Generation model

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

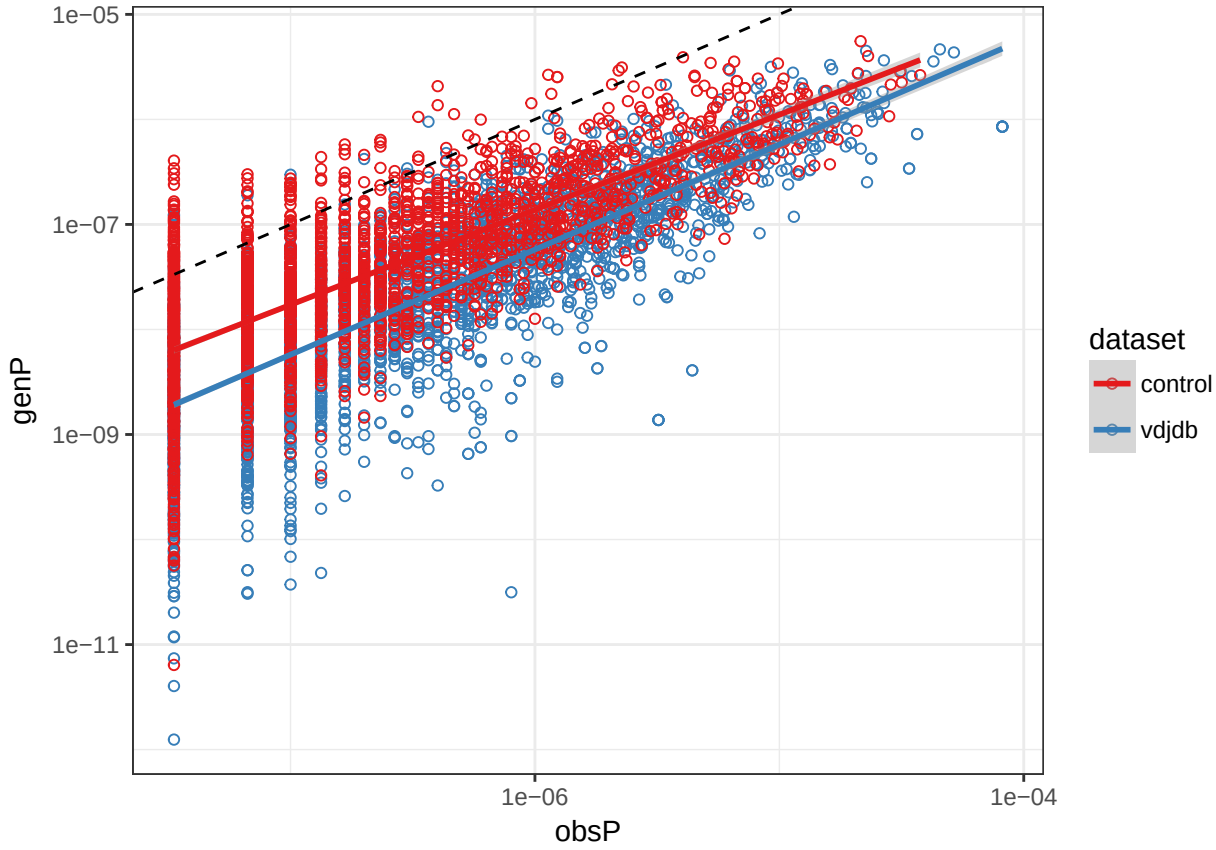
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

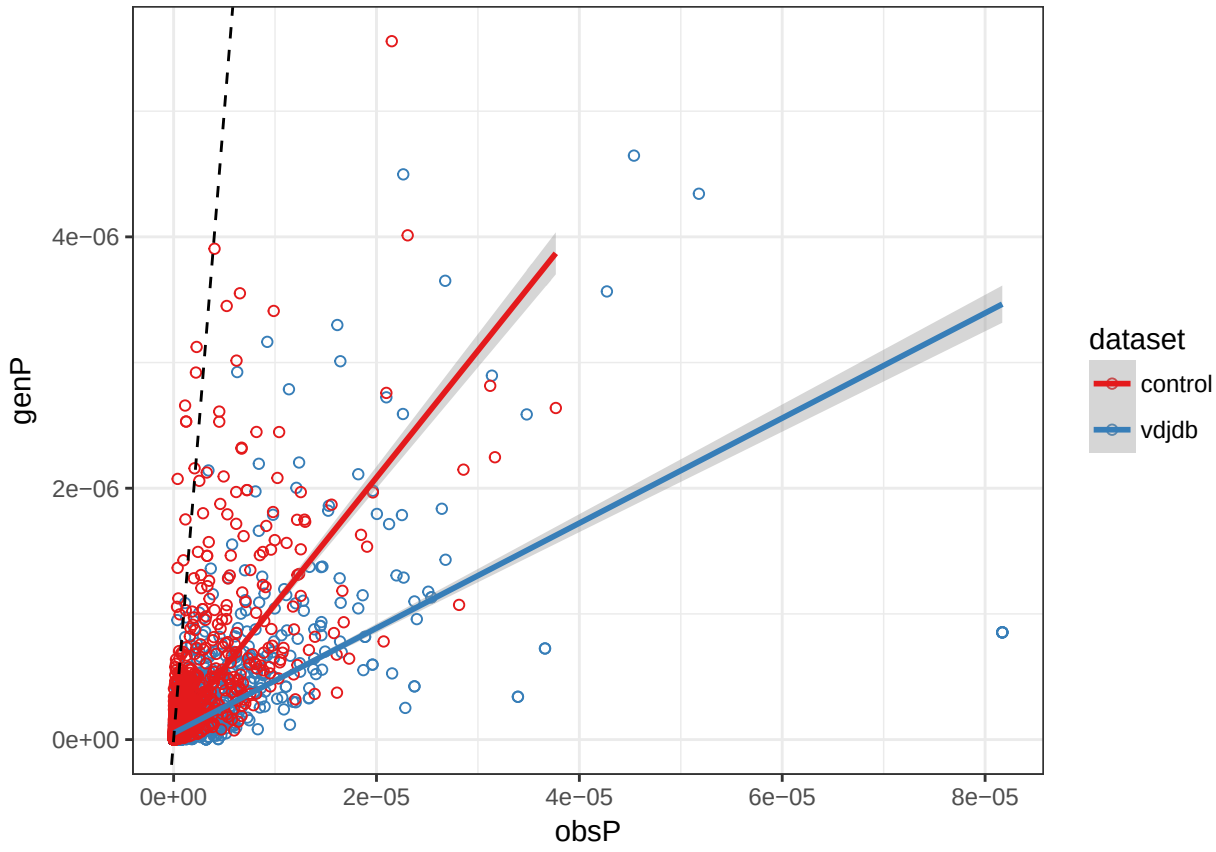
```
library(ggplot2)
TOTAL_REARRANGEMENTS_AGING = 29989055

df = read.table("VDJDB_fullP.txt", header=T, sep="\t")
df$obsP = df$ageing_occur / TOTAL_REARRANGEMENTS_AGING
df.control = read.table("control_probs.txt", header=T, sep="\t") %>%
  mutate(occurrences = ifelse(is.na(occurrences), 0, occurrences), genP = fullP)
df.control$obsP = df.control$occurrences / TOTAL_REARRANGEMENTS_AGING

df.1 = df %>% select(obsP, genP, antigen.epitope, antigen.gene)
df.1$dataset = "vdjdb"
df.2 = df.control %>% select(obsP, genP)
df.2$antigen.epitope = "NA"
df.2$antigen.gene = "NA"
df.2$dataset = "control"
df.3 = rbind(df.1, df.2) %>% filter(obsP>0 & genP>0)
```

Correlation between genP and obsP

```
ggplot(df.3, aes(x=obsP, y=genP, color=dataset)) +
  geom_point(shape=21) +
  geom_smooth(method="lm") +
  geom_abline(linetype="dashed") +
  scale_x_log10() +
  scale_y_log10() +
  scale_color_brewer(palette = "Set1") +
  theme_bw()
```

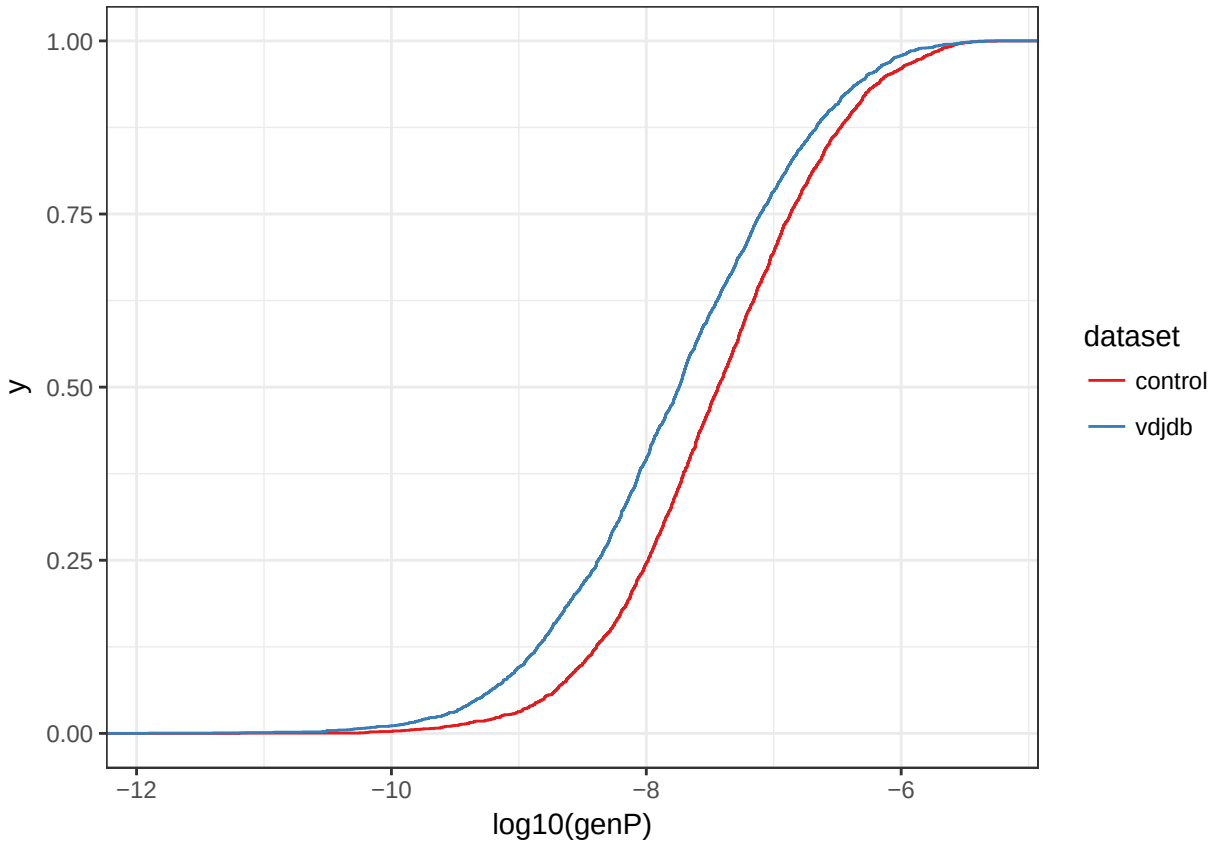



```
ggplot(df.3, aes(x=obsP, y=genP, color=dataset)) +
  geom_point(shape=21) +
  geom_smooth(method="lm") +
  geom_abline(linetype="dashed") +
  scale_color_brewer(palette = "Set1") +
  theme_bw()
```

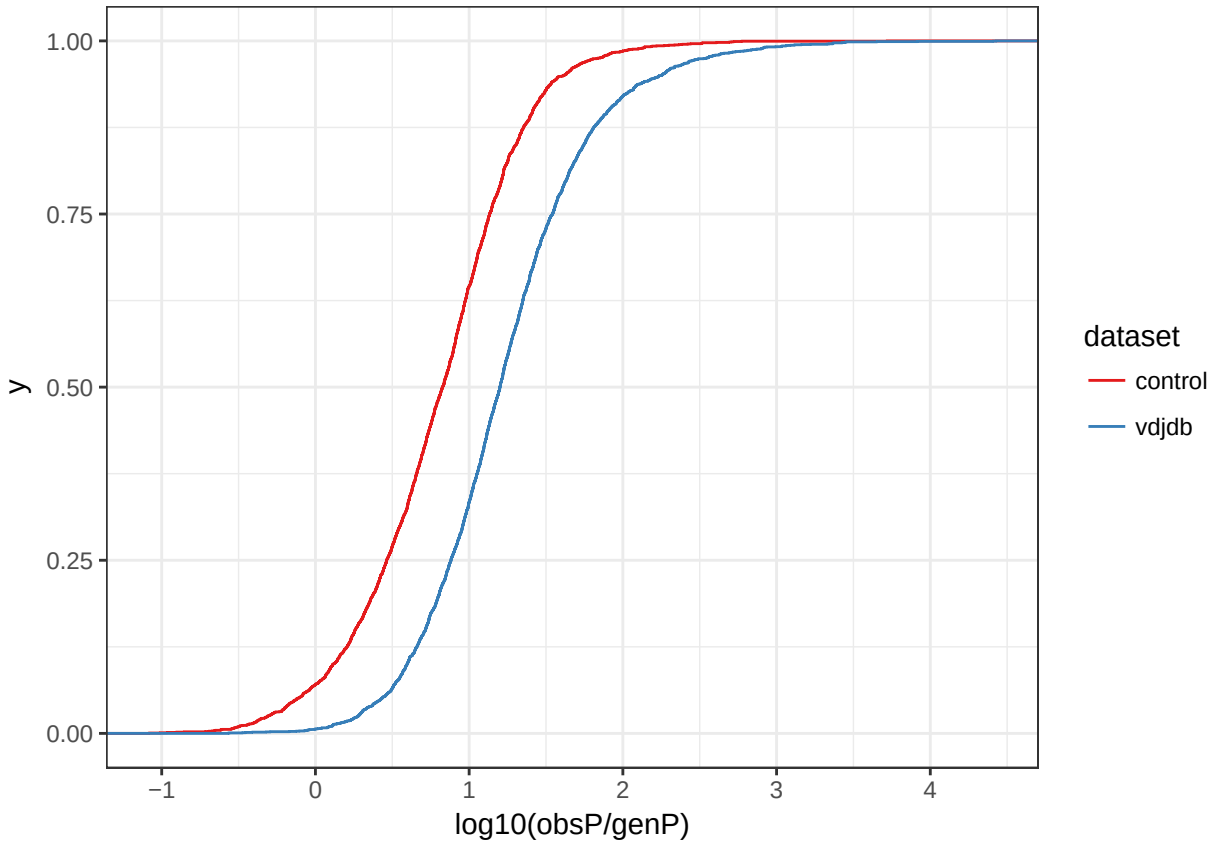


Selection effect

```
ggplot(df.3, aes(x=log10(genP), color = dataset)) +
  stat_ecdf() +
  scale_color_brewer(palette = "Set1") +
  theme_bw()
```




```
ggplot(df.3, aes(x=log10(obsP/genP), color = dataset)) +
  stat_ecdf() +
  scale_color_brewer(palette = "Set1") +
  theme_bw()
```


1.00
0.75
0.50
0.25
0.00
y
−1
0
1
2
3
4
log10(obsP/genP)
dataset
control
vdjdb