

The basic use of ggplot2 and SureTypeSCR

Lishan Cai

November 2019

This article provides the basic guide of how to make use of R packages ggplot2 and SureTypeSCR.

1 Callrate

```
# import SureTypeSCR and ggplot2 packages

library(SureTypeSCR)
library(ggplot2)

# process raw data (GTCs)

df <- scbasic(manifest.bmp, cluster.egt, samplesheet.csv)

#Calculate callrates with different threshold

callrate15 <- callrate(df,th=0.15)
callrate5 <- callrate(df,th=0.5)
callrate7 <- callrate(df,th=0.7)

# plot the three different sets of callrate with different threshold

cols <- c("callrate15"= "red", "callrate5"="green","callrate7"="blue")

callrateplot <- ggplot(callrate15,aes(rownames(callrate15),Call_rate))
+ geom_point(data = callrate15,size=2,aes(colour="callrate15"))
+ geom_point(data=callrate5,aes(colour="callrate5"))
+ geom_point(data=callrate7,aes(colour="callrate7"))
+ scale_colour_manual(name="threshold",values=cols)
+ ylim(0,1) + ggtitle("Callrate") + xlab("Sample_name") + ylab("callrate")
+ theme(axis.text.x = element_text(angle = 45, hjust = 0.5, vjust = 0.5))
```

2 callrate from different chromosomes

```

# Calculate callrates of different chromosomes

chr1 <- callrate_chr(df,"1",th=0.15)
chr21 <- callrate_chr(df,"21",th=0.15)
chry <- callrate_chr(df,"Y", th=0.15)

# plot

cols <- c("chr1"= "red", "chr21"="green","chry"="blue")

chrplot <- ggplot(chr1,aes(rownames(chr1),Call_rate))
+ geom_point(data = chr1,size=2,aes(colour="chr1"))
+ geom_point(data=chr21,aes(colour="chr21"))
+ geom_point(data=chry,aes(colour="chry"))
+ scale_colour_manual(name="chromosomes",values=cols)
+ ylim(0,1) + ggtitle("Callrate of different chromosomes") + xlab("Sample_name") + ylab("Call_rate")
+ theme(axis.text.x = element_text(angle = 45, hjust = 0.5, vjust = 0.5))

```

3 MA plot of chromosomes

```

# get M and A features of targeted chromosome of specific sample

am <- sample_ma(df,"gDNA_1","Y")

MAplot <- ggplot(data=am, aes(x=A, y= M)) + geom_point()
+ xlab("A") + ylab("M") + ggtitle("MA plot of Y of gDNA_1")

```

4 PCA plot

```

library(ggrepel)

# Apply PCA over samples

samples <- pca_samples(df, th = 0.15)

# plot
pcaplot <- ggplot(data=samples,aes(PC1,PC2),label = rownames(samples))
+ geom_point()+ geom_text_repel(data=samples, aes(x=PC1,y=PC2),label = rownames(samples))
+ xlab("PC1") + ylab("PC2")
+ ggtitle("PC1 and PC2 over all samples")

```