

Supplementary file_R script

Less-invasive age estimation using hair based on DNA methylation in brown bears

Shiori Nakamura, Jumpei Yamazaki, Naoya Matsumoto, Kyougo Hagino, Hideyuki Sakamoto, Masami Yamanaka, Masanao Nakanishi, Mina Jimbo, Yojiro Yanagawa, Hideyuki Ito, Toshio Tsubota, Michito Shimozuru

- [Load in packages](#)
- [Data input](#)
- [The correlation between methylation level and age](#)
- [Age estimation model 【Single regression】](#)
 - [Single regression \(VGF-3\)](#)
 - [Single regression \(KCNK12-3\)](#)
 - [Single regression \(ELOVL2-3\)](#)
- [Age estimation model 【Elastic net regression】](#)
 - [Elastic net regression \(VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-1, -2, -3\)](#)
- [Age estimation model 【Support vector regression】](#)
 - [Support vector regression \(VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-1, -2, -3\)](#)
 - [Support vector regression \(VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-2, -3\)](#)
 - [Support vector regression \(VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-3\)](#)
 - [Support vector regression \(VGF-1, -3, KCNK12-1, -2, -3, ELOVL2-3\)](#)
 - [Support vector regression \(VGF-1, -3, KCNK12-2, -3, ELOVL2-3\)](#)
 - [Support vector regression \(VGF-1, -3, KCNK12-2, -3\)](#)
 - [Support vector regression \(VGF-1, -3, KCNK12-3\)](#)
 - [Support vector regression \(VGF-3, KCNK12-3\)](#)
 - [Support vector regression \(VGF-1, -2, -3\)](#)
 - [Support vector regression \(KCNK12-1, -2, -3\)](#)
 - [Support vector regression \(ELOVL2-1, -2, -3\)](#)
- [Influences of interaction among age, sex, and growth environment](#)
 - [Support vector regression \(VGF-1, -3, KCNK12-3\)](#)
- [How to apply to the models](#)
 - [Single regression \(KCNK12-3\)](#)
 - [Elastic net regression \(VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-1, -2, -3\)](#)
 - [Support vector regression \(VGF-1, -3, KCNK12-3\)](#)
 - [Output to a csv file](#)

Load in packages

```
library(dplyr)
library(MuMIn)
library(glmnet)
library(e1071)
library(ggplot2)
library(car)
library(lme4)
```

Data input

```
BBH<-read.csv("brown_bear_hair.csv")
BBHS<-read.csv("brown_bear_hair_standardized.csv")
```

The correlation between methylation level and age

```
#VGF-1
cor.test(BBH$age,BBH$VGF_1_methylation_rate_ave)
```

Pearson's product-moment correlation

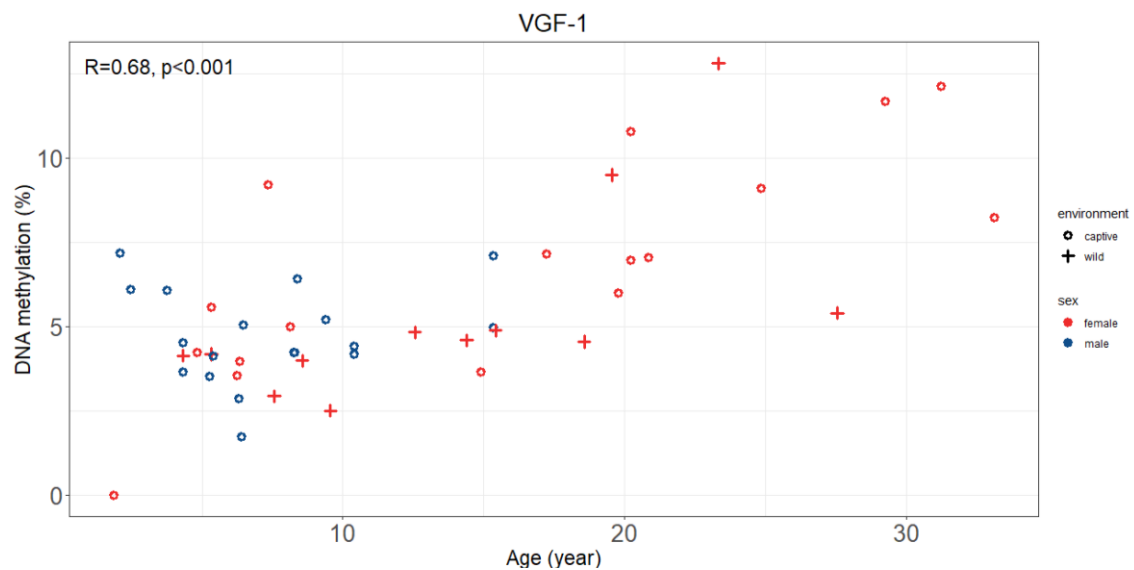
```
data: BBH$age and BBH$VGF_1_methylation_rate_ave
t = 6.249, df = 45, p-value = 1.332e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4904513 0.8101983
sample estimates:
      cor
0.6816181
```

```
VGF1<-ggplot(BBH,aes(x=age,y=VGF_1_methylation_rate_ave))+theme_bw()+
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
  labs(x="Age (year)",y="DNA methylation (%)")+
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1,"W"=3))+
```

```

scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick
2","M"="dodgerblue4"))+
theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
annotate("text",size=6,x=-Inf,y=Inf,hjust=-.1,vjust=2,label="R=0.68, p<0.001")+
labs (title="VGF-1")+
theme(plot.title=element_text(size=20,hjust = 0.5)) +
guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))

```



```

#VGF-2
cor.test(BBH$age,BBH$VGF_2_methylation_rate_ave)

```

Pearson's product-moment correlation

```

data: BBH$age and BBH$VGF_2_methylation_rate_ave
t = 7.0905, df = 45, p-value = 7.483e-09
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5550330 0.8386416
sample estimates:
cor
0.7264182

```

```

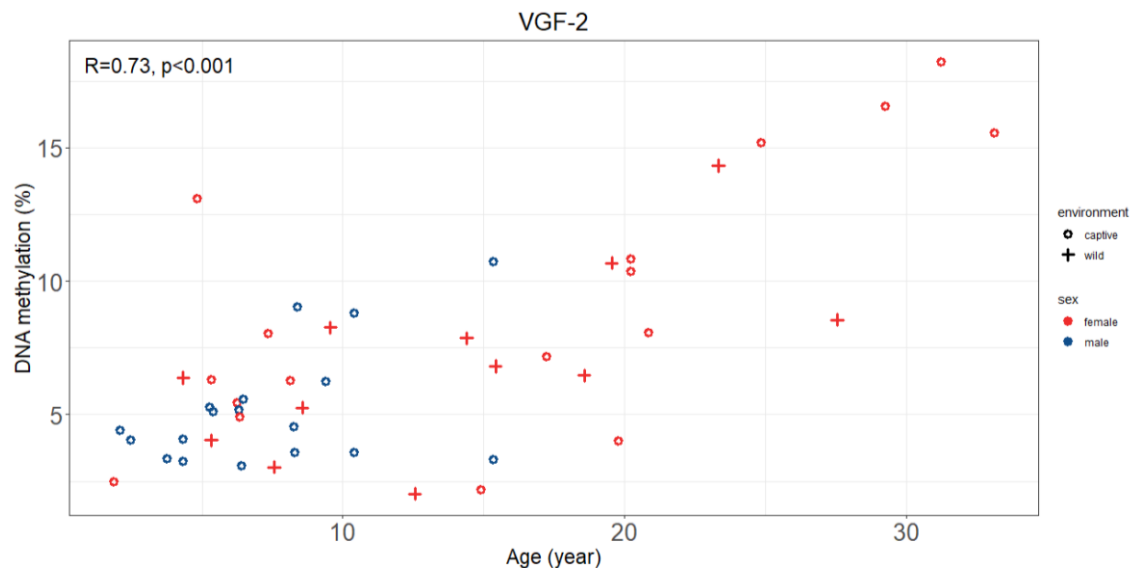
VGF2<-ggplot(BBH,aes(x=age,y=VGF_2_methylation_rate_ave))+theme_bw()+

```

```

geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
labs(x="Age (year)",y="DNA methylation (%)")+
scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1,"W"=3))+
scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick2","M"="dodgerblue4"))+
theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
annotate("text",size=6,x=-Inf,y=Inf,hjust=-.1,vjust=2,label="R=0.73, p<0.001")+
labs (title="VGF-2")+
theme(plot.title=element_text(size=20,hjust = 0.5)) +
guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))

```



```

#VGF-3
cor.test(BBH$age,BBH$VGF_3_methylation_rate_ave)

```

Pearson's product-moment correlation

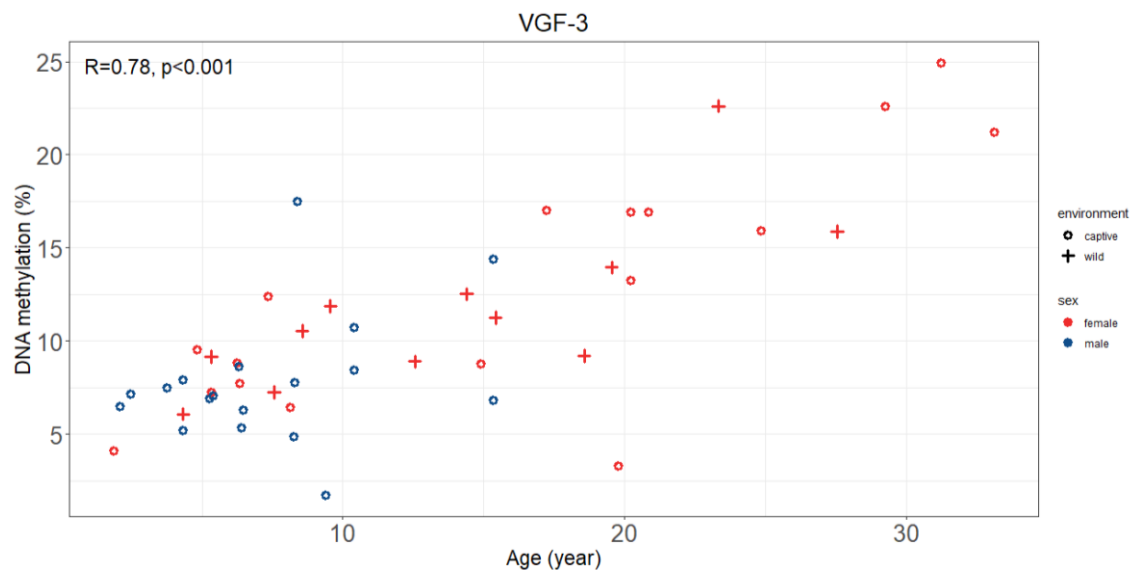
```

data: BBH$age and BBH$VGF_3_methylation_rate_ave
t = 8.4627, df = 45, p-value = 7.383e-11
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6406716 0.8741155
sample estimates:

```

```
cor  
0.7836608
```

```
VGf3<-ggplot(BBH,aes(x=age,y=VGf_3_methylation_rate_ave))+theme_bw()+  
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+  
  labs(x="Age (year)",y="DNA methylation (%)")+  
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1  
  ,"W"=3))+  
  scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick  
  2","M"="dodgerblue4"))+  
  theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+  
  theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+  
  annotate("text",size=6,x=-Inf,y=Inf,hjust=-.1,vjust=2,label="R=0.78, p<0.001")+  
  labs (title="VGf-3")+  
  theme(plot.title=element_text(size=20,hjust = 0.5)) +  
  guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))
```



```
#KCNK12-1  
cor.test(BBH$age,BBH$KCNK12_1_methylation_rate_ave)
```

Pearson's product-moment correlation

```
data: BBH$age and BBH$KCNK12_1_methylation_rate_ave  
t = 7.7282, df = 45, p-value = 8.597e-10
```

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

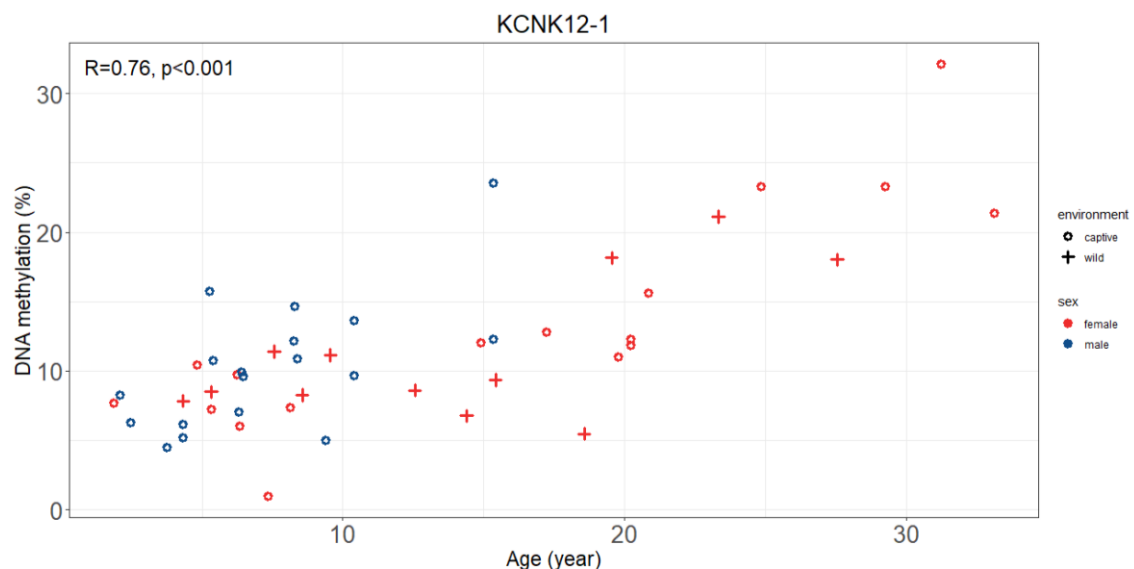
0.5976185 0.8565876

sample estimates:

cor

0.7551836

```
KCNK1<-ggplot(BBH,aes(x=age,y=KCNK12_1_methylation_rate_ave))+theme_bw()+
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
  labs(x="Age (year)",y="DNA methylation (%)")+
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1
,"W"=3))+
  scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick
2","M"="dodgerblue4"))+
  theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
  theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
  annotate("text",size=6,x=-Inf,y=Inf,hjust=-.1,vjust=2,label="R=0.76, p<0.001")+
  labs (title="KCNK12-1")+
  theme(plot.title=element_text(size=20,hjust = 0.5)) +
  guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))
```



```
#KCNK12-2
```

```
cor.test(BBH$age,BBH$KCNK12_2_methylation_rate_ave)
```

Pearson's product-moment correlation

data: BBH\$age and BBH\$KCNK12_2_methylation_rate_ave

t = 6.683, df = 45, p-value = 3.012e-08

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

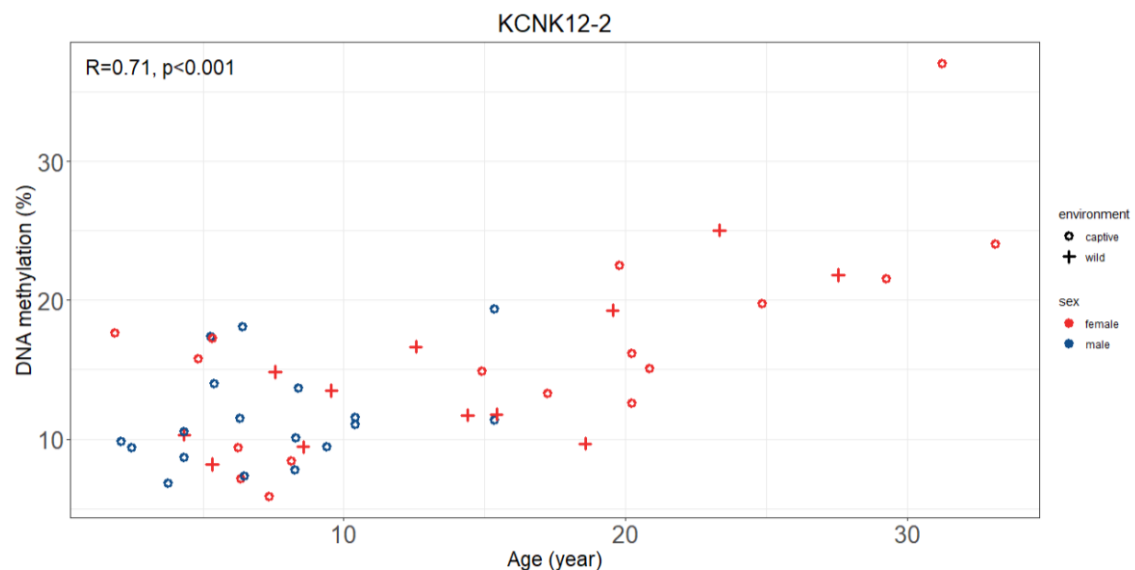
0.5250188 0.8256123

sample estimates:

cor

0.705777

```
KCNK2<-ggplot(BBH,aes(x=age,y=KCNK12_2_methylation_rate_ave))+theme_bw()+
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
  labs(x="Age (year)",y="DNA methylation (%)")+
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1
,"W"=3))+
  scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick
2","M"="dodgerblue4"))+
  theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
  theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
  annotate("text",size=6,x=-Inf,y=Inf,hjust=-.1,vjust=2,label="R=0.71, p<0.001")+
  labs (title="KCNK12-2")+
  theme(plot.title=element_text(size=20,hjust = 0.5)) +
  guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))
```

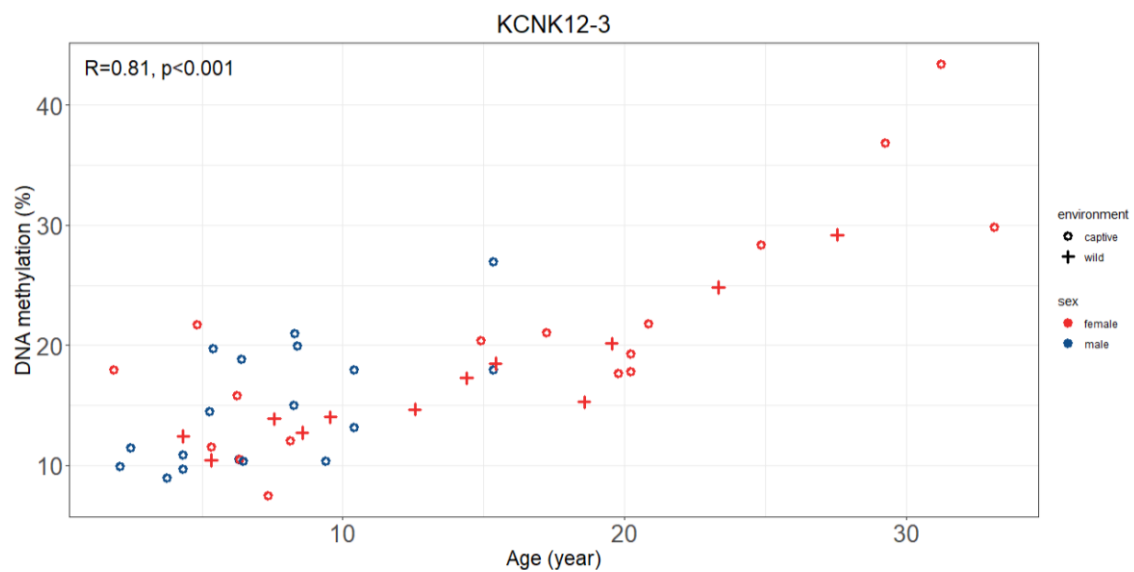


```
#KCNK12-3
cor.test(BBH$age,BBH$KCNK12_3_methylation_rate_ave)
```

Pearson's product-moment correlation

```
data: BBH$age and BBH$KCNK12_3_methylation_rate_ave
t = 9.2591, df = 45, p-value = 5.478e-12
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6810019 0.8900012
sample estimates:
      cor
0.8098028
```

```
KCNK3<-ggplot(BBH,aes(x=age,y=KCNK12_3_methylation_rate_ave))+theme_bw()+
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
  labs(x="Age (year)",y="DNA methylation (%)")+
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1
,"W"=3))+
  scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick
2","M"="dodgerblue4"))+
  theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
  theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
  annotate("text",size=6,x=-Inf,y=Inf,hjust=-.1,vjust=2,label="R=0.81, p<0.001")+
  labs (title="KCNK12-3")+
  theme(plot.title=element_text(size=20,hjust = 0.5)) +
  guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))
```

```
#ELOVL2-1
cor.test(BBH$age,BBH$ELOVL2_1_methylation_rate_ave)
```

Pearson's product-moment correlation

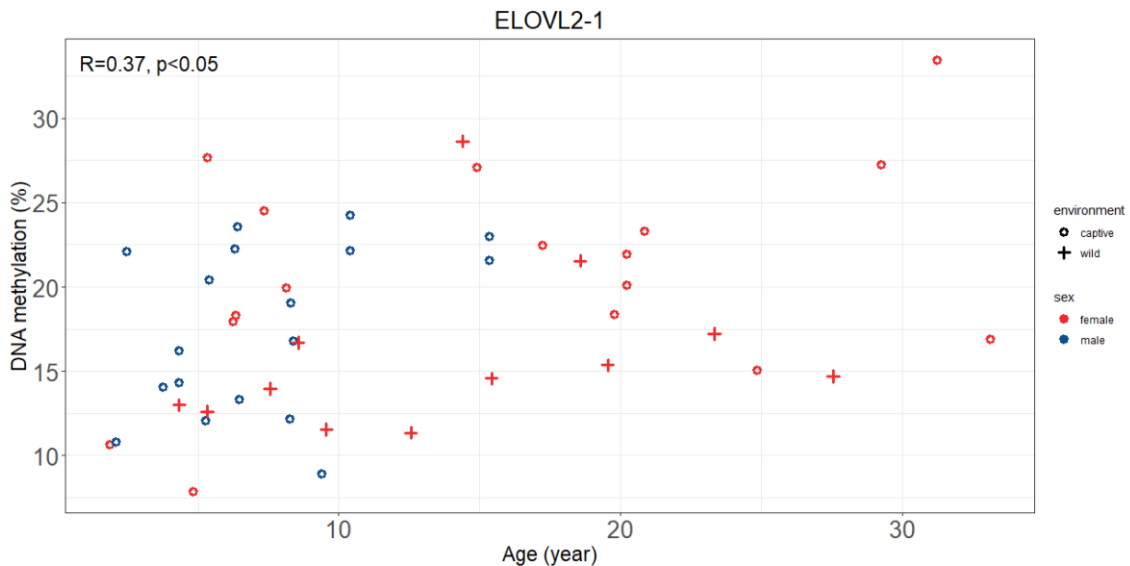
```
data: BBH$age and BBH$ELOVL2_1_methylation_rate_ave
t = 2.6801, df = 45, p-value = 0.01025
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.09384802 0.59480967
sample estimates:
cor
0.3710159
```

```
ELOVL1<-ggplot(BBH,aes(x=age,y=ELOVL2_1_methylation_rate_ave))+theme_bw()+
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
  labs(x="Age (year)",y="DNA methylation (%)")+
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1,"W"=3))+
  scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick2","M"="dodgerblue4"))+
  theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
  theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
```

```

annotate("text",size=6,x=-Inf,y=Inf,hjust=-.1,vjust=2,label="R=0.37, p<0.05")+
labs (title="ELOVL2-1")+
theme(plot.title=element_text(size=20,hjust = 0.5)) +
guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))

```



```

#ELOVL2-2
cor.test(BBH$age,BBH$ELOVL2_2_methylation_rate_ave)

```

Pearson's product-moment correlation

```

data: BBH$age and BBH$ELOVL2_2_methylation_rate_ave
t = 3.4533, df = 45, p-value = 0.001217
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1963429 0.6583395
sample estimates:
cor
0.4577029

```

```

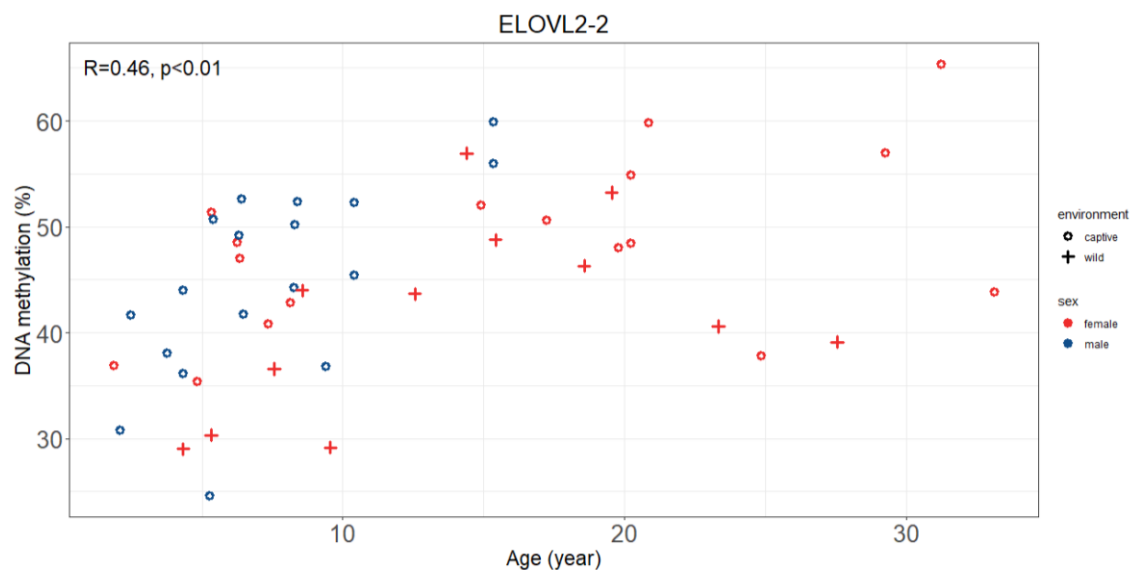
ELOVL2<-ggplot(BBH,aes(x=age,y=ELOVL2_2_methylation_rate_ave))+theme_bw()+
geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
labs(x="Age (year)",y="DNA methylation (%)")+
scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1,"W"=3))+

```

```

scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick
2","M"="dodgerblue4"))+
theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
annotate("text",size=6,x=-Inf,y=Inf,hjust=-.1,vjust=2,label="R=0.46, p<0.01")+
labs (title="ELOVL2-2")+
theme(plot.title=element_text(size=20,hjust = 0.5)) +
guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))

```



```

#ELOVL2-3
cor.test(BBH$age,BBH$ELOVL2_3_methylation_rate_ave)

```

Pearson's product-moment correlation

```

data: BBH$age and BBH$ELOVL2_3_methylation_rate_ave
t = 4.0671, df = 45, p-value = 0.0001892
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2717354 0.7012156
sample estimates:
cor
0.5184461

```

```

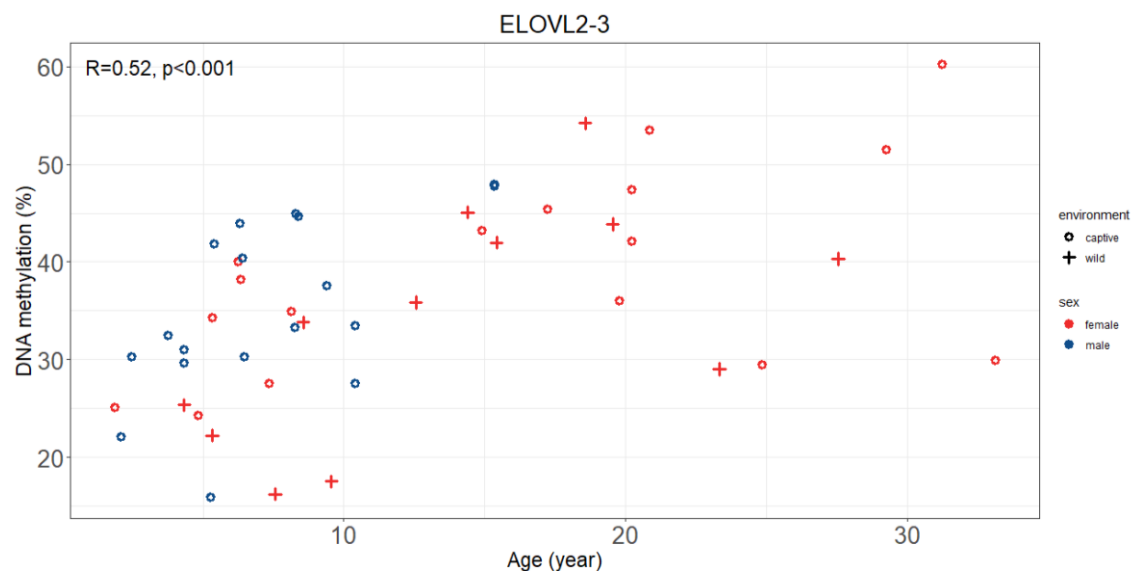
ELOVL3<-ggplot(BBH,aes(x=age,y=ELOVL2_3_methylation_rate_ave))+theme_bw()+

```

```

geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
labs(x="Age (year)",y="DNA methylation (%)")+
scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1,"W"=3))+
scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick2","M"="dodgerblue4"))+
theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
annotate("text",size=6,x=-Inf,y=Inf,hjust=-.1,vjust=2,label="R=0.52, p<0.001")+
labs (title="ELOVL2-3")+
theme(plot.title=element_text(size=20,hjust = 0.5)) +
guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))

```



Age estimation model 【Single regression】

```

YS <- BBHS$age
vgf1S <- BBHS$VGF_1_methylation_rate_ave
vgf2S <- BBHS$VGF_2_methylation_rate_ave
vgf3S <- BBHS$VGF_3_methylation_rate_ave
kcnk1S <- BBHS$KCNK12_1_methylation_rate_ave
kcnk2S <- BBHS$KCNK12_2_methylation_rate_ave
kcnk3S <- BBHS$KCNK12_3_methylation_rate_ave
elovl1S <- BBHS$ELOVL2_1_methylation_rate_ave
elovl2S <- BBHS$ELOVL2_2_methylation_rate_ave

```

```
elovl3S <- BBHS$ELOVL2_3_methylation_rate_ave
AGE <- BBH$age
```

```
#RemoveOne function
removeOne <- function(dat,x) {
  if(x<dat){
    list=seq(1,dat)
    x1=x-1;x2=x+1
    v1=c(list[0:x1]);v2=c(list[x2:dat])
    data=c(v1,v2)}
  else {data=seq(1,dat-1)}
  return (data)}
```

Single regression (VGF-3)

```
SRM_VGF_3<-lm(formula=YS~vgf3S,data=BBHS)
coef(SRM_VGF_3)
```

```
(Intercept)      vgf3S
2.127681e-11 7.836608e-01
```

```
#LOOCV
nSamples<-nrow(BBHS)
predict_SRM_VGF_3_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-data.frame(BBHS$age[indices],BBHS$VGF_3_methylation_rate_ave[indices])
  colnames(dr)<-c("age","methylvgf_3")
  bestmodel_SRM_VGF_3_loocv<-lm(age~methylvgf_3, data=dr)
  newdata<-data.frame(methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z])
  p<-predict(bestmodel_SRM_VGF_3_loocv,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SRM_VGF_3_loocv[z]<-p}

BBH_predict_SRM_VGF_3_loocv<-cbind(BBH,predict_SRM_VGF_3_loocv)
MAE_SRM_VGF_3_loocv<-mean(abs(BBH_predict_SRM_VGF_3_loocv$predict_SRM_VGF_3_loocv-
BBH_predict_SRM_VGF_3_loocv$age))
```

```

MedianAE_SRM_VGF_3_loocv<-
median(abs(BBH_predict_SRM_VGF_3_loocv$predict_SRM_VGF_3_loocv-
BBH_predict_SRM_VGF_3_loocv$age))
RMSE_SRM_VGF_3_loocv<-sqrt(mean((BBH_predict_SRM_VGF_3_loocv$predict_SRM_VGF_3_loocv-
BBH_predict_SRM_VGF_3_loocv$age)^2))
cat("MAE:", MAE_SRM_VGF_3_loocv, "\nMed AE:", MedianAE_SRM_VGF_3_loocv, "\nRMSE:",
RMSE_SRM_VGF_3_loocv, "\n")

```

MAE: 4.137939

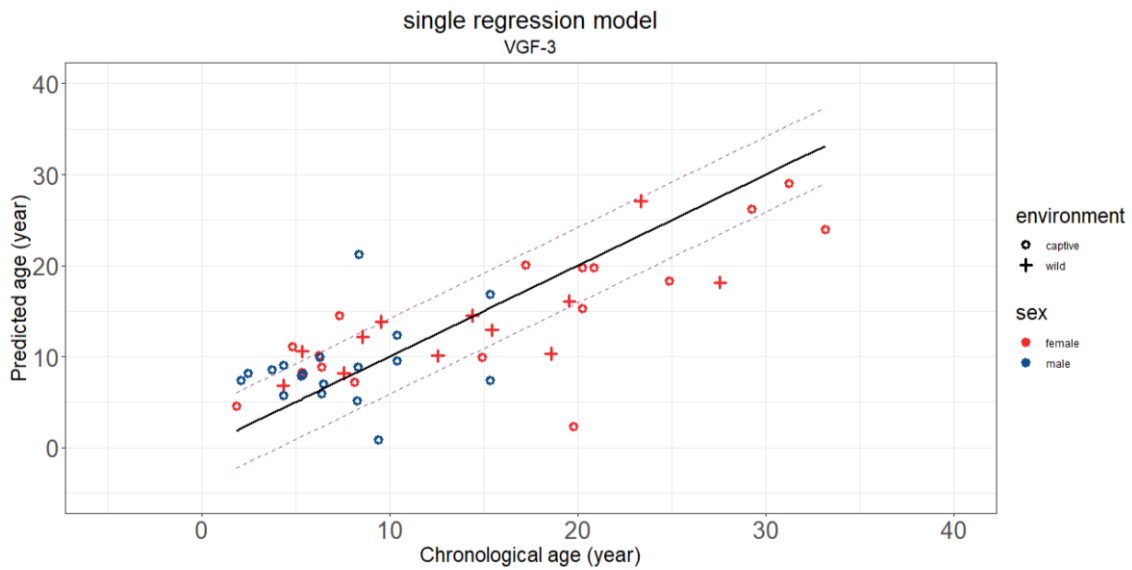
Med AE: 3.127337

RMSE: 5.353131

```

g_SRM_VGF_3_loocv<-
ggplot(BBH_predict_SRM_VGF_3_loocv,aes(age,predict_SRM_VGF_3_loocv))+theme_bw()+
  annotate("segment",x=min(BBH$age),xend=max(BBH$age),y=min(BBH$age)+4.137939,yend=max(BBH$age)+4.137939,colour="orchid4",linetype=2, linewidth =0.7)+
  annotate("segment",x=min(BBH$age),xend=max(BBH$age),y=min(BBH$age)-4.137939,yend=max(BBH$age)-4.137939,colour="orchid4",linetype=2, linewidth =0.7)+
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
  labs(x="Chronological age (year)",y="Predicted age (year)")+
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1,"W"=3))+
  scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick2","M"="dodgerblue4"))+
  theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
  theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
  geom_line(aes(y =age), linewidth=1)+
  labs(title="single regression model")+
  theme(title=element_text(size=17),plot.title=element_text(hjust=0.5))+
  scale_y_continuous(limits=c(-5,40))+
  scale_x_continuous(limits=c(-5,40))+
  labs(subtitle="VGF-3")+
  theme(plot.subtitle=element_text(size=15,hjust=0.5))+
  guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))

```



Single regression (KCNK12-3)

```
SRM_KCNK_3<-lm(formula=YS~kcnk3S,data=BBHS)
```

```
coef(SRM_KCNK_3)
```

```
(Intercept)      kcnk3S
-1.318289e-11  8.098028e-01
```

```
#LOOCV
nSamples<-nrow(BBHS)
predict_SRM_KCNK_3_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-data.frame(BBHS$age[indices],BBHS$KCNK12_3_methylation_rate_ave[indices])
  colnames(dr)<-c("age","methylnkcnk_3")
  bestmodel_SRM_KCNK_3_loocv<-lm(age~methylnkcnk_3, data=dr)
  newdata<-data.frame(methylnkcnk_3=BBHS$KCNK12_3_methylation_rate_ave[z])
  p<-predict(bestmodel_SRM_KCNK_3_loocv,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SRM_KCNK_3_loocv[z]<-p}

BBH_predict_SRM_KCNK_3_loocv<-cbind(BBH,predict_SRM_KCNK_3_loocv)
MAE_SRM_KCNK_3_loocv<-mean(abs(BBH_predict_SRM_KCNK_3_loocv$predict_SRM_KCNK_3_loocv-
BBH_predict_SRM_KCNK_3_loocv$age))
```

```

MedianAE_SRM_KCNK_3_loocv<-
median(abs(BBH_predict_SRM_KCNK_3_loocv$predict_SRM_KCNK_3_loocv-
BBH_predict_SRM_KCNK_3_loocv$age))
RMSE_SRM_KCNK_3_loocv<-sqrt(mean((BBH_predict_SRM_KCNK_3_loocv$predict_SRM_KCNK_3_loocv-
BBH_predict_SRM_KCNK_3_loocv$age)^2))
cat("MAE:", MAE_SRM_KCNK_3_loocv, "\nMed AE:", MedianAE_SRM_KCNK_3_loocv, "\nRMSE:",
RMSE_SRM_KCNK_3_loocv, "\n")

```

MAE: 3.97837

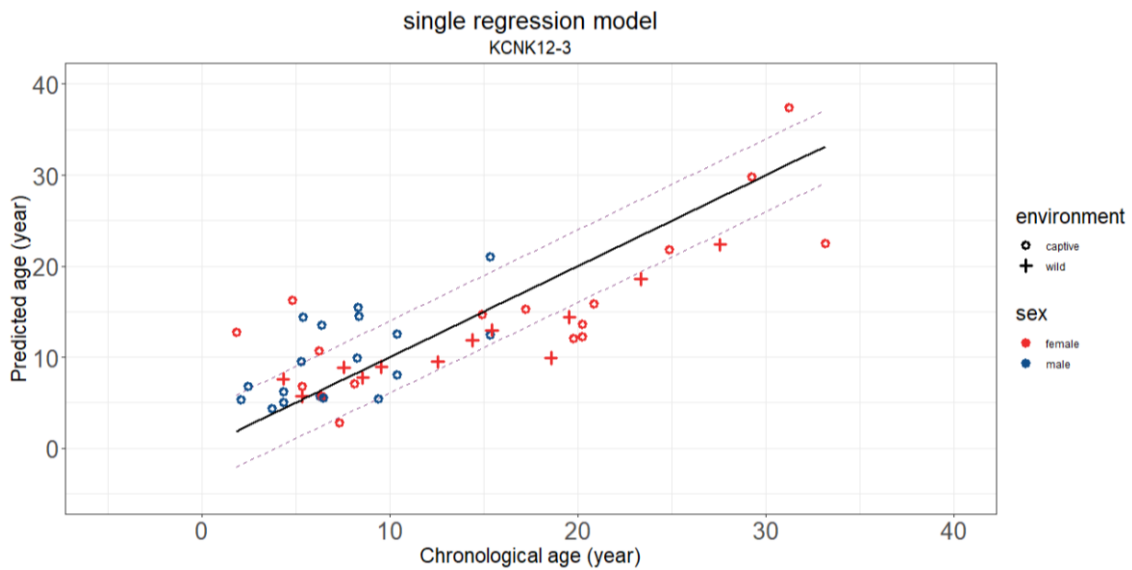
Med AE: 3.26955

RMSE: 5.010119

```

g_SRM_KCNK_3_loocv<-
ggplot(BBH_predict_SRM_KCNK_3_loocv,aes(age,predict_SRM_KCNK_3_loocv))+theme_bw()+
  annotate("segment",x=min(BBH$age),xend=max(BBH$age),y=min(BBH$age)+3.97837,yend=max(BBH
$age)+3.97837,colour="orchid4",linetype=2, linewidth =0.7)+
  annotate("segment",x=min(BBH$age),xend=max(BBH$age),y=min(BBH$age)-
3.97837,yend=max(BBH$age)-3.97837,colour="orchid4",linetype=2, linewidth =0.7)+
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
  labs(x="Chronological age (year)",y="Predicted age (year)")+
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1
,"W"=3))+
  scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick
2","M"="dodgerblue4"))+
  theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
  theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
  geom_line(aes(y =age), linewidth=1)+
  labs(title="single regression model")+
  theme(title=element_text(size=17),plot.title=element_text(hjust=0.5))+
  scale_y_continuous(limits=c(-5,40))+
  scale_x_continuous(limits=c(-5,40))+
  labs(subtitle="KCNK12-3")+
  theme(plot.subtitle=element_text(size=15,hjust=0.5))+
  guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))

```

Single regression (ELOVL2-3)

```
SRM_ELOVL_3<-lm(formula=YS~elovl3S,data=BBHS)
```

```
coef(SRM_ELOVL_3)
```

```
(Intercept)      elovl3S
1.024588e-11 5.184461e-01
```

```
#LOOCV
nSamples<-nrow(BBHS)
predict_SRM_ELOVL_3_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-data.frame(BBHS$age[indices],BBHS$ELOVL2_3_methylation_rate_ave[indices])
  colnames(dr)<-c("age","methylelov1_3")
  bestmodel_SRM_ELOVL_3_loocv<-lm(age~methylelov1_3, data=dr)
  newdata<-data.frame(methylelov1_3=BBHS$ELOVL2_3_methylation_rate_ave[z])
  p<-predict(bestmodel_SRM_ELOVL_3_loocv,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SRM_ELOVL_3_loocv[z]<-p}

BBH_predict_SRM_ELOVL_3_loocv<-cbind(BBH,predict_SRM_ELOVL_3_loocv)
MAE_SRM_ELOVL_3_loocv<-mean(abs(BBH_predict_SRM_ELOVL_3_loocv$predict_SRM_ELOVL_3_loocv-
BBH_predict_SRM_ELOVL_3_loocv$age))
```

```

MedianAE_SRM_ELOVL_3_loocv<-
median(abs(BBH_predict_SRM_ELOVL_3_loocv$predict_SRM_ELOVL_3_loocv-
BBH_predict_SRM_ELOVL_3_loocv$age))
RMSE_SRM_ELOVL_3_loocv<-
sqrt(mean((BBH_predict_SRM_ELOVL_3_loocv$predict_SRM_ELOVL_3_loocv-
BBH_predict_SRM_ELOVL_3_loocv$age)^2))
cat("MAE:", MAE_SRM_ELOVL_3_loocv, "\nMed AE:", MedianAE_SRM_ELOVL_3_loocv, "\nRMSE:",
RMSE_SRM_ELOVL_3_loocv, "\n")

```

MAE: 5.524564

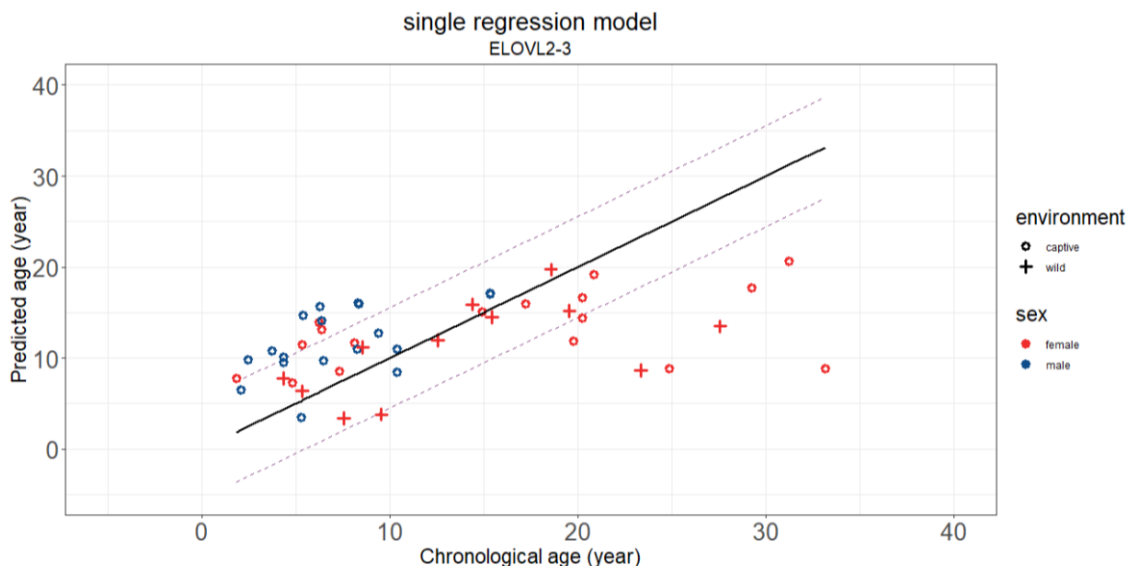
Med AE: 4.359233

RMSE: 7.278303

```

g_SRM_ELOVL_3_loocv<-
ggplot(BBH_predict_SRM_ELOVL_3_loocv,aes(age,predict_SRM_ELOVL_3_loocv))+theme_bw()+
  annotate("segment",x=min(BBH$age),xend=max(BBH$age),y=min(BBH$age)+5.524564,yend=max(BBH$age)+5.524564,colour="orchid4",linetype=2, linewidth =0.7)+
  annotate("segment",x=min(BBH$age),xend=max(BBH$age),y=min(BBH$age)-5.524564,yend=max(BBH$age)-5.524564,colour="orchid4",linetype=2, linewidth =0.7)+
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
  labs(x="Chronological age (year)",y="Predicted age (year)")+
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1,"W"=3))+
  scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick2","M"="dodgerblue4"))+
  theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
  theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
  geom_line(aes(y =age), linewidth=1)+
  labs(title="single regression model")+
  theme(title=element_text(size=17),plot.title=element_text(hjust=0.5))+
  scale_y_continuous(limits=c(-5,40))+
  scale_x_continuous(limits=c(-5,40))+
  labs(subtitle="ELOVL2-3")+
  theme(plot.subtitle=element_text(size=15,hjust=0.5))+
  guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))

```



Age estimation model 【Elastic net regression】

```
YS <- BBHS$age
vgf1S <- BBHS$VGF_1_methylation_rate_ave
vgf2S <- BBHS$VGF_2_methylation_rate_ave
vgf3S <- BBHS$VGF_3_methylation_rate_ave
kcnk1S <- BBHS$KCNK12_1_methylation_rate_ave
kcnk2S <- BBHS$KCNK12_2_methylation_rate_ave
kcnk3S <- BBHS$KCNK12_3_methylation_rate_ave
elovl1S <- BBHS$ELOVL2_1_methylation_rate_ave
elovl2S <- BBHS$ELOVL2_2_methylation_rate_ave
elovl3S <- BBHS$ELOVL2_3_methylation_rate_ave
AGE <- BBH$age
```

#RemoveOne function

```
removeOne <- function(dat,x) {
  if(x<dat){
    list=seq(1,dat)
    x1=x-1;x2=x+1
    v1=c(list[0:x1]);v2=c(list[x2:dat])
    data=c(v1,v2)}
  else {data=seq(1,dat-1)}
  return (data)}
```

Elastic net regression (VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-1, -2, -3)

```
XS <- cbind(vgf1S,vgf2S,vgf3S,kcnk1S,kcnk2S,kcnk3S,elovl1S,elovl2S,elovl3S)
alpha <- seq(0.01, 0.99, 0.01)
mse.df <- NULL
for (i in 1:length(alpha)) {
  m <- cv.glmnet(x = XS, y = YS, family = "gaussian", alpha = alpha[i], standardize =
FALSE)
  mse.df <- rbind(mse.df, data.frame(alpha = alpha[i], mse = min(m$cvm)))
}
best.alpha <- mse.df$alpha[mse.df$mse == min(mse.df$mse)]
m <- cv.glmnet(x = XS, y = YS, family = "gaussian", alpha = best.alpha, standardize =
FALSE)
best.lambda <- m$lambda.min
##alpha= 0.04, lambda= 0.4369317

ENM <- glmnet(x = XS, y = YS, family = "gaussian", lambda = best.lambda, alpha =
best.alpha, standardize = FALSE)
coef(ENM,s=best.lambda)
```

10 x 1 sparse Matrix of class "dgCMatrix"

```
          s1
(Intercept) 1.240302e-11
vgf1S      1.346730e-01
vgf2S      9.775300e-02
vgf3S      1.602995e-01
kcnk1S      1.212136e-01
kcnk2S      1.312119e-01
kcnk3S      1.564545e-01
elovl1S      .
elovl2S      2.685446e-02
elovl3S      1.310693e-01
```

```
ENM1 <- glmnet(x = XS, y = YS, family = "gaussian",
               lambda = 0.4369317, alpha = 0.04, standardize = FALSE)
```

```

#LOOCV
nSamples<-nrow(BBHS)
predict_ENM_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-
data.frame(BBHS$age[indices],BBHS$VGF_1_methylation_rate_ave[indices],BBHS$VGF_2_methylation_rate_ave[indices],BBHS$VGF_3_methylation_rate_ave[indices],BBHS$KCNK12_1_methylation_rate_ave[indices],BBHS$KCNK12_2_methylation_rate_ave[indices],BBHS$KCNK12_3_methylation_rate_ave[indices],BBHS$ELOVL2_1_methylation_rate_ave[indices],BBHS$ELOVL2_2_methylation_rate_ave[indices],BBHS$ELOVL2_3_methylation_rate_ave[indices])
  colnames(dr)<-
c("age","methylvgf_1","methylvgf_2","methylvgf_3","methylkcnk_1","methylkcnk_2","methylkcnk_3","methylelovl_1","methylelovl_2","methylelovl_3")
  X1 <-
cbind(BBHS$VGF_1_methylation_rate_ave[indices],BBHS$VGF_2_methylation_rate_ave[indices],BBHS$VGF_3_methylation_rate_ave[indices],BBHS$KCNK12_1_methylation_rate_ave[indices],BBHS$KCNK12_2_methylation_rate_ave[indices],BBHS$KCNK12_3_methylation_rate_ave[indices],BBHS$ELOVL2_1_methylation_rate_ave[indices],BBHS$ELOVL2_2_methylation_rate_ave[indices],BBHS$ELOVL2_3_methylation_rate_ave[indices])
  Y1 <- BBHS$age[indices]
  ENM2 <- glmnet(x = X1, y = Y1, family = "gaussian", lambda = 0.4369317, alpha = 0.04,
standardize = FALSE)
  Xnew <-
cbind(methylvgf_1=BBHS$VGF_1_methylation_rate_ave[z],methylvgf_2=BBHS$VGF_2_methylation_rate_ave[z],methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z],methylkcnk_1=BBHS$KCNK12_1_methylation_rate_ave[z],methylkcnk_2=BBHS$KCNK12_2_methylation_rate_ave[z],methylkcnk_3=BBHS$KCNK12_3_methylation_rate_ave[z],methylelovl_1=BBHS$ELOVL2_1_methylation_rate_ave[z],methylelovl_2=BBHS$ELOVL2_2_methylation_rate_ave[z],methylelovl_3=BBHS$ELOVL2_3_methylation_rate_ave[z])
  p<-predict(ENM2,Xnew,s=0.4369317) *sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_ENM_loocv[z]<-p}

BBH_predict_ENM_loocv<-cbind(BBH,predict_ENM_loocv)

```

```

MAE_ENM_loocv<-mean(abs(BBH_predict_ENM_loocv$predict_ENM_loocv-
BBH_predict_ENM_loocv$age))
MedianAE_ENM_loocv<-median(abs(BBH_predict_ENM_loocv$predict_ENM_loocv-
BBH_predict_ENM_loocv$age))
RMSE_ENM_loocv<- sqrt(mean((BBH_predict_ENM_loocv$predict_ENM_loocv-
BBH_predict_ENM_loocv$age)^2))
cat("MAE:", MAE_ENM_loocv, "\nMed AE:", MedianAE_ENM_loocv, "\nRMSE:", RMSE_ENM_loocv,
"\n")

```

MAE: 3.6273

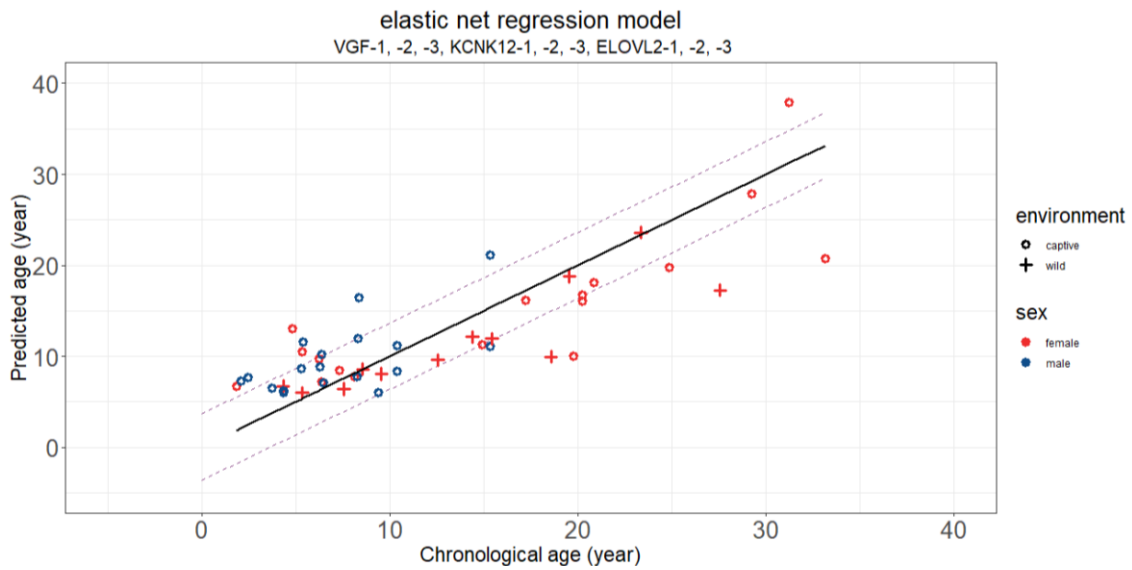
Med AE: 3.374564

RMSE: 4.630456

```

g_ENM_loocv<-ggplot(BBH_predict_ENM_loocv,aes(age,predict_ENM_loocv))+theme_bw()+
  annotate("segment",x=0,xend=max(BBH$age),y=3.6273,yend=max(BBH$age)+3.6273,colour="orch
id4",linetype=2,linewidth =0.7)+
  annotate("segment",x=0,xend=max(BBH$age),y=-3.6273,yend=max(BBH$age)-
3.6273,colour="orchid4",linetype=2,linewidth =0.7)+
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
  labs(x="Chronological age (year)",y="Predicted age (year)")+
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1
,"W"=3))+
  scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick
2","M"="dodgerblue4"))+
  theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
  theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
  geom_line(aes(y =age), linewidth=1)+
  labs(title="elastic net regression model")+
  theme(title=element_text(size=17),plot.title=element_text(hjust=0.5))+
  scale_y_continuous(limits=c(-5,40))+
  scale_x_continuous(limits=c(-5,40))+
  labs(subtitle=" VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-1, -2, -3")+
  theme(plot.subtitle=element_text(size=15,hjust=0.5)) +
  guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))

```



Age estimation model 【Support vector regression】

```
AGE <- BBH$age
```

```
#RemoveOne function
```

```
removeOne <- function(dat,x) {
  if(x<dat){
    list=seq(1,dat)
    x1=x-1;x2=x+1
    v1=c(list[0:x1]);v2=c(list[x2:dat])
    data=c(v1,v2)}
  else {data=seq(1,dat-1)}
  return (data)}

```

Support vector regression (VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-1, -2, -3)

```
tuneResult<-
```

```
tune(svm,age~VGF_1_methylation_rate_ave+VGF_2_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_1_methylation_rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methylation_rate_ave+ELOVL2_1_methylation_rate_ave+ELOVL2_2_methylation_rate_ave+ELOVL2_3_methylation_rate_ave,data=BBHS,
```

```

ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)

tunedModel <- tuneResult$best.model

```

Call:

```

best.tune(METHOD = svm, train.x = age ~ VGF_1_methylation_rate_ave +
  VGF_2_methylation_rate_ave + VGF_3_methylation_rate_ave +
  KCNK12_1_methylation_rate_ave +
  KCNK12_2_methylation_rate_ave + KCNK12_3_methylation_rate_ave +
  ELOVL2_1_methylation_rate_ave + ELOVL2_2_methylation_rate_ave +
  ELOVL2_3_methylation_rate_ave, data = BBHS, ranges = list(cost = 10^(seq(-4,
  5, 0.1)), gamma = 10^(seq(-5, 4, 0.1))), tunecontrol = tune.control(sampling = "cross",
  cross = 10), scale = FALSE)

```

Parameters:

```

SVM-Type: eps-regression
SVM-Kernel: radial
  cost: 3.981072
  gamma: 0.005011872
  epsilon: 0.1

```

Number of Support Vectors: 36

```

SVRM<-
svm(age~VGF_1_methylation_rate_ave+VGF_2_methylation_rate_ave+VGF_3_methylation_rate_ave
+KCNK12_1_methylation_rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methylation_rate_a
ve+ELOVL2_1_methylation_rate_ave+ELOVL2_2_methylation_rate_ave+ELOVL2_3_methylation_rate
_ave, data=BBHS,
cost=3.981072,gamma=0.005011872,epsilon=0.1, scale = FALSE)

#LOOCV
nSamples<-nrow(BBHS)
predict_SVRM_loocv<-numeric(nSamples)

```



```

for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-
data.frame(BBHS$age[indices],BBHS$VGF_1_methylation_rate_ave[indices],BBHS$VGF_2_methylation_rate_ave[indices],BBHS$VGF_3_methylation_rate_ave[indices],BBHS$KCNK12_1_methylation_rate_ave[indices],BBHS$KCNK12_2_methylation_rate_ave[indices],BBHS$KCNK12_3_methylation_rate_ave[indices],BBHS$ELOVL2_1_methylation_rate_ave[indices],BBHS$ELOVL2_2_methylation_rate_ave[indices],BBHS$ELOVL2_3_methylation_rate_ave[indices])
  colnames(dr)<-
c("age","methylvgf_1","methylvgf_2","methylvgf_3","methylkcnk_1","methylkcnk_2","methylkcnk_3","methylelovl_1","methylelovl_2","methylelovl_3")
  bestmodel_SVRM<-
svm(age~methylvgf_1+methylvgf_2+methylvgf_3+methylkcnk_1+methylkcnk_2+methylkcnk_3+methylelovl_1+methylelovl_2+methylelovl_3, data=dr,
cost=3.981072,gamma=0.005011872,epsilon=0.1, scale = FALSE)
  newdata<-
data.frame(methylvgf_1=BBHS$VGF_1_methylation_rate_ave[z],methylvgf_2=BBHS$VGF_2_methylation_rate_ave[z],methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z],methylkcnk_1=BBHS$KCNK12_1_methylation_rate_ave[z],methylkcnk_2=BBHS$KCNK12_2_methylation_rate_ave[z],methylkcnk_3=BBHS$KCNK12_3_methylation_rate_ave[z],methylelovl_1=BBHS$ELOVL2_1_methylation_rate_ave[z],methylelovl_2=BBHS$ELOVL2_2_methylation_rate_ave[z],methylelovl_3=BBHS$ELOVL2_3_methylation_rate_ave[z])
  p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SVRM_loocv[z]<-p}

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)

MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
RMSE_SVRM_loocv, "\n")

```

MAE: 3.397692

Med AE: 2.91568

RMSE: 4.409281

Support vector regression (VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-2, -3)

```
tuneResult<-  
tune(svm,age~VGF_1_methylation_rate_ave+VGF_2_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_1_methylation_rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methylation_rate_ave+ELOVL2_2_methylation_rate_ave+ELOVL2_3_methylation_rate_ave,data=BBHS,  
ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),  
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)  
  
tunedModel <- tuneResult$best.model
```

Call:

```
best.tune(METHOD = svm, train.x = age ~ VGF_1_methylation_rate_ave +  
      VGF_2_methylation_rate_ave + VGF_3_methylation_rate_ave +  
      KCNK12_1_methylation_rate_ave +  
      KCNK12_2_methylation_rate_ave + KCNK12_3_methylation_rate_ave +  
      ELOVL2_2_methylation_rate_ave + ELOVL2_3_methylation_rate_ave,  
      data = BBHS, ranges = list(cost = 10^(seq(-4, 5, 0.1)), gamma = 10^(seq(-5,  
      4, 0.1))), tunecontrol = tune.control(sampling = "cross",  
      cross = 10), scale = FALSE)
```

Parameters:

SVM-Type: eps-regression

SVM-Kernel: radial

cost: 5.011872

gamma: 0.003981072

epsilon: 0.1

```
SVRM<-  
svm(age~VGF_1_methylation_rate_ave+VGF_2_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_1_methylation_rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methylation_rate_ave+ELOVL2_2_methylation_rate_ave+ELOVL2_3_methylation_rate_ave, data=BBHS,  
cost=5.011872,gamma=0.003981072,epsilon=0.1, scale = FALSE)
```

```

#LOOCV
nSamples<-nrow(BBHS)
predict_SVRM_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-
data.frame(BBHS$age[indices],BBHS$VGF_1_methylation_rate_ave[indices],BBHS$VGF_2_methylation_rate_ave[indices],BBHS$VGF_3_methylation_rate_ave[indices],BBHS$KCNK12_1_methylation_rate_ave[indices],BBHS$KCNK12_2_methylation_rate_ave[indices],BBHS$KCNK12_3_methylation_rate_ave[indices],BBHS$ELOVL2_2_methylation_rate_ave[indices],BBHS$ELOVL2_3_methylation_rate_ave[indices])
  colnames(dr)<-
c("age","methylvgf_1","methylvgf_2","methylvgf_3","methylkcnk_1","methylkcnk_2","methylkcnk_3","methylelov1_2","methylelov1_3")
  bestmodel_SVRM<-
svm(age~methylvgf_1+methylvgf_2+methylvgf_3+methylkcnk_1+methylkcnk_2+methylkcnk_3+methylelov1_2+methylelov1_3, data=dr,
cost=5.011872,gamma=0.003981072,epsilon=0.1, scale = FALSE)
  newdata<-
data.frame(methylvgf_1=BBHS$VGF_1_methylation_rate_ave[z],methylvgf_2=BBHS$VGF_2_methylation_rate_ave[z],methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z],methylkcnk_1=BBHS$KCNK12_1_methylation_rate_ave[z],methylkcnk_2=BBHS$KCNK12_2_methylation_rate_ave[z],methylkcnk_3=BBHS$KCNK12_3_methylation_rate_ave[z],methylelov1_2=BBHS$ELOVL2_2_methylation_rate_ave[z],methylelov1_3=BBHS$ELOVL2_3_methylation_rate_ave[z])
  p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SVRM_loocv[z]<-p}

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)

MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
RMSE_SVRM_loocv, "\n")

```

MAE: 3.341879

Med AE: 2.914074

RMSE: 4.328937

Support vector regression (VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-3)

```
tuneResult<-  
tune(svm,age~VGF_1_methylation_rate_ave+VGF_2_methylation_rate_ave+VGF_3_methylation_r  
ate_ave+KCNK12_1_methylation_rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methylation_r  
ate_ave+ELOVL2_3_methylation_rate_ave,data=BBHS,  
ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),  
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)  
  
tunedModel <- tuneResult$best.model
```

Call:

```
best.tune(METHOD = svm, train.x = age ~ VGF_1_methylation_rate_ave +  
      VGF_2_methylation_rate_ave + VGF_3_methylation_rate_ave +  
      KCNK12_1_methylation_rate_ave +  
      KCNK12_2_methylation_rate_ave + KCNK12_3_methylation_rate_ave +  
      ELOVL2_3_methylation_rate_ave, data = BBHS, ranges = list(cost = 10^(seq(-4,  
5, 0.1)), gamma = 10^(seq(-5, 4, 0.1))), tunecontrol = tune.control(sampling = "cross",  
cross = 10), scale = FALSE)
```

Parameters:

SVM-Type: eps-regression

SVM-Kernel: radial

cost: 39.81072

gamma: 0.05011872

epsilon: 0.1

Number of Support Vectors: 40

SVRM<-

```

svm(age~VGF_1_methylation_rate_ave+VGF_2_methylation_rate_ave+VGF_3_methylation_rate_ave
+KCNK12_1_methylation_rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methylation_rate_a
ve+ELOVL2_3_methylation_rate_ave, data=BBHS,
cost=39.81072,gamma=0.05011872,epsilon=0.1, scale = FALSE)

#LOOCV
nSamples<-nrow(BBHS)
predict_SVRM_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-
data.frame(BBHS$age[indices],BBHS$VGF_1_methylation_rate_ave[indices],BBHS$VGF_2_methyla
tion_rate_ave[indices],BBHS$VGF_3_methylation_rate_ave[indices],BBHS$KCNK12_1_methylatio
n_rate_ave[indices],BBHS$KCNK12_2_methylation_rate_ave[indices],BBHS$KCNK12_3_methylatio
n_rate_ave[indices],BBHS$ELOVL2_3_methylation_rate_ave[indices])
  colnames(dr)<-
c("age","methylvgf_1","methylvgf_2","methylvgf_3","methylkcnk_1","methylkcnk_2","methylk
cnk_3","methylelov1_3")
  bestmodel_SVRM<-
svm(age~methylvgf_1+methylvgf_2+methylvgf_3+methylkcnk_1+methylkcnk_2+methylkcnk_3+methy
lelov1_3, data=dr,
cost=39.81072,gamma=0.05011872,epsilon=0.1, scale = FALSE)
  newdata<-
data.frame(methylvgf_1=BBHS$VGF_1_methylation_rate_ave[z],methylvgf_2=BBHS$VGF_2_methyla
tion_rate_ave[z],methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z],methylkcnk_1=BBHS$KCNK12
_1_methylation_rate_ave[z],methylkcnk_2=BBHS$KCNK12_2_methylation_rate_ave[z],methylkcnk
_3=BBHS$KCNK12_3_methylation_rate_ave[z],methylelov1_3=BBHS$ELOVL2_3_methylation_rate_av
e[z])
  p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SVRM_loocv[z]<-p}

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)

MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))

```

```
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
RMSE_SVRM_loocv, "\n")
```

MAE: 3.254831

Med AE: 2.511308

RMSE: 4.138454

Support vector regression (VGF-1, -3, KCNK12-1, -2, -3, ELOVL2-3)

```
tuneResult<-
tune(svm,age~VGF_1_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_1_methylation_
rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methylation_rate_ave+ELOVL2_3_methylatio
n_rate_ave,data=BBHS,
ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)

tunedModel <- tuneResult$best.model
```

Call:

```
best.tune(METHOD = svm, train.x = age ~ VGF_1_methylation_rate_ave +
  VGF_3_methylation_rate_ave + KCNK12_1_methylation_rate_ave +
  KCNK12_2_methylation_rate_ave + KCNK12_3_methylation_rate_ave +
  ELOVL2_3_methylation_rate_ave, data = BBHS, ranges = list(cost = 10^(seq(-4,
  5, 0.1)), gamma = 10^(seq(-5, 4, 0.1))), tunecontrol = tune.control(sampling = "cross",
  cross = 10), scale = FALSE)
```

Parameters:

SVM-Type: eps-regression

SVM-Kernel: radial

cost: 1.258925

gamma: 0.01258925

epsilon: 0.1

Number of Support Vectors: 37

```
SVRM<-
svm(age~VGF_1_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_1_methylation_rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methylation_rate_ave+ELOVL2_3_methylation_rate_ave, data=BBHS,
cost=1.258925,gamma=0.01258925,epsilon=0.1, scale = FALSE)

#LOOCV
nSamples<-nrow(BBHS)
predict_SVRM_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-
data.frame(BBHS$age[indices],BBHS$VGF_1_methylation_rate_ave[indices],BBHS$VGF_3_methylation_rate_ave[indices],BBHS$KCNK12_1_methylation_rate_ave[indices],BBHS$KCNK12_2_methylation_rate_ave[indices],BBHS$KCNK12_3_methylation_rate_ave[indices],BBHS$ELOVL2_3_methylation_rate_ave[indices])
  colnames(dr)<-
c("age","methylvgf_1","methylvgf_3","methylkcnk_1","methylkcnk_2","methylkcnk_3","methylelovl_3")
  bestmodel_SVRM<-
svm(age~methylvgf_1+methylvgf_3+methylkcnk_1+methylkcnk_2+methylkcnk_3+methylelovl_3,
data=dr,
cost=1.258925,gamma=0.01258925,epsilon=0.1, scale = FALSE)
  newdata<-
data.frame(methylvgf_1=BBHS$VGF_1_methylation_rate_ave[z],methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z],methylkcnk_1=BBHS$KCNK12_1_methylation_rate_ave[z],methylkcnk_2=BBHS$KCNK12_2_methylation_rate_ave[z],methylkcnk_3=BBHS$KCNK12_3_methylation_rate_ave[z],methylelovl_3=BBHS$ELOVL2_3_methylation_rate_ave[z])
  p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SVRM_loocv[z]<-p}

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)
```

```
MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
    RMSE_SVRM_loocv, "\n")
```

MAE: 3.35938

Med AE: 2.646612

RMSE: 4.240585

Support vector regression (VGF-1, -3, KCNK12-2, -3, ELOVL2-3)

```
tuneResult<-
tune(svm,age~VGF_1_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_2_methylation_
rate_ave+KCNK12_3_methylation_rate_ave+ELOVL2_3_methylation_rate_ave,data=BBHS,
ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)

tunedModel <- tuneResult$best.model
```

Call:

```
best.tune(METHOD = svm, train.x = age ~ VGF_1_methylation_rate_ave +
  VGF_3_methylation_rate_ave + KCNK12_2_methylation_rate_ave +
  KCNK12_3_methylation_rate_ave + ELOVL2_3_methylation_rate_ave,
  data = BBHS, ranges = list(cost = 10^(seq(-4, 5, 0.1)), gamma = 10^(seq(-5,
  4, 0.1))), tunecontrol = tune.control(sampling = "cross",
  cross = 10), scale = FALSE)
```

Parameters:

SVM-Type: eps-regression

SVM-Kernel: radial

cost: 1.258925

gamma: 0.01584893

epsilon: 0.1

Number of Support Vectors: 42

```
SVRM<-
svm(age~VGF_1_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methylation_rate_ave+ELOVL2_3_methylation_rate_ave, data=BBHS,
cost=1.258925,gamma=0.01584893,epsilon=0.1, scale = FALSE)

#LOOCV
nSamples<-nrow(BBHS)
predict_SVRM_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-
data.frame(BBHS$age[indices],BBHS$VGF_1_methylation_rate_ave[indices],BBHS$VGF_3_methylation_rate_ave[indices],BBHS$KCNK12_2_methylation_rate_ave[indices],BBHS$KCNK12_3_methylation_rate_ave[indices],BBHS$ELOVL2_3_methylation_rate_ave[indices])
  colnames(dr)<-
c("age","methylvgf_1","methylvgf_3","methylkcnk_2","methylkcnk_3","methylelovl_3")
  bestmodel_SVRM<-
svm(age~methylvgf_1+methylvgf_3+methylkcnk_2+methylkcnk_3+methylelovl_3, data=dr,
cost=1.258925,gamma=0.01584893,epsilon=0.1, scale = FALSE)
  newdata<-
data.frame(methylvgf_1=BBHS$VGF_1_methylation_rate_ave[z],methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z],methylkcnk_2=BBHS$KCNK12_2_methylation_rate_ave[z],methylkcnk_3=BBHS$KCNK12_3_methylation_rate_ave[z],methylelovl_3=BBHS$ELOVL2_3_methylation_rate_ave[z])
  p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SVRM_loocv[z]<-p}

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)

MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
RMSE_SVRM_loocv, "\n")
```

MAE: 3.367143

Med AE: 2.718321

RMSE: 4.246292

Support vector regression (VGF-1, -3, KCNK12-2, -3)

```
tuneResult<-  
tune(svm,age~VGF_1_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_2_methylation_  
rate_ave+KCNK12_3_methylation_rate_ave,data=BBHS,  
ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),  
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)  
  
tunedModel <- tuneResult$best.model
```

Call:

```
best.tune(METHOD = svm, train.x = age ~ VGF_1_methylation_rate_ave +  
  VGF_3_methylation_rate_ave + KCNK12_2_methylation_rate_ave +  
  KCNK12_3_methylation_rate_ave, data = BBHS, ranges = list(cost = 10^(seq(-4,  
  5, 0.1)), gamma = 10^(seq(-5, 4, 0.1))), tunecontrol = tune.control(sampling = "cross",  
  cross = 10), scale = FALSE)
```

Parameters:

SVM-Type: eps-regression

SVM-Kernel: radial

cost: 1.995262

gamma: 0.01584893

epsilon: 0.1

Number of Support Vectors: 37

```
SVRM<-  
svm(age~VGF_1_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_2_methylation_rate_  
ave+KCNK12_3_methylation_rate_ave, data=BBHS,  
cost=1.995262,gamma=0.01584893,epsilon=0.1, scale = FALSE)
```

```

#LOOCV
nSamples<-nrow(BBHS)
predict_SVRM_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-
data.frame(BBHS$age[indices],BBHS$VGF_1_methylation_rate_ave[indices],BBHS$VGF_3_methylation_rate_ave[indices],BBHS$KCNK12_2_methylation_rate_ave[indices],BBHS$KCNK12_3_methylation_rate_ave[indices])
  colnames(dr)<-c("age","methylvgf_1","methylvgf_3","methylkcnk_2","methylkcnk_3")
  bestmodel_SVRM<-svm(age~methylvgf_1+methylvgf_3+methylkcnk_2+methylkcnk_3, data=dr,
cost=1.995262,gamma=0.01584893,epsilon=0.1, scale = FALSE)
  newdata<-
data.frame(methylvgf_1=BBHS$VGF_1_methylation_rate_ave[z],methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z],methylkcnk_2=BBHS$KCNK12_2_methylation_rate_ave[z],methylkcnk_3=BBHS$KCNK12_3_methylation_rate_ave[z])
  p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SVRM_loocv[z]<-p}

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)

MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
RMSE_SVRM_loocv, "\n")

```

MAE: 3.634067

Med AE: 3.01526

RMSE: 4.533188

Support vector regression (VGF-1, -3, KCNK12-3)

```
tuneResult<-
```

```
tune(svm,age~VGF_1_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_3_methylation_
rate_ave,data=BBHS,
ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)

tunedModel <- tuneResult$best.model
```

Call:

```
best.tune(METHOD = svm, train.x = age ~ VGF_1_methylation_rate_ave +
VGF_3_methylation_rate_ave + KCNK12_3_methylation_rate_ave, data = BBHS,
ranges = list(cost = 10^(seq(-4, 5, 0.1)), gamma = 10^(seq(-5,
4, 0.1))), tunecontrol = tune.control(sampling = "cross",
cross = 10), scale = FALSE)
```

Parameters:

SVM-Type: eps-regression

SVM-Kernel: radial

cost: 50.11872

gamma: 0.006309573

epsilon: 0.1

Number of Support Vectors: 42

```
SVRM<-
svm(age~VGF_1_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_3_methylation_rate_
ave, data=BBHS,
cost=50.11872,gamma=0.006309573,epsilon=0.1, scale = FALSE)

#LOOCV
nSamples<-nrow(BBHS)
predict_SVRM_loocv<-numeric(nSamples)
for (z in 1:nSamples){
indices<-removeOne(nSamples,z)
```

```

dr<-
data.frame(BBHS$age[indices],BBHS$VGF_1_methylation_rate_ave[indices],BBHS$VGF_3_methylation_rate_ave[indices],BBHS$KCNK12_3_methylation_rate_ave[indices])
colnames(dr)<-c("age","methylvgf_1","methylvgf_3","methylkcnk_3")
bestmodel_SVRM<-svm(age~methylvgf_1+methylvgf_3+methylkcnk_3, data=dr,
cost=50.11872,gamma=0.006309573,epsilon=0.1, scale = FALSE)
newdata<-
data.frame(methylvgf_1=BBHS$VGF_1_methylation_rate_ave[z],methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z],methylkcnk_3=BBHS$KCNK12_3_methylation_rate_ave[z])
p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
if (p<0){p=0}
predict_SVRM_loocv[z]<-p}

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)

MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
RMSE_SVRM_loocv, "\n")

```

MAE: 3.191852

Med AE: 2.188271

RMSE: 4.238056

```

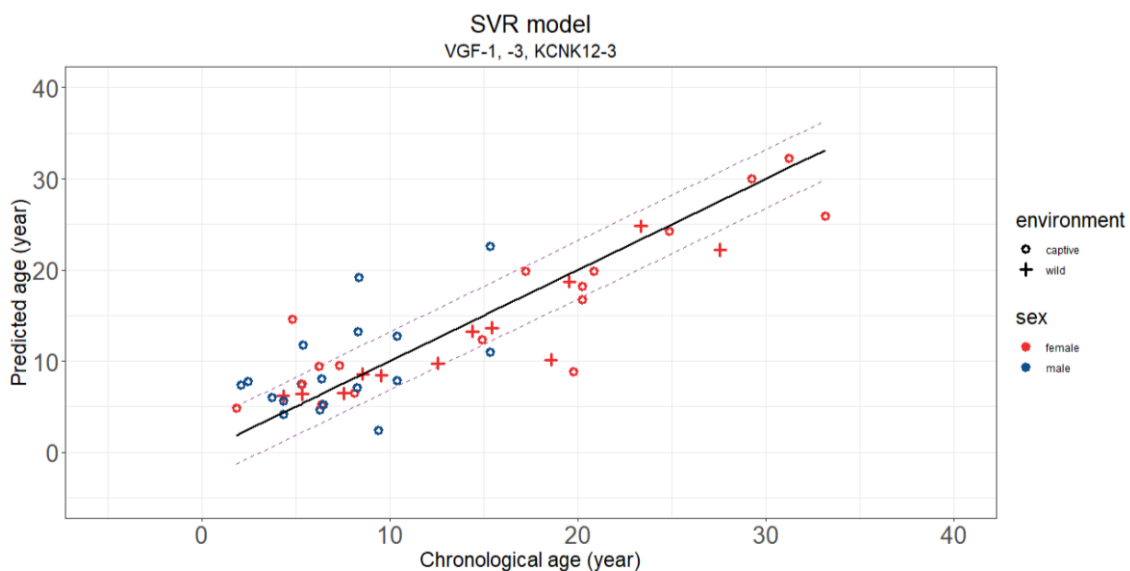
g_SVRM_loocv<-ggplot(BBH_SVRM_loocv,aes(age,predict_SVRM_loocv))+theme_bw()+
  annotate("segment",x=min(BBH$age),xend=max(BBH$age),y=min(BBH$age)+3.191852,yend=max(BBH$age)+3.191852,colour="orchid4",linetype=2,linewidth =0.7)+
  annotate("segment",x=min(BBH$age),xend=max(BBH$age),y=min(BBH$age)-3.191852,yend=max(BBH$age)-3.191852,colour="orchid4",linetype=2,linewidth =0.7)+
  geom_point(aes(shape=environment,color=sex),size=2,stroke=2)+
  labs(x="Chronological age (year)",y="Predicted age (year)")+
  scale_shape_manual(name="environment",labels=c("C"="captive","W"="wild"),values=c("C"=1,"W"=3))+
  scale_color_manual(name="sex",labels=c("F"="female","M"="male"),values=c("F"="firebrick2","M"="dodgerblue4"))+

```

```

theme(axis.text.x=element_text(size=20),axis.text.y=element_text(size=20))+
theme(axis.title.x=element_text(size=17),axis.title.y=element_text(size=17))+
geom_line(aes(y =age), linewidth=1)+
labs(title="SVR model")+
theme(title=element_text(size=17),plot.title=element_text(hjust=0.5))+
scale_y_continuous(limits=c(-5,40))+
scale_x_continuous(limits=c(-5,40))+
labs(subtitle="VGF-1, -3, KCNK12-3")+
theme(plot.subtitle=element_text(size=15,hjust=0.5))+
guides(shape= guide_legend(order = 1), color= guide_legend(order = 2))

```



Support vector regression (VGF-3, KCNK12-3)

```

tuneResult<-
tune(svm,age~VGF_3_methylation_rate_ave+KCNK12_3_methylation_rate_ave,data=BBHS,
ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)

tunedModel <- tuneResult$best.model

```

Call:

```

best.tune(METHOD = svm, train.x = age ~ VGF_3_methylation_rate_ave +
KCNK12_3_methylation_rate_ave, data = BBHS, ranges = list(cost = 10^(seq(-4,

```

```
5, 0.1)), gamma = 10^(seq(-5, 4, 0.1))), tunecontrol = tune.control(sampling = "cross",
cross = 10), scale = FALSE)
```

Parameters:

SVM-Type: eps-regression

SVM-Kernel: radial

cost: 0.7943282

gamma: 0.03162278

epsilon: 0.1

Number of Support Vectors: 40

```
SVRM<-
svm(age~VGF_3_methylation_rate_ave+KCNK12_3_methylation_rate_ave, data=BBHS,
cost=0.7943282,gamma=0.03162278,epsilon=0.1, scale = FALSE)

#LOOCV
nSamples<-nrow(BBHS)
predict_SVRM_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-
data.frame(BBHS$age[indices],BBHS$VGF_3_methylation_rate_ave[indices],BBHS$KCNK12_3_meth
ylation_rate_ave[indices])
  colnames(dr)<-c("age","methylvgf_3","methylkcnk_3")
  bestmodel_SVRM<-svm(age~methylvgf_3+methylkcnk_3, data=dr,
cost=0.7943282,gamma=0.03162278,epsilon=0.1, scale = FALSE)
  newdata<-
data.frame(methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z],methylkcnk_3=BBHS$KCNK12_3_met
hylation_rate_ave[z])
  p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SVRM_loocv[z]<-p}
```

```

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)

MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
    RMSE_SVRM_loocv, "\n")

```

```

MAE: 3.513377
Med AE: 2.76578
RMSE: 4.468467

```

Support vector regression (VGF-1, -2, -3)

```

tuneResult<-
tune(svm,age~VGF_1_methylation_rate_ave+VGF_2_methylation_rate_ave+VGF_3_methylation_rate_ave,data=BBHS,
ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)

tunedModel <- tuneResult$best.model

```

Call:

```

best.tune(METHOD = svm, train.x = age ~ VGF_1_methylation_rate_ave +
  VGF_2_methylation_rate_ave + VGF_3_methylation_rate_ave, data = BBHS,
  ranges = list(cost = 10^(seq(-4, 5, 0.1)), gamma = 10^(seq(-5,
    4, 0.1))), tunecontrol = tune.control(sampling = "cross",
    cross = 10), scale = FALSE)

```

Parameters:

SVM-Type: eps-regression

SVM-Kernel: radial

cost: 79.43282

gamma: 0.07943282

epsilon: 0.1

Number of Support Vectors: 35

```
SVRM<-
svm(age~VGF_1_methylation_rate_ave+VGF_2_methylation_rate_ave+VGF_3_methylation_rate_ave
, data=BBHS,
cost=79.43282,gamma=0.07943282,epsilon=0.1, scale = FALSE)

#LOOCV
nSamples<-nrow(BBHS)
predict_SVRM_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-
data.frame(BBHS$age[indices],BBHS$VGF_1_methylation_rate_ave[indices],BBHS$VGF_2_methylation_rate_ave[indices],BBHS$VGF_3_methylation_rate_ave[indices])
  colnames(dr)<-c("age","methylvgf_1","methylvgf_2","methylvgf_3")
  bestmodel_SVRM<-svm(age~methylvgf_1+methylvgf_2+methylvgf_3, data=dr,
cost=79.43282,gamma=0.07943282,epsilon=0.1, scale = FALSE)
  newdata<-
data.frame(methylvgf_1=BBHS$VGF_1_methylation_rate_ave[z],methylvgf_2=BBHS$VGF_2_methylation_rate_ave[z],methylvgf_3=BBHS$VGF_3_methylation_rate_ave[z])
  p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SVRM_loocv[z]<-p}

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)

MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
RMSE_SVRM_loocv, "\n")
```

MAE: 3.461789

Med AE: 2.438894

RMSE: 5.019953

Support vector regression (KCNK12-1, -2, -3)

```
tuneResult<-  
tune(svm,age~KCNK12_1_methylation_rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methyl  
ation_rate_ave,data=BBHS,  
ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),  
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)  
  
tunedModel <- tuneResult$best.model
```

Call:

```
best.tune(METHOD = svm, train.x = age ~ KCNK12_1_methylation_rate_ave +  
  KCNK12_2_methylation_rate_ave + KCNK12_3_methylation_rate_ave,  
  data = BBHS, ranges = list(cost = 10^(seq(-4, 5, 0.1)), gamma = 10^(seq(-5,  
    4, 0.1))), tunecontrol = tune.control(sampling = "cross",  
    cross = 10), scale = FALSE)
```

Parameters:

SVM-Type: eps-regression

SVM-Kernel: radial

cost: 1.258925

gamma: 0.03162278

epsilon: 0.1

Number of Support Vectors: 39

```
SVRM<-  
svm(age~KCNK12_1_methylation_rate_ave+KCNK12_2_methylation_rate_ave+KCNK12_3_methylation  
_rate_ave, data=BBHS,  
cost=1.258925,gamma=0.03162278,epsilon=0.1, scale = FALSE)  
  
#LOOCV  
nSamples<-nrow(BBHS)
```

```

predict_SVRM_loocv<-numeric(nSamples)
for (z in 1:nSamples){
  indices<-removeOne(nSamples,z)
  dr<-
data.frame(BBHS$age[indices],BBHS$KCNK12_1_methylation_rate_ave[indices],BBHS$KCNK12_2_m
ethylation_rate_ave[indices],BBHS$KCNK12_3_methylation_rate_ave[indices])
  colnames(dr)<-c("age","methylkcnk_1","methylkcnk_2","methylkcnk_3")
  bestmodel_SVRM<-svm(age~methylkcnk_1+methylkcnk_2+methylkcnk_3, data=dr,
cost=1.258925,gamma=0.03162278,epsilon=0.1, scale = FALSE)
  newdata<-
data.frame(methylkcnk_1=BBHS$KCNK12_1_methylation_rate_ave[z],methylkcnk_2=BBHS$KCNK12_2
_methylation_rate_ave[z],methylkcnk_3=BBHS$KCNK12_3_methylation_rate_ave[z])
  p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
  if (p<0){p=0}
  predict_SVRM_loocv[z]<-p}

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)

MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
RMSE_SVRM_loocv, "\n")

```

MAE: 3.844654

Med AE: 3.214305

RMSE: 4.909006

Support vector regression (ELOVL2-1, -2, -3)

```

tuneResult<-
tune(svm,age~ELOVL2_1_methylation_rate_ave+ELOVL2_2_methylation_rate_ave+ELOVL2_3_methyl
ation_rate_ave,data=BBHS,
ranges=list(cost=10^(seq(-4,5,0.1)),gamma=10^(seq(-5,4,0.1))),
tunecontrol = tune.control(sampling = "cross", cross = 10), scale = FALSE)

tunedModel <- tuneResult$best.model

```

Call:

```
best.tune(METHOD = svm, train.x = age ~ ELOVL2_1_methylation_rate_ave +
  ELOVL2_2_methylation_rate_ave + ELOVL2_3_methylation_rate_ave,
  data = BBHS, ranges = list(cost = 10^(seq(-4, 5, 0.1)), gamma = 10^(seq(-5,
  4, 0.1))), tunecontrol = tune.control(sampling = "cross",
  cross = 10), scale = FALSE)
```

Parameters:

SVM-Type: eps-regression

SVM-Kernel: radial

cost: 251.1886

gamma: 0.001584893

epsilon: 0.1

Number of Support Vectors: 41

SVRM<-

```
svm(age~ELOVL2_1_methylation_rate_ave+ELOVL2_2_methylation_rate_ave+ELOVL2_3_methylation
_rate_ave, data=BBHS,
cost=251.1886,gamma=0.001584893,epsilon=0.1, scale = FALSE)
```

#LOOCV

```
nSamples<-nrow(BBHS)
```

```
predict_SVRM_loocv<-numeric(nSamples)
```

```
for (z in 1:nSamples){
```

```
  indices<-removeOne(nSamples,z)
```

```
  dr<-
```

```
data.frame(BBHS$age[indices],BBHS$ELOVL2_1_methylation_rate_ave[indices],BBHS$ELOVL2_2_m
ethylation_rate_ave[indices],BBHS$ELOVL2_3_methylation_rate_ave[indices])
```

```
colnames(dr)<-c("age","methylelovl_1","methylelovl_2","methylelovl_3")
```

```
bestmodel_SVRM<-svm(age~methylelovl_1+methylelovl_2+methylelovl_3, data=dr,
```

```
cost=251.1886,gamma=0.001584893,epsilon=0.1, scale = FALSE)
```

```

newdata<-
data.frame(methylelovl_1=BBHS$ELOVL2_1_methylation_rate_ave[z],methylelovl_2=BBHS$ELOVL2_2_methylation_rate_ave[z],methylelovl_3=BBHS$ELOVL2_3_methylation_rate_ave[z])
p<-predict(bestmodel_SVRM,newdata)*sd(AGE)+mean(AGE)
if (p<0){p=0}
predict_SVRM_loocv[z]<-p}

BBH_SVRM_loocv<-cbind(BBH,predict_SVRM_loocv)

MAE_SVRM_loocv<-mean(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
MedianAE_SVRM_loocv<-median(abs(BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age))
RMSE_SVRM_loocv<- sqrt(mean((BBH_SVRM_loocv$predict_SVRM_loocv-BBH$age)^2))
cat("MAE:", MAE_SVRM_loocv, "\nMed AE:", MedianAE_SVRM_loocv, "\nRMSE:",
RMSE_SVRM_loocv, "\n")

```

MAE: 5.047747

Med AE: 3.864458

RMSE: 7.259932

Influences of interaction among age, sex, and growth environment

Support vector regression (VGF-1, -3 , KCNK12-3)

```

SVRM3<-
svm(age~VGF_1_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_3_methylation_rate_ave, data=BBHS,cost=50.11872,gamma=0.006309573,epsilon=0.1, scale = FALSE)

predict_SVRM3_s <- predict(SVRM3)

#Δ age
deltaage_SVRM3<- predict_SVRM3_s-BBHS$age
modeldelta_SVRM3<-lm(formula =deltaage_SVRM3~BBHS$age*BBHS$sex*BBHS$environment)
options(na.action = "na.fail")
modellist_SVRM3<-dredge(modeldelta_SVRM3,rank="AIC")
bestmodel_SVRM3<- get.models(dredge(modeldelta_SVRM3,rank="AIC"),subset=1)

```

```
bestmodel_SVRM3
```

```
$`4`
```

Call:

```
lm(formula = deltaage_SVRM3 ~ BBHS$age + BBHS$environment + 1)
```

Coefficients:

(Intercept)	BBHS\$age	BBHS\$environmentW
0.07009	-0.21261	-0.20370

```
attr("rank")
function (x)
do.call("rank", list(x))
<environment: 0x0000025fb7262648>
attr("call")
AIC(x)
attr("class")
[1] "function"      "rankFunction"
attr("beta")
[1] "none"
```

```
summary(lm(formula = deltaage_SVRM3 ~ BBHS$age + BBHS$environment + 1))
```

Call:

```
lm(formula = deltaage_SVRM3 ~ BBHS$age + BBHS$environment + 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.06505	-0.28872	-0.01493	0.20515	1.08435

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.07009	0.07242	0.968	0.33846
BBHS\$age	-0.21261	0.06354	-3.346	0.00169 **

```
BBHS$environmentW -0.20370    0.14417 -1.413  0.16470
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.4276 on 44 degrees of freedom
```

```
Multiple R-squared:  0.2491,    Adjusted R-squared:  0.215
```

```
F-statistic: 7.299 on 2 and 44 DF,  p-value: 0.00183
```

```
#|Δage|
```

```
absdeltaage_SVRM3<- abs(predict_SVRM3_s-BBHS$age)
```

```
modeldelta_absSVRM3<-glm(absdeltaage_SVRM3 ~ age*sex*environment, data = BBHS, family =  
Gamma(link = "log"))
```

```
options(na.action = "na.fail")
```

```
modellist_absSVRM3<-dredge(modeldelta_absSVRM3,rank="AIC")
```

```
bestmodel_absSVRM3<- get.models(dredge(modeldelta_absSVRM3,rank="AIC"),subset=1)
```

```
bestmodel_absSVRM3
```

```
$`1`
```

```
Call: glm(formula = absdeltaage_SVRM3 ~ 1, family = Gamma(link = "log"),  
data = BBHS)
```

```
Coefficients:
```

```
(Intercept)
```

```
-1.041
```

```
Degrees of Freedom: 46 Total (i.e. Null); 46 Residual
```

```
Null Deviance:    43.54
```

```
Residual Deviance: 43.54    AIC: -0.525
```

```
attr("rank")
```

```
function (x)
```

```
do.call("rank", list(x))
```

```
<environment: 0x0000025fb22acc08>
```

```
attr("call")
```

```
AIC(x)
```

```
attr("class")
[1] "function"      "rankFunction"
attr("beta")
[1] "none"
```

How to apply to the models

```
HMD <- read.csv("hair_measurement_data.csv")
HMD[,2] <- (HMD[,2] - 5.6163) / 2.6814
HMD[,3] <- (HMD[,3] - 7.0590) / 4.0436
HMD[,4] <- (HMD[,4] - 10.572) / 5.4001
HMD[,5] <- (HMD[,5] - 11.605) / 6.0872
HMD[,6] <- (HMD[,6] - 14.020) / 5.9832
HMD[,7] <- (HMD[,7] - 17.529) / 7.3715
HMD[,8] <- (HMD[,8] - 18.322) / 5.7400
HMD[,9] <- (HMD[,9] - 45.225) / 9.0431
HMD[,10] <- (HMD[,10] - 36.266) / 10.283

write.csv(HMD, "hair_measurement_data_standardized.csv", row.names = FALSE)
HMDS <- read.csv("hair_measurement_data_standardized.csv")

BBHS<-read.csv("brown_bear_hair_standardized.csv")
SD <- 8.2771
MEAN <- 12.149
```

The file “hair_measurement_data.csv” is a template file, available in Dryad; doi: 10.5061/dryad.h44j0zpz. Write the sample IDs and measured methylation levels on it.

The “hair_measurement_data_standardized.csv” file contains values of standardized methylation levels of “hair_measurement_data.csv”.

The standardized values are calculated by the following equation.

“standardized value” = (“original value” – “mean of training data”) ÷ “standard deviation of training data”)

Single regression (KCNK12-3)

```
SRM <-lm(formula=age~KCNK12_3_methylation_rate_ave,data=BBHS)
```



```

predicted_age_SRM <- predict(SRM,HMDS)*SD+MEAN
predicted_age_SRM[predicted_age_SRM < 0] <- 0

data.frame(Sample_ID = HMDS$Sample_ID, Predicted_Age = predicted_age_SRM)

```

Elastic net regression (VGF-1, -2, -3, KCNK12-1, -2, -3, ELOVL2-1, -2, -3)

```

library(glmnet)

ENM <- glmnet(x = cbind(BBHS$VGF_1_methylation_rate_ave,BBHS$VGF_2_methylation_rate_ave,
BBHS$VGF_3_methylation_rate_ave,BBHS$KCNK12_1_methylation_rate_ave,BBHS$KCNK12_2_methylation_rate_ave,BBHS$KCNK12_3_methylation_rate_ave,BBHS$ELOVL2_1_methylation_rate_ave,BBHS$ELOVL2_2_methylation_rate_ave,BBHS$ELOVL2_3_methylation_rate_ave),
y = BBHS$age, family = "gaussian", lambda = 0.4369317, alpha = 0.04, standardize = FALSE)

HMDS_ENM <- cbind(HMDS$VGF_1_methylation_rate_ave,HMDS$VGF_2_methylation_rate_ave,
HMDS$VGF_3_methylation_rate_ave,HMDS$KCNK12_1_methylation_rate_ave,HMDS$KCNK12_2_methylation_rate_ave,HMDS$KCNK12_3_methylation_rate_ave,HMDS$ELOVL2_1_methylation_rate_ave,HMDS$ELOVL2_2_methylation_rate_ave,HMDS$ELOVL2_3_methylation_rate_ave)

predicted_age_ENM <- c(predict(ENM,HMDS_ENM,s=0.4369317)*SD+MEAN)
predicted_age_ENM[predicted_age_ENM < 0] <- 0

data.frame(Sample_ID = HMDS$Sample_ID, Predicted_Age = predicted_age_ENM)

```

Support vector regression (VGF-1, -3, KCNK12-3)

```

library(e1071)

SVRM<-
svm(age~VGF_1_methylation_rate_ave+VGF_3_methylation_rate_ave+KCNK12_3_methylation_rate_ave,data=BBHS, cost=50.11872, gamma=0.006309573, epsilon=0.1, scale = FALSE)

predicted_age_SVRM<-predict(SVRM,HMDS)*SD+MEAN
predicted_age_SVRM[predicted_age_SVRM < 0] <- 0

data.frame(Sample_ID = HMDS$Sample_ID, Predicted_Age = predicted_age_SVRM)

```

Output to a csv file

```

sample_ID <- HMDS[, 1, drop = FALSE]

predicted_age<-cbind(sample_ID,predicted_age_SRM,predicted_age_ENM,predicted_age_SVRM)
write.csv(predicted_age, "predicted_age_result.csv", row.names = FALSE)

```

At this point, two files, “hair_measurement_data_standardized.csv” and “predicted_age_result.csv”, should have been generated. Leaving them as they are may cause errors in subsequent runs, so please either rename the files or move them to a different folder.