

# R code and data for – Microenvironmental niche divergence shapes BRCA1-related cancer morphological plasticity and disease progression in ovarian cancer –

## Contents

<b>1</b>	<b>Preparation</b>	<b>2</b>
1.1	Load data and functions . . . . .	2
1.2	Clinical parameters of the ovarian cancer samples . . . . .	2
<b>2</b>	<b>Survival analysis</b>	<b>4</b>
2.1	Pan subtypes . . . . .	4
2.2	Microenvironmentally defined subtypes . . . . .	5
2.3	Immunoreactive subtype . . . . .	12
<b>3</b>	<b>Cell abundance distributions</b>	<b>14</b>
<b>4</b>	<b>Discovery and Validation set</b>	<b>15</b>
4.1	Survival analysis for discovery and validation set . . . . .	16
<b>5</b>	<b>Prognostic value of the microenvironment</b>	<b>17</b>
<b>6</b>	<b>Session Info</b>	<b>20</b>

# 1 Preparation

## 1.1 Load data and functions

To reproduce our result, first source the R functions needed for analysis, plotting and load the data for the ovarian cancer samples, which is available as part of the paper supplement.

```
library(survival)
PUBDIR<-'D:\\Ovarian-Sweve\\sweave\\'
setwd(PUBDIR)
load(paste0(PUBDIR, 'trait.rdata'))
source('YYfuns.R')
```

## 1.2 Clinical parameters of the ovarian cancer samples

Each row in this clinical file is a sample/patient.

```
# Show clinical information data
head(trait)

##      PATIENT_ID hasNiche      SUBTYPE      c_BS1
## 6482 TCGA-04-1331    FALSE Proliferative 0.5664574
## 6483 TCGA-04-1332     TRUE  Mesenchymal 0.5480628
## 6485 TCGA-04-1336     TRUE Differentiated 0.6781114
## 6486 TCGA-04-1337     TRUE  Mesenchymal 0.6110431
## 6487 TCGA-04-1338     TRUE  Mesenchymal 0.6208161
## 6488 TCGA-04-1341     TRUE Proliferative 0.6710611
##      l_BS1      o_BS1      l      c      o
## 6482 0.2151915 0.19246773 0.2151915 0.5664574 0.19246773
## 6483 0.1641943 0.27514810 0.1641943 0.5480628 0.27514810
## 6485 0.2204599 0.08545062 0.2204599 0.6781114 0.08545062
## 6486 0.1448838 0.22646021 0.1448838 0.6110431 0.22646021
## 6487 0.1694143 0.20585924 0.1694143 0.6208161 0.20585924
## 6488 0.1290206 0.16955025 0.1290206 0.6710611 0.16955025
##      MutBRCA1 MutBRCA2 BRCA1METH BRCA2METH tumor.purity
## 6482      0      1 0.05339720 0.8804598      0.88
## 6483      0      0 0.06456207 0.9024191      0.48
## 6485      0      1 0.16629090 0.7742857      0.92
## 6486      0     NA 0.05528291 0.8079238      0.50
## 6487      0      0 0.05428533 0.1547611      0.75
## 6488     NA     NA 0.04682960 0.8878061      0.82
##      ESTIMATE.score percent_tumor_nuclei_BS1
## 6482      3536.80      88
## 6483      5281.64      89
## 6485      4032.06      91
## 6486      5945.10      84
## 6487      6415.36      85
## 6488      3142.08      80
##      percent_tumor_cells_BS1 percent_stromal_cells_BS1
## 6482      30      65
## 6483      40      52
## 6485      83      0
## 6486      50      50
## 6487      45      40
```

```

## 6488          55          28
##      Microenvironmental.Subtype ABSOLUTE Stromal.score
## 6482          Immunolike      0.88      2800.13
## 6483      HighLymHighStromal      0.50      3568.16
## 6485          Immunolike      0.92      1705.66
## 6486          Mesenchymallike      0.50      3844.68
## 6487      HighLymHighStromal      0.74      3647.83
## 6488          LowLymLowStromal      0.82      2330.79
##      Immune.score Arm.SCNA.Level
## 6482      736.68      -1.0128
## 6483      1713.47      -1.0128
## 6485      2326.41      0.1100
## 6486      2100.42      0.3907
## 6487      2767.53      0.1100
## 6488      811.29      <NA>
##      SCNA.Level.normalized.by.size Chrom.Arm.SCNA.Level
## 6482          -0.6261          -1.7723
## 6483          -1.0535          -1.1977
## 6485          -0.4999          -0.9367
## 6486          -0.5139          -0.0815
## 6487          1.1540          0.4996
## 6488          <NA>          <NA>
##      Focal.SCNA.Level Chrom.SCNA.Level Chrom.SCNA.Level.1
## 6482          0.3733          -0.7594          -0.7594
## 6483          -0.1251          -0.1849          -0.1849
## 6485          -0.6236          -1.0467          -1.0467
## 6486          -1.6205          -0.4722          -0.4722
## 6487          2.0111          0.3896          0.3896
## 6488          <NA>          <NA>          <NA>
##      CellCycle.Signature.Score CNA Mutation.Count
## 6482          0.6861 0.437      78
## 6483          0.4294 0.371      24
## 6485          0.4886 0.523      51
## 6486          0.1143 0.329      84
## 6487          0.6168 0.659     118
## 6488          <NA> 0.722      NA
##      TotNofMutations.in.exons Immune.Signature.Score
## 6482          94          0.1961
## 6483          32          0.3198
## 6485          <NA>          0.6321
## 6486          97          0.7499
## 6487          147         0.9433
## 6488          <NA>          <NA>
##      S_10year OS10 DFS10 TISSUE.SOURCE.SITE
## 6482      43.9 43.9 15.1      4
## 6483      41.0 41.0 12.9      4
## 6485      49.1+ 49.1+ 49.1+      4
## 6486      2.0 2.0 NA?      4
## 6487      46.6+ 46.6+ 12.5      4
## 6488      NA+ NA+ NA+      4

```

The summary statistics of OS survival:

```
summary(trait$OS10)
```

```
##      time      status
## Min.   : 0.26   Min.   :0.0000
## 1st Qu.: 11.70  1st Qu.:0.0000
## Median : 29.10  Median :1.0000
## Mean   : 33.15  Mean   :0.5322
## 3rd Qu.: 47.65  3rd Qu.:1.0000
## Max.   :120.00  Max.   :1.0000
## NA's   :3      NA's   :1
```

The summary statistics of RFS survival:

```
summary(trait$DFS10)
```

```
##      time      status
## Min.   : 0.530   Min.   :0.0000
## 1st Qu.: 8.273   1st Qu.:0.0000
## Median : 13.800  Median :1.0000
## Mean   : 19.250  Mean   :0.6674
## 3rd Qu.: 23.250  3rd Qu.:1.0000
## Max.   :120.000  Max.   :1.0000
## NA's   :74      NA's   :72
```

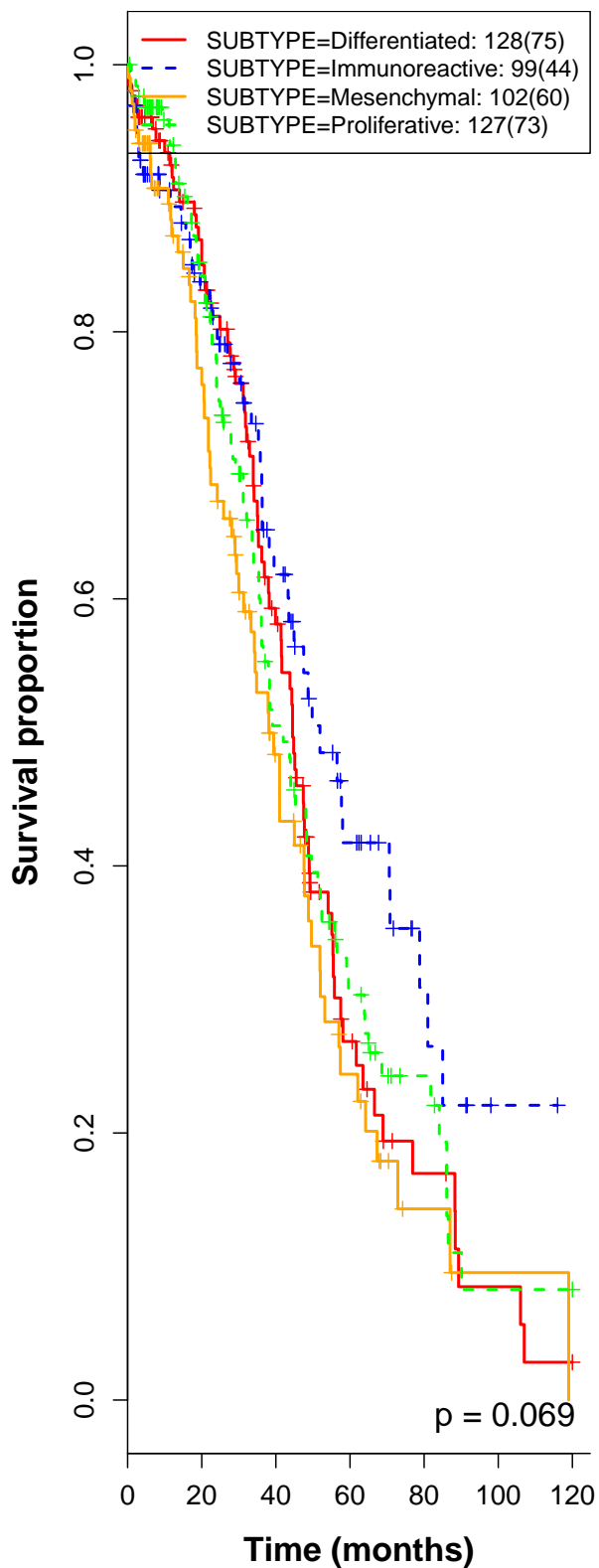
## 2 Survival analysis

### 2.1 Pan subtypes

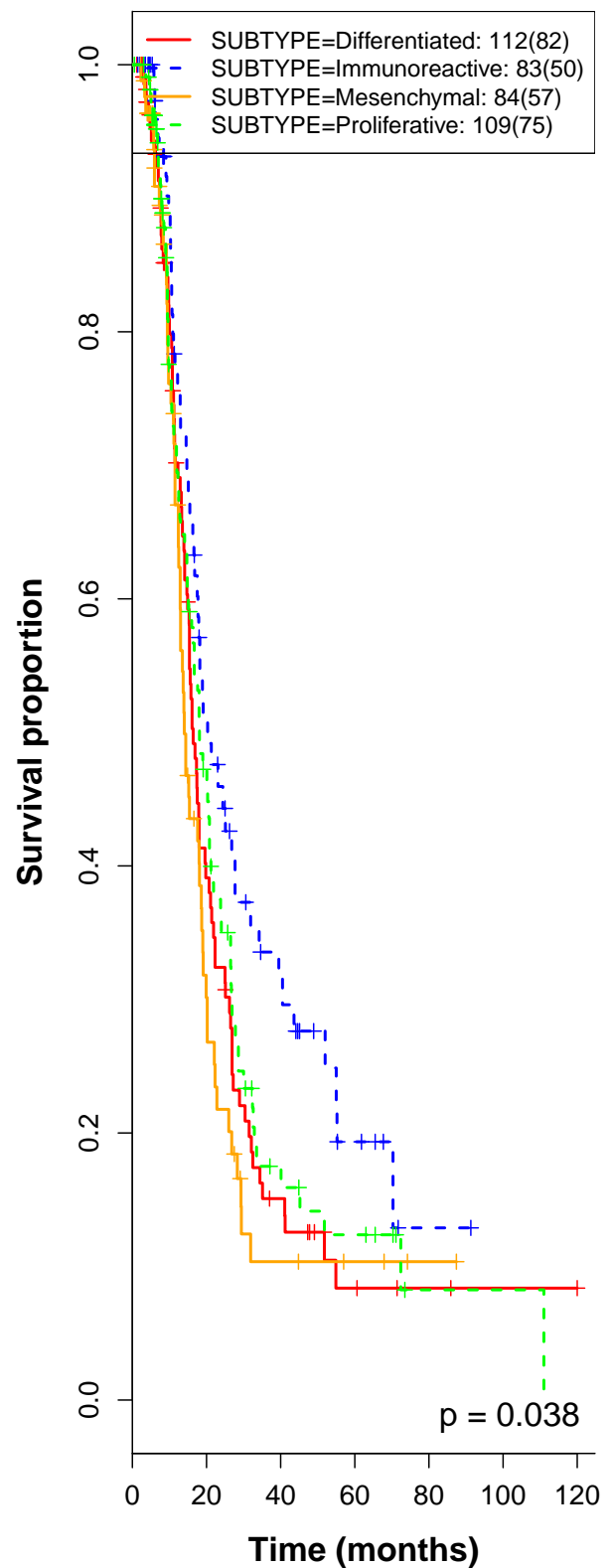
```
# OS10
par(mfrow = c(1, 2))
SurvDIFF <- survdiff(formula = OS10 ~ SUBTYPE, data = trait)
p.val <- 1 - pchisq(SurvDIFF$chisq, length(SurvDIFF$n) - 1)
SurvFIT_KM <- survfit(formula = OS10 ~ SUBTYPE, type = "kaplan-meier",
  data = trait)
ColourCode <- c("red", "blue", "orange", "green")
plotKM(SurvDIFF, SurvFIT_KM, main = paste0("OS10~molecular subtypes"),
  measure = "f", ThresholdLegend = 0.05, pval = TRUE, colourCurves = ColourCode,
  lineType = c(1, 2, 1, 2), Legend.Names = names(SurvDIFF$n))

# DFS10
SurvDIFF <- survdiff(formula = DFS10 ~ SUBTYPE, data = trait)
p.val <- 1 - pchisq(SurvDIFF$chisq, length(SurvDIFF$n) - 1)
SurvFIT_KM <- survfit(formula = DFS10 ~ SUBTYPE, type = "kaplan-meier",
  data = trait)
ColourCode <- c("red", "blue", "orange", "green")
plotKM(SurvDIFF, SurvFIT_KM, main = paste0("RFS10~molecular subtypes"),
  measure = "f", ThresholdLegend = 0.05, pval = TRUE, colourCurves = ColourCode,
  lineType = c(1, 2, 1, 2), Legend.Names = names(SurvDIFF$n))
```

## OS10~molecular subtypes



## RFS10~molecular subtypes

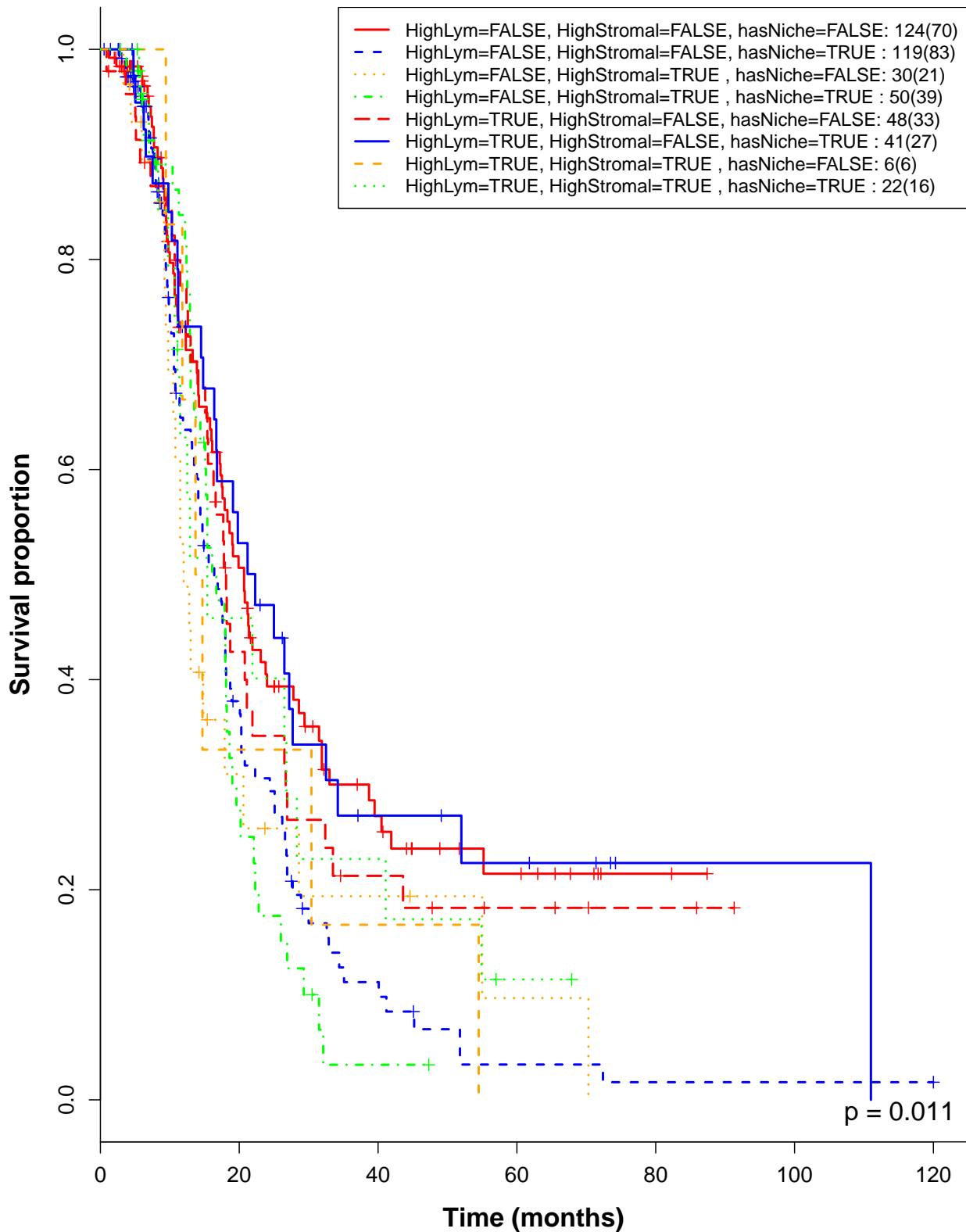


## 2.2 Microenvironmentally defined subtypes

```
# Microenvironmental subtype RFS10  
HighLym <- Groups3(trait$l_BS1, c(0.25, 0.75)) == 3
```

```
HighStromal <- Groups3(trait$o_BS1, c(0.25, 0.75)) == 3
SurvDIFF <- survdiff(formula = DFS10 ~ HighLym + HighStromal +
  hasNiche, data = trait)
SurvFIT_KM <- survfit(formula = DFS10 ~ HighLym + HighStromal +
  hasNiche, type = "kaplan-meier", data = trait)
plotKM(SurvDIFF, SurvFIT_KM, main = paste0("Microenv. subtypes ~ RFS10"),
  measure = iStatsFeature, ThresholdLegend = "", pval = TRUE,
  colourCurves = c("red", "blue", "orange", "green"), lineType = c(1:5),
  Legend.Names = names(SurvDIFF$n))
```

## Microenv. subtypes ~ RFS10



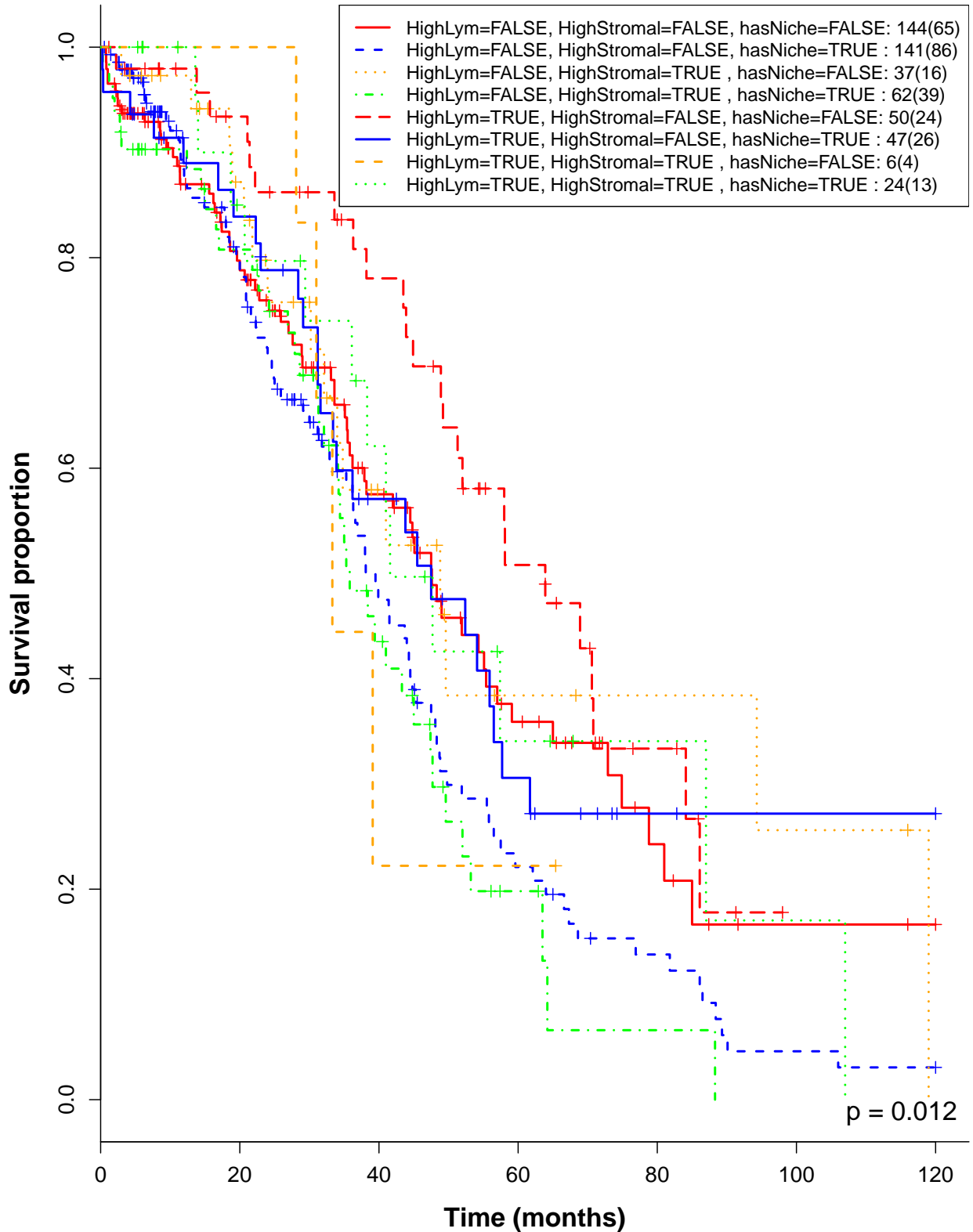
```
HighLym <- Groups3(trait$l_BS1, c(0.25, 0.75)) == 3
HighStromal <- Groups3(trait$o_BS1, c(0.25, 0.75)) == 3
SurvvDIFF <- survdiff(formula = OS10 ~ HighLym + HighStromal +
  hasNiche, data = trait)
SurvFIT_KM <- survfit(formula = OS10 ~ HighLym + HighStromal +
```

```

hasNiche, type = "kaplan-meier", data = trait)
plotKM(SurvDIFF, SurvFIT_KM, main = paste0("Microenv. subtypes ~ OS10"),
measure = iStatsFeature, ThresholdLegend = "", pval = TRUE,
colourCurves = c("red", "blue", "orange", "green"), lineType = c(1:5),
Legend.Names = names(SurvDIFF$n))

```

### Microenv. subtypes ~ OS10





```

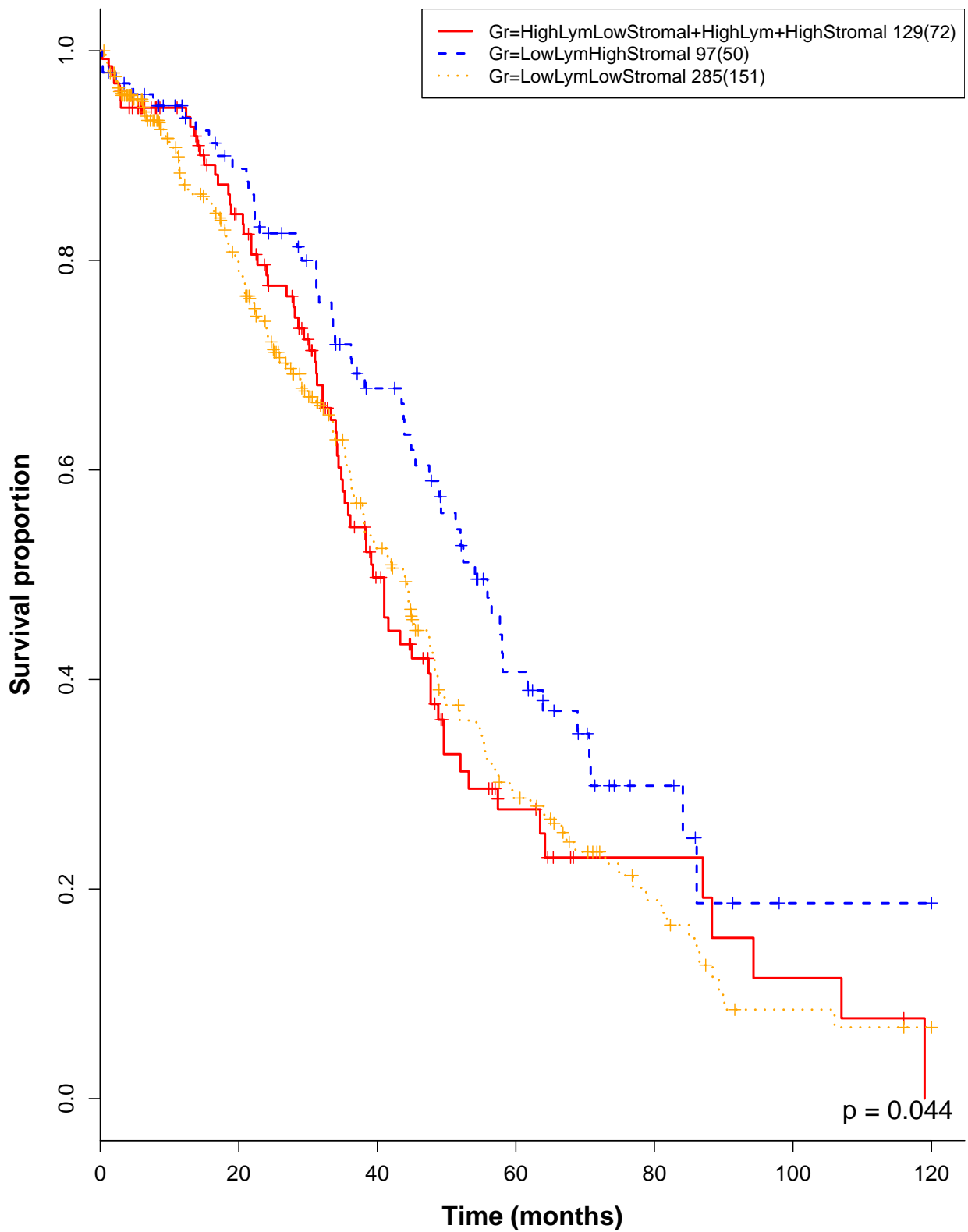
# Merged

HighStromal <- Groups3(trait$o_BS1, c(0.25, 0.75)) == 3
HighLym <- Groups3(trait$l_BS1, c(0.25, 0.75)) == 3
Gr <- interaction(HighStromal, HighLym)
Gr <- as.character(Gr)
Gr[Gr == "TRUE.FALSE" | Gr == "TRUE.TRUE"] <- "HighLymLowStromal+HighLym+HighStromal"
Gr[Gr == "FALSE.FALSE"] <- "LowLymLowStromal"
Gr[Gr == "FALSE.TRUE"] <- "LowLymHighStromal"

SurvDIFF <- survdiff(formula = OS10 ~ Gr, data = trait)
SurvFIT_KM <- survfit(formula = OS10 ~ Gr, type = "kaplan-meier",
  data = trait)
plotKM(SurvDIFF, SurvFIT_KM, main = paste0("Microenv. subtypes ~ OS10"),
  measure = iStatsFeature, ThresholdLegend = "", pval = TRUE,
  colourCurves = c("red", "blue", "orange", "green"), lineType = c(1:5),
  Legend.Names = names(SurvDIFF$n))

```

## Microenv. subtypes ~ OS10



```
summary(coxph(formula = OS10 ~ Gr == "LowLymHighStromal", data = trait)) #p=0.0104
```

```
## Call:
```

```
## coxph(formula = OS10 ~ Gr == "LowLymHighStromal", data = trait)
```

```
##
```

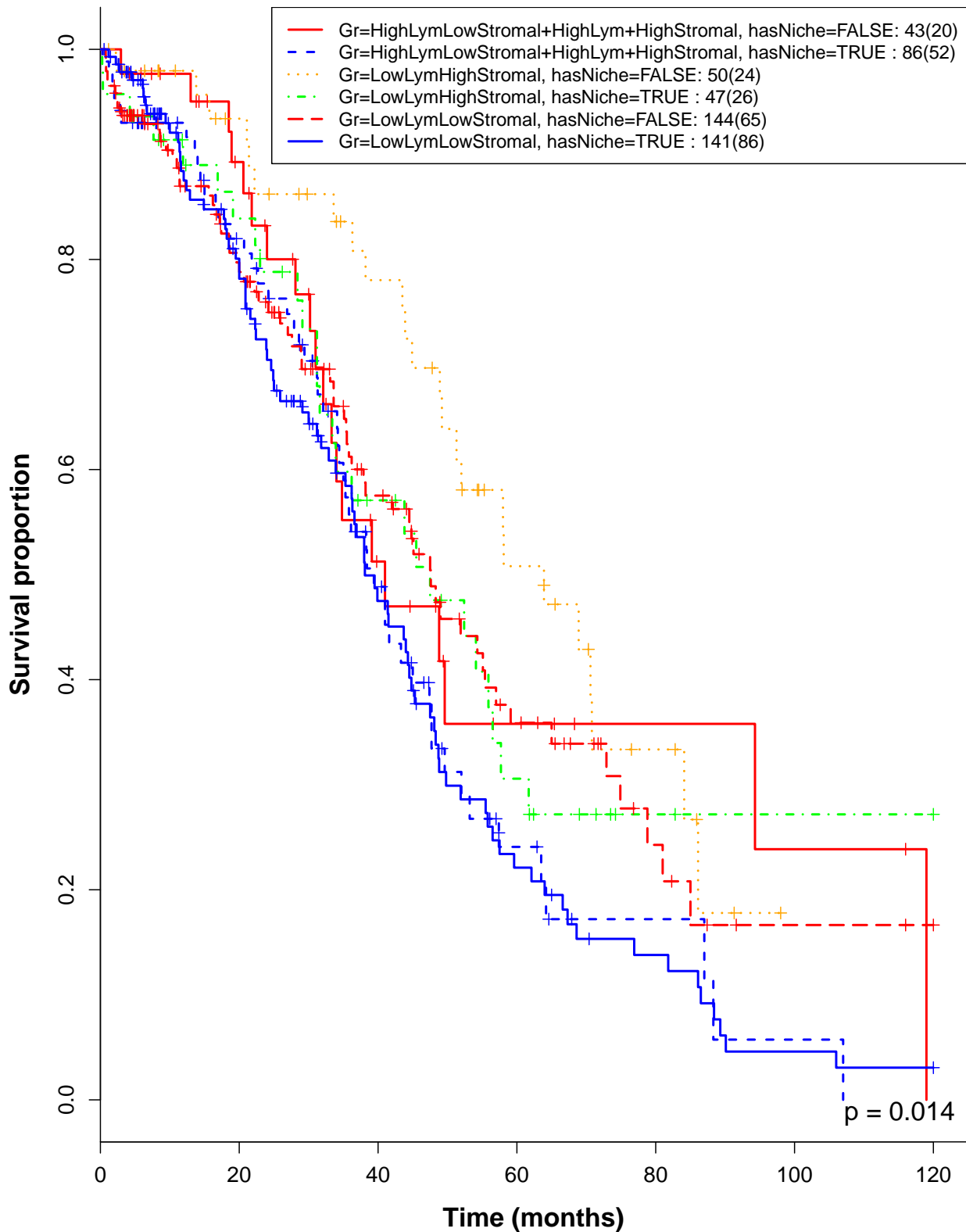
```

## n= 511, number of events= 273
## (3 observations deleted due to missingness)
##
##               coef exp(coef) se(coef)
## Gr == "LowLymHighStromal"TRUE -0.3897  0.6773  0.1572
##               z Pr(>|z|)
## Gr == "LowLymHighStromal"TRUE -2.479  0.0132 *
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef)
## Gr == "LowLymHighStromal"TRUE  0.6773  1.476
##               lower .95 upper .95
## Gr == "LowLymHighStromal"TRUE  0.4977  0.9217
##
## Concordance= 0.535 (se = 0.014 )
## Rsquare= 0.013 (max possible= 0.996 )
## Likelihood ratio test= 6.64 on 1 df, p=0.01
## Wald test = 6.14 on 1 df, p=0.01
## Score (logrank) test = 6.22 on 1 df, p=0.01

SurvDIFF <- survdiff(formula = OS10 ~ Gr + hasNiche, data = trait)
SurvFIT_KM <- survfit(formula = OS10 ~ Gr + hasNiche, type = "kaplan-meier",
  data = trait)
plotKM(SurvDIFF, SurvFIT_KM, main = paste0("Microenv. subtypes ~ OS10"),
  measure = iStatsFeature, ThresholdLegend = "", pval = TRUE,
  colourCurves = c("red", "blue", "orange", "green"), lineType = c(1:5),
  Legend.Names = names(SurvDIFF$n))

```

## Microenv. subtypes ~ OS10



### 2.3 Immunoreactive subtype

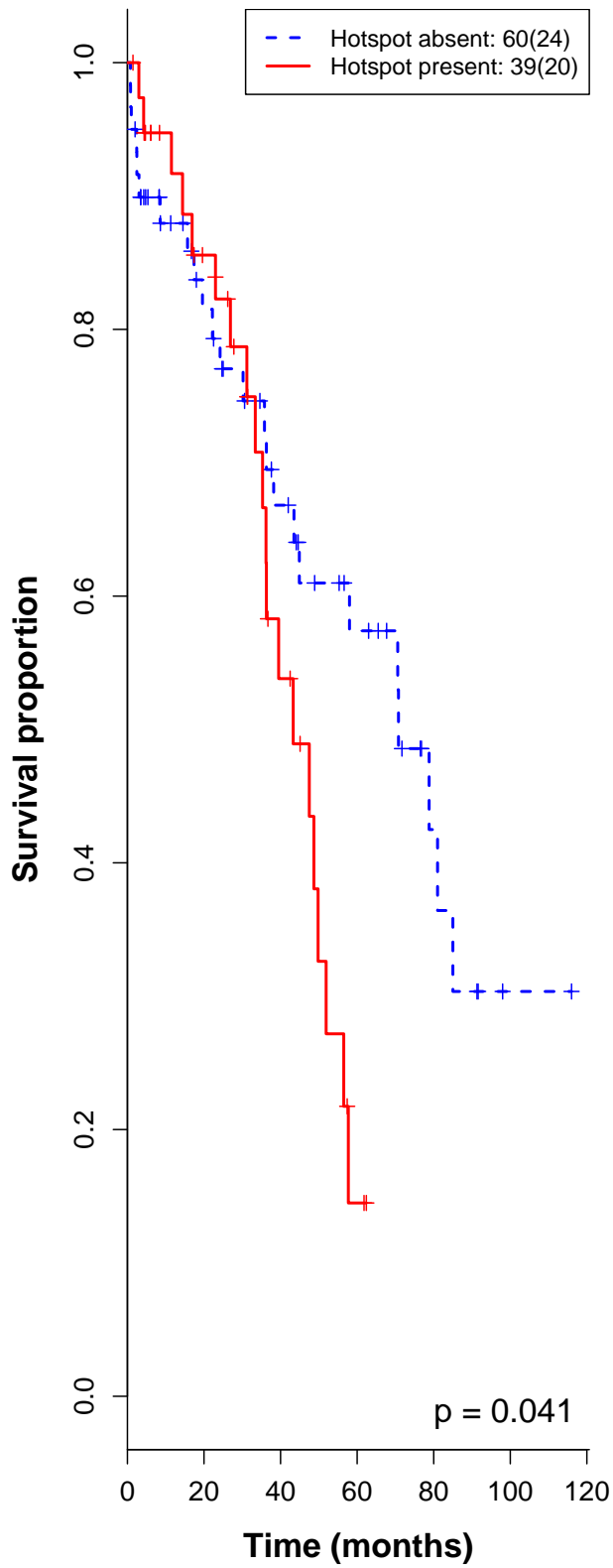
```
par(mfrow = c(1, 2))  
S.Types <- c("OS10", "DFS10")
```

```

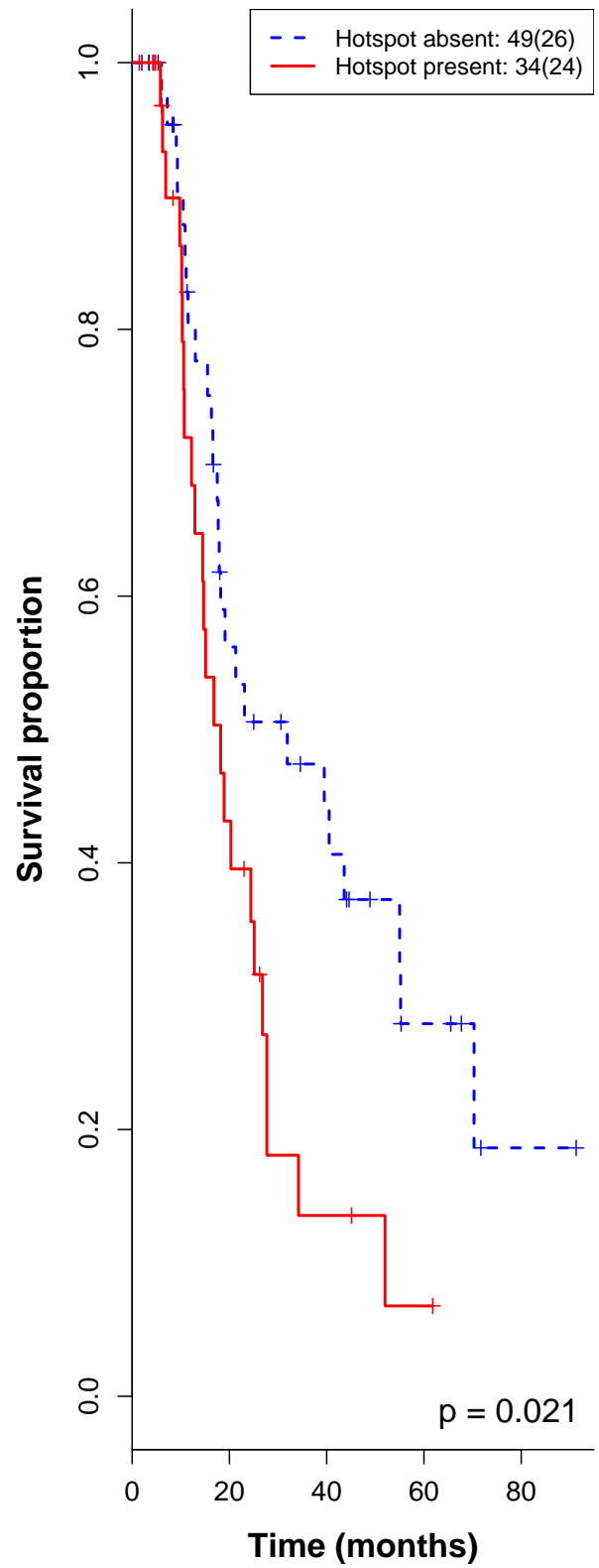
idx.Immuno <- (trait$SUBTYPE == "Immunoreactive" & !is.na(trait$SUBTYPE))
gr.imm <- trait$hasNiche[idx.Immuno]
for (iS.Type in S.Types) {
  trait3.imm <- trait[idx.Immuno, ]
  SurvDIFF <- eval(parse(text = paste0("survdif(formula = ",
    iS.Type, " ~ ", "gr.imm ,data = trait3.imm)")))
  p.val <- 1 - pchisq(SurvDIFF$chisq, length(SurvDIFF$n) -
    1)
  SurvFIT_KM <- eval(parse(text = paste0("survfit(formula = ",
    iS.Type, " ~ ", "gr.imm, type=\"kaplan-meier\",data = trait3.imm)")))
  # ColourCode Has niche red, has no niche green
  ColourCode <- c("blue", "red", "orange", "green")
  plotKM(SurvDIFF, SurvFIT_KM, main = paste0("SUBTYPE ~ ",
    iS.Type), pval = TRUE, ThresholdLegend = c(NULL, NULL),
    Legend.Names = c("Hotspot absent", "Hotspot present"),
    colourCurves = ColourCode, Legend.Symbol = c("", ""),
    lineType = c(2, 1), Legend = T)
}

```

### SUBTYPE ~ OS10



### SUBTYPE ~ DFS10



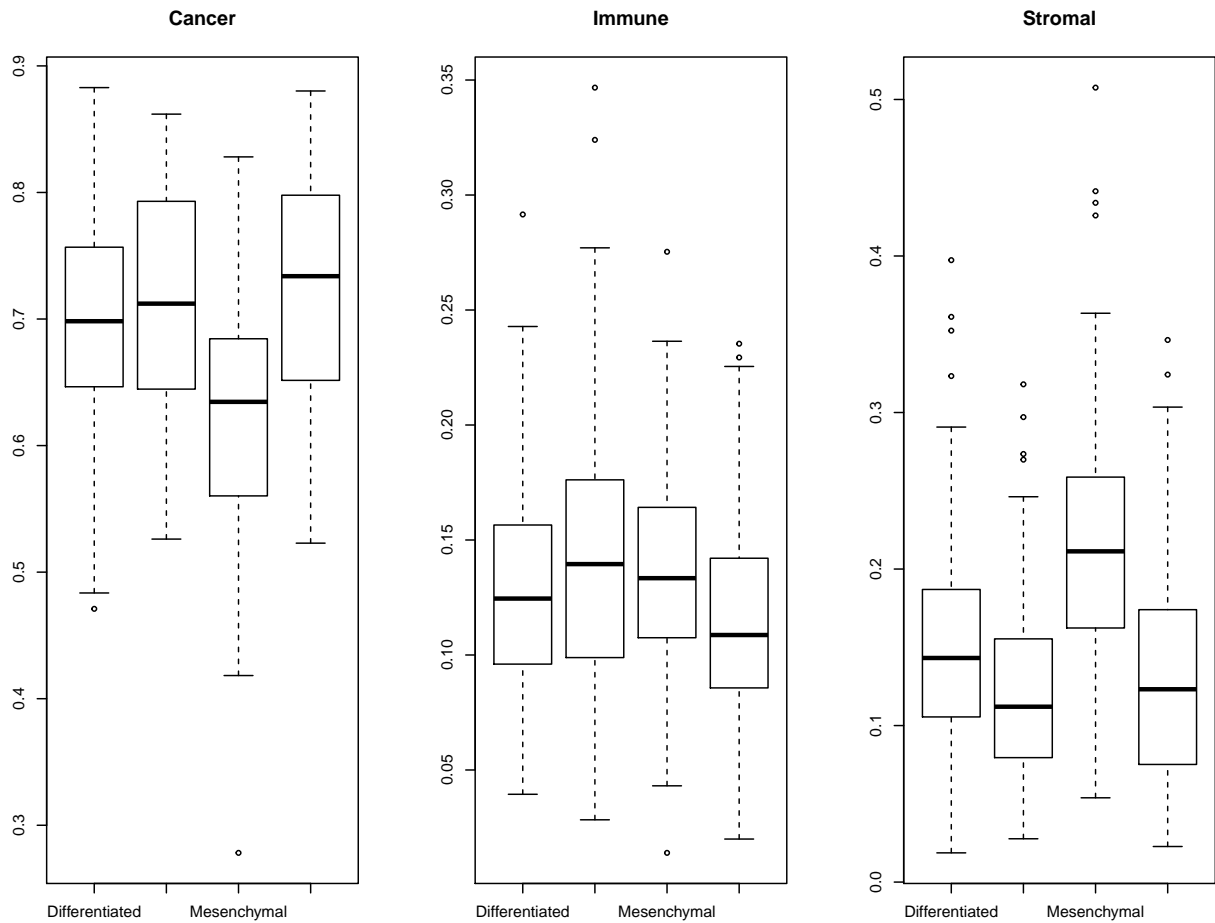
## 3 Cell abundance distributions

Cell ratios per molecular subtype

```

par(mfrow = c(1, 3))
boxplot(trait$c_BS1 ~ trait$SUBTYPE, main = "Cancer")
boxplot(trait$l_BS1 ~ trait$SUBTYPE, main = "Immune")
boxplot(trait$o_BS1 ~ trait$SUBTYPE, main = "Stromal")

```



## 4 Discovery and Validation set

The discovery cohort was generated by merging samples from two hospitals (University of Pittsburgh and Memorial Sloan Kettering) and test set was generated by merging the remaining sites.

```

## Divide data in training and test set
Test <- names(sort(table(trait$TISSUE.SOURCE.SITE)))[c(1:17,
19)]
Training <- names(sort(table(trait$TISSUE.SOURCE.SITE)))[c(18,
20)]
Training.set <- subset(trait, trait$TISSUE.SOURCE.SITE %in% Training)
Test.set <- subset(trait, trait$TISSUE.SOURCE.SITE %in% Test)
print(dim(Training.set))

## [1] 156 37

print(dim(Test.set))

## [1] 358 37

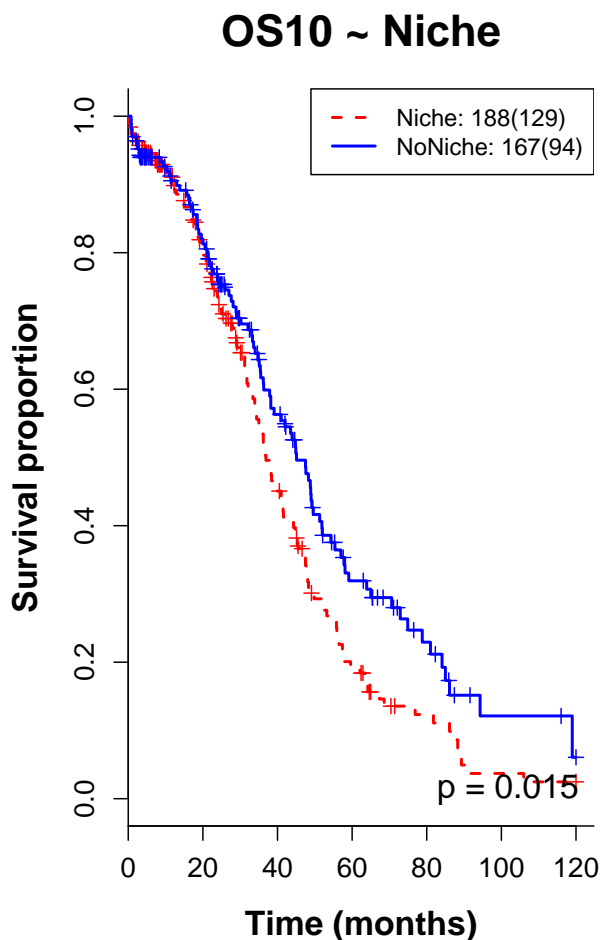
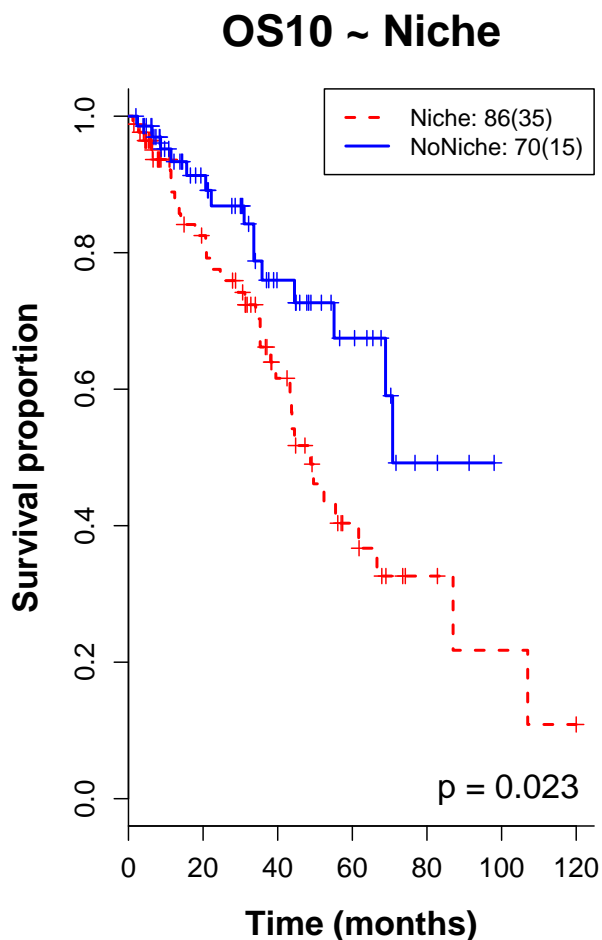
```

## 4.1 Survival analysis for discovery and validation set

```
# Trainings set
par(mfrow = c(1, 2))

SurvDIFF <- survdiff(formula = Training.set$OS10 ~ factor(Training.set$hasNiche,
  levels = c("TRUE", "FALSE")))
SurvFIT_KM <- survfit(formula = Training.set$OS10 ~ factor(Training.set$hasNiche,
  levels = c("TRUE", "FALSE")), type = "kaplan-meier")
plotKM(SurvDIFF, SurvFIT_KM, main = "OS10 ~ Niche", measure = "",
  ThresholdLegend = "", pval = T, colourCurves = c("red", "blue"),
  Legend.Names = c("Niche", "NoNiche"), lineType = c(2, 1),
  Legend.Symbol = c("", ""))

# Test set
SurvDIFF <- survdiff(formula = Test.set$OS10 ~ factor(Test.set$hasNiche,
  levels = c("TRUE", "FALSE")))
SurvFIT_KM <- survfit(formula = Test.set$OS10 ~ factor(Test.set$hasNiche,
  levels = c("TRUE", "FALSE")), type = "kaplan-meier")
plotKM(SurvDIFF, SurvFIT_KM, main = "OS10 ~ Niche", measure = "",
  ThresholdLegend = "", pval = T, colourCurves = c("red", "blue"),
  Legend.Names = c("Niche", "NoNiche"), lineType = c(2, 1),
  Legend.Symbol = c("", ""))
```



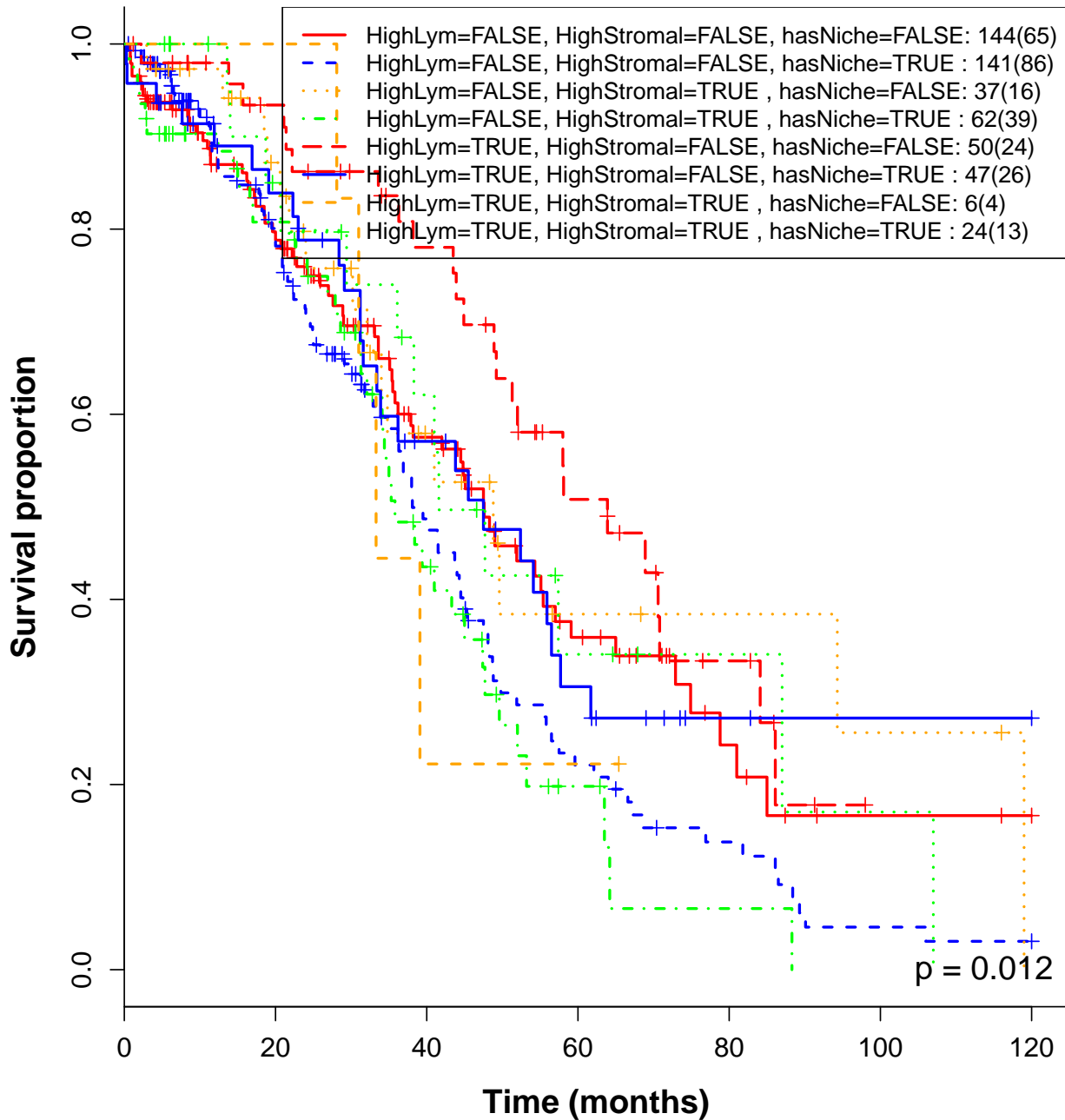


## 5 Prognostic value of the microenvironment

```
# par(mfrow=c(2,1))
HighLym <- Groups3(trait$l_BS1, c(0.25, 0.75)) == 3
HighStromal <- Groups3(trait$o_BS1, c(0.25, 0.75)) == 3

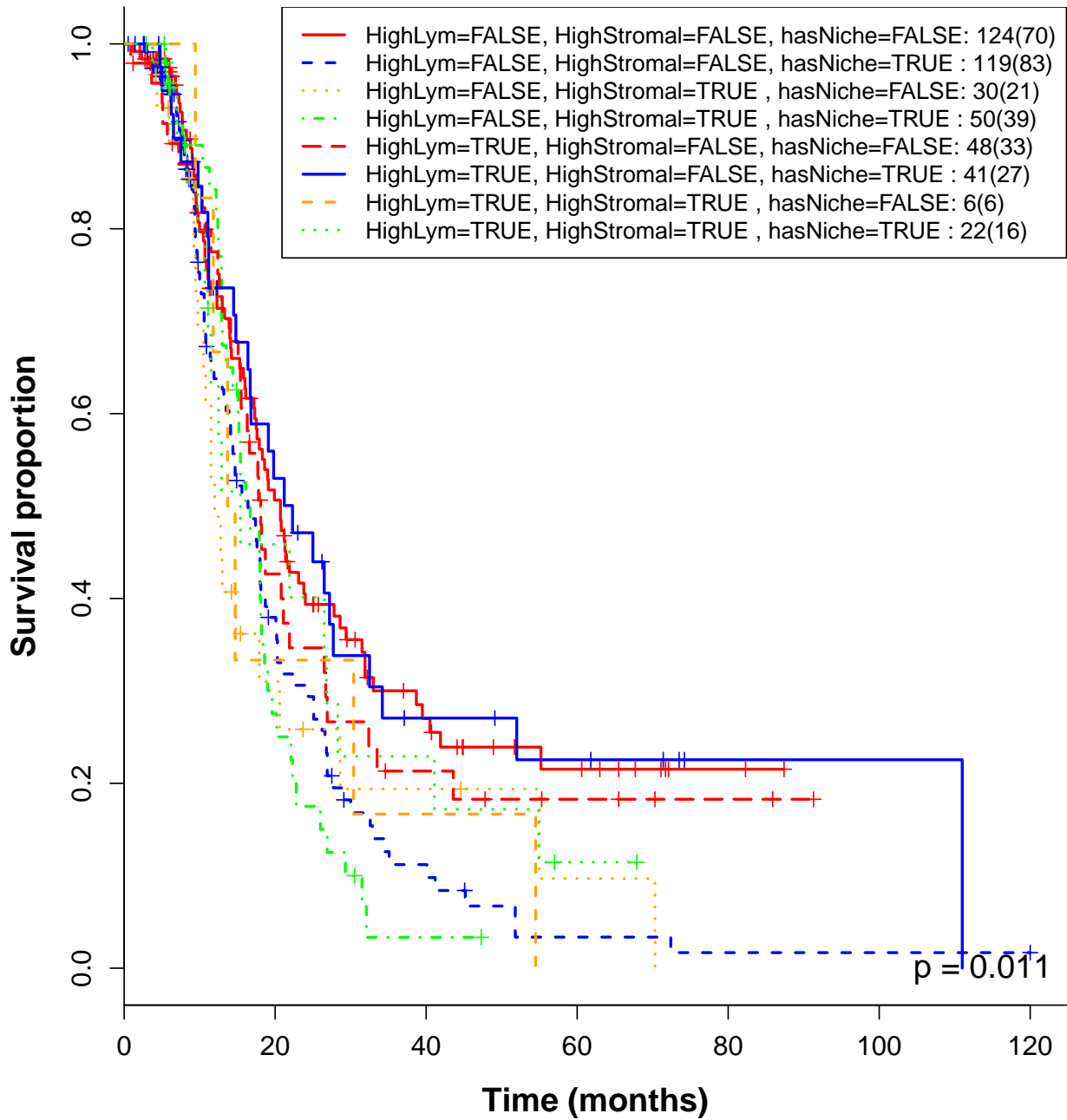
SurvDIFF <- survdiff(formula = OS10 ~ HighLym + HighStromal +
  hasNiche, data = trait)
SurvFIT_KM <- survfit(formula = OS10 ~ HighLym + HighStromal +
  hasNiche, type = "kaplan-meier", data = trait)
plotKM(SurvDIFF, SurvFIT_KM, main = paste0("Microenv. subtypes ~ OS10"),
  measure = iStatsFeature, ThresholdLegend = "", pval = TRUE,
  colourCurves = c("red", "blue", "orange", "green"), lineType = c(1:5),
  Legend.Names = names(SurvDIFF$n))
```

## Microenv. subtypes ~ OS10



```
SurvDIFF <- survdiff(formula = DFS10 ~ HighLym + HighStromal +
  hasNiche, data = trait)
SurvFIT_KM <- survfit(formula = DFS10 ~ HighLym + HighStromal +
  hasNiche, type = "kaplan-meier", data = trait)
plotKM(SurvDIFF, SurvFIT_KM, main = paste0("Microenv. subtypes ~ RFS0"),
  measure = iStatsFeature, ThresholdLegend = "", pval = TRUE,
  colourCurves = c("red", "blue", "orange", "green"), lineType = c(1:5),
  Legend.Names = names(SurvDIFF$n))
```

## Microenv. subtypes ~ RFS0



## 6 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 3.5.1 (2018-07-02)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17134)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United Kingdom.1252
## [2] LC_CTYPE=English_United Kingdom.1252
## [3] LC_MONETARY=English_United Kingdom.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United Kingdom.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets
## [6] methods   base
##
## other attached packages:
## [1] survival_2.42-3 knitr_1.20
##
## loaded via a namespace (and not attached):
## [1] compiler_3.5.1 Matrix_1.2-14 magrittr_1.5
## [4] formatR_1.5    tools_3.5.1    splines_3.5.1
## [7] stringi_1.1.7  highr_0.7      grid_3.5.1
## [10] stringr_1.3.1  lattice_0.20-35 evaluate_0.11
```