

CD138+ spectra and clinical risk

Rosalie Griffin Waller

27-OCT-2020

Code to investigate associations between CD138+ spectra and clinical risk cytogenetics and disease stage

0. Setup

```
# Install and load required R packages
library(ggplot2)
library(MASS)
library(survivalAnalysis)
library(dplyr)
library(data.table)
```

Define data directory

```
data_dir = "/path/to/data" # exclude ending "/"
```

Load transcriptome spectra (PC1-PC)

```
# read in spectra and clinical data
spectra_clinical = read.csv(file = paste0(data_dir,
                                           "/baseline-clinical-spectra-sd.csv")
                           ) %>% data.table()
```

1. Cytogenetics

Large somatic chromosomal DNA aberrations detected by cytogenetics are used to define prognostic risk groups in myeloma. Clinical risk categories defined by mSMART7 include: high risk, del(17p) and t(14;16); intermediate risk, amp(1q) and t(4;14); and standard risk, t(11;14).

1.1. del(17p): D_TRI_CF_ABNORMALITYPR11 & D_TRI_CF_17PABNORMALCE (abundance)

```
DAT = spectra_clinical %>%
  dplyr::select("D_TRI_CF_ABNORMALITYPR11", starts_with("PC")) %>%
  dplyr::filter(D_TRI_CF_ABNORMALITYPR11 !="Not Done" &
               D_TRI_CF_ABNORMALITYPR11 !="")

print(paste("Baseline samples with data:", nrow(DAT), "(total)", ",
          nrow(DAT[DAT$D_TRI_CF_ABNORMALITYPR11==\"Yes\",])", "(+)", ",
          nrow(DAT[DAT$D_TRI_CF_ABNORMALITYPR11==\"No\",])", "(-)"))
```

```
## [1] "Baseline samples with data: 436 (total), 52 (+), 384 (-)"
```

```
DAT$D_TRI_CF_ABNORMALITYPR11 = as.factor(DAT$D_TRI_CF_ABNORMALITYPR11)
```

```
# Logistic regression
```

```
MOD = glm(data = DAT, formula = D_TRI_CF_ABNORMALITYPR11 ~ ., family = "binomial")  
summary(MOD)
```

```
##
```

```
## Call:
```

```
## glm(formula = D_TRI_CF_ABNORMALITYPR11 ~ ., family = "binomial",  
##      data = DAT)
```

```
##
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -1.5460  -0.5434  -0.3682  -0.2166   3.2274
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -2.33931    0.19885 -11.765  <2e-16 ***  
## PC1_SD      -0.14872    0.16816  -0.884   0.3765  
## PC2_SD       0.05974    0.15799   0.378   0.7053  
## PC3_SD       0.37944    0.18161   2.089   0.0367 *  
## PC4_SD      -0.06789    0.16062  -0.423   0.6726  
## PC5_SD      -0.31276    0.16150  -1.937   0.0528 .  
## PC6_SD      -0.36238    0.16624  -2.180   0.0293 *  
## PC7_SD       0.31209    0.17164   1.818   0.0690 .  
## PC8_SD       0.09814    0.16563   0.593   0.5535  
## PC9_SD       0.02164    0.16421   0.132   0.8951  
## PC10_SD      0.08407    0.16914   0.497   0.6192  
## PC11_SD      0.28001    0.16681   1.679   0.0932 .  
## PC12_SD     -0.08424    0.16878  -0.499   0.6177  
## PC13_SD     -0.22907    0.16579  -1.382   0.1671  
## PC14_SD     -0.08211    0.16564  -0.496   0.6201  
## PC15_SD      0.17542    0.16406   1.069   0.2849  
## PC16_SD     -0.08960    0.15796  -0.567   0.5706  
## PC17_SD     -0.04821    0.16229  -0.297   0.7664  
## PC18_SD      0.29209    0.17394   1.679   0.0931 .  
## PC19_SD      0.17987    0.15566   1.156   0.2479  
## PC20_SD     -0.04769    0.17202  -0.277   0.7816  
## PC21_SD     -0.07767    0.15997  -0.486   0.6273  
## PC22_SD     -0.18464    0.15887  -1.162   0.2452  
## PC23_SD     -0.04358    0.15095  -0.289   0.7728  
## PC24_SD      0.10254    0.15678   0.654   0.5131  
## PC25_SD     -0.06833    0.15963  -0.428   0.6686  
## PC26_SD     -0.04335    0.16423  -0.264   0.7918  
## PC27_SD     -0.11872    0.16213  -0.732   0.4640  
## PC28_SD     -0.13992    0.16361  -0.855   0.3924  
## PC29_SD     -0.28691    0.15851  -1.810   0.0703 .  
## PC30_SD     -0.17156    0.15881  -1.080   0.2800  
## PC31_SD      0.07626    0.16050   0.475   0.6347  
## PC32_SD     -0.15844    0.15977  -0.992   0.3213  
## PC33_SD      0.13535    0.15422   0.878   0.3801  
## PC34_SD      0.21123    0.17583   1.201   0.2296
```

```
## PC35_SD      0.11168    0.17134    0.652    0.5145
## PC36_SD      0.09137    0.15100    0.605    0.5451
## PC37_SD      0.12802    0.16963    0.755    0.4504
## PC38_SD      0.23489    0.15565    1.509    0.1313
## PC39_SD      0.33491    0.15942    2.101    0.0357 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 318.68  on 435  degrees of freedom
## Residual deviance: 275.47  on 396  degrees of freedom
## AIC: 355.47
##
## Number of Fisher Scoring iterations: 6
```

```
# Count number of significant spectra
nsg = data.table(summary(MOD)$coeff[-1,"Pr(>|z|)"]) %>%
  subset(V1<0.05) %>% nrow()
print(paste0(nsg, " of 39 spectra significant (p<.05)"))
```

```
## [1] "3 of 39 spectra significant (p<.05)"
```

```
# Save model
mod.risk = list(d17p=MOD)

# Overall p-value
NLL = glm(data = DAT, formula = D_TRI_CF_ABNORMALITYPR11 ~ 1, family = "binomial")

mod.risk$d17p$p = pchisq(deviance(NLL)-deviance(MOD),
                        df.residual(NLL)-df.residual(MOD),
                        lower.tail=FALSE)

mod.risk$d17p$p
```

```
## [1] 0.2961634
```

Run model with significant spectra only

```
md2 = glm(data = DAT,
          formula = D_TRI_CF_ABNORMALITYPR11 ~ PC3_SD + PC6_SD + PC39_SD,
          family = "binomial")
summary(md2)
```

```
##
## Call:
## glm(formula = D_TRI_CF_ABNORMALITYPR11 ~ PC3_SD + PC6_SD + PC39_SD,
##      family = "binomial", data = DAT)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9347  -0.5392  -0.4560  -0.3828   2.5694
##
```

```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.0722     0.1575 -13.153  <2e-16 ***
## PC3_SD        0.2493     0.1510   1.651  0.0988 .
## PC6_SD       -0.2745     0.1506  -1.823  0.0683 .
## PC39_SD       0.3362     0.1491   2.254  0.0242 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 318.68  on 435  degrees of freedom
## Residual deviance: 308.03  on 432  degrees of freedom
## AIC: 316.03
##
## Number of Fisher Scoring iterations: 5
```

```
pchisq(deviance(NLL)-deviance(md2),
        df.residual(NLL)-df.residual(md2),lower.tail=FALSE)
```

```
## [1] 0.01378549
```

1.2. t(14;16): D_TRI_CF_ABNORMALITYPR8 & D_TRI_CF_T1416ABNORMAL (abundance)

```
DAT = spectra_clinical %>%
  dplyr::select("D_TRI_CF_ABNORMALITYPR8",starts_with("PC")) %>%
  dplyr::filter(D_TRI_CF_ABNORMALITYPR8!="Not Done" &
    D_TRI_CF_ABNORMALITYPR8!="")

print(paste("Baseline samples with data:",nrow(DAT),"(total)",
  nrow(DAT[DAT$D_TRI_CF_ABNORMALITYPR8=="Yes",]),"(+)",
  nrow(DAT[DAT$D_TRI_CF_ABNORMALITYPR8=="No",]),"(-)"))
```

```
## [1] "Baseline samples with data: 554 (total), 53 (+), 501 (-)"
```

```
DAT$D_TRI_CF_ABNORMALITYPR8 = as.factor(DAT$D_TRI_CF_ABNORMALITYPR8)

# Logistic regression
MOD <- glm(data = DAT,formula = D_TRI_CF_ABNORMALITYPR8 ~ .,family = "binomial")
summary(MOD)
```

```
##
## Call:
## glm(formula = D_TRI_CF_ABNORMALITYPR8 ~ ., family = "binomial",
##      data = DAT)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.90797  -0.35808  -0.18434  -0.08671   3.11720
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
```

```

## (Intercept) -3.68550    0.34193 -10.779 < 2e-16 ***
## PC1_SD      -0.30672    0.21028  -1.459 0.144664
## PC2_SD      -0.55403    0.20849  -2.657 0.007877 **
## PC3_SD       0.32627    0.20724   1.574 0.115408
## PC4_SD      -0.24505    0.18243  -1.343 0.179186
## PC5_SD      -0.46980    0.17898  -2.625 0.008669 **
## PC6_SD       0.36046    0.20485   1.760 0.078472 .
## PC7_SD      -0.01306    0.19061  -0.069 0.945358
## PC8_SD       0.04631    0.19583   0.236 0.813053
## PC9_SD      -0.24567    0.20440  -1.202 0.229401
## PC10_SD     0.50221    0.20863   2.407 0.016076 *
## PC11_SD     -0.08352    0.17871  -0.467 0.640253
## PC12_SD     0.60753    0.19491   3.117 0.001827 **
## PC13_SD     -0.41082    0.17773  -2.311 0.020806 *
## PC14_SD     -0.31970    0.18560  -1.723 0.084969 .
## PC15_SD     0.02333    0.19011   0.123 0.902330
## PC16_SD     0.80028    0.22187   3.607 0.000310 ***
## PC17_SD     0.04611    0.19884   0.232 0.816632
## PC18_SD     -0.22539    0.17719  -1.272 0.203369
## PC19_SD     -0.89008    0.22412  -3.971 7.14e-05 ***
## PC20_SD     0.25206    0.20866   1.208 0.227043
## PC21_SD     -0.22007    0.17950  -1.226 0.220210
## PC22_SD     0.02942    0.20551   0.143 0.886173
## PC23_SD     -0.10984    0.19196  -0.572 0.567199
## PC24_SD     -0.16723    0.18541  -0.902 0.367077
## PC25_SD     -0.72354    0.20843  -3.471 0.000518 ***
## PC26_SD     0.30567    0.18513   1.651 0.098721 .
## PC27_SD     -0.12525    0.19136  -0.655 0.512773
## PC28_SD     0.56144    0.17975   3.123 0.001787 **
## PC29_SD     0.08356    0.19796   0.422 0.672958
## PC30_SD     -0.17086    0.17983  -0.950 0.342045
## PC31_SD     -0.27408    0.18853  -1.454 0.146007
## PC32_SD     0.17124    0.17858   0.959 0.337619
## PC33_SD     0.19644    0.18058   1.088 0.276670
## PC34_SD     -0.12919    0.19111  -0.676 0.499049
## PC35_SD     -0.37854    0.19025  -1.990 0.046623 *
## PC36_SD     -0.07250    0.16971  -0.427 0.669217
## PC37_SD     0.40193    0.19317   2.081 0.037465 *
## PC38_SD     -0.01291    0.17264  -0.075 0.940380
## PC39_SD     -0.40072    0.19526  -2.052 0.040147 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 349.53  on 553  degrees of freedom
## Residual deviance: 227.27  on 514  degrees of freedom
## AIC: 307.27
##
## Number of Fisher Scoring iterations: 7

```

```

# Count sig spectra in model
nsig = data.table(summary(MOD)$coeff[-1,"Pr(>|z|)"]) %>%
  subset(V1<0.05) %>% nrow()

```

```
print(paste0(nsig," of 39 spectra significant (p<.05)"))
```

```
## [1] "12 of 39 spectra significant (p<.05)"
```

```
# Save model
```

```
mod.risk = c(mod.risk,list(t1416=MOD))
```

```
# Overall p-value
```

```
NLL = glm(data = DAT,formula = D_TRI_CF_ABNORMALITYPR8 ~ 1,family = "binomial")
```

```
mod.risk$t1416$p = pchisq(deviance(NLL)-deviance(MOD),
                           df.residual(NLL)-df.residual(MOD),
                           lower.tail=FALSE)
```

```
mod.risk$t1416$p
```

```
## [1] 1.608411e-10
```

1.3. amp(1q): D_TRI_CF_ABNORMALITYPR13 & D_TRI_CF_1PAMPLIFICATI2 (abundance)

```
DAT = spectra_clinical %>%
```

```
  dplyr::select("D_TRI_CF_ABNORMALITYPR13",starts_with("PC")) %>%
```

```
  dplyr::filter(D_TRI_CF_ABNORMALITYPR13!="Not Done" &
                D_TRI_CF_ABNORMALITYPR13!="")
```

```
print(paste("Baseline samples with data:",nrow(DAT),"(total)",
            nrow(DAT[DAT$D_TRI_CF_ABNORMALITYPR13=="Yes",]),"(+)",
            nrow(DAT[DAT$D_TRI_CF_ABNORMALITYPR13=="No",]),"(-)"))
```

```
## [1] "Baseline samples with data: 525 (total), 202 (+), 323 (-)"
```

```
DAT$D_TRI_CF_ABNORMALITYPR13 = as.factor(DAT$D_TRI_CF_ABNORMALITYPR13)
```

```
# Logistic regression
```

```
MOD <- glm(data = DAT,formula = D_TRI_CF_ABNORMALITYPR13 ~ .,family = "binomial")
```

```
summary(MOD)
```

```
##
```

```
## Call:
```

```
## glm(formula = D_TRI_CF_ABNORMALITYPR13 ~ ., family = "binomial",
##      data = DAT)
```

```
##
```

```
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.4620  -0.6314  -0.3313   0.6043   3.0382
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.690992   0.126291  -5.471 4.46e-08 ***
## PC1_SD      -0.007713   0.122192  -0.063 0.949672
## PC2_SD      -0.610630   0.126239  -4.837 1.32e-06 ***
## PC3_SD       1.024640   0.138244   7.412 1.25e-13 ***
```

```

## PC4_SD      0.109691  0.125437  0.874 0.381861
## PC5_SD     -0.802967  0.135274 -5.936 2.92e-09 ***
## PC6_SD      0.482744  0.127773  3.778 0.000158 ***
## PC7_SD      0.117559  0.119897  0.980 0.326841
## PC8_SD     -0.230686  0.122601 -1.882 0.059890 .
## PC9_SD     -0.244367  0.123028 -1.986 0.047004 *
## PC10_SD    -0.408167  0.125689 -3.247 0.001164 **
## PC11_SD    -0.284010  0.123191 -2.305 0.021141 *
## PC12_SD     0.148954  0.121227  1.229 0.219178
## PC13_SD     0.144725  0.116883  1.238 0.215641
## PC14_SD    -0.446631  0.124381 -3.591 0.000330 ***
## PC15_SD     0.178899  0.118319  1.512 0.130534
## PC16_SD     0.127909  0.120903  1.058 0.290083
## PC17_SD    -0.076199  0.118711 -0.642 0.520947
## PC18_SD    -0.193285  0.123084 -1.570 0.116334
## PC19_SD    -0.202939  0.125611 -1.616 0.106177
## PC20_SD     0.048262  0.125938  0.383 0.701557
## PC21_SD    -0.287935  0.117308 -2.455 0.014107 *
## PC22_SD    -0.163345  0.124526 -1.312 0.189610
## PC23_SD    -0.308160  0.119422 -2.580 0.009868 **
## PC24_SD     0.393140  0.128223  3.066 0.002169 **
## PC25_SD     0.288189  0.127105  2.267 0.023370 *
## PC26_SD    -0.105200  0.123295 -0.853 0.393528
## PC27_SD    -0.453369  0.124358 -3.646 0.000267 ***
## PC28_SD     0.015377  0.116773  0.132 0.895233
## PC29_SD    -0.065995  0.128013 -0.516 0.606179
## PC30_SD    -0.065103  0.113445 -0.574 0.566053
## PC31_SD    -0.010491  0.115031 -0.091 0.927330
## PC32_SD    -0.183180  0.128897 -1.421 0.155280
## PC33_SD    -0.082332  0.124583 -0.661 0.508701
## PC34_SD     0.221452  0.124709  1.776 0.075773 .
## PC35_SD     0.291681  0.122200  2.387 0.016991 *
## PC36_SD     0.065219  0.120091  0.543 0.587076
## PC37_SD    -0.252369  0.127726 -1.976 0.048170 *
## PC38_SD     0.140748  0.117811  1.195 0.232208
## PC39_SD     0.038282  0.125684  0.305 0.760681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 699.66  on 524  degrees of freedom
## Residual deviance: 443.73  on 485  degrees of freedom
## AIC: 523.73
##
## Number of Fisher Scoring iterations: 5

```

```

# Count sig spectra in model
nsig = data.table(summary(MOD)$coeff[-1,"Pr(>|z|)"]) %>%
  subset(V1<0.05) %>% nrow()
print(paste0(nsig, " of 39 spectra significant (p<.05)"))

```

```
## [1] "15 of 39 spectra significant (p<.05)"
```

```
# Save model
mod.risk = c(mod.risk,list(a1q=MOD))

# Overall p-value
NLL = glm(data = DAT,formula = D_TRI_CF_ABNORMALITYPR13 ~ 1,family = "binomial")

mod.risk$a1q$p = pchisq(deviance(NLL)-deviance(MOD),
                        df.residual(NLL)-df.residual(MOD),
                        lower.tail=FALSE)

mod.risk$a1q$p
```

```
## [1] 1.071633e-33
```

1.4. t(4;14): D_TRI_CF_ABNORMALITYPR3 & D_TRI_CF_T414ABNORMALC (abundance)

```
DAT = spectra_clinical %>%
  dplyr::select("D_TRI_CF_ABNORMALITYPR3",starts_with("PC")) %>%
  dplyr::filter(D_TRI_CF_ABNORMALITYPR3!="Not Done" &
                D_TRI_CF_ABNORMALITYPR3!="")

print(paste("Baseline samples with data:",nrow(DAT),"(total)",",",
            nrow(DAT[DAT$D_TRI_CF_ABNORMALITYPR3=="Yes",]),"(+)",",",
            nrow(DAT[DAT$D_TRI_CF_ABNORMALITYPR3=="No",]),"(-)"))
```

```
## [1] "Baseline samples with data: 594 (total), 108 (+), 486 (-)"
```

```
DAT$D_TRI_CF_ABNORMALITYPR3 = as.factor(DAT$D_TRI_CF_ABNORMALITYPR3)

# Logistic regression
MOD <- glm(data = DAT,formula = D_TRI_CF_ABNORMALITYPR3 ~ .,family = "binomial")
summary(MOD)
```

```
##
## Call:
## glm(formula = D_TRI_CF_ABNORMALITYPR3 ~ ., family = "binomial",
##      data = DAT)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4577  -0.4685  -0.3038  -0.1423   2.9473
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.286504   0.184652 -12.383  < 2e-16 ***
## PC1_SD       0.288035   0.145436   1.980  0.047648 *
## PC2_SD      -0.570193   0.153941  -3.704  0.000212 ***
## PC3_SD       0.246057   0.144001   1.709  0.087504 .
## PC4_SD       0.055605   0.149027   0.373  0.709060
## PC5_SD       0.148729   0.137047   1.085  0.277816
## PC6_SD       0.235319   0.138274   1.702  0.088787 .
## PC7_SD       0.234987   0.146427   1.605  0.108536
## PC8_SD       0.532455   0.156388   3.405  0.000662 ***
```



```
## PC9_SD      -0.789087    0.160637   -4.912    9e-07 ***
## PC10_SD     -0.590438    0.158077   -3.735    0.000188 ***
## PC11_SD      0.066273    0.142892    0.464    0.642794
## PC12_SD      0.008056    0.142266    0.057    0.954846
## PC13_SD     -0.410435    0.142481   -2.881    0.003969 **
## PC14_SD      0.040884    0.146781    0.279    0.780602
## PC15_SD      0.286835    0.141114    2.033    0.042088 *
## PC16_SD     -0.067236    0.145373   -0.463    0.643717
## PC17_SD     -0.110613    0.149501   -0.740    0.459371
## PC18_SD      0.111623    0.156538    0.713    0.475802
## PC19_SD     -0.261750    0.152306   -1.719    0.085690 .
## PC20_SD      0.292234    0.142093    2.057    0.039721 *
## PC21_SD     -0.220324    0.153225   -1.438    0.150459
## PC22_SD     -0.025535    0.152651   -0.167    0.867150
## PC23_SD      0.202122    0.144174    1.402    0.160935
## PC24_SD     -0.010786    0.143987   -0.075    0.940289
## PC25_SD     -0.085848    0.151531   -0.567    0.571027
## PC26_SD      0.014047    0.141659    0.099    0.921013
## PC27_SD      0.321734    0.146496    2.196    0.028078 *
## PC28_SD      0.056878    0.145160    0.392    0.695181
## PC29_SD      0.126763    0.143918    0.881    0.378427
## PC30_SD     -0.237363    0.144678   -1.641    0.100875
## PC31_SD      0.244815    0.144839    1.690    0.090979 .
## PC32_SD      0.224236    0.151767    1.478    0.139541
## PC33_SD      0.224455    0.143789    1.561    0.118524
## PC34_SD      0.040400    0.146833    0.275    0.783205
## PC35_SD     -0.228223    0.146554   -1.557    0.119408
## PC36_SD     -0.195065    0.143619   -1.358    0.174398
## PC37_SD      0.210835    0.146764    1.437    0.150844
## PC38_SD     -0.383261    0.139350   -2.750    0.005953 **
## PC39_SD     -0.289167    0.154489   -1.872    0.061239 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 563.28  on 593  degrees of freedom
## Residual deviance: 355.97  on 554  degrees of freedom
## AIC: 435.97
##
## Number of Fisher Scoring iterations: 6
```

```
# Count sig spectra in model
nsig = data.table(summary(MOD)$coeff[-1,"Pr(>|z|)"]) %>%
  subset(V1<0.05) %>% nrow()
print(paste0(nsig, " of 39 spectra significant (p<.05)"))
```

```
## [1] "10 of 39 spectra significant (p<.05)"
```

```
# Save model
mod.risk = c(mod.risk,list(t414=MOD))

# Overall p-value
```

```

NLL = glm(data = DAT, formula = D_TRI_CF_ABNORMALITYPR3 ~ 1, family = "binomial")

mod.risk$t414$p = pchisq(deviance(NLL)-deviance(MOD),
                        df.residual(NLL)-df.residual(MOD),
                        lower.tail=FALSE)

mod.risk$t414$p

```

```
## [1] 8.207055e-25
```

1.5. t(11;14): D_TRI_CF_ABNORMALITYPR6 & D_TRI_CF_T1114ABNORMAL (abundance)

```

DAT = spectra_clinical %>%
  dplyr::select("D_TRI_CF_ABNORMALITYPR6", starts_with("PC")) %>%
  dplyr::filter(D_TRI_CF_ABNORMALITYPR6!="Not Done" &
               D_TRI_CF_ABNORMALITYPR6!="")

print(paste("Baseline samples with data:", nrow(DAT), "(total)",
            nrow(DAT[DAT$D_TRI_CF_ABNORMALITYPR6=="Yes",]), "(+)",
            nrow(DAT[DAT$D_TRI_CF_ABNORMALITYPR6=="No",]), "(-)"))

```

```
## [1] "Baseline samples with data: 594 (total), 153 (+), 441 (-)"
```

```

DAT$D_TRI_CF_ABNORMALITYPR6 = as.factor(DAT$D_TRI_CF_ABNORMALITYPR6)

# Logistic regression
MOD <- glm(data = DAT, formula = D_TRI_CF_ABNORMALITYPR6 ~ ., family = "binomial")
summary(MOD)

```

```

##
## Call:
## glm(formula = D_TRI_CF_ABNORMALITYPR6 ~ ., family = "binomial",
##      data = DAT)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6025  -0.5794  -0.3609   0.4043   2.9983
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.629511   0.144002 -11.316 < 2e-16 ***
## PC1_SD       0.757514   0.123825   6.118 9.5e-10 ***
## PC2_SD      -0.334711   0.121617  -2.752 0.00592 **
## PC3_SD      -1.009172   0.136773  -7.378 1.6e-13 ***
## PC4_SD       0.173529   0.129039   1.345 0.17869
## PC5_SD      -0.124781   0.126141  -0.989 0.32256
## PC6_SD      -0.335438   0.136356  -2.460 0.01389 *
## PC7_SD      -0.384628   0.126124  -3.050 0.00229 **
## PC8_SD       0.003280   0.125356   0.026 0.97912
## PC9_SD       0.219560   0.139354   1.576 0.11513
## PC10_SD      0.293407   0.132796   2.209 0.02714 *
## PC11_SD      0.243246   0.134069   1.814 0.06963 .

```

```
## PC12_SD      -0.075078    0.122723   -0.612    0.54069
## PC13_SD      -0.131966    0.127691   -1.033    0.30138
## PC14_SD      -0.152805    0.128630   -1.188    0.23486
## PC15_SD      -0.009517    0.130661   -0.073    0.94193
## PC16_SD       0.211483    0.130026    1.626    0.10385
## PC17_SD      -0.050028    0.129911   -0.385    0.70017
## PC18_SD       0.360956    0.125101    2.885    0.00391 **
## PC19_SD      -0.104462    0.125673   -0.831    0.40585
## PC20_SD      -0.108394    0.129358   -0.838    0.40206
## PC21_SD      -0.267156    0.128115   -2.085    0.03704 *
## PC22_SD       0.067931    0.122739    0.553    0.57995
## PC23_SD      -0.149114    0.127474   -1.170    0.24210
## PC24_SD       0.126206    0.126297    0.999    0.31766
## PC25_SD      -0.124085    0.121039   -1.025    0.30529
## PC26_SD       0.030517    0.126957    0.240    0.81004
## PC27_SD      -0.142821    0.121754   -1.173    0.24079
## PC28_SD       0.276628    0.127636    2.167    0.03021 *
## PC29_SD       0.142448    0.128801    1.106    0.26875
## PC30_SD       0.074267    0.120390    0.617    0.53731
## PC31_SD      -0.149381    0.121115   -1.233    0.21743
## PC32_SD      -0.179306    0.123783   -1.449    0.14747
## PC33_SD      -0.058201    0.125441   -0.464    0.64267
## PC34_SD      -0.079668    0.124263   -0.641    0.52144
## PC35_SD       0.116446    0.123757    0.941    0.34674
## PC36_SD       0.277852    0.125477    2.214    0.02680 *
## PC37_SD      -0.139475    0.135170   -1.032    0.30214
## PC38_SD       0.119632    0.124680    0.960    0.33730
## PC39_SD       0.073817    0.119798    0.616    0.53777
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 677.76  on 593  degrees of freedom
## Residual deviance: 453.15  on 554  degrees of freedom
## AIC: 533.15
##
## Number of Fisher Scoring iterations: 5
```

```
# Count sig spectra in model
nsig = data.table(summary(MOD)$coeff[-1,"Pr(>|z|)"]) %>%
  subset(V1<0.05) %>% nrow()
print(paste0(nsig," of 39 spectra significant (p<.05)"))
```

```
## [1] "10 of 39 spectra significant (p<.05)"
```

```
# Save model
mod.risk = c(mod.risk,list(t1114=MOD))

# Overall p-value
NLL = glm(data = DAT,formula = D_TRI_CF_ABNORMALITYPR6 ~ 1,family = "binomial")

mod.risk$t1114$p = pchisq(deviance(NLL)-deviance(MOD),
```

```

df.residual(NLL)-df.residual(MOD),
lower.tail=FALSE)
mod.risk$t1114$p

```

```
## [1] 6.201022e-28
```

2. International staging system (ISS)

2.1. RUN LOGISTIC REGRESSION

```

DAT = spectra_clinical %>%
  dplyr::select("D_PT_iss",starts_with("PC"))

# Order dependent variable
DAT$ISS = factor(DAT$D_PT_iss, levels = c(1,2,3), ordered = TRUE)

# Ordinal logistic regression model
MOD = polr(data = DAT[, -c("D_PT_iss")], formula = ISS ~ ., Hess = T)
summary(MOD)

```

```

## Call:
## polr(formula = ISS ~ ., data = DAT[, -c("D_PT_iss")], Hess = T)
##
## Coefficients:
##              Value Std. Error   t value
## PC1_SD   -0.2749662   0.07106  -3.869573
## PC2_SD   -0.1333559   0.07127  -1.871037
## PC3_SD    0.3549374   0.07179   4.944326
## PC4_SD    0.3126878   0.07203   4.341113
## PC5_SD   -0.1743363   0.07120  -2.448579
## PC6_SD   -0.1477172   0.07224  -2.044932
## PC7_SD    0.0166838   0.07019   0.237700
## PC8_SD    0.1602979   0.07054   2.272335
## PC9_SD   -0.1729565   0.07093  -2.438448
## PC10_SD  -0.0107528   0.07059  -0.152337
## PC11_SD   0.0517981   0.07172   0.722239
## PC12_SD  -0.1726008   0.07115  -2.425862
## PC13_SD  -0.1743827   0.06957  -2.506712
## PC14_SD  -0.0209329   0.07013  -0.298501
## PC15_SD  -0.1563124   0.07029  -2.223775
## PC16_SD   0.0820734   0.07022   1.168744
## PC17_SD   0.2543831   0.07047   3.609659
## PC18_SD  -0.0401684   0.07203  -0.557635
## PC19_SD   0.2633690   0.07220   3.647976
## PC20_SD  -0.0066336   0.07045  -0.094160
## PC21_SD   0.0927783   0.07108   1.305260
## PC22_SD  -0.0071847   0.07087  -0.101386
## PC23_SD  -0.0005706   0.07006  -0.008144
## PC24_SD   0.0082876   0.07171   0.115568
## PC25_SD   0.1693466   0.07045   2.403712
## PC26_SD   0.2382966   0.07029   3.390307
## PC27_SD  -0.1344839   0.07060  -1.904751

```

```
## PC28_SD 0.1398285 0.07131 1.960725
## PC29_SD 0.1387677 0.07090 1.957315
## PC30_SD 0.0289022 0.07145 0.404529
## PC31_SD -0.0464069 0.07063 -0.657088
## PC32_SD -0.0366594 0.07173 -0.511082
## PC33_SD 0.0126673 0.07093 0.178598
## PC34_SD 0.0243803 0.07073 0.344691
## PC35_SD 0.1347160 0.07013 1.921014
## PC36_SD -0.0364027 0.07093 -0.513211
## PC37_SD -0.0230455 0.07037 -0.327481
## PC38_SD 0.1395633 0.07135 1.956152
## PC39_SD 0.0207842 0.07123 0.291788
##
## Intercepts:
##      Value Std. Error t value
## 1|2 -0.7310 0.0834 -8.7688
## 2|3 1.0550 0.0886 11.9131
##
## Residual Deviance: 1479.279
## AIC: 1561.279
## (21 observations deleted due to missingness)
```

```
# Count sig spectra in model
sig = pnorm(abs(coef(summary(MOD))[1:39,"t value"]),lower.tail = FALSE)*2
nsig = data.table(sig) %>% subset(sig<0.05) %>% nrow()
print(paste0(nsig, " of 39 spectra significant (p<.05)"))
```

```
## [1] "15 of 39 spectra significant (p<.05)"
```

```
# Save model
mod.risk = c(mod.risk,list(iss=MOD))

# Overall p-value
NLL = polr(data = DAT[, -c("D_PT_iss")], formula = ISS ~ 1, Hess = T)

mod.risk$iss$p = pchisq(deviance(NLL)-deviance(MOD),
                        df.residual(NLL)-df.residual(MOD),
                        lower.tail=FALSE)

mod.risk$iss$p
```

```
## [1] 2.999079e-15
```

Save model results

```
save(mod.risk,file = "rdata/mod.clinical-risk.rdata")
```