

main

July 23, 2021

1 Quality Control for the Caudate Nucleus

```
[1]: suppressMessages({library(repr)
                      library(ggpubr)
                      library(tidyverse)
                      library(moderndive)})
```

1.1 Functions

```
[2]: save_img <- function(image, fn, w, h){
  for(ext in c(".svg", ".pdf", ".png")){
    ggsave(file=paste0(fn, ext), plot=image, width=w, height=h)
  }
}

get_pheno <- function(){
  phenotypes <- data.table::fread(paste0("/ceph/projects/v4_phase3_paper/
→inputs/", 
                                         "phenotypes/_m/merged_phenotypes.
→csv")) %>%
  mutate_if(is.character, as.factor)
  return(phenotypes)
}

memPHENO <- memoise::memoise(get_pheno)
```

1.2 Covariates examination

1.2.1 Examine if there are significant differences between diagnosis in continuous covariates

```
[3]: covarsCont = memPHENO() %>%
  select(-c('RNum', 'Race', 'Sex', 'Dx', 'Region', 'BrNum',
          "antipsychotics", "lifetime_antipsych", "Protocol")) %>% colnames
options(repr.plot.width=12, repr.plot.height=6)
dir.create("covariate_plots")
for(covar in covarsCont){
  set.seed(20210723)
```

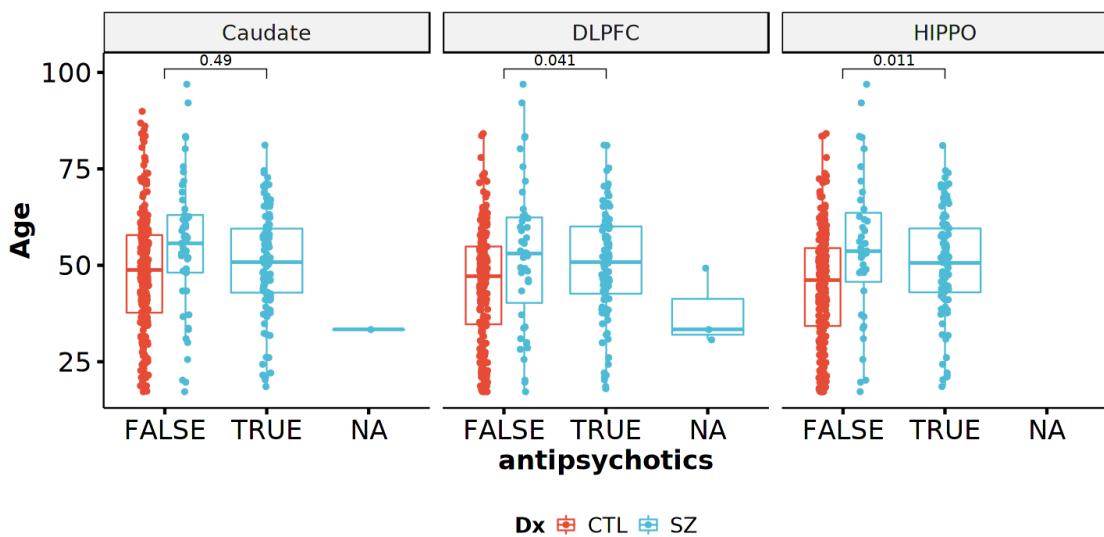
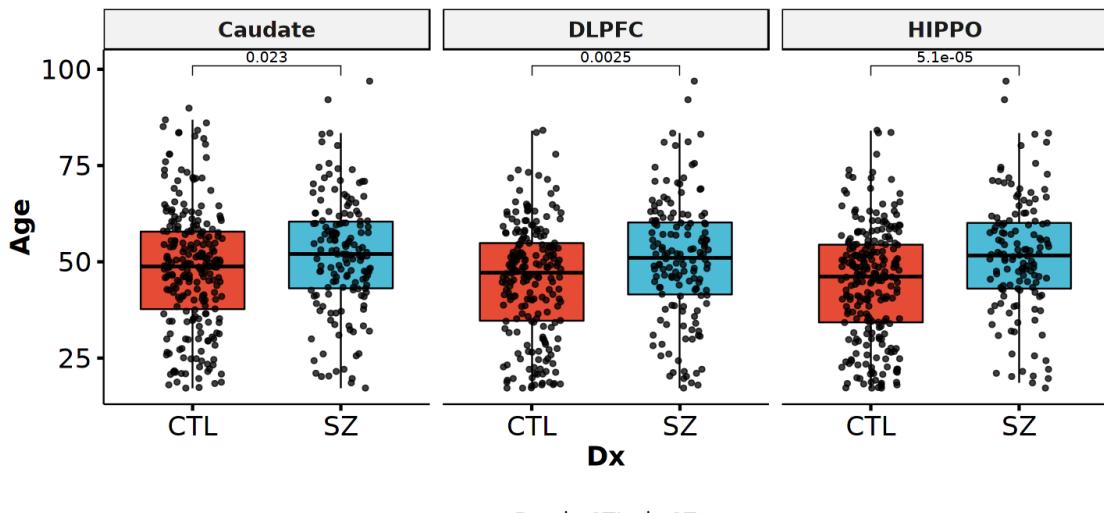
```

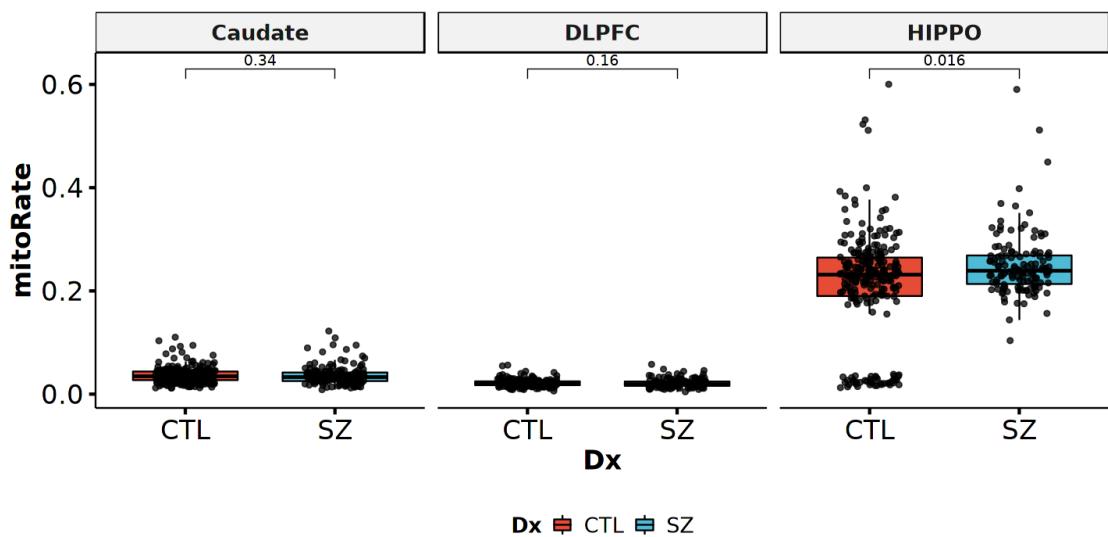
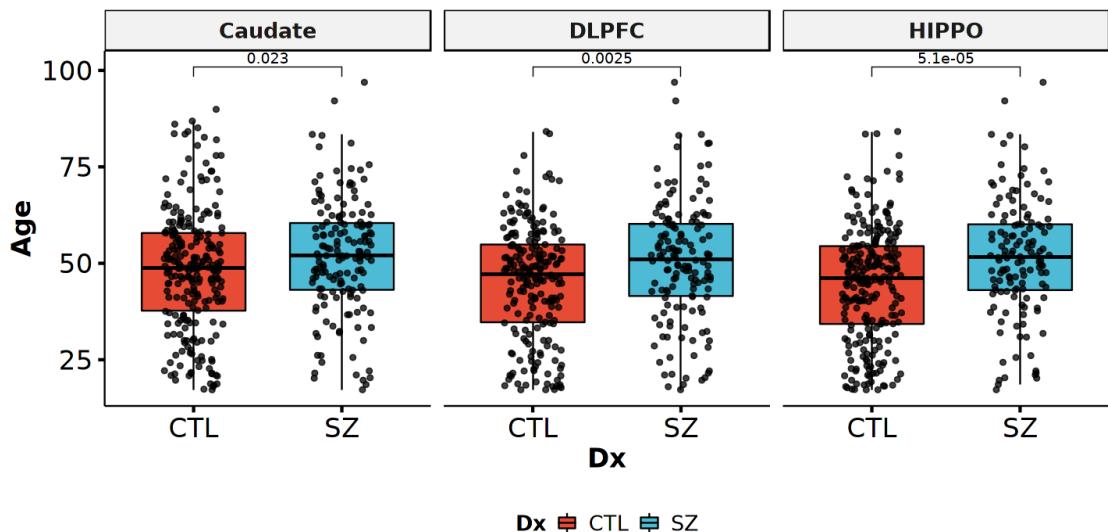
flush.console()
bxp = memPHENO() %>%
  filter(Age > 17, Race %in% c("AA", "EA"), Dx %in% c("CTL", "SZ")) %>%
  ggboxplot(x='Dx', y=covar, fill='Dx', add='jitter', palette="npg",
             facet.by="Region", legend='bottom', panel.labs.
  ~font=list(face='bold'),
             add.params=list(alpha=0.75), ~
  ~ggtheme=theme_pubr(base_size=20)) +
  stat_compare_means(comparisons=list(c("CTL", "SZ")),
                     aes(label=..p.signif..), method="wilcox.test") +
  font("title", color="black", face="bold") +
  font("xy.title", face="bold") + #font("xy.text", size=14) +
  font("legend.title", size=16, face="bold") #font("legend.text", size=14)
  save_img(bxp, paste0("covariate_plots/boxplot_dx_filtered_", covar), w=12, ~
  ~h=6)
  print(bxp)
bxp = memPHENO() %>%
  filter(Age > 17, Race %in% c("AA", "EA"), Dx %in% c("CTL", "SZ")) %>%
  ggboxplot(x='antipsychotics', y=covar, color='Dx', add='jitter',
             palette="npg", facet.by="Region", legend='bottom',
             ggtheme=theme_pubr(base_size=20)) +
  stat_compare_means(comparisons=list(c("TRUE", "FALSE")),
                     aes(label=..p.signif..), method="wilcox.test") +
  font("title", color="black", face="bold") +
  font("xy.title", face="bold") +
  font("legend.title", size=16, face="bold")
  save_img(bxp, paste0("covariate_plots/
  ~boxplot_dxBYantipsychotics_filtered_", covar),
             w=12, h=6)
  print(bxp)
bxp = memPHENO() %>%
  filter(Age > 17, Race %in% c("AA", "EA"), Dx %in% c("CTL", "SZ")) %>%
  ggboxplot(x='Dx', y=covar, fill='Dx', add='jitter', palette="npg",
             facet.by="Region", legend='bottom', panel.labs.
  ~font=list(face='bold'),
             add.params=list(alpha=0.75), ~
  ~ggtheme=theme_pubr(base_size=20)) +
  stat_compare_means(comparisons=list(c("CTL", "SZ")),
                     aes(label=..p.signif..), method="wilcox.test") +
  font("title", color="black", face="bold") +
  font("xy.title", face="bold") + #font("xy.text", size=14) +
  font("legend.title", size=16, face="bold") #font("legend.text", size=14)
print(bxp)
save_img(bxp, paste0("covariate_plots/boxplot_dx_all_", covar), w=12, h=6)
}

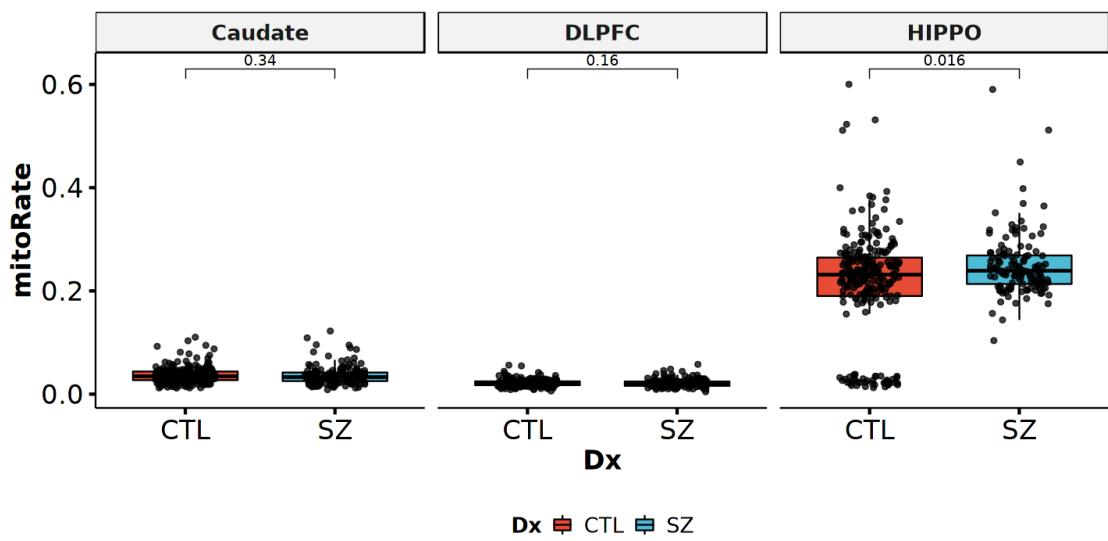
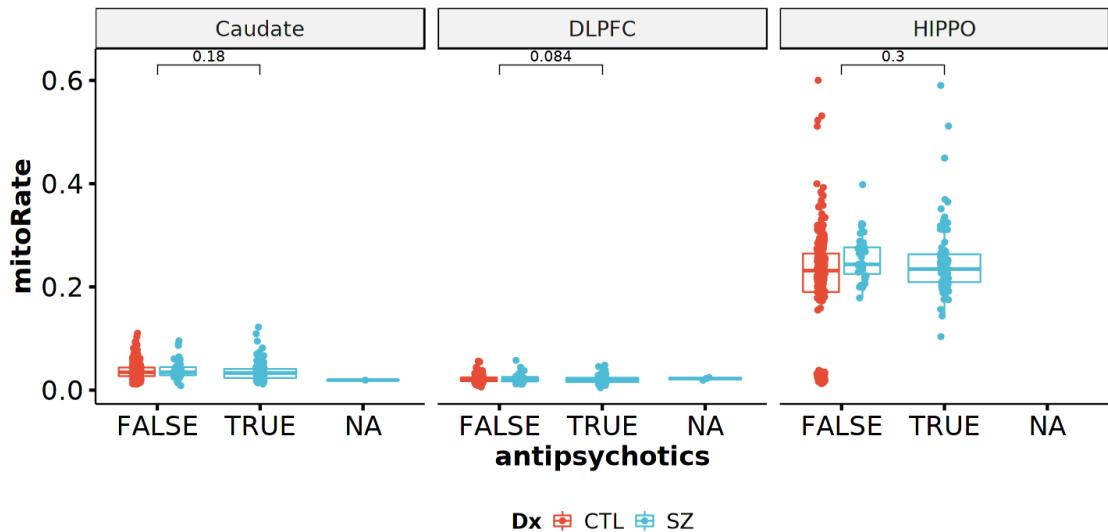
```

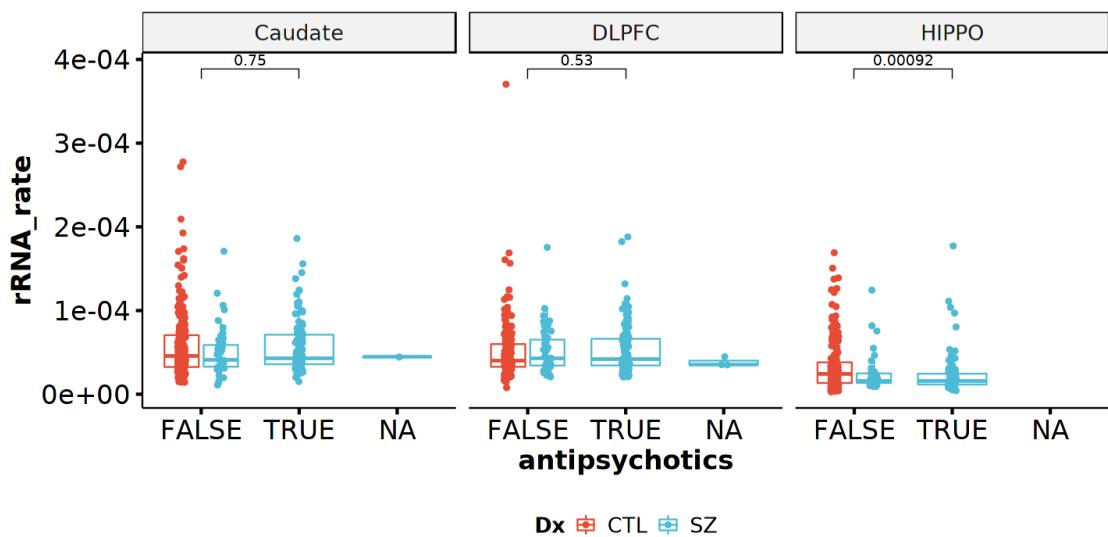
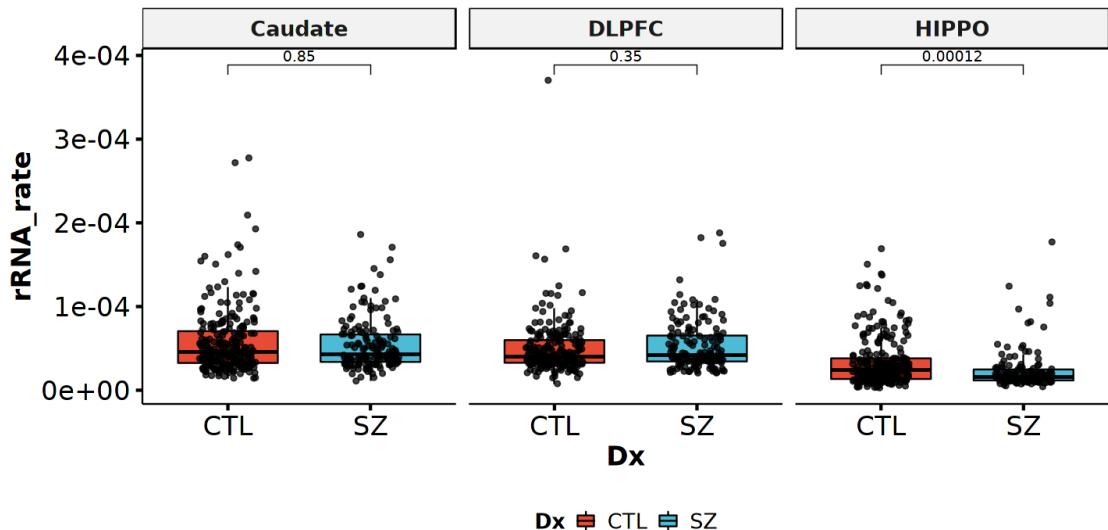
Warning message:

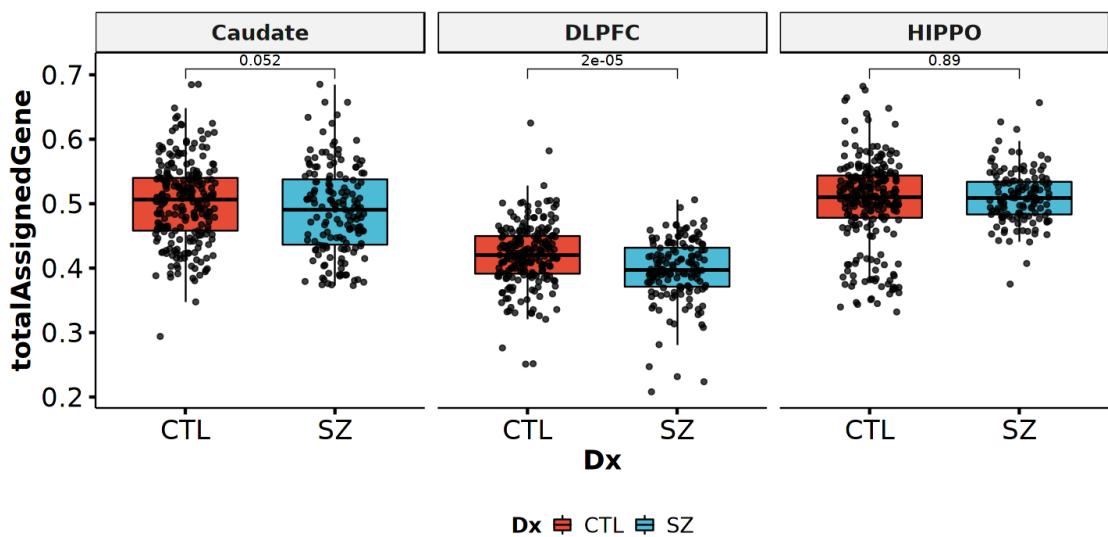
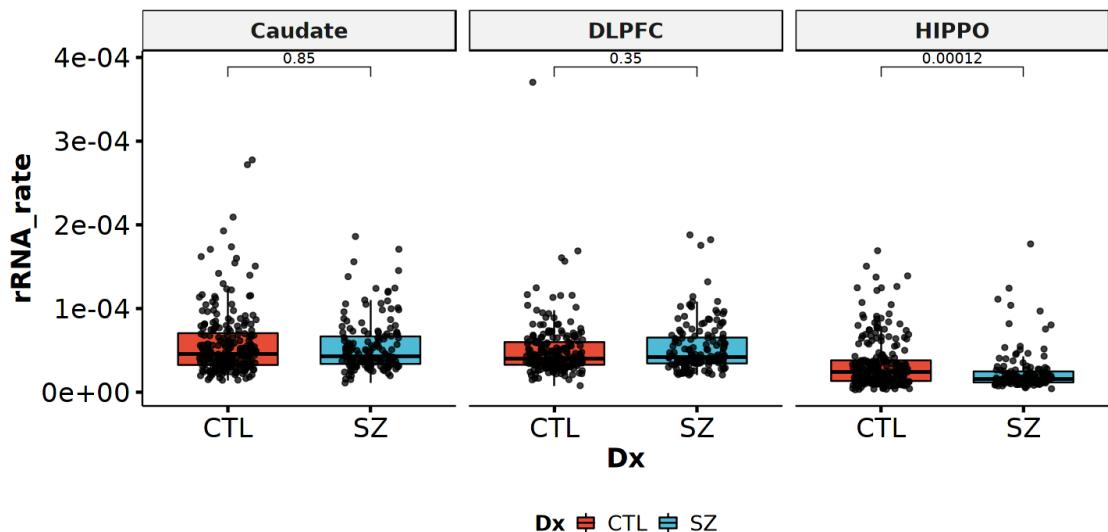
"Using `as.character()` on a quosure is deprecated as of rlang 0.3.0.
 Please use `as_label()` or `as_name()` instead.
 This warning is displayed once per session."

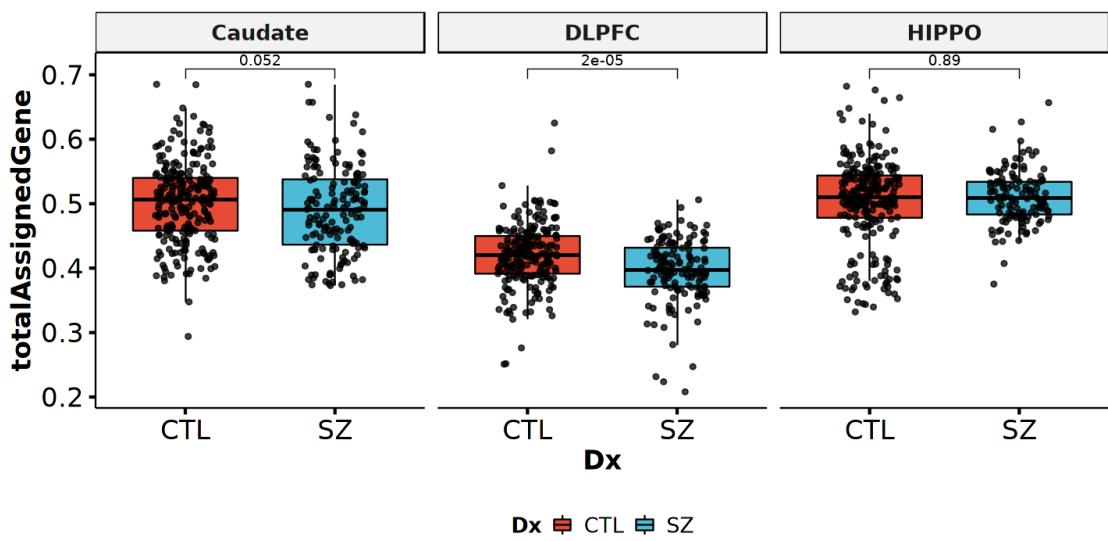
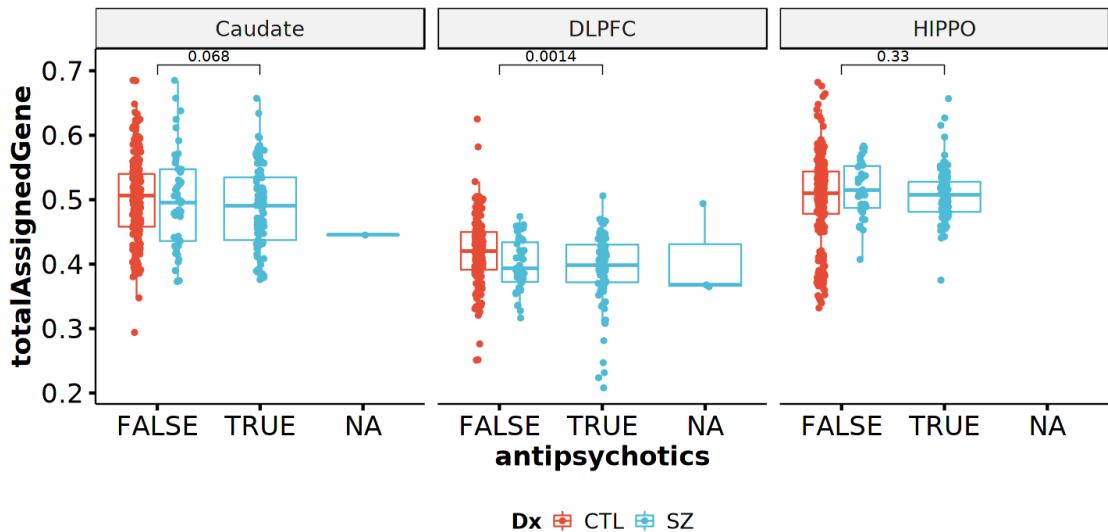


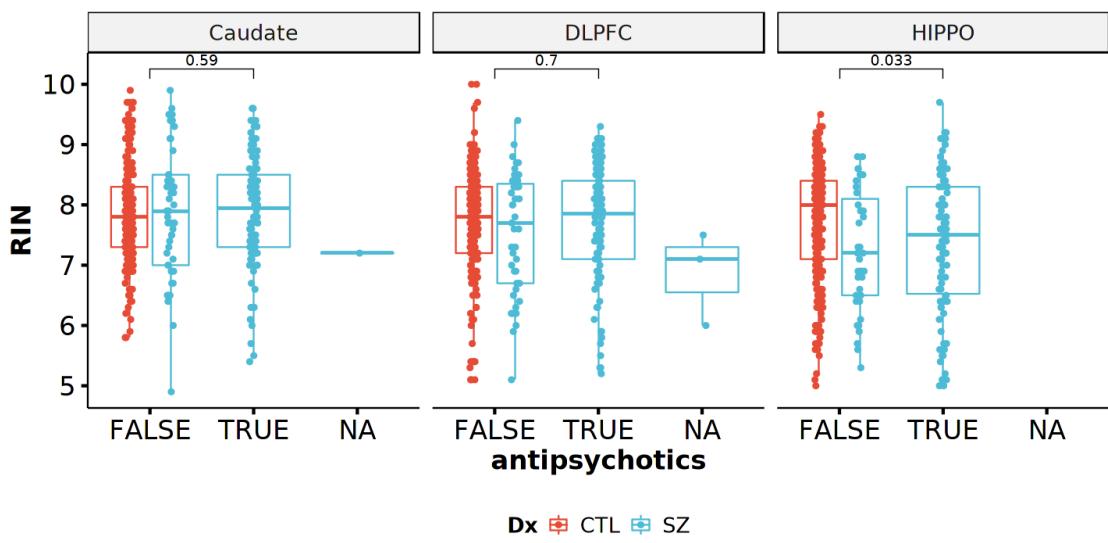
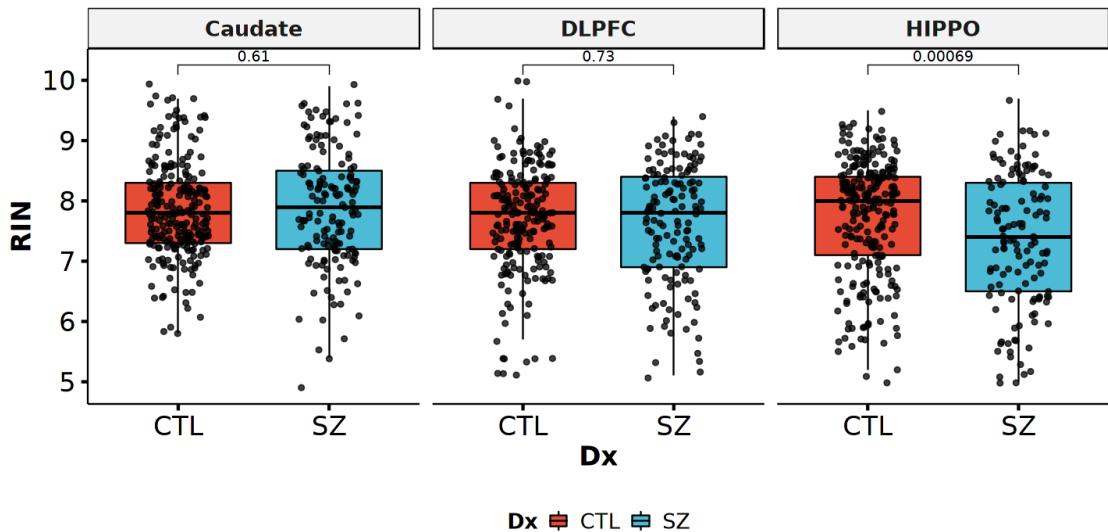


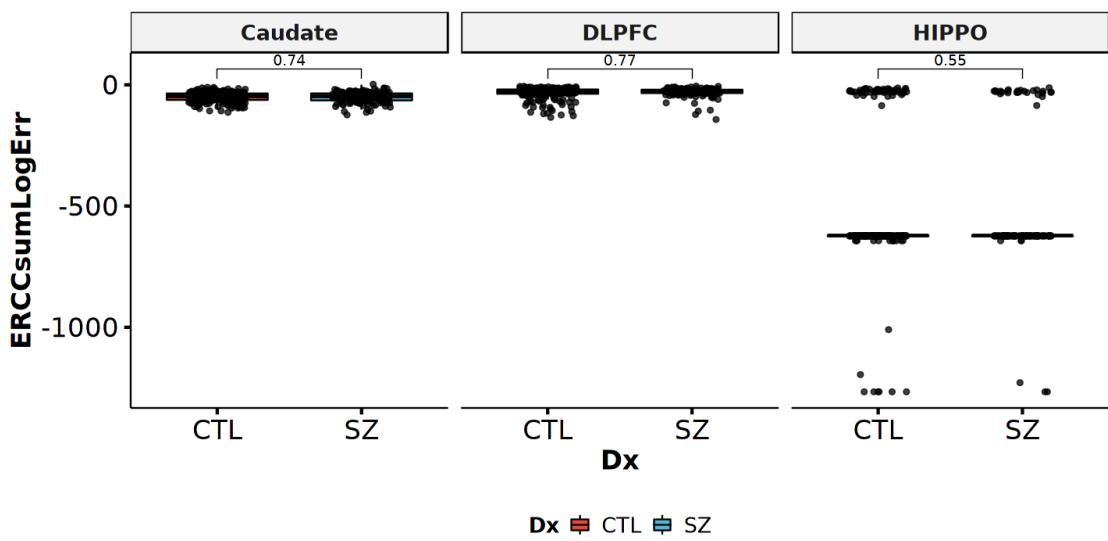
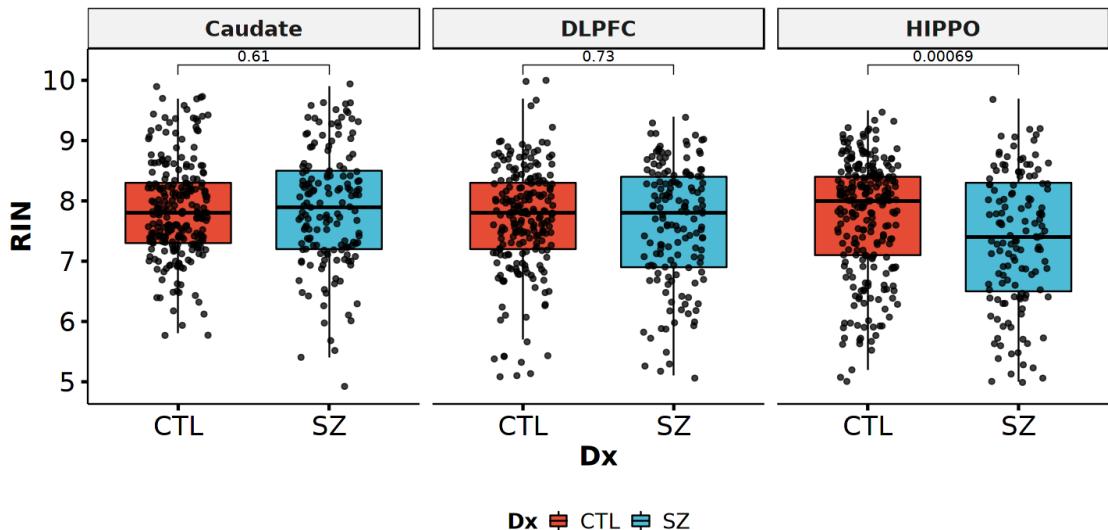


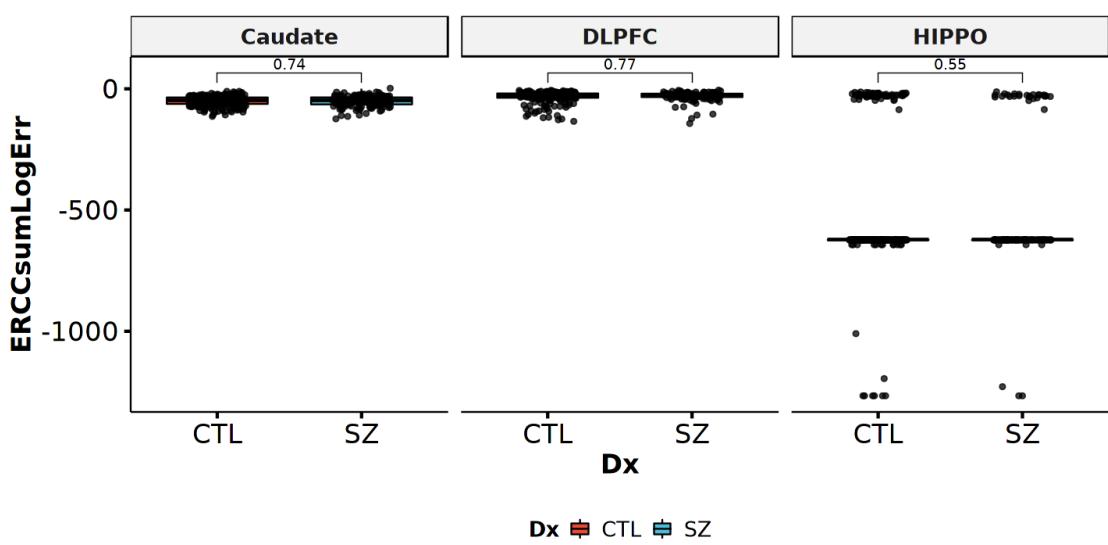
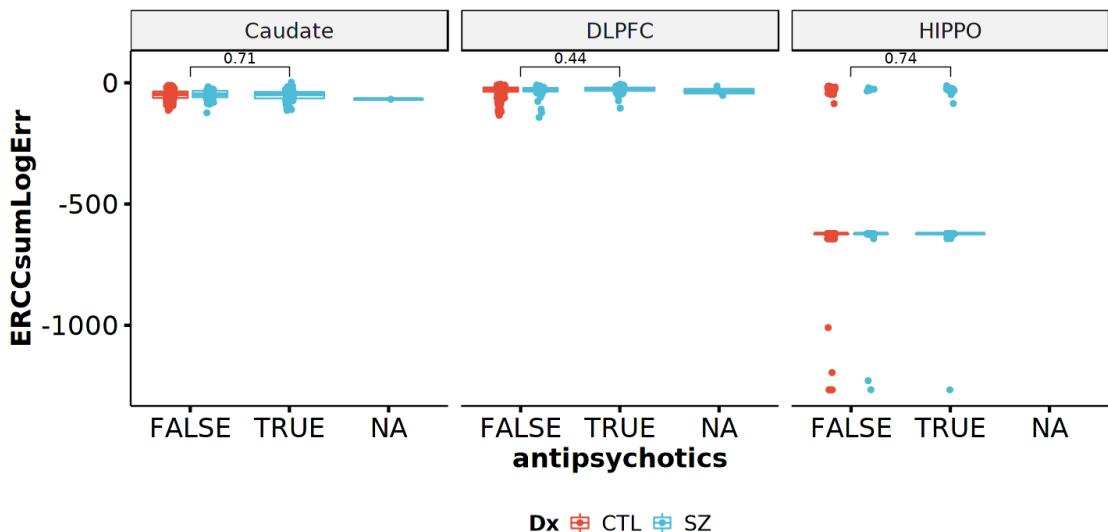


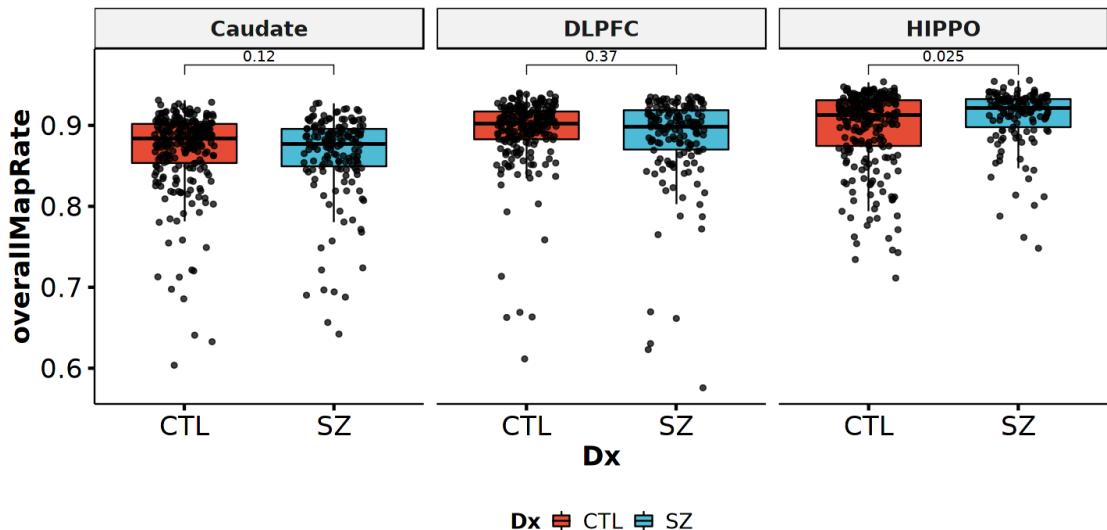












Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

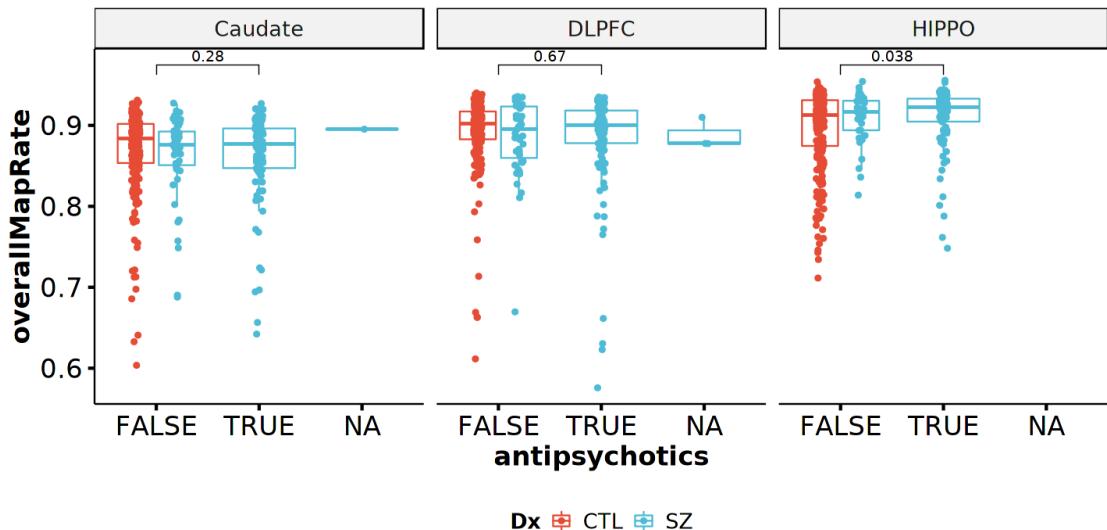
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."



Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

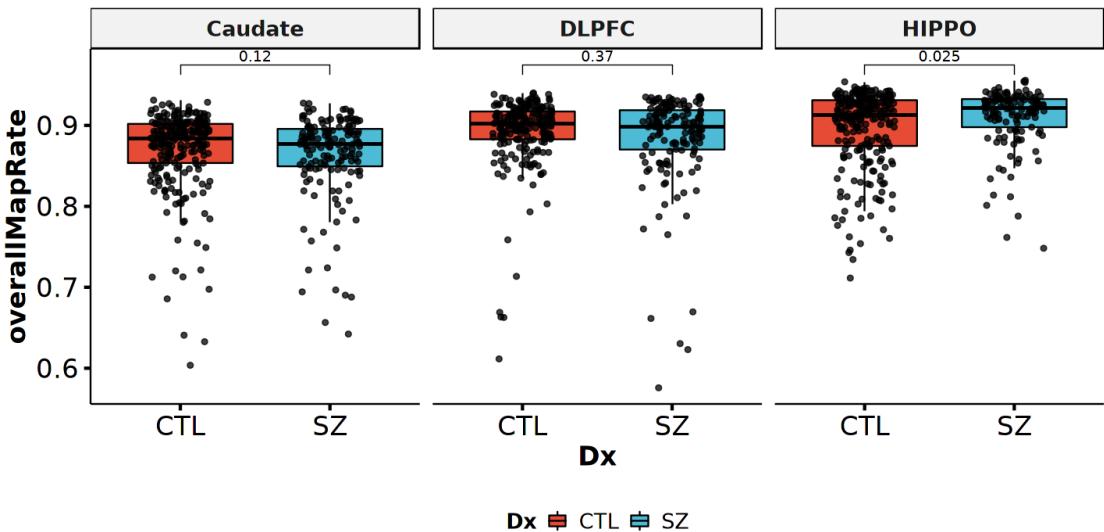
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."



Warning message:

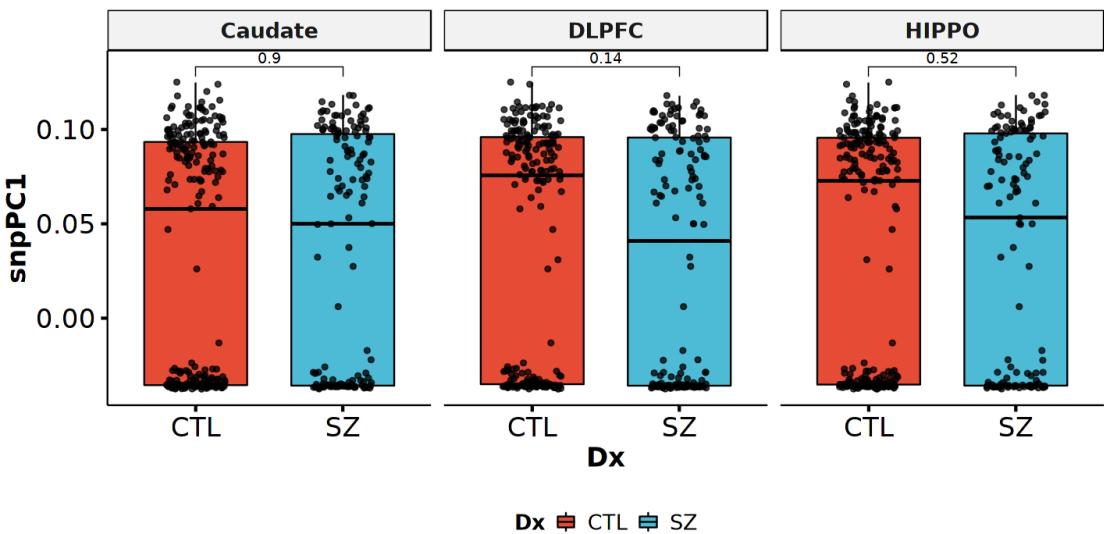
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

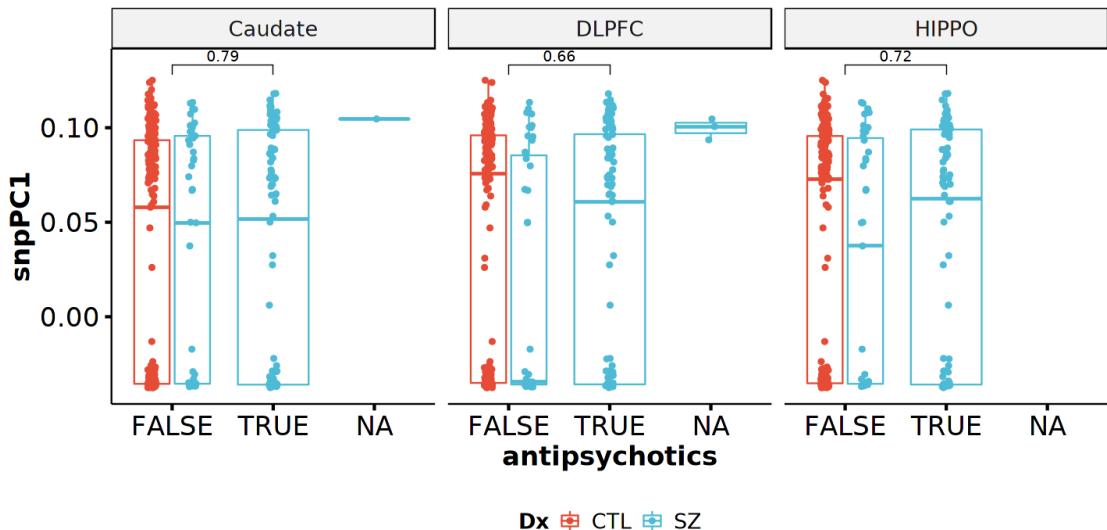


Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

```
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."
```



Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

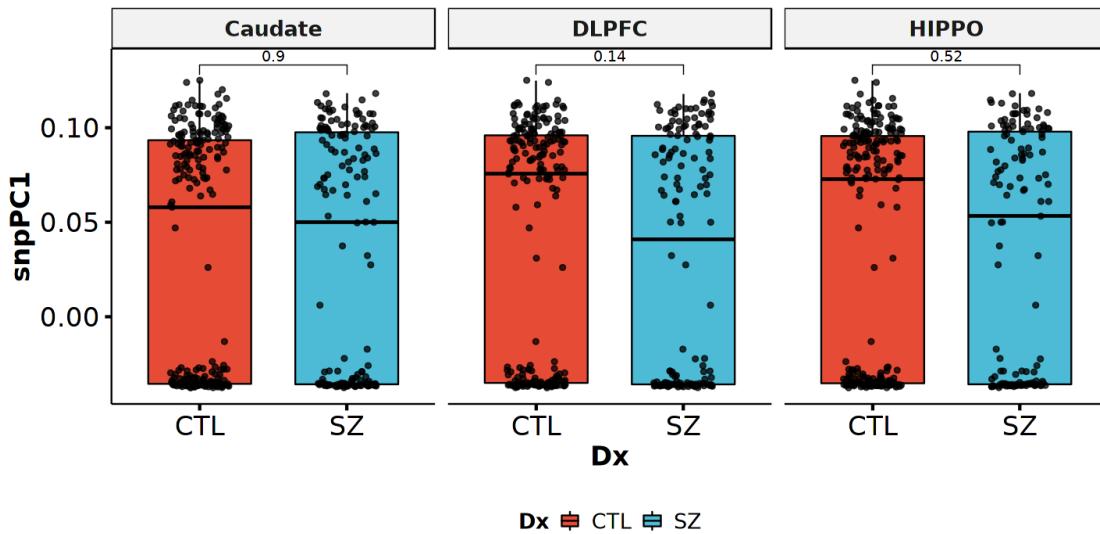
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."



Warning message:

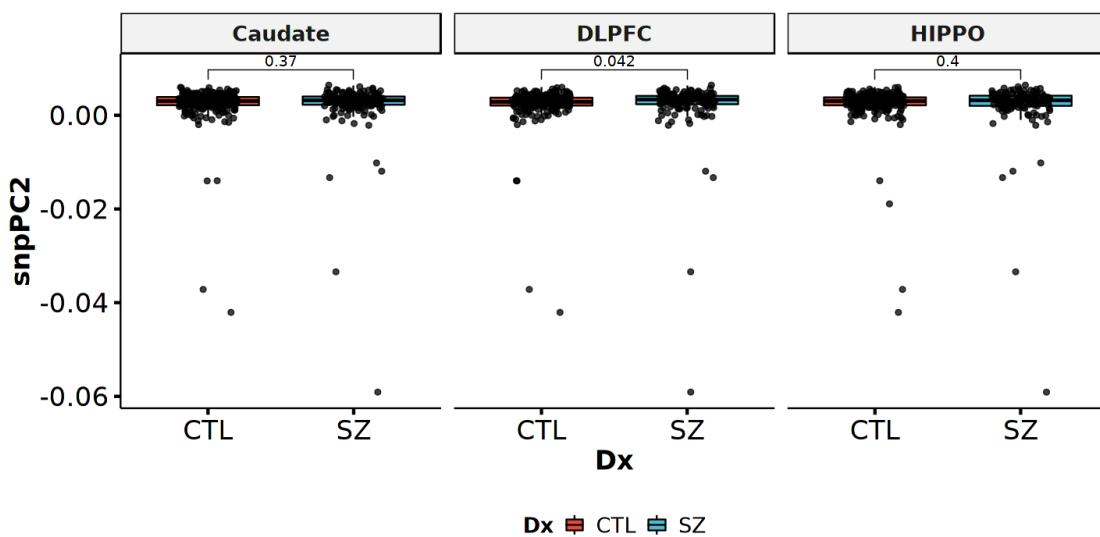
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

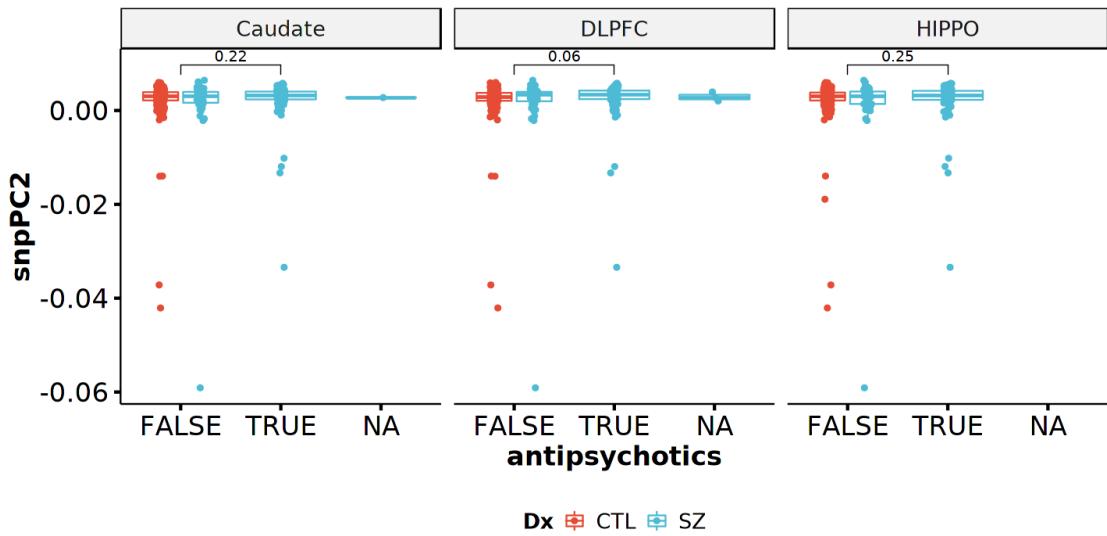


Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

```
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
```



Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

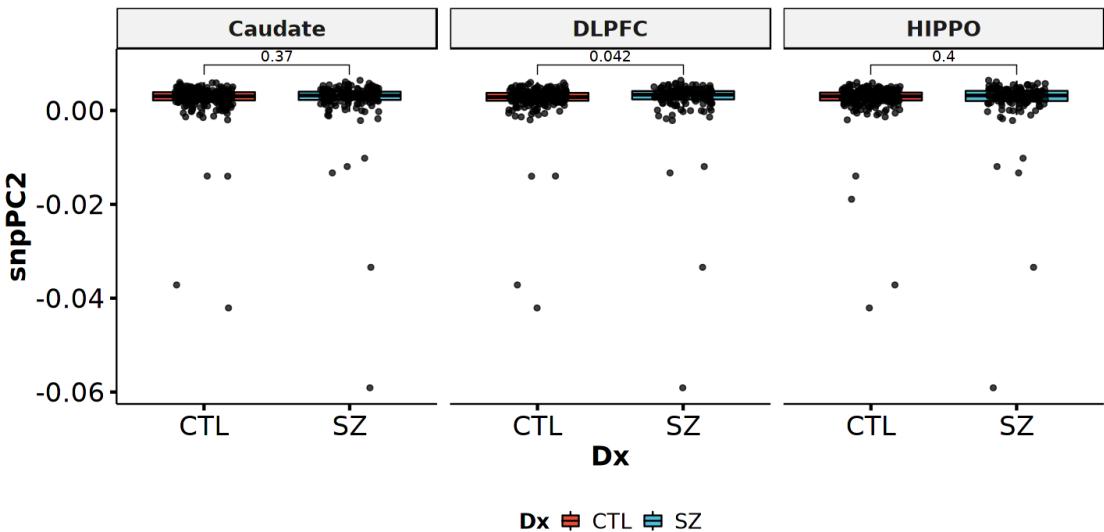
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."



Warning message:

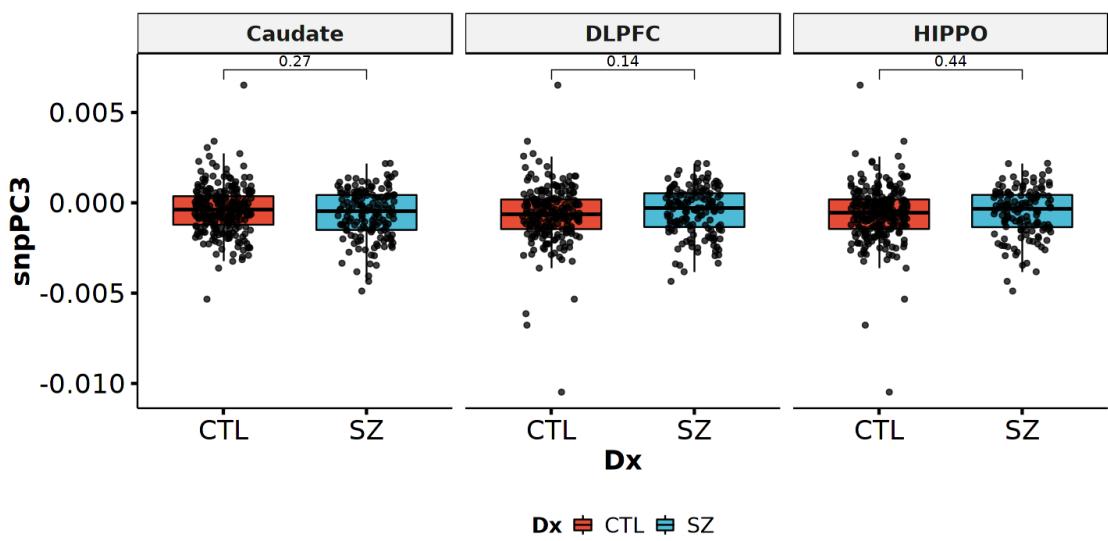
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

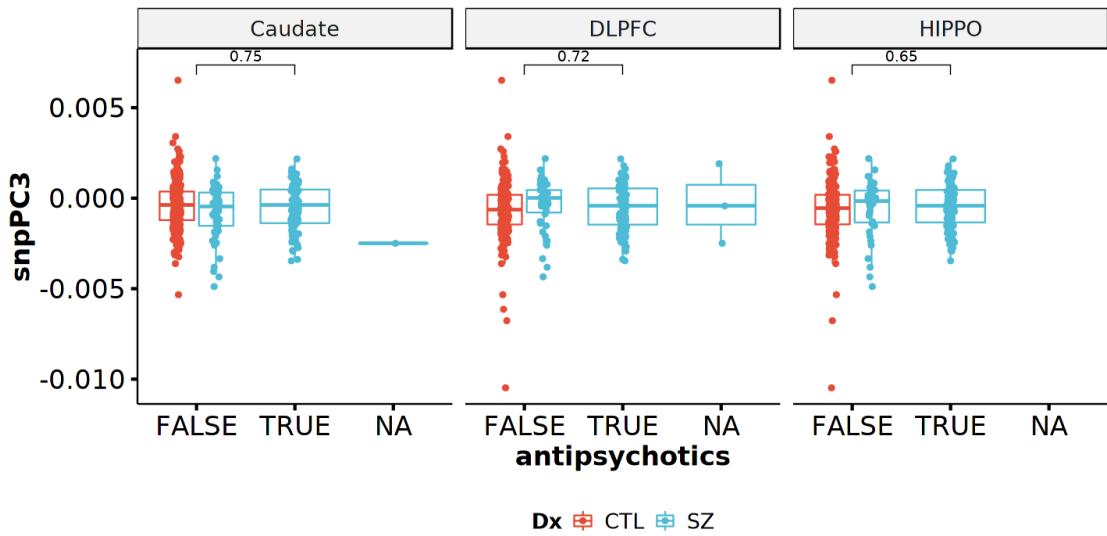


Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

```
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_boxplot)."  
Warning message:  
"Removed 3 rows containing non-finite values (stat_signif)."  
Warning message:  
"Removed 3 rows containing missing values (geom_point)."
```



Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

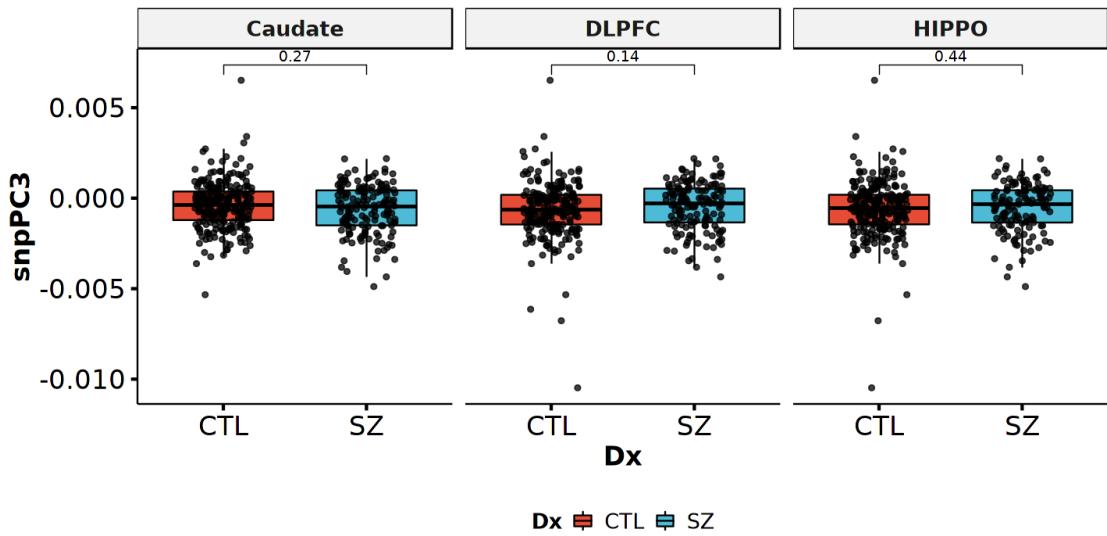
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."



Warning message:

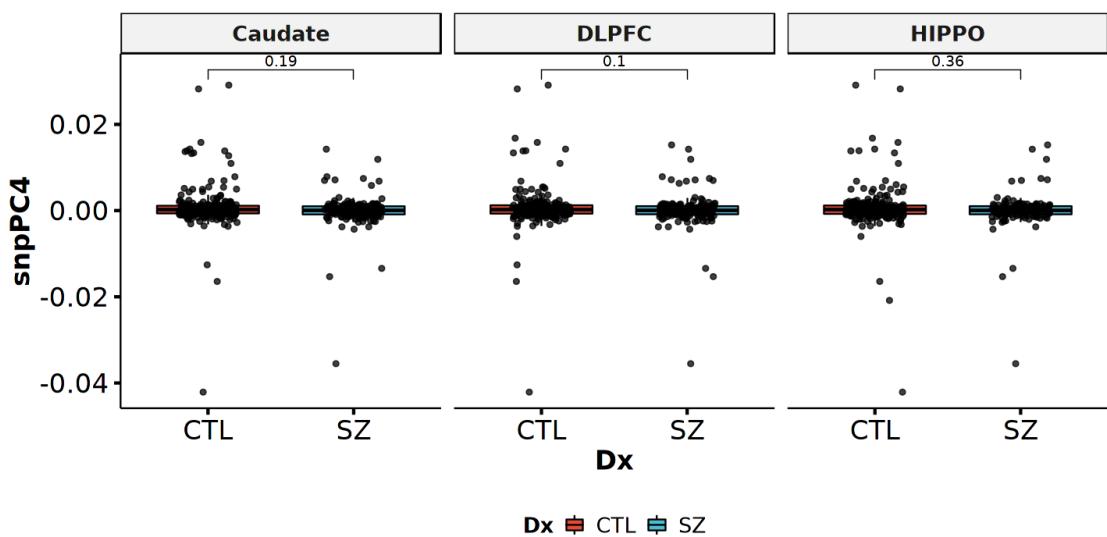
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

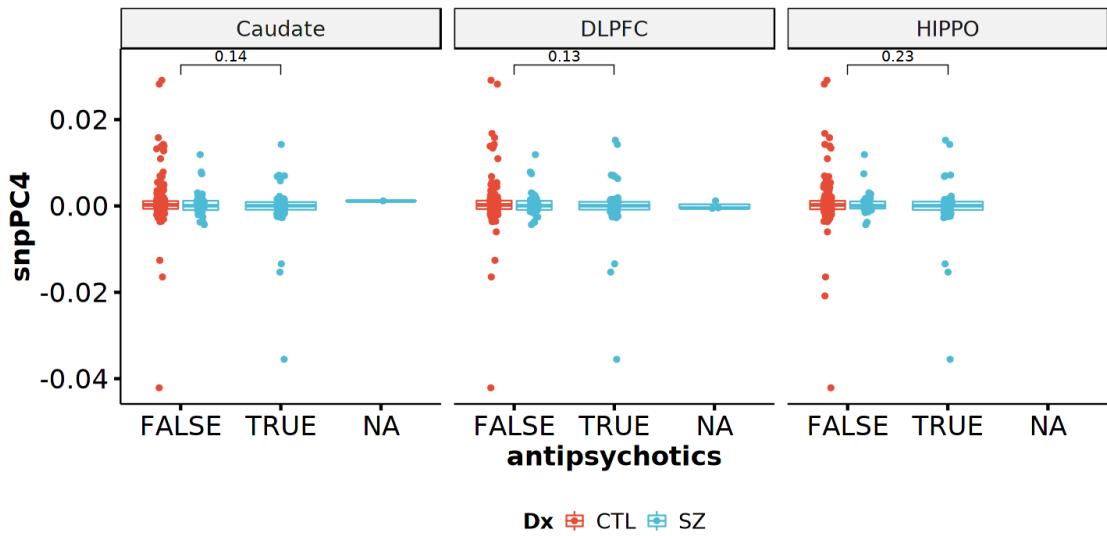


Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

```
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
```



Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

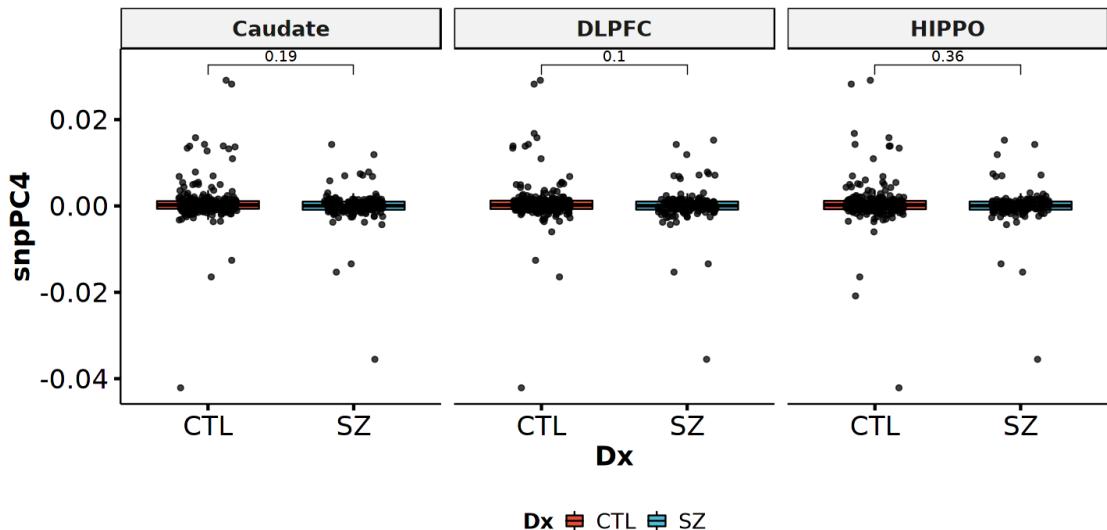
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."



Warning message:

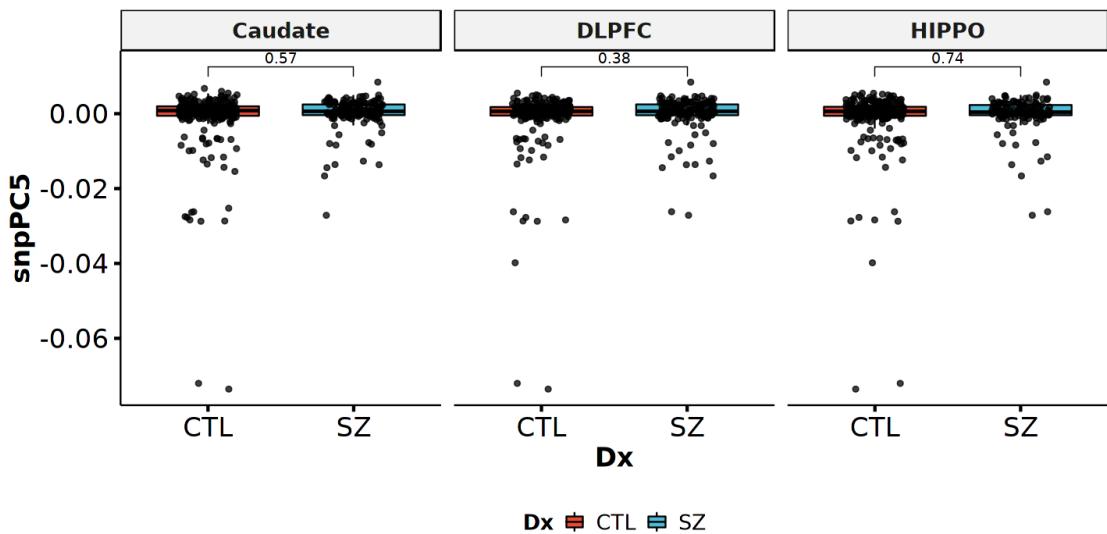
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."



Warning message:

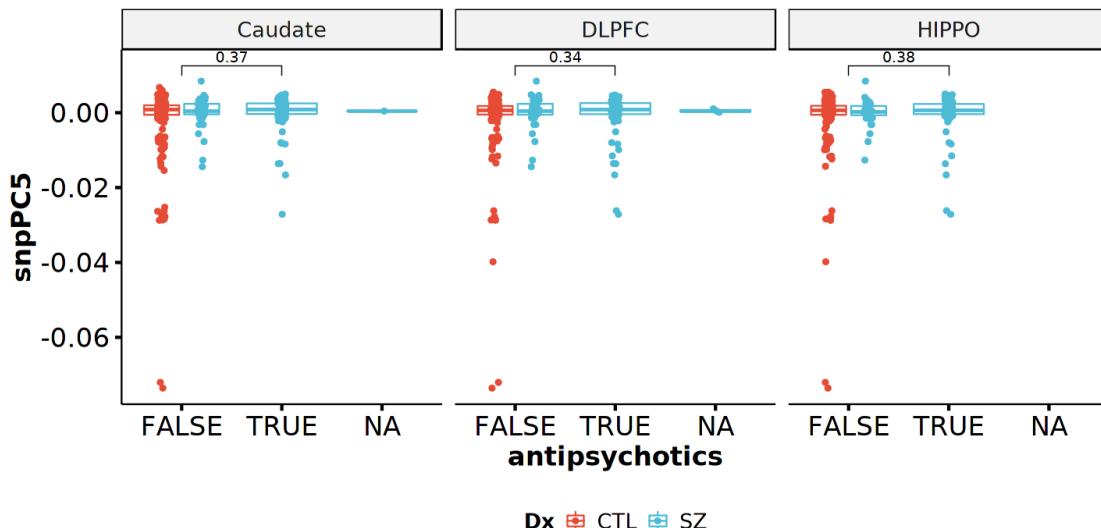
"Removed 3 rows containing non-finite values (stat_boxplot)."

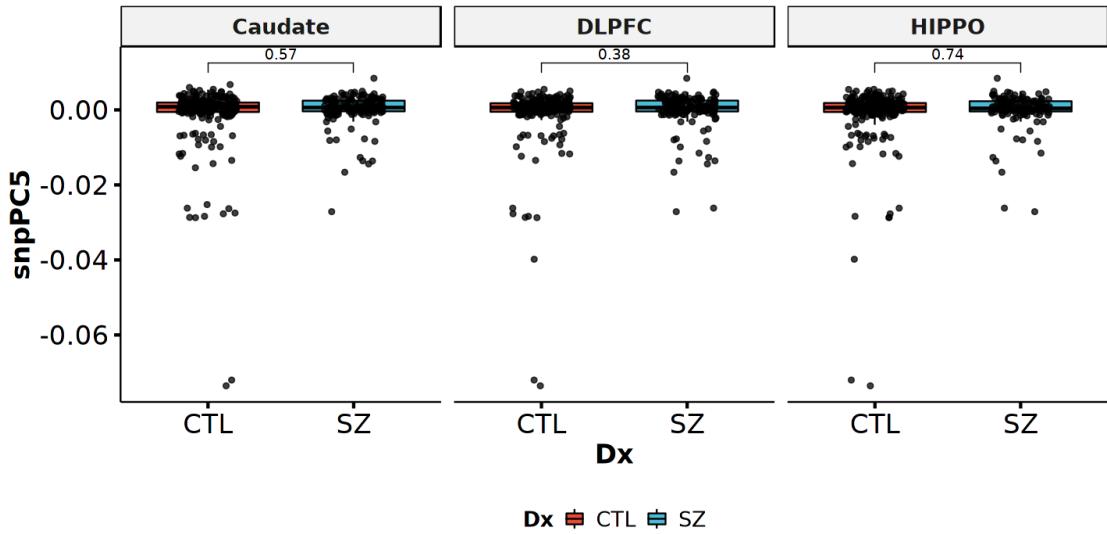
Warning message:

```

"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."
Warning message:
"Removed 3 rows containing non-finite values (stat_boxplot)."
Warning message:
"Removed 3 rows containing non-finite values (stat_signif)."
Warning message:
"Removed 3 rows containing missing values (geom_point)."

```





1.3 Examine gene expression relating to covariates

1.3.1 Load TPM (log2)

```
[4]: caudate = data.table::fread("/ceph/projects/v4_phase3_paper/inputs/counts/
  ↪text_files_counts/tpm/_m/caudate/gene/log2tpm.csv") %>%
  column_to_rownames(var="names") %>% t %>% as.data.frame
dlpfc = data.table::fread("/ceph/projects/v4_phase3_paper/inputs/counts/
  ↪text_files_counts/tpm/_m/dlpfc/gene/log2tpm.csv") %>%
  column_to_rownames(var="names") %>% t %>% as.data.frame
hippo = data.table::fread("/ceph/projects/v4_phase3_paper/inputs/counts/
  ↪text_files_counts/tpm/_m/hippocampus/gene/log2tpm.csv") %>%
  column_to_rownames(var="names") %>% t %>% as.data.frame
```

1.3.2 Dimensional reduction

```
[5]: pca_cc = prcomp(caudate, center = TRUE)$x
pca_dd = prcomp(dlpfc, center = TRUE)$x
pca_hh = prcomp(hippo, center = TRUE)$x
```

1.3.3 Merge tissues

```
[6]: dtc = pca_cc %>% as.data.frame %>% rownames_to_column() %>%
  select(c(rowname, PC1, PC2, PC3, PC4, PC5)) %>%
  pivot_longer(-rowname, names_to="PC", values_to="PC_values") %>%
  mutate(tissue="Caudate")
dtd = pca_dd %>% as.data.frame %>% rownames_to_column() %>%
  select(c(rowname, PC1, PC2, PC3, PC4, PC5)) %>%
```

```

pivot_longer(-rowname, names_to="PC", values_to="PC_values") %>%
  mutate(tissue="DLPFC")
dth = pca_hh %>% as.data.frame %>% rownames_to_column() %>%
  select(c(rowname, PC1, PC2, PC3, PC4, PC5)) %>%
  pivot_longer(-rowname, names_to="PC", values_to="PC_values") %>%
  mutate(tissue="Hippocampus")
dt = bind_rows(dtc, dtd, dth)

```

1.3.4 Plot scatters and correlations with covariates

[7]: colors = get_palette(palette = "npg", 3)
colors

1. '#E64B35FF' 2. '#4DBBD5FF' 3. '#00A087FF'

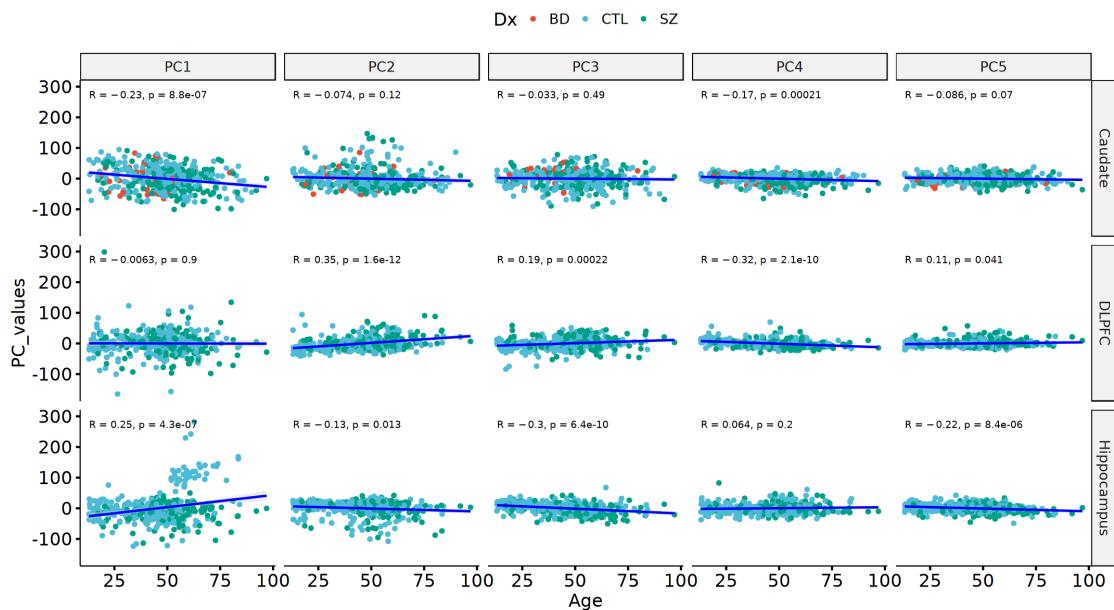
[8]: options(repr.plot.width=18, repr.plot.height=10)
dir.create("pca_scatter")
for(covar in covarsCont){
 flush.console()
 sca1 = dt %>% inner_join(memPHENO(), by=c("rowname"="RNum")) %>%
 ggscatter(y="PC_values", x=covar, color="Dx", palette="npg",
 facet.by=c('tissue', 'PC'), ncol=5,
 add='reg.line', conf.int=TRUE, cor.coef=TRUE,
 add.params=list(color="blue", fill="lightgray"),
 ggtheme=theme_pubr(base_size=20))
 save_img(sca1, paste0("pca_scatter/scatter_log2tpm_dx_5pcs_",covar), w=18, □
 ↳h=10)
 print(sca1)
 sca2 = dt %>% inner_join(memPHENO(), by=c("rowname"="RNum")) %>%
 filter(antipsychotics == "TRUE", Dx == "SZ") %>%
 ggscatter(y="PC_values", x=covar, color=colors[2], facet.by=c('tissue', □
 ↳'PC'),
 ncol=5, add='reg.line', conf.int=TRUE, cor.coef=TRUE,
 palette="npg", add.params=list(color="blue", □
 ↳fill="lightgray"),
 ggtheme=theme_pubr(base_size=20))
 save_img(sca2, paste0("pca_scatter/scatter_log2tpm_AP_5pcs_",covar), w=18, □
 ↳h=10)
 print(sca2)
 sca3 = dt %>% inner_join(memPHENO(), by=c("rowname"="RNum")) %>%
 filter(antipsychotics == "FALSE", Dx == "SZ") %>%
 ggscatter(y="PC_values", x=covar, color=colors[1], facet.by=c('tissue', □
 ↳'PC'),
 ncol=5, add='reg.line', conf.int=TRUE, cor.coef=TRUE,
 palette="npg", add.params=list(color="blue", □
 ↳fill="lightgray"),
 ggtheme=theme_pubr(base_size=20))
}

```

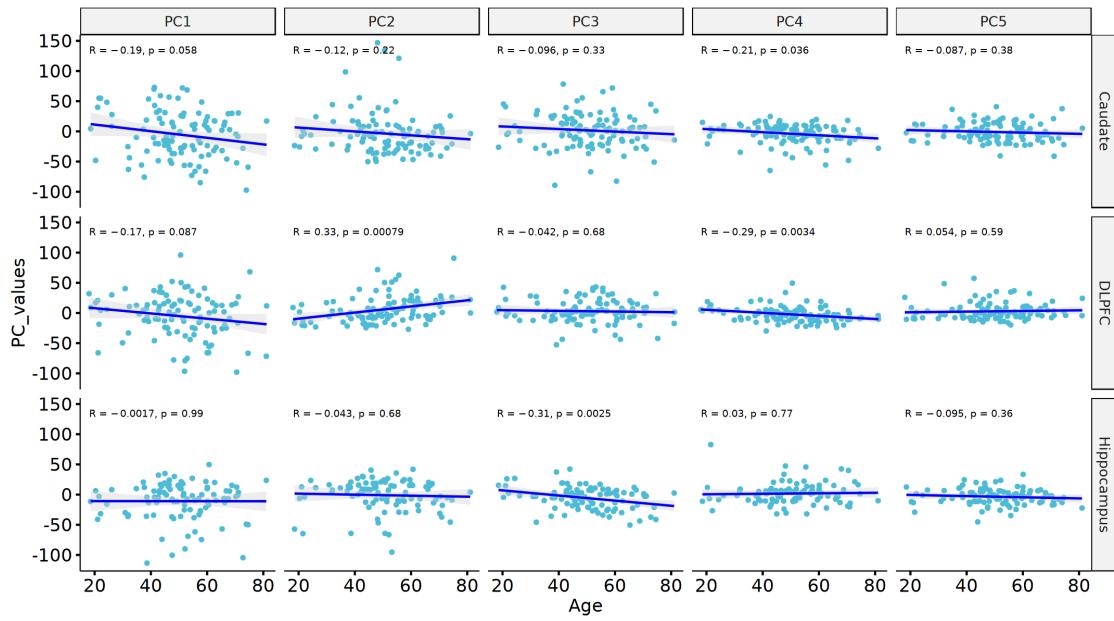
    save_img(sca3, paste0("pca_scatter/scatter_log2tpm_noAP_5pcs_",covar),u
    ↵w=18, h=10)
    print(sca3)
}

`geom_smooth()` using formula 'y ~ x'

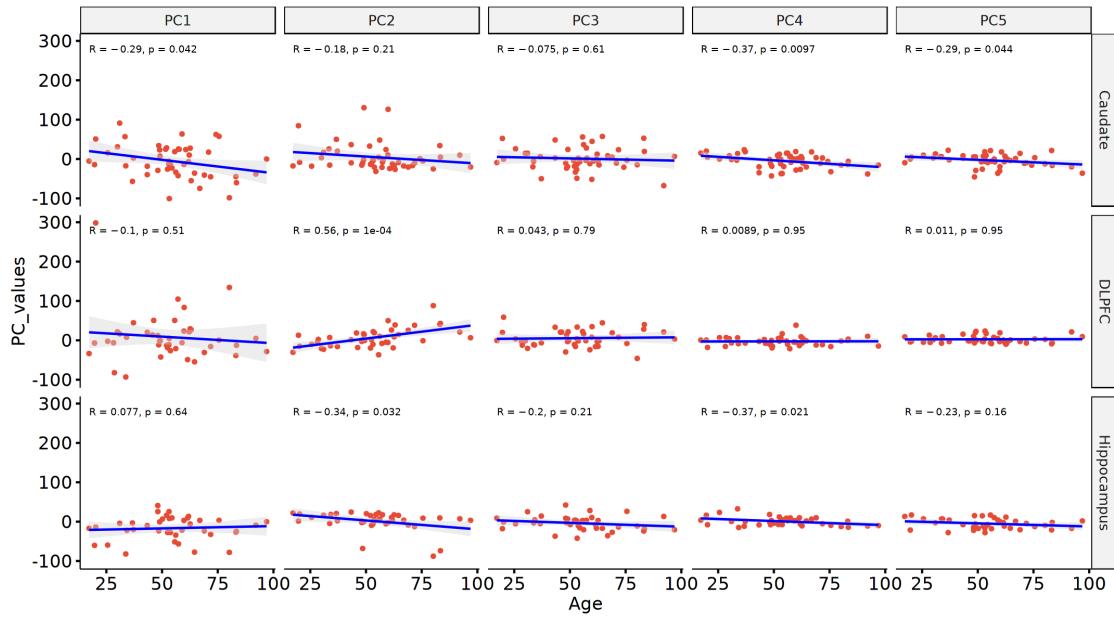
```



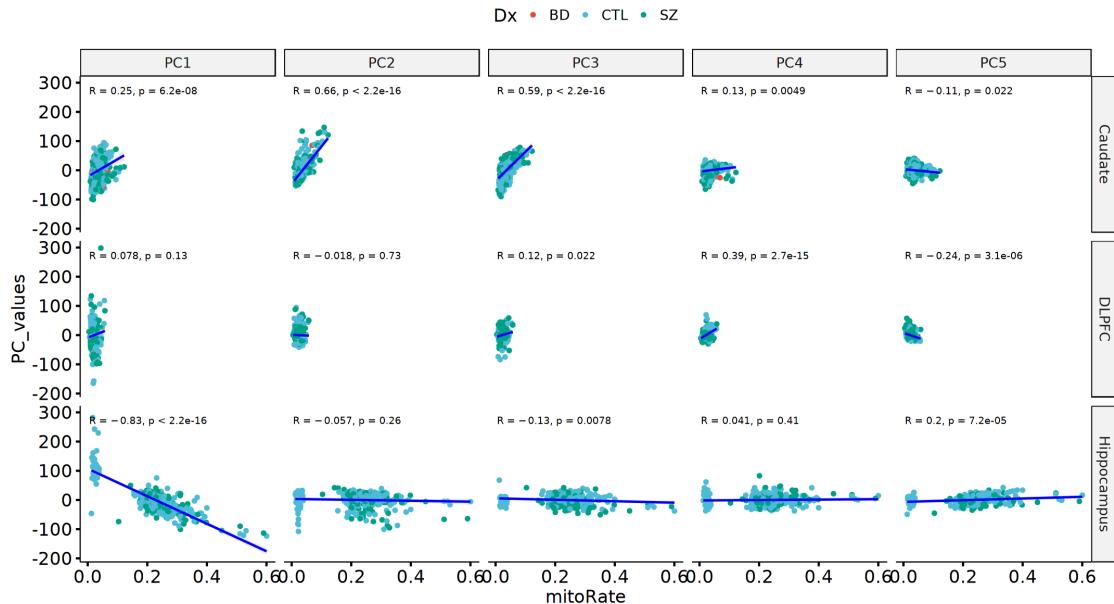
```
`geom_smooth()` using formula 'y ~ x'
```



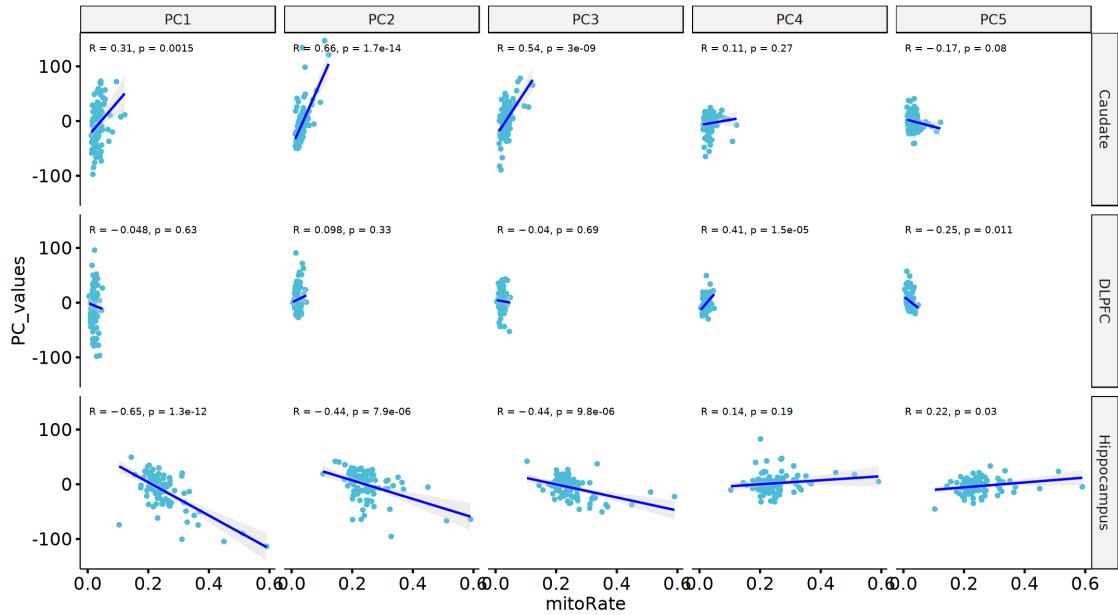
```
`geom_smooth()` using formula 'y ~ x'
```



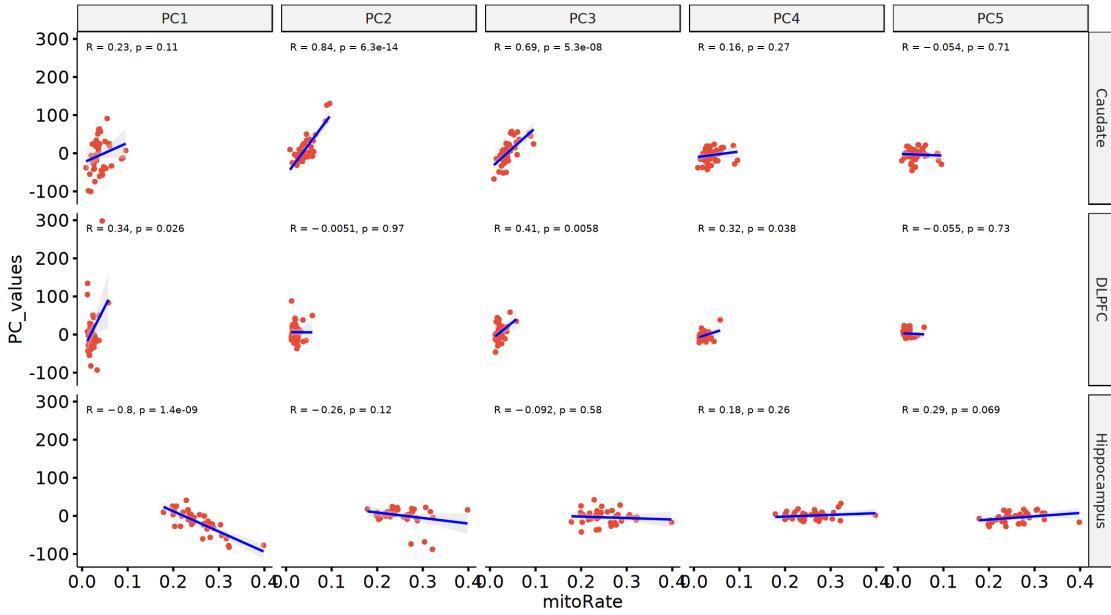
```
`geom_smooth()` using formula 'y ~ x'
```



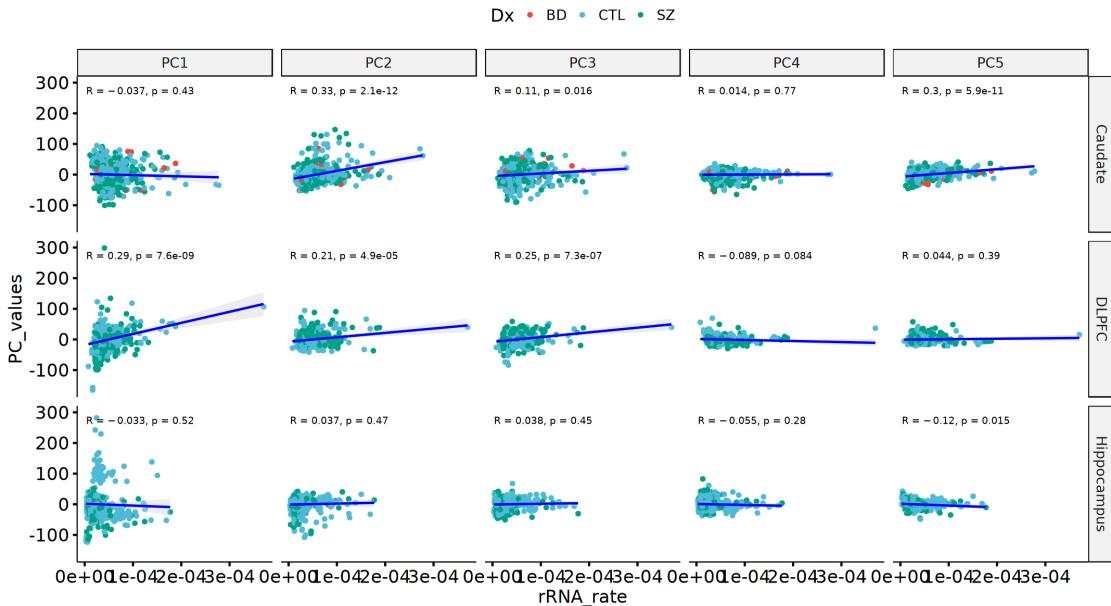
```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```



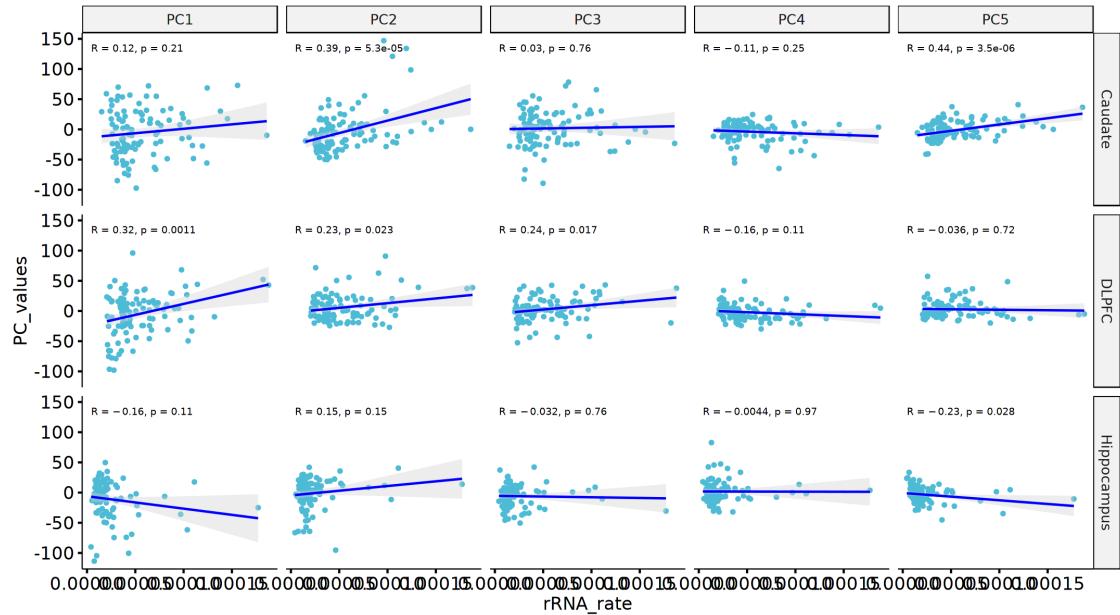
```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```



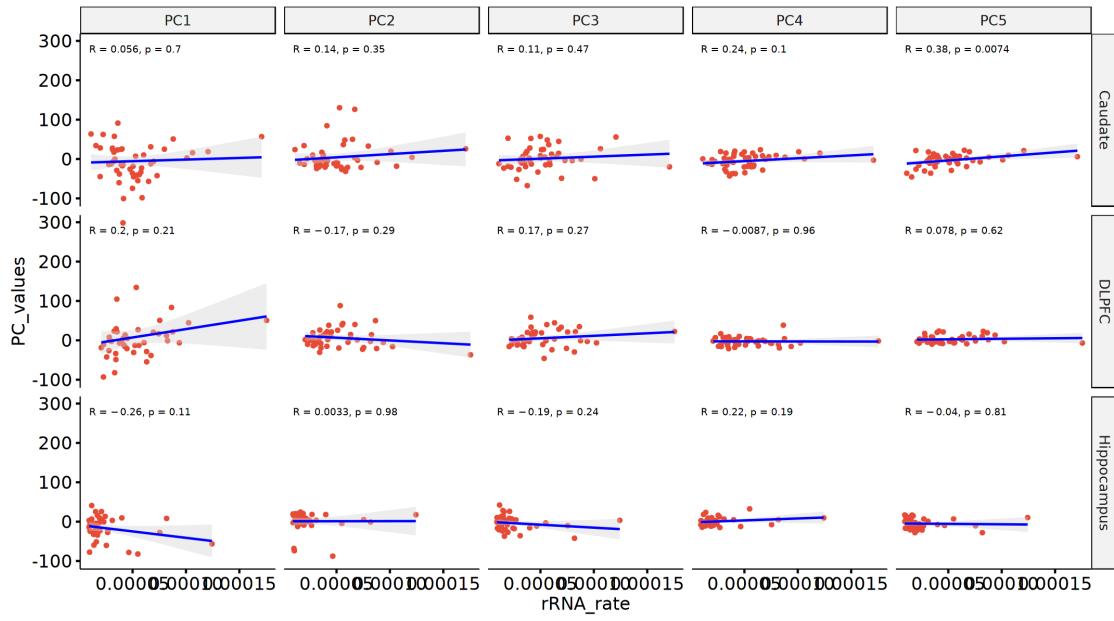
```
`geom_smooth()` using formula 'y ~ x'
```



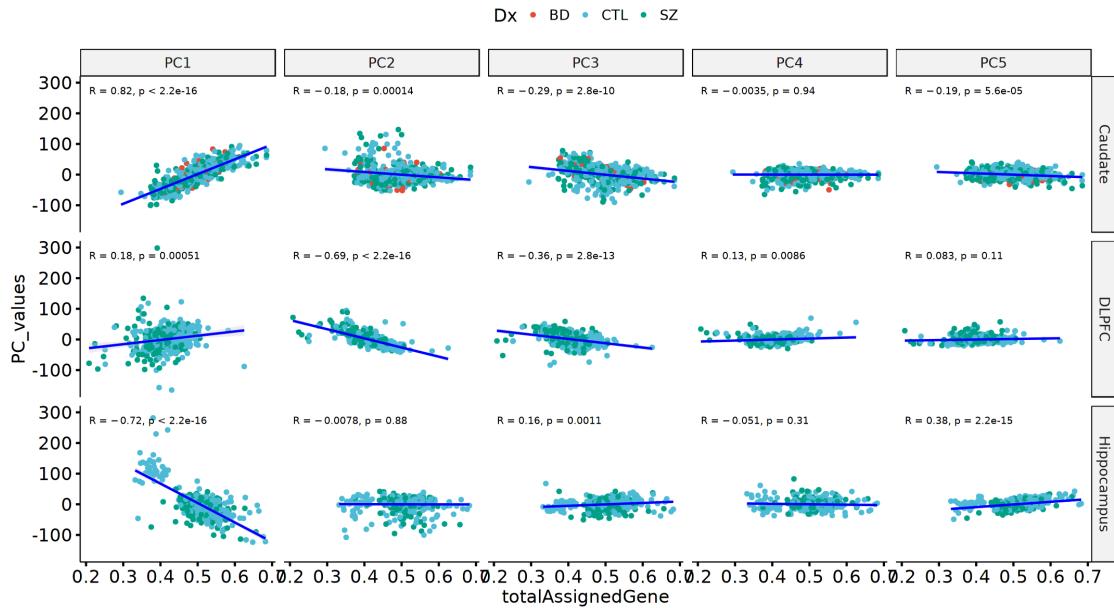
```
`geom_smooth()` using formula 'y ~ x'
```



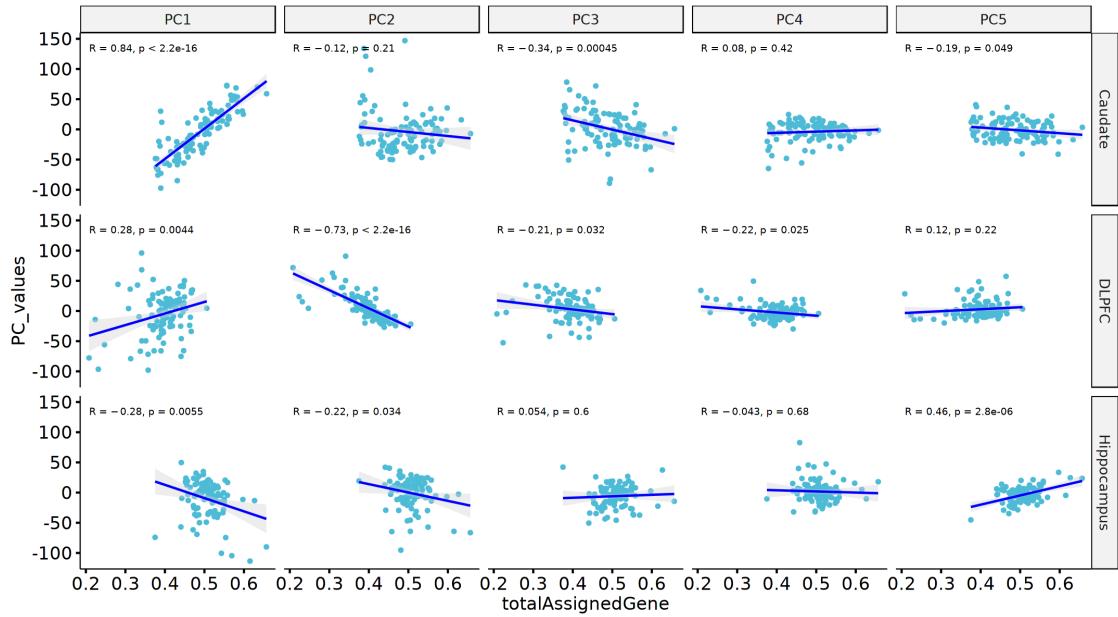
```
`geom_smooth()` using formula 'y ~ x'
```



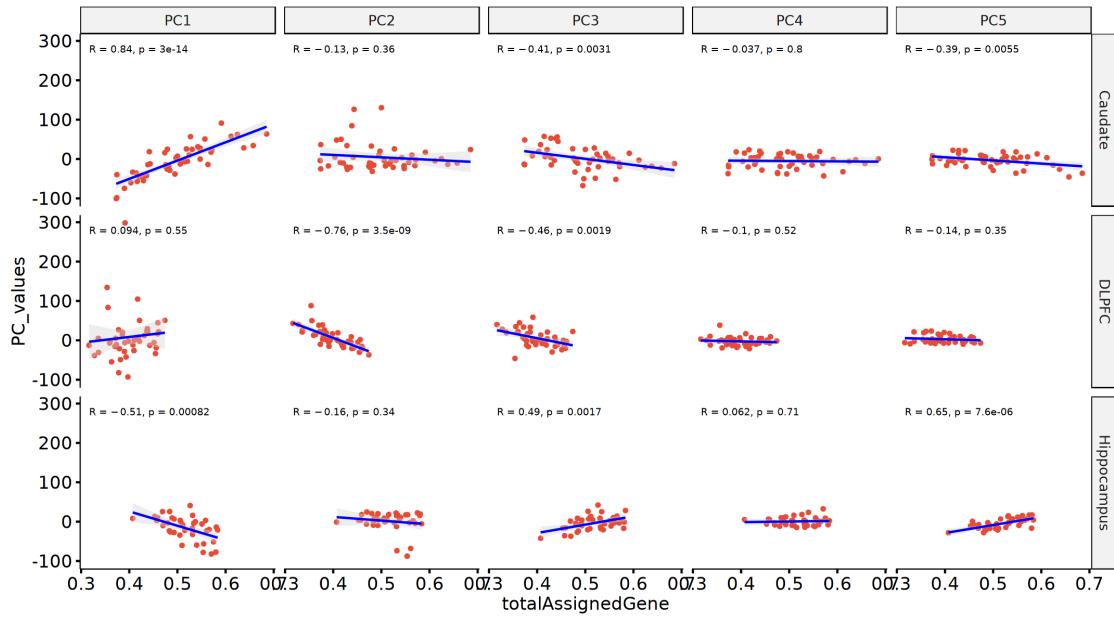
```
`geom_smooth()` using formula 'y ~ x'
```



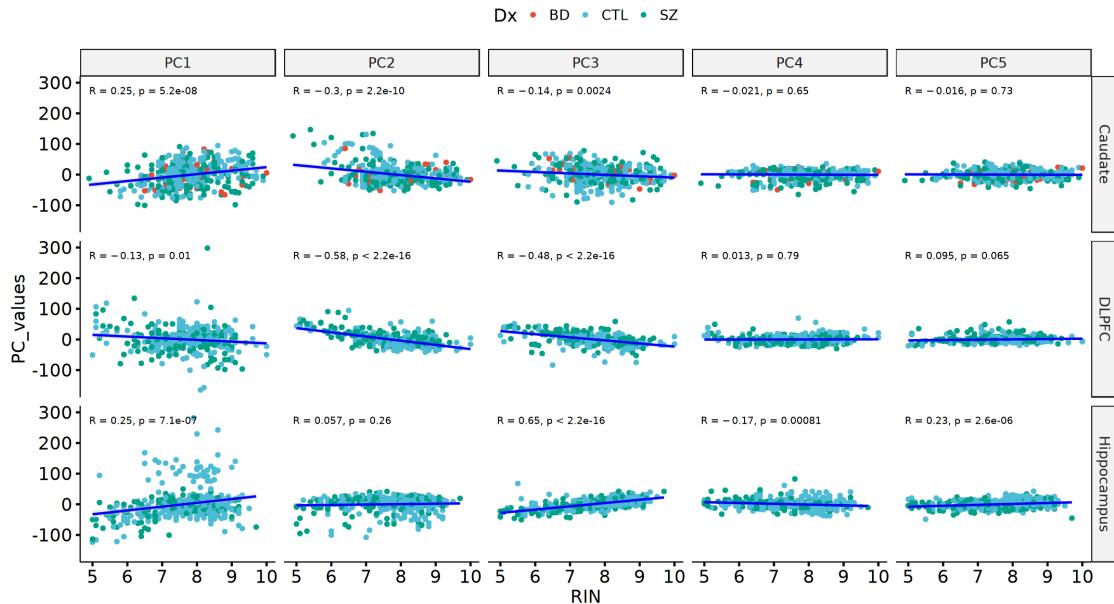
```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```



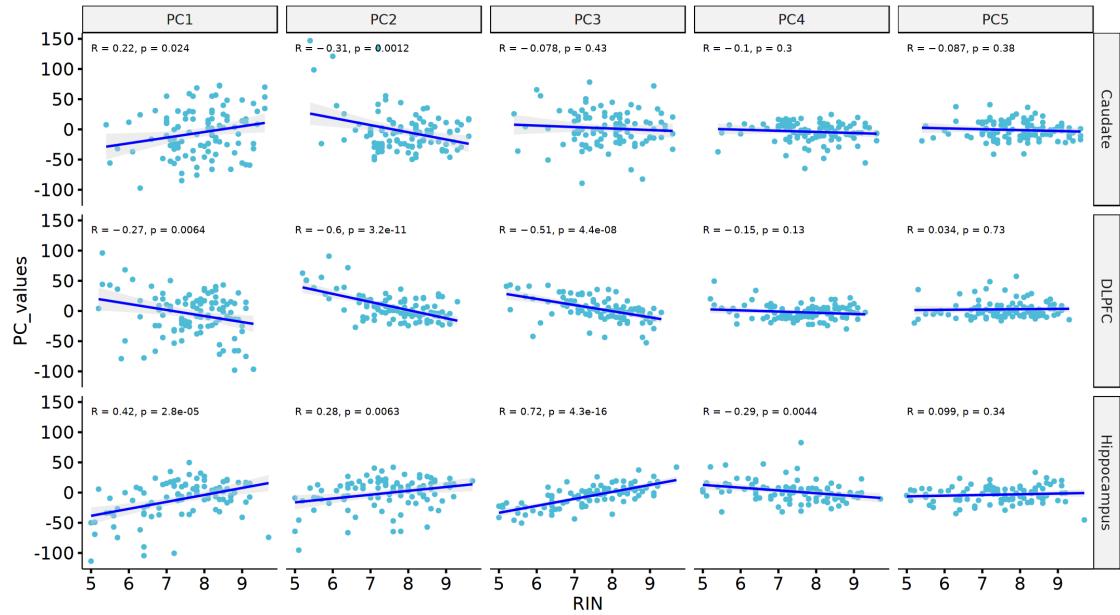
```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```



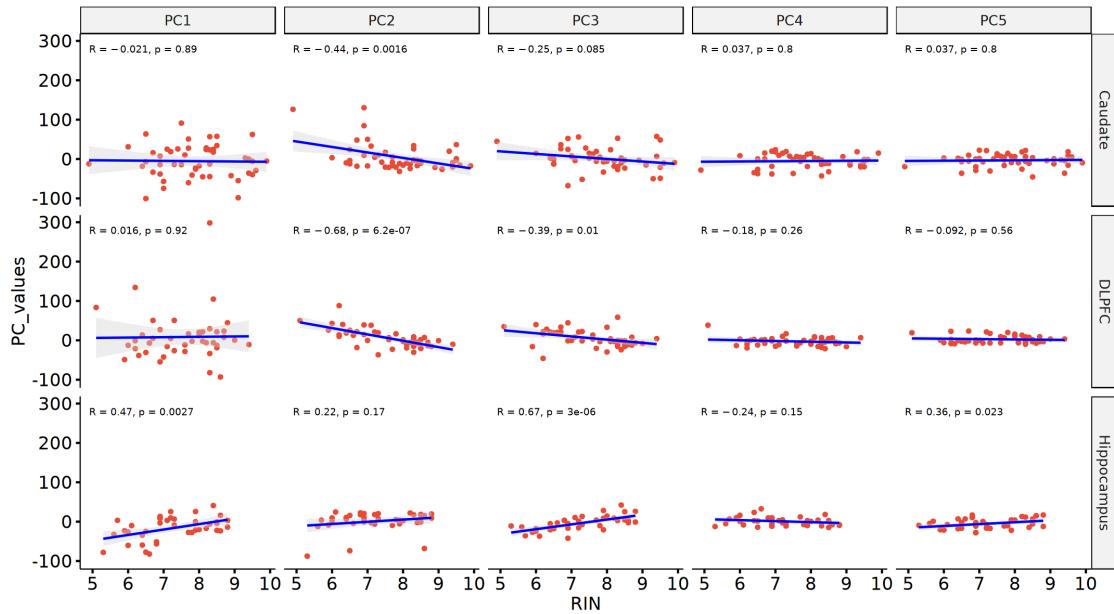
```
`geom_smooth()` using formula 'y ~ x'
```



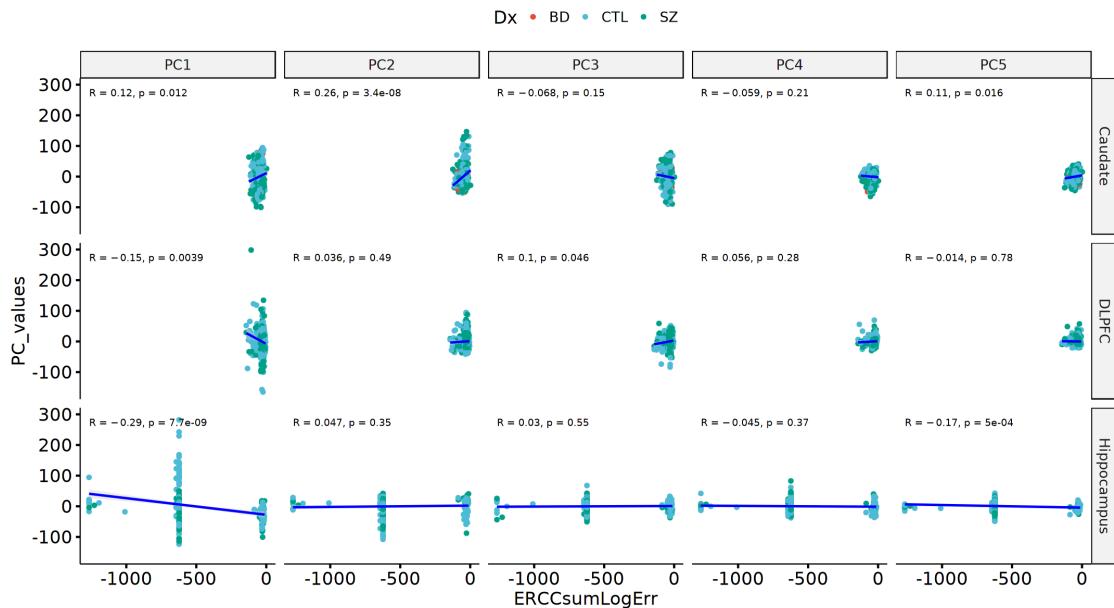
```
`geom_smooth()` using formula 'y ~ x'
```



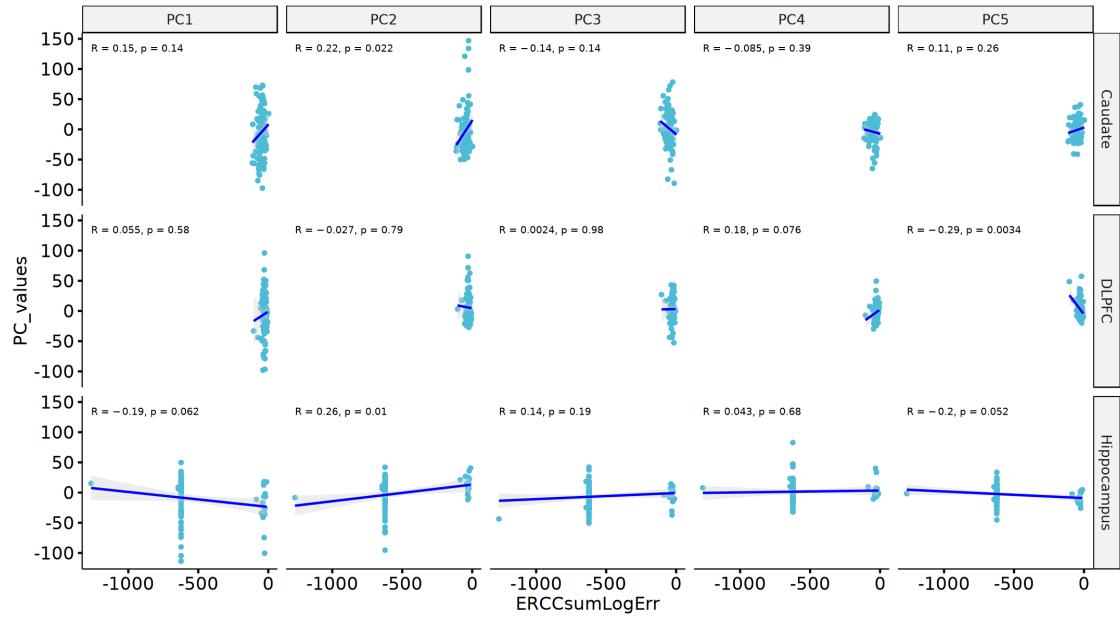
```
`geom_smooth()` using formula 'y ~ x'
```



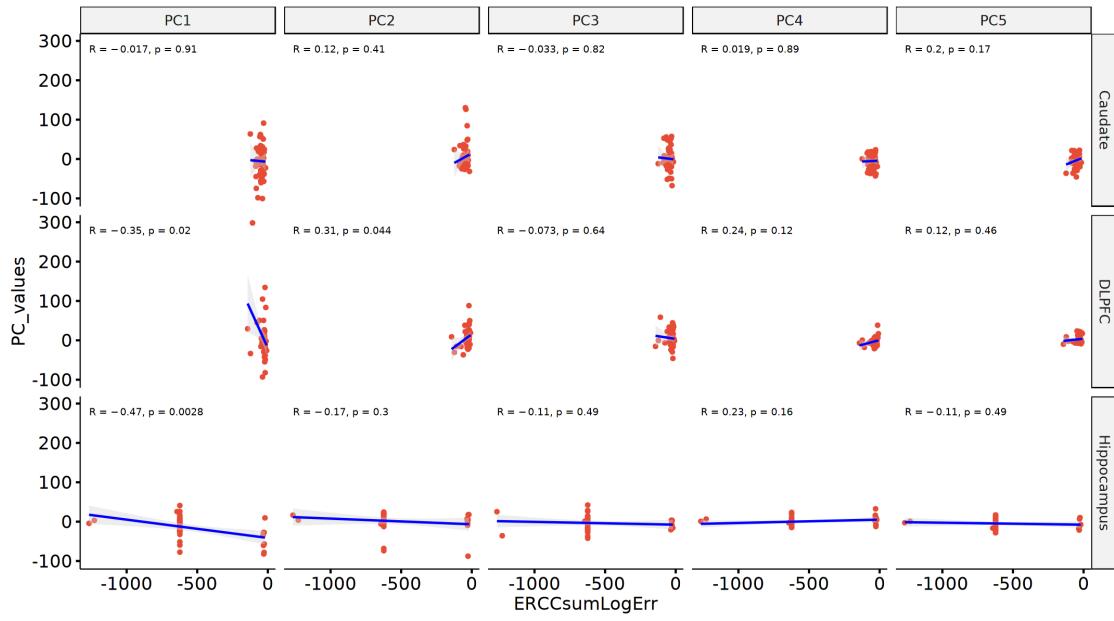
`geom_smooth()` using formula 'y ~ x'



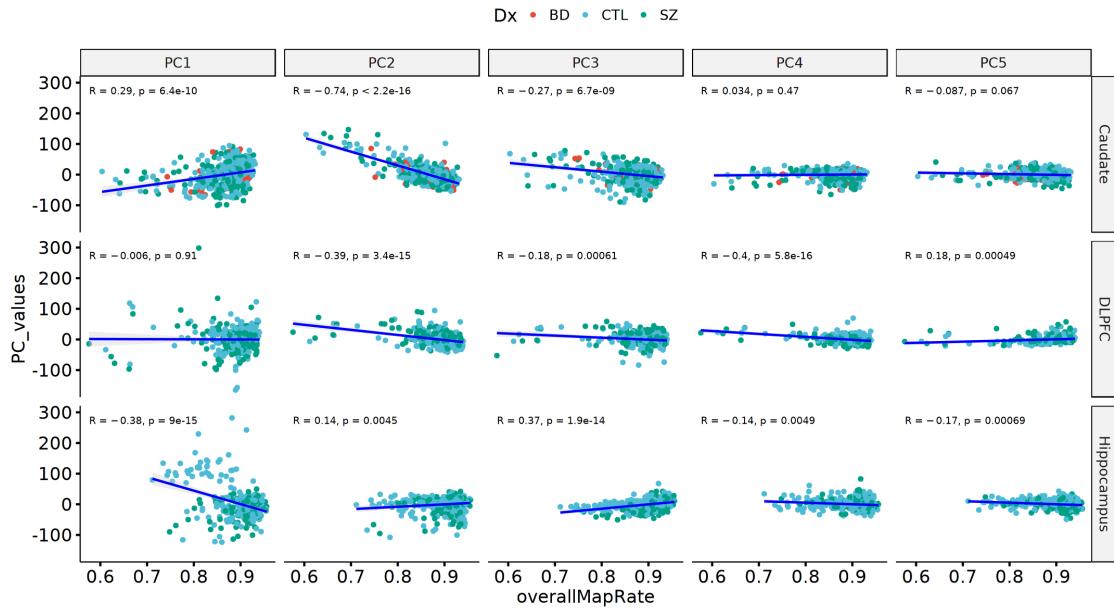
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



```

`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'

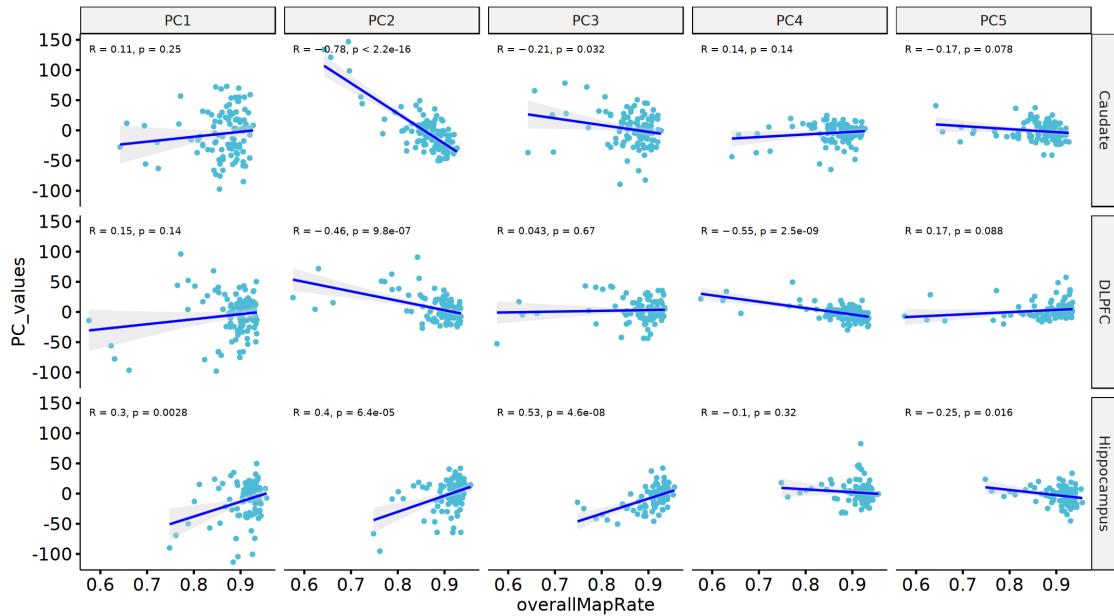
Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."

```



```
`geom_smooth()` using formula 'y ~ x'
```

Warning message:

```
"Removed 15 rows containing non-finite values (stat_smooth)."
```

Warning message:

```
"Removed 15 rows containing non-finite values (stat_cor)."
```

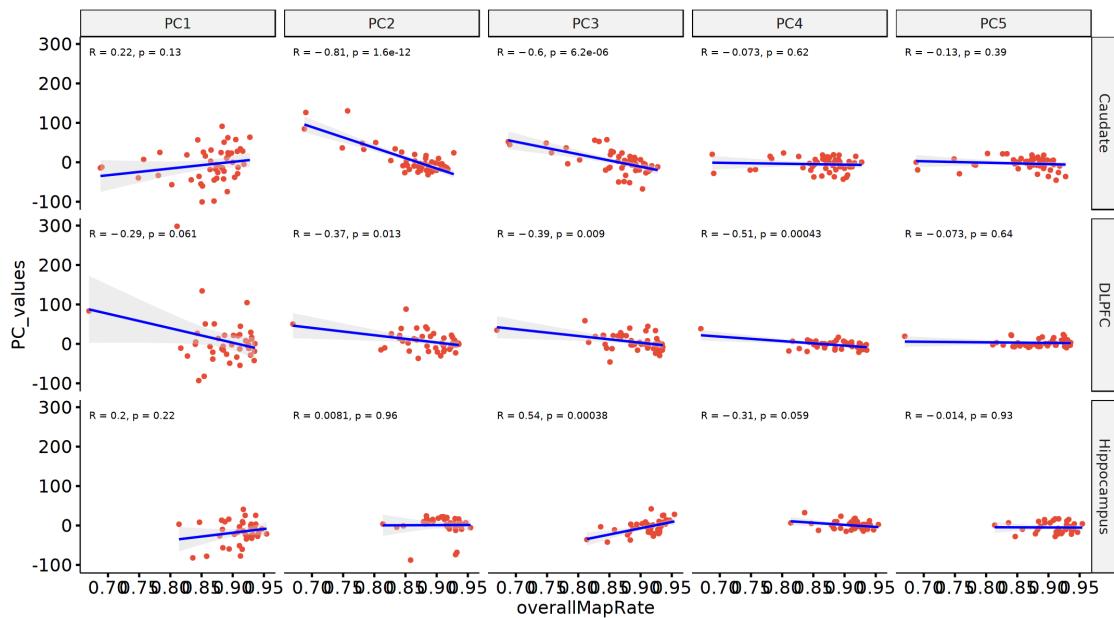
Warning message:

```
"Removed 15 rows containing missing values (geom_point)."
```

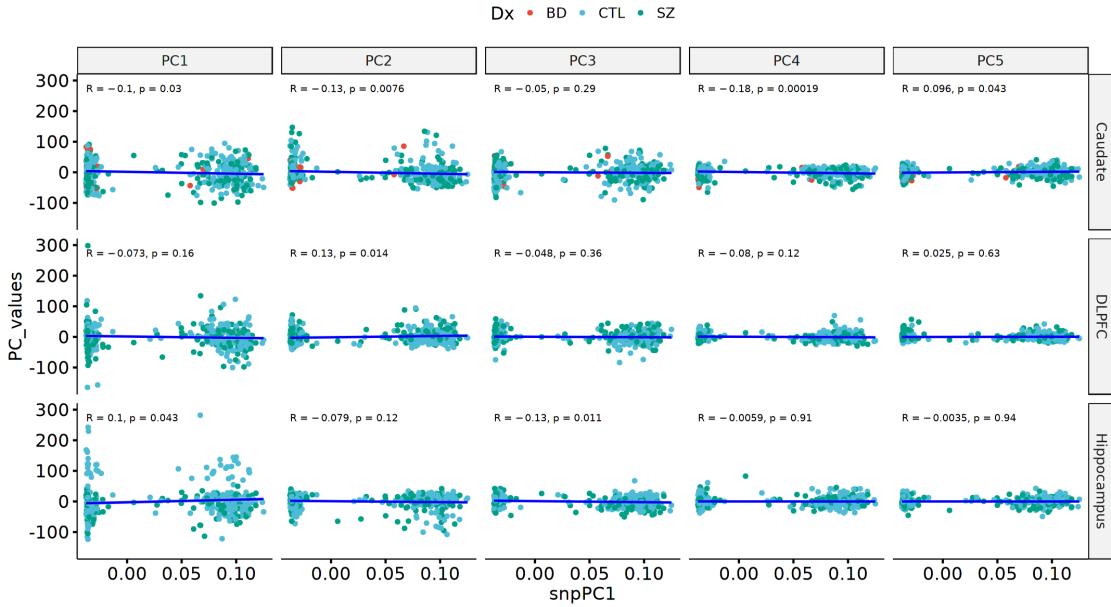
```
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



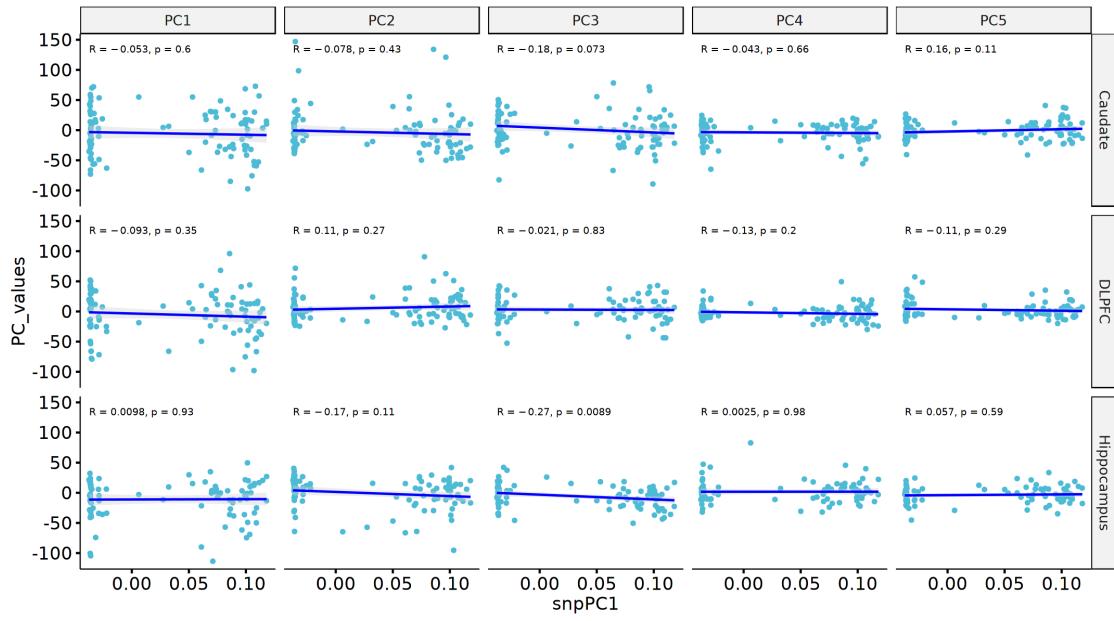
```
`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
```



```
`geom_smooth()` using formula 'y ~ x'
```

Warning message:

```
"Removed 15 rows containing non-finite values (stat_smooth)."
```

Warning message:

```
"Removed 15 rows containing non-finite values (stat_cor)."
```

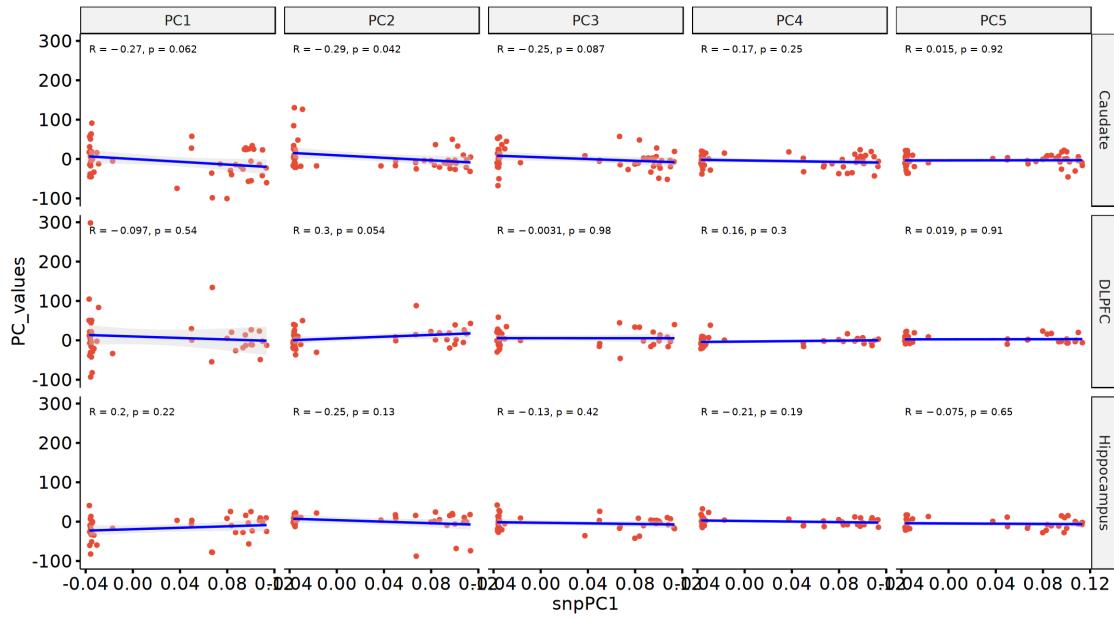
Warning message:

```
"Removed 15 rows containing missing values (geom_point)."
```

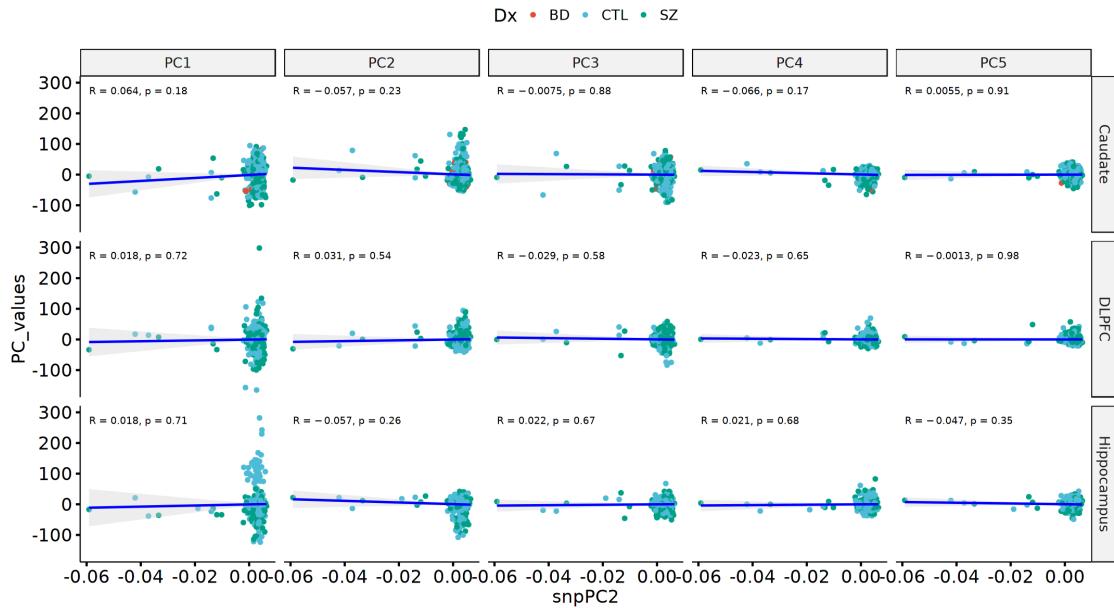
```
`geom_smooth()` using formula 'y ~ x'
```

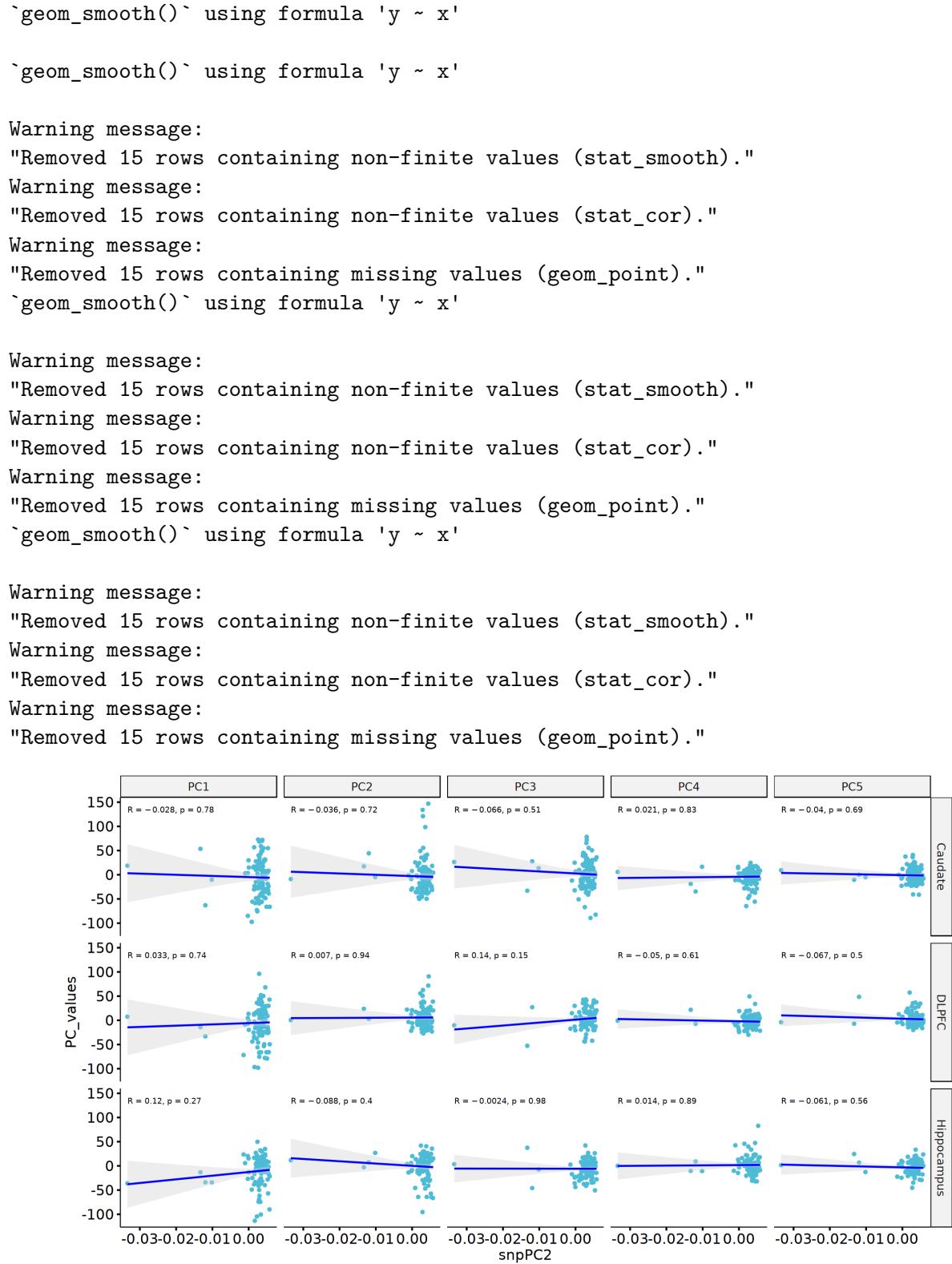
```
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```





```
`geom_smooth()` using formula 'y ~ x'
```

Warning message:

```
"Removed 15 rows containing non-finite values (stat_smooth)."
```

Warning message:

```
"Removed 15 rows containing non-finite values (stat_cor)."
```

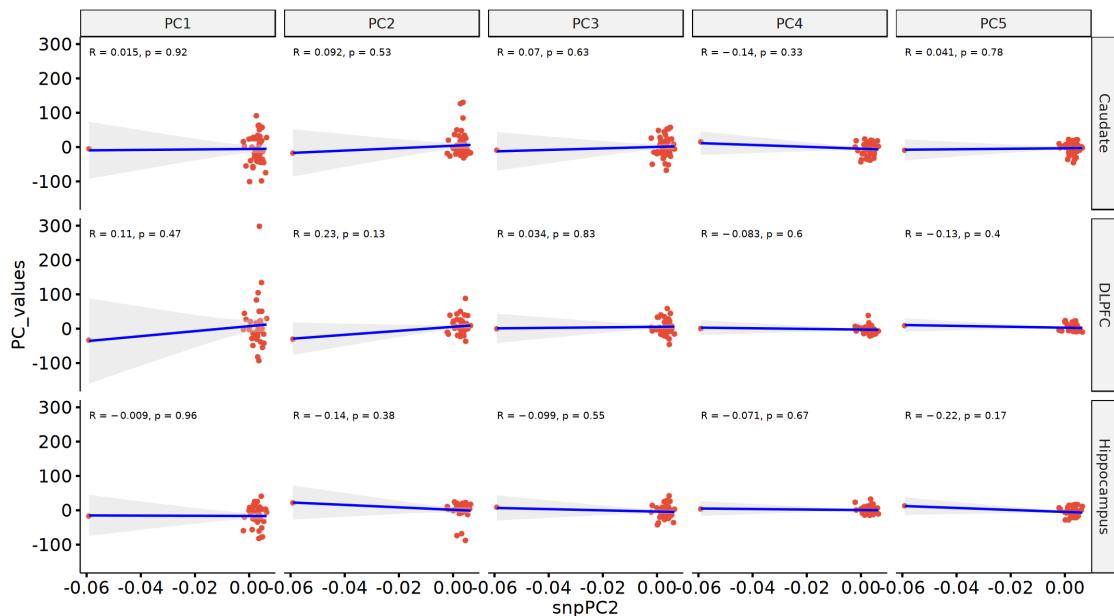
Warning message:

```
"Removed 15 rows containing missing values (geom_point)."
```

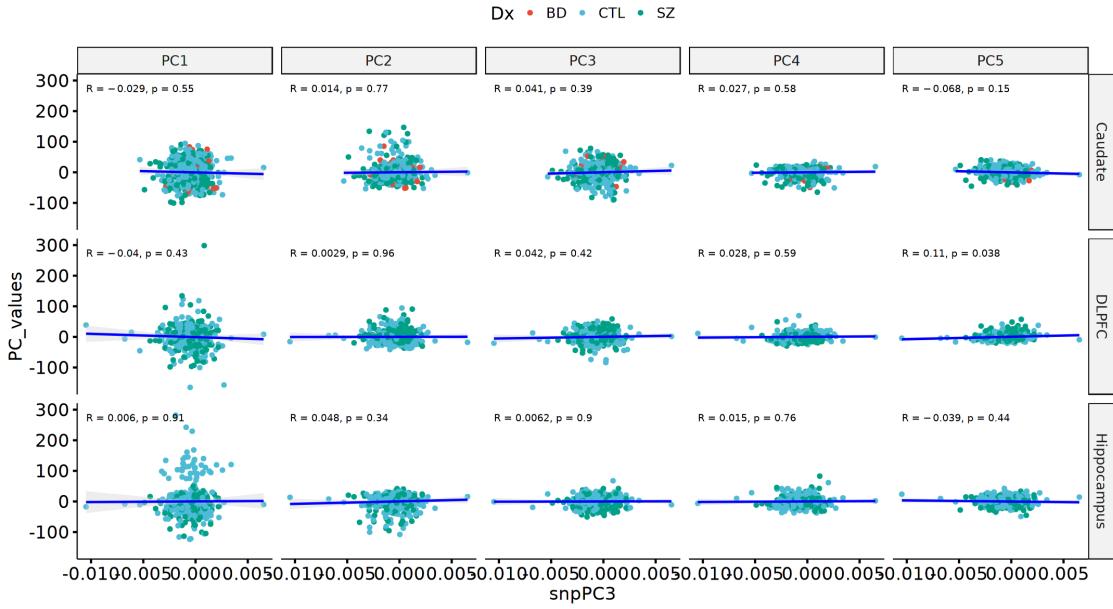
```
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



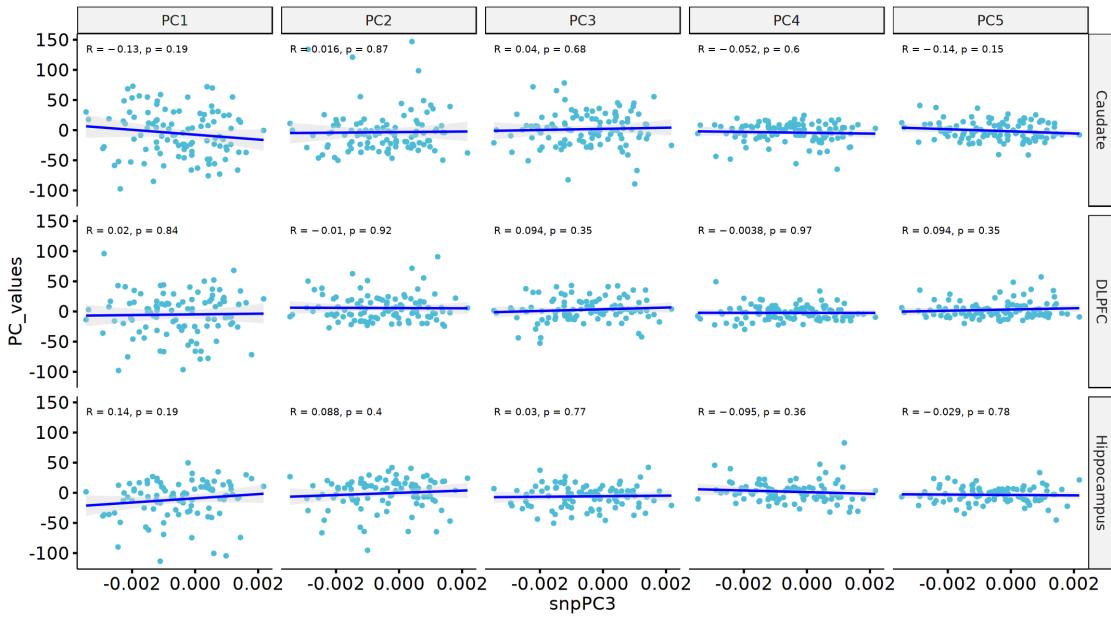
```
`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
```



```
`geom_smooth()` using formula 'y ~ x'
```

Warning message:

```
"Removed 15 rows containing non-finite values (stat_smooth)."
```

Warning message:

```
"Removed 15 rows containing non-finite values (stat_cor)."
```

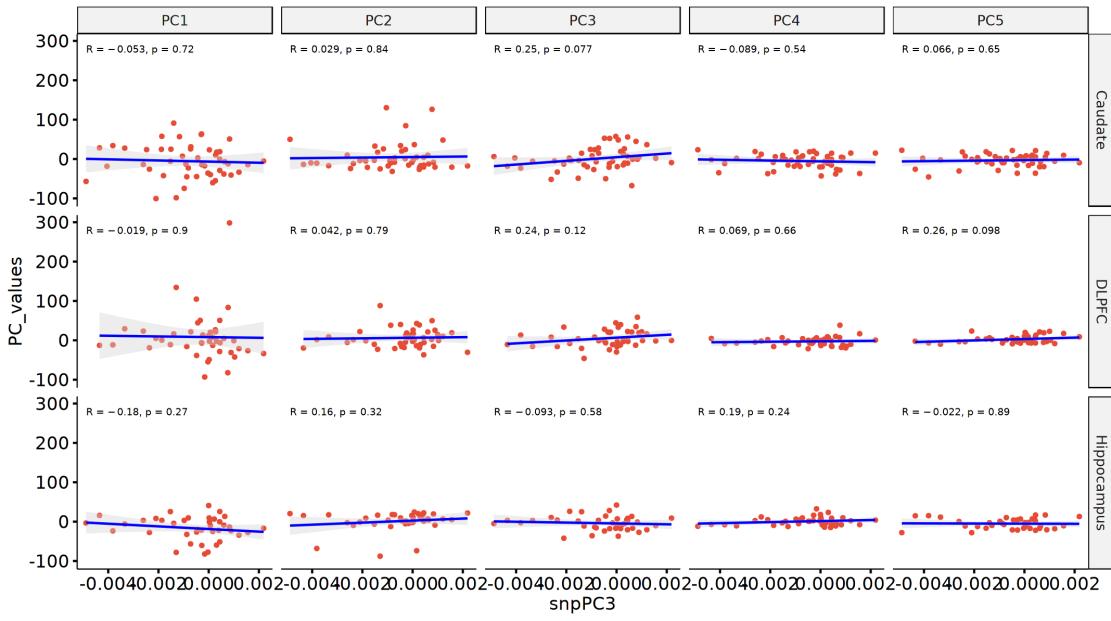
Warning message:

```
"Removed 15 rows containing missing values (geom_point)."
```

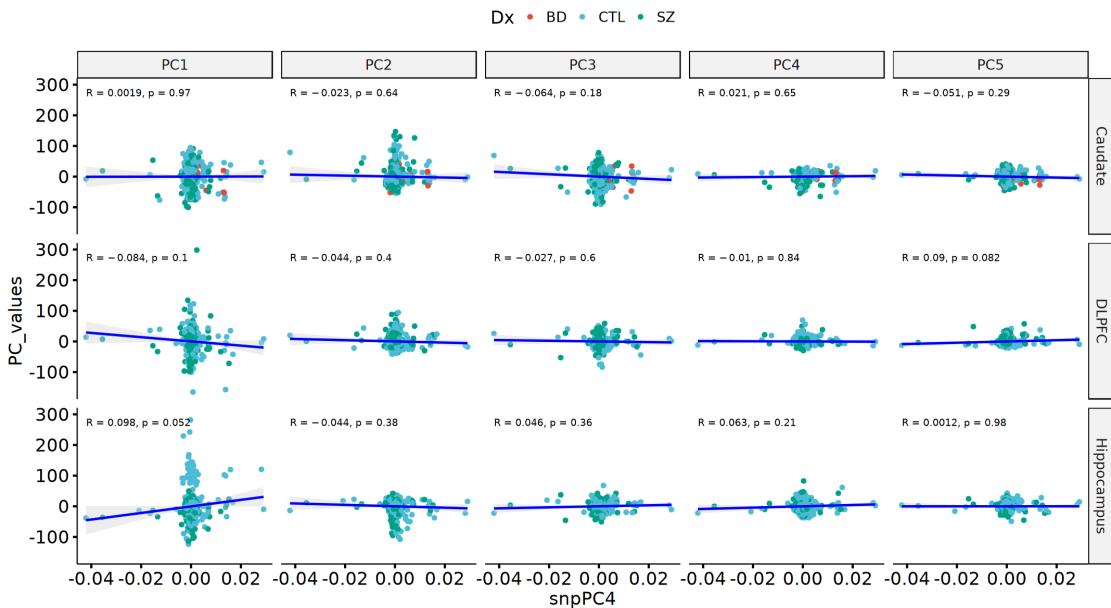
```
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



```

`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'

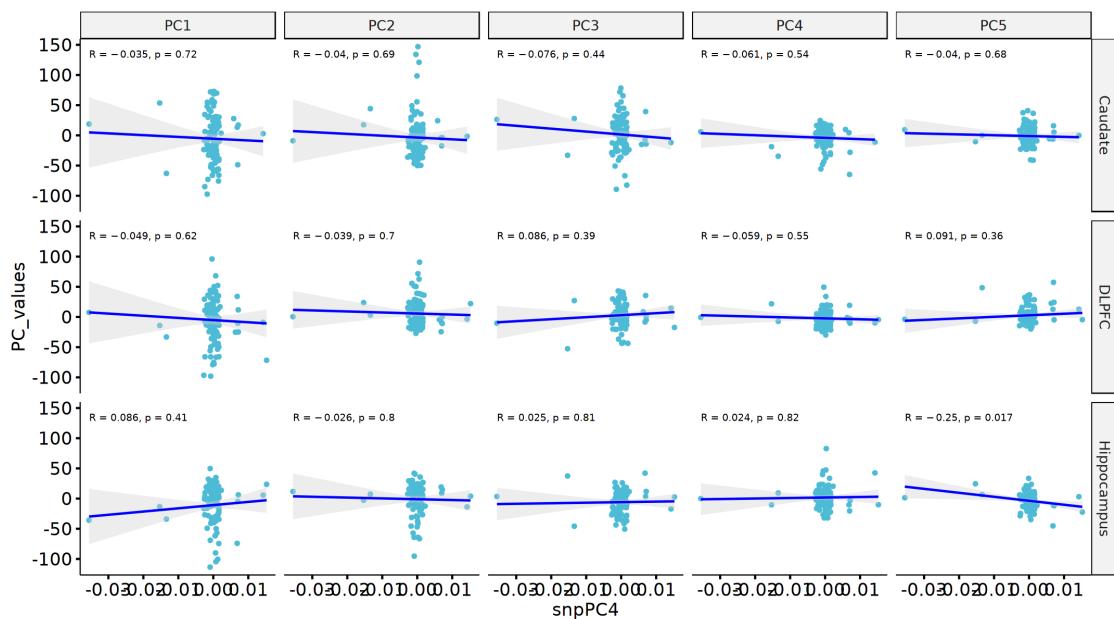
Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."
`geom_smooth()` using formula 'y ~ x'

Warning message:
"Removed 15 rows containing non-finite values (stat_smooth)."
Warning message:
"Removed 15 rows containing non-finite values (stat_cor)."
Warning message:
"Removed 15 rows containing missing values (geom_point)."

```



```
`geom_smooth()` using formula 'y ~ x'
```

Warning message:

```
"Removed 15 rows containing non-finite values (stat_smooth)."
```

Warning message:

```
"Removed 15 rows containing non-finite values (stat_cor)."
```

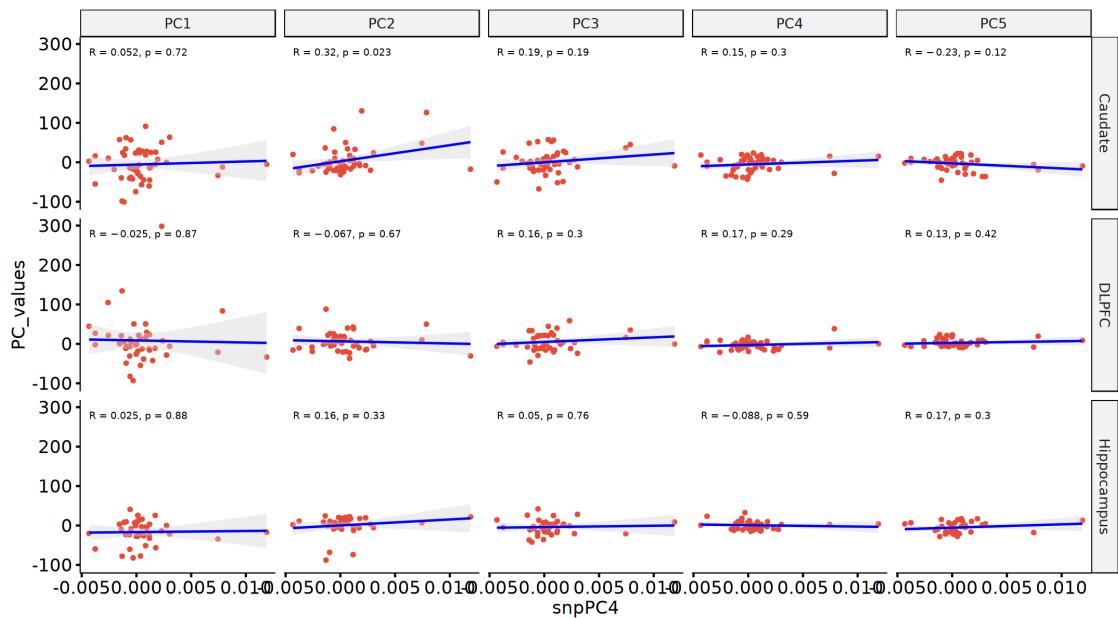
Warning message:

```
"Removed 15 rows containing missing values (geom_point)."
```

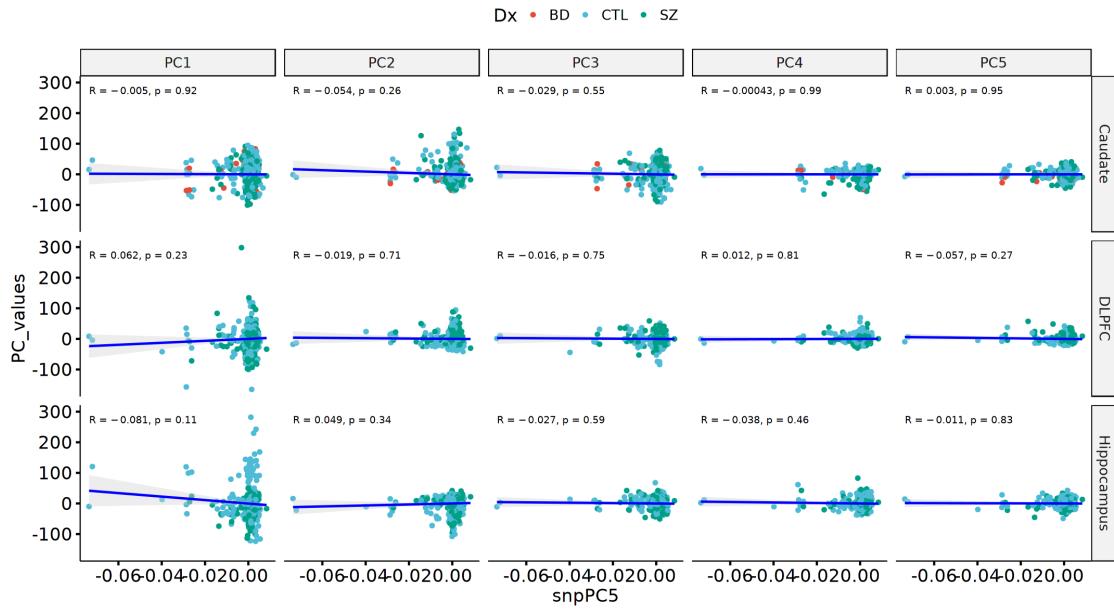
```
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```

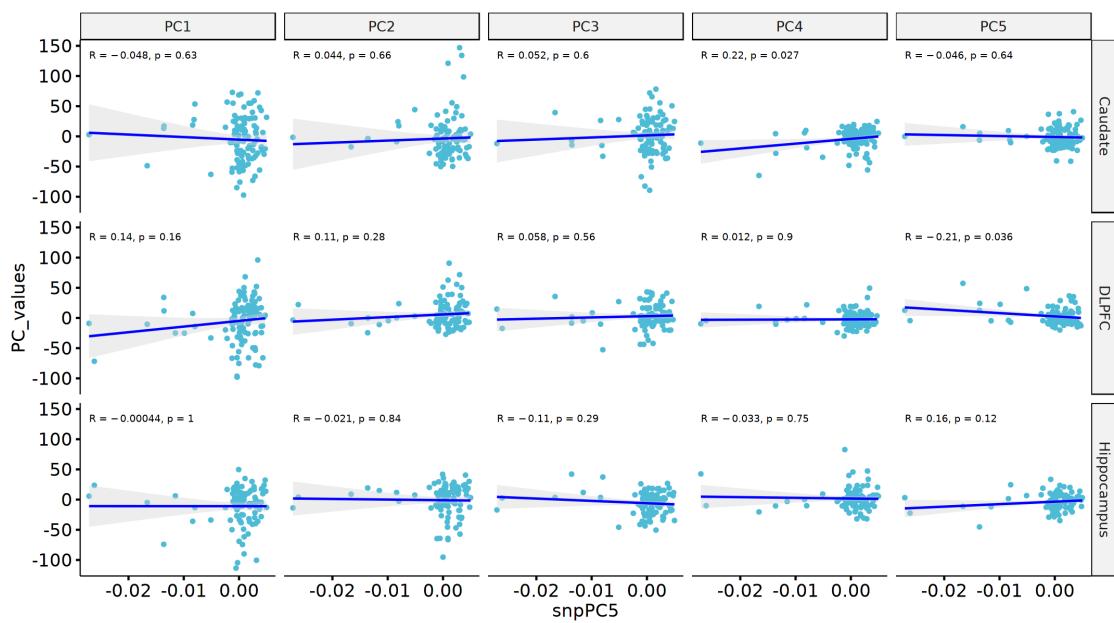
```
`geom_smooth()` using formula 'y ~ x'
```

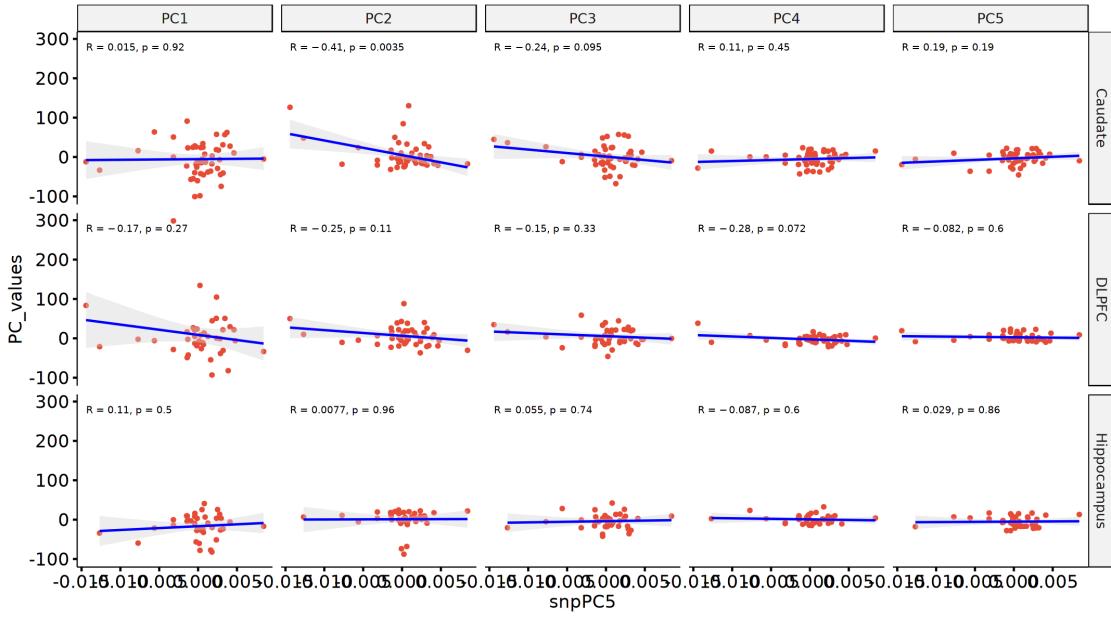


```
`geom_smooth()` using formula 'y ~ x'
```



``geom_smooth()`` using formula `'y ~ x'`





```
[9]: rm(caudate, dlpfc, dtc, dtd, dth, hippo)
```

1.4 Test covariates with Caudate

```
[10]: suppressMessages({library(SummarizedExperiment)
  library(limma)
  library(edgeR)
  library(sva)})
```

Function from jaffelab github

```
merge_rse_metrics <- function(rse) {
  stopifnot(is(rse, 'RangedSummarizedExperiment'))
  stopifnot(
    c('concordMapRate', 'overallMapRate', 'mitoRate', 'rRNA_rate',
      'totalAssignedGene', 'numMapped', 'numReads', 'numUnmapped',
      'mitoMapped', 'totalMapped') %in%
    colnames(SummarizedExperiment::colData(rse)))
  )

  stopifnot(all(sapply(c(
    'concordMapRate', 'overallMapRate', 'mitoRate', 'rRNA_rate',
    'totalAssignedGene', 'numMapped', 'numReads', 'numUnmapped',
    'mitoMapped', 'totalMapped'), function(var) {
      is(colData(rse)[, var], 'List')
    }))
  ))
```

```

rse$concordMapRate = mapply(function(r, n) {
  sum(r*n)/sum(n)
}, rse$concordMapRate, rse$numReads)
rse$overallMapRate = mapply(function(r, n) {
  sum(r*n)/sum(n)
}, rse$overallMapRate, rse$numReads)
rse$mitoRate = mapply(function(r, n) {
  sum(r*n)/sum(n)
}, rse$mitoRate, rse$numMapped)
rse$rRNA_rate = mapply(function(r, n) {
  sum(r*n)/sum(n)
}, rse$rRNA_rate, rse$numMapped)
rse$totalAssignedGene = mapply(function(r, n) {
  sum(r*n)/sum(n)
}, rse$totalAssignedGene, rse$numMapped)

rse$numMapped = sapply(rse$numMapped, sum)
rse$numReads = sapply(rse$numReads, sum)
rse$numUnmapped = sapply(rse$numUnmapped, sum)
rse$mitoMapped = sapply(rse$mitoMapped, sum)
rse$totalMapped = sapply(rse$totalMapped, sum)
return(rse)
}

```

```

[11]: library(memoise)
prep_data <- function(count_fn){
  mds_file = "/ceph/projects/v4_phase3_paper/inputs/genotypes/mds/_m/
  ↳LIBD_Brain_TopMed.mds"
  mds = data.table:::fread(mds_file) %>%
    rename("C1"="snpPC1", "C2"="snpPC2", "C3"="snpPC3",
           "C4"="snpPC4", "C5"="snpPC5") %>%
    mutate_if(is.character, as.factor)
  load(count_fn)
  keepIndex = which((rse_gene$Dx %in% c("Control", "Schizo")) &
                     rse_gene$Age > 17 & rse_gene$Race %in% c("AA", "CAUC"))
  rse_gene = rse_gene[, keepIndex]
  rse_gene$Dx = factor(rse_gene$Dx, levels = c("Control", "Schizo"))
  rse_gene$Sex <- factor(rse_gene$Sex)
  rse_gene <- merge_rse_metrics(rse_gene)
  rse_gene$ERCCsumLogErr <- mapply(function(r, n) {
    sum(r * n)/sum(n)
}, rse_gene$ERCCsumLogErr, rse_gene$numReads)
  colData(rse_gene)$RIN = sapply(colData(rse_gene)$RIN, "[", 1)
  pheno = colData(rse_gene) %>% as.data.frame %>%
    inner_join(mds, by=c("BrNum"="FID")) %>%
    distinct(RNum, .keep_all = TRUE)
  # Generate DGE list

```

```

x <- DGEList(counts=assays(rse_gene)$counts[,pheno$RNum],
              genes=rowData(rse_gene), samples=pheno)
# Filter by expression
design0 <- model.matrix(~Dx, data=x$samples)
keep.x <- filterByExpr(x, design=design0)
x <- x[keep.x, , keep.lib.sizes=FALSE]
print(paste('There are:', sum(keep.x), 'features left!', sep=' '))
# Normalize library size
x <- calcNormFactors(x, method="TMM")
return(x)
}

memo_prepData <- memoise(prep_data)

qsv_model <- function(count_fn, qsv_file){
  x <- memo_prepData(count_fn)
  # Design matrix
  mod = model.matrix(~Dx + Age + Sex + mitoRate + rRNA_rate +
    totalAssignedGene + RIN + overallMapRate +
    snpPC1 + snpPC2 + snpPC3, data = x$samples)
  colnames(mod) <- gsub("Dx", "", colnames(mod))
  colnames(mod) <- gsub("SexM", "Male", colnames(mod))
  colnames(mod) <- gsub("\\\\(Intercept\\\\)", "Intercept", colnames(mod))
  # qSV
  modQsva <- mod %>% as.data.frame %>% rownames_to_column() %>%
    inner_join(data.table::fread(qsv_file), by=c("rowname"="V1")) %>%
    rename_all(list(~str_replace_all(., 'PC', 'qPC'))) %>%
    column_to_rownames("rowname") %>% as.matrix
  return(modQsva)
}

memo_qsvModel <- memoise(qsv_model)

pca_select <- function(count_fn){
  ### Dimensional reduction (PCA)
  x <- memo_prepData(count_fn)
  log2cpm = cpm(x, log=TRUE) %>% t
  pca_df = prcomp(log2cpm, center=TRUE)$x
  dt = pca_df %>% as.data.frame %>% rownames_to_column() %>%
    select(c(rowname, PC1, PC2, PC3, PC4, PC5)) %>%
    pivot_longer(-rowname, names_to="PC", values_to="PC_values")
  return(dt)
}

memo_pcaSelect <- memoise(pca_select)

save_norm <- function(count_fn){

```

```

x <- memo_prepData(count_fn)
return(cpm(x, log=TRUE))
}

memNORM <- memoise(save_norm)

get_voom <- function(count_fn, qsv_file){
  ### Preform voom
  x <- memo_prepData(count_fn)
  modQsva <- memo_qsvModel(count_fn, qsv_file)
  v <- voom(x, modQsva, plot=FALSE)
  return(v)
}

memo_voom <- memoise(get_voom)

cal_res <- function(count_fn, qsv_file){
  ### Calculate residuals
  v <- memo_voom(count_fn, qsv_file)
  null_model <- v$design %>% as.data.frame %>% select(-c("Schizo")) %>% as.
  ↪matrix
  fit_res <- lmFit(v, design=null_model)
  res = v$E - (fit_res$coefficients %*% t(null_model) )
  res_sd = apply(res, 1, sd)
  res_mean = apply(res, 1, mean)
  res_norm = (res - res_mean) / res_sd
  return(res_norm)
}

memo_res <- memoise(cal_res)

## Normalize residuals
pca_res <- function(count_fn, qsv_file){
  res_df = memo_res(count_fn, qsv_file) %>% t
  pca_df = prcomp(res_df, center=TRUE)$x
  dt = pca_df %>% as.data.frame %>% rownames_to_column %>%
    select(c(rowname, PC1, PC2, PC3, PC4, PC5)) %>%
    pivot_longer(-rowname, names_to="PC", values_to="PC_values")
  return(dt)
}

memo_pcaRES <- memoise(pca_res)

```

1.5 Differential Expression Analysis

```
[12]: counts_lt = list("Caudate"=paste0("/ceph/projects/v3_phase3_paper/inputs/phase3/
→_m/count_data/",

→"caudate_brainseq_phase3_hg38_rseGene_merged_n464.rda"),
"DLPPFC"=paste0("/ceph/projects/v3_phase3_paper/inputs/phase2/
→_m/count_data/",

→"dlppfc_ribozero_brainseq_phase2_hg38_rseGene_merged_n453.rda"),
"HIPPO"=paste0("/ceph/projects/v3_phase3_paper/inputs/phase2/
→_m/count_data/",

→"hippo_brainseq_phase2_hg38_rseGene_merged_n447.
→rda"))

qsv_lt = list("Caudate"="/ceph/projects/v4_phase3_paper/inputs/counts/
→text_files_counts/_m/caudate/qSV_caudate.csv",
"DLPPFC"="/ceph/projects/v4_phase3_paper/inputs/counts/
→text_files_counts/_m/dlppfc/qSV_dlppfc.csv",
"HIPPO"="/ceph/projects/v4_phase3_paper/inputs/counts/
→text_files_counts/_m/hippocampus/qSV_hippo.csv")
```

1.5.1 Plot scatters and correlations with covariates

```
[13]: for(tissue in c('Caudate', 'DLPPFC', 'HIPPO')){
  flush.console()
  lotissue = paste0(tolower(tissue), "/")
  dir.create(lotissue)
  options(repr.plot.width=18, repr.plot.height=8)
  modQsva <- memo_qsvModel(counts_lt[[tissue]], qsv_lt[[tissue]])
  dt <- memo_pcaSelect(counts_lt[[tissue]])
  covar_model = modQsva %>% as.data.frame %>% rownames_to_column
  cols = modQsva %>% as.data.frame %>%
    select(-c(Intercept, Male, Schizo)) %>% colnames
  for(covar in cols){
    flush.console()
    sca = dt %>% inner_join(covar_model, by="rowname") %>%
      inner_join(memPHENO()[, c("RNum", "Sex", "Dx")], □
    →by=c("rowname"="RNum")) %>%
      ggscatter(y="PC_values", x=covar, color="Dx", facet.by=c('PC'), □
    →ncol=5,
              add='reg.line', conf.int=TRUE, cor.coef=TRUE, □
    →palette="npg",
              add.params=list(color="blue", fill="lightgray"),
              ggtheme=theme_pubr(base_size=20))
      save_img(sca, paste0(lotissue,"scatter_log2cpm_dx_5pcs_",covar), w=18, □
    →h=8)
      print(sca)
```

```

}

cols = modQsva %>% as.data.frame %>%
  select(-c(Intercept, Male, Schizo)) %>% colnames
memNORM(counts_lt[[tissue]]) %>% as.data.frame %>%
  data.table::fwrite(paste0(lotissue, "normalized_expression.tsv"),
                      sep='\t', row.names=TRUE)
memo_res(counts_lt[[tissue]], qsv_lt[[tissue]]) %>% as.data.frame %>%
  data.table::fwrite(paste0(lotissue, "residualized_expression.tsv"),
                      sep='\t', row.names=TRUE)
dt <- memo_pcaRES(counts_lt[[tissue]], qsv_lt[[tissue]])
for(covar in cols){
  flush.console()
  sca = dt %>% inner_join(covar_model, by="rowname") %>%
    inner_join(memPHENO()[, c("RNum", "Sex", "Dx")], □
  ↪by=c("rowname"="RNum")) %>%
    ggscatter(y="PC_values", x=covar, color="Dx", facet.by=c('PC'), □
  ↪ncol=5,
              add='reg.line', conf.int=TRUE, cor.coef=TRUE, □
  ↪palette="npg",
              add.params=list(color="blue", fill="lightgray"),
              ggtheme=theme_pubr(base_size=20))
  save_img(sca, paste0(lotissue, "scatter_normres_dx_5pcs_", covar), w=18, □
  ↪h=8)
  print(sca)
}
}

```

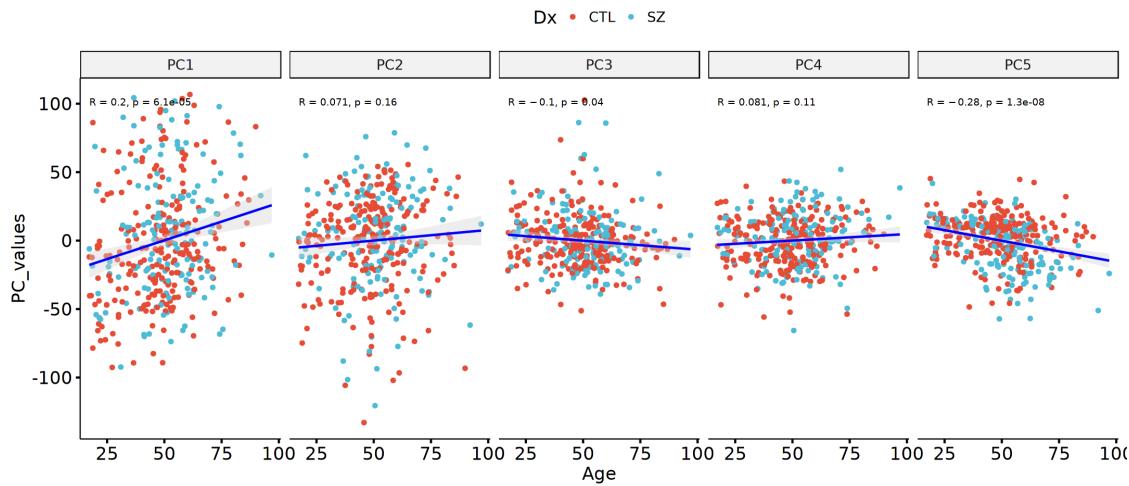
```

[1] "There are: 22958 features left!"

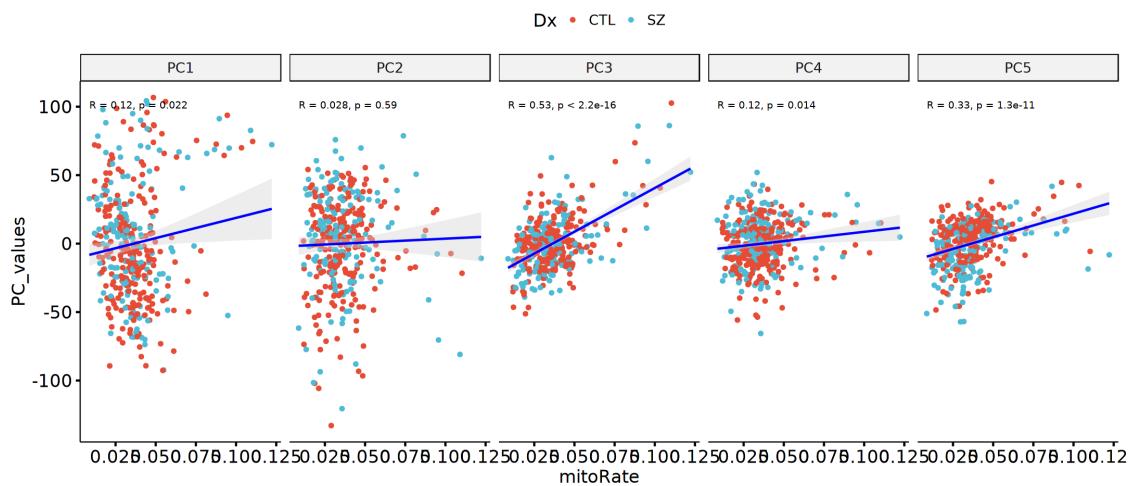
`geom_smooth()` using formula 'y ~ x'

```

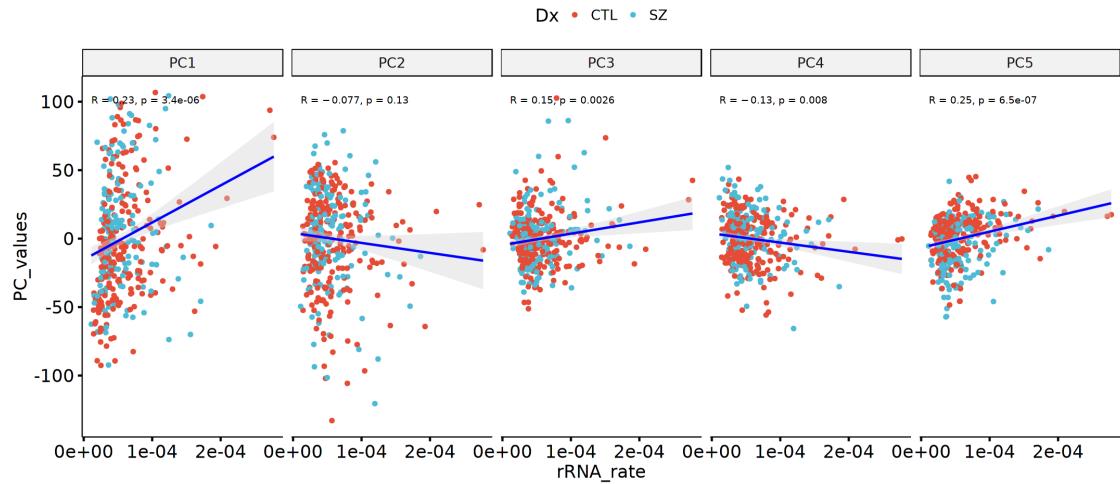
```
`geom_smooth()` using formula 'y ~ x'
```



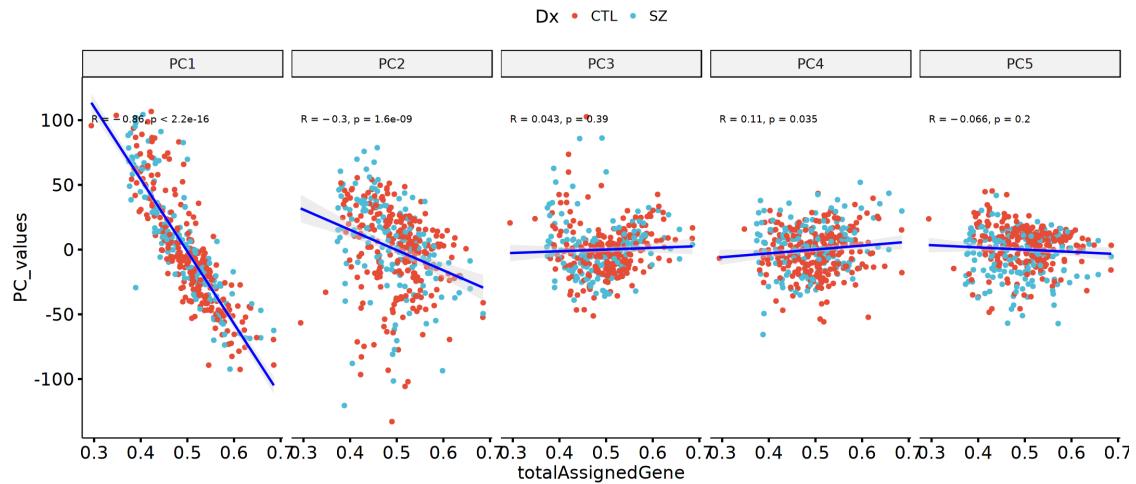
```
`geom_smooth()` using formula 'y ~ x'
```



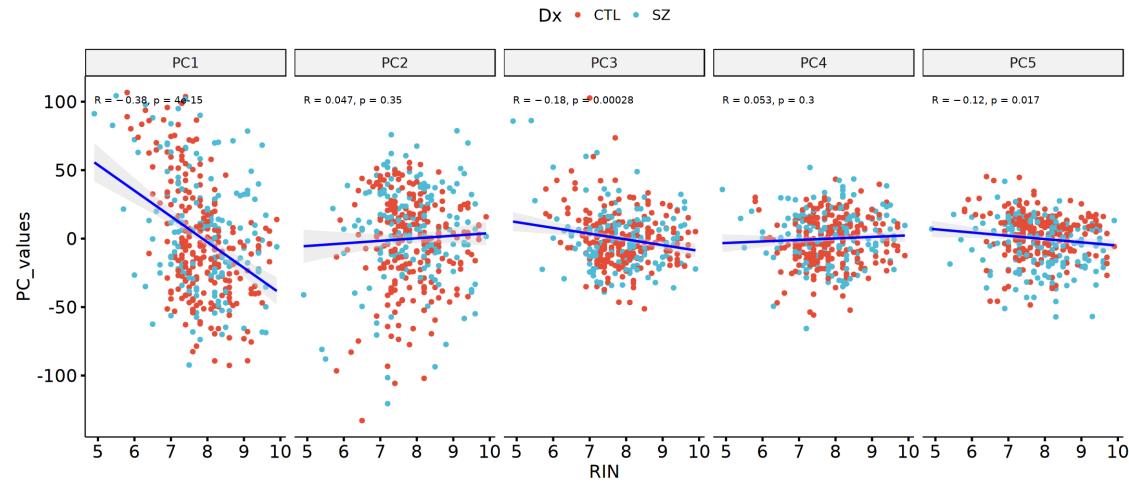
```
`geom_smooth()` using formula 'y ~ x'
```



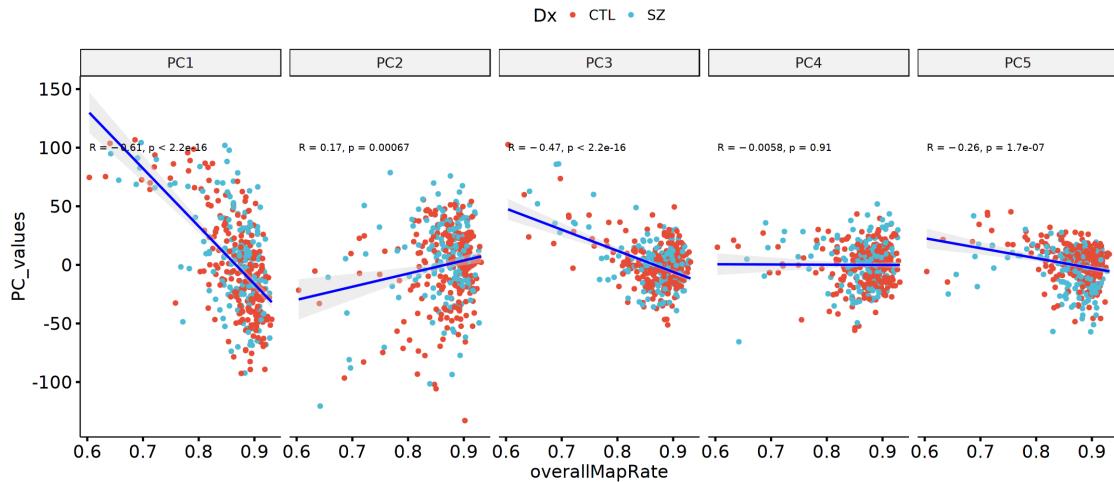
```
`geom_smooth()` using formula 'y ~ x'
```



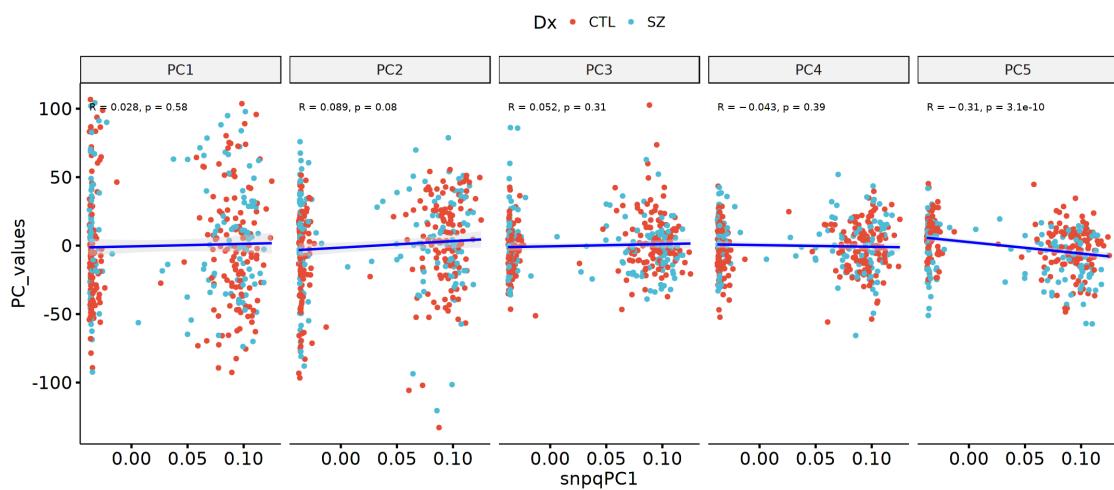
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

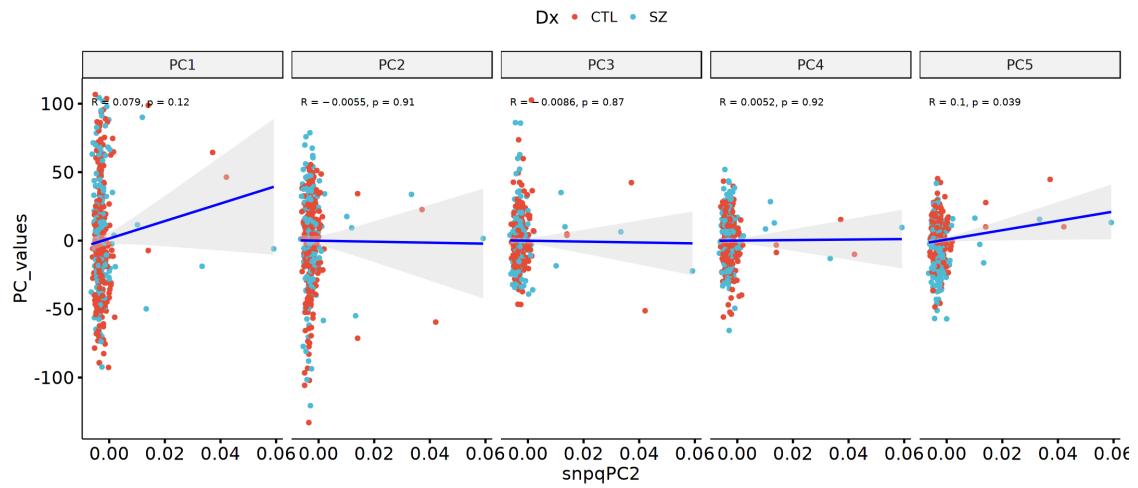


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

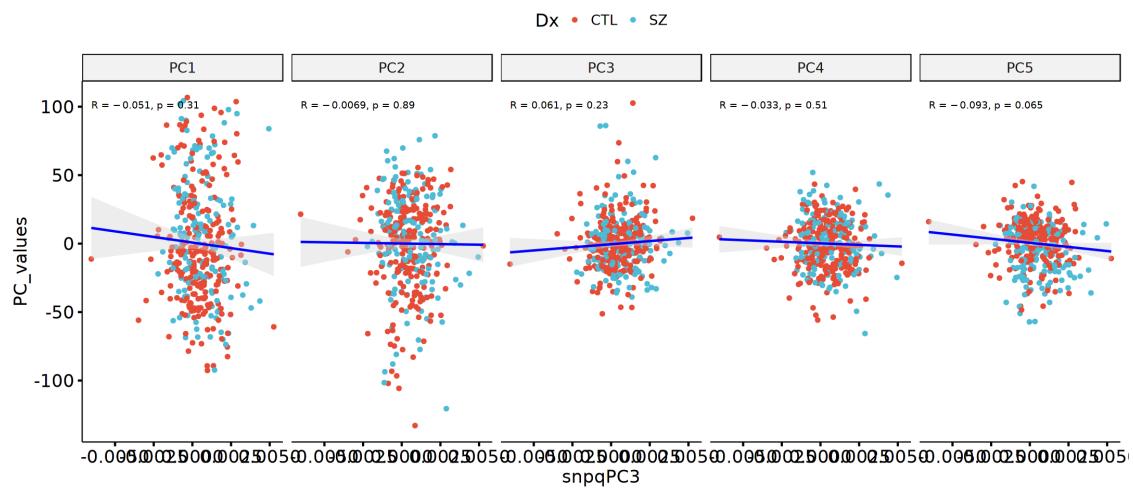


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

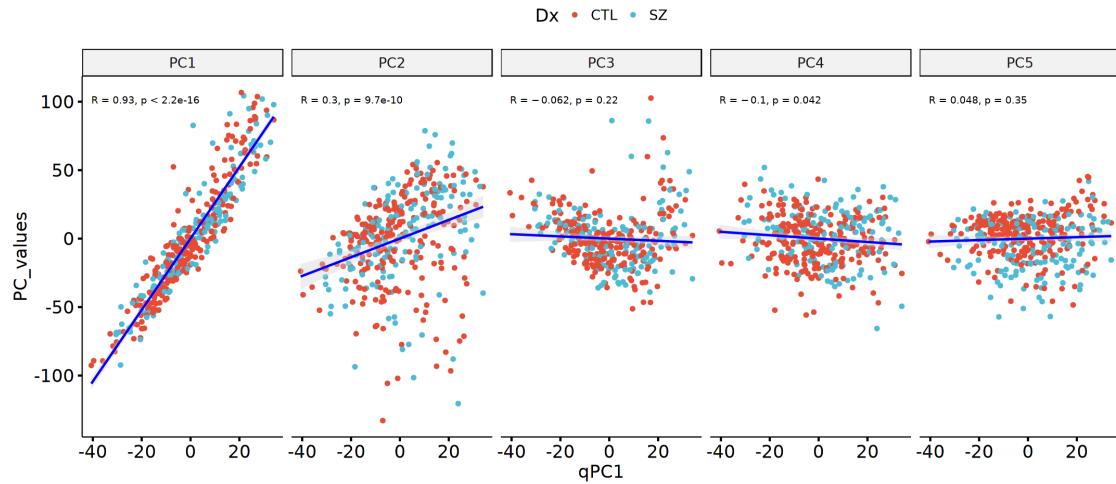
```
`geom_smooth()` using formula 'y ~ x'
```



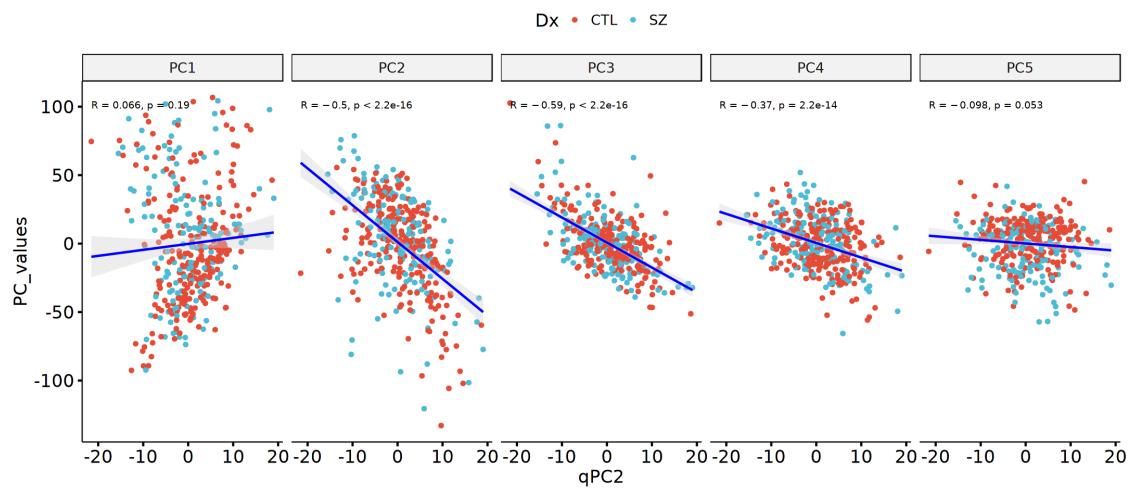
```
`geom_smooth()` using formula 'y ~ x'
```



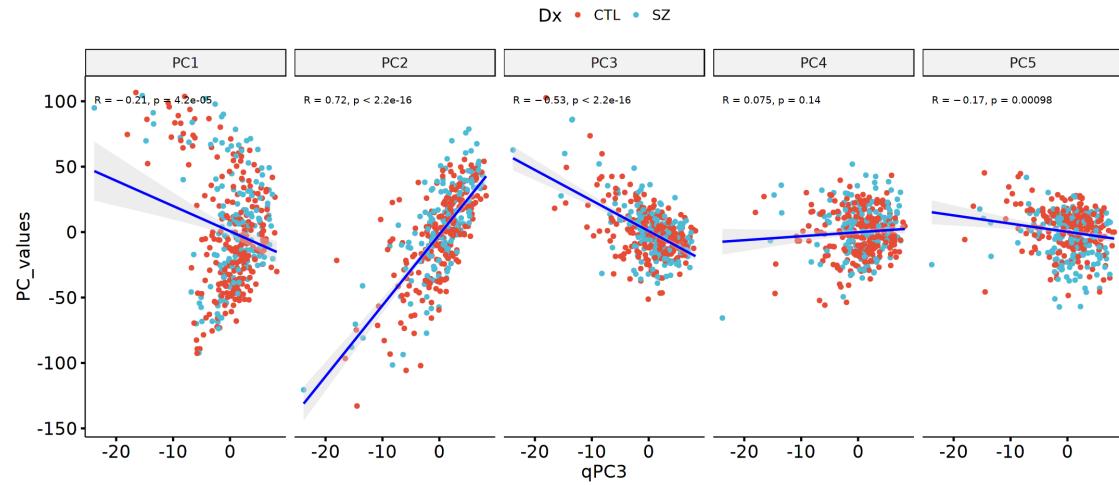
```
`geom_smooth()` using formula 'y ~ x'
```



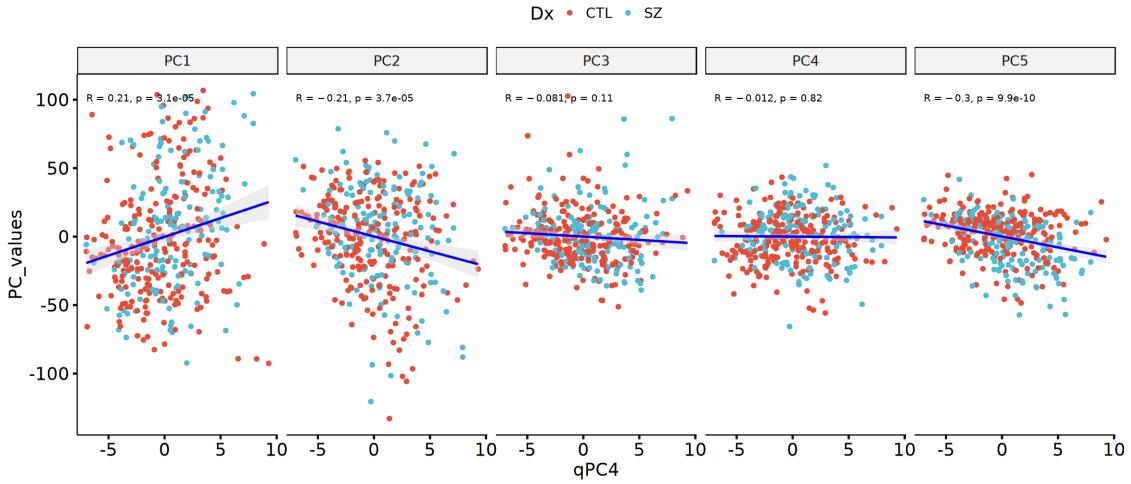
```
`geom_smooth()` using formula 'y ~ x'
```



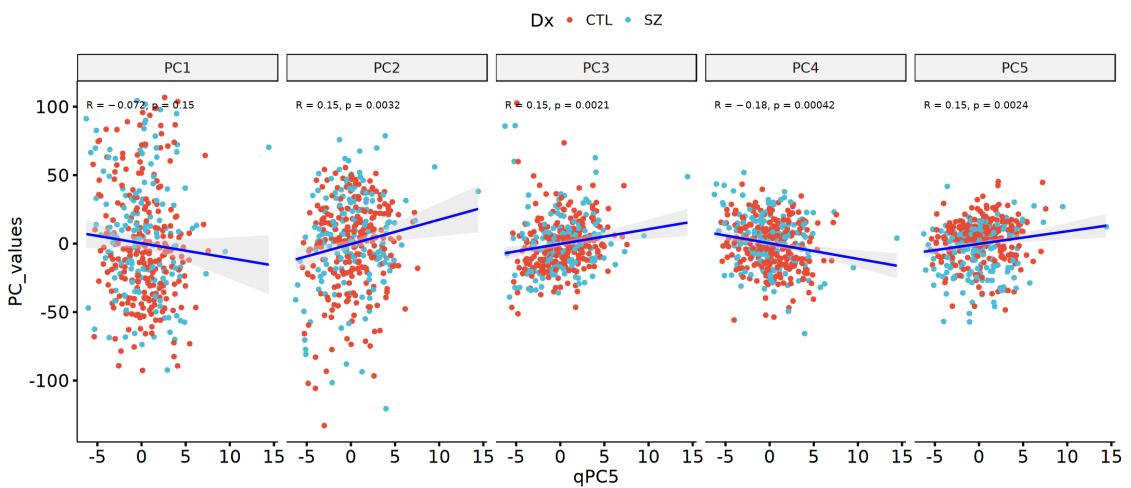
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

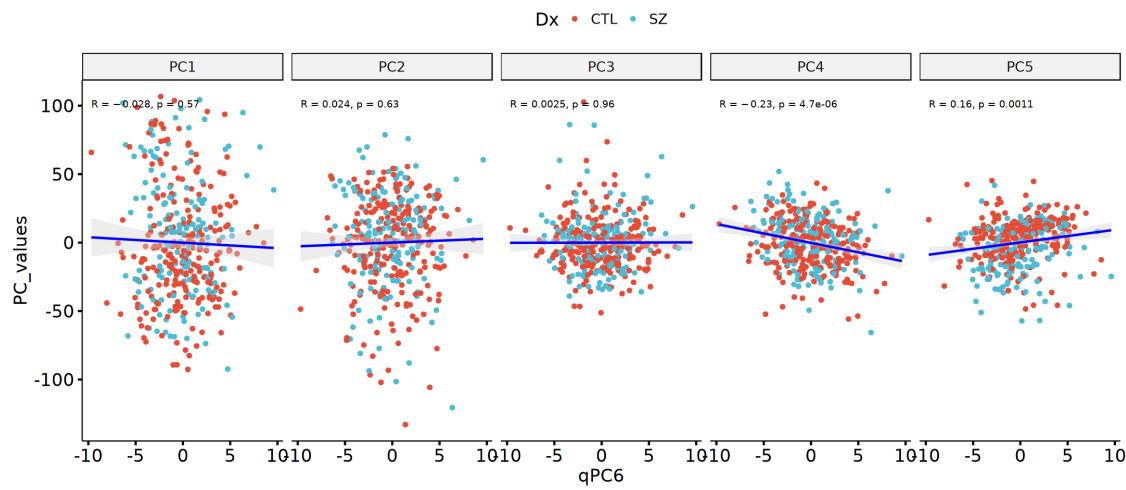


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

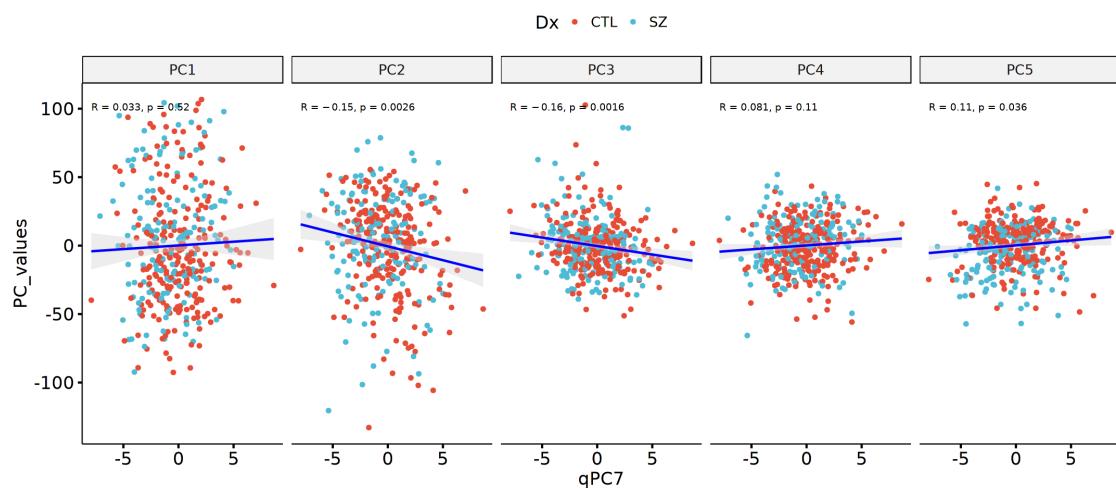


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

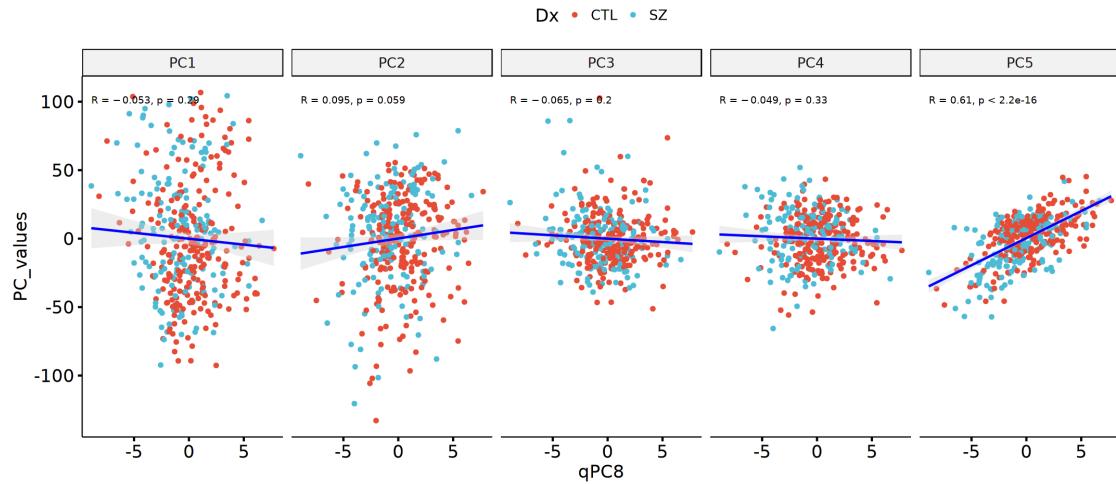
```
`geom_smooth()` using formula 'y ~ x'
```



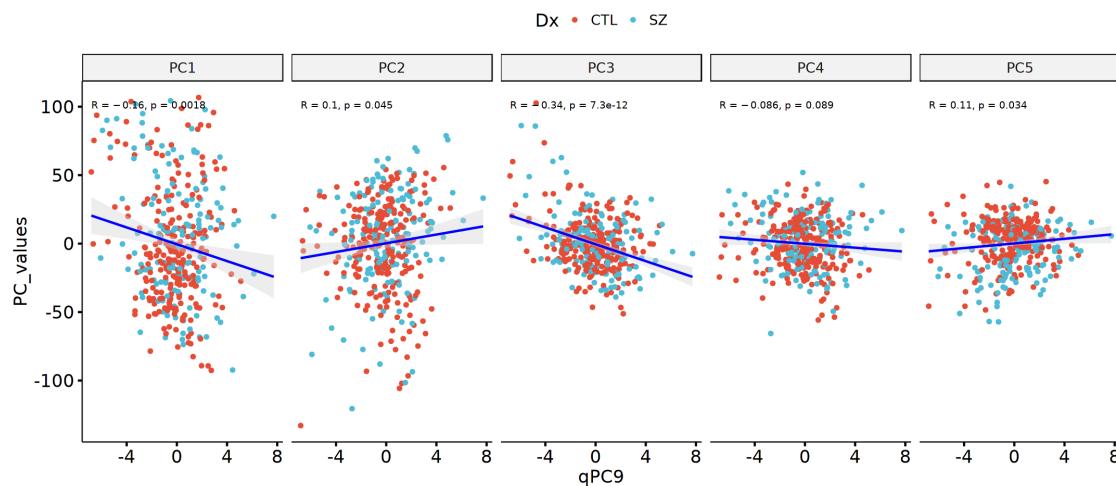
```
`geom_smooth()` using formula 'y ~ x'
```



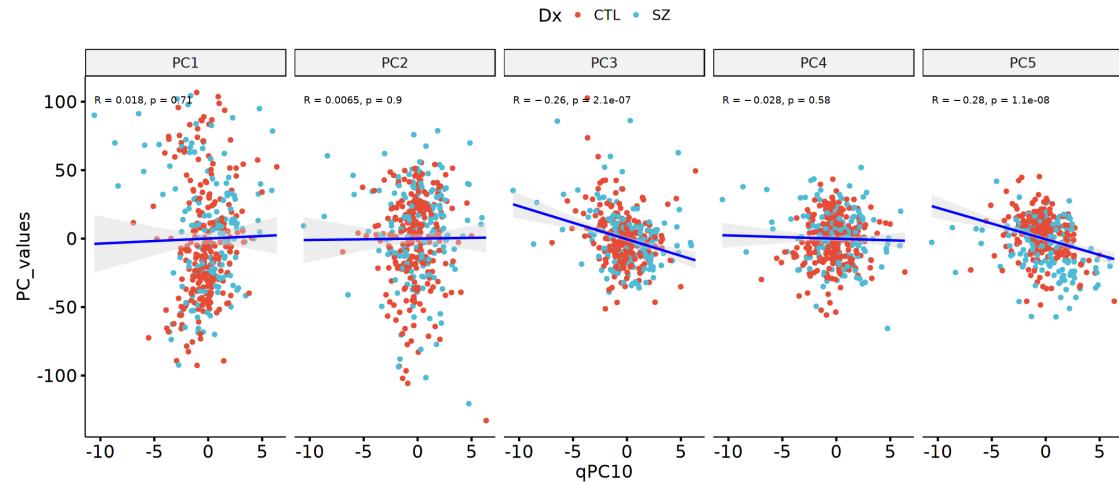
```
`geom_smooth()` using formula 'y ~ x'
```



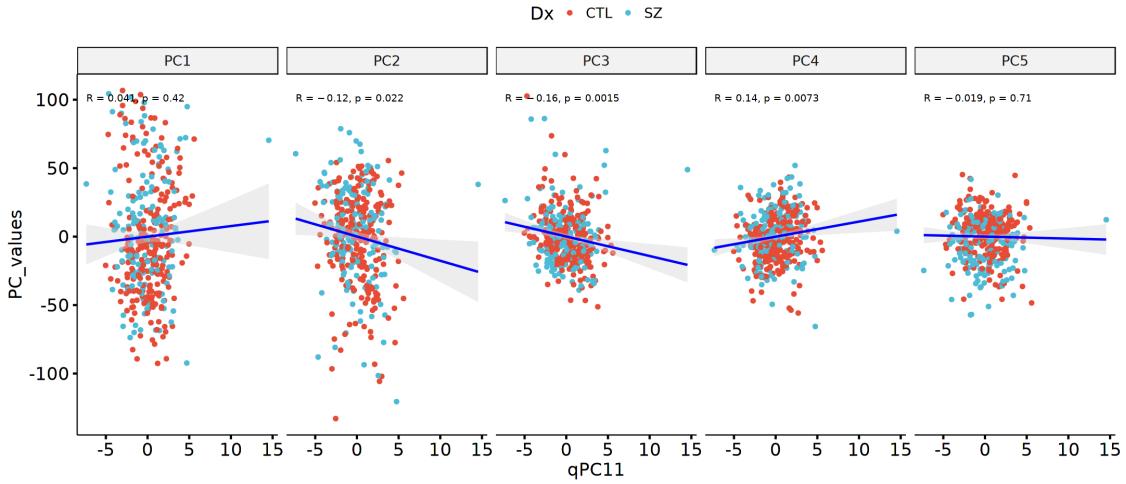
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

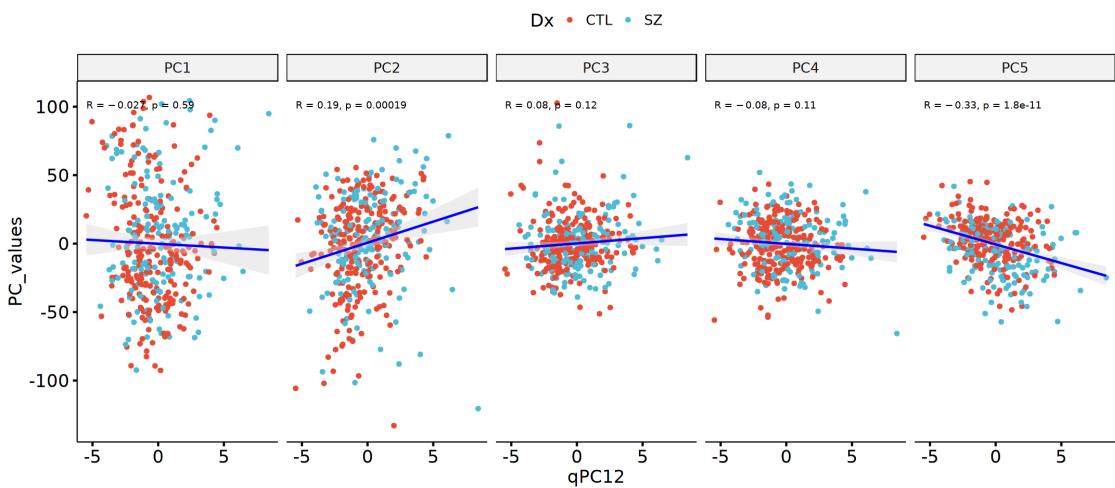


```
`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'
```

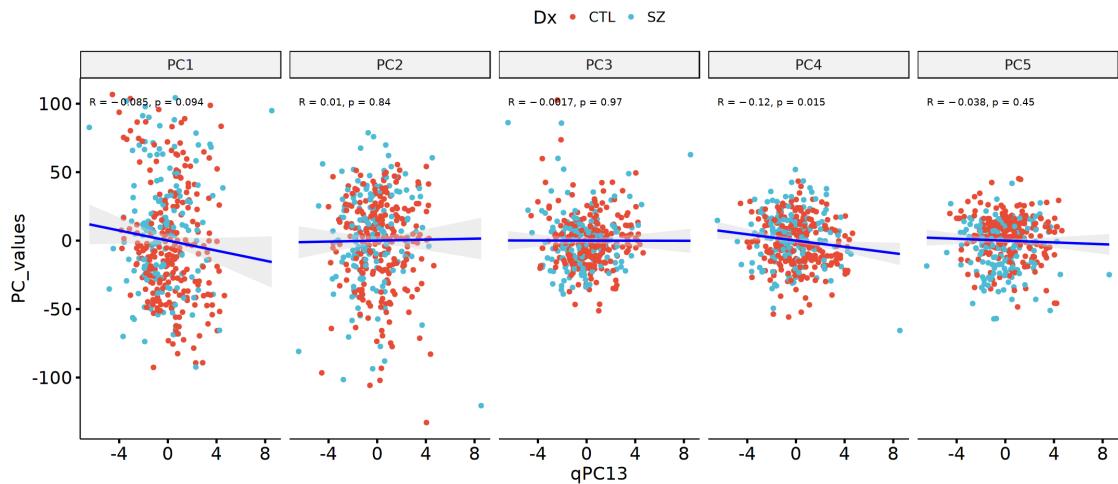


```
`geom_smooth()` using formula 'y ~ x'  

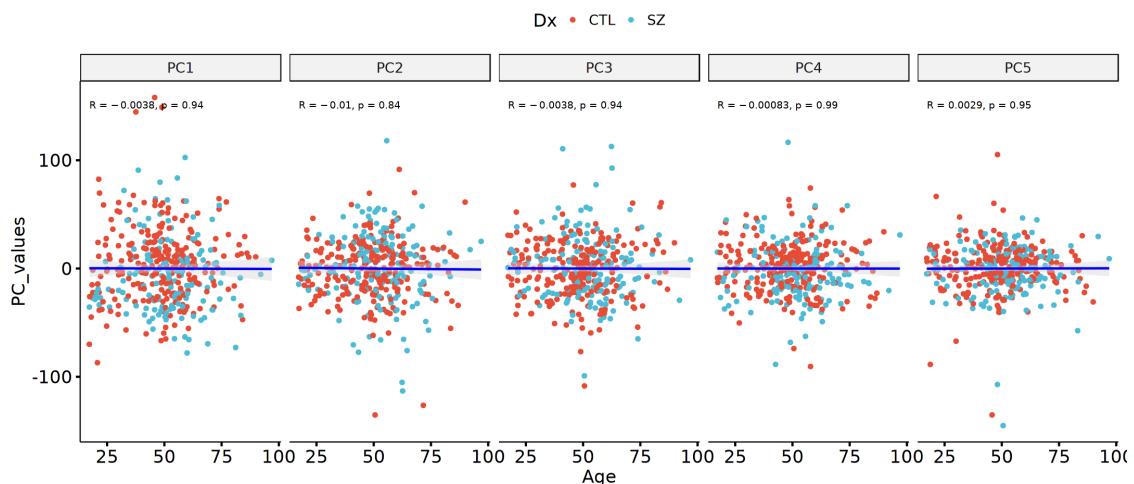
`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

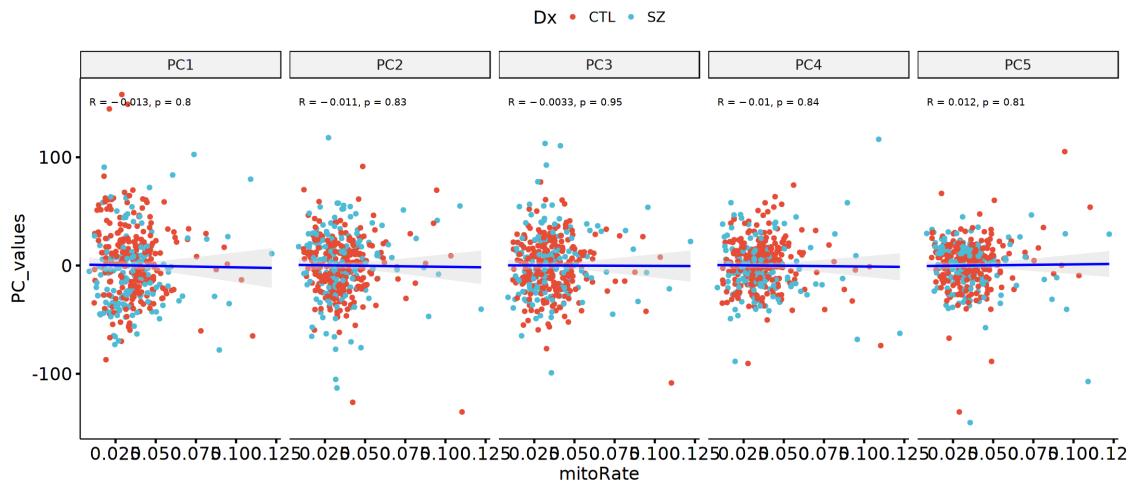


```
`geom_smooth()` using formula 'y ~ x'
```

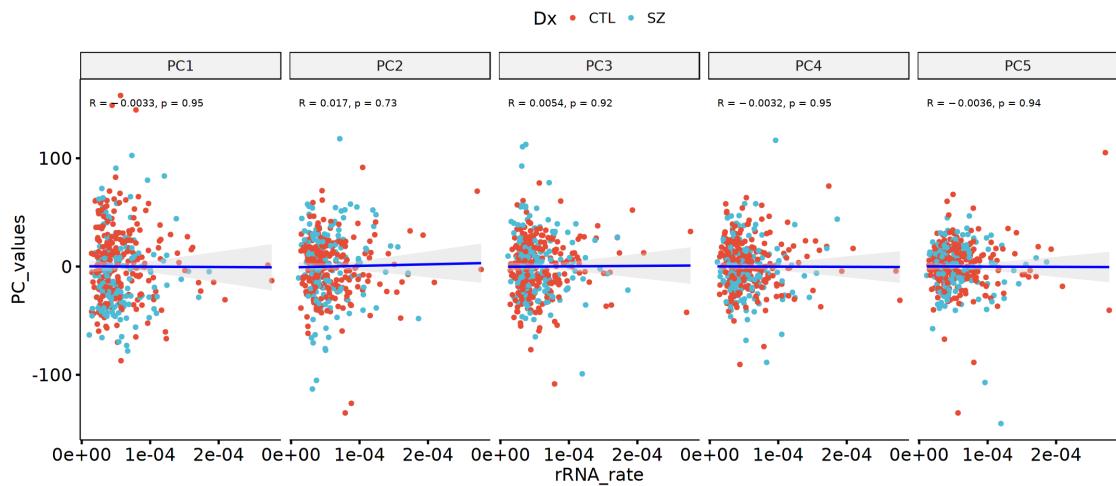
``geom_smooth()`` using formula `'y ~ x'`

``geom_smooth()` using formula 'y ~ x'`

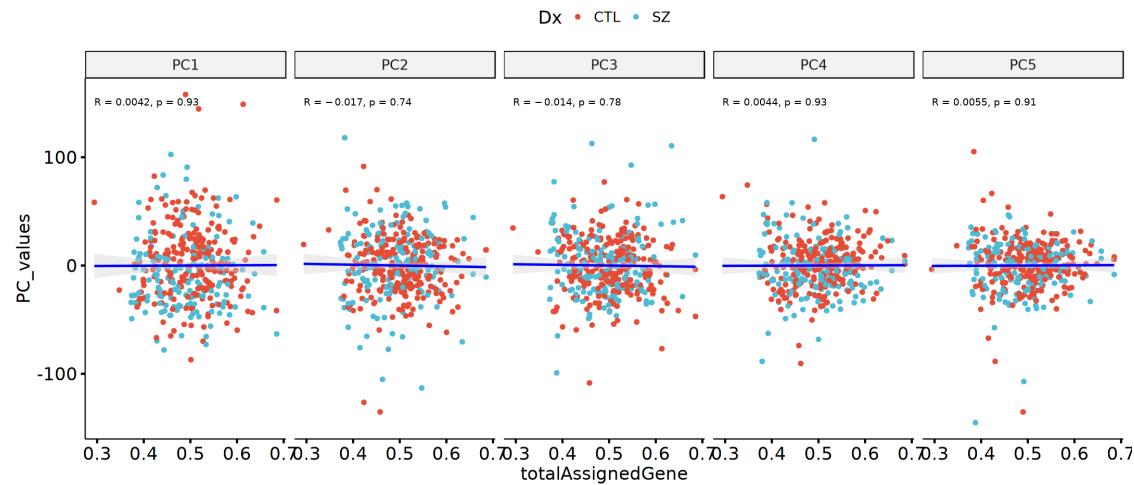
``geom_smooth()`` using formula ' $y \sim x$ '



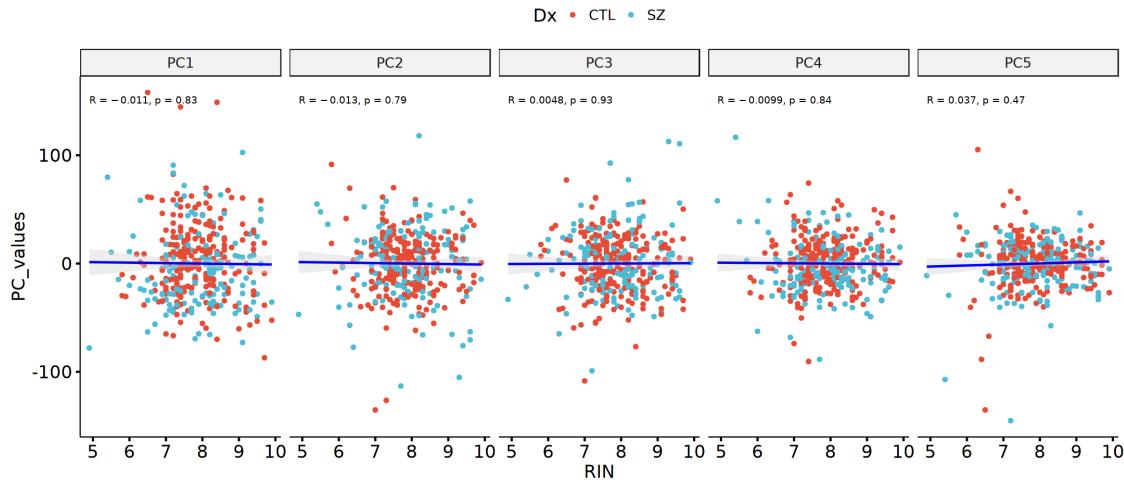
`geom_smooth()` using formula 'y ~ x'



```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

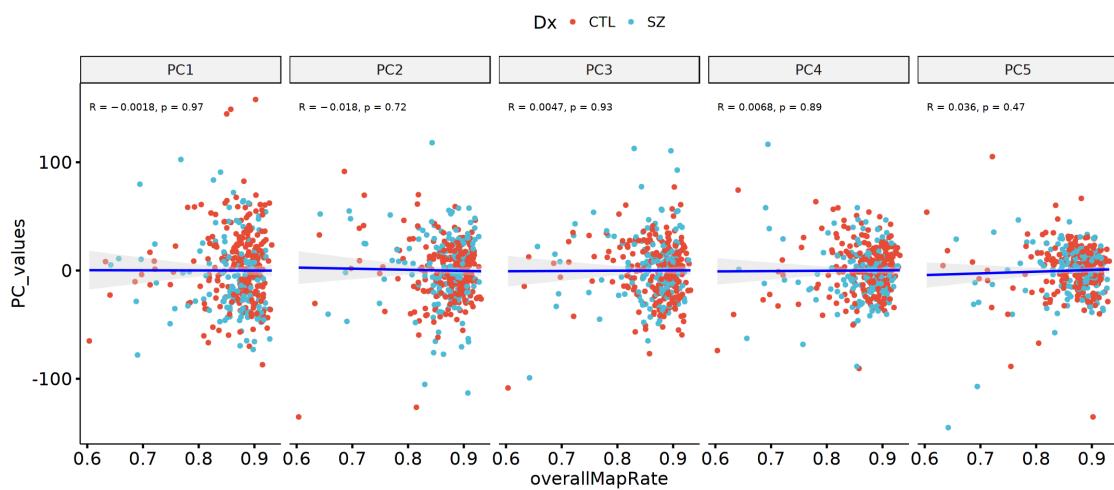


```
`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'
```

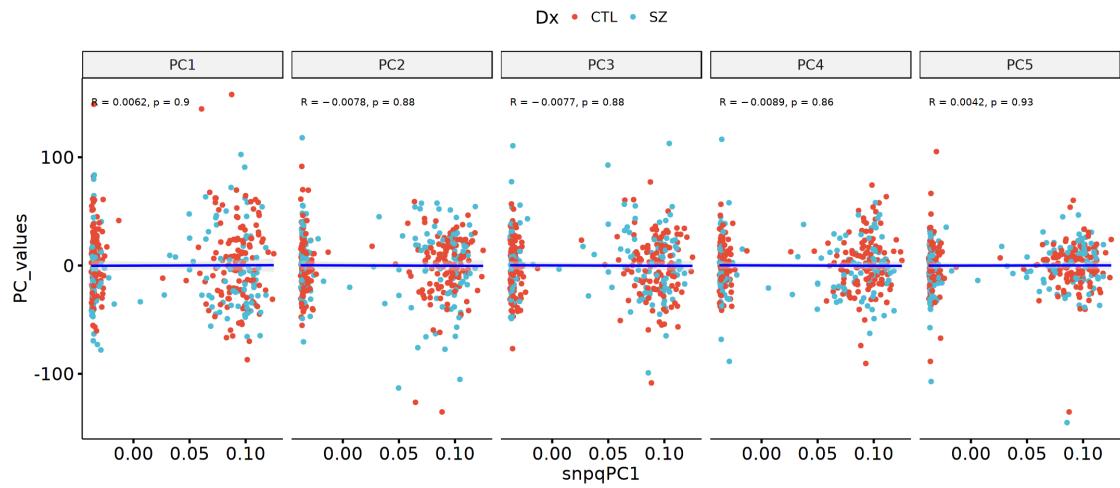


```
`geom_smooth()` using formula 'y ~ x'  

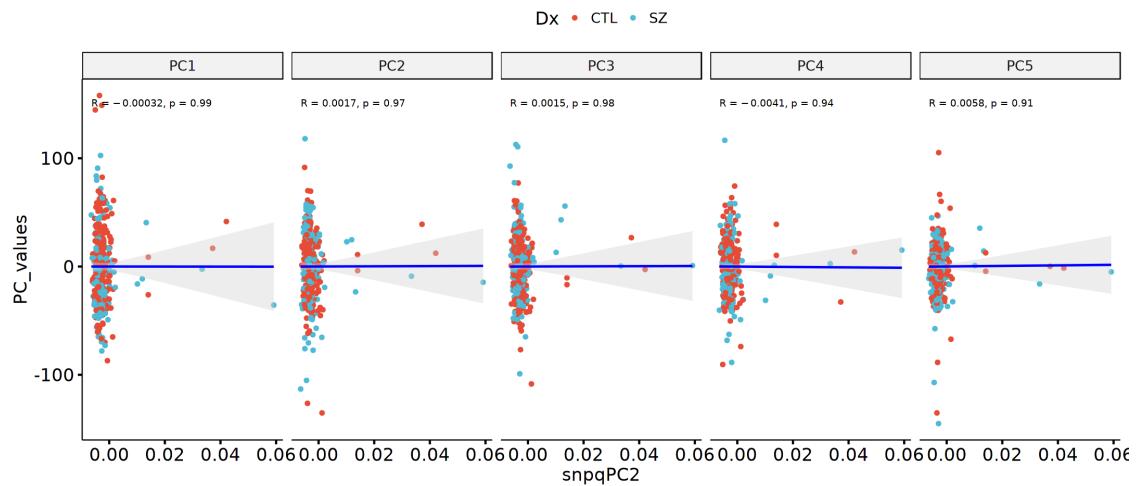
`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'
```

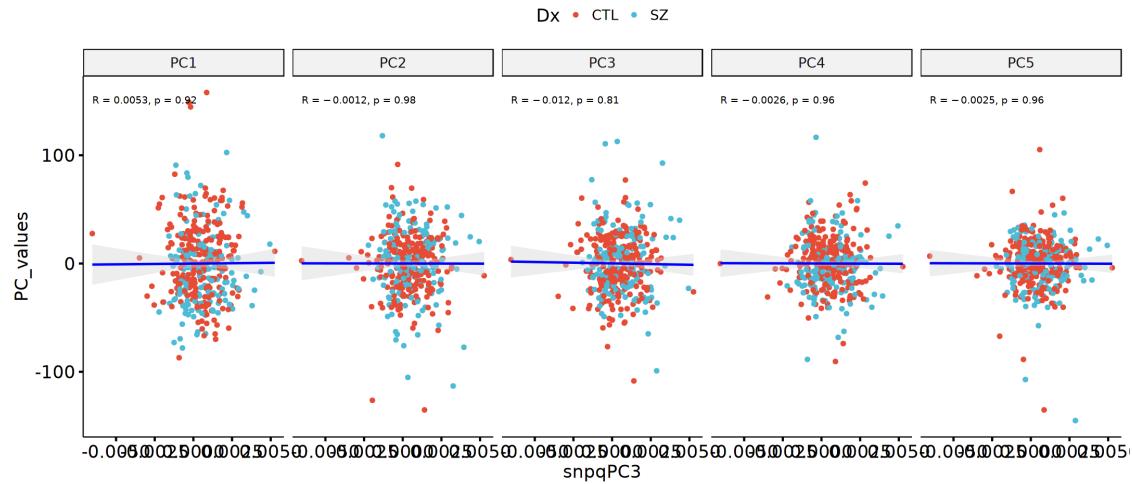
```
`geom_smooth()` using formula 'y ~ x'
```



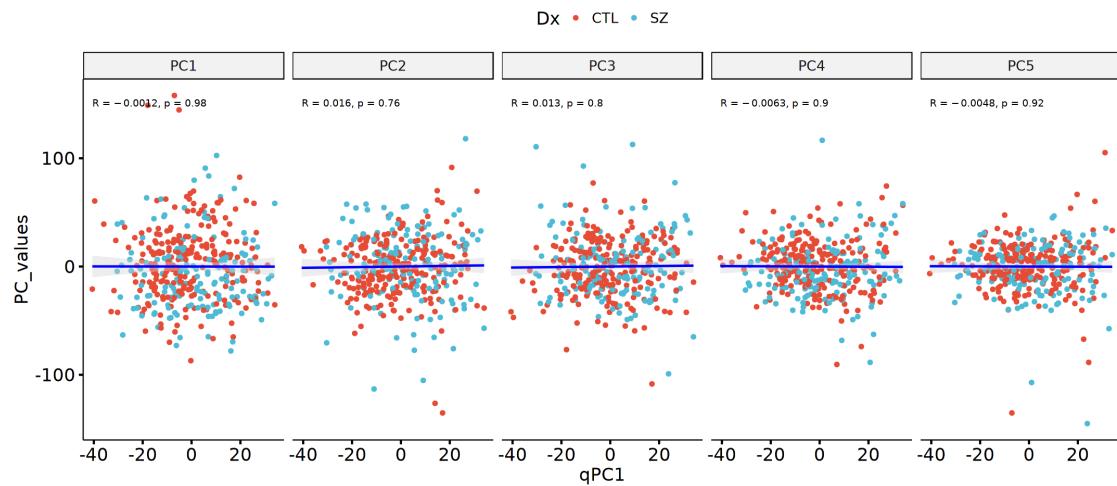
```
`geom_smooth()` using formula 'y ~ x'
```



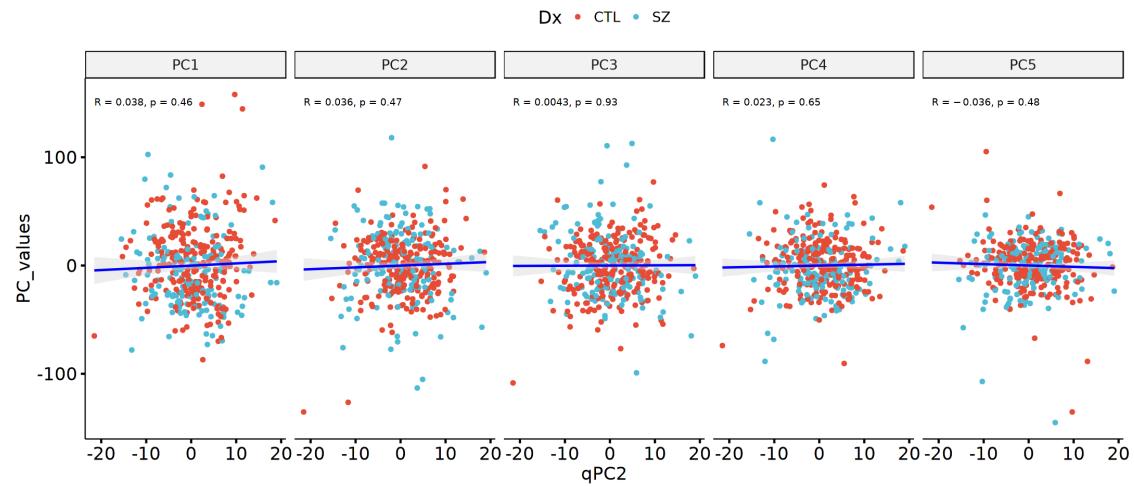
```
`geom_smooth()` using formula 'y ~ x'
```



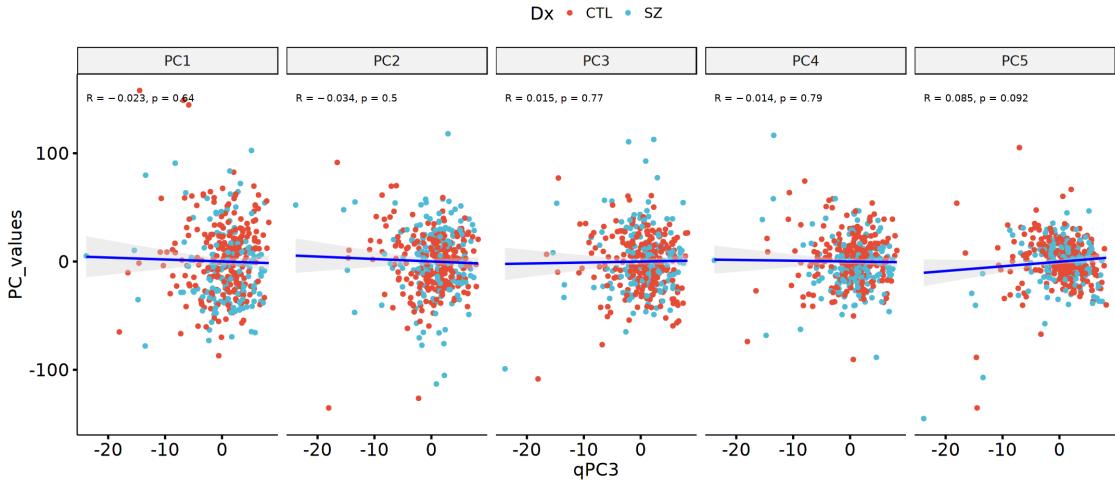
```
`geom_smooth()` using formula 'y ~ x'
```



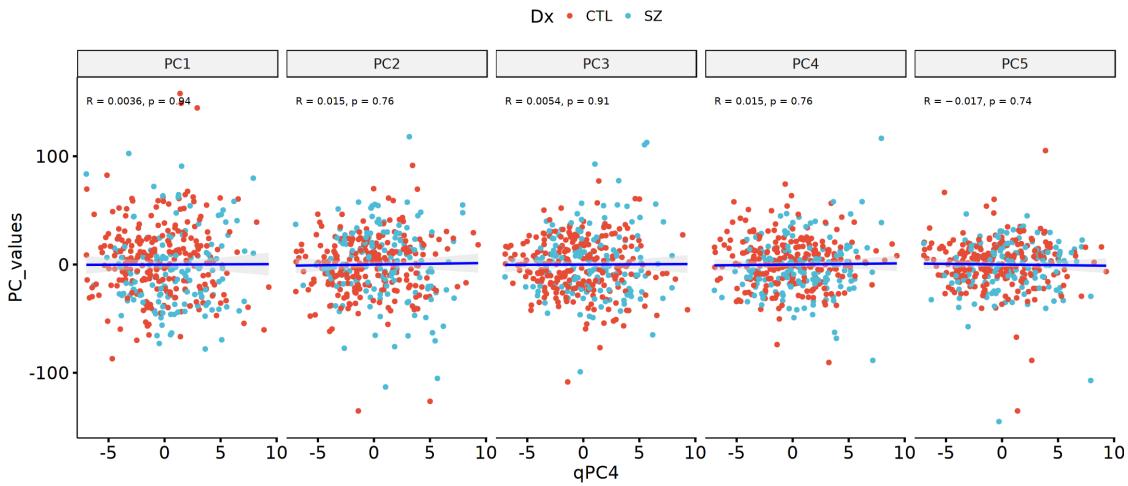
```
`geom_smooth()` using formula 'y ~ x'
```



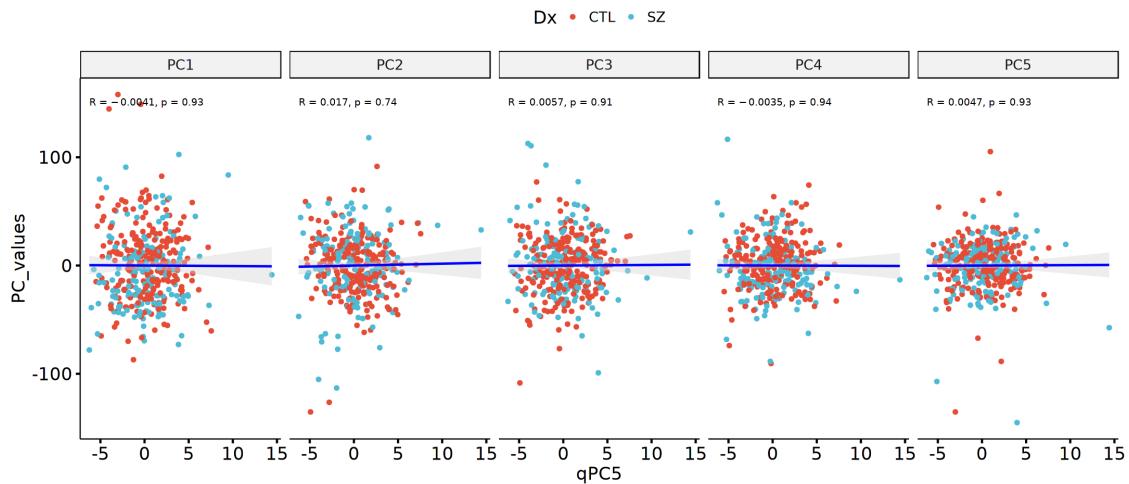
```
`geom_smooth()` using formula 'y ~ x'
```



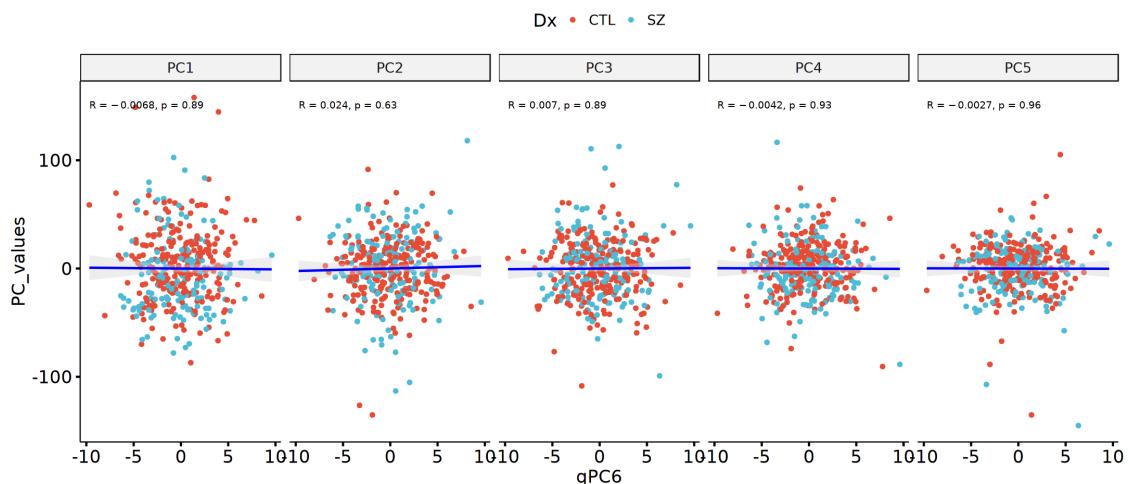
```
`geom_smooth()` using formula 'y ~ x'
```



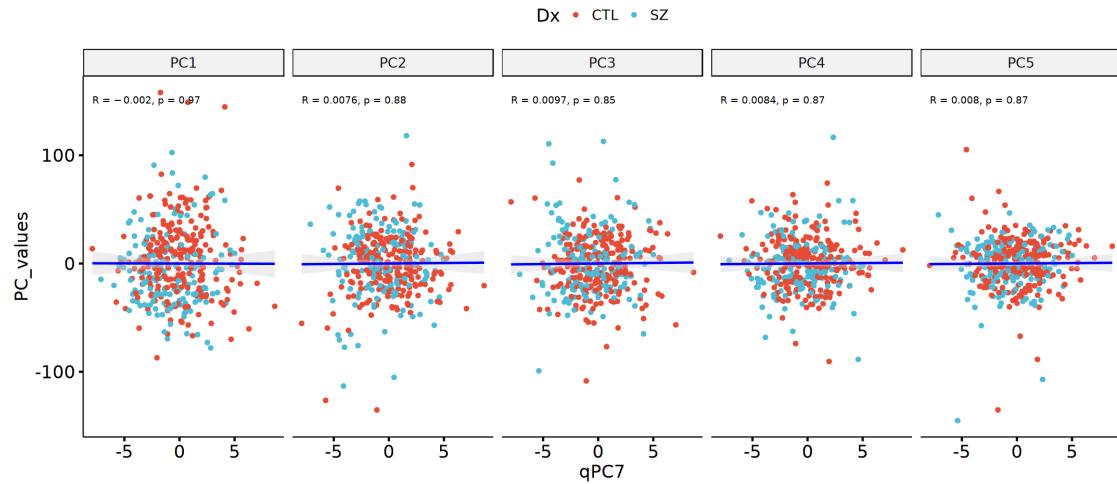
```
`geom_smooth()` using formula 'y ~ x'
```



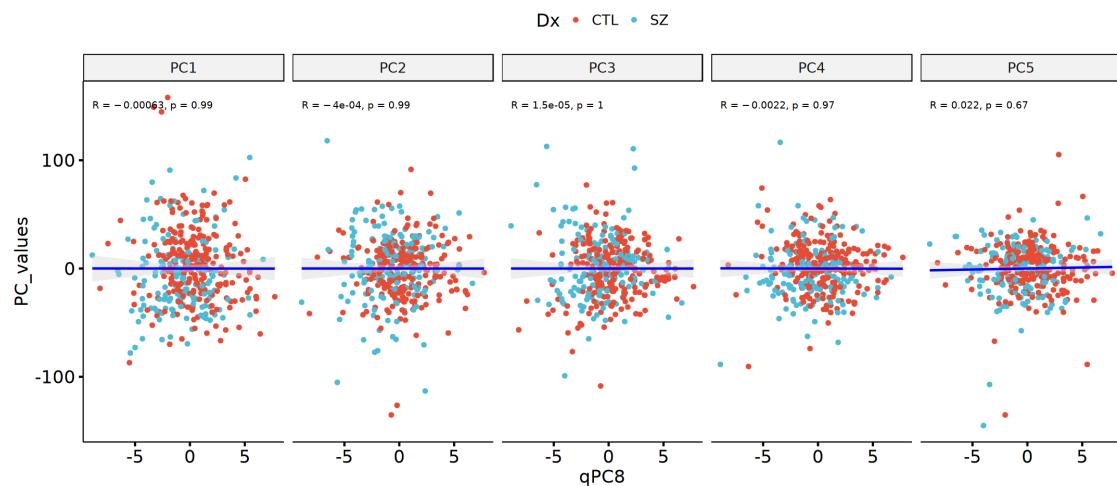
```
`geom_smooth()` using formula 'y ~ x'
```



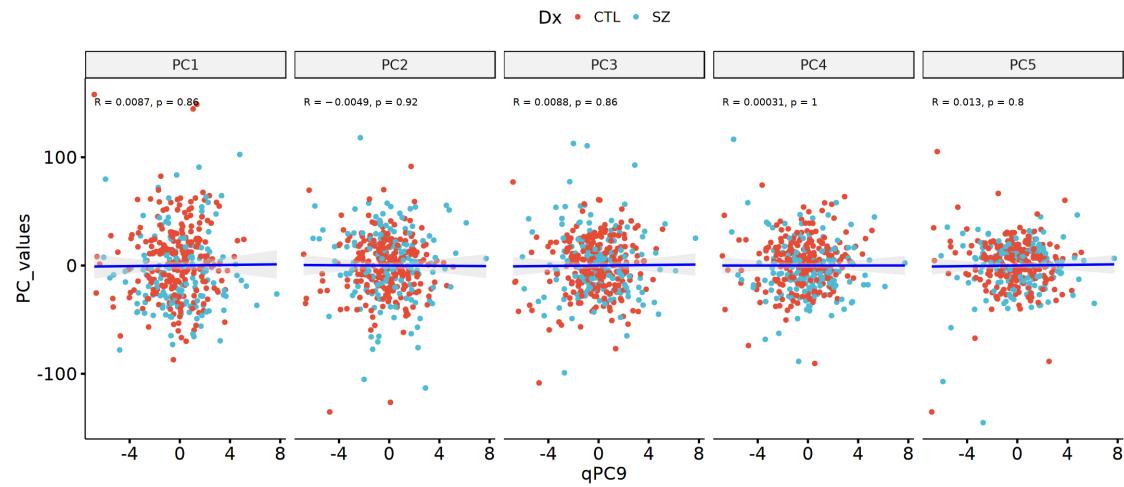
```
`geom_smooth()` using formula 'y ~ x'
```



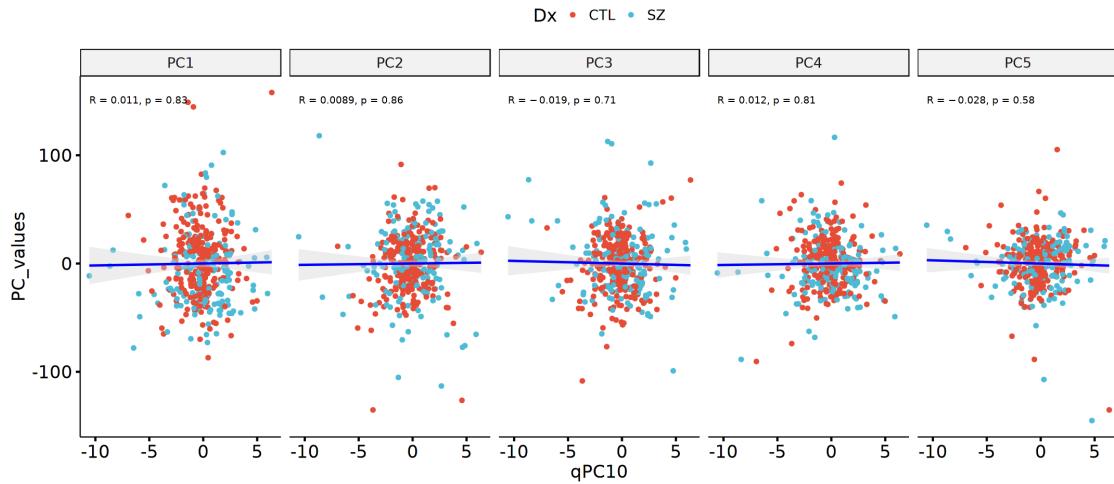
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



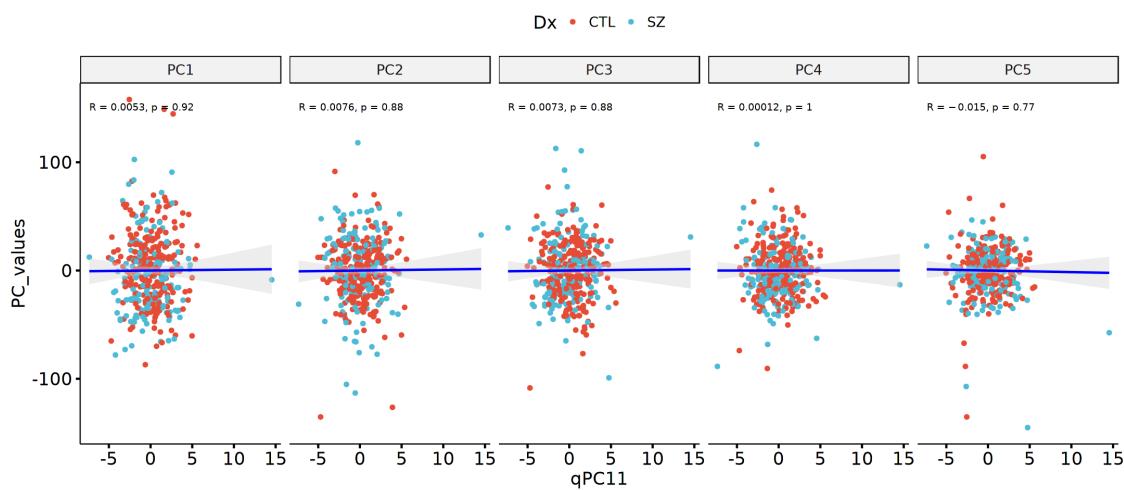
```

`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'

`geom_smooth()` using formula 'y ~ x'
  
```



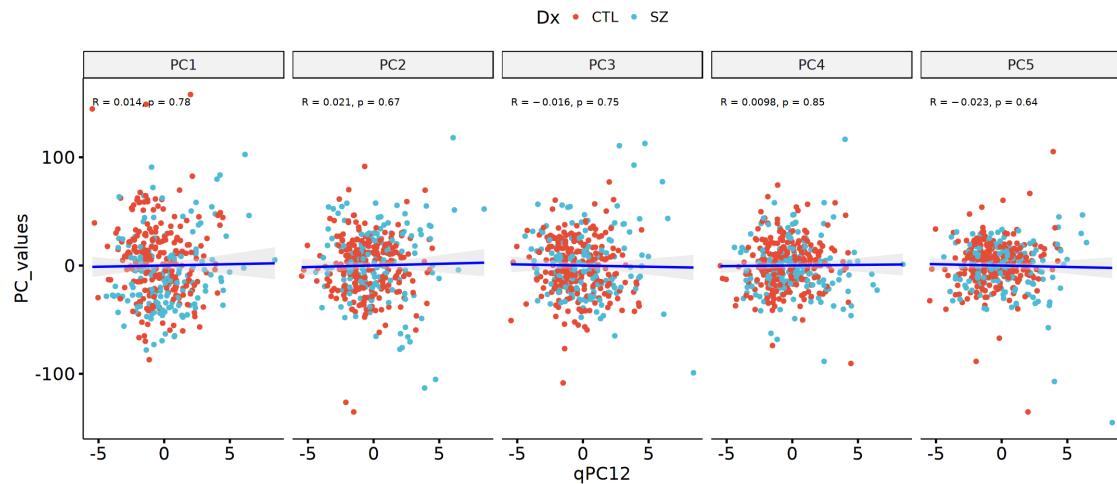
```
`geom_smooth()` using formula 'y ~ x'
```

```
[1] "There are: 22523 features left!"
```

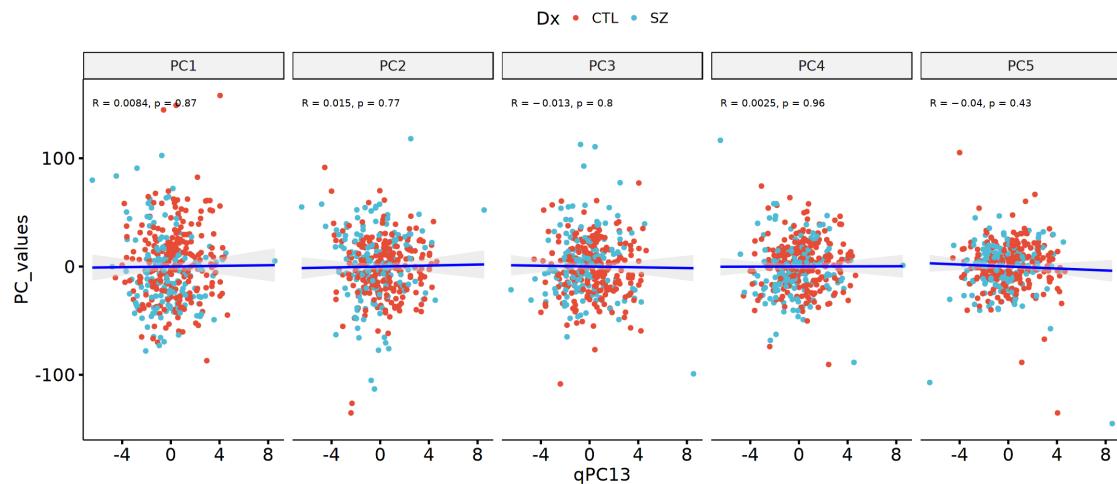
```
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```

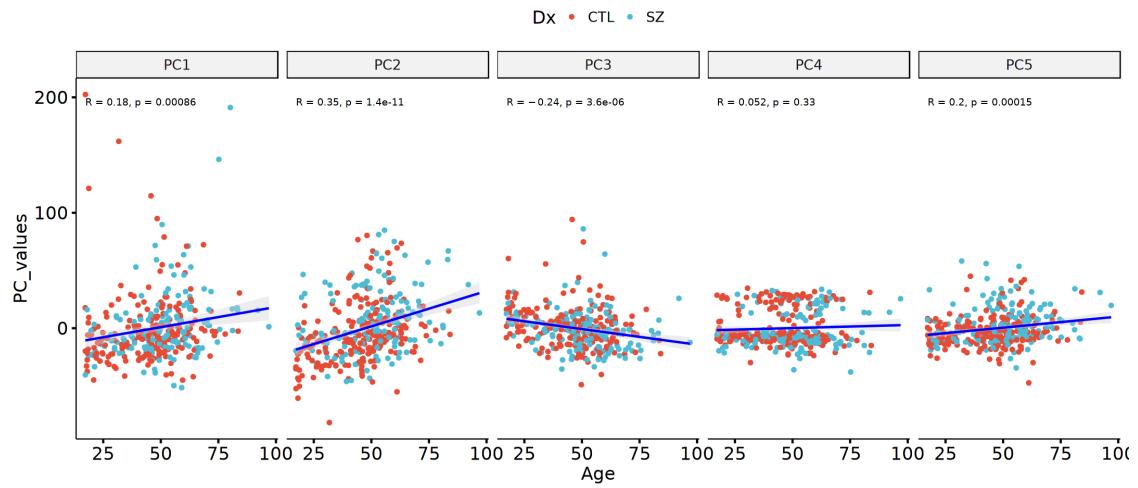
```
`geom_smooth()` using formula 'y ~ x'
```



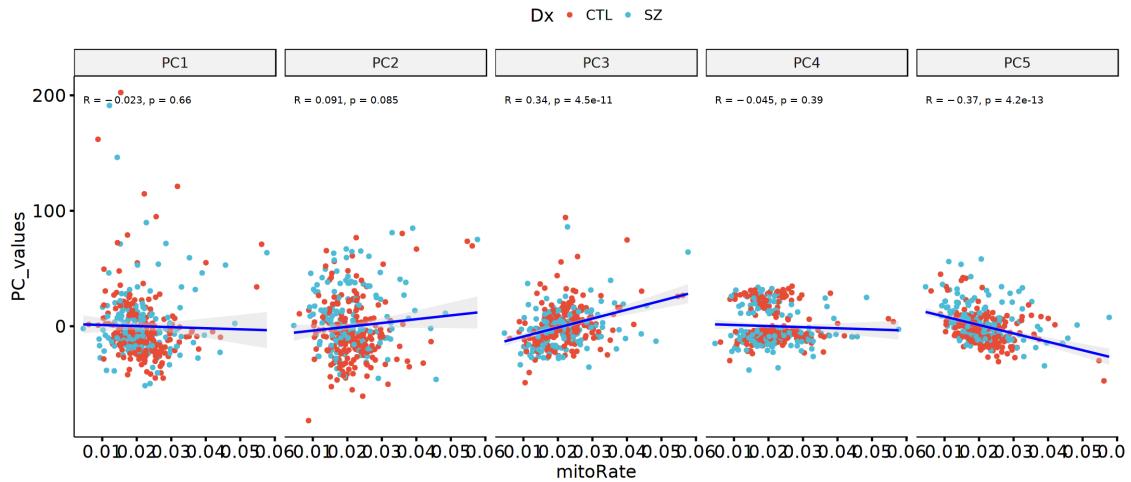
```
`geom_smooth()` using formula 'y ~ x'
```



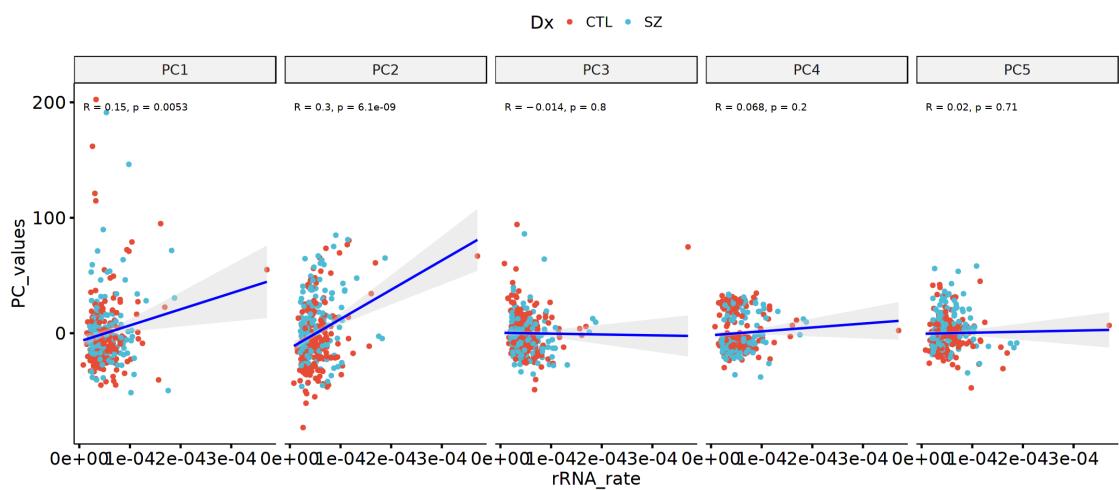
```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

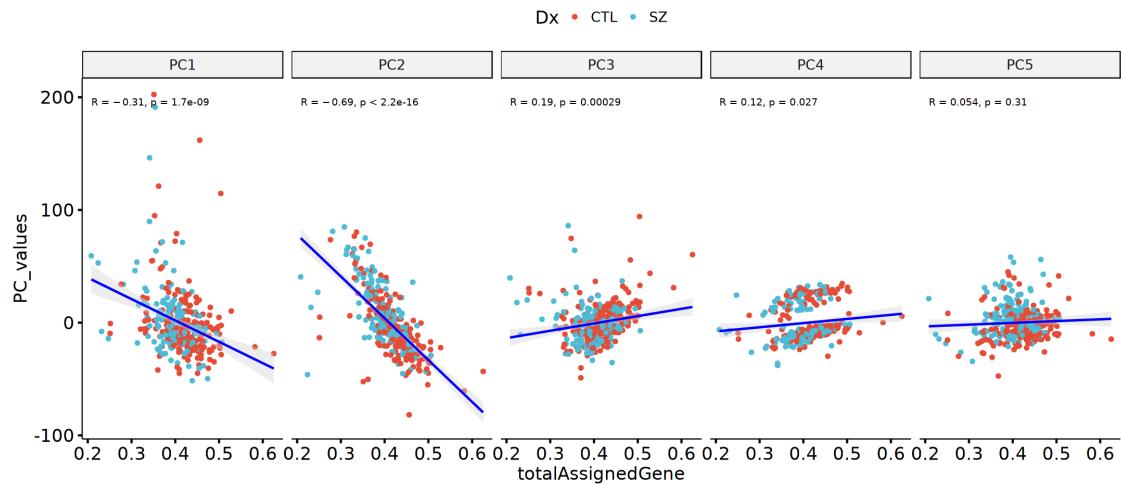


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

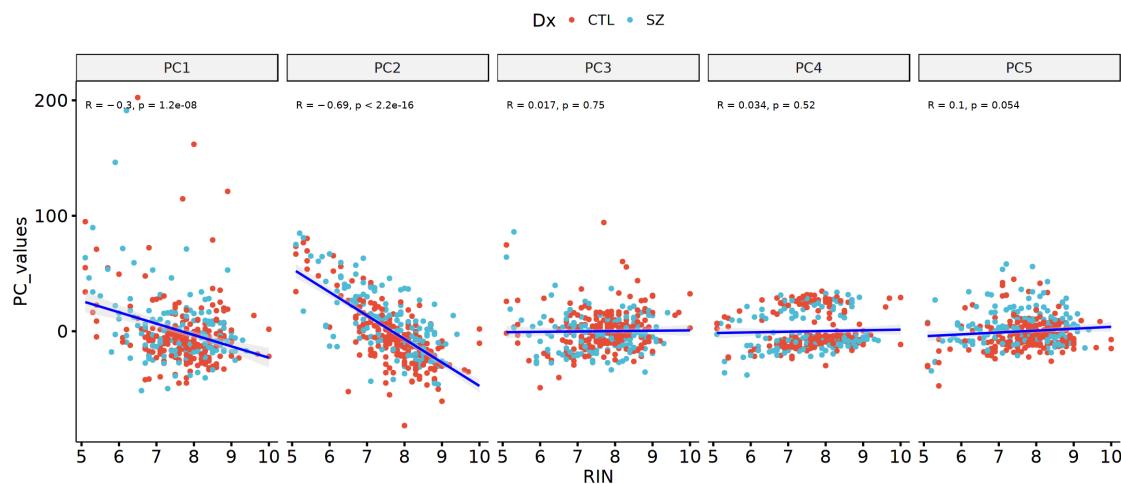


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

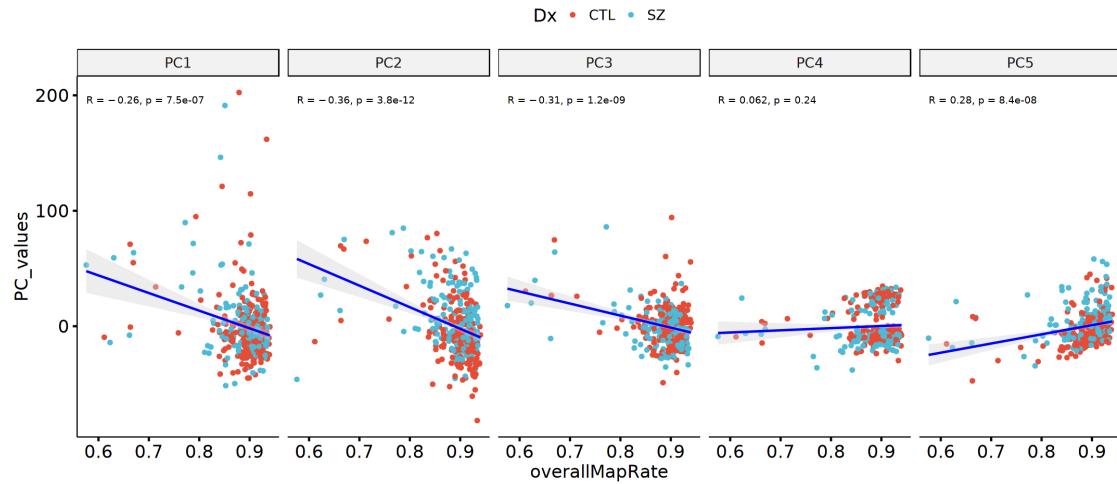
```
`geom_smooth()` using formula 'y ~ x'
```



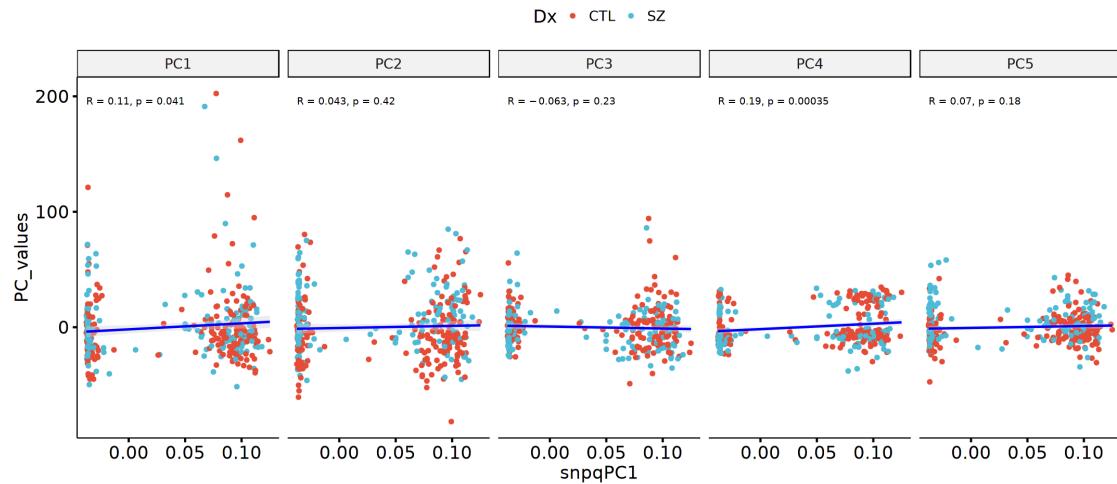
```
`geom_smooth()` using formula 'y ~ x'
```



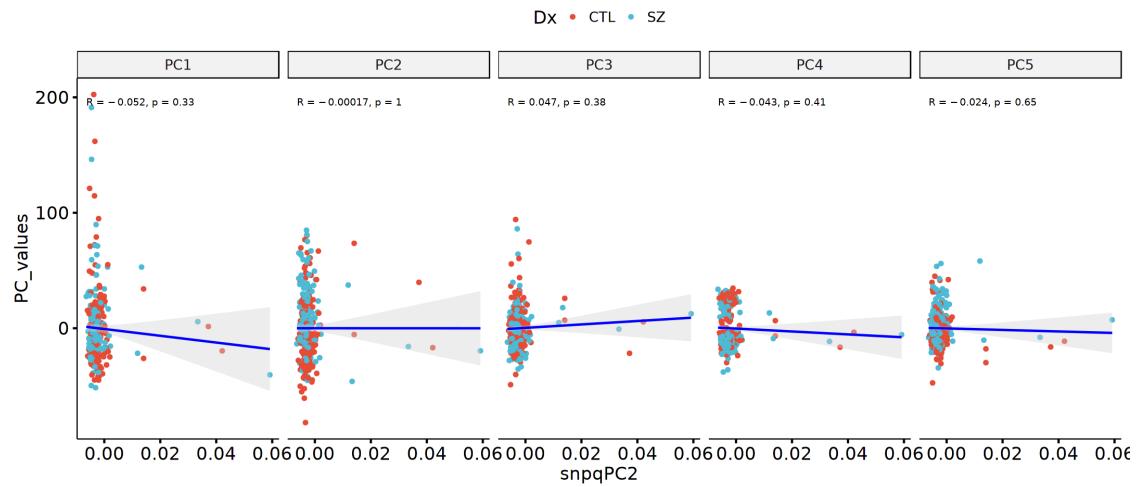
```
`geom_smooth()` using formula 'y ~ x'
```



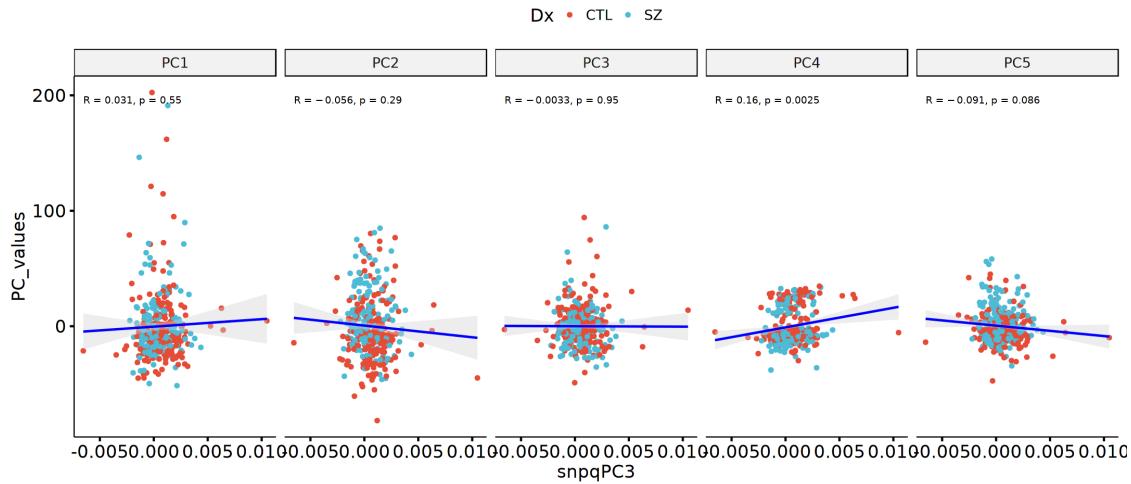
```
`geom_smooth()` using formula 'y ~ x'
```



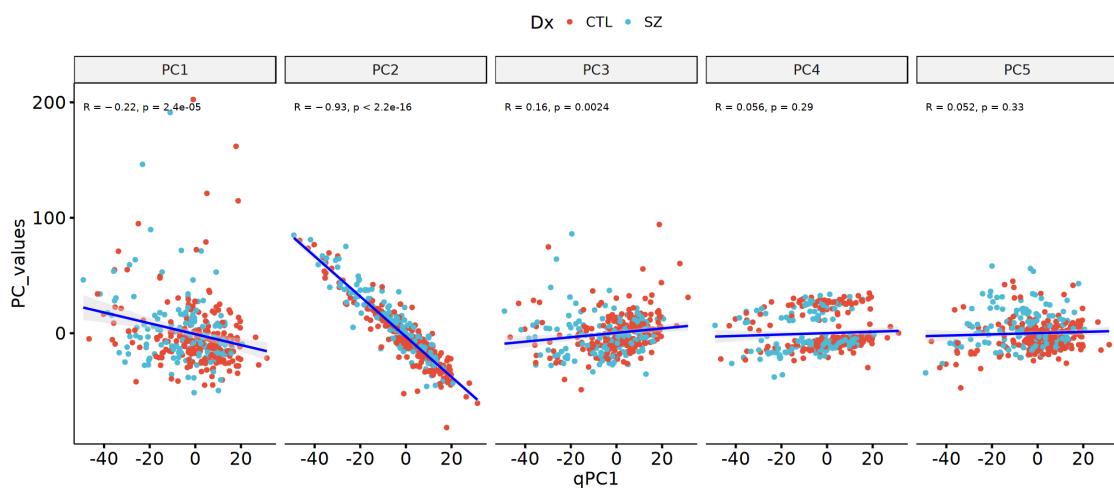
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

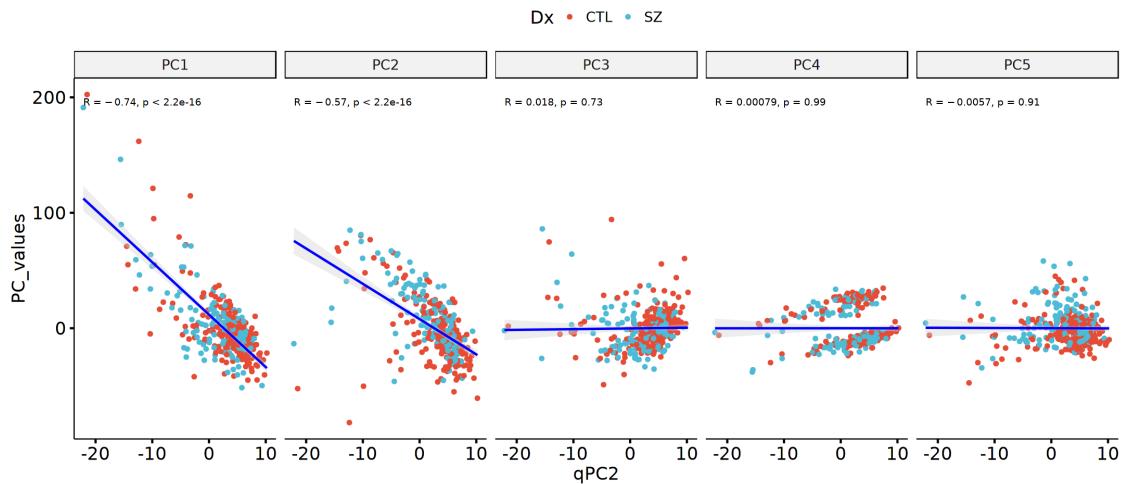


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

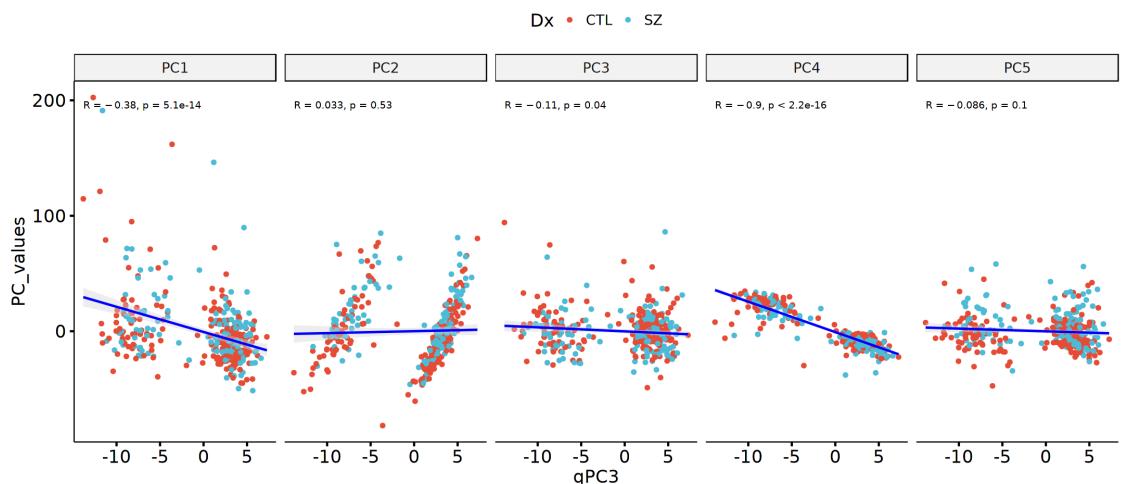


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

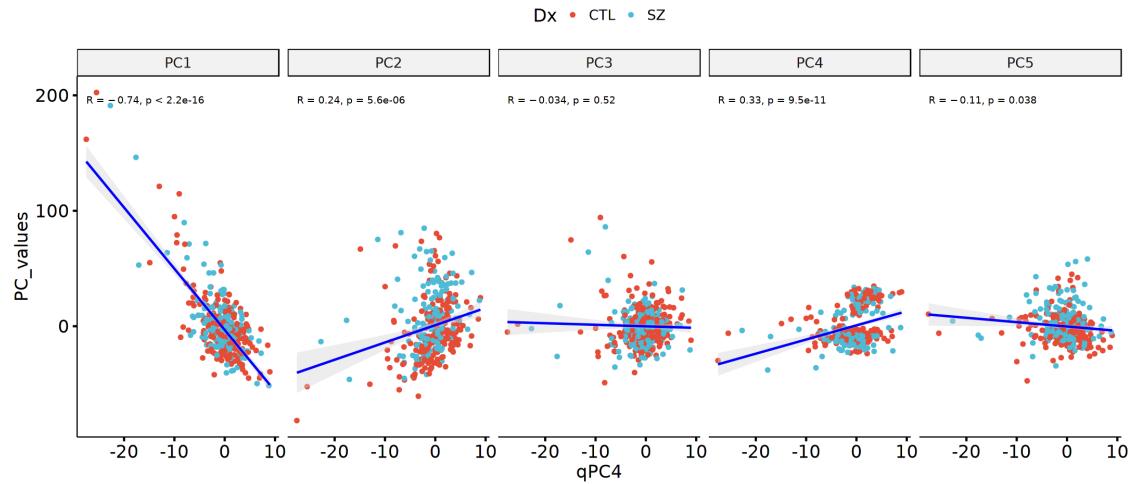
```
`geom_smooth()` using formula 'y ~ x'
```



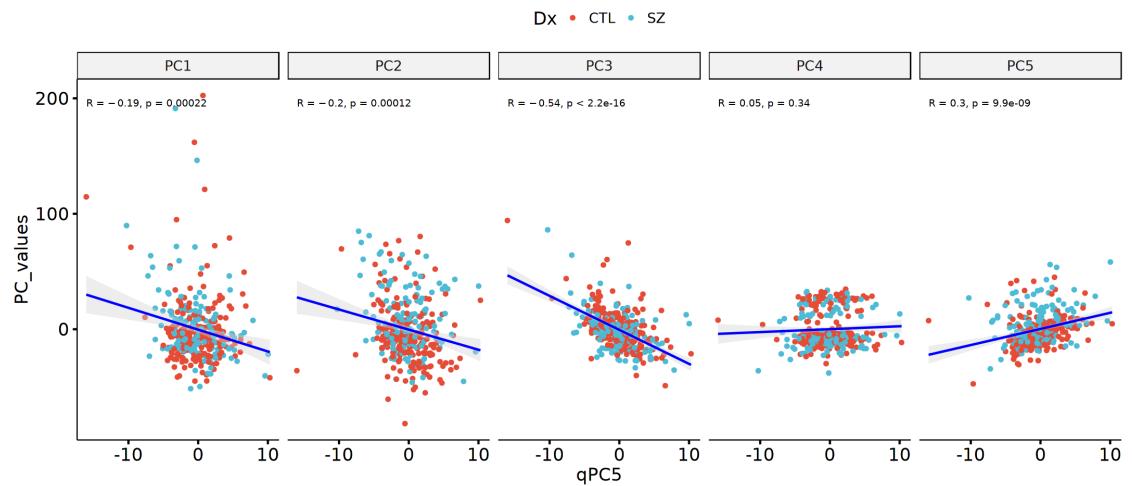
```
`geom_smooth()` using formula 'y ~ x'
```



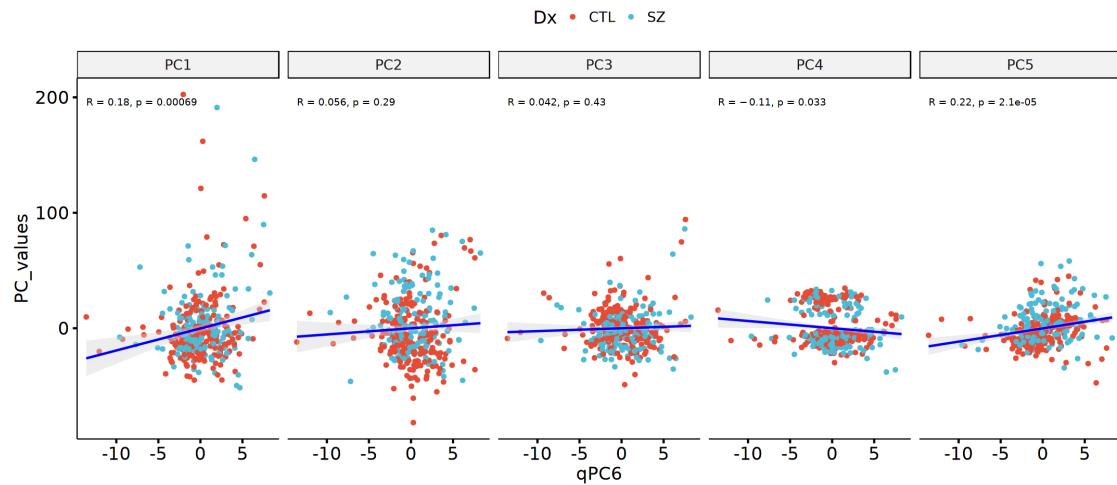
```
`geom_smooth()` using formula 'y ~ x'
```



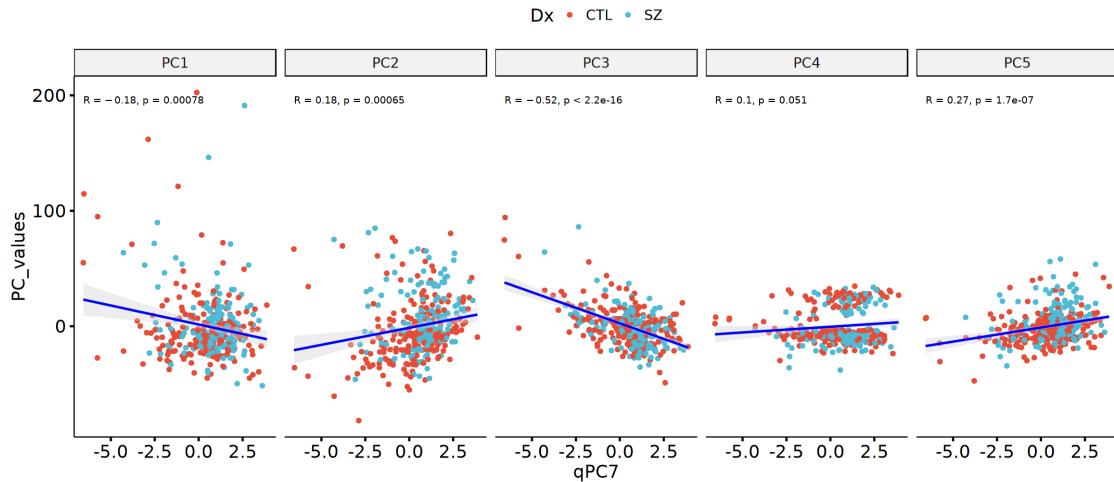
```
`geom_smooth()` using formula 'y ~ x'
```



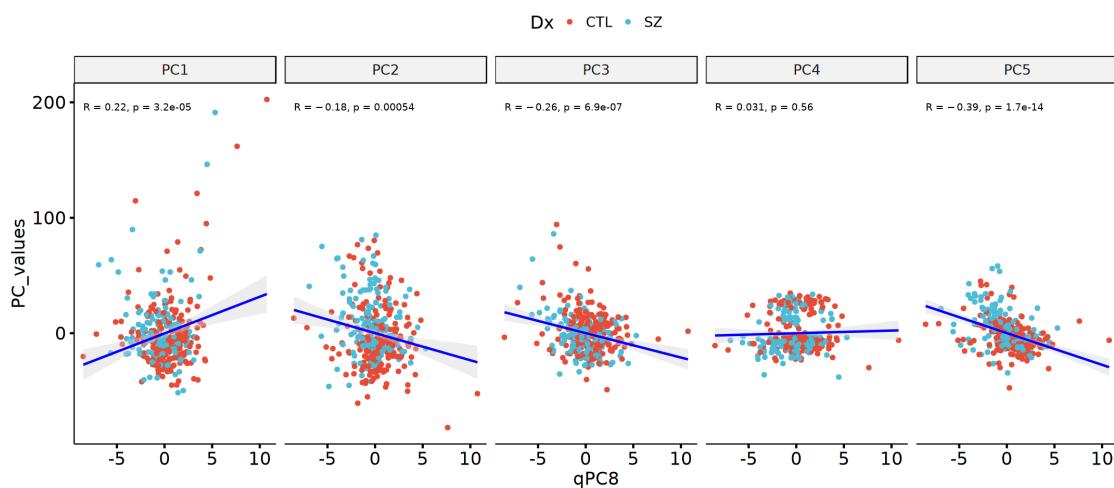
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

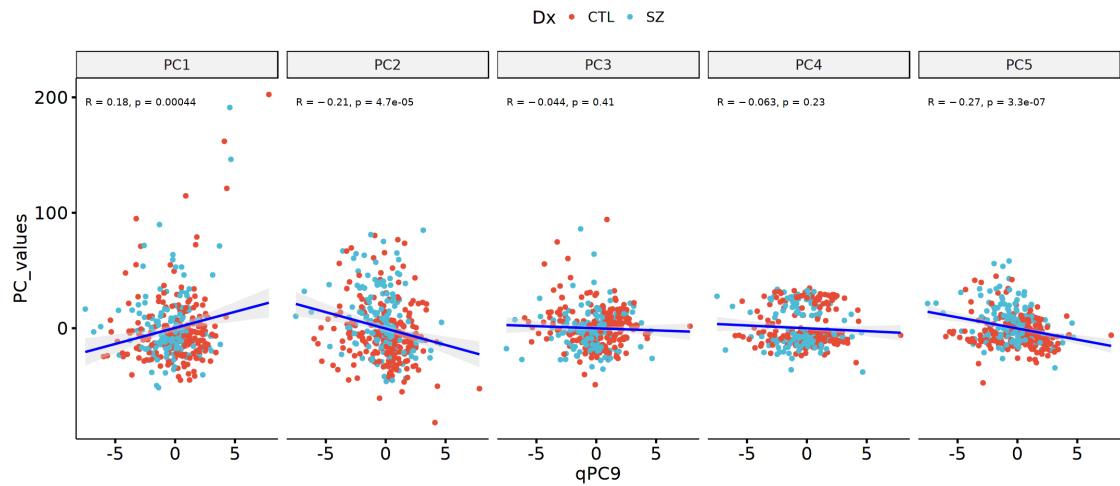


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

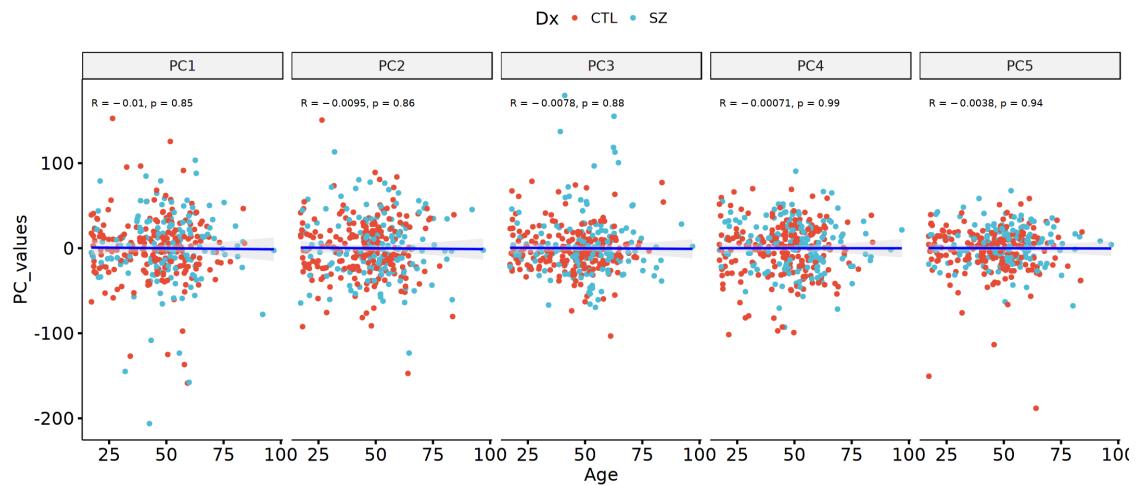


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

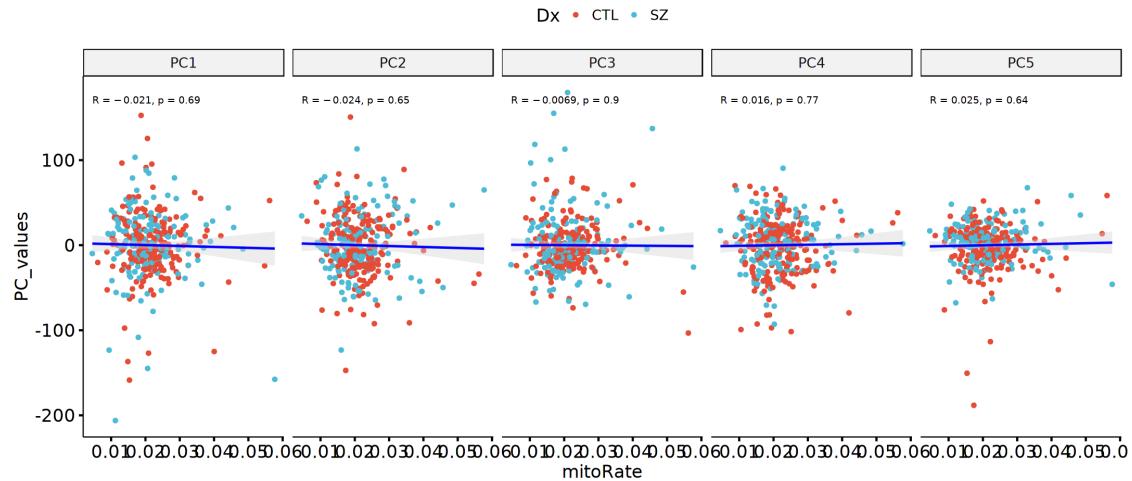
```
`geom_smooth()` using formula 'y ~ x'
```



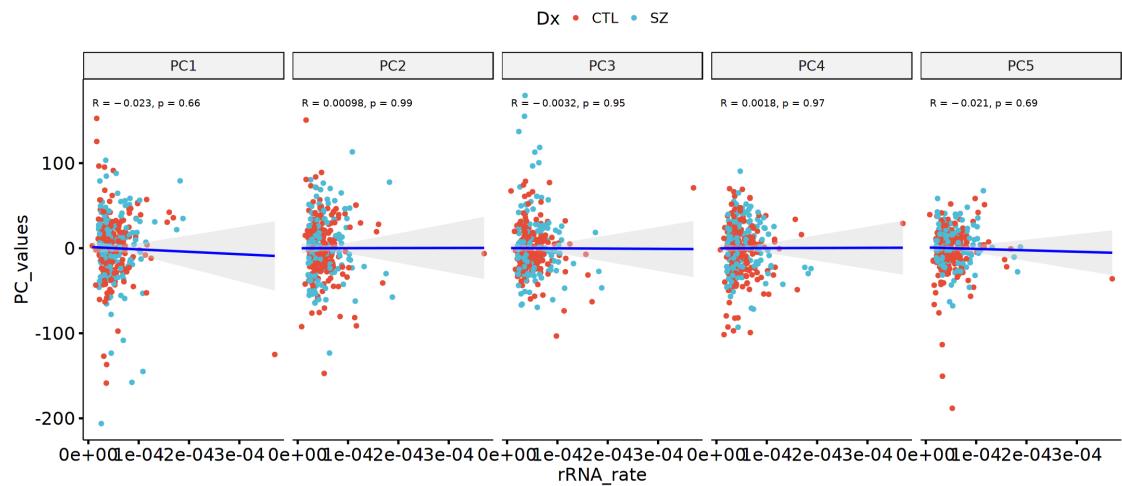
```
`geom_smooth()` using formula 'y ~ x'
```



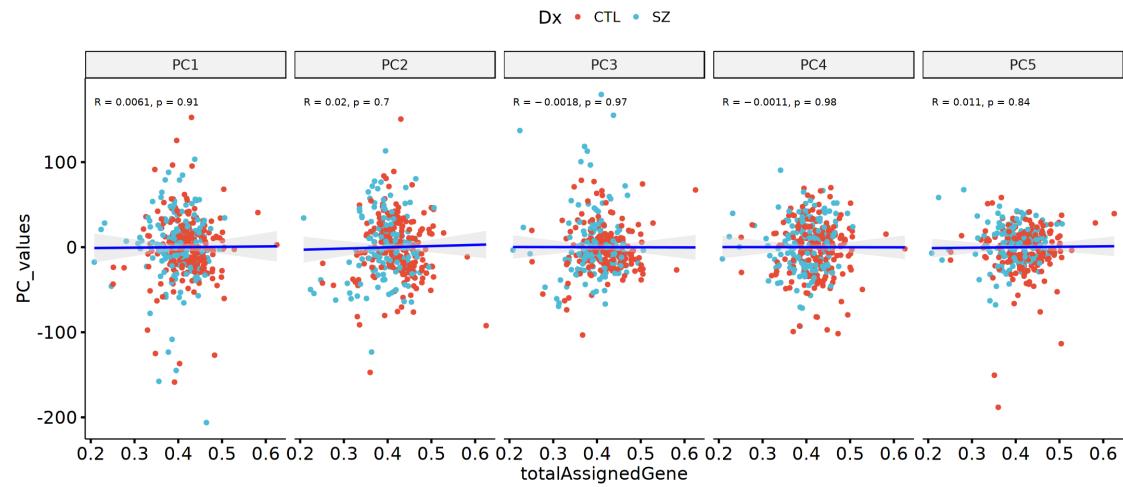
```
`geom_smooth()` using formula 'y ~ x'
```



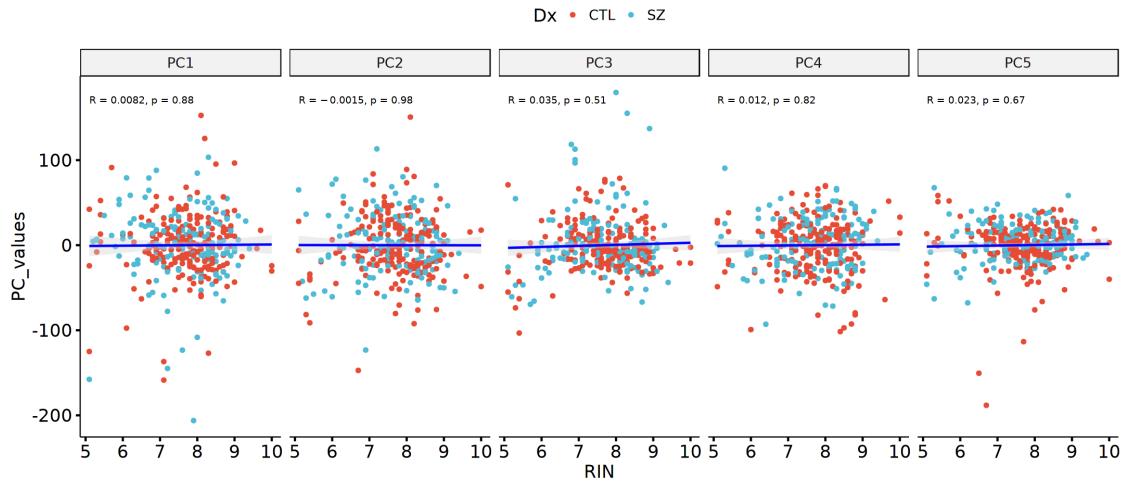
```
`geom_smooth()` using formula 'y ~ x'
```



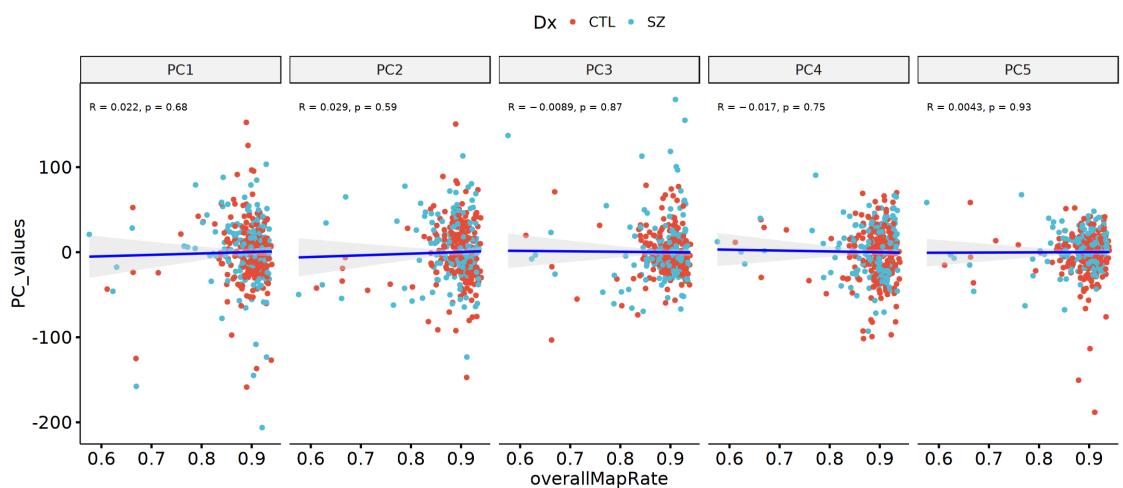
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

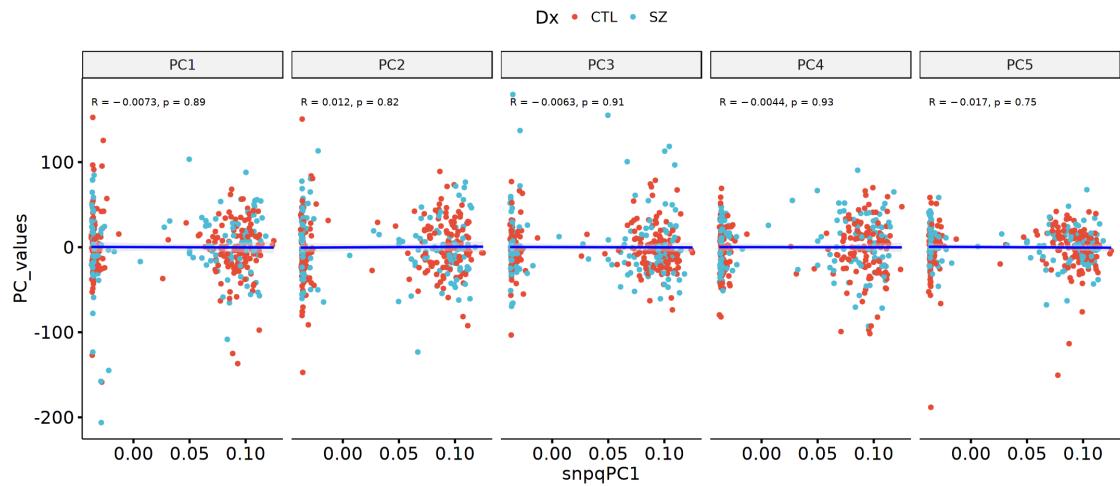


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

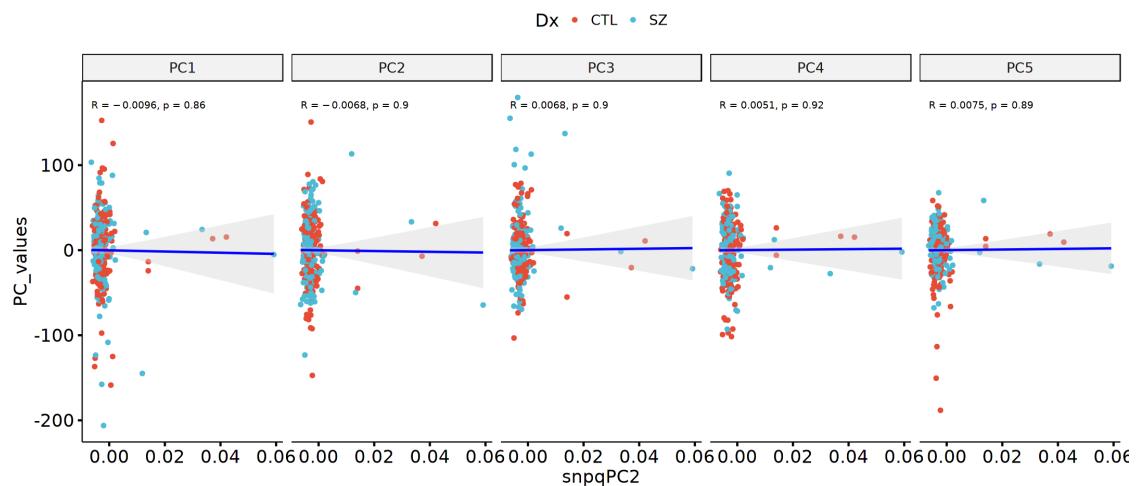


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

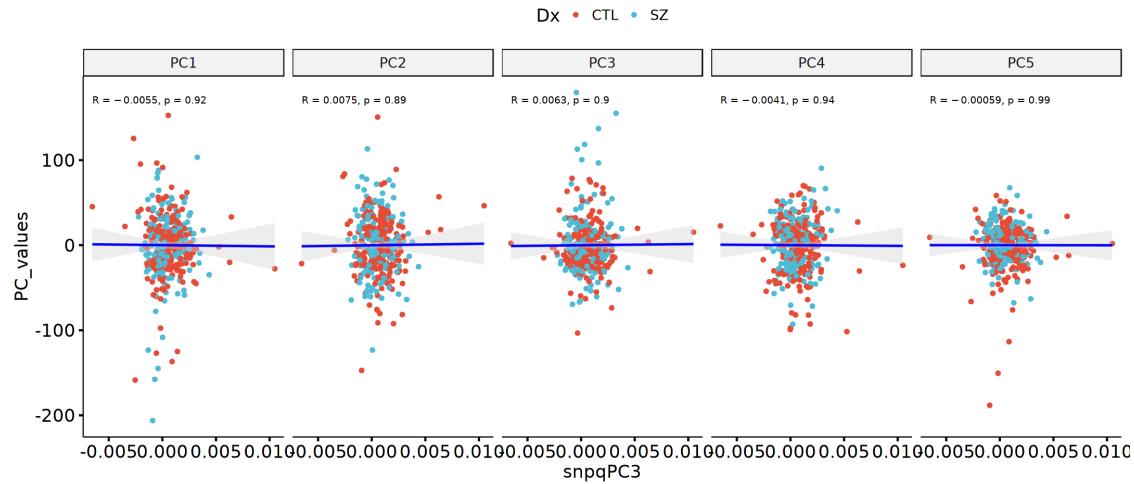
```
`geom_smooth()` using formula 'y ~ x'
```



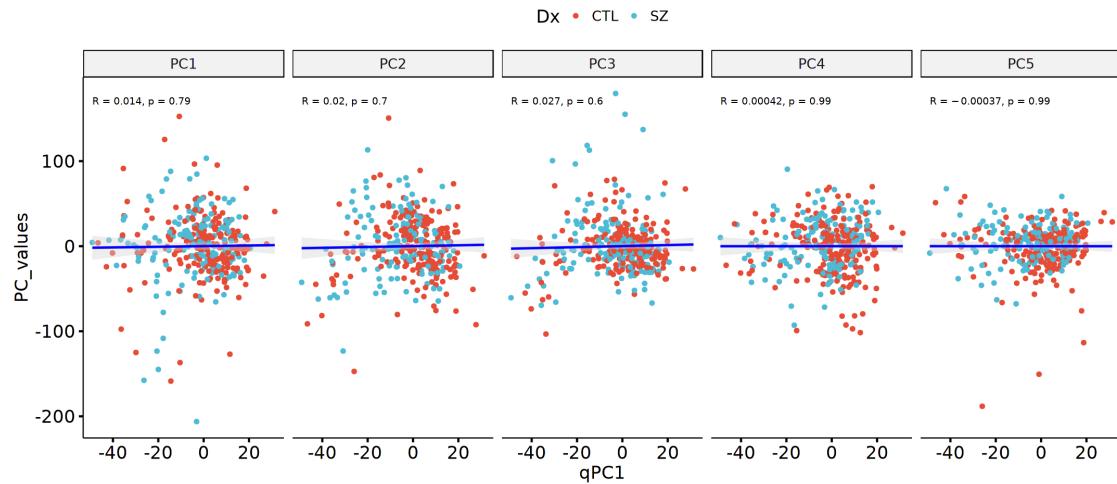
```
`geom_smooth()` using formula 'y ~ x'
```



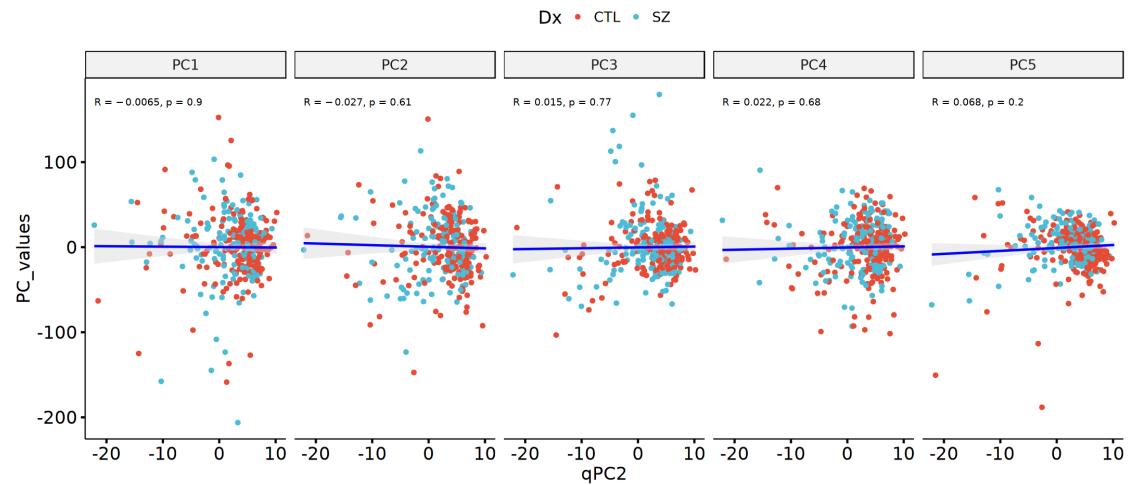
```
`geom_smooth()` using formula 'y ~ x'
```



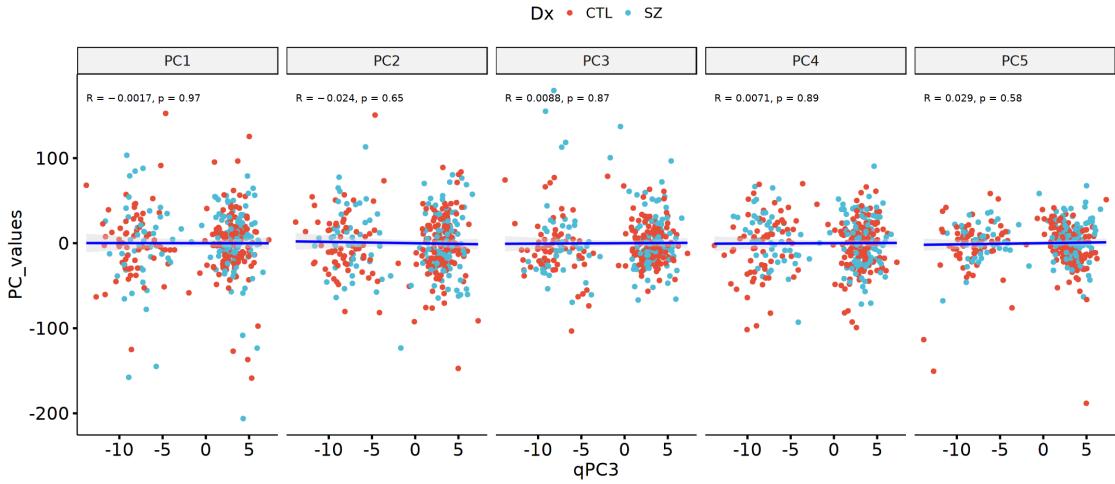
```
`geom_smooth()` using formula 'y ~ x'
```



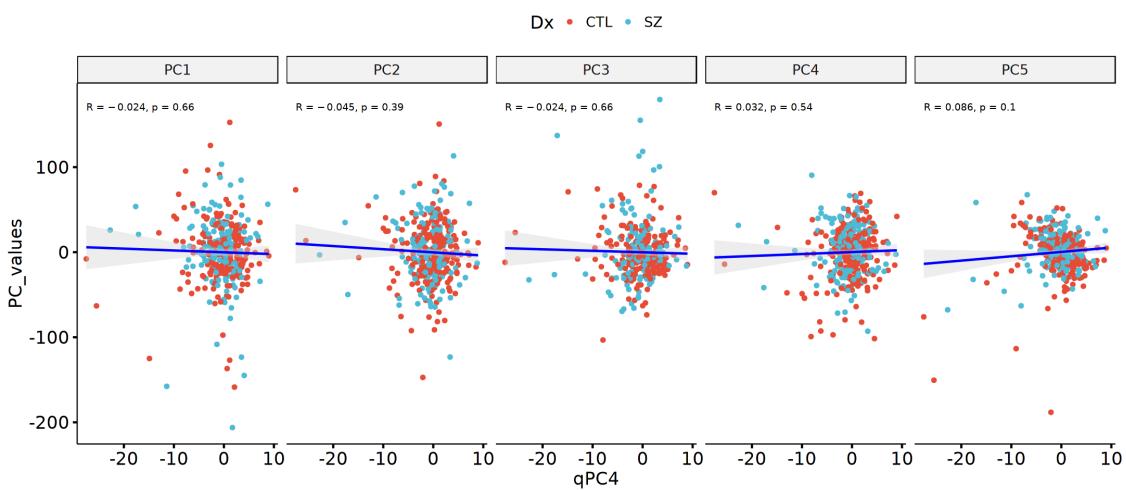
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

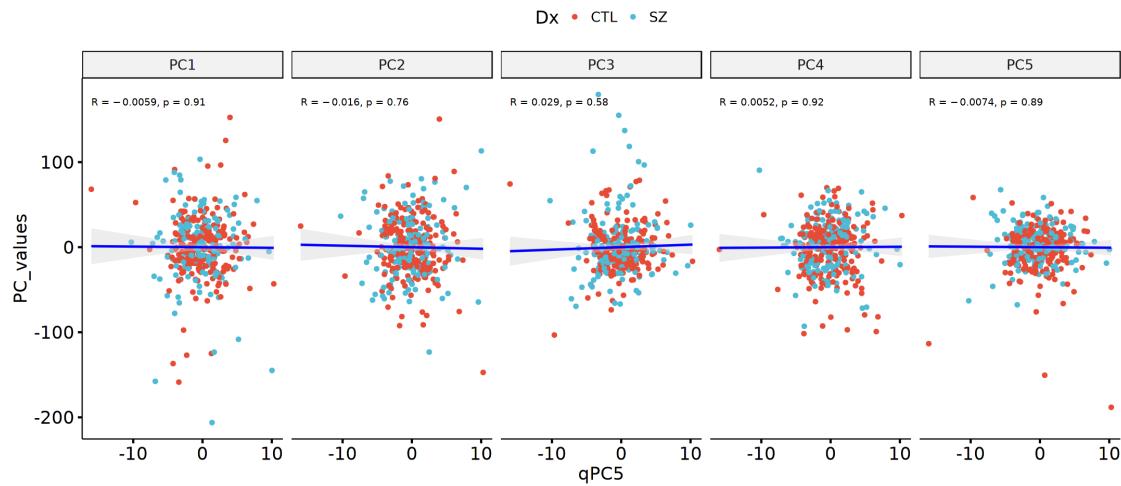


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

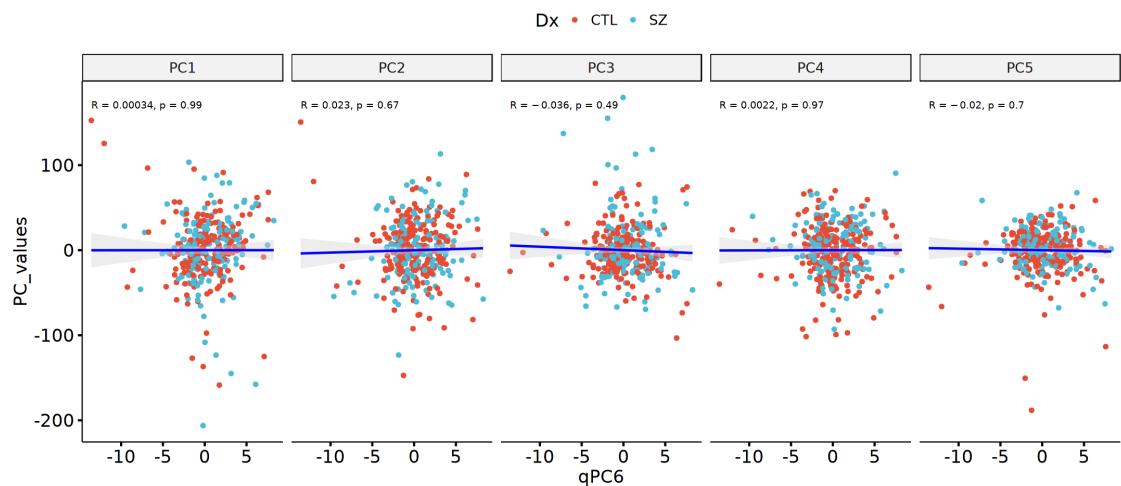


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'
```

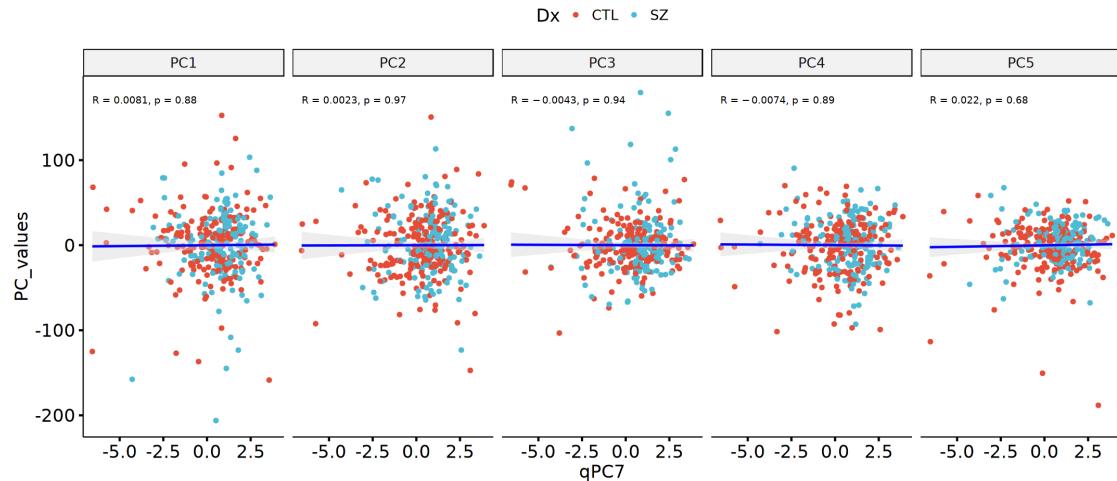


```
`geom_smooth()` using formula 'y ~ x'
```

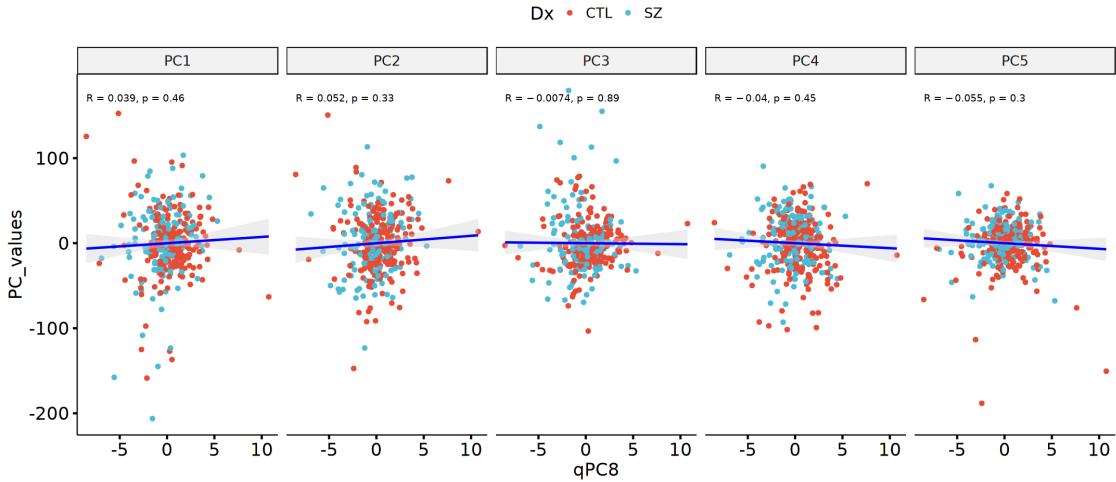


```
`geom_smooth()` using formula 'y ~ x'
```

```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'  
[1] "There are: 22769 features left!"  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

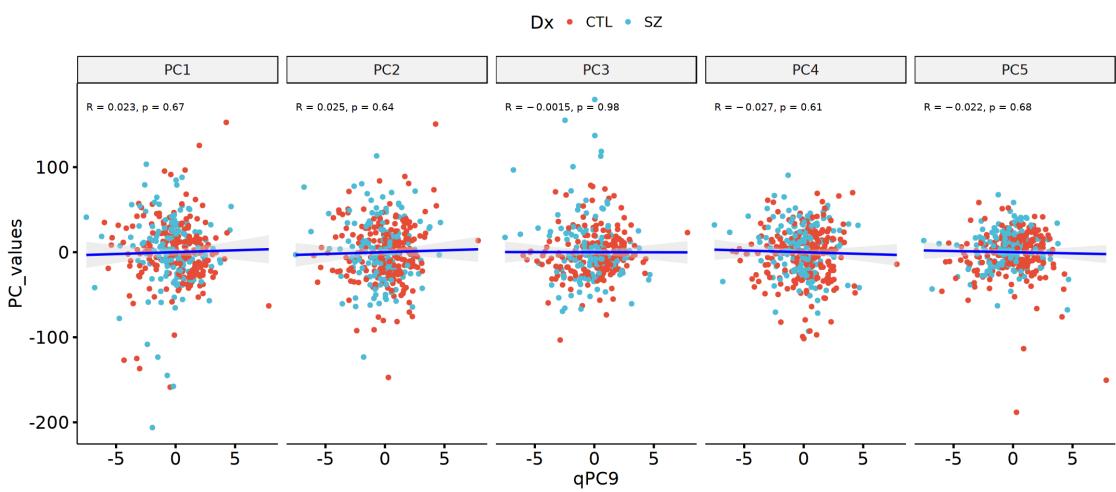


```
`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'
```

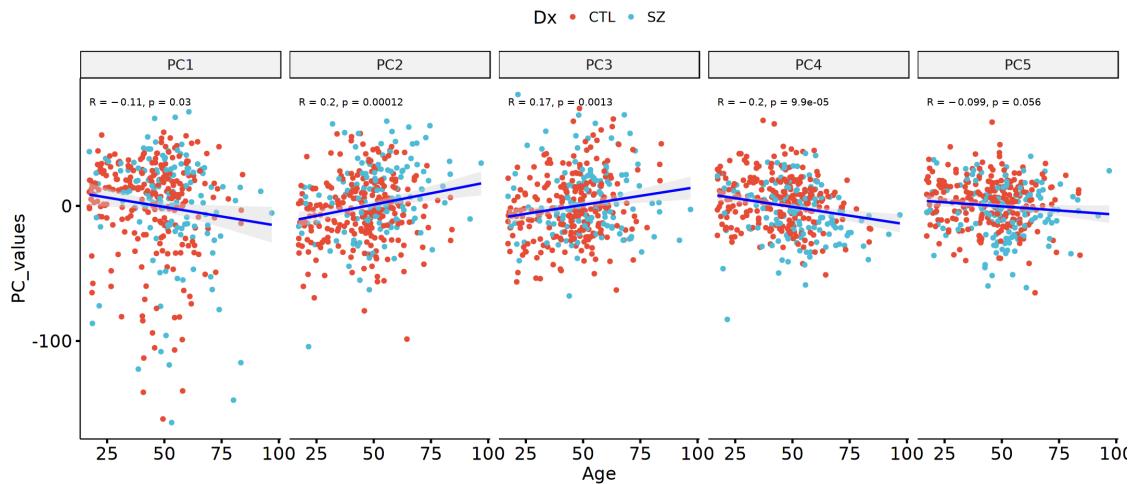


```
`geom_smooth()` using formula 'y ~ x'  

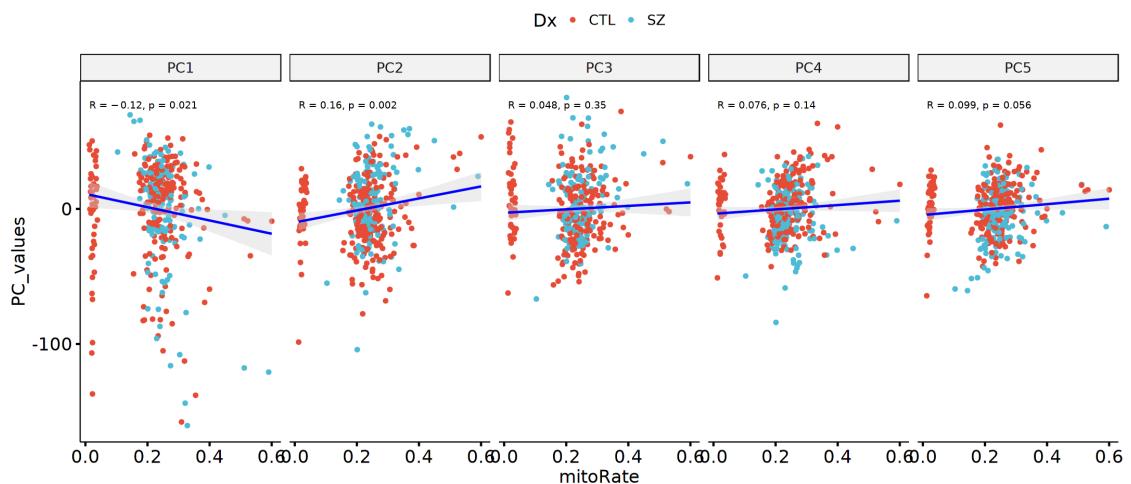
`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'
```

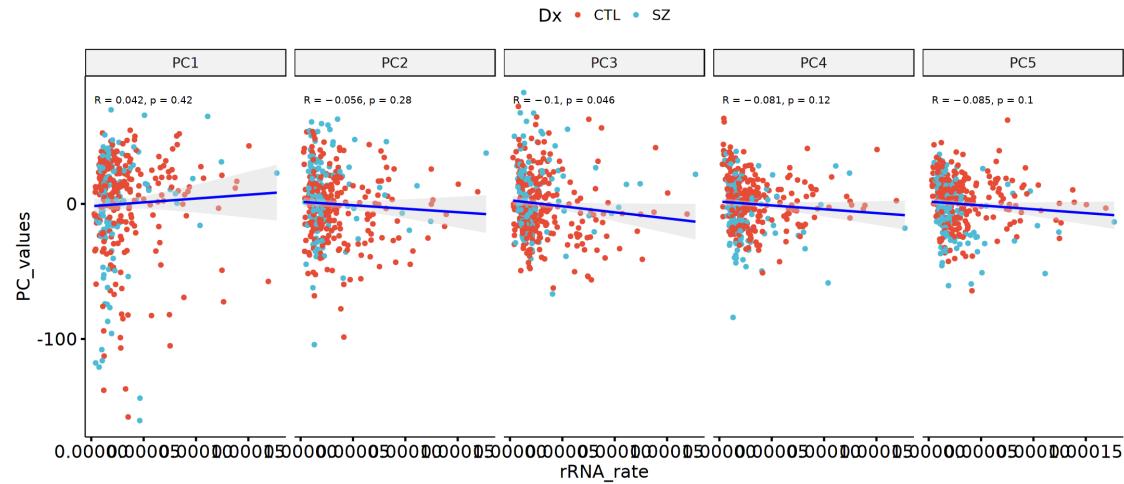
```
`geom_smooth()` using formula 'y ~ x'
```



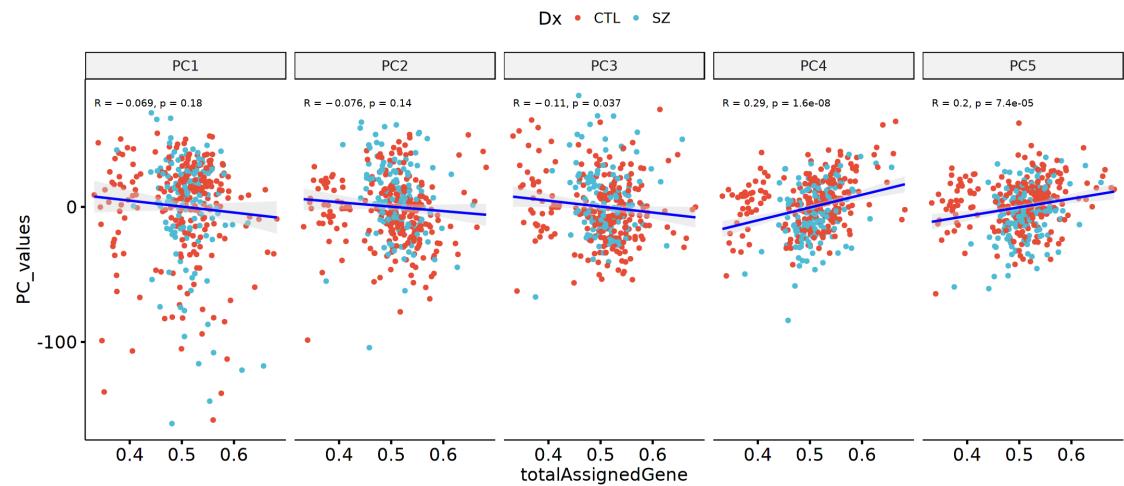
```
`geom_smooth()` using formula 'y ~ x'
```



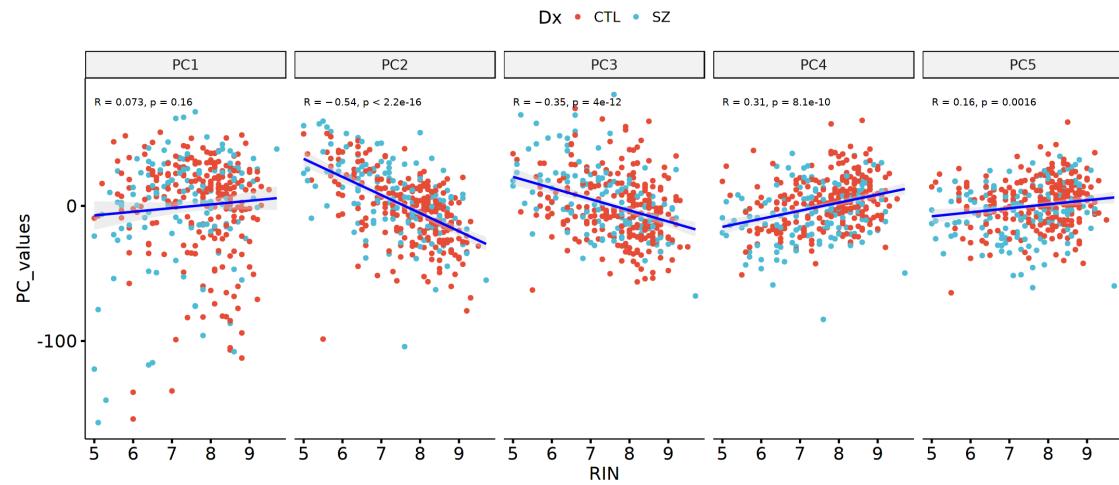
```
`geom_smooth()` using formula 'y ~ x'
```



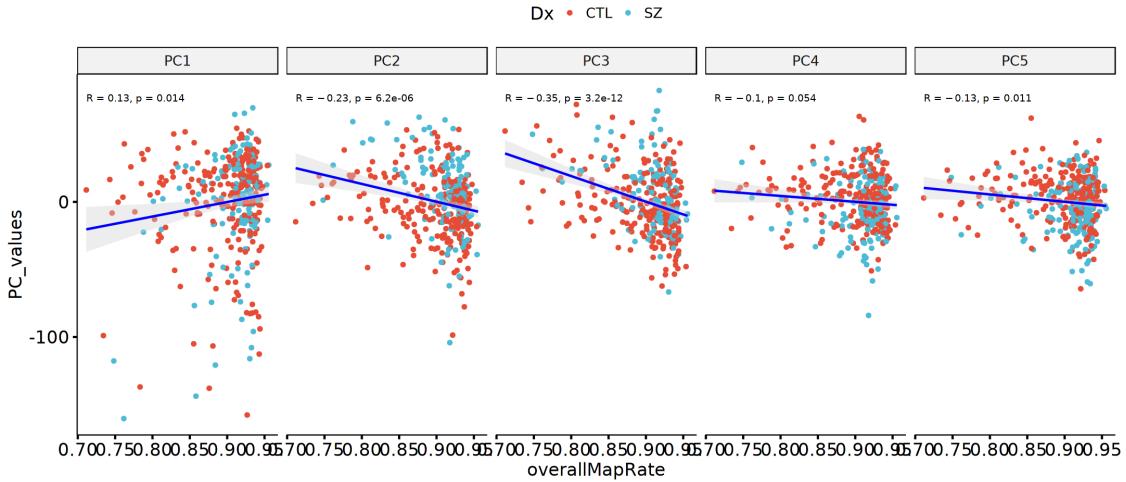
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

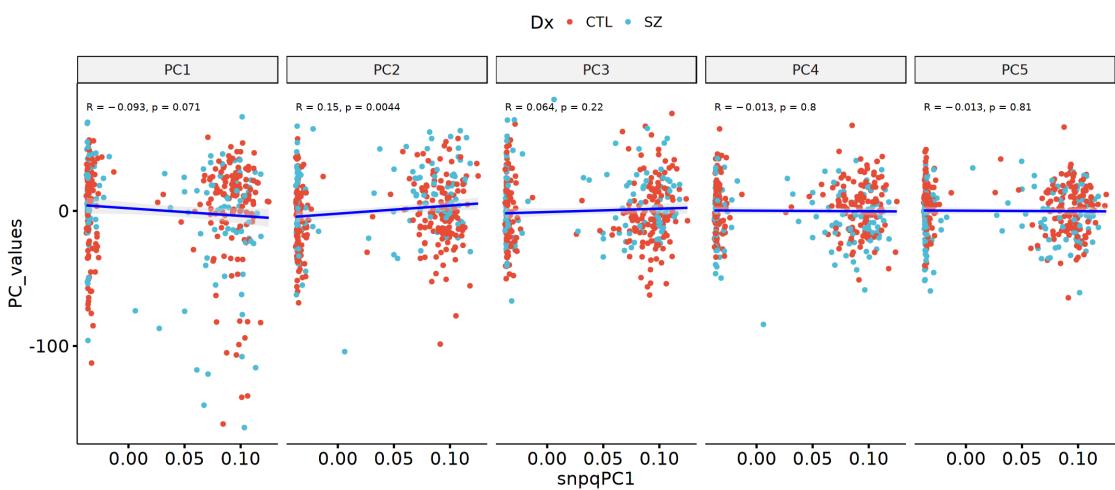


```
`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'
```

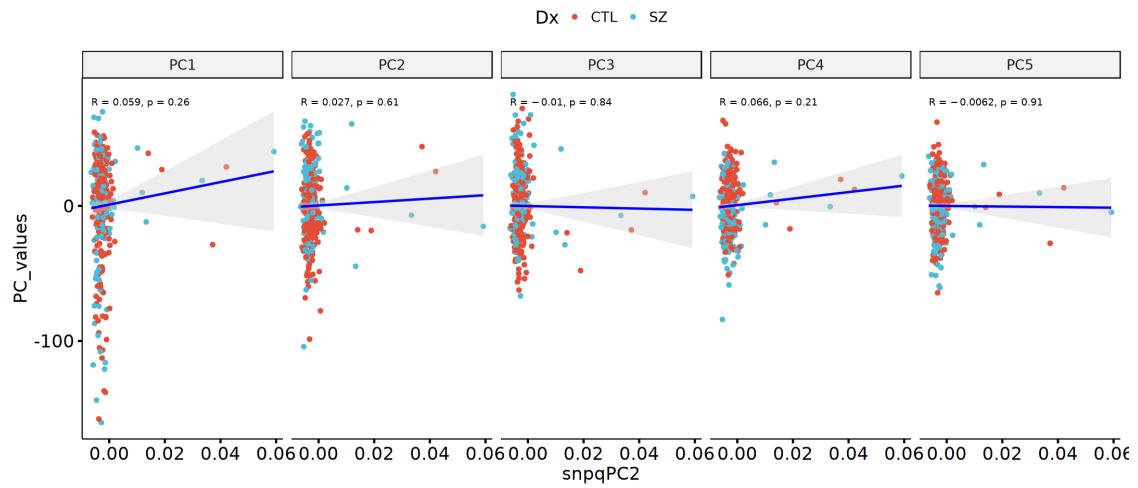


```
`geom_smooth()` using formula 'y ~ x'  

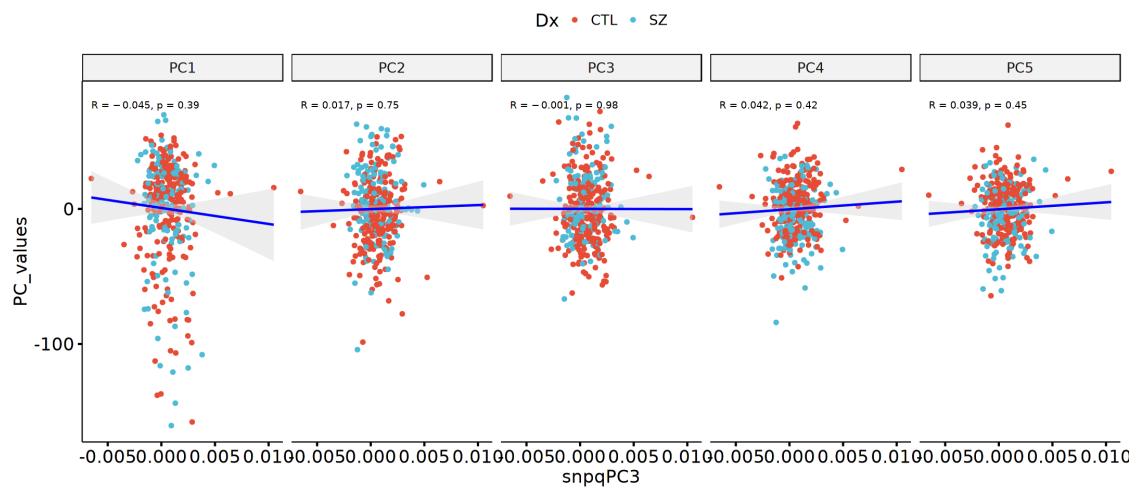
`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'
```

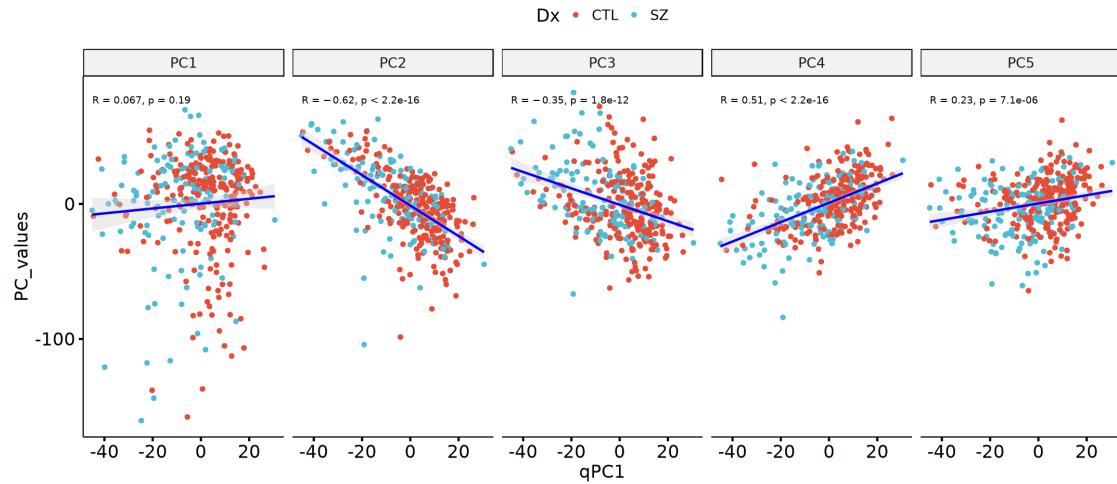
```
`geom_smooth()` using formula 'y ~ x'
```



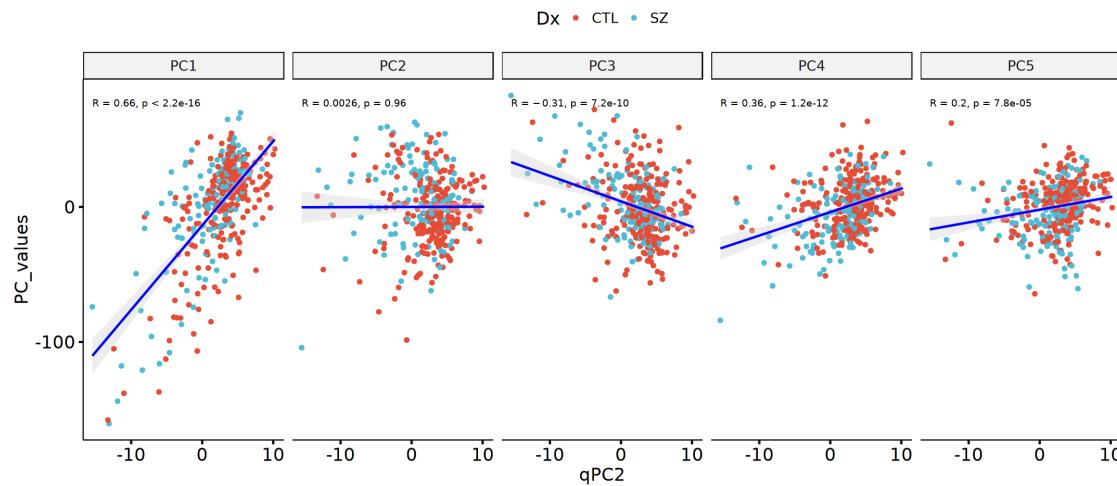
```
`geom_smooth()` using formula 'y ~ x'
```



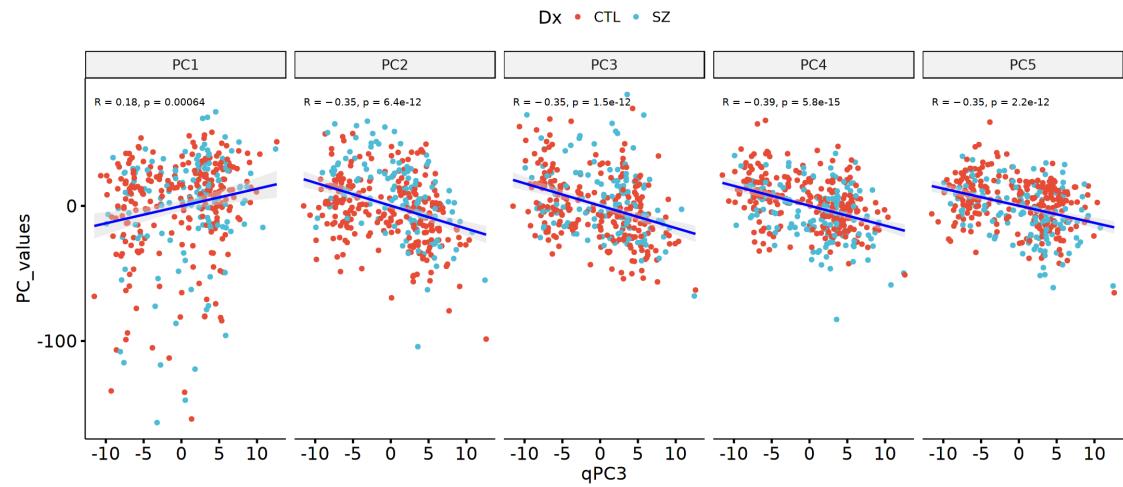
```
`geom_smooth()` using formula 'y ~ x'
```



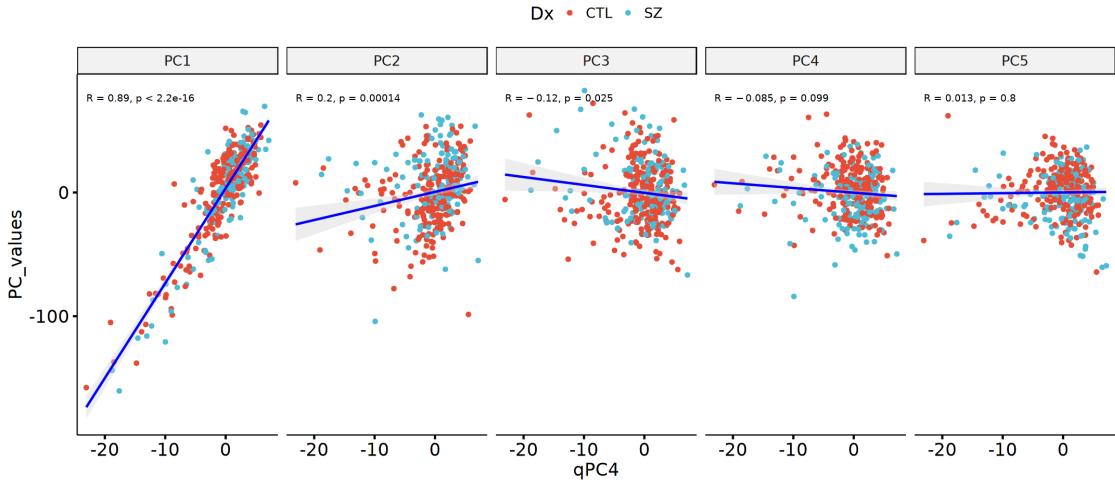
```
`geom_smooth()` using formula 'y ~ x'
```



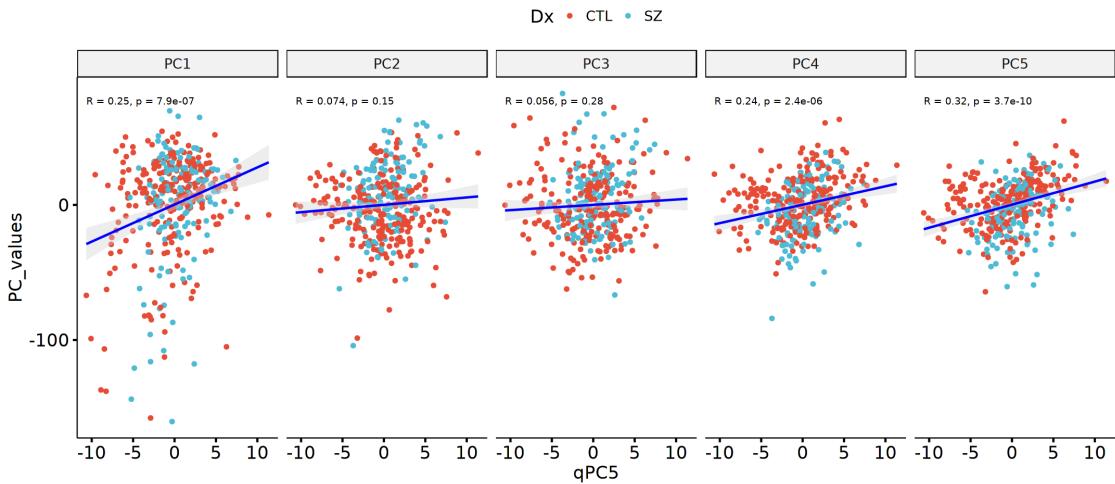
```
`geom_smooth()` using formula 'y ~ x'
```



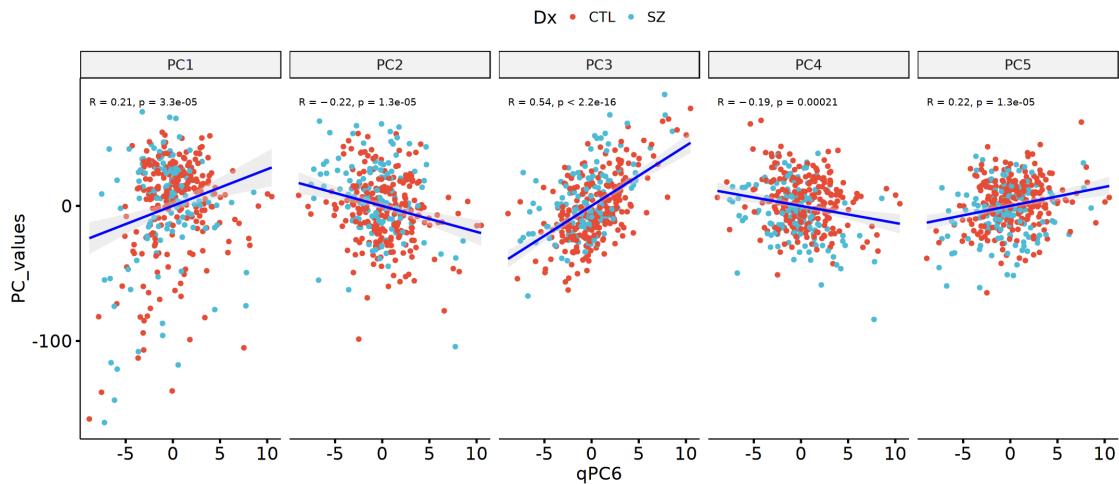
```
`geom_smooth()` using formula 'y ~ x'
```



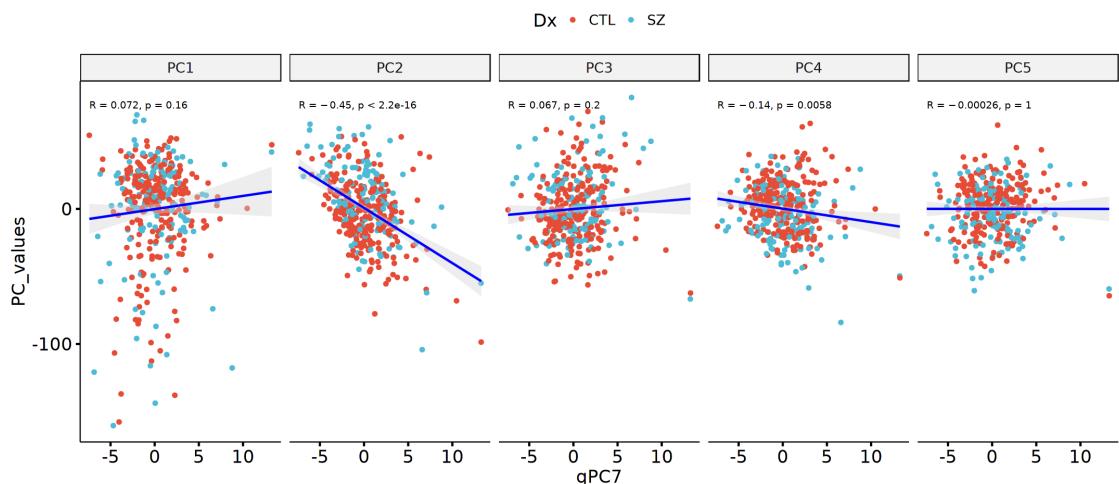
```
`geom_smooth()` using formula 'y ~ x'
```



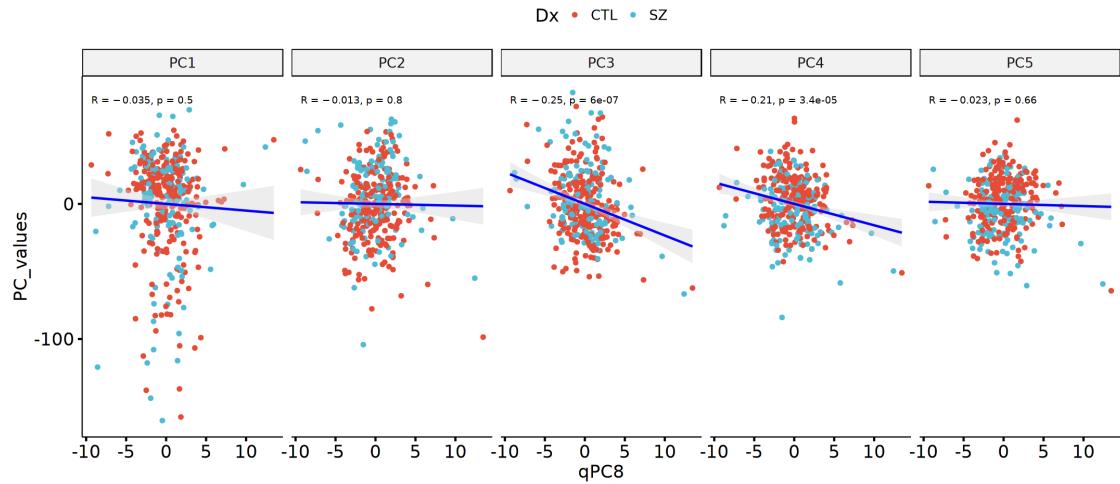
```
`geom_smooth()` using formula 'y ~ x'
```



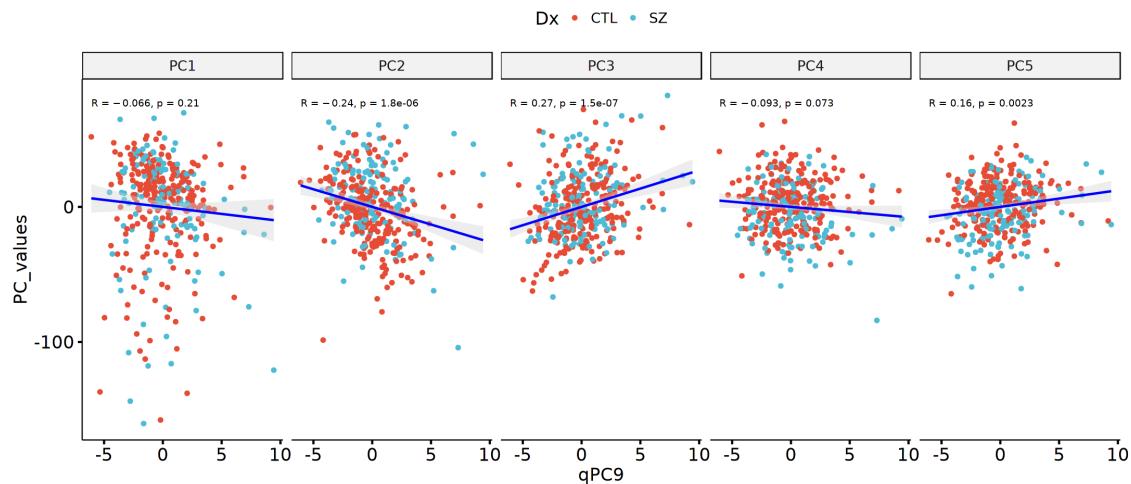
```
`geom_smooth()` using formula 'y ~ x'
```



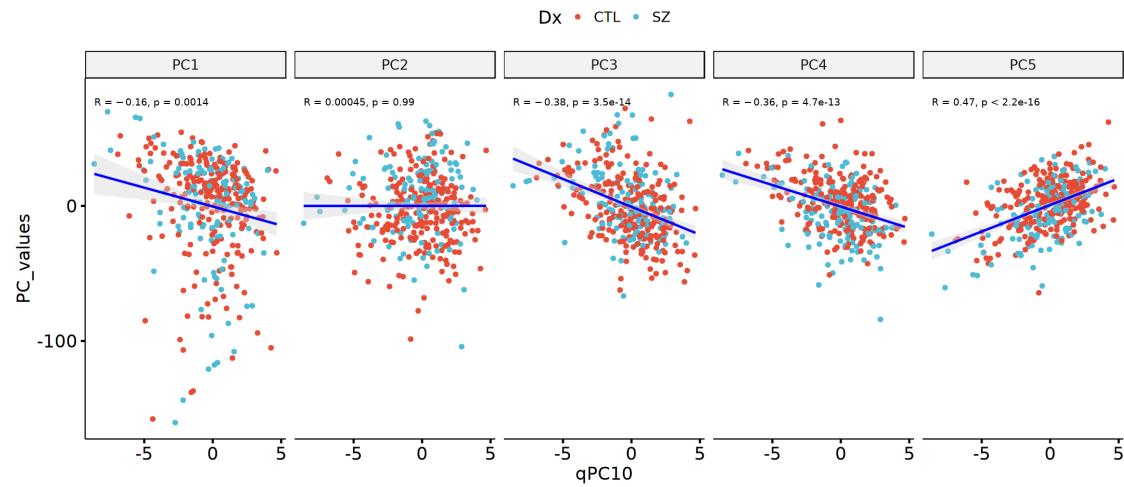
```
`geom_smooth()` using formula 'y ~ x'
```



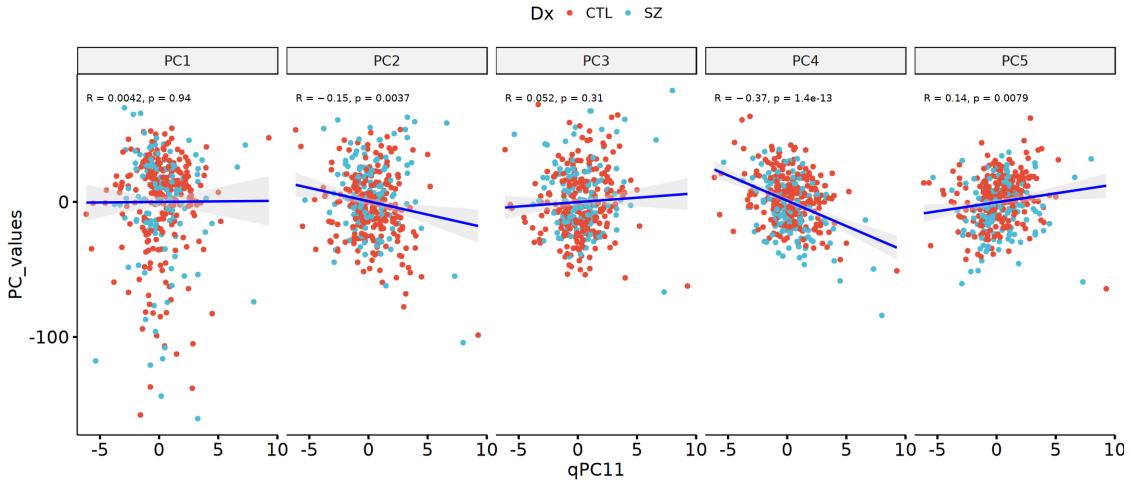
```
`geom_smooth()` using formula 'y ~ x'
```



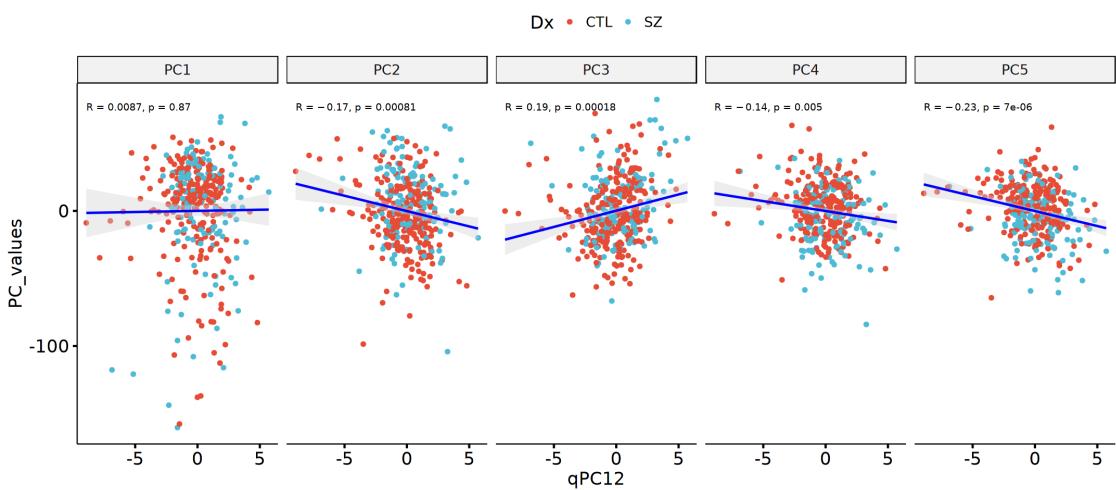
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

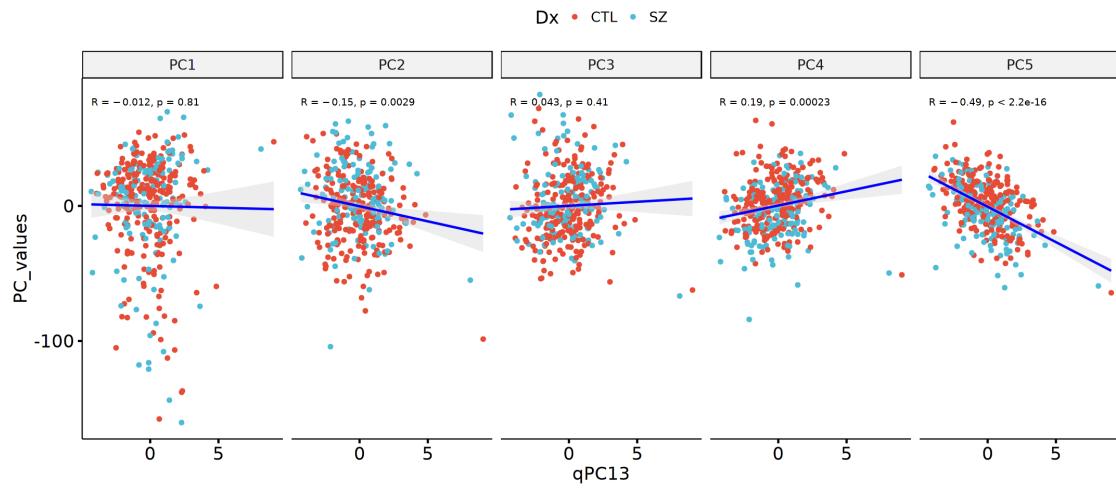


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

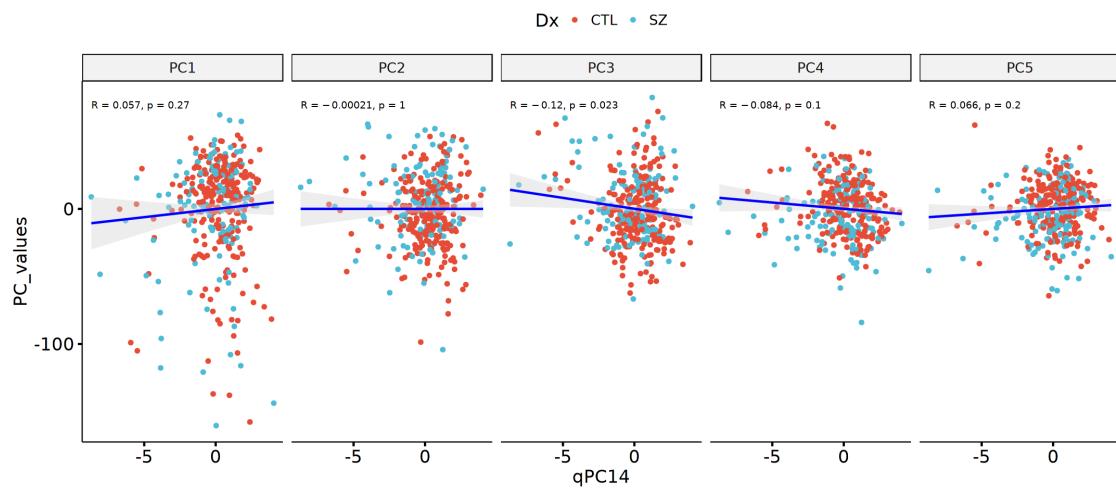


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

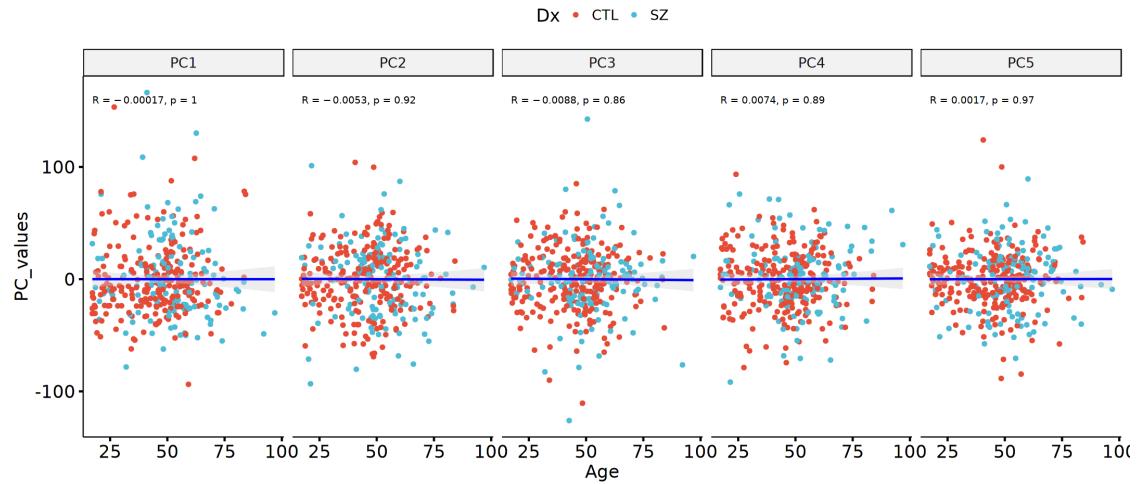
```
`geom_smooth()` using formula 'y ~ x'
```



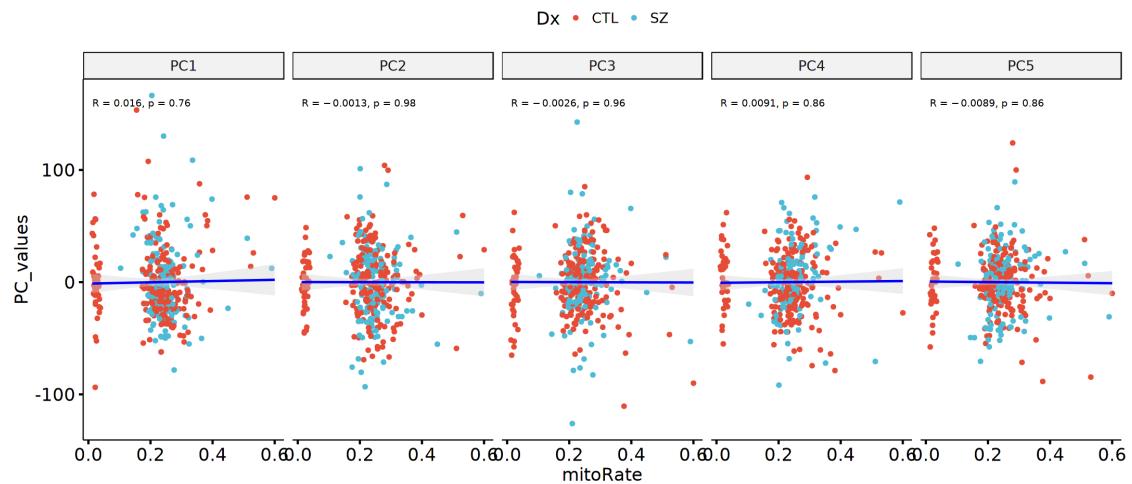
```
`geom_smooth()` using formula 'y ~ x'
```



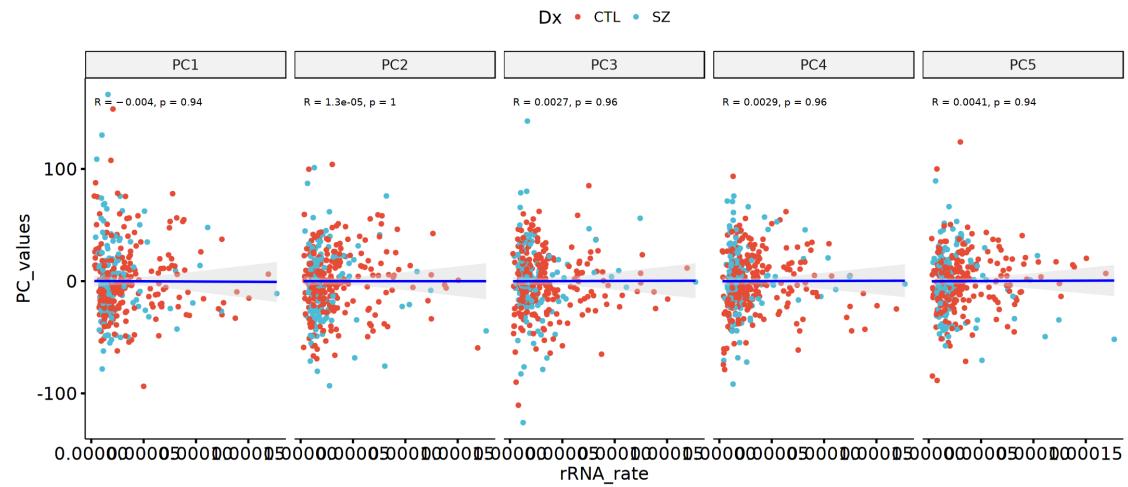
```
`geom_smooth()` using formula 'y ~ x'
```



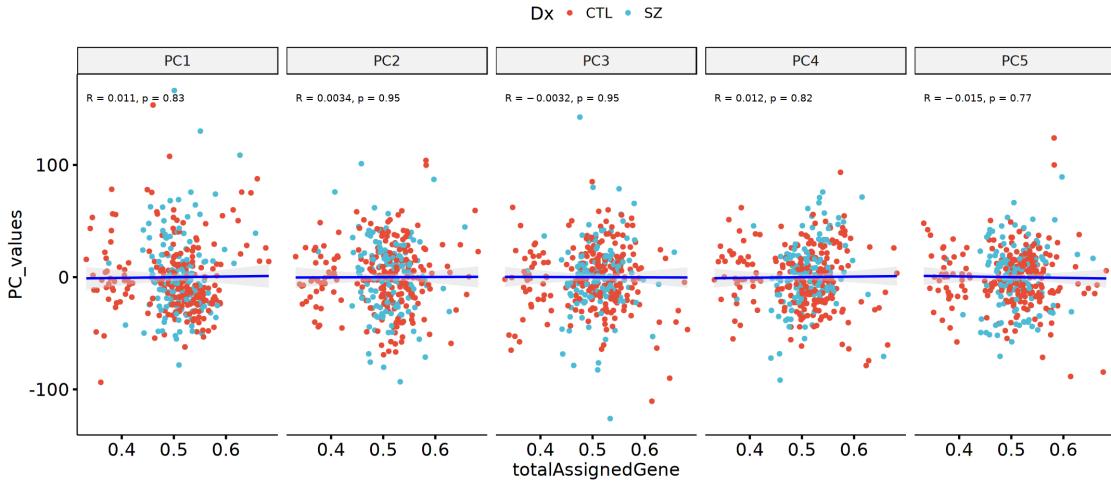
```
`geom_smooth()` using formula 'y ~ x'
```



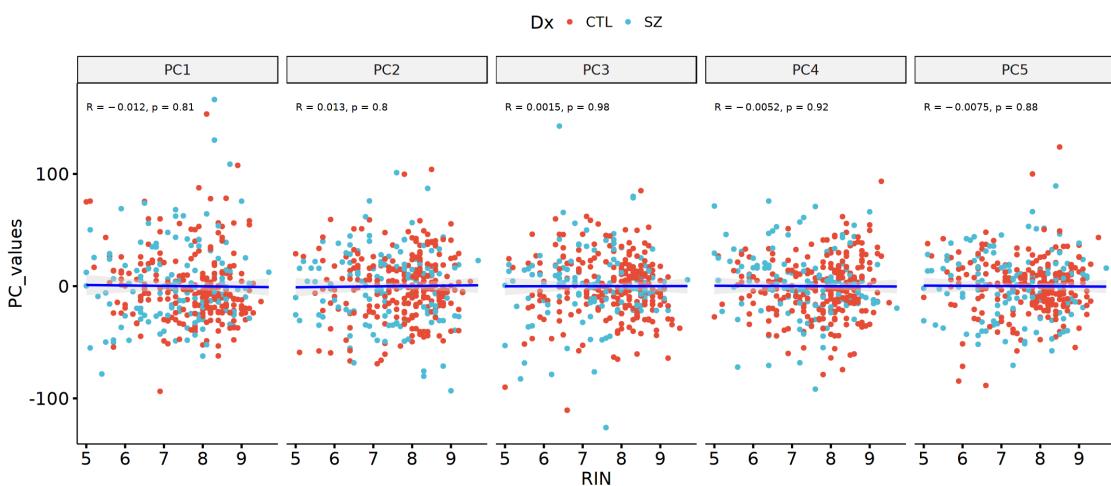
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

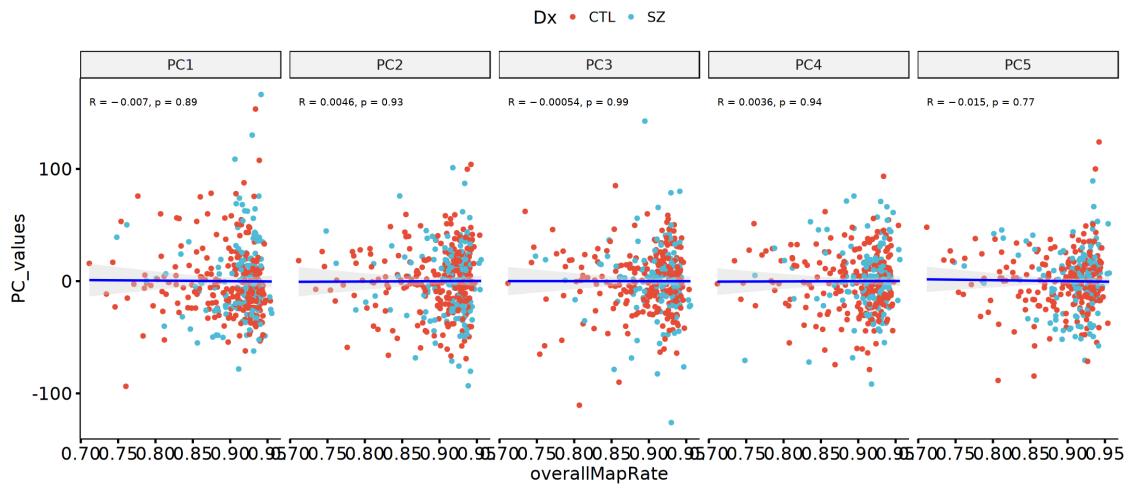


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

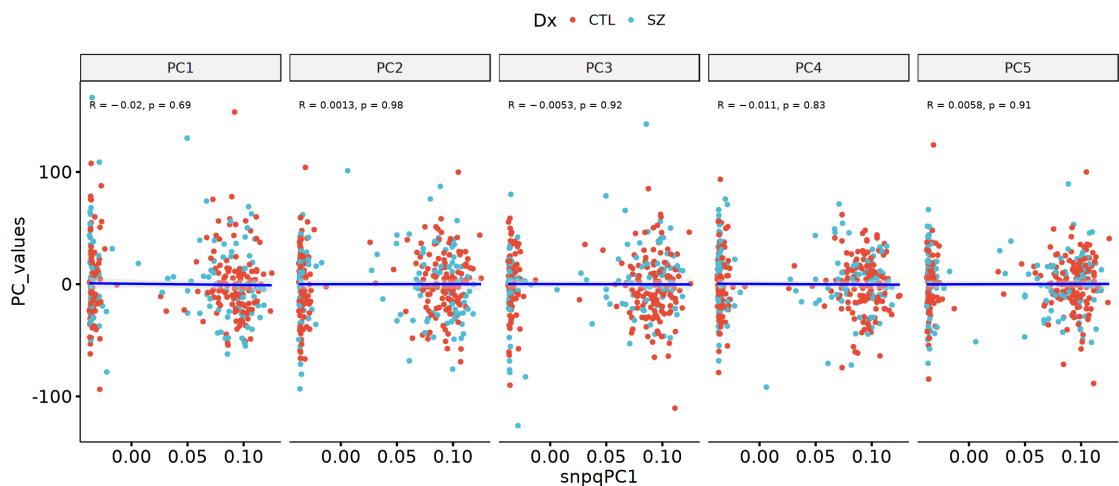


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

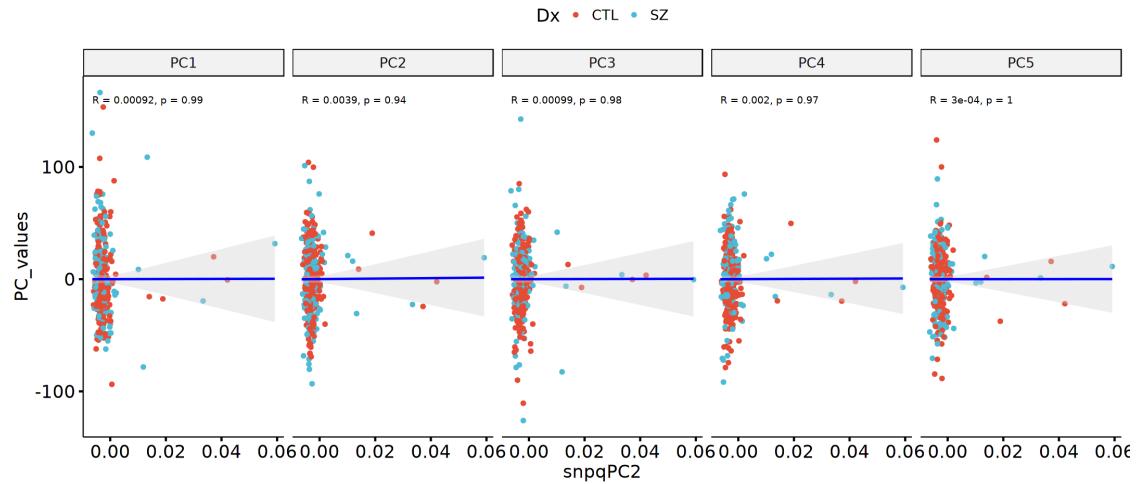
```
`geom_smooth()` using formula 'y ~ x'
```



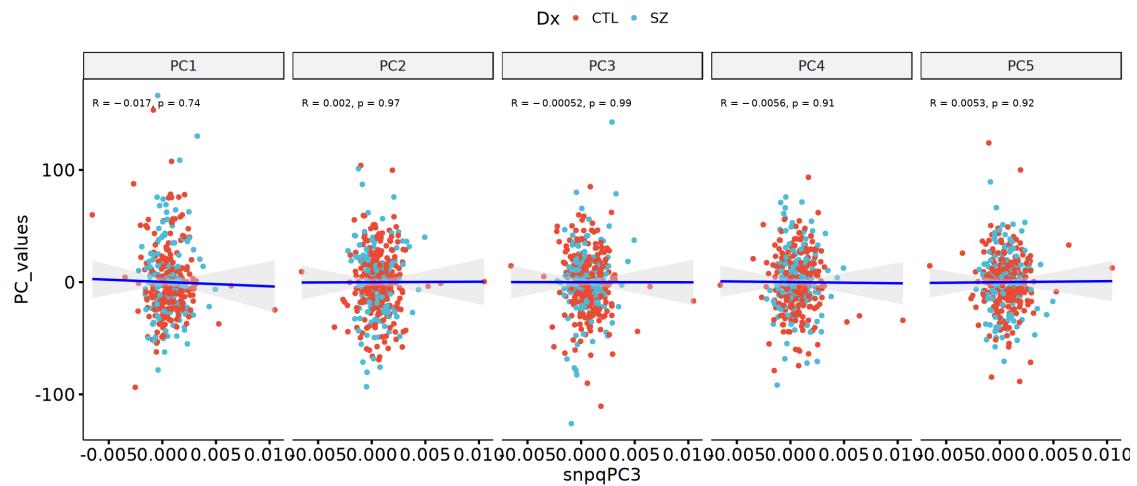
```
`geom_smooth()` using formula 'y ~ x'
```



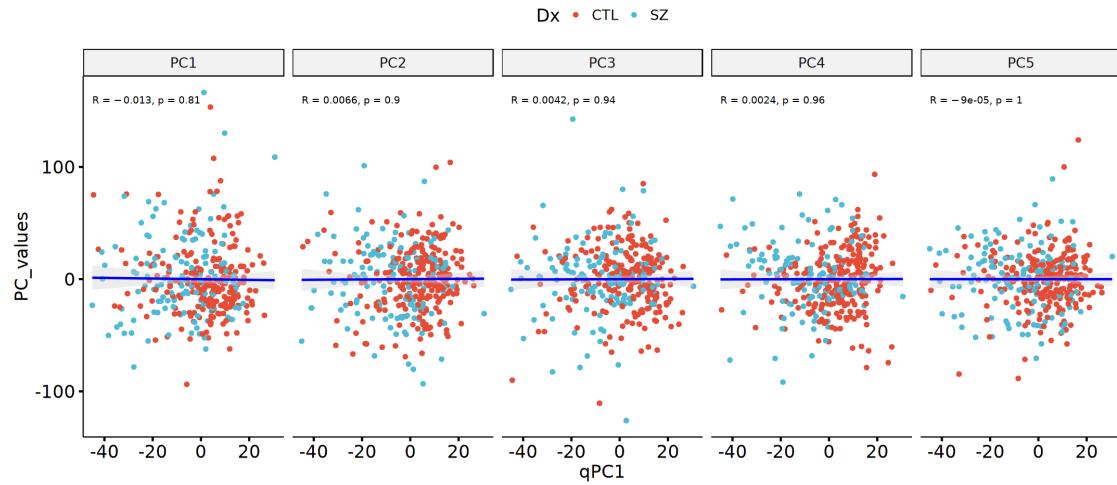
```
`geom_smooth()` using formula 'y ~ x'
```



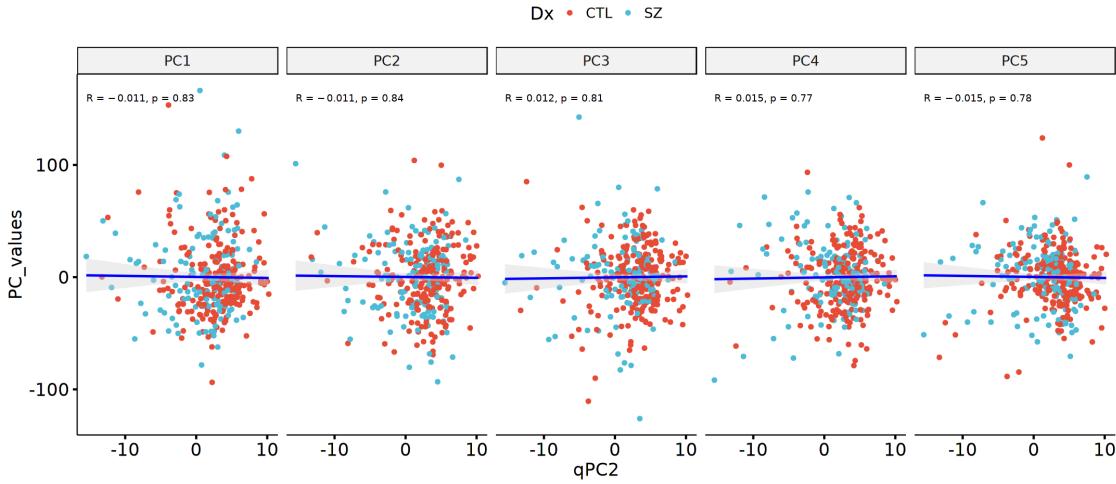
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

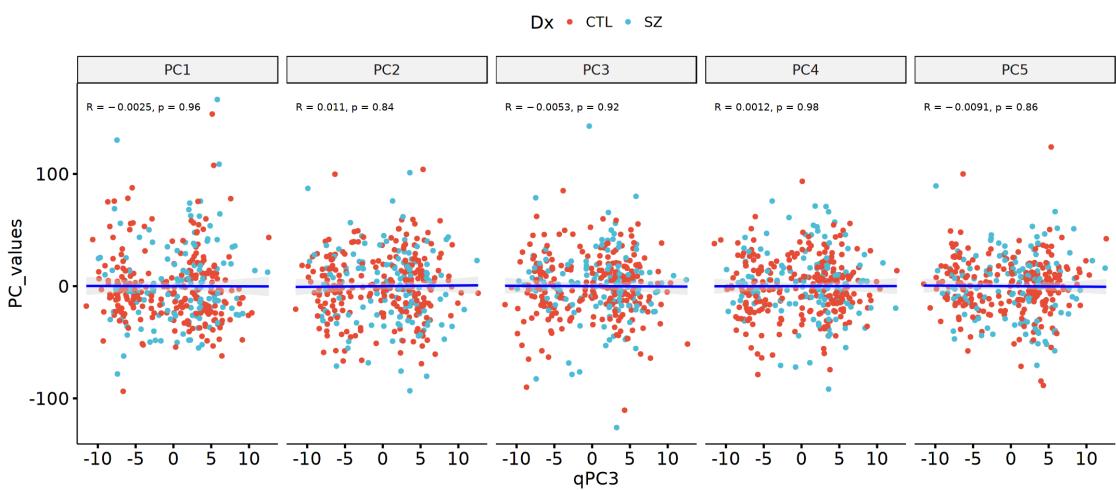


```
`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'
```

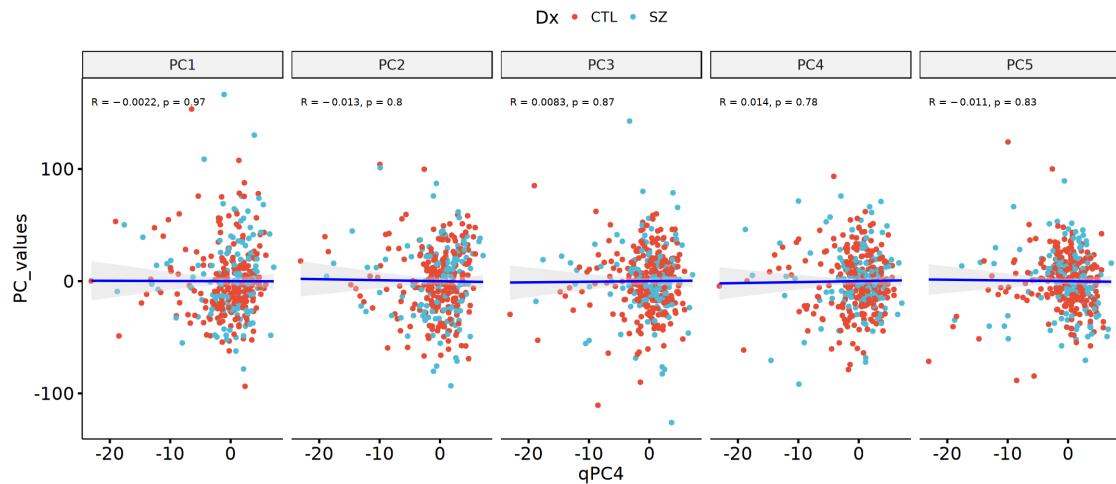


```
`geom_smooth()` using formula 'y ~ x'  

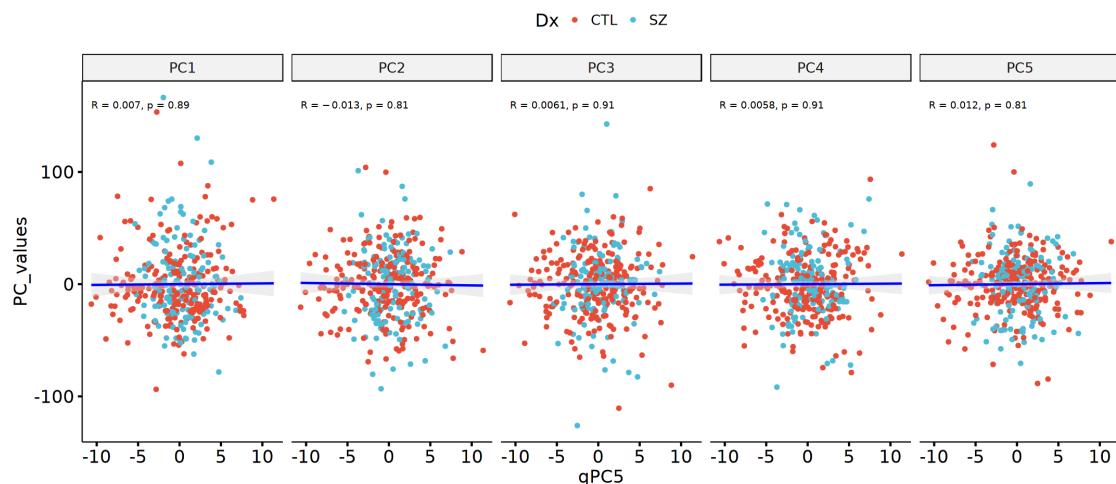
`geom_smooth()` using formula 'y ~ x'  

`geom_smooth()` using formula 'y ~ x'
```

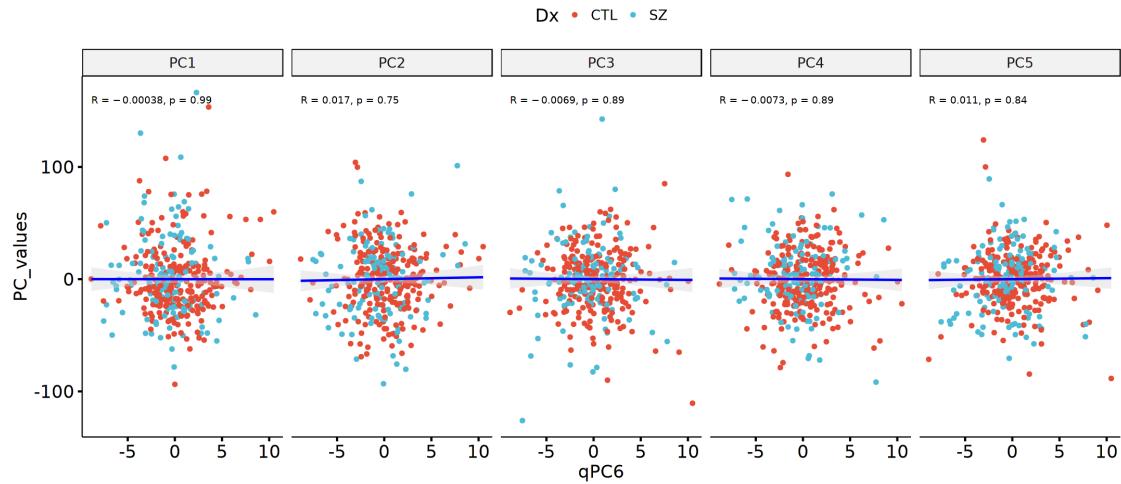
```
`geom_smooth()` using formula 'y ~ x'
```



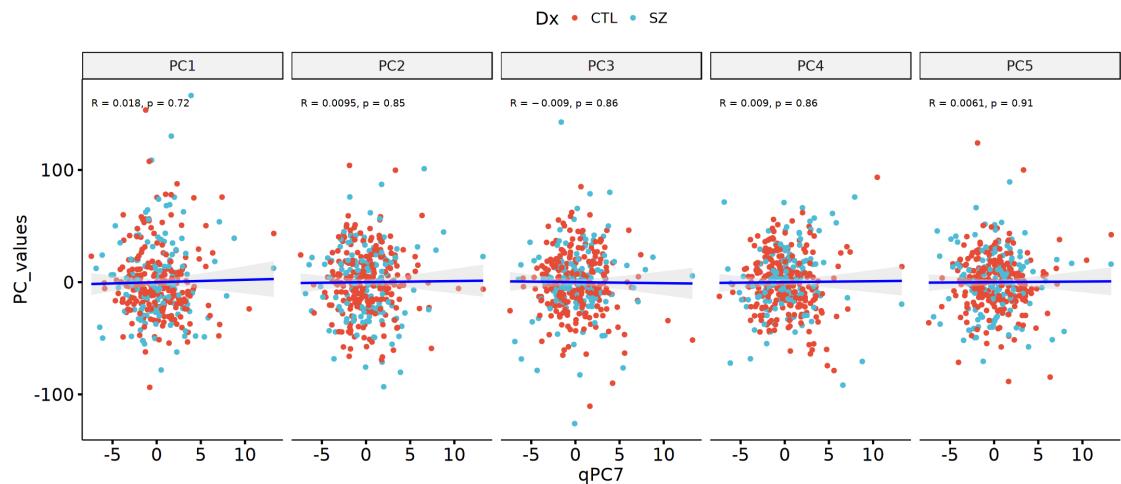
```
`geom_smooth()` using formula 'y ~ x'
```



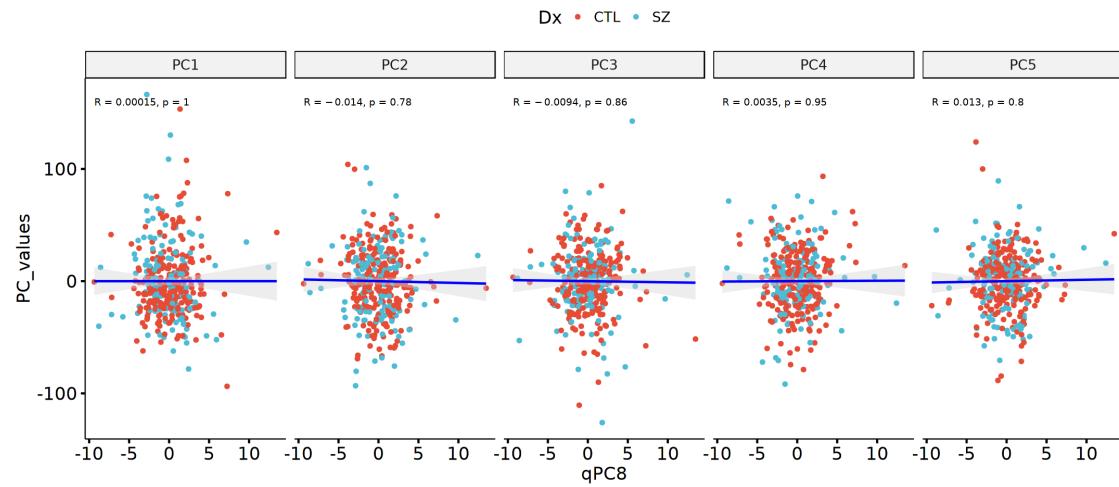
```
`geom_smooth()` using formula 'y ~ x'
```



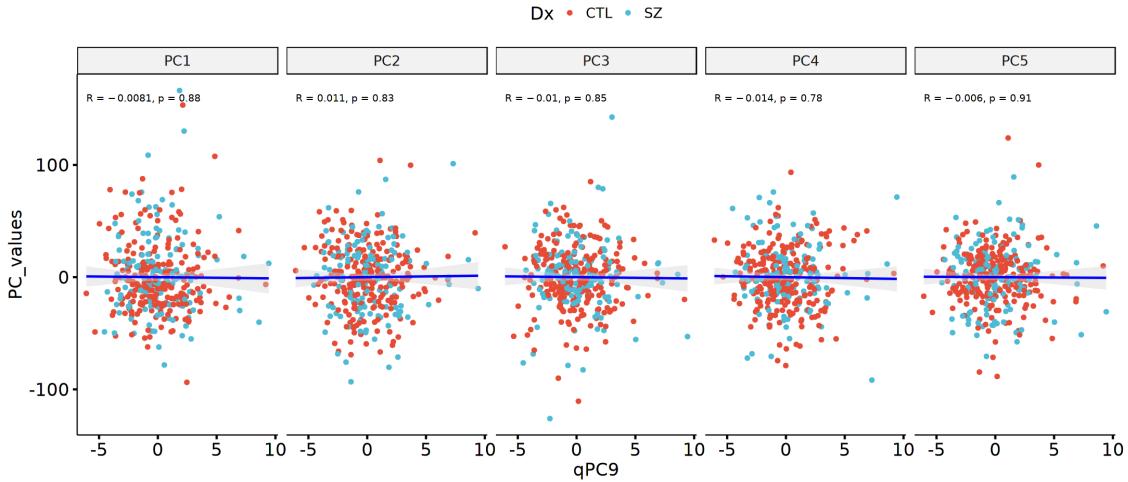
```
`geom_smooth()` using formula 'y ~ x'
```



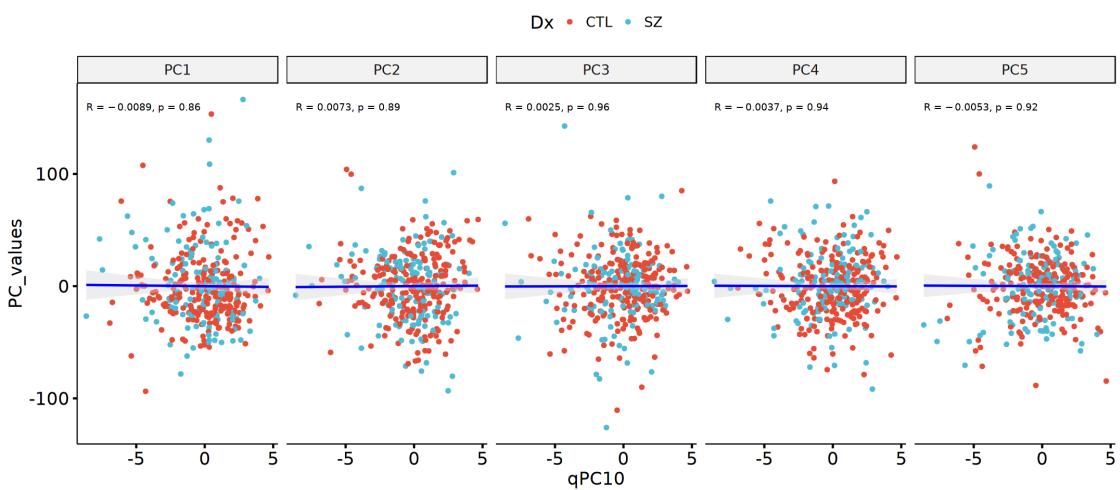
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```

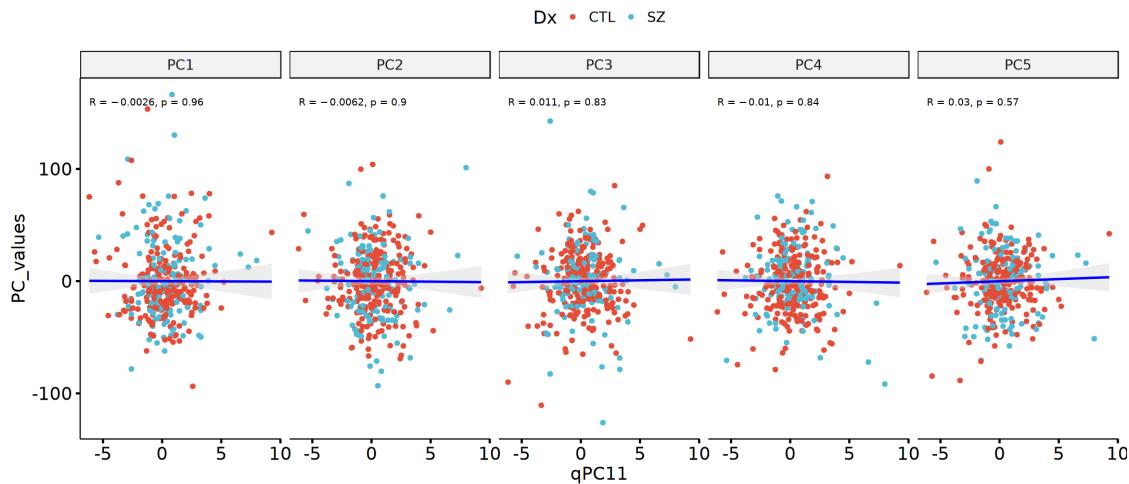


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

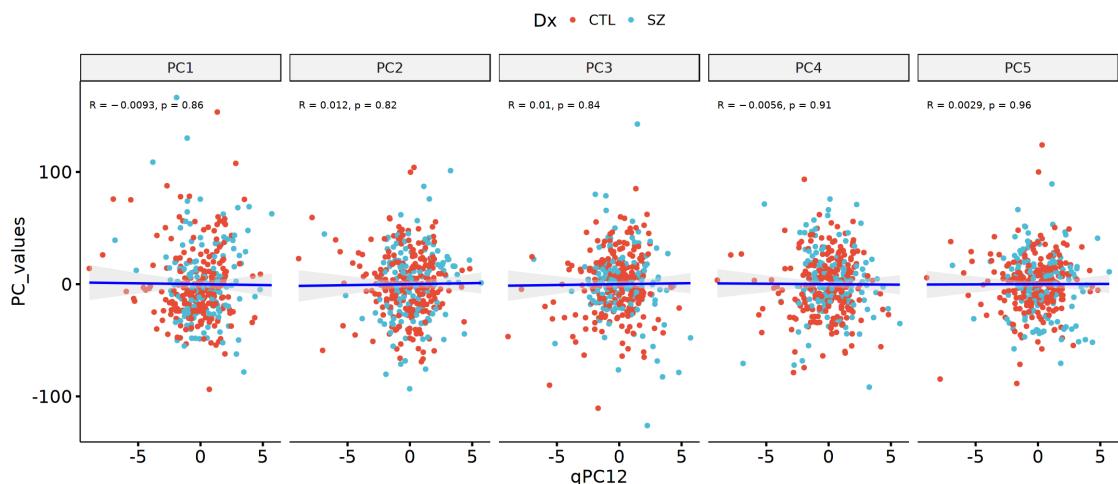


```
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'  
`geom_smooth()` using formula 'y ~ x'
```

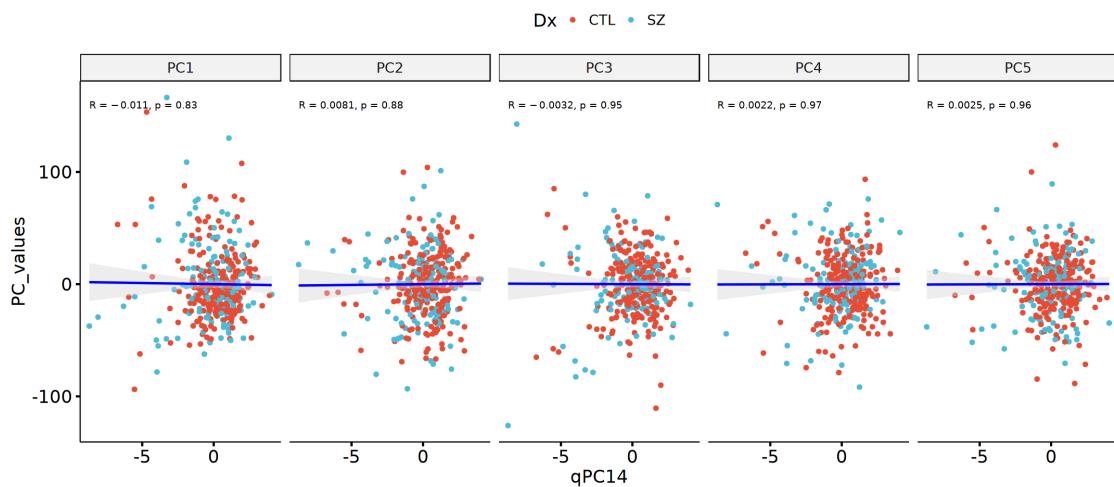
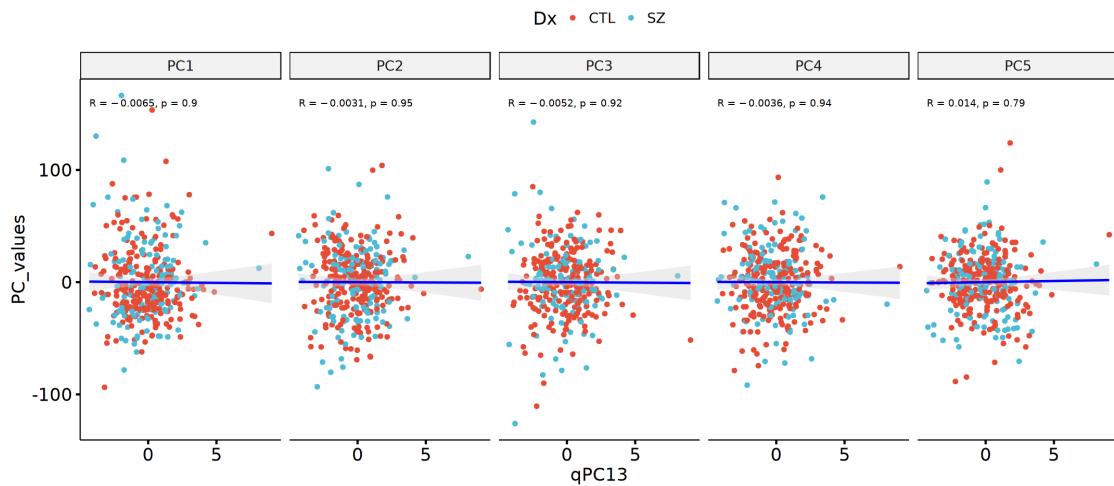
```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



```
`geom_smooth()` using formula 'y ~ x'
```



1.6 Reproducibility Information

```
[14]: Sys.time()
proc.time()
options(width = 120)
sessioninfo::session_info()
```

```
[1] "2021-07-23 10:52:58 EDT"
```

```
    user   system  elapsed
3552.128 2011.560 1044.568
```

```

Session info
setting  value
version  R version 4.0.3 (2020-10-10)
os       Arch Linux
system   x86_64, linux-gnu
ui       X11
language (EN)
collate  en_US.UTF-8
ctype    en_US.UTF-8
tz       America/New_York
date     2021-07-23

```

Packages

package	* version	date	lib	source
abind	1.4-5	2016-07-21	[1]	CRAN (R 4.0.2)
annotate	1.68.0	2020-10-27	[1]	Bioconductor
AnnotationDbi	1.52.0	2020-10-27	[1]	Bioconductor
assertthat	0.2.1	2019-03-21	[1]	CRAN (R 4.0.2)
backports	1.2.1	2020-12-09	[1]	CRAN (R 4.0.2)
base64enc	0.1-3	2015-07-28	[1]	CRAN (R 4.0.2)
Biobase	* 2.50.0	2020-10-27	[1]	Bioconductor
BiocGenerics	* 0.36.1	2021-04-16	[1]	Bioconductor
BiocParallel	* 1.24.1	2020-11-06	[1]	Bioconductor
bit	4.0.4	2020-08-04	[1]	CRAN (R 4.0.2)
bit64	4.0.5	2020-08-30	[1]	CRAN (R 4.0.2)
bitops	1.0-7	2021-04-24	[1]	CRAN (R 4.0.3)
blob	1.2.1	2020-01-20	[1]	CRAN (R 4.0.2)
broom	0.7.8	2021-06-24	[1]	CRAN (R 4.0.3)
cachem	1.0.5	2021-05-15	[1]	CRAN (R 4.0.3)
Cairo	1.5-12.2	2020-07-07	[1]	CRAN (R 4.0.2)
car	3.0-11	2021-06-27	[1]	CRAN (R 4.0.3)
carData	3.0-4	2020-05-22	[1]	CRAN (R 4.0.2)
cellranger	1.1.0	2016-07-27	[1]	CRAN (R 4.0.2)
cli	3.0.0	2021-06-30	[1]	CRAN (R 4.0.3)
colorspace	2.0-2	2021-06-24	[1]	CRAN (R 4.0.3)
crayon	1.4.1	2021-02-08	[1]	CRAN (R 4.0.3)
curl	4.3.2	2021-06-23	[1]	CRAN (R 4.0.3)
data.table	1.14.0	2021-02-21	[1]	CRAN (R 4.0.3)
DBI	1.1.1	2021-01-15	[1]	CRAN (R 4.0.2)
dbplyr	2.1.1	2021-04-06	[1]	CRAN (R 4.0.3)
DelayedArray	0.16.3	2021-03-24	[1]	Bioconductor
digest	0.6.27	2020-10-24	[1]	CRAN (R 4.0.2)
dplyr	* 1.0.7	2021-06-18	[1]	CRAN (R 4.0.3)
edgeR	* 3.32.1	2021-01-14	[1]	Bioconductor
ellipsis	0.3.2	2021-04-29	[1]	CRAN (R 4.0.3)
evaluate	0.14	2019-05-28	[1]	CRAN (R 4.0.2)
fansi	0.5.0	2021-05-25	[1]	CRAN (R 4.0.3)
farver	2.1.0	2021-02-28	[1]	CRAN (R 4.0.3)

fastmap	1.1.0	2021-01-25	[1]	CRAN	(R 4.0.2)
forcats	* 0.5.1	2021-01-27	[1]	CRAN	(R 4.0.2)
foreign	0.8-80	2020-05-24	[2]	CRAN	(R 4.0.3)
formula.tools	1.7.1	2018-03-01	[1]	CRAN	(R 4.0.2)
fs	1.5.0	2020-07-31	[1]	CRAN	(R 4.0.2)
genefilter	* 1.72.1	2021-01-21	[1]	Bioconductor	
generics	0.1.0	2020-10-31	[1]	CRAN	(R 4.0.2)
GenomeInfoDb	* 1.26.7	2021-04-08	[1]	Bioconductor	
GenomeInfoDbData	1.2.4	2021-02-02	[1]	Bioconductor	
GenomicRanges	* 1.42.0	2020-10-27	[1]	Bioconductor	
ggplot2	* 3.3.5	2021-06-25	[1]	CRAN	(R 4.0.3)
ggpubr	* 0.4.0	2020-06-27	[1]	CRAN	(R 4.0.2)
ggsci	2.9	2018-05-14	[1]	CRAN	(R 4.0.2)
ggsignif	0.6.2	2021-06-14	[1]	CRAN	(R 4.0.3)
glue	1.4.2	2020-08-27	[1]	CRAN	(R 4.0.2)
gttable	0.3.0	2019-03-25	[1]	CRAN	(R 4.0.2)
haven	2.4.1	2021-04-23	[1]	CRAN	(R 4.0.3)
hms	1.1.0	2021-05-17	[1]	CRAN	(R 4.0.3)
htmltools	0.5.1.1	2021-01-22	[1]	CRAN	(R 4.0.2)
httr	1.4.2	2020-07-20	[1]	CRAN	(R 4.0.2)
infer	0.5.4	2021-01-13	[1]	CRAN	(R 4.0.2)
IRanges	* 2.24.1	2020-12-12	[1]	Bioconductor	
IRdisplay	1.0	2021-01-20	[1]	CRAN	(R 4.0.2)
IRkernel	1.2	2021-05-11	[1]	CRAN	(R 4.0.3)
janitor	2.1.0	2021-01-05	[1]	CRAN	(R 4.0.2)
jsonlite	1.7.2	2020-12-09	[1]	CRAN	(R 4.0.2)
knitr	1.33	2021-04-24	[1]	CRAN	(R 4.0.3)
labeling	0.4.2	2020-10-20	[1]	CRAN	(R 4.0.2)
lattice	0.20-41	2020-04-02	[2]	CRAN	(R 4.0.3)
lifecycle	1.0.0	2021-02-15	[1]	CRAN	(R 4.0.3)
limma	* 3.46.0	2020-10-27	[1]	Bioconductor	
locfit	1.5-9.4	2020-03-25	[1]	CRAN	(R 4.0.2)
lubridate	1.7.10	2021-02-26	[1]	CRAN	(R 4.0.3)
magrittr	2.0.1	2020-11-17	[1]	CRAN	(R 4.0.2)
Matrix	1.3-4	2021-06-01	[1]	CRAN	(R 4.0.3)
MatrixGenerics	* 1.2.1	2021-01-30	[1]	Bioconductor	
matrixStats	* 0.59.0	2021-06-01	[1]	CRAN	(R 4.0.3)
memoise	* 2.0.0	2021-01-26	[1]	CRAN	(R 4.0.2)
mgcv	* 1.8-33	2020-08-27	[2]	CRAN	(R 4.0.3)
modelr	0.1.8	2020-05-19	[1]	CRAN	(R 4.0.2)
moderndive	* 0.5.1	2021-01-09	[1]	CRAN	(R 4.0.2)
munsell	0.5.0	2018-06-12	[1]	CRAN	(R 4.0.2)
nlme	* 3.1-152	2021-02-04	[1]	CRAN	(R 4.0.3)
openxlsx	4.2.4	2021-06-16	[1]	CRAN	(R 4.0.3)
operator.tools	1.6.3	2017-02-28	[1]	CRAN	(R 4.0.2)
pbdZMQ	0.3-5	2021-02-10	[1]	CRAN	(R 4.0.3)
pillar	1.6.1	2021-05-16	[1]	CRAN	(R 4.0.3)
pkgconfig	2.0.3	2019-09-22	[1]	CRAN	(R 4.0.2)

purrr	* 0.3.4	2020-04-17	[1]	CRAN	(R 4.0.2)
R6	2.5.0	2020-10-28	[1]	CRAN	(R 4.0.2)
Rcpp	1.0.7	2021-07-07	[1]	CRAN	(R 4.0.3)
RCurl	1.98-1.3	2021-03-16	[1]	CRAN	(R 4.0.3)
readr	* 1.4.0	2020-10-05	[1]	CRAN	(R 4.0.2)
readxl	1.3.1	2019-03-13	[1]	CRAN	(R 4.0.2)
repr	* 1.1.3	2021-01-21	[1]	CRAN	(R 4.0.2)
reprex	2.0.0	2021-04-02	[1]	CRAN	(R 4.0.3)
rio	0.5.27	2021-06-21	[1]	CRAN	(R 4.0.3)
rlang	0.4.11	2021-04-30	[1]	CRAN	(R 4.0.3)
RSQLite	2.2.7	2021-04-22	[1]	CRAN	(R 4.0.3)
rstatix	0.7.0	2021-02-13	[1]	CRAN	(R 4.0.3)
rstudioapi	0.13	2020-11-12	[1]	CRAN	(R 4.0.2)
rvest	1.0.0	2021-03-09	[1]	CRAN	(R 4.0.3)
S4Vectors	* 0.28.1	2020-12-09	[1]	Bioconductor	
scales	1.1.1	2020-05-11	[1]	CRAN	(R 4.0.2)
sessioninfo	1.1.1	2018-11-05	[1]	CRAN	(R 4.0.2)
snakecase	0.11.0	2019-05-25	[1]	CRAN	(R 4.0.2)
stringi	1.7.3	2021-07-16	[1]	CRAN	(R 4.0.3)
stringr	* 1.4.0	2019-02-10	[1]	CRAN	(R 4.0.2)
SummarizedExperiment	* 1.20.0	2020-10-27	[1]	Bioconductor	
survival	3.2-7	2020-09-28	[2]	CRAN	(R 4.0.3)
sva	* 3.38.0	2020-10-27	[1]	Bioconductor	
svglite	2.0.0	2021-02-20	[1]	CRAN	(R 4.0.3)
systemfonts	1.0.2	2021-05-11	[1]	CRAN	(R 4.0.3)
tibble	* 3.1.2	2021-05-16	[1]	CRAN	(R 4.0.3)
tidyverse	* 1.1.3	2021-03-03	[1]	CRAN	(R 4.0.3)
tidyselect	1.1.1	2021-04-30	[1]	CRAN	(R 4.0.3)
tidyverse	* 1.3.1	2021-04-15	[1]	CRAN	(R 4.0.3)
utf8	1.2.1	2021-03-12	[1]	CRAN	(R 4.0.3)
uuid	0.1-4	2020-02-26	[1]	CRAN	(R 4.0.2)
vctrs	0.3.8	2021-04-29	[1]	CRAN	(R 4.0.3)
withr	2.4.2	2021-04-18	[1]	CRAN	(R 4.0.3)
xfun	0.24	2021-06-15	[1]	CRAN	(R 4.0.3)
XML	3.99-0.6	2021-03-16	[1]	CRAN	(R 4.0.3)
xml2	1.3.2	2020-04-23	[1]	CRAN	(R 4.0.2)
xtable	1.8-4	2019-04-21	[1]	CRAN	(R 4.0.2)
XVector	0.30.0	2020-10-27	[1]	Bioconductor	
zip	2.2.0	2021-05-31	[1]	CRAN	(R 4.0.3)
zlibbioc	1.36.0	2020-10-27	[1]	Bioconductor	

```
[1] /home/jbenja13/R/x86_64-pc-linux-gnu-library/4.0
[2] /usr/lib/R/library
```