap(. ~ sample\_type,ncol = 1)+  
 ylab("Log10 copies")+  
 xlab("BMI")+  
 theme(legend.position = "bottom")

 ## Days past onset

#Days past onset  
plot\_cor%>%  
 filter(sample\_type%in%c("G-II Cone Swab","G-II Coarse Aerosol","G-II Fine Aerosol","Midturbinate Swab","Saliva"))%>%  
 rename(`S-gene target failure:`=S.gene.dropout)%>%  
 ggplot(aes(x=days\_past\_onset,y=av\_ngene))+  
 geom\_point(aes(color=`S-gene target failure:`),size=1,alpha=0.9,position=position\_jitter(h=.1,w=.1))+  
 geom\_smooth(method = "loess")+  
 stat\_cor(method = "spearman",cor.coef.name = "rho")+  
 facet\_wrap(. ~ sample\_type,ncol = 1)+  
 ylab("Log10 copies")+  
 xlab("Days past onset of illness")+  
 theme(legend.position = "bottom")



# Distribution of the number samples contributed by 1 subject

plot\_cor%>%  
 group\_by(sample\_type,study\_id)%>%  
 summarize(n=length(unique(study\_id)))

## `summarise()` regrouping output by 'sample\_type' (override with `.groups` argument)

## # A tibble: 375 x 3  
## # Groups: sample\_type [9]  
## sample\_type study\_id n  
## <fct> <int> <int>  
## 1 G-II Coarse Aerosol 56 1  
## 2 G-II Coarse Aerosol 57 1  
## 3 G-II Coarse Aerosol 58 1  
## 4 G-II Coarse Aerosol 88 1  
## 5 G-II Coarse Aerosol 107 1  
## 6 G-II Coarse Aerosol 144 1  
## 7 G-II Coarse Aerosol 180 1  
## 8 G-II Coarse Aerosol 222 1  
## 9 G-II Coarse Aerosol 227 1  
## 10 G-II Coarse Aerosol 237 1  
## # ... with 365 more rows

#i.e. X subjects provided n samples of a given sample type  
plot\_cor%>%  
 group\_by(study\_id,sample\_type)%>%  
 summarize(`no of samples provided`=n())%>%  
 group\_by(`no of samples provided`,sample\_type)%>%  
 summarize(`no of subjects`=n())%>%  
 ggplot(aes(x=factor(`no of samples provided`),y=`no of subjects`))+geom\_bar(stat = "identity")+facet\_wrap(.~sample\_type)

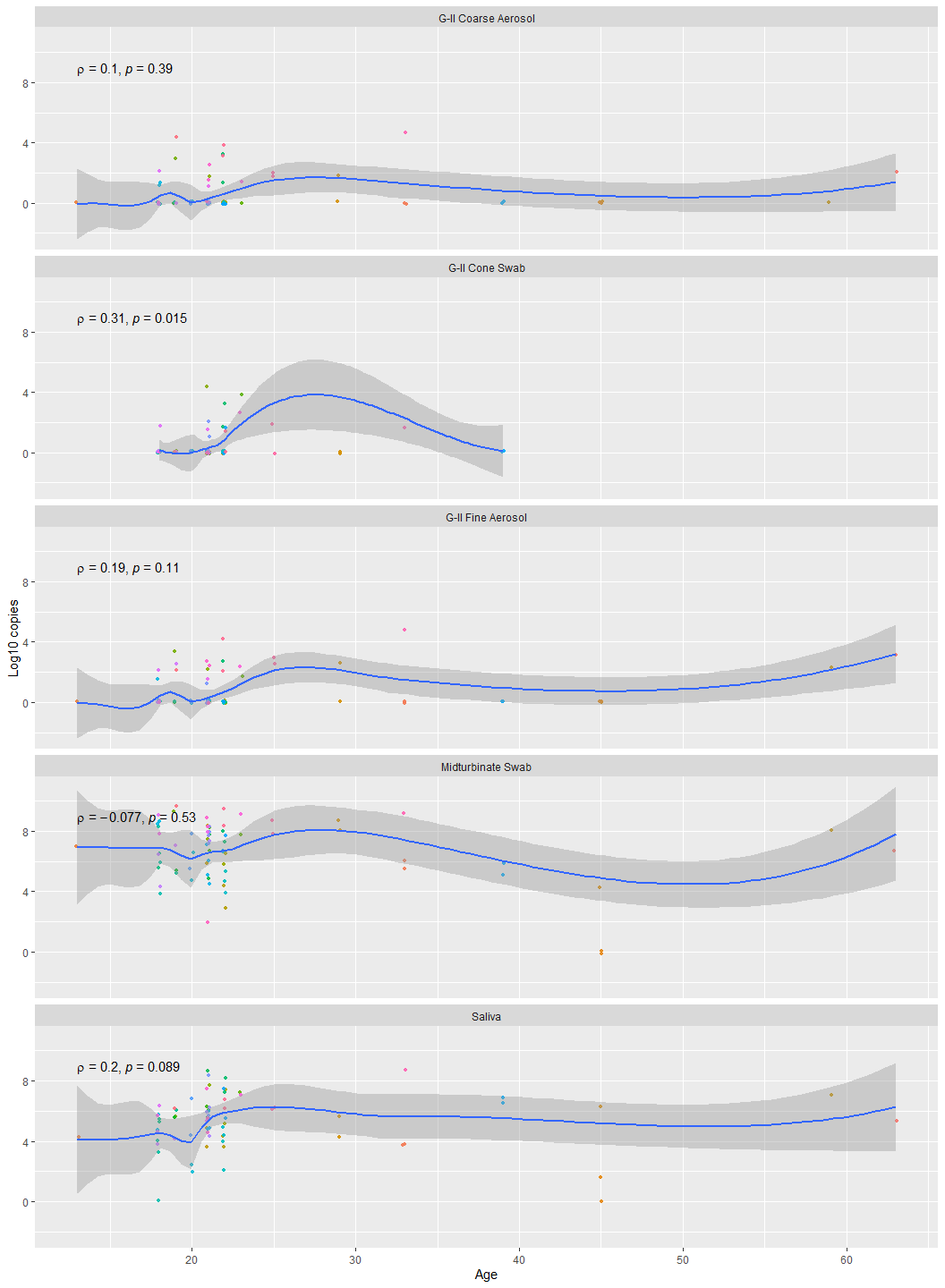
## `summarise()` regrouping output by 'study\_id' (override with `.groups` argument)

## `summarise()` regrouping output by 'no of samples provided' (override with `.groups` argument)

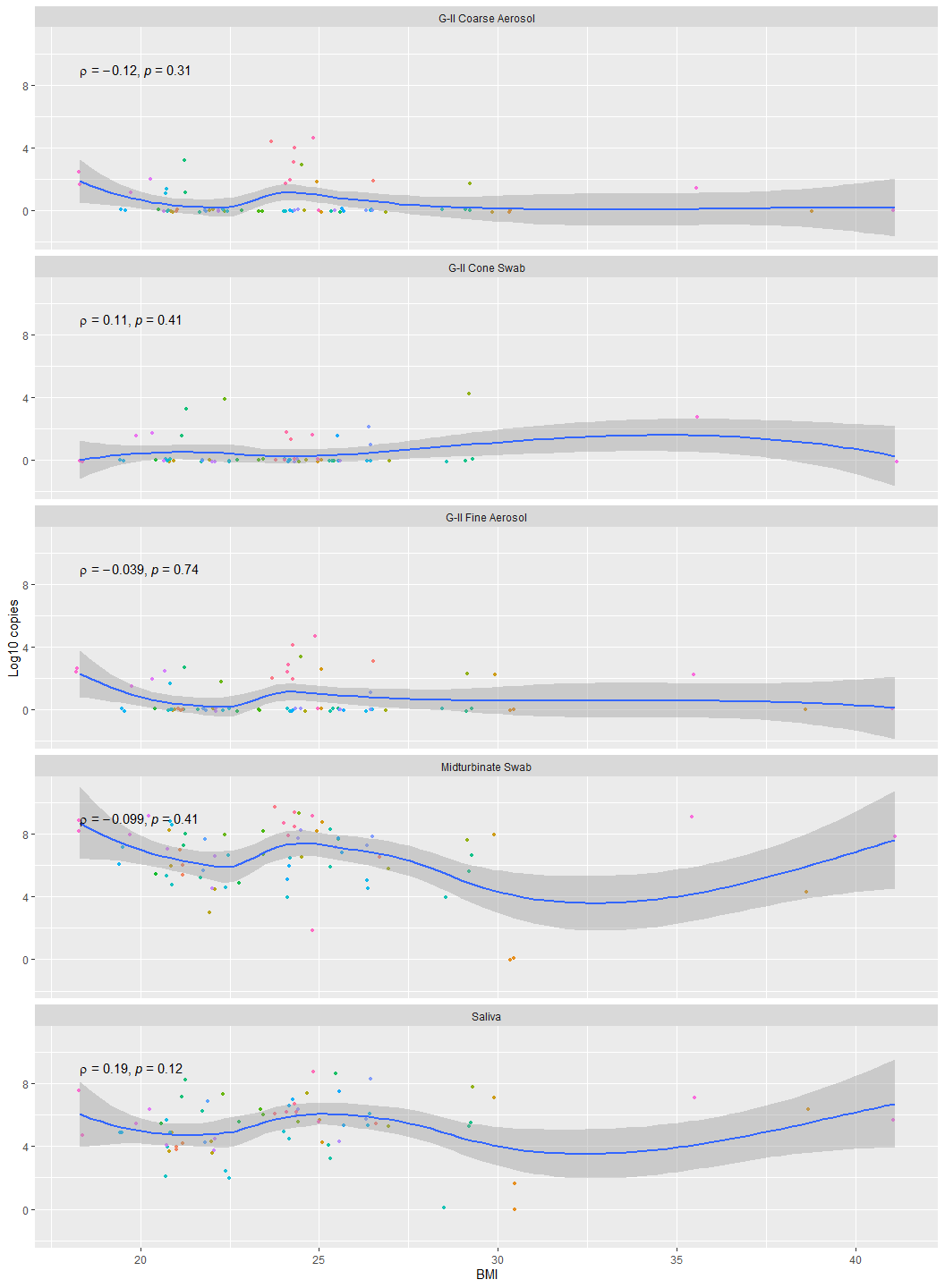


# Cor plots with study ids colored

#Age   
plot\_cor%>%  
 filter(sample\_type%in%c("G-II Cone Swab","G-II Coarse Aerosol","G-II Fine Aerosol","Midturbinate Swab","Saliva"))%>%  
 # rename(`S-gene target failure:`=S.gene.dropout)%>%  
 ggplot(aes(x=age\_entry,y=av\_ngene))+  
 geom\_point(aes(color=as.factor(study\_id)),size=1,alpha=0.9,position=position\_jitter(h=.1,w=.1))+  
 geom\_smooth(method = "loess")+  
 stat\_cor(method = "spearman",cor.coef.name = "rho")+  
 facet\_wrap(. ~ sample\_type,ncol = 1)+  
 ylab("Log10 copies")+  
 xlab("Age")+  
 theme(legend.position = "none")



#BMI  
plot\_cor%>%  
 filter(sample\_type%in%c("G-II Cone Swab","G-II Coarse Aerosol","G-II Fine Aerosol","Midturbinate Swab","Saliva"))%>%  
 rename(`S-gene target failure:`=S.gene.dropout)%>%  
 ggplot(aes(x=bmi,y=av\_ngene))+  
 geom\_point(aes(color=as.factor(study\_id)),size=1,alpha=0.9,position=position\_jitter(h=.1,w=.1))+  
 geom\_smooth(method = "loess")+  
 stat\_cor(method = "spearman",cor.coef.name = "rho")+  
 facet\_wrap(. ~ sample\_type,ncol = 1)+  
 ylab("Log10 copies")+  
 xlab("BMI")+  
 theme(legend.position = "none")



#Days past onset  
plot\_cor%>%  
 filter(sample\_type%in%c("G-II Cone Swab","G-II Coarse Aerosol","G-II Fine Aerosol","Midturbinate Swab","Saliva"))%>%  
 rename(`S-gene target failure:`=S.gene.dropout)%>%  
 ggplot(aes(x=days\_past\_onset,y=av\_ngene))+  
 geom\_point(aes(color=as.factor(study\_id)),size=1,alpha=0.9,position=position\_jitter(h=.1,w=.1))+  
 geom\_smooth(method = "loess")+  
 stat\_cor(method = "spearman",cor.coef.name = "rho")+  
 facet\_wrap(. ~ sample\_type,ncol = 1)+  
 ylab("Log10 copies")+  
 xlab("Days past onset of illness")+  
 theme(legend.position = "none")

