

CD138+ spectra and demographic risk groups

Code to investigate associations between patient demographics (age, gender, self-reported race and ethnicity) and transcriptome spectra using analysis of variance.

0. Setup

Define data directory

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

Load packages

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

Load transcriptome spectra (PC1-PC39)

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

1. Age

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

# Linear regression
lm.age = lm(data = DAT, formula = D_PT_age ~ .)
summary(lm.age)
```

```
##
## Call:
## lm(formula = D_PT_age ~ ., data = DAT)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.8921  -6.3635   0.0883   6.3609  30.6409
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 62.77184    0.36209 173.361 < 2e-16 ***
## PC1_SD      -1.48430    0.36232  -4.097 4.66e-05 ***
## PC2_SD       0.12813    0.36232   0.354 0.723705
## PC3_SD       0.66423    0.36232   1.833 0.067175 .
## PC4_SD      -0.92603    0.36232  -2.556 0.010796 *
## PC5_SD      -0.29912    0.36232  -0.826 0.409324
## PC6_SD       0.32925    0.36232   0.909 0.363805
## PC7_SD       0.94218    0.36232   2.600 0.009501 **
## PC8_SD       0.17277    0.36232   0.477 0.633620
## PC9_SD       0.01177    0.36232   0.032 0.974091
## PC10_SD     -1.66962    0.36232  -4.608 4.80e-06 ***
## PC11_SD     -0.73956    0.36232  -2.041 0.041595 *
## PC12_SD     -0.90340    0.36232  -2.493 0.012875 *
## PC13_SD      0.92601    0.36232   2.556 0.010798 *
## PC14_SD     -0.60176    0.36232  -1.661 0.097179 .
## PC15_SD     -0.90335    0.36232  -2.493 0.012881 *
## PC16_SD      0.05095    0.36232   0.141 0.888219
## PC17_SD      0.57335    0.36232   1.582 0.113986
## PC18_SD     -0.46581    0.36232  -1.286 0.198982
## PC19_SD      1.00681    0.36232   2.779 0.005597 **
## PC20_SD     -1.54263    0.36232  -4.258 2.34e-05 ***
## PC21_SD      0.79250    0.36232   2.187 0.029040 *
## PC22_SD     -0.67360    0.36232  -1.859 0.063416 .
## PC23_SD      0.01675    0.36232   0.046 0.963142
## PC24_SD      1.30580    0.36232   3.604 0.000335 ***
## PC25_SD     -0.29181    0.36232  -0.805 0.420856
## PC26_SD     -0.80834    0.36232  -2.231 0.025986 *
## PC27_SD     -0.39035    0.36232  -1.077 0.281683
## PC28_SD      0.21170    0.36232   0.584 0.559218
## PC29_SD     -0.19330    0.36232  -0.533 0.593851
## PC30_SD      0.29737    0.36232   0.821 0.412072
## PC31_SD      0.89996    0.36232   2.484 0.013220 *
## PC32_SD     -0.72314    0.36232  -1.996 0.046325 *
## PC33_SD     -0.30753    0.36232  -0.849 0.396294
## PC34_SD     -0.29324    0.36232  -0.809 0.418584
## PC35_SD      0.09050    0.36232   0.250 0.802833
## PC36_SD      0.55632    0.36232   1.535 0.125114
## PC37_SD      0.02817    0.36232   0.078 0.938041
## PC38_SD      0.01374    0.36232   0.038 0.969750
## PC39_SD     -0.59990    0.36232  -1.656 0.098212 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.03 on 727 degrees of freedom
## Multiple R-squared:  0.179, Adjusted R-squared:  0.1349
## F-statistic: 4.063 on 39 and 727 DF, p-value: 1.989e-14
```

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

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

```
# Overall p-value
df1 = summary(lm.age)$fstatistic[2]
df2 = summary(lm.age)$fstatistic[3]
f = summary(lm.age)$fstatistic[1]
lm.age$p = pf(f,df1,df2,lower.tail = F) # Compute p-value from f-statistic
```

2. Gender

```
DAT = spectra_clinical %>%
  dplyr::select("D_PT_gender",starts_with("PC"))
DAT$D_PT_gender = as.factor(DAT$D_PT_gender)

# Logistic regression
glm.gender = glm(data = DAT,formula = D_PT_gender ~ .,family = "binomial")
summary(glm.gender)
```

```
##
## Call:
## glm(formula = D_PT_gender ~ ., family = "binomial", data = DAT)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9799  -0.9661  -0.6382   1.1340   2.3527
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.422256   0.080079  -5.273 1.34e-07 ***
## PC1_SD       0.036010   0.078288   0.460 0.645537
## PC2_SD      -0.072280   0.080530  -0.898 0.369424
## PC3_SD       0.274289   0.079512   3.450 0.000561 ***
## PC4_SD       0.110024   0.079390   1.386 0.165789
## PC5_SD       0.059432   0.080218   0.741 0.458766
## PC6_SD      -0.065574   0.080012  -0.820 0.412475
## PC7_SD      -0.132248   0.079955  -1.654 0.098123 .
## PC8_SD      -0.067240   0.078832  -0.853 0.393683
## PC9_SD       0.310437   0.081385   3.814 0.000136 ***
## PC10_SD      0.012423   0.079575   0.156 0.875945
## PC11_SD     -0.274311   0.080467  -3.409 0.000652 ***
## PC12_SD      0.283500   0.081133   3.494 0.000475 ***
## PC13_SD      0.145449   0.079834   1.822 0.068471 .
## PC14_SD     -0.163407   0.081204  -2.012 0.044188 *
## PC15_SD      0.075851   0.080070   0.947 0.343480
## PC16_SD      0.117868   0.080389   1.466 0.142588
## PC17_SD      0.092468   0.079835   1.158 0.246765
## PC18_SD     -0.074449   0.080354  -0.927 0.354183
## PC19_SD     -0.070864   0.080201  -0.884 0.376920
## PC20_SD      0.096550   0.079726   1.211 0.225886
## PC21_SD     -0.109125   0.079472  -1.373 0.169715
## PC22_SD     -0.109524   0.080680  -1.358 0.174619
```

```
## PC23_SD      -0.151552    0.080071   -1.893  0.058396 .
## PC24_SD       0.067214    0.079681    0.844  0.398922
## PC25_SD      -0.020035    0.080621   -0.249  0.803742
## PC26_SD       0.146534    0.081139    1.806  0.070924 .
## PC27_SD       0.065396    0.079295    0.825  0.409534
## PC28_SD      -0.029238    0.079575   -0.367  0.713304
## PC29_SD      -0.065196    0.081220   -0.803  0.422141
## PC30_SD       0.095563    0.080078    1.193  0.232725
## PC31_SD      -0.126141    0.079100   -1.595  0.110782
## PC32_SD      -0.082797    0.080191   -1.032  0.301841
## PC33_SD      -0.004691    0.078633   -0.060  0.952432
## PC34_SD      -0.216274    0.080101   -2.700  0.006934 **
## PC35_SD      -0.096846    0.079786   -1.214  0.224820
## PC36_SD      -0.137452    0.079761   -1.723  0.084832 .
## PC37_SD      -0.126494    0.079947   -1.582  0.113598
## PC38_SD       0.063249    0.080043    0.790  0.429419
## PC39_SD      -0.291919    0.081480   -3.583  0.000340 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1037.96  on 766  degrees of freedom
## Residual deviance:  925.15  on 727  degrees of freedom
## AIC: 1005.1
##
## Number of Fisher Scoring iterations: 4
```

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

```
## [1] "7 of 39 spectra significant (p<.05)"
```

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

glm.gender$p = pchisq(deviance(NLL)-deviance(glm.gender),
                      df.residual(NLL)-df.residual(glm.gender),
                      lower.tail=FALSE)
```

3. Self-reported race

```
DAT = spectra_clinical %>%
  dplyr::select("D_PT_race", starts_with("PC")) %>%
  dplyr::filter(D_PT_race%in%c(1,2))
DAT$D_PT_race = as.factor(DAT$D_PT_race)

# Logistic regression
glm.race = glm(data = DAT, formula = D_PT_race ~ ., family = "binomial")
summary(glm.race)
```

```

##
## Call:
## glm(formula = D_PT_race ~ ., family = "binomial", data = DAT)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6635  -0.5945  -0.3457  -0.1467   2.8155
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.12667    0.16722 -12.718  < 2e-16 ***
## PC1_SD       0.16123    0.13454   1.198  0.230761
## PC2_SD       0.10287    0.12361   0.832  0.405290
## PC3_SD      -0.01948    0.12518  -0.156  0.876328
## PC4_SD       0.14230    0.13166   1.081  0.279773
## PC5_SD       0.33210    0.13449   2.469  0.013540 *
## PC6_SD      -0.17244    0.13253  -1.301  0.193216
## PC7_SD       0.22071    0.13754   1.605  0.108564
## PC8_SD      -0.05869    0.12389  -0.474  0.635678
## PC9_SD       0.29580    0.13684   2.162  0.030642 *
## PC10_SD      0.29809    0.12720   2.343  0.019106 *
## PC11_SD     -0.46550    0.13441  -3.463  0.000534 ***
## PC12_SD     -0.15707    0.12853  -1.222  0.221700
## PC13_SD     -0.27996    0.12989  -2.155  0.031126 *
## PC14_SD     -0.40673    0.13556  -3.000  0.002696 **
## PC15_SD      0.14717    0.13412   1.097  0.272509
## PC16_SD     -0.10943    0.12745  -0.859  0.390580
## PC17_SD     -0.37349    0.13755  -2.715  0.006620 **
## PC18_SD      0.17397    0.13280   1.310  0.190190
## PC19_SD     -0.21388    0.12427  -1.721  0.085229 .
## PC20_SD     -0.06517    0.12775  -0.510  0.609962
## PC21_SD     -0.08175    0.13243  -0.617  0.537001
## PC22_SD      0.31139    0.12766   2.439  0.014722 *
## PC23_SD      0.13727    0.12582   1.091  0.275278
## PC24_SD     -0.31488    0.12819  -2.456  0.014036 *
## PC25_SD      0.39171    0.13828   2.833  0.004614 **
## PC26_SD     -0.01679    0.12920  -0.130  0.896592
## PC27_SD      0.18929    0.13316   1.421  0.155177
## PC28_SD      0.15773    0.13610   1.159  0.246470
## PC29_SD      0.17183    0.13233   1.298  0.194120
## PC30_SD     -0.08536    0.13085  -0.652  0.514184
## PC31_SD      0.01500    0.12869   0.117  0.907238
## PC32_SD     -0.23318    0.12678  -1.839  0.065880 .
## PC33_SD     -0.04656    0.12554  -0.371  0.710713
## PC34_SD     -0.38807    0.12949  -2.997  0.002726 **
## PC35_SD      0.17470    0.13462   1.298  0.194369
## PC36_SD      0.13366    0.13269   1.007  0.313758
## PC37_SD     -0.13100    0.13184  -0.994  0.320411
## PC38_SD     -0.73933    0.14180  -5.214  1.85e-07 ***
## PC39_SD      0.36080    0.13782   2.618  0.008846 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)

```

```
##
## Null deviance: 541.04 on 600 degrees of freedom
## Residual deviance: 423.77 on 561 degrees of freedom
## AIC: 503.77
##
## Number of Fisher Scoring iterations: 6

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

## [1] "13 of 39 spectra significant (p<.05)"

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

glm.race$p = pchisq(deviance(NLL)-deviance(glm.race),
                    df.residual(NLL)-df.residual(glm.race),
                    lower.tail=FALSE)
```

4. Self-reported ethnicity

```
DAT = spectra_clinical %>%
  dplyr::select("D_PT_ethnic",starts_with("PC")) %>%
  dplyr::filter(D_PT_ethnic%in%c(1,2))
DAT$D_PT_ethnic = as.factor(DAT$D_PT_ethnic)

# Logistic regression
glm.ethnic = glm(data = DAT,formula = D_PT_ethnic ~ .,family = "binomial")
summary(glm.ethnic)

##
## Call:
## glm(formula = D_PT_ethnic ~ ., family = "binomial", data = DAT)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.9469   0.1690   0.2913   0.4507   1.3267
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.100130   0.245687  12.618 < 2e-16 ***
## PC1_SD       0.407986   0.205337   1.987  0.04693 *
## PC2_SD       0.276568   0.173751   1.592  0.11144
## PC3_SD       0.179213   0.192690   0.930  0.35234
## PC4_SD       0.405026   0.170305   2.378  0.01740 *
## PC5_SD      -0.081381   0.158782  -0.513  0.60828
## PC6_SD      -0.076875   0.152472  -0.504  0.61413
## PC7_SD       0.310664   0.165116   1.881  0.05990 .
## PC8_SD       0.119571   0.167668   0.713  0.47576
```

```
## PC9_SD      -0.117273    0.162672   -0.721    0.47096
## PC10_SD     0.432800    0.172560    2.508    0.01214 *
## PC11_SD     0.216388    0.153421    1.410    0.15842
## PC12_SD    -0.325697    0.169518   -1.921    0.05469 .
## PC13_SD    -0.007241    0.161120   -0.045    0.96416
## PC14_SD    -0.498837    0.170121   -2.932    0.00337 **
## PC15_SD     0.327955    0.161982    2.025    0.04290 *
## PC16_SD     0.494932    0.164275    3.013    0.00259 **
## PC17_SD    -0.466888    0.160176   -2.915    0.00356 **
## PC18_SD    -0.092183    0.168737   -0.546    0.58485
## PC19_SD    -0.277022    0.149252   -1.856    0.06344 .
## PC20_SD     0.278688    0.151716    1.837    0.06622 .
## PC21_SD    -0.128437    0.160520   -0.800    0.42364
## PC22_SD     0.031529    0.158660    0.199    0.84248
## PC23_SD    -0.238077    0.165442   -1.439    0.15014
## PC24_SD     0.078455    0.160903    0.488    0.62584
## PC25_SD     0.045299    0.155979    0.290    0.77150
## PC26_SD    -0.167313    0.159118   -1.052    0.29303
## PC27_SD     0.146909    0.166184    0.884    0.37669
## PC28_SD    -0.305626    0.156574   -1.952    0.05094 .
## PC29_SD    -0.006187    0.150410   -0.041    0.96719
## PC30_SD    -0.021131    0.152101   -0.139    0.88951
## PC31_SD    -0.385929    0.152592   -2.529    0.01143 *
## PC32_SD     0.235482    0.160711    1.465    0.14285
## PC33_SD     0.026765    0.160168    0.167    0.86729
## PC34_SD    -0.066407    0.147978   -0.449    0.65360
## PC35_SD    -0.076210    0.156301   -0.488    0.62585
## PC36_SD     0.202460    0.153940    1.315    0.18845
## PC37_SD     0.152140    0.169381    0.898    0.36907
## PC38_SD     0.341932    0.159397    2.145    0.03194 *
## PC39_SD    -0.278720    0.151667   -1.838    0.06611 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 390.75  on 624  degrees of freedom
## Residual deviance: 319.37  on 585  degrees of freedom
## AIC: 399.37
##
## Number of Fisher Scoring iterations: 6
```

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

```
## [1] "9 of 39 spectra significant (p<.05)"
```

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

glm.ethnic$p = pchisq(deviance(NLL)-deviance(glm.ethnic),
```

```
df.residual(NLL)-df.residual(glm.ethnic),  
lower.tail=FALSE)
```

Save model results

```
save(lm.age,glm.gender,glm.race,glm.ethnic,file = "rdata/mod.demographic-risk.rdata")
```