

# Supplementary materials, R code for downstream analysis

## Contents

<b>1</b>	<b>Preparation</b>	<b>2</b>
1.1	Loading data and functions . . . . .	2
1.2	Clinical parameters . . . . .	2
<b>2</b>	<b>Whole-slide spatial organization features (SOFs)</b>	<b>3</b>
2.1	Distributions of tissues . . . . .	3
2.2	Generating whole-slide level SOFs . . . . .	4
2.3	Distributions and associations with CMSs of whole-slide SOFs . . . . .	4
2.4	Distributions and associations with CMSs of ESTIMATE scores . . . . .	8
2.5	Correlation of Whole-slide SOFs and ESTIMATE scores . . . . .	9
<b>3</b>	<b>Infiltrating spatial organization features</b>	<b>9</b>
3.1	Generating infiltrating SOFs . . . . .	9
3.2	Univariate cox regression analyses . . . . .	10
3.3	Survival analyses of infiltrating SOFs . . . . .	11
<b>4</b>	<b>Session Info</b>	<b>12</b>

# 1 Preparation

## 1.1 Loading data and functions

This document presents R codes and data for reproducing our analysis result as described in the paper. To start, first load the R functions needed for analysis and plotting and load the data for the TCGA colorectal cancer samples, which is available as part of the paper supplement.

```
library(ggplot2)
library(cowplot)
library(survival)
library(stringr)
load("./ProcessedDataAndClinical.rdata")
source("./getInfiltrateRatio.R")
source("./plotCorrelation.R")
source("./plot_KMCurve.R")
head(getInfiltrateRatio)

##
## 1 function (windows = 6, overlap = 0.5, locationMatrix)
## 2 {
## 3     Ratio_aver <- sapply(1:length(locationMatrix), function(i) {
## 4         tmp_mat <- locationMatrix[[i]]
## 5         squnR <- seq(1, nrow(tmp_mat) - windows * overlap, windows *
## 6             overlap)
## 7     })
## 8     head(plotCorrelation)

##
## 1 function (x, y, groups = NULL, legend.pos = "none", maintitle = NULL)
## 2 {
## 3     pval <- paste(sprintf("Correlation = %.3f\\nP", Hmisc::rcorr(x,
## 4         y)$r[2, 1]), paste0("= ", signif(Hmisc::rcorr(x, y)$P[2,
## 5         1], 3)))
## 6     cor_plot <- data.frame(a = x, b = y)
## 7     head(plot_KMCurve)

##
## 1 function (clinical, labels, limit = NULL, annot = NULL, color = NULL,
## 2     font = "serif", xlab = "Follow up", ylab = "Survival Probability",
## 3     title = NULL, legend.pos = "top", palette = "jama_classic",
## 4     risk.table = T, risk.table.ratio = 0.4, anno.pos = "bottom",
## 5     anno.x.shift = 0.5)
## 6 {
```

## 1.2 Clinical parameters

Each row in this clinical file is a sample/patient from the TCGA cohort. Columns indicate: public patient ID (sample), gender, age, TNM stage, tumor location, relapse-free survival (rfs.event, rfs.delay), the status of microsatellite instability (mmr.status), consensus cluster subtype classification (cms), sample subsets (dataset), the fractions of stromal, immune, and tumor cells which were inferred from gene expression profiles by 'ESTIMATE' (StromalScore, ImmuneScore and TumorPurity).

```
head(Clinical_info)
```

```
##               sample sex age tnm.stage tumor.location rfs.event rfs.delay
## TCGA-G4-6299 TCGA-G4-6299 M  69         3             L          0  75.60000
## TCGA-D5-5537 TCGA-D5-5537 M  83         2             R          1  24.06667
## TCGA-G4-6627 TCGA-G4-6627 M  84         2             R          1  33.43333
## TCGA-CM-5860 TCGA-CM-5860 M  44         2             R          0  32.46667
## TCGA-CA-6717 TCGA-CA-6717 M  57         2             R          0  12.93333
## TCGA-CM-6163 TCGA-CM-6163 M  74         1             L          0  14.23333
##               mmr.status cms dataset StromalScore ImmuneScore TumorPurity
## TCGA-G4-6299      mss CMS1   TCGA   -212.814614  1073.742857   0.7442658
## TCGA-D5-5537      mss CMS2   TCGA   -577.955207    9.018398   0.8670859
## TCGA-G4-6627      msi CMS4   TCGA    6.104147  1167.730166   0.7128151
## TCGA-CM-5860      mss CMS4   TCGA   282.634393   700.233982   0.7321925
## TCGA-CA-6717      mss CMS4   TCGA   476.618101  1427.056730   0.6337347
## TCGA-CM-6163      mss CMS2   TCGA  -185.274656   920.772889   0.7564358
```

## 2 Whole-slide spatial organization features (SOFs)

### 2.1 Distributions of tissues

```
distributionOfPatch<- sapply(1:length(CRC_locationMatrix),function(i){
  mat <- CRC_locationMatrix[[i]]
  tmp <- sapply(c(1:9), function(j){
    return(length(which(mat==j)))
  })
  return(tmp)
})
rownames(distributionOfPatch) <-c('ADI','BACK','DEB','LYM','MUC',
                                'MUS','NORM','STR','TUM')
colnames(distributionOfPatch) <- names(CRC_locationMatrix)
distributionOfPatch <- as.data.frame(t(distributionOfPatch))

Distribution <- distributionOfPatch[,1:9]/rowSums(distributionOfPatch[,1:9])
distriPlot <- data.frame(num=c(as.matrix(Distribution)))
distriPlot$class <- rep(c('ADI','BACK','DEB','LYM','MUC','MUS','NORM','STR','TUM'),
                        each=nrow(distributionOfPatch))
boxplot( num ~ class, distriPlot)
```

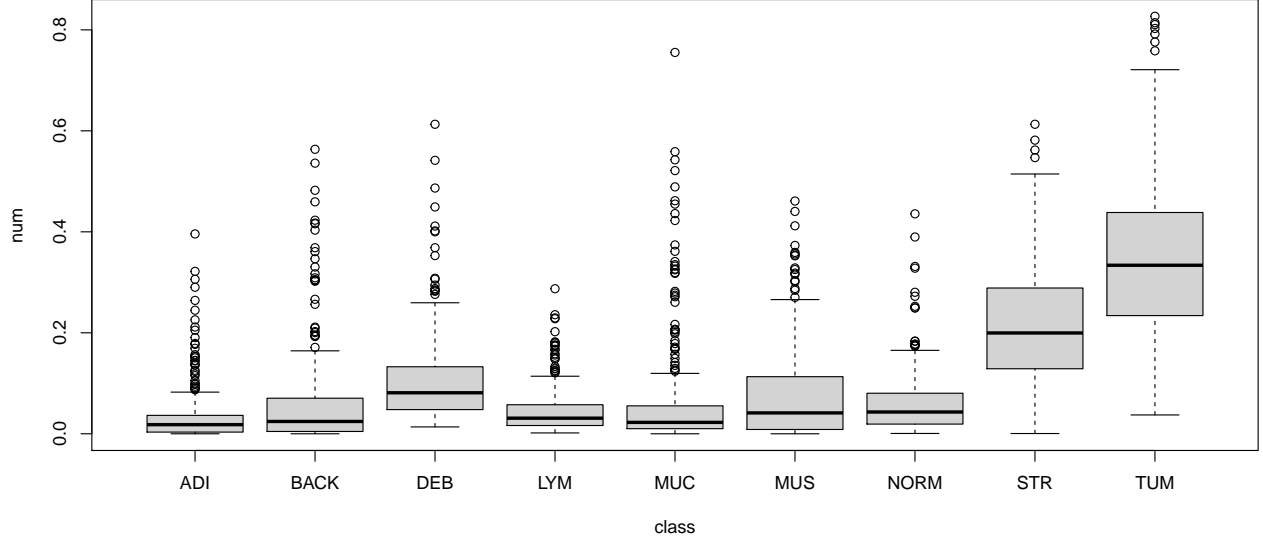


Figure 2a. Distributions of whole-slide SOFs in the TCGA cohort.

## 2.2 Generating whole-slide level SOFs

To evaluate the potential association between spatial organization features (SOFs) and the Consensus Molecular Subtypes (CMSs) of colorectal cancer, we calculated the relative abundance of image patches assigned to different tissue types in the whole-slide images as whole-slide tissue organization features (SOFs). More specifically, we first remove patches classified as background (BACK) and debris (DEB) for each whole-slide image in the TCGA cohort. The relative proportions of image patches classified to the other seven tissue types were subsequently calculated and compared between tumors belonging to different CMSs.

```
WholeSlide_ratio <- distributionOfPatch[, -c(2,3)] / rowSums(distributionOfPatch[, -c(2,3)])
head(WholeSlide_ratio)
```

```
##           ADI           LYM           MUC           MUS           NORM
## TCGA-G4-6299 0.056409866 0.04834044 0.02253350 0.030146163 0.002892814
## TCGA-D5-5537 0.032137118 0.06534547 0.01562221 0.134351009 0.045795394
## TCGA-G4-6627 0.043205635 0.04964783 0.01022161 0.116474833 0.005239650
## TCGA-CM-5860 0.037047870 0.08121283 0.03977771 0.072145852 0.075753144
## TCGA-CA-6717 0.008450163 0.02746303 0.01248320 0.007489917 0.001536393
## TCGA-CM-6163 0.023880978 0.03760935 0.04737884 0.043611519 0.195772939
##           STR           TUM
## TCGA-G4-6299 0.5044915 0.3351857
## TCGA-D5-5537 0.2662917 0.4404571
## TCGA-G4-6627 0.3254595 0.4497509
## TCGA-CM-5860 0.3585844 0.3354782
## TCGA-CA-6717 0.3906280 0.5519493
## TCGA-CM-6163 0.4372007 0.2145457
```

## 2.3 Distributions and associations with CMSs of whole-slide SOFs

Each CMS has its characteristic tissue type that is significantly more abundant than the other CMSs

```
all.equal(as.character(Clinical_info$sample), rownames(WholeSlide_ratio))
```

```
## [1] TRUE
```

```

wilcox.test(WholeSlide_ratio$LYM[which(Clinical_info$cms=="CMS1")],
            WholeSlide_ratio$LYM[which(Clinical_info$cms!="CMS1")])

##
## Wilcoxon rank sum test with continuity correction
##
## data: WholeSlide_ratio$LYM[which(Clinical_info$cms == "CMS1")] and WholeSlide_ratio$LYM[which(Clinical_info$cms != "CMS1")]
## W = 8570, p-value = 0.01856
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(WholeSlide_ratio$TUM[which(Clinical_info$cms=="CMS2")],
            WholeSlide_ratio$TUM[which(Clinical_info$cms!="CMS2")])

##
## Wilcoxon rank sum test with continuity correction
##
## data: WholeSlide_ratio$TUM[which(Clinical_info$cms == "CMS2")] and WholeSlide_ratio$TUM[which(Clinical_info$cms != "CMS2")]
## W = 16153, p-value = 0.0002742
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(WholeSlide_ratio$MUC[which(Clinical_info$cms=="CMS3")],
            WholeSlide_ratio$MUC[which(Clinical_info$cms!="CMS3")])

##
## Wilcoxon rank sum test with continuity correction
##
## data: WholeSlide_ratio$MUC[which(Clinical_info$cms == "CMS3")] and WholeSlide_ratio$MUC[which(Clinical_info$cms != "CMS3")]
## W = 9619, p-value = 0.0007595
## alternative hypothesis: true location shift is not equal to 0
wilcox.test(WholeSlide_ratio$STR[which(Clinical_info$cms=="CMS4")],
            WholeSlide_ratio$STR[which(Clinical_info$cms!="CMS4")])

##
## Wilcoxon rank sum test with continuity correction
##
## data: WholeSlide_ratio$STR[which(Clinical_info$cms == "CMS4")] and WholeSlide_ratio$STR[which(Clinical_info$cms != "CMS4")]
## W = 16895, p-value = 4.853e-06
## alternative hypothesis: true location shift is not equal to 0
LYM_average <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(mean(WholeSlide_ratio$LYM[which(Clinical_info$cms==i)],na.rm=TRUE))
})
LYM_se <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(sd(WholeSlide_ratio$LYM[which(Clinical_info$cms==i)],na.rm=TRUE)/
         sqrt(length(which(Clinical_info$cms==i))))
})
LYM <- data.frame(Average=LYM_average,SE=LYM_se)*100
LYM$CMS <- c("CMS1","CMS2","CMS3","CMS4")
LYM <- ggplot(LYM,aes(x=CMS, y=Average,fill=CMS)) +
  geom_bar(stat = "identity",position = "dodge",colour="black",width = 0.6)+
  scale_fill_manual(values=c("Dimgray","gray","gray","gray"))+
  geom_errorbar(aes(ymin=Average-SE, ymax=Average+SE), width=.2)+
  coord_cartesian(ylim=c(2, 8))+
  ggtitle("Whole-slide Lymphocyte Ratio")

```

```

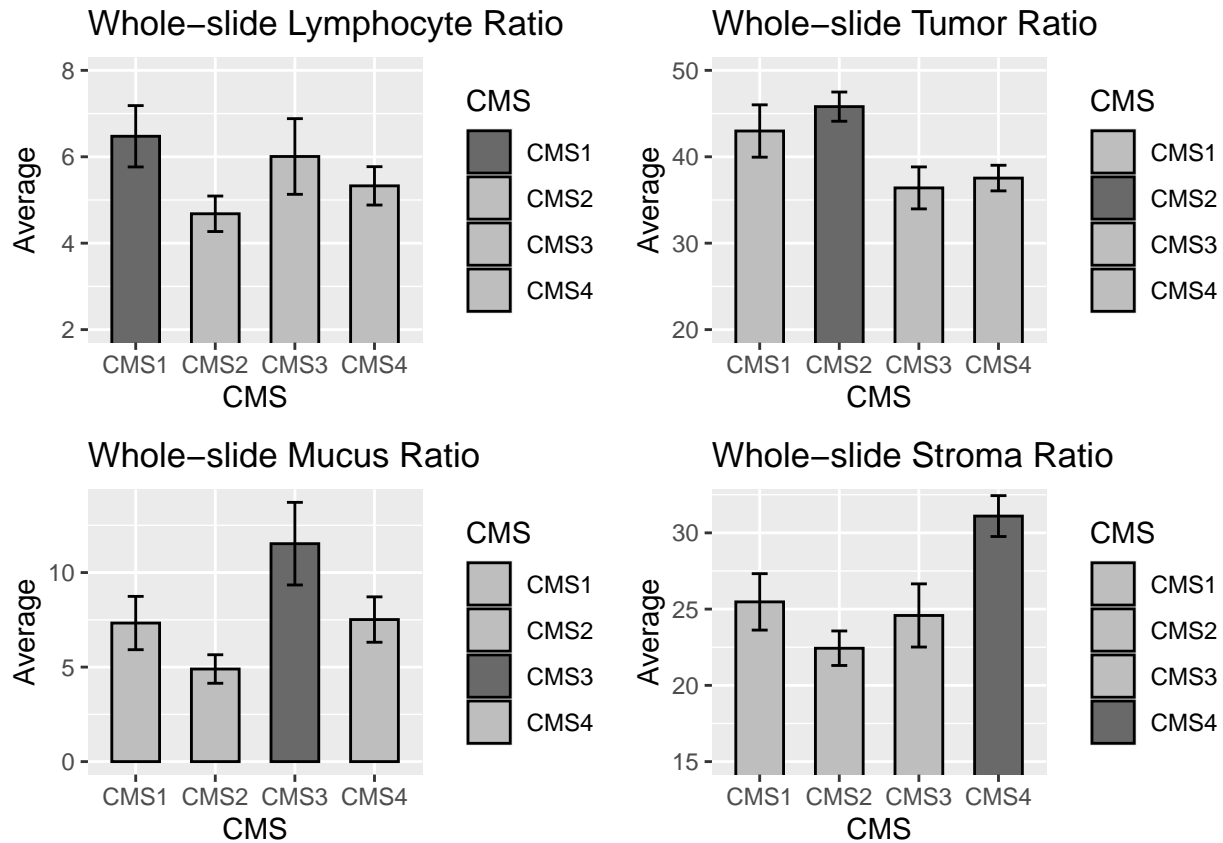
TUM_average <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(mean(WholeSlide_ratio$TUM[which(Clinical_info$cms==i)],na.rm=TRUE))
})
TUM_se <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(sd(WholeSlide_ratio$TUM[which(Clinical_info$cms==i)],na.rm=TRUE)/
    sqrt(length(which(Clinical_info$cms==i))))
})
TUM <- data.frame(Average=TUM_average,SE=TUM_se)*100
TUM$CMS <- c("CMS1","CMS2","CMS3","CMS4")
TUM <- ggplot(TUM,aes(x=CMS, y=Average,fill=CMS)) +
  geom_bar(stat = "identity",position = "dodge",colour="black",width = 0.6)+
  scale_fill_manual(values=c("gray","Dimgray","gray","gray"))+
  geom_errorbar(aes(ymin=Average-SE, ymax=Average+SE), width=.2)+
  coord_cartesian(ylim=c(20, 50))+
  ggtitle("Whole-slide Tumor Ratio")

MUC_average <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(mean(WholeSlide_ratio$MUC[which(Clinical_info$cms==i)],na.rm=TRUE))
})
MUC_se <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(sd(WholeSlide_ratio$MUC[which(Clinical_info$cms==i)],na.rm=TRUE)/
    sqrt(length(which(Clinical_info$cms==i))))
})
MUC <- data.frame(Average=MUC_average,SE=MUC_se)*100
MUC$CMS <- c("CMS1","CMS2","CMS3","CMS4")
MUC <- ggplot(MUC,aes(x=CMS, y=Average,fill=CMS)) +
  geom_bar(stat = "identity",position = "dodge",colour="black",width = 0.6)+
  scale_fill_manual(values=c("gray","gray","Dimgray","gray"))+
  geom_errorbar(aes(ymin=Average-SE, ymax=Average+SE), width=.2)+
  ggtitle("Whole-slide Mucus Ratio")

STR_average <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(mean(WholeSlide_ratio$STR[which(Clinical_info$cms==i)],na.rm=TRUE))
})
STR_se <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(sd(WholeSlide_ratio$STR[which(Clinical_info$cms==i)],na.rm=TRUE)/
    sqrt(length(which(Clinical_info$cms==i))))
})
STR <- data.frame(Average=STR_average,SE=STR_se)*100
STR$CMS <- c("CMS1","CMS2","CMS3","CMS4")
STR <- ggplot(STR,aes(x=CMS, y=Average,fill=CMS)) +
  geom_bar(stat = "identity",position = "dodge",colour="black",width = 0.6)+
  scale_fill_manual(values=c("gray","gray","gray","Dimgray"))+
  geom_errorbar(aes(ymin=Average-SE, ymax=Average+SE), width=.2)+
  coord_cartesian(ylim=c(15, 32))+
  ggtitle("Whole-slide Stroma Ratio")

plot_grid(LYM,TUM,MUC,STR,ncol = 2,align = "hv")

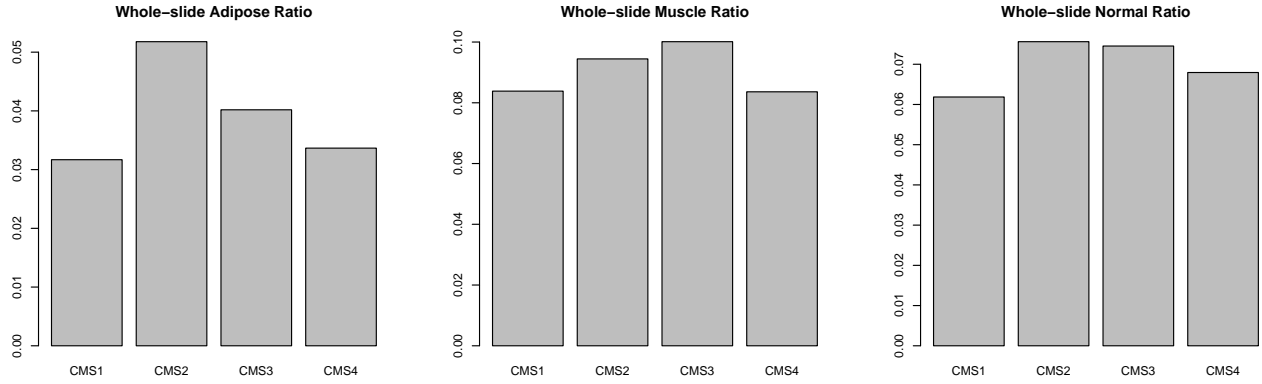
```



**Figure 2c. Associations of whole-slide SPFs and molecular subtypes of colorectal cancer.**

Moreover, the whole-slide adipose, muscle, and normal ratios did not show significant differences between the four CMSs.

```
par(mfrow = c(1, 3))
ADI_average <- sapply(c("CMS1", "CMS2", "CMS3", "CMS4"), function(i){
  return(mean(WholeSlide_ratio$ADI[which(Clinical_info$cms==i)], na.rm=TRUE))
})
MUS_average <- sapply(c("CMS1", "CMS2", "CMS3", "CMS4"), function(i){
  return(mean(WholeSlide_ratio$MUS[which(Clinical_info$cms==i)], na.rm=TRUE))
})
NORM_average <- sapply(c("CMS1", "CMS2", "CMS3", "CMS4"), function(i){
  return(mean(WholeSlide_ratio$NORM[which(Clinical_info$cms==i)], na.rm=TRUE))
})
barplot(ADI_average, names.arg = c("CMS1", "CMS2", "CMS3", "CMS4"),
  col = rep("gray", 4), beside=TRUE, main = "Whole-slide Adipose Ratio")
barplot(MUS_average, names.arg = c("CMS1", "CMS2", "CMS3", "CMS4"),
  col = rep("gray", 4), beside=TRUE, main = "Whole-slide Muscle Ratio")
barplot(NORM_average, names.arg = c("CMS1", "CMS2", "CMS3", "CMS4"),
  col = rep("gray", 4), beside=TRUE, main = "Whole-slide Normal Ratio")
```



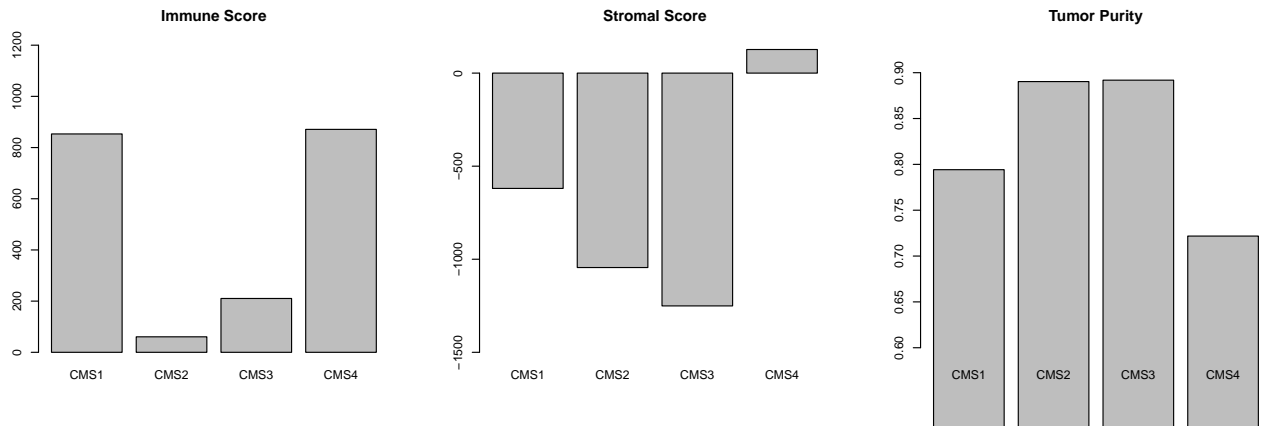
Supplementary Figure 2a. Distribution of whole-slide adipose, muscle, and normal ratios.

## 2.4 Distributions and associations with CMSs of ESTIMATE scores

The fractions of stromal, immune, and tumor cells can also be inferred from bulk tumor gene expression profiles using their corresponding gene expression signatures by ESTIMATE.

```
par(mfrow = c(1, 3))
ImmuneScore <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(mean(Clinical_info$ImmuneScore[which(Clinical_info$cms==i)],na.rm=TRUE))
})
StromalScore <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(mean(Clinical_info$StromalScore[which(Clinical_info$cms==i)],na.rm=TRUE))
})
TumorPurity <- sapply(c("CMS1","CMS2","CMS3","CMS4"), function(i){
  return(mean(Clinical_info$TumorPurity[which(Clinical_info$cms==i)],na.rm=TRUE))
})

barplot(ImmuneScore,names.arg = c("CMS1","CMS2","CMS3","CMS4"),
  col = rep("gray",4), beside=TRUE,main = "Immune Score",ylim = c(0, 1200))
barplot(StromalScore,names.arg = c("CMS1","CMS2","CMS3","CMS4"),
  col = rep("gray",4), beside=TRUE,main = "Stromal Score",ylim = c(-1500, 150))
barplot(TumorPurity,names.arg = c("CMS1","CMS2","CMS3","CMS4"),
  col = rep("gray",4), beside=TRUE,main = "Tumor Purity",ylim = c(0.595, 0.93))
```



Supplementary Figure 2b. Distribution of immune scores, stromal scores and tumor purity levels.



## 2.5 Correlation of Whole-slide SOFs and ESTIMATE scores

The whole-slide spatial organization features (SOFs) and ESTIMATE scores are significantly correlated

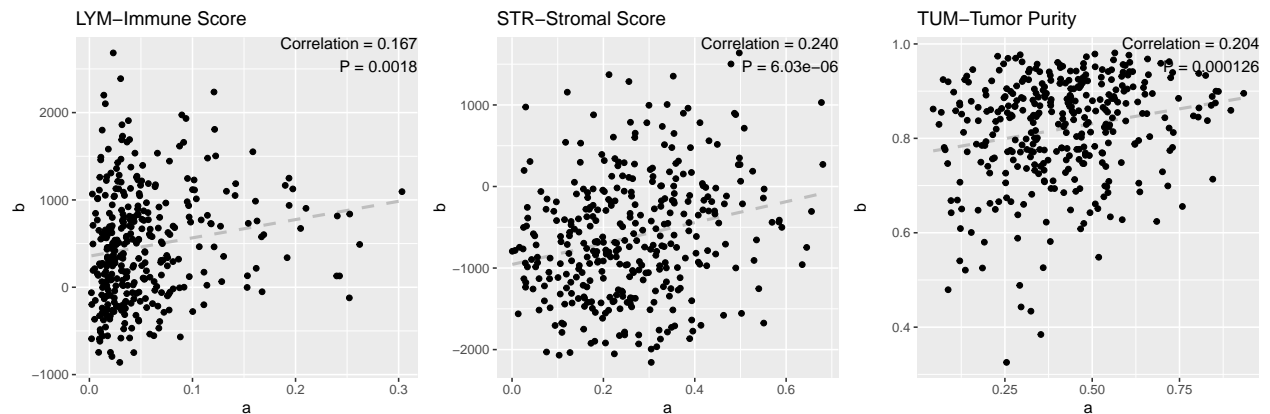
```
head(plotCorrelation)
```

```
##
## 1 function (x, y, groups = NULL, legend.pos = "none", maintitle = NULL)
## 2 {
## 3     pval <- paste(sprintf("Correlation = %.3f\\nP", Hmisc::rcorr(x,
## 4         y)$r[2, 1]), paste0(" = ", signif(Hmisc::rcorr(x, y)$P[2,
## 5         1], 3)))
## 6     cor_plot <- data.frame(a = x, b = y)
LYM_cor <- data.frame(a=as.numeric(WholeSlide_ratio$LYM),
                      b=as.numeric(Clinical_info$ImmuneScore))
LYM_cor_plot <- plotCorrelation(x=LYM_cor$a, y=LYM_cor$b, groups = NULL,
                              legend.pos = "none", maintitle="LYM-Immune Score")

STR_cor <- data.frame(a=as.numeric(WholeSlide_ratio$STR),
                      b=as.numeric(Clinical_info$StromalScore))
STR_cor_plot <- plotCorrelation(x=STR_cor$a, y=STR_cor$b, groups = NULL,
                              legend.pos = "none", maintitle="STR-Stromal Score")

TUM_cor <- data.frame(a=as.numeric(WholeSlide_ratio$TUM),
                      b=as.numeric(Clinical_info$TumorPurity))
TUM_cor_plot <- plotCorrelation(x=TUM_cor$a, y=TUM_cor$b, groups = NULL,
                              legend.pos = "none", maintitle="TUM-Tumor Purity")

plot_grid(LYM_cor_plot, STR_cor_plot, TUM_cor_plot, ncol = 3, align = "hv")
```



Supplementary Figure 3. Correlation of Whole-slide SOFs and ESTIMATE scores.

## 3 Infiltrating spatial organization features

### 3.1 Generating infiltrating SOFs

Whole-slide H&E stained images often contain large areas of tissues without tumor cells, which may not be functionally relevant and associated with clinical outcomes. To calculate SOFs in the proximity of tumor

tissues, a  $6 \times 6$  tissue patches window ( $1536 \times 1536$  pixels) with a 50% overlap (768-pixel step size) was used to scan the whole-slide image to identify regions of interests (ROIs).

```
InfiltrateRatio <- getInfiltrateRatio(windows=6,overlap=0.5,
                                     locationMatrix=CRC_locationMatrix)
InfiltrateRatio[,1:10]

##      TCGA-G4-6299 TCGA-D5-5537 TCGA-G4-6627 TCGA-CM-5860 TCGA-CA-6717
## ADI    0.014692484  0.004076541  0.002441924  0.0006254568  0.0004662303
## DEB    0.048981443  0.067021391  0.077074067  0.0429222866  0.3046551026
## LYM    0.058840433  0.019079081  0.011025295  0.0452484806  0.0077971592
## MUC    0.004969578  0.011489757  0.001984605  0.0187784270  0.0033998341
## MUS    0.006763237  0.013004925  0.032035513  0.0012644873  0.0011473395
## NORM   0.002686176  0.017262383  0.001919852  0.0136823166  0.0004114931
## STR    0.201378745  0.186224698  0.168297472  0.2079293170  0.1901127101
## TUM    0.571657943  0.605722637  0.644192064  0.6229111183  0.4711557189
##      TCGA-CM-6163 TCGA-CK-4952 TCGA-DM-A1D9 TCGA-D5-6530 TCGA-CA-5254
## ADI    0.001701865  0.004676458  0.0008846526  0.006491380  6.025987e-04
## DEB    0.077677024  0.075975281  0.0560346450  0.076422399  7.219081e-02
## LYM    0.055456466  0.020436218  0.0494136265  0.081147155  4.582562e-03
## MUC    0.025734287  0.053358216  0.0005722737  0.006159127  8.769769e-03
## MUS    0.002330367  0.003205521  0.0013710601  0.001122017  8.520504e-05
## NORM   0.043210805  0.055338607  0.0156673357  0.030276997  1.186715e-03
## STR    0.164928654  0.057045163  0.1110544901  0.184398771  2.204521e-01
## TUM    0.594711524  0.630540244  0.6895569925  0.559949737  6.149703e-01

all.equal(rownames(Clinical_info),colnames(InfiltrateRatio))

## [1] TRUE

Clinical_info$ILR <- InfiltrateRatio['LYM',]
Clinical_info$ISR <- InfiltrateRatio['STR',]
```

### 3.2 Univariate cox regression analyses

We found the TNM stage, infiltrating lymphocyte ratio (ILR), and infiltrating stroma ratio (ISR) were significantly associated with relapse-free survival. However, the stromal and immune scores calculated by ESTIMATE did not show significant associations with RFS.

```
index <- which(is.na(Clinical_info$rfs.delay)==TRUE |
              is.na(Clinical_info$rfs.event)==TRUE)
CoxCli <- Clinical_info[-index,]
factor_name <- c("sex", "age", "tumor.location", "tnm.stage",
               "mmr.status", "StromalScore", "ImmuneScore", "TumorPurity", "ILR", "ISR" )
clin_factors <- CoxCli[, factor_name]

dfs_time <- CoxCli$rfs.delay
dfs_event <- CoxCli$rfs.event
res <- Surv(dfs_time,dfs_event)

univ_formulas <- sapply(factor_name,function(x) as.formula(paste('res ~', x)))
univ_models <- lapply(univ_formulas, function(x){summary(coxph(x, data=CoxCli) ) })
univ_results <- sapply(1:length(univ_models),
                      function(i){
                        x <- univ_models[[i]]
```

```

        p <- x$logtest[3]
        HR <- x$conf.int[1,c(1,3,4)]
        p <- as.numeric(p)
        aa <- c(HR,p)
        return(aa)
    })
colnames(univ_results) <- factor_name
rownames(univ_results) <- c("HR", "L95", "H95", "logtest_P")
univ_results <- t(univ_results)
univ_results

```

##		HR	L95	H95	logtest_P
## sex		1.487183e+00	9.892930e-01	2.235652	5.378588e-02
## age		9.963307e-01	9.802691e-01	1.012655	6.578958e-01
## tumor.location		9.925716e-01	6.676921e-01	1.475528	9.705958e-01
## tnm.stage		2.450831e+00	1.896263e+00	3.167585	9.976250e-13
## mmr.status		8.681012e-01	5.724797e-01	1.316378	5.086756e-01
## StromalScore		1.000170e+00	9.999076e-01	1.000433	2.073830e-01
## ImmuneScore		9.999516e-01	9.996380e-01	1.000265	7.617121e-01
## TumorPurity		6.094148e-01	1.066465e-01	3.482405	5.815811e-01
## ILR		2.302741e-04	6.147393e-08	0.862580	3.120918e-02
## ISR		5.964958e+01	8.012285e+00	444.077094	1.291893e-04

Table 2. Univariate analyses of spatial organization features, ESTIMATE scores, clinical, and pathologic factors in the TCGA cohort.

### 3.3 Survival analyses of infiltrating SOFs

```

ILR <- CoxCli$ILR
ILR_label_F <- function(score,cut){
  tmp <- score >=cut
  tmp <- str_replace_all(tmp,"TRUE","LH")
  tmp <- str_replace_all(tmp,"FALSE","LL")
  return(tmp)
}
ILR_label <- as.factor(as.character(ILR_label_F(ILR,summary(ILR)[3])))

ISR_label_H <- function(score,cut){
  tmp <- score >=cut
  tmp <- str_replace_all(tmp,"TRUE","LHSH")
  tmp <- str_replace_all(tmp,"FALSE","LHSL")
  return(tmp)
}
ISR_label_L <- function(score,cut){
  tmp <- score >=cut
  tmp <- str_replace_all(tmp,"TRUE","LLSH")
  tmp <- str_replace_all(tmp,"FALSE","LLSL")
  return(tmp)
}

score1LH <- CoxCli$ISR[which(ILR_label=="LH")]
label1LH <- as.factor(as.character(ISR_label_H(score1LH,summary(score1LH)[3])))
score1LL <- CoxCli$ISR[which(ILR_label=="LL")]

```

```

labelLL <- as.factor(as.character(ISR_label_L(score1LL,summary(score1LL)[3])))
info <- rbind(CoxCli[which(ILR_label=="LH"),],CoxCli[which(ILR_label=="LL"),])
info$pa_labels <- c(as.character(labelLH) ,as.character(labelLL) )
labels <- info$pa_labels
labels <- factor(labels,levels=c("LLSL","LHSL","LLSH","LHSH"))
legend.labs <- as.vector(na.omit(unique(labels)))
input <- as.data.frame( cbind(info$rfs.delay,info$rfs.event))
plot_KMCurve(input,labels,ylab="Relapse Free Survival",
  color = c("#056DB7", "#F89F23","grey","red"),
  legend.pos = c(0.8,0.8),xlab="Months")

```

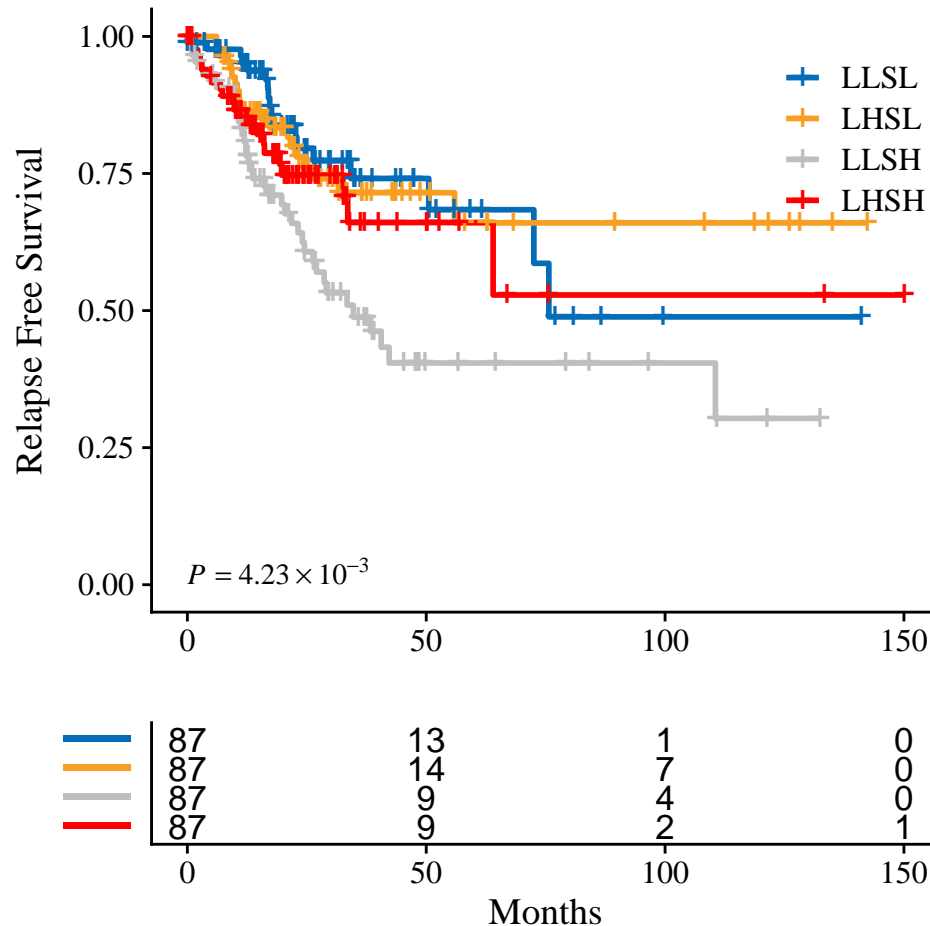


Figure 3. Infiltrating lymphocyte and stroma ratios are significant prognosticators of CRC in TCGA CRC cohort.

## 4 Session Info

This document was prepared using R package knitr. Function knit2pdf("sweave.rnw") was used to compile the sweave file and generate the pdf file.

```
sessionInfo()
```

```
## R version 4.0.0 (2020-04-24)
## Platform: x86_64-pc-linux-gnu (64-bit)
```

```

## Running under: Ubuntu 16.04.6 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/openblas-base/libblas.so.3
## LAPACK: /usr/lib/libopenblas-r0.2.18.so
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
## [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
## [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
## [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] stringr_1.4.0  survival_3.2-7 cowplot_1.1.0 ggplot2_3.3.2
##
## loaded via a namespace (and not attached):
## [1] tidyr_1.1.2      splines_4.0.0      carData_3.0-4
## [4] Formula_1.2-3    latticeExtra_0.6-29 cellranger_1.1.0
## [7] yaml_2.2.1       pillar_1.4.6       backports_1.1.10
## [10] lattice_0.20-41  glue_1.4.2         digest_0.6.25
## [13] RColorBrewer_1.1-2 ggsignif_0.6.0     checkmate_2.0.0
## [16] colorspace_1.4-1 htmltools_0.5.0    Matrix_1.2-18
## [19] pkgconfig_2.0.3  broom_0.7.1        haven_2.3.1
## [22] xtable_1.8-4     purrr_0.3.4        scales_1.1.1
## [25] km.ci_0.5-2      jpeg_0.1-8.1       openxlsx_4.2.2
## [28] rio_0.5.16       KMSurv_0.1-5       htmlTable_2.1.0
## [31] tibble_3.0.3     mgcv_1.8-33        generics_0.0.2
## [34] farver_2.0.3     car_3.0-10         ellipsis_0.3.1
## [37] ggpubr_0.4.0     withr_2.3.0        nnet_7.3-14
## [40] magrittr_1.5     crayon_1.3.4       readxl_1.3.1
## [43] evaluate_0.14    nlme_3.1-149       rstatix_0.6.0
## [46] forcats_0.5.0    foreign_0.8-80     tools_4.0.0
## [49] data.table_1.13.0 hms_0.5.3          lifecycle_0.2.0
## [52] munsell_0.5.0    cluster_2.1.0      zip_2.1.1
## [55] compiler_4.0.0   survminer_0.4.8    rlang_0.4.8
## [58] grid_4.0.0       rstudioapi_0.11    htmlwidgets_1.5.2
## [61] base64enc_0.1-3  labeling_0.3       rmarkdown_2.4
## [64] gtable_0.3.0     abind_1.4-5        curl_4.3
## [67] R6_2.4.1         zoo_1.8-8          gridExtra_2.3
## [70] knitr_1.30       dplyr_1.0.2        survMisc_0.5.5
## [73] Hmisc_4.4-1     stringi_1.5.3      Rcpp_1.0.5
## [76] vctrs_0.3.4     rpart_4.1-15       png_0.1-7
## [79] tidyselect_1.1.0 xfun_0.18

```