

# CodeS1: Treatments

Sigurgeir Ólafsson

9/19/2022

## Introduction

This document describes the treatment metadata and how we looked for treatment effects on the somatic mutation landscape of the skin in psoriasis patients.

```
.libPaths("/lustre/scratch126/humgen/projects/psoriasis/R_packages_farm5_R4.1.0_install/")
library(ggplot2)
library(reshape2)
library(cowplot)
library(ggsignif)
library(ggExtra)
library(MutationalPatterns)
library(GenomicRanges)
library(BSgenome)
ref_genome <- "BSgenome.Hsapiens.UCSC.hg38"
library(ref_genome, character.only = TRUE)
library(dplyr)
library(nlme)
library(scales)

## DEFINE PLOTTING VARIABLES
#####
BASESIZE=18

working_dir="/nfs/users/nfs_s/sol1/phd/psoriasis/bsub_jupyter_lab/psoriasis/manuscript_data_and_figures/"

microd_meta <- read.table(paste(working_dir, "Microdissection_metaData.txt", sep=""), h=T)
patient_meta <- read.table(paste(working_dir, "Patient_metaData.txt", sep=""), h=T)
treatment_data <- read.table(paste(working_dir, "treatment_data_combined.txt", sep=""), h=T)

patient_meta$Shows_Psoralen_signature <- ifelse(patient_meta$anyCloneGT50PUVA=="Yes" | patient_meta$Patient.ID %i
n% c("patient89","patient56","patient105"), T, F)
patient_meta <- merge(patient_meta, treatment_data, by.x="Patient.ID", by.y="PatientID", all.x=T)
patient_meta$Sex[patient_meta$Sex=="female"] <- "Female"
patient_meta$Sex[patient_meta$Sex=="male"] <- "Male"
```

## Visualize the treatment data

We see that we have no dosage information for methotrexate or topical steroids. For the phototreatments, we have information about the cumulative doses

```

puva <- patient_meta[,c("Patient.ID", "AmountPuva", "EverPuva")]
puva$Treatment <- "PUVA"
colnames(puva)[2] <- "NrSessions"
colnames(puva)[3] <- "Treated"
uvb <- patient_meta[,c("Patient.ID", "AmountUVB", "EverUVB")]
uvb$Treatment <- "UVB"
colnames(uvb)[2] <- "NrSessions"
colnames(uvb)[3] <- "Treated"

methotrexate <- patient_meta[, c("Patient.ID", "EverMethotrexate")]
colnames(methotrexate)[2] <- "Treated"
methotrexate$Treatment <- "Methotrexate"

steroids <- patient_meta[, c("Patient.ID", "EverSteroids")]
colnames(steroids)[2] <- "Treated"
steroids$Treatment <- "Topical Steroids"

photo <- rbind(puva, uvb)
photo$NrSessions[photo$NrSessions=="unknown"] <- "Unknown (>0)"
photo$NrSessions <- factor(photo$NrSessions, levels=c("None", "Unknown (>0)", "<=50", "51-200", ">200"))

treatments <- rbind(puva[,c(1,3,4)], uvb[,c(1,3,4)], methotrexate, steroids)

treatments$Treated <- ifelse(treatments$Treated, "Treatment history", "No history")
treatments$Treated <- factor(treatments$Treated, levels=c("Treatment history", "No history"))
p1 <- ggplot(treatments, aes(x=Treatment, fill=Treated)) + geom_bar() + theme_classic(base_size = BASESIZE) +
  labs(x="", y="# Patients", fill="") +
  theme(axis.text.x = element_text(angle=90)) + scale_y_continuous(expand = c(0,0)) +
  theme(legend.position = "bottom") + scale_fill_brewer(palette = "Set1")

p2 <- ggplot(photo, aes(x=NrSessions)) + geom_bar(fill="gray") + theme_classic(base_size = BASESIZE) +
  facet_wrap(~Treatment) + labs(x="# Treatment sessions", y="# Patients") +
  theme(axis.text.x = element_text(angle=90)) + scale_y_continuous(expand = c(0,0), limits = c(0,111)) +
  scale_fill_brewer(palette = "Set2")

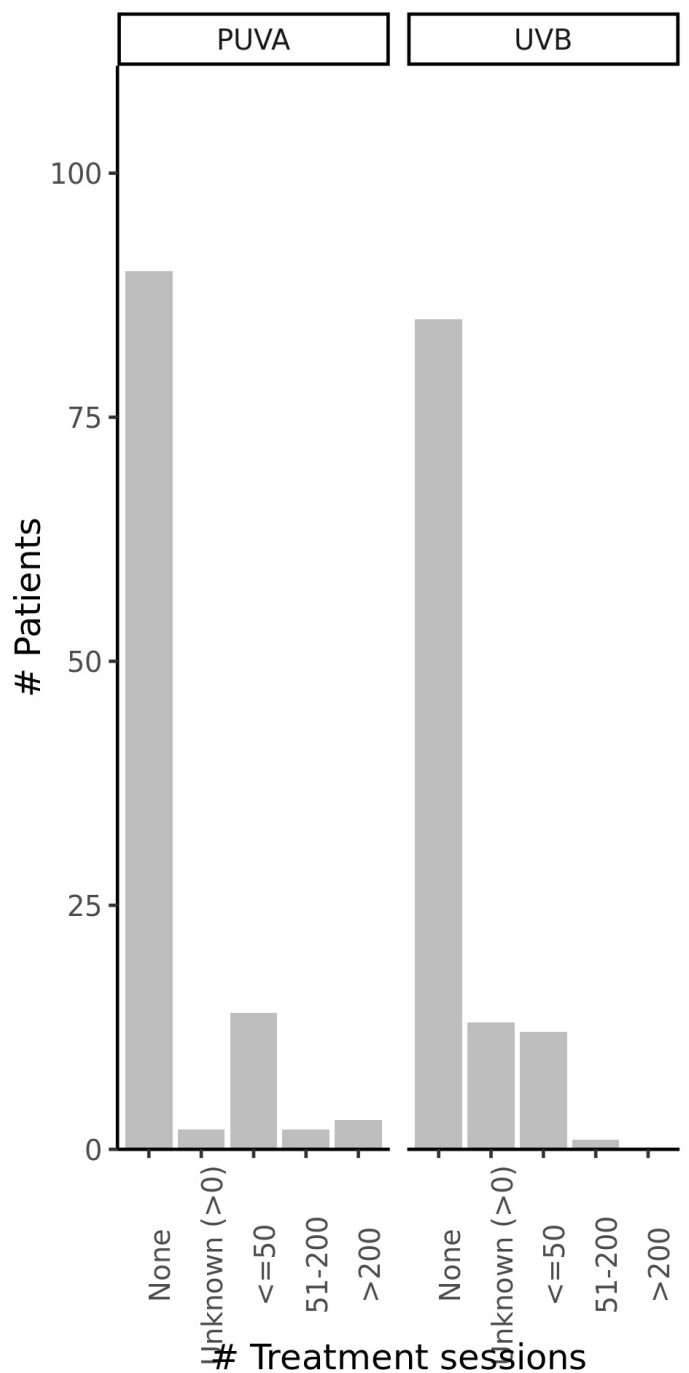
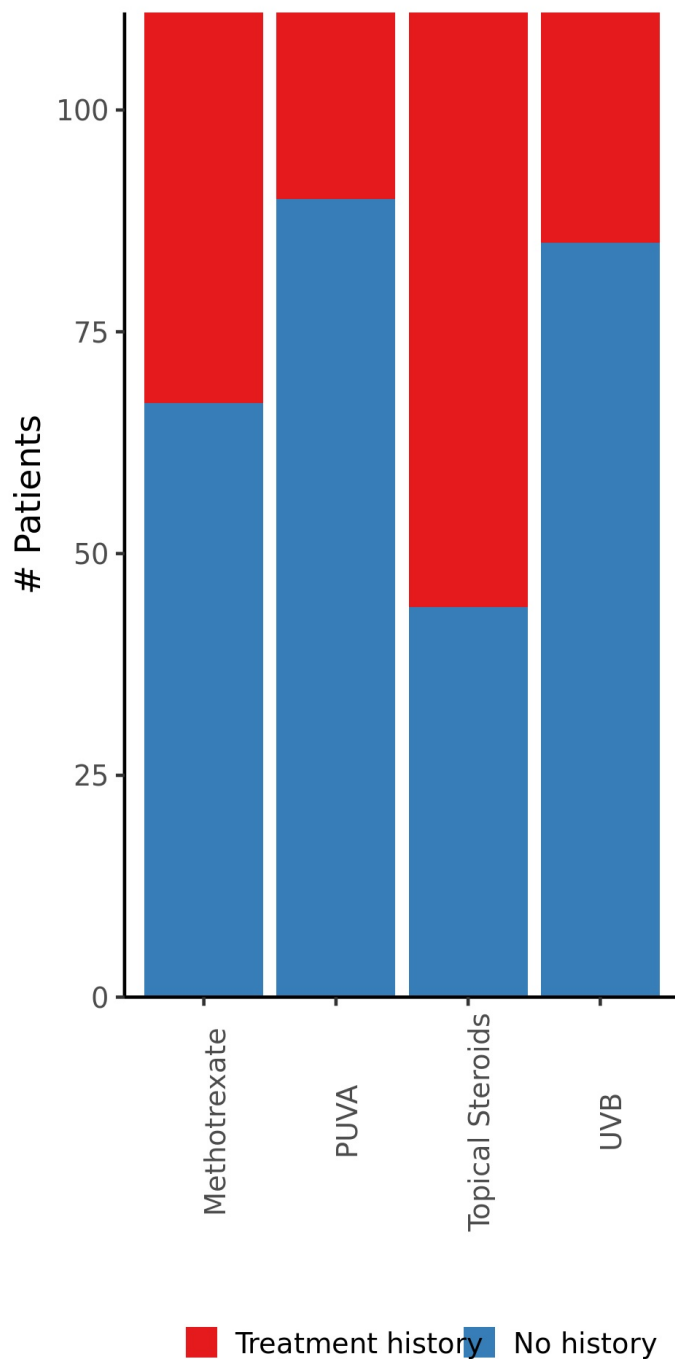
plot_grid(p1, p2, align="v")

```

```

## Warning: Graphs cannot be vertically aligned unless the axis parameter is set.
## Placing graphs unaligned.

```



## Effects of treatments on mutational spectra

Only a single unknown mutational signature was found during mutational signature extraction. We attribute this to psoralen exposure. PUVA treatment is likely a significant source of psoralen exposure in psoriasis patients. That only a single novel mutational signature was found is probably the strongest indication that other treatments don't affect the mutation spectra. However, to look further into the potential effects of the different treatments, I am going to compare the mutational spectra of patients who have or have not received the treatments. I want to pool the mutations from patients who have received different treatments. To keep any individuals with high mutation burdens from dominating the spectra though, I'm sampling 300 mutations from each patient.

### PUVA

We can see the effects of the Psoralen signature on the mutation spectrum of patients with history of PUVA treatment.

```

all_mutations$puva_history <- ifelse(all_mutations$patient %in% patient_meta$Patient.ID[patient_meta$EverPuva], "
PUVA-History", "No-PUVA-History")

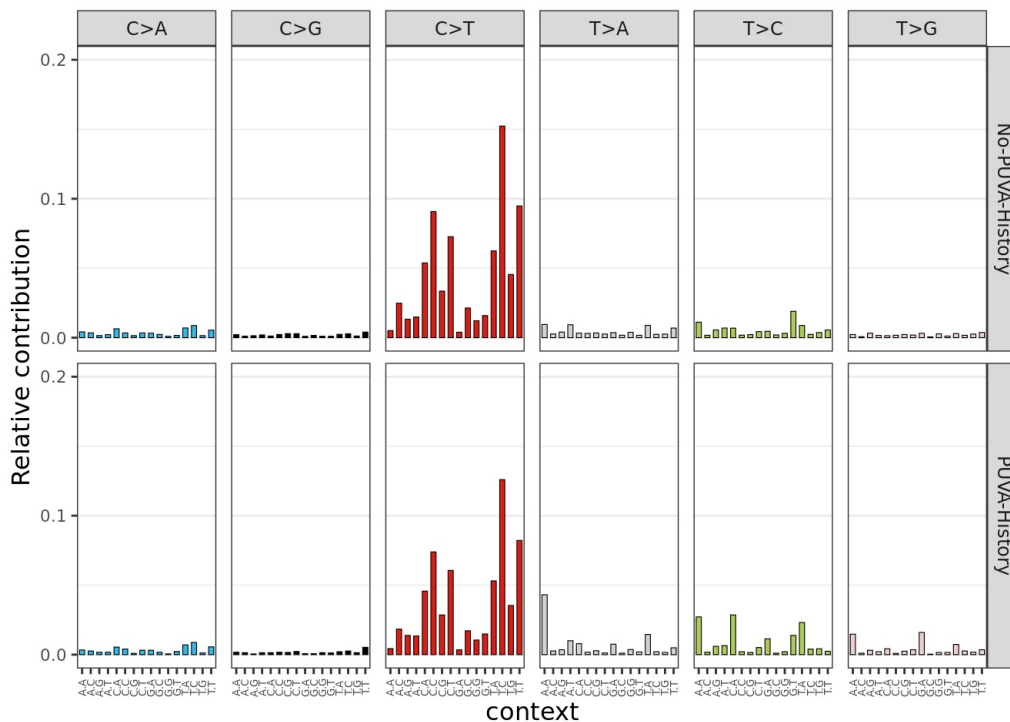
random_mutations <- all_mutations %>% group_by(patient) %>% sample_n(size = 300)

grange_obj = makeGRangesListFromDataFrame(random_mutations, split.field = "puva_history", keep.extra.columns = T,
ignore.strand = T, seqnames.field = "chr",
start.field = "pos", end.field = "pos")

GenomeInfoDb::genome(grange_obj) = "hg38"
subs_grl <- get_mut_type(grange_obj)
mut_mat <- mut_matrix(subs_grl, ref_genome)

plot_96_profile(mut_mat)

```



```
cos_sim(mut_mat[,1], mut_mat[,2])
```

```
## [1] 0.9677008
```

## UV-B

Plot the effects of UV-B on the mutational spectrum. UV-B might be especially likely to affect the burden of the UV-signature SBS7, which is the dominant mutational signature in the dataset. We formally test for an effect of this treatment on SBS7 burden below but here show that the mutational spectra of UV-B treated and untreated patients look nearly identical. In particular, there is no hint of a difference in the C>T class.

```

all_mutations$uvb_history <- ifelse(all_mutations$patient %in% patient_meta$Patient.ID[patient_meta$EverUVB], "UV
B-History", "No-UVB-History")

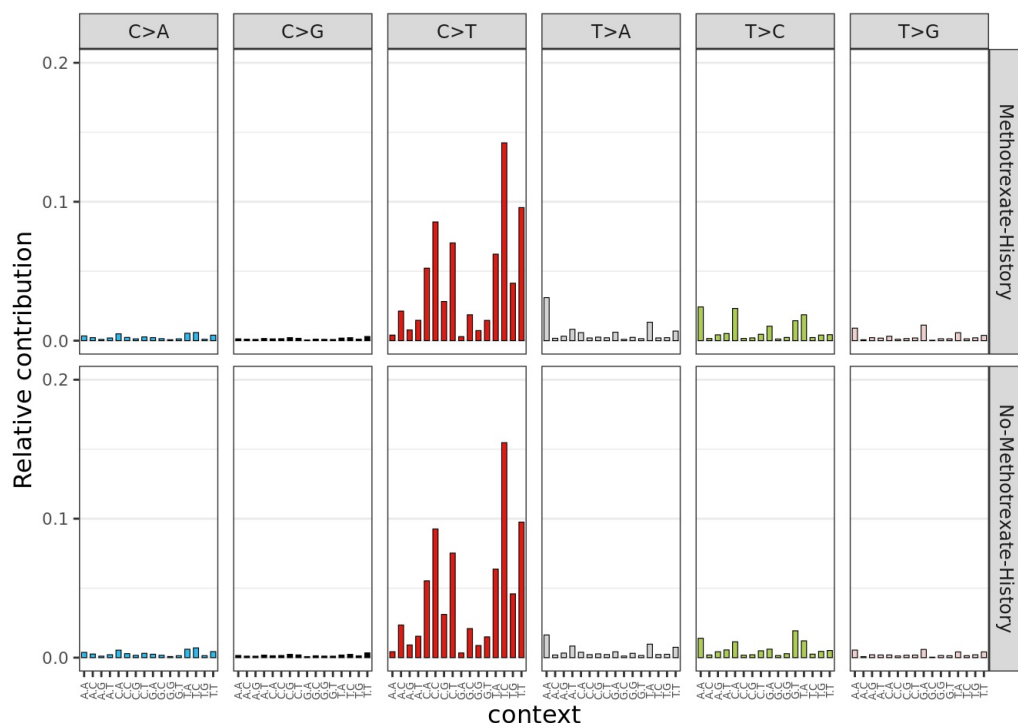
random_mutations <- all_mutations %>% group_by(patient) %>% sample_n(size = 300)
grange_obj = makeGRangesListFromDataFrame(all_mutations, split.field = "uvb_history", keep.extra.columns = T, ign
ore.strand = T, seqnames.field = "chr",
start.field = "pos", end.field = "pos")

GenomeInfoDb::genome(grange_obj) = "hg38"
subs_grl <- get_mut_type(grange_obj)
mut_mat <- mut_matrix(subs_grl, ref_genome)

plot_96_profile(mut_mat)

```





```
cos_sim(mut_mat[,1], mut_mat[,2])
```

```
## [1] 0.9933854
```

## Topical steroids

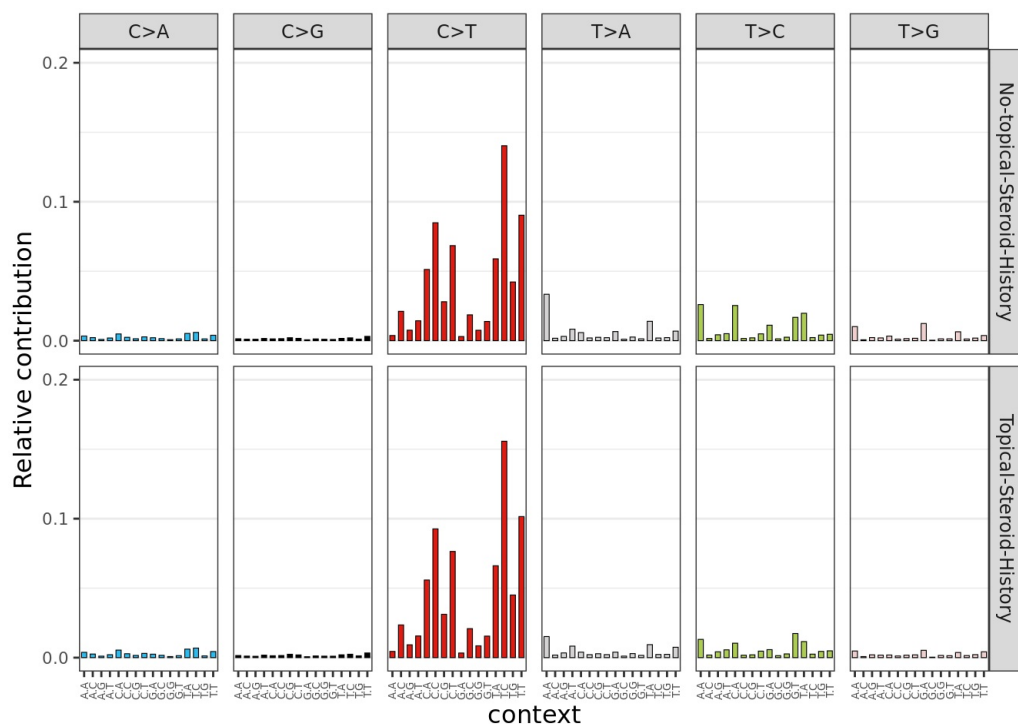
Compare the mutational spectra of patients with and without a history of topical steroid use.

```
all_mutations$steroid_history <- ifelse(all_mutations$patient %in% patient_meta$Patient.ID[patient_meta$EverSteroids], "Topical-Steroid-History", "No-topical-Steroid-History")
random_mutations <- all_mutations %>% group_by(patient) %>% sample_n(size = 300)

grange_obj = makeGRangesListFromDataFrame(all_mutations, split.field = "steroid_history", keep.extra.columns = T,
                                           ignore.strand = T, seqnames.field = "chr",
                                           start.field = "pos", end.field = "pos")

GenomeInfoDb::genome(grange_obj) = "hg38"
subs_grl <- get_mut_type(grange_obj)
mut_mat <- mut_matrix(subs_grl, ref_genome)

plot_96_profile(mut_mat)
```



```
cos_sim(mut_mat[,1], mut_mat[,2])
```

```
## [1] 0.9896338
```

## Effect of treatments on mutation burden of particular signatures

Another way to test for potential effects of treatments is to include the treatment status or dose in the linear models of mutation burden of different signatures. We can then test if including these variables improves the fit of the model.

```
working_dir="/nfs/users/nfs_s/sol1/phd/psoriasis/bsub_jupyter_lab/psoriasis/manuscript_data_and_figures/Supplementary_material/"

clone_burden <- read.table(paste(working_dir, "Supplementary_Table3_clone_mutationBurden.txt", sep=""), h=T)
clone_burden <- merge(clone_burden, microd_meta[,c("SampleID", "BiopsyID", "MetaLocation",
                                                  "PatientID", "DiseaseStatus")], by.x="HighCellFrac_sample", by
.y="SampleID")
clone_burden <- merge(clone_burden, patient_meta[,c("Patient.ID", "Age_at_sampling", "Disease_duration",
                                                  "Sex", "BMI", "Smoking", "PASI")], by.x="PatientID", by.y="Patient.ID")

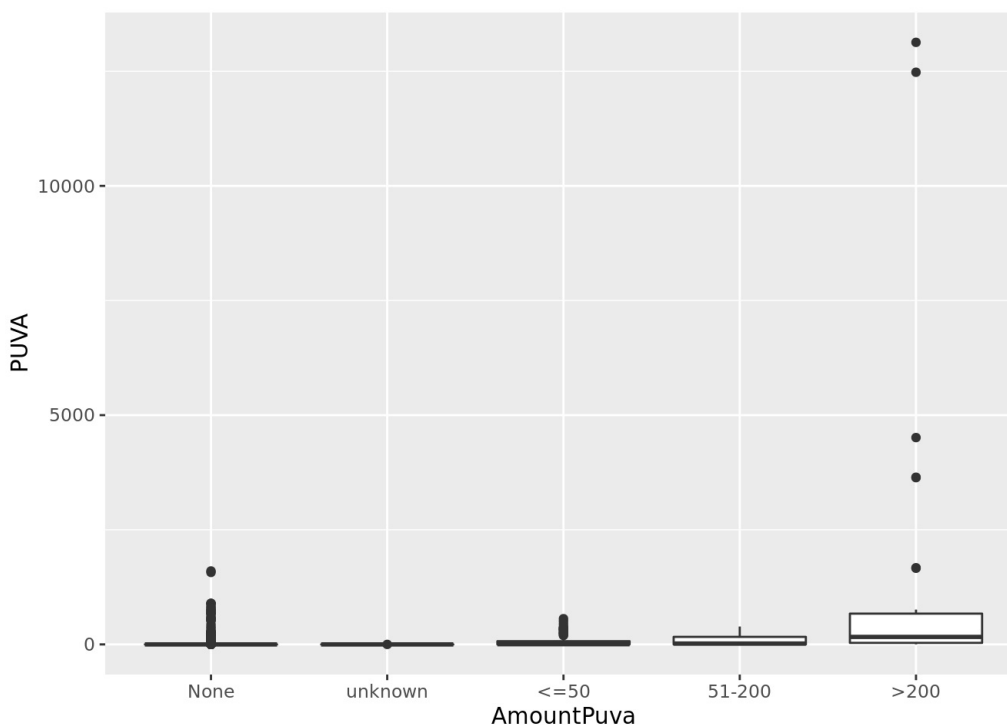
clone_burden$Disease_duration[clone_burden$DiseaseStatus=="Non-lesional" & !is.na(clone_burden$Disease_duration)]
<- 0

clone_burden <- merge(clone_burden, patient_meta[,c(1,22:27)], by.y="Patient.ID", by.x="PatientID")
clone_burden$AmountUVB <- factor(clone_burden$AmountUVB, levels=c("None","unknown", "<=50","51-200"))
clone_burden$AmountPuva <- factor(clone_burden$AmountPuva, levels=c("None","unknown", "<=50","51-200",">200"))
```

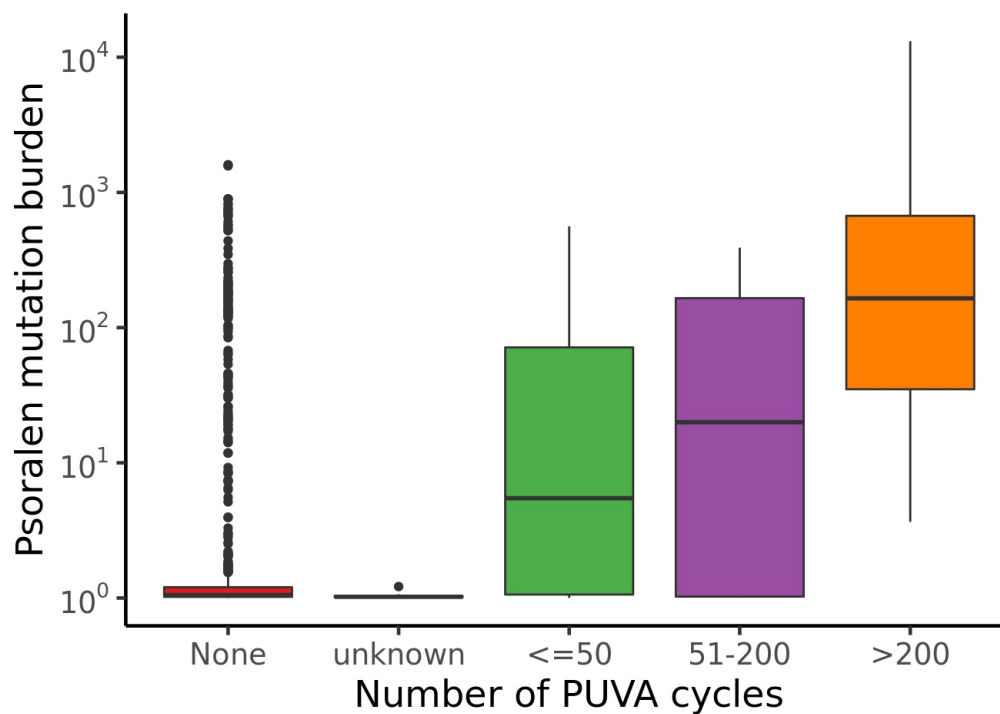
## PUVA

The mutation burden of psoralen-associated mutations seems to increase with the number of PUVA cycles the patient has received. However, there are also some individuals who show a high burden of the psoralen signature without any history of PUVA treatment.

```
ggplot(clone_burden, aes(x=AmountPuva, y=PUVA)) + geom_boxplot()
```



```
## Better to plot this on a log-scale
ggplot(clone_burden, aes(x=AmountPuva, y=PUVA+1, fill=AmountPuva)) +
  geom_boxplot() + theme_classic(base_size = BASESIZE) + scale_fill_brewer(palette="Set1") + theme(legend.position
= "none") +
  labs(x="Number of PUVA cycles", y="Psoralen mutation burden") +
  scale_y_log10(breaks = trans_breaks("log10", function(x) 10^x),
               labels = trans_format("log10", math_format(10^.x)))
```



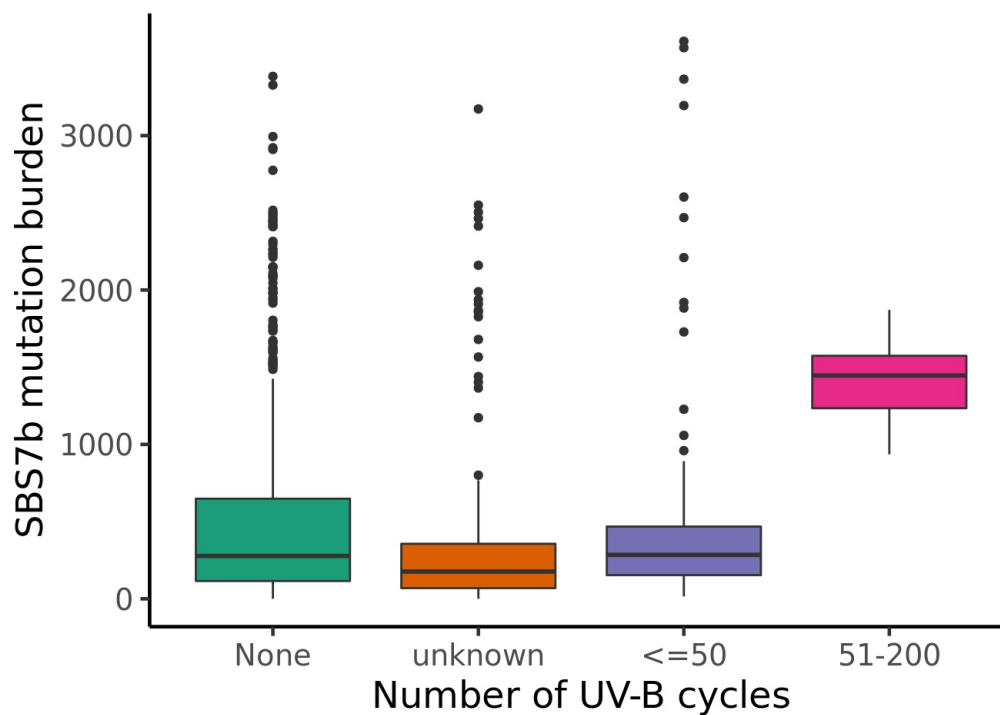
```
tapply(clone_burden$PUVA, clone_burden$AmountPuva, median, na.rm=T)
```

	None	unknown	<=50	51-200	>200
##	0.05257151	0.02121345	4.47220392	18.94403196	163.64534000

## UV-B

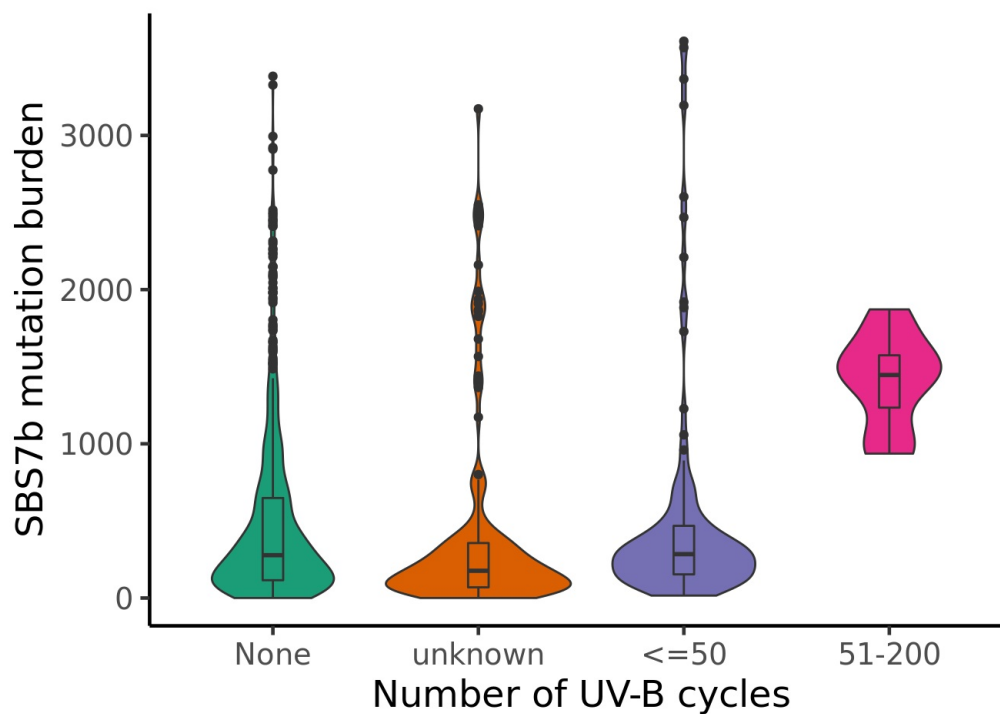
For UV-B, the story is less clear. Only a single individual has had >50 cycles of UV-B. This individual does show a higher-than-average burden of UV-related mutations but this is not sufficient for us to make any definitive statements about the mutagenic nature of this treatment.

```
ggplot(clone_burden, aes(x=AmountUVB, y=SBS7b, fill=AmountUVB))+
  geom_boxplot() + theme_classic(base_size = BASESIZE) + scale_fill_brewer(palette="Dark2") + theme(legend.position = "none") +
  labs(x="Number of UV-B cycles", y="SBS7b mutation burden")
```



```
ggplot(clone_burden, aes(x=AmountUVB, y=SBS7b, fill=AmountUVB)) + geom_violin() +
  geom_boxplot(width=0.1) + theme_classic(base_size = BASESIZE) + scale_fill_brewer(palette="Dark2") + theme(legend.position = "none") +
  labs(x="Number of UV-B cycles", y="SBS7b mutation burden")
```





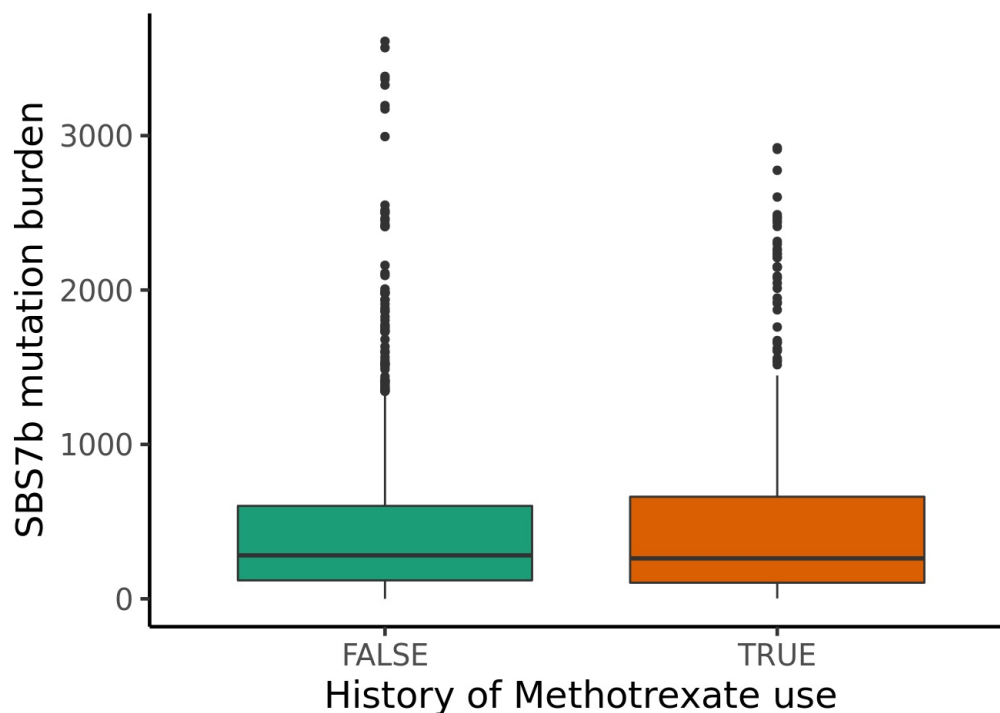
```
tapply(clone_burden$SBS7b, clone_burden$AmountUVB, median, na.rm=T)
```

```
##      None   unknown   <=50   51-200
## 277.6872 176.6552 284.5389 1446.3273
```

## Methotrexate

History of treatment with methotrexate is not associated with increased mutation burden, neither of the UV-related SBS7 nor of the clock-like SBS1/5.

```
ggplot(clone_burden, aes(x=EverMethotrexate, y=SBS7b, fill=EverMethotrexate))+
  geom_boxplot() + theme_classic(base_size = BASESIZE) + scale_fill_brewer(palette="Dark2") + theme(legend.position = "none") +
  labs(x="History of Methotrexate use", y="SBS7b mutation burden")
```



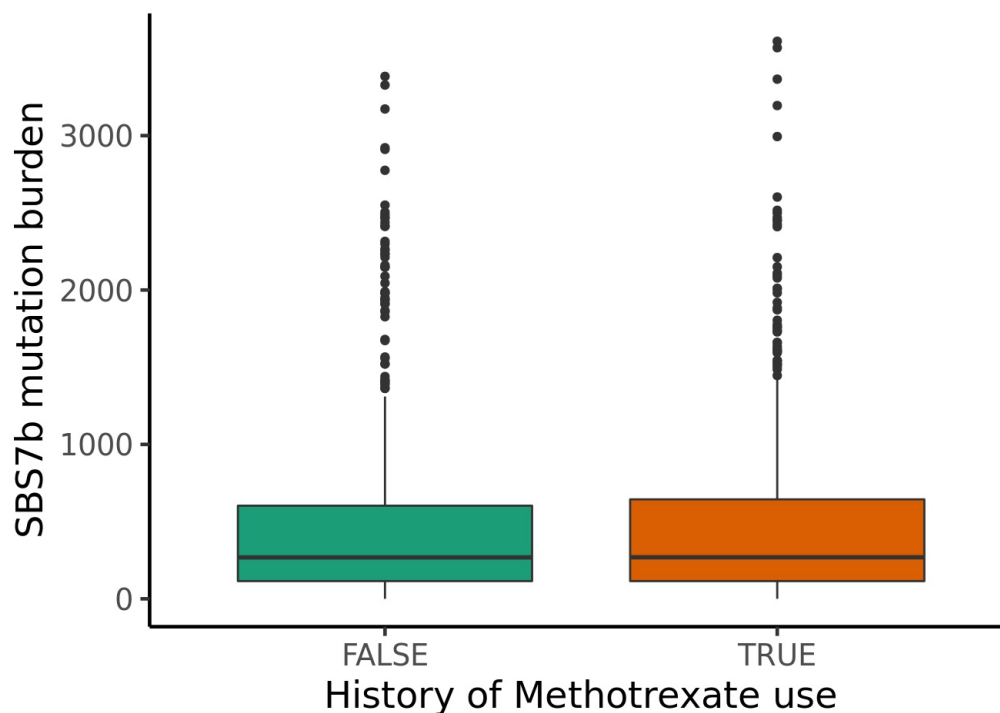
```
ggplot(clone_burden, aes(x=EverMethotrexate, y=SBS1.5, fill=EverMethotrexate))+
  geom_boxplot() + theme_classic(base_size = BASESIZE) + scale_fill_brewer(palette="Dark2") + theme(legend.position = "none") +
  labs(x="History of Methotrexate use", y="SBS1/5 mutation burden")
```



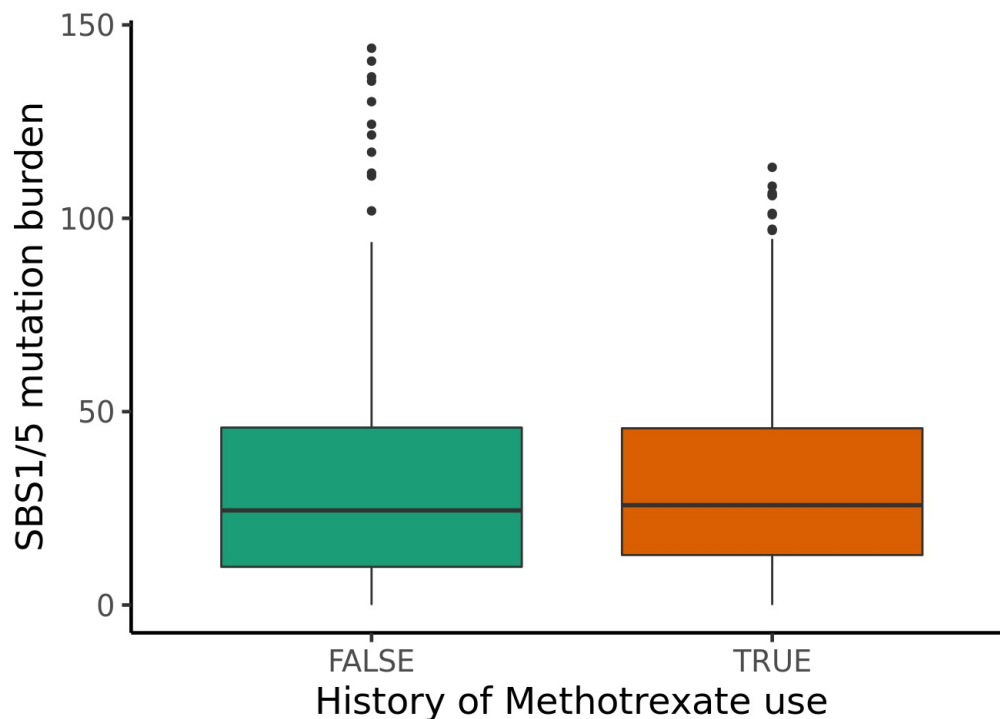
## Topical steroids

History of treatment with topical steroids is not associated with increased mutation burden, neither of the UV-related SBS7 nor of the clock-like SBS1/5.

```
ggplot(clone_burden, aes(x=EverSteroids, y=SBS7b, fill=EverSteroids))+
  geom_boxplot() + theme_classic(base_size = BASESIZE) + scale_fill_brewer(palette="Dark2") + theme(legend.position = "none") +
  labs(x="History of Methotrexate use", y="SBS7b mutation burden")
```



```
ggplot(clone_burden, aes(x=EverSteroids, y=SBS1.5, fill=EverSteroids))+
  geom_boxplot() + theme_classic(base_size = BASESIZE) + scale_fill_brewer(palette="Dark2") + theme(legend.position = "none") +
  labs(x="History of Methotrexate use", y="SBS1/5 mutation burden")
```



## Effects of treatments on clonal composition of the tissue

```
microd_meta <- merge(microd_meta, patient_meta[, c("Patient.ID", "Shows_Psoralen_signature", "EverPuva", "AmountPuva", "EverUVB", "AmountUVB", "EverMethotrexate", "EverSteroids")], by.x="PatientID", by.y="Patient.ID", all.x=T)

microd_meta$AmountUVB <- factor(microd_meta$AmountUVB, levels=c("None", "unknown", "<=50", "51-200"))
microd_meta$AmountPuva <- factor(microd_meta$AmountPuva, levels=c("None", "unknown", "<=50", "51-200", ">200"))

mVAFs <- data.frame(tapply(microd_meta$MedianVAF, microd_meta$PatientID, median, na.rm=T))
mVAFs$PatientID <- rownames(mVAFs)
colnames(mVAFs)[1] <- "MedianVAF"
patient_meta <- merge(patient_meta, mVAFs, by.x="Patient.ID", by.y="PatientID")
```

## PUVA

There are a few ways to look at the effects of treatment on the clonal structure of the tissue. We could do a simple Wilcoxon test to check for higher VAFs among microbiopsies that are dissected from patients who have a history of PUVA treatment.

```
puva_vaf_test <- wilcox.test(microd_meta$MedianVAF~microd_meta$EverPuva)
puva_vaf_test
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: microd_meta$MedianVAF by microd_meta$EverPuva
## W = 104968, p-value = 0.005047
## alternative hypothesis: true location shift is not equal to 0
```

However, the Wilcoxon test assumes independent sampling, which we haven't got. Samples taken from the same biopsy and from the same patient are likely to be correlated. Furthermore, we must take into account that patients who have received PUVA treatment are older than patients with no history of treatment. Older age often translates to higher VAF.

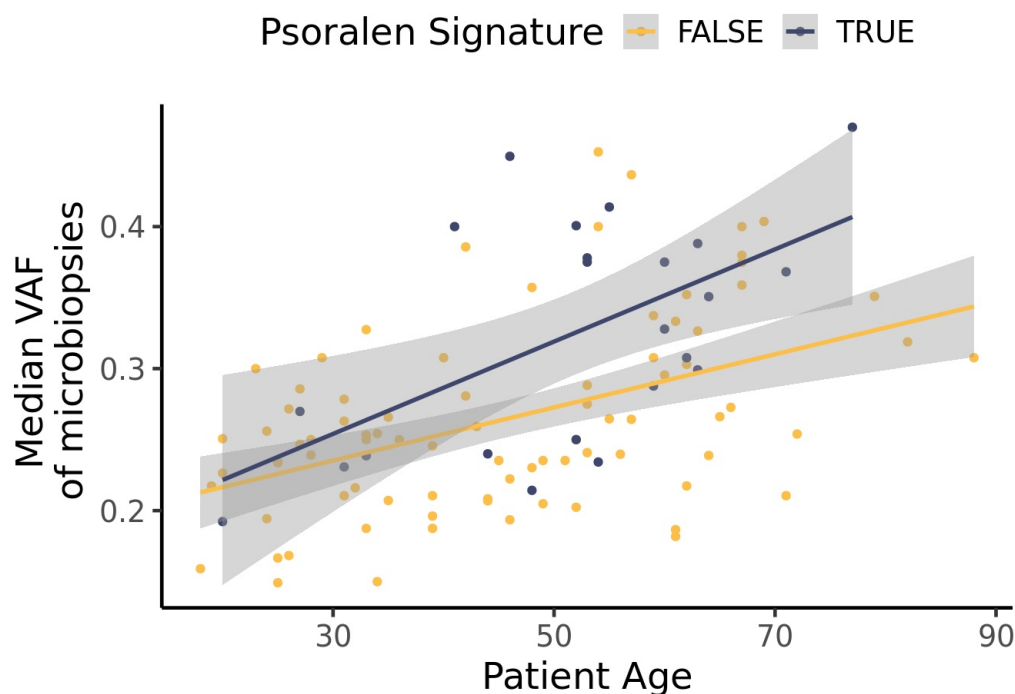
One way around this is to just take the median VAF of all the microbiopsies from the patient and regress this against the patient age and whether or not the psoralen signature is observed.

```
ggplot(patient_meta, aes(x=Age_at_sampling, y=MedianVAF, colour=Shows_Psoralen_signature)) + geom_point() + geom_smooth(method="lm") + theme_classic(base_size = 18) +
  labs(y="Median VAF \n of microbiopsies", x="Patient Age", colour="Psoralen Signature") +
  theme(legend.position = "top") + scale_colour_manual(values=c("#FCBF49", "#40476D"))
```

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

```
## Warning: Removed 5 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 5 rows containing missing values (geom_point).
```



```
summary(lm(MedianVAF~Age_at_sampling + Shows_Psoralen_signature, data=patient_meta))
```

```
##
## Call:
## lm(formula = MedianVAF ~ Age_at_sampling + Shows_Psoralen_signature,
##     data = patient_meta)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.114810 -0.047430  0.000359  0.039905  0.170611
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.1691106   0.0183757    9.203 4.37e-15 ***
## Age_at_sampling  0.0020905   0.0003737    5.593 1.84e-07 ***
## Shows_Psoralen_signatureTRUE 0.0473183   0.0147209    3.214 0.00175 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06175 on 103 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.3186, Adjusted R-squared:  0.3053
## F-statistic: 24.08 on 2 and 103 DF,  p-value: 2.639e-09
```

Finally, we could do the same kind of linear-mixed-effects modeling as we've done for the mutation burden. We can use a likelihood ratio test to see if including a binary variable for the presence of the psoralen signature improves the fit of the model. We see that it does.

```
microd_meta <- merge(microd_meta, patient_meta, by.x="PatientID", by.y="Patient.ID")

model_vaf.null <- lme(fixed = MedianVAF.x ~ Age_at_sampling,
                     random = list(PatientID = pdSymm(form = ~ Age_at_sampling - 1), BiopsyID = pdSymm(form = ~
Age_at_sampling - 1)),
                     data = microd_meta[!is.na(microd_meta$Age_at_sampling) & !is.na(microd_meta$MedianVAF.x)],
                     method="ML")

model_vaf.psoralen <- lme(fixed = MedianVAF.x ~ Age_at_sampling + Shows_Psoralen_signature.x,
                        random = list(PatientID = pdSymm(form = ~ Age_at_sampling - 1), BiopsyID = pdSymm(form = ~
Age_at_sampling - 1)),
                        data = microd_meta[!is.na(microd_meta$Age_at_sampling) & !is.na(microd_meta$MedianVAF.x)],
                        method="ML")

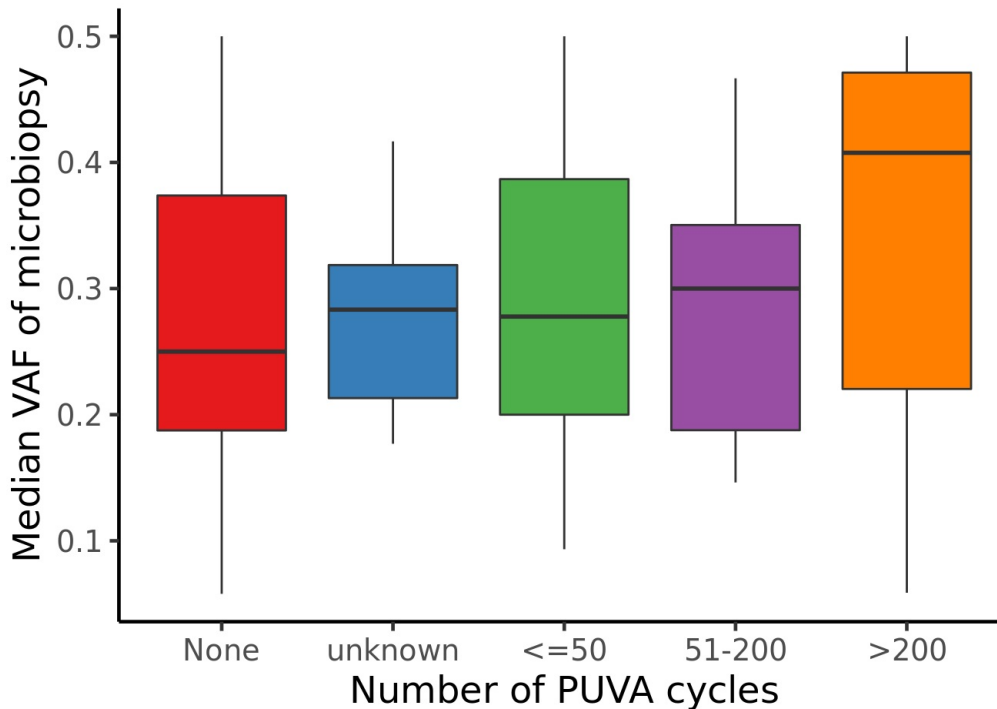
anova(model_vaf.null,model_vaf.psoralen, test=T)$"p-value"[2]
```

```
## [1] 0.02836863
```

The above has simply tested a binary variable for whether or not the psoralen signature is observed. We can see however, that there does also seem to be a trend for increased clonality with greater number of PUVA cycles. This has the caveat of potentially different age distributions, as discussed above.

```
ggplot(microd_meta[!is.na(microd_meta$AmountPuva.x),], aes(x=AmountPuva.x, y=MedianVAF.x, fill=AmountPuva.x)) +
  geom_boxplot() +
  theme_classic(base_size = 18) +
  scale_fill_brewer(palette="Set1") + theme(legend.position = "none") +
  labs(x="Number of PUVA cycles", y="Median VAF of microbiopsy")
```

```
## Warning: Removed 6 rows containing non-finite values (stat_boxplot).
```



## UVB

Is there evidence that treatment with UVB increases the VAFs of microbiopsies? It doesn't look like there is.

```
uvb_vaf_test <- wilcox.test(microd_meta$MedianVAF.x~microd_meta$EverUVB.x)
uvb_vaf_test
```

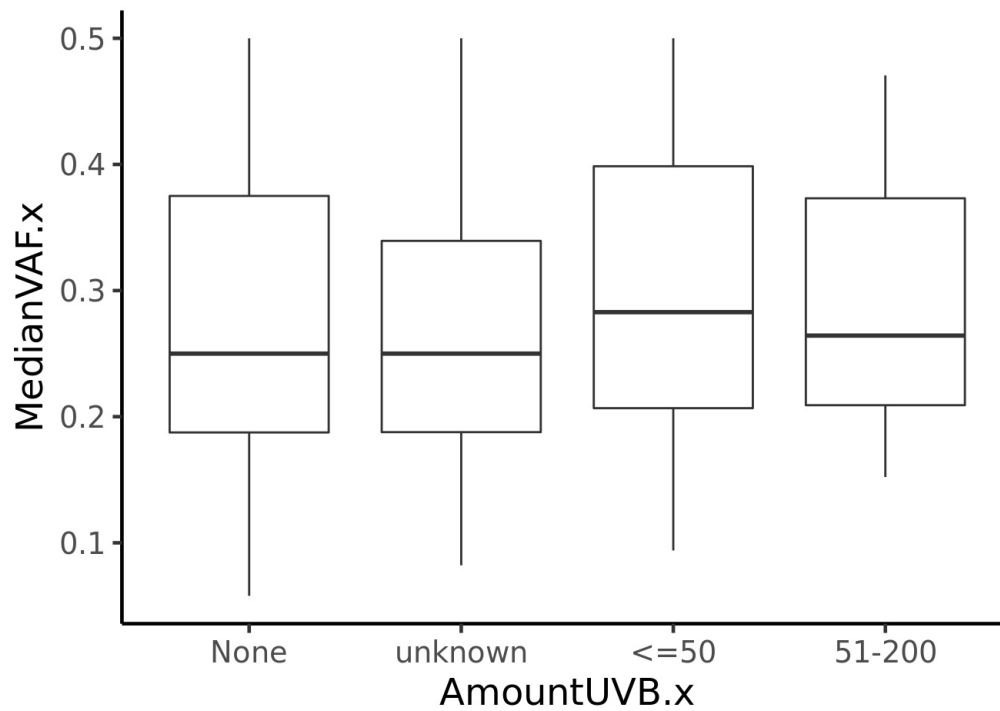
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: microd_meta$MedianVAF.x by microd_meta$EverUVB.x
## W = 127561, p-value = 0.6868
## alternative hypothesis: true location shift is not equal to 0
```

```
summary(lm(MedianVAF~Age_at_sampling + EverUVB, data=patient_meta))
```

```
##
## Call:
## lm(formula = MedianVAF ~ Age_at_sampling + EverUVB, data = patient_meta)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.124791 -0.047696 -0.002962  0.037881  0.177992
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.163842   0.020299   8.071 1.36e-12 ***
## Age_at_sampling 0.002340   0.000391   5.986 3.17e-08 ***
## EverUVBTRUE     0.015445   0.014754   1.047  0.298
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06444 on 103 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.2581, Adjusted R-squared:  0.2437
## F-statistic: 17.92 on 2 and 103 DF, p-value: 2.103e-07
```

```
ggplot(microd_meta[!is.na(microd_meta$AmountUVB.x),], aes(x=AmountUVB.x, y=MedianVAF.x)) + geom_boxplot() +  
  theme_classic(base_size = 18)
```

```
## Warning: Removed 6 rows containing non-finite values (stat_boxplot).
```



## Methotrexate and topical steroids

The analysis for methotrexate and topical steroids was done in the same way as shown above for PUVA and UVB, but there was no evidence that these treatments affect the clonal structure of the epidermis.