Analysis: Expression levels of PD-L1 (CD274) in lower-grade gliomas (LGG - TCGA publicly available data)

Sebastian Kurscheid

23 January 2016

Contents

1 Introduction **2**

2 Results **2**

2.1 Data preparation . . . 2

2.2 Comparing PD-L1 expression between Grade II and III samples: . . . 3

2.2.1 Boxplot . . . 3

2.2.2 Conclusions . . . 3

2.3 Comparing LGG samples stratified by IDH1/2 mutation status . . . 4

2.3.1 Data preparation . . . 4

2.3.2 Boxplot . . . 4

2.3.3 Statistical Test (t-Test) . . . 5

2.3.4 Conclusions . . . 5

2.4 Comparing LGG CIMP positive to negative samples . . . 6

2.4.1 Boxplot . . . 6

2.4.2 Statistical Test (t-Test) . . . 6

2.4.3 Conclusions . . . 6

2.5 Comparing LGG samples stratified by IDH and 1p19q status to GBM . . . 7

2.5.1 Boxplot . . . 7

2.5.2 Statistical Test (ANOVA) . . . 8

2.5.3 Conclusions . . . 8

2.6 Comparing LGG samples stratified by CIMP and 1p19q status to GBM . . . 8

2.6.1 Boxplot . . . 8

2.6.2 Statistical Test (ANOVA) . . . 10

2.6.3 Conclusions . . . 11

3 Appendix **11**

3.1 Data pre-processing . . . 11

1 Introduction

Purpose of the analysis: Investigate differences in PD-L1 (CD274) expression levels betweeen molecular & pathohistological subgroups of lower grade gliomas (LGG) as well as glioblastoma tumors (GBM), using publicly available from The Canger Genome Atlas (TCGA).

TCGA data for LGG and GBM samples has been downloaded between November 2015 and January 2016, and has been pre-processed to produce single data frames of expression valus for each dataset. Additionally, Pierre Bady ([link]pierre.bady@unil.ch) has compiled information about somatic mutations (incl. IDH1/2), CNV status, and CIMP status for the LGG dataset and kindly made these available to allow stratification of LGG samples.

---

2 Results

2.1 Data preparation

```
#----------  load the prepared clinical LGG data --------------
load("~/Data/Collaborations/LGG_PDL1/clinpatient.rda")
# some data re-formatting
clinpatient$bcr_patient_barcode <- as(clinpatient$bcr_patient_barcode, "character")

#---------- load LGG RNA-Seq data from pre-processing step --------------
# TCGA Level 3 RNA-Seq gene-level data was used, which is stored as FPKM
# (estimated fragments, normalized for transcript length) this data has been
# log2 transformed with an offset of 1 to avoid -Inf and approximate normal
# distribution of the data
load("~/Data/Collaborations/LGG_PDL1/lgg.rna_seq.norm.log2.rda")

#---------- load GBM RNA-Seq data from pre-processing step --------------
# TCGA Level 3 RNA-Seq gene-level data was used, which is stored as FPKM
# (estimated fragments, normalized for transcript length) this data has been
# log2 transformed with an offset of 1 to avoid -Inf and approximate normal
# distribution of the data
load("~/Data/Collaborations/LGG_PDL1/gbm.rna_seq.norm.log2.rda")
```

Number of available samples for LGG and GBM, respectively

```
# LGG samples
ncol(lgg.rna_seq.norm.log2)
```

```
## [1] 430
```

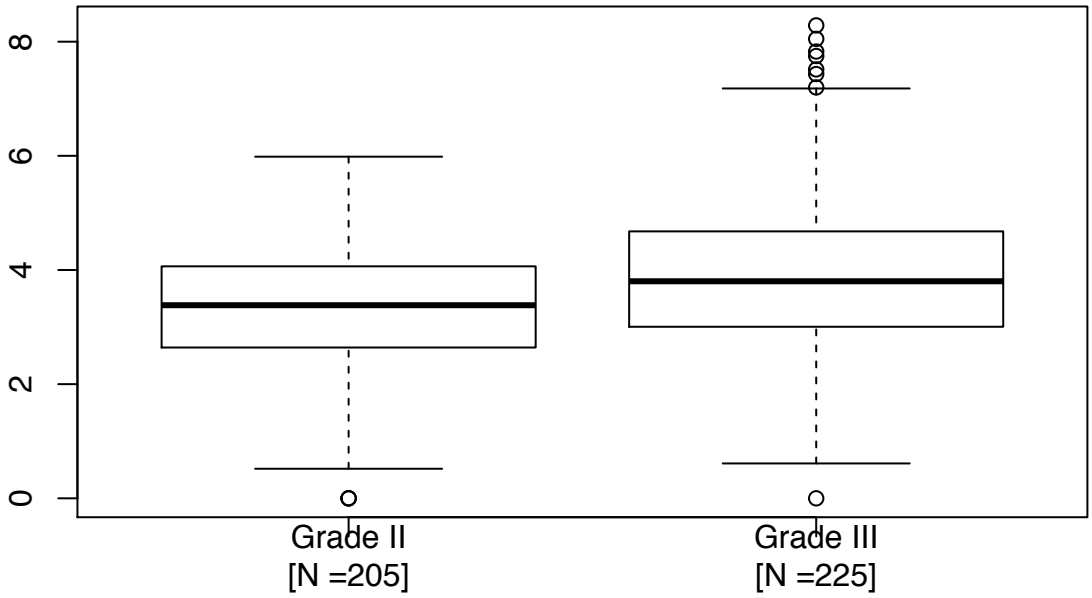
```
# GBM samples
ncol(gbm.rna_seq.norm.log2)
```

```
## [1] 166
```

2.2 Comparing PD-L1 expression between Grade II and III samples:

2.2.1 Boxplot

```
#-------------- Boxplot of PD-L1 expression levels, comparing LGG Grade II & III --------------
# The EntrezID for PD-L1/CD274 is 29126
PDL1 <- grep(29126, rownames(lgg.rna_seq.norm.log2))
gradeII <- clinpatient[which(clinpatient$grade == 2), "bcr_patient_barcode"]
gradeIII <- clinpatient[which(clinpatient$grade == 3), "bcr_patient_barcode"]
boxplot(unlist(lgg.rna_seq.norm.log2[PDL1, gradeII]), unlist(lgg.rna_seq.norm.log2[PDL1,
    gradeIII]), names = c(paste("Grade II\n[N =", length(gradeII), "]", sep = ""),
    paste("Grade III\n[N =", length(gradeIII), "]", sep = "")))
```




```
# T-test to compare the two population means
t.test(lgg.rna_seq.norm.log2[PDL1, gradeII], lgg.rna_seq.norm.log2[PDL1, gradeIII])
```

```
##
##  Welch Two Sample t-test
##
## data:  lgg.rna_seq.norm.log2[PDL1, gradeII] and lgg.rna_seq.norm.log2[PDL1, gradeIII]
## t = -4.6754, df = 417.96, p-value = 3.966e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8241116 -0.3362594
## sample estimates:
## mean of x mean of y
##  3.356597  3.936782
```

2.2.2 Conclusions

There is a small, but statistically significant difference in mean PD-L1 expression levels between Grade II and III tumors.

2.3 Comparing LGG samples stratified by IDH1/2 mutation status

2.3.1 Data preparation

```
#--------------- collate PD-L1 expression levels and survival data for patients with known IDH1 status -
IDHmut <- clinpatient[which(clinpatient$mIDHtot == 1), "bcr_patient_barcode"]
IDHwt <- clinpatient[which(clinpatient$mIDHtot == 0), "bcr_patient_barcode"]
IDHmut1p19qcodel <- clinpatient[which(clinpatient$mIDHtot == 1 & clinpatient$co1p19q ==
    "cd"), "bcr_patient_barcode"]
IDHmut1p19qnorm <- clinpatient[which(clinpatient$mIDHtot == 1 & clinpatient$co1p19q ==
    "n"), "bcr_patient_barcode"]
CIMPpos <- clinpatient[which(clinpatient$hCIMP == 1), "bcr_patient_barcode"]
CIMPneg <- clinpatient[which(clinpatient$hCIMP == 0), "bcr_patient_barcode"]

PDL1.IDHstatusKnownPatients <- c(unlist(lgg.rna_seq.norm.log2[PDL1, IDHmut1p19qcodel]),
    unlist(lgg.rna_seq.norm.log2[PDL1, IDHmut1p19qnorm]), unlist(lgg.rna_seq.norm.log2[PDL1,
        IDHwt]))
PDL1.IDHstatusKnownPatients <- data.frame(PDL1exp = PDL1.IDHstatusKnownPatients,
    status = c(rep("IDHmut1p19qcodel", length(IDHmut1p19qcodel)), rep("IDHmut1p19qnorm",
        length(IDHmut1p19qnorm)), rep("IDHwt", length(IDHwt))), months_to_death = clinpatient[c(IDHmut1
        IDHmut1p19qnorm, IDHwt), "days_to_death"]/30)
PDL1.IDHstatusKnownPatients$death <- 0
PDL1.IDHstatusKnownPatients[!is.na(PDL1.IDHstatusKnownPatients$months_to_death),
    "death"] <- 1
```

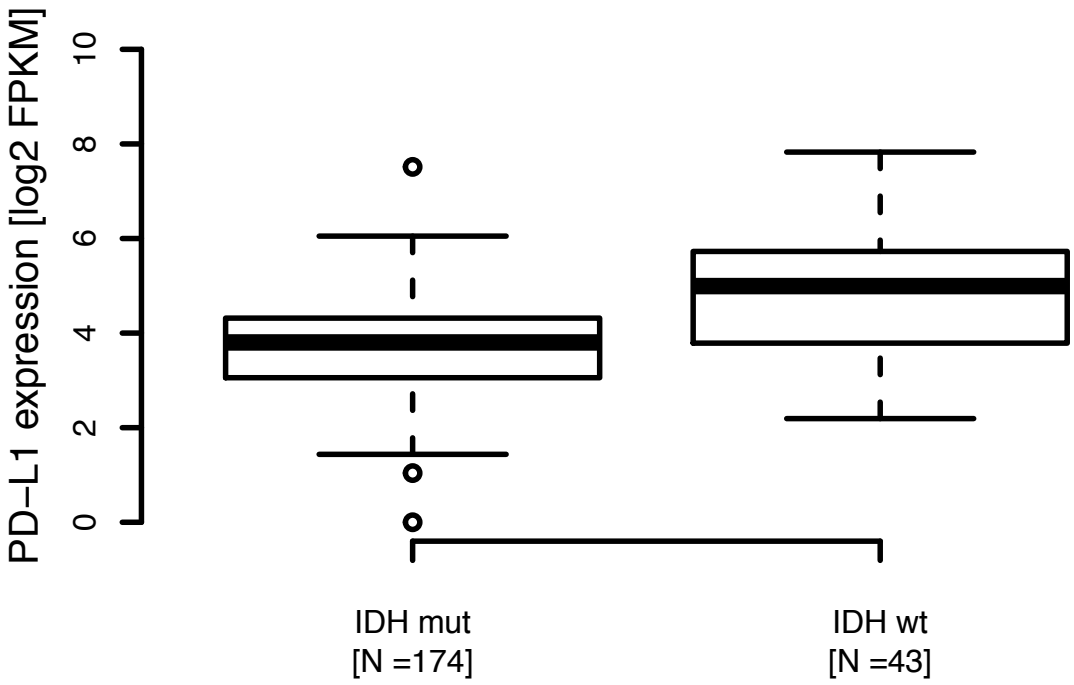
2.3.2 Boxplot

```
#--------------- PD-L1 expression levels, comparing IDH mut to wt, irrespective of grade --------------
boxplot(PDL1.IDHstatusKnownPatients[IDHmut, "PDL1exp"], PDL1.IDHstatusKnownPatients[IDHwt,
    "PDL1exp"], axes = FALSE, frame = FALSE, lwd = 2.75, ylim = c(0, 10))

axis(1, lwd = 2.75, at = c(1, 2), labels = c(paste("IDH mut\n[N =", length(IDHmut),
    "]", sep = ""), paste("IDH wt\n[N =", length(IDHwt), "]", sep = "")), padj = 1)

axis(2, lwd = 2.75)

mtext(side = 2, "PD-L1 expression [log2 FPKM]", line = 2.5, cex = 1.25)
```



2.3.3 Statistical Test (t-Test)

```
# T-test to compare the two population means
t.test(PDL1.IDHstatusKnownPatients[IDHmut, "PDL1exp"],
       PDL1.IDHstatusKnownPatients[IDHwt, "PDL1exp"])
```

```
##
##  Welch Two Sample t-test
##
## data:  PDL1.IDHstatusKnownPatients[IDHmut, "PDL1exp"] and PDL1.IDHstatusKnownPatients[IDHwt, "PDL1exp
## t = -5.2216, df = 52.544, p-value = 3.086e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.6953375 -0.7542191
## sample estimates:
## mean of x mean of y
##  3.698944  4.923722
```

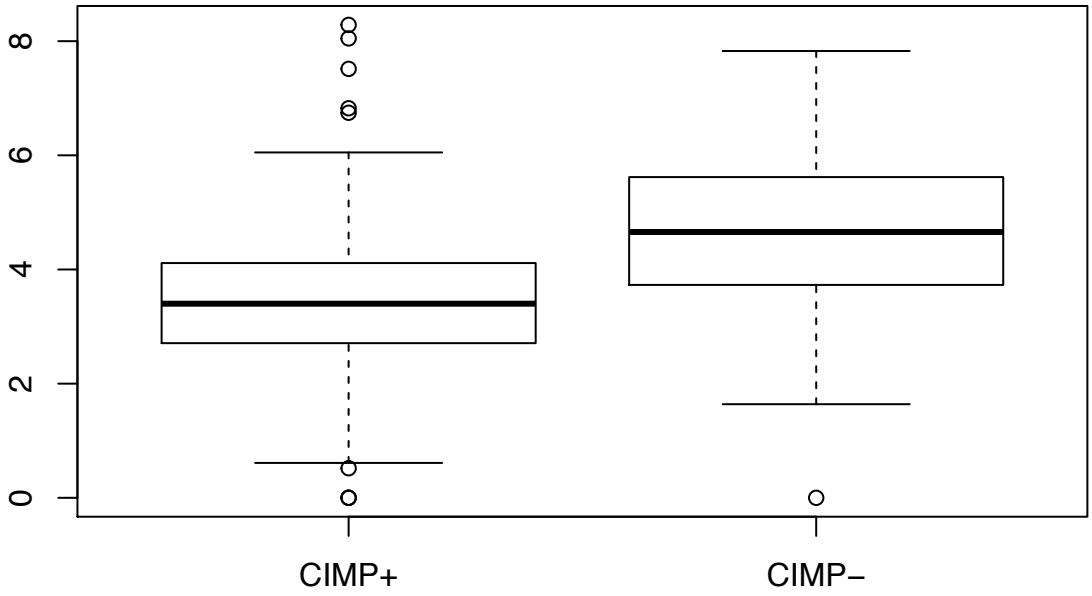
2.3.4 Conclusions

There is a statistically significant difference in PD-L1 expression when comparing IDH1/2^{mut} to IDH^{wt} samples, with higher expression levels found in IDH^{wt} samples.

2.4 Comparing LGG CIMP positive to negative samples

2.4.1 Boxplot

```
#--------------- PD-L1 expression levels, comparing CIMP pos to neg, irrespective of grade -------------
boxplot(unlist(lgg.rna_seq.norm.log2[PDL1, CIMPpos]), unlist(lgg.rna_seq.norm.log2[PDL1,
    CIMPneg]), names = c("CIMP+", "CIMP-"))
```



2.4.2 Statistical Test (t-Test)

```
# T-test to compare the two population means
t.test(lgg.rna_seq.norm.log2[PDL1, CIMPpos],
       lgg.rna_seq.norm.log2[PDL1, CIMPneg])
```

```
##
##  Welch Two Sample t-test
##
## data:  lgg.rna_seq.norm.log2[PDL1, CIMPpos] and lgg.rna_seq.norm.log2[PDL1, CIMPneg]
## t = -6.7587, df = 98.742, p-value = 9.815e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.6170476 -0.8830508
## sample estimates:
## mean of x mean of y
##  3.426332  4.676381
```

2.4.3 Conclusions

The previous finding is replicated when looking at $CIMP^{+/-}$ status instead of $IDH^{mut/wt}$ - $CIMP^{+}$ status is causally linked with IDH1/2 mutation and could therefore be used as a

proxy. This is important, as CIMP status is available for nearly twice as many samples as IDH mutation status.

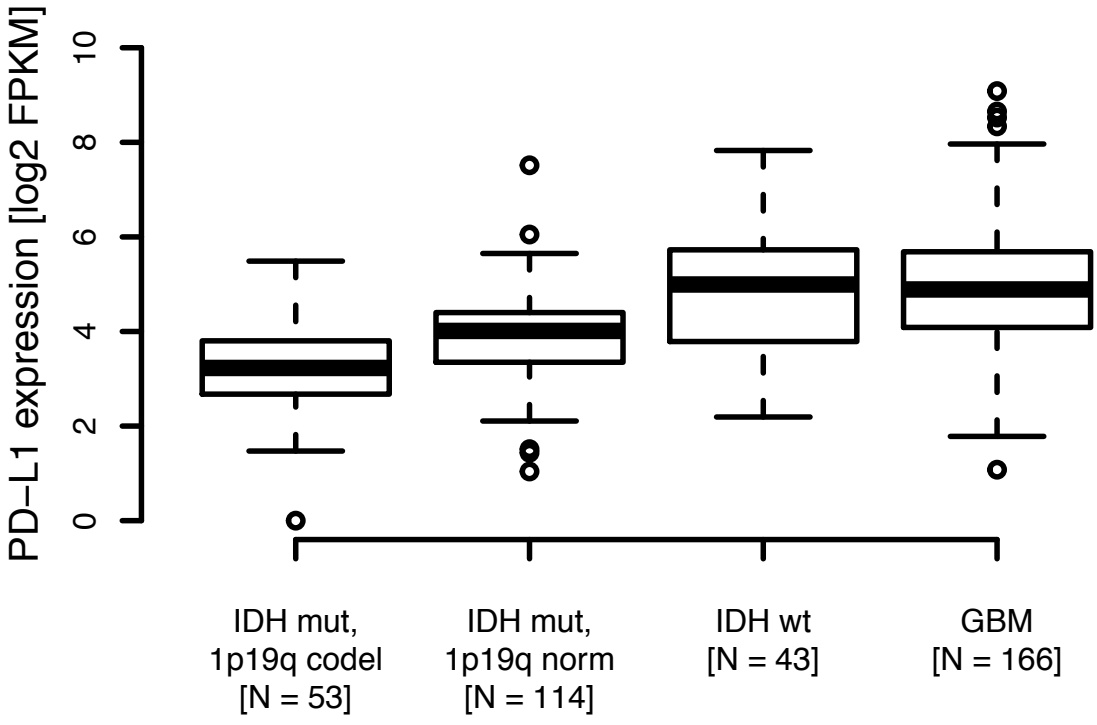
2.5 Comparing LGG samples stratified by IDH and 1p19q status to GBM

2.5.1 Boxplot

```
#--------------- PD-L1 expression levels, comparing IDHmut/1p19qcodel to IDmut/1p19qn to IDHwt, includin
tab1 <- PDL1.IDHstatusKnownPatients[, c("PDL1exp", "status")]
tab2 <- data.frame(PDL1exp = unlist(gbm.rna_seq.norm.log2[PDL1, ]), status = rep("GBM",
    ncol(gbm.rna_seq.norm.log2)))
tab1 <- rbind(tab1, tab2)
tab1$status <- factor(as.character(tab1$status), levels = c("IDHmut1p19qcodel",
    "IDHmut1p19qnorm", "IDHwt", "GBM"))


boxplot(tab1$PDL1exp ~ tab1$status, axes = FALSE, frame = FALSE, lwd = 2.75,
    ylim = c(0, 10))

axis(1, lwd = 2.75, at = c(1, 2, 3, 4), labels = c(paste("IDH mut,\n1p19q codel\n[N = ",
    length(IDHmut1p19qcodel), "]", sep = ""), paste("IDH mut,\n1p19q norm\n[N = ",
    length(IDHmut1p19qnorm), "]", sep = ""), paste("IDH wt\n[N = ", length(IDHwt),
    "]", sep = ""), paste("GBM\n[N = ", ncol(gbm.rna_seq.norm.log2), "]", sep = "")),
    padj = 1)
axis(2, lwd = 2.75)
mtext(side = 2, "PD-L1 expression [log2 FPKM]", line = 2.5, cex = 1.25)
```



2.5.2 Statistical Test (ANOVA)

```
# performing ANOVA to check if differences between groups are significant
aov2 <- aov(formula = PDL1exp ~ status, data = tab1)
# post-hoc analysis shows that they are

TukeyHSD(aov2)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = PDL1exp ~ status, data = tab1)
##
## $status
##                                        diff        lwr      upr     p adj
## IDHmut1p19qnorm-IDHmut1p19qcodel  0.67417979  0.1376299 1.210730 0.0070430
## IDHwt-IDHmut1p19qcodel            1.68499686  1.0226191 2.347375 0.0000000
## GBM-IDHmut1p19qcodel              1.64180675  1.1326256 2.150988 0.0000000
## IDHwt-IDHmut1p19qnorm             1.01081707  0.4332460 1.588388 0.0000499
## GBM-IDHmut1p19qnorm               0.96762696  0.5750590 1.360195 0.0000000
## GBM-IDHwt                        -0.04319011 -0.5954292 0.509049 0.9970861
```

```
bartlett.test(tab1$PDL1exp ~ tab1$status)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  tab1$PDL1exp by tab1$status
## Bartlett's K-squared = 31.197, df = 3, p-value = 7.728e-07
```

```
# but Bartlett test for homogeneity of variances shows that the assumption is violated
# so the ANOVA has to be taken with a grain of salt
# [possibly due to uneven sized groups? or because values are derived from count data?]
```

2.5.3 Conclusions

These esults show that we observe statistically significant difference in PD-L1 expression levels between the three LGG sub-groups (IDH^{mut} & 1p19q co-del, IDH^{mut} & 1p19q normal, IDH^{wt}), while there is no difference between LGG IDH^{wt} and GBM samples.*

2.6 Comparing LGG samples stratified by CIMP and 1p19q status to GBM

2.6.1 Boxplot

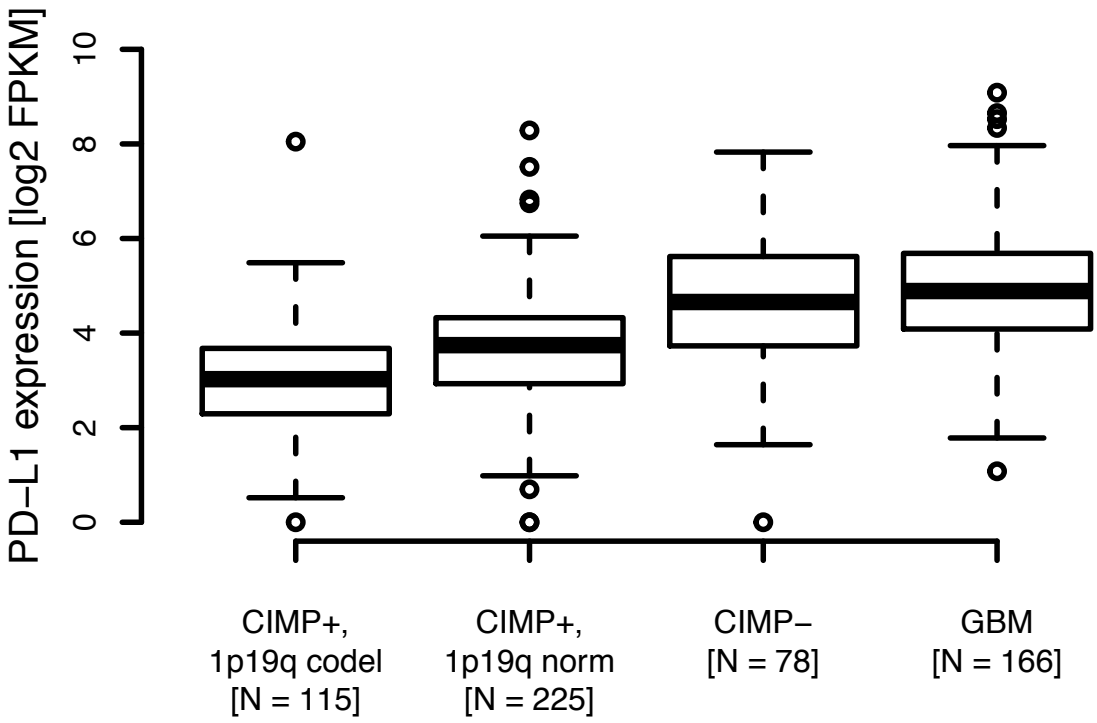
```
#-------------- collate PD-L1 expression levels and survival data for patients with known CIMP status [
PDL1 <- grep(29126, rownames(lgg.rna_seq.norm.log2))
CIMPpos1p19qcodel <- clinpatient[which(clinpatient$hCIMP == 1 & clinpatient$co1p19q ==
    "cd"), "bcr_patient_barcode"]
CIMPpos1p19qnorm <- clinpatient[which(clinpatient$hCIMP == 1 & clinpatient$co1p19q ==
    "n"), "bcr_patient_barcode"]
CIMPpos <- clinpatient[which(clinpatient$hCIMP == 1), "bcr_patient_barcode"]
CIMPneg <- clinpatient[which(clinpatient$hCIMP == 0), "bcr_patient_barcode"]

PDL1.CIMPstatusKnownPatients <- c(unlist(lgg.rna_seq.norm.log2[PDL1, CIMPpos1p19qcodel]),
    unlist(lgg.rna_seq.norm.log2[PDL1, CIMPpos1p19qnorm]), unlist(lgg.rna_seq.norm.log2[PDL1,
        CIMPneg]))
PDL1.CIMPstatusKnownPatients <- data.frame(PDL1exp = PDL1.CIMPstatusKnownPatients,
    status = c(rep("CIMPpos1p19qcodel", length(CIMPpos1p19qcodel)), rep("CIMPpos1p19qnorm",
        length(CIMPpos1p19qnorm)), rep("CIMPneg", length(CIMPneg))), months_to_death = clinpatient[c(CI
        CIMPpos1p19qnorm, CIMPneg), "days_to_death"]/30)
PDL1.CIMPstatusKnownPatients$death <- 0
PDL1.CIMPstatusKnownPatients[!is.na(PDL1.CIMPstatusKnownPatients$months_to_death),
    "death"] <- 1

#-------------- PD-L1 expression levels, comparing CIMPpos/1p19qcodel to CIMPpos/1p19qn to CIMPneg, inc
tab1 <- PDL1.CIMPstatusKnownPatients[, c("PDL1exp", "status")]
tab2 <- data.frame(PDL1exp = unlist(gbm.rna_seq.norm.log2[PDL1, ]), status = rep("GBM",
    ncol(gbm.rna_seq.norm.log2)))
tab1 <- rbind(tab1, tab2)
tab1$status <- factor(as.character(tab1$status), levels = c("CIMPpos1p19qcodel",
    "CIMPpos1p19qnorm", "CIMPneg", "GBM"))

boxplot(tab1$PDL1exp ~ tab1$status, axes = FALSE, frame = FALSE, lwd = 2.75,
    ylim = c(0, 10))

axis(1, lwd = 2.75, at = c(1, 2, 3, 4), labels = c(paste("CIMP+,\n1p19q codel\n[N = ",
    length(CIMPpos1p19qcodel), "]", sep = ""), paste("CIMP+,\n1p19q norm\n[N = ",
    length(CIMPpos1p19qnorm), "]", sep = ""), paste("CIMP-\n[N = ", length(CIMPneg),
    "]", sep = ""), paste("GBM\n[N = ", ncol(gbm.rna_seq.norm.log2), "]", sep = "")),
    padj = 1)
axis(2, lwd = 2.75)
mtext(side = 2, "PD-L1 expression [log2 FPKM]", line = 2.5, cex = 1.25)
```



2.6.2 Statistical Test (ANOVA)

```
# performing ANOVA to check if differences between groups are significant
aov2 <- aov(formula = PDL1exp ~ status, data = tab1)
# post-hoc analysis shows that they are

TukeyHSD(aov2)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = PDL1exp ~ status, data = tab1)
##
## $status
##                                         diff        lwr       upr
## CIMPpos1p19qnorm-CIMPpos1p19qcodel 0.6627982  0.2820391 1.0435573
## CIMPneg-CIMPpos1p19qcodel          1.6881851  1.2009558 2.1754145
## GBM-CIMPpos1p19qcodel              1.8923361  1.4893396 2.2953325
## CIMPneg-CIMPpos1p19qnorm           1.0253870  0.5889377 1.4618362
## GBM-CIMPpos1p19qnorm               1.2295379  0.8896825 1.5693932
## GBM-CIMPneg                        0.2041509 -0.2518278 0.6601297
##                                        p adj
## CIMPpos1p19qnorm-CIMPpos1p19qcodel 0.0000519
## CIMPneg-CIMPpos1p19qcodel          0.0000000
## GBM-CIMPpos1p19qcodel              0.0000000
## CIMPneg-CIMPpos1p19qnorm           0.0000000
## GBM-CIMPpos1p19qnorm               0.0000000
## GBM-CIMPneg                        0.6564049
```

```r
bartlett.test(tab1$PDL1exp ~ tab1$status)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  tab1$PDL1exp by tab1$status
## Bartlett's K-squared = 20.112, df = 3, p-value = 0.0001609
```

```r
# but the Bartlett test for homogeneity of variances shows that the assumption is violated
# so the ANOVA has to be taken with a grain of salt
# [possibly due to uneven sized groups? or because values are derived from count data?]
```

2.6.3 Conclusions

These results indicate that there is a statistically significant difference when comparing mean PD-L1 expression levels between CIMP$^+$ [both 1p19 normal and co-deleted] and CIMP$^-$ LGG samples as well as GBM samples. However, there is significant difference between the expression levels in GBMs and CIMP$^-$ LGGs. After correcting a coding error in the analysis, I can now confirm that the trend the same trend in PD-L1 expression differences is observed between 1p19q co-deleted and normal subgroups, irrespective of using IDHmut or CIMP status.

---

3 Appendix

3.1 Data pre-processing

Included for documentation purpose.

```r
#---------- Expression & DNA methylation data on GDU cluster
path <- "/Volumes/gduserv/Data/TCGA/LGG"

#---------- Import list of LGG RNA-Seq samples --------------
lgg.RNASeq.samples <- read.table(paste(path, "METADATA", "UNC__IlluminaHiSeq_RNASeqV2",
    "unc.edu_LGG.IlluminaHiSeq_RNASeqV2.1.15.0.sdrf.txt", sep = "/"), header = T,
    as.is = T, sep = "\t")
lgg.RNASeq.samples$sample_id <- unlist(lapply(strsplit(lgg.RNASeq.samples$Comment..TCGA.Barcode.,
    "-"), function(x) paste(x[1:3], collapse = "-")))

# only retain rows with RSEM gene-level normalized data files & filter out
# duplicated IDs
lgg.RNASeq.samples <- lgg.RNASeq.samples[grep("RSEM_genes_normalized", lgg.RNASeq.samples$Protocol.REF.
    ]
lgg.RNASeq.samples <- lgg.RNASeq.samples[!duplicated(lgg.RNASeq.samples$sample_id),
    ]

# get the unique files
lgg.sampleIDs <- lgg.RNASeq.samples$sample_id
lgg.rna_seq_files <- lgg.RNASeq.samples$Derived.Data.File
names(lgg.rna_seq_files) <- lgg.sampleIDs
```

```
#---------- TCGA LGG RNA-Seq data processing --------------
path.rna_seq <- paste(path, "/RNASeqV2/UNC__IlluminaHiSeq_RNASeqV2/Level_3/",
    sep = "")

# using normalized data PD-L1 (CD274) Entrez gene ID: 29126
PDL1 <- grep(29126, rownames(lgg.rna_seq.norm.log2))
for (i in 1:length(lgg.rna_seq_files)) {
    if (i == 1) {
        temp <- read.table(paste(path.rna_seq, lgg.rna_seq_files[i], sep = ""),
            header = T, as.is = T, sep = "\t")
    } else {
        temp2 <- read.table(paste(path.rna_seq, lgg.rna_seq_files[i], sep = ""),
            header = T, as.is = T, sep = "\t")
        temp <- cbind(temp, temp2$normalized_count)
    }
}
lgg.rna_seq.norm <- temp
rm(temp)
rownames(lgg.rna_seq.norm) <- lgg.rna_seq.norm$gene_id
lgg.rna_seq.norm <- lgg.rna_seq.norm[, -1]
colnames(lgg.rna_seq.norm) <- names(lgg.rna_seq_files)
# save(lgg.rna_seq.norm, file =
# '~/Data/Collaborations/LGG_PDL1/lgg.rna_seq.rda') we only retain data for
# patients with clinical data
lgg.rna_seq.norm <- lgg.rna_seq.norm[, which(colnames(lgg.rna_seq.norm) %in%
    rownames(clinpatient))]
# log2 transformation of data, adding offset of 1 to avoid -Inf
lgg.rna_seq.norm.log2 <- log2(lgg.rna_seq.norm + 1)
# save(lgg.rna_seq.norm.log2, file =
# '~/Data/Collaborations/LGG_PDL1/lgg.rna_seq.norm.log2.rda') PD-L1
# expression levels: summary(unlist(lgg.rna_seq.norm.log2[grep(29126,
# rownames(lgg.rna_seq.norm.log2)),])) Min. 1st Qu.  Median Mean 3rd Qu.
# Max.  0.000 2.830 3.674 3.660 4.358 8.285 human PD-L1 (CD274, EntrezID:
# 29126) is located on chr 9p24

#---------- Import list of GBM RNA-Seq samples --------------
gbmPath <- "/home/skurscheid/Data/TCGA/GBM"
gbm.RNASeq.samples <- read.table(paste(gbmPath, "METADATA", "UNC__IlluminaHiSeq_RNASeqV2",
    "unc.edu_GBM.IlluminaHiSeq_RNASeqV2.1.4.0.sdrf.txt", sep = "/"), header = T,
    as.is = T, sep = "\t")
gbm.RNASeq.samples$sample_id <- unlist(lapply(strsplit(gbm.RNASeq.samples$Comment..TCGA.Barcode.,
    "-"), function(x) paste(x[1:3], collapse = "-")))

# only retain rows with RSEM gene-level normalized data files & filter out
# duplicated IDs
gbm.RNASeq.samples <- gbm.RNASeq.samples[grep("RSEM_genes_normalized", gbm.RNASeq.samples$Protocol.REF.4
    ]
gbm.RNASeq.samples <- gbm.RNASeq.samples[!duplicated(gbm.RNASeq.samples$sample_id),
    ]

# get the unique files
gbm.sampleIDs <- gbm.RNASeq.samples$sample_id
gbm.rna_seq_files <- gbm.RNASeq.samples$Derived.Data.File
```

```
names(gbm.rna_seq_files) <- gbm.sampleIDs

#---------- TCGA LGG RNA-Seq data processing --------------
path.rna_seq <- paste(gbmPath, "/RNASeqV2/UNC__IlluminaHiSeq_RNASeqV2/Level_3/",
    sep = "")

# using normalized data PD-L1 (CD274) Entrez gene ID: 29126
for (i in 1:length(gbm.rna_seq_files)) {
    if (i == 1) {
        temp <- read.table(paste(path.rna_seq, gbm.rna_seq_files[i], sep = ""),
            header = T, as.is = T, sep = "\t")
    } else {
        temp2 <- read.table(paste(path.rna_seq, gbm.rna_seq_files[i], sep = ""),
            header = T, as.is = T, sep = "\t")
        temp <- cbind(temp, temp2$normalized_count)
    }
}
gbm.rna_seq.norm <- temp
rm(temp)
rownames(gbm.rna_seq.norm) <- gbm.rna_seq.norm$gene_id
gbm.rna_seq.norm <- gbm.rna_seq.norm[, -1]
colnames(gbm.rna_seq.norm) <- names(gbm.rna_seq_files)
# save(lgg.rna_seq.norm, file =
# '~/Data/Collaborations/LGG_PDL1/lgg.rna_seq.rda') we only retain data for
# patients with clinical data log2 transformation of data, adding offset of
# 1 to avoid -Inf
gbm.rna_seq.norm.log2 <- log2(gbm.rna_seq.norm + 1)
# save(lgg.rna_seq.norm.log2, file =
# '~/Data/Collaborations/LGG_PDL1/lgg.rna_seq.norm.log2.rda') PD-L1
# expression levels: summary(unlist(lgg.rna_seq.norm.log2[grep(29126,
# rownames(lgg.rna_seq.norm.log2)),])) Min. 1st Qu.  Median Mean 3rd Qu.
# Max.  0.000 2.830 3.674 3.660 4.358 8.285 human PD-L1 (CD274, EntrezID:
# 29126) is located on chr 9p24
```