

```

library(data.table)
library(ggplot2)
library(ggpubr)

## Loading required package: magrittr

library(ggpointdensity)
theme_set(theme_bw())

# -----
# load results of deconvolution
# -----

load("deconv_expr_COAD.RData")
ls()

## [1] "deconv_expr_COAD"

load("deconv_methy_COAD.RData")
ls()

## [1] "deconv_expr_COAD" "rho_COAD"

dim(deconv_expr_COAD)

## [1] 185 7

dim(rho_COAD)

## [1] 185 7 6

deconv_expr_COAD[1:2,]

##           CD4T      CD8T  Monocyte      B      NK Neutrophil
## 5656 0.07627453 0.01670232 0.02561258 0.04381193 0.01982880 0.00652889
## 6781 0.06005796 0.12891577 0.10364729 0.07942259 0.05740941 0.04947972
##           Treg
## 5656 0.01124095
## 6781 0.02106727

rho_COAD[1,,]

##           LaplaceEM      OriEM      svr      ls      rls
## CD4T      0.02056698 6.894222e-03 0.019335272 0.00000000 0.000000000
## CD8T      0.00000000 -4.019258e-18 0.067540167 0.12346010 0.119316714
## Monocyte  0.03787344 5.501763e-02 0.056029383 0.03145094 0.032486828
## B         0.02621915 2.488973e-02 0.046140866 0.02314771 0.024156154
## NK        0.02221589 1.069264e-03 0.005883450 0.00000000 0.001562089
## Neutrophil 0.02102997 5.308029e-04 0.005070863 0.02194124 0.022478216
## Treg      0.07209457 1.115984e-01 0.000000000 0.00000000 0.000000000
##           qp
## CD4T      2.000000e-01
## CD8T      -8.686388e-18
## Monocyte  1.205043e-16
## B         0.000000e+00
## NK        1.524040e-16
## Neutrophil -1.019300e-16
## Treg      -3.707308e-17

```

```
dimnames(rho_COAD)
```

```
## [[1]]
## [1] "5656" "6781" "5659" "3809" "5665" "2684" "3655" "6535" "6888" "6898"
## [11] "6138" "6751" "6534" "3506" "6890" "A285" "2675" "6321" "5916" "A28M"
## [21] "6965" "6548" "6651" "6166" "6162" "5537" "6806" "6142" "A1DB" "3660"
## [31] "A1D7" "6678" "5868" "2686" "6304" "5541" "A1D8" "5820" "6855" "6927"
## [41] "A1DA" "6704" "A28K" "4323" "6459" "6317" "6569" "6926" "5913" "6539"
## [51] "6625" "6674" "6627" "6004" "6922" "3662" "6307" "6648" "6929" "4616"
## [61] "6931" "A28F" "A0X9" "3495" "6703" "4615" "5860" "6680" "6540" "6920"
## [71] "6586" "6677" "6295" "6807" "6598" "6161" "6164" "6533" "4747" "4682"
## [81] "6899" "5407" "7000" "6461" "6165" "3489" "A282" "6530" "2679" "3492"
## [91] "5341" "5862" "6923" "5667" "6541" "A288" "2681" "6746" "3663" "6163"
## [101] "4751" "6196" "6315" "4315" "A280" "4105" "6628" "4614" "A28G" "6532"
## [111] "6932" "6924" "4313" "A28E" "3713" "6311" "6963" "A0XD" "6679" "3496"
## [121] "6603" "6856" "4748" "6626" "A1D6" "6303" "5539" "2671" "6782" "6588"
## [131] "6140" "4951" "6310" "6171" "5864" "5662" "6463" "5912" "A28C" "5348"
## [141] "5821" "A28H" "6964" "6805" "6529" "3488" "6320" "6538" "5403" "6601"
## [151] "3712" "4947" "A0XF" "3494" "6531" "5538" "6605" "5349" "3779" "6137"
## [161] "6854" "6172" "6928" "A1D9" "6930" "6676" "4107" "6294" "6314" "5915"
## [171] "6299" "6808" "5666" "3510" "6895" "A28A" "6747" "6606" "4308" "3502"
## [181] "3697" "A1HA" "6170" "6607" "6460"
##
## [[2]]
## [1] "CD4T" "CD8T" "Monocyte" "B" "NK"
## [6] "Neutrophil" "Treg"
##
## [[3]]
## [1] "LaplaceEM" "OriEM" "svr" "ls" "rls" "qp"
```

```
table(rownames(deconv_expr_COAD) == dimnames(rho_COAD)[[1]])
```

```
##
## TRUE
## 185
```

```
cor(deconv_expr_COAD[, "B"], rho_COAD[, "B", "OriEM"])
```

```
## [1] 0.8258726
```

```
# -----
# load results of clinical information
# -----
```

```
sc = readRDS("clinical_data/COAD_somatic_clinic.rds")
dim(sc)
```

```
## [1] 384 38
```

```
names(sc)
```

```
## [1] "barcode"
## [2] "SMASH_S_hg38"
## [3] "SMASH_oE_hg38"
## [4] "SMASH_wE_hg38"
## [5] "SMASH_oNE_hg38"
## [6] "SMASH_wNE_hg38"
```

```

## [7] "AscatPurity"
## [8] "AscatPloidy"
## [9] "raw_MB_hg38_SNV"
## [10] "raw_MB_hg38_INDEL"
## [11] "IDH_CNV_status_hg38"
## [12] "tCN_burden"
## [13] "tCN_burden_ap"
## [14] "tumor_type.x"
## [15] "num_clonal"
## [16] "num_subclonal"
## [17] "prop_clonal"
## [18] "gender"
## [19] "age"
## [20] "tumor_type.y"
## [21] "initial_pathologic_diagnosis_method"
## [22] "histological_type"
## [23] "pathologic_stage"
## [24] "gleason_score"
## [25] "psa_level"
## [26] "hpv_status_by_ish_testing"
## [27] "hpv_status_by_p16_testing"
## [28] "prior_glioma"
## [29] "history_of_neoadjuvant_treatment"
## [30] "breast_carcinoma_estrogen_receptor_status"
## [31] "breast_carcinoma_progesterone_receptor_status"
## [32] "neoplasm_histologic_grade"
## [33] "yr_of_tobacco_smoking_onset"
## [34] "pathologic_T"
## [35] "stage"
## [36] "Time"
## [37] "Delta"
## [38] "radiation_therapy"

sc = sc[,-(23:32)]
dim(sc)

## [1] 384 28

sc[1:2,]

##          barcode SMASH_S_hg38 SMASH_oE_hg38 SMASH_wE_hg38 SMASH_oNE_hg38
## 1 TCGA-3L-AA1B          2          0.69      0.6765692      0.2783978
## 2 TCGA-4N-A93T          3          0.95      1.0292827      0.1656512
## SMASH_wNE_hg38 AscatPurity AscatPloidy raw_MB_hg38_SNV raw_MB_hg38_INDEL
## 1      0.1958878      0.41      3.712254          94          2
## 2      0.1803782      0.82      1.847516          69          4
## IDH_CNV_status_hg38 tCN_burden tCN_burden_ap tumor_type.x num_clonal
## 1      IDH wild type 1.7672127      0.8270333      COAD          88
## 2      IDH wild type 0.3889078      0.4157566      COAD          56
## num_subclonal prop_clonal gender age tumor_type.y
## 1          6      0.9361702 FEMALE 61      Colon
## 2         13      0.8115942  MALE 67      Colon
## initial_pathologic_diagnosis_method histological_type
## 1                                <NA> Colon Adenocarcinoma
## 2                                <NA> Colon Adenocarcinoma
## yr_of_tobacco_smoking_onset pathologic_T stage Time Delta radiation_therapy

```

```
## 1          NA          T2      I 475.01      0          NO
## 2          NA          T4a     III 146.01      0          NO

emInfo = fread("clinical_data/patient_coad_M_info_hyperMeth.txt")
dim(emInfo)

## [1] 213  28

emInfo[1, ]

##      race gender          ethnicity bcr_patient_barcode patient_id
## 1: WHITE  MALE NOT HISPANIC OR LATINO      TCGA-A6-2671      2671
##      tissue_source_site birth_days_to last_contact_days_to death_days_to
## 1:          A6          -31329          648          1331
##      vital_status tumor_status ajcc_pathologic_tumor_stage
## 1:          Dead   WITH TUMOR          Stage IV
##      age_at_initial_pathologic_diagnosis      noHM_PC1      noHM_PC2      noHM_PC3
## 1:          85 -0.009249593 0.004130469 0.01562243
##      noHM_PC4          methylation_barcode
## 1: 0.003222672 TCGA-A6-2671-01A-01D-1407-05
##
##                                     methylation_file
## 1: jhu-usc.edu_COAD.HumanMethylation450.1.lvl-3.TCGA-A6-2671-01A-01D-1407-05.txt
##      platform          array
## 1: HumanMethylation450 RARER_p_TCGA_MixedRedos_N_GenomeWideSNP_6_D05_747712
##      segment_count abs_call abs_purity abs_ploidy abs_doublings hyperMeth
## 1:          182      called          0.73          3.57          1          0
##      CIMP.Status
## 1:      Negative

table(emInfo$bcr_patient_barcode %in% sc$barcode)

##
## FALSE  TRUE
##      9    204

names(emInfo)[which(names(emInfo)=="bcr_patient_barcode")] = "barcode"

emInfo = merge(emInfo, sc, by="barcode")
dim(emInfo)

## [1] 204  55

# -----
# merge clinical information and cell type composition estimates
# -----

colnames(deconv_expr_COAD) = paste0(colnames(deconv_expr_COAD), ".E")

ct = data.frame(deconv_expr_COAD, rho_COAD[, "OriEM"])
dim(ct)

## [1] 185  14

ct[1:2, ]

##      CD4T.E      CD8T.E Monocyte.E      B.E      NK.E Neutrophil.E
## 5656 0.07627453 0.01670232 0.02561258 0.04381193 0.01982880 0.00652889
## 6781 0.06005796 0.12891577 0.10364729 0.07942259 0.05740941 0.04947972
##      Treg.E      CD4T      CD8T      Monocyte      B      NK
```

```
## 5656 0.01124095 0.006894222 -4.019258e-18 0.05501763 0.02488973 0.001069264
## 6781 0.02106727 0.015058624 6.213568e-02 0.14423068 0.02443667 0.025293428
##      Neutrophil      Treg
## 5656 0.0005308029 0.11159835
## 6781 0.1362817944 0.09256313

table(rownames(ct) %in% emInfo$patient_id)

##
## FALSE TRUE
##      8    177

ct$patient_id = rownames(ct)
ct = merge(ct, emInfo, by="patient_id")
dim(ct)

## [1] 177 69

ct[1:2,]

##      patient_id      CD4T.E      CD8T.E Monocyte.E      B.E      NK.E
## 1      2671 0.07669056 0.02889390 0.03723064 0.05887684 0.01678819
## 2      2675 0.14310630 0.05848576 0.09817728 0.14408533 0.03417972
##      Neutrophil.E      Treg.E      CD4T      CD8T      Monocyte      B
## 1 0.02455244 0.02696743 0.03263128 -3.096069e-20 0.06598839 0.02346060
## 2 0.04044031 0.06152530 0.30011767 7.320417e-18 0.07620356 0.06326375
##      NK Neutrophil      Treg      barcode      race gender.x
## 1 2.112018e-19 0.04600237 0.10191736 TCGA-A6-2671 WHITE      MALE
## 2 3.694618e-02 0.02214920 0.08131964 TCGA-A6-2675 WHITE      MALE
##      ethnicity tissue_source_site birth_days_to last_contact_days_to
## 1 NOT HISPANIC OR LATINO      A6      -31329      648
## 2 NOT HISPANIC OR LATINO      A6      -28813      1321
##      death_days_to vital_status tumor_status ajcc_pathologic_tumor_stage
## 1      1331      Dead      WITH TUMOR      Stage IV
## 2      NA      Alive      TUMOR FREE      Stage IIA
##      age_at_initial_pathologic_diagnosis      noHM_PC1      noHM_PC2      noHM_PC3
## 1      85 -0.009249593 0.004130469 0.01562243
## 2      78 0.098222522 0.023360098 0.04020544
##      noHM_PC4      methylation_barcode
## 1 0.0032226720 TCGA-A6-2671-01A-01D-1407-05
## 2 0.0009877077 TCGA-A6-2675-01A-02D-1721-05
##      methylation_file
## 1 jhu-usc.edu_COAD.HumanMethylation450.1.lv1-3.TCGA-A6-2671-01A-01D-1407-05.txt
## 2 jhu-usc.edu_COAD.HumanMethylation450.3.lv1-3.TCGA-A6-2675-01A-02D-1721-05.txt
##      platform      array
## 1 HumanMethylation450 RARER_p_TCGA_MixedRedos_N_GenomeWideSNP_6_D05_747712
## 2 HumanMethylation450 GRIPS_p_TCGA_b116_SNP_N_GenomeWideSNP_6_C07_781418
##      segment_count abs_call abs_purity abs_ploidy abs_doublings hyperMeth
## 1      182      called      0.73      3.57      1      0
## 2      230      called      0.42      3.37      1      0
##      CIMP.Status SMASH_S_hg38 SMASH_oE_hg38 SMASH_wE_hg38 SMASH_oNE_hg38
## 1      Negative      3      1.06      0.9274599      0.5531430
## 2      Negative      2      0.68      0.7213201      0.6600493
##      SMASH_wNE_hg38 AscatPurity AscatPloidy raw_MB_hg38_SNV raw_MB_hg38_INDEL
## 1      0.4846527      0.68      3.609337      49      7
## 2      0.6585205      0.49      3.351309      72      2
```

```
## IDH_CNV_status_hg38 tCN_burden tCN_burden_ap tumor_type.x num_clonal
## 1 IDH wild type 2.228662 1.0712969 COAD 39
## 2 IDH wild type 1.572784 0.8041944 COAD 46
## num_subclonal prop_clonal gender.y age tumor_type.y
## 1 10 0.7959184 MALE 85 Colon
## 2 26 0.6388889 MALE 78 Colon
## initial_pathologic_diagnosis_method histological_type
## 1 <NA> Colon Adenocarcinoma
## 2 <NA> Colon Adenocarcinoma
## yr_of_tobacco_smoking_onset pathologic_T stage Time Delta
## 1 NA T3 IV 1331.01 1
## 2 NA T3 II 1321.01 0
## radiation_therapy
## 1 NO
## 2 NO
```

```
names(ct)
```

```
## [1] "patient_id" "CD4T.E"
## [3] "CD8T.E" "Monocyte.E"
## [5] "B.E" "NK.E"
## [7] "Neutrophil.E" "Treg.E"
## [9] "CD4T" "CD8T"
## [11] "Monocyte" "B"
## [13] "NK" "Neutrophil"
## [15] "Treg" "barcode"
## [17] "race" "gender.x"
## [19] "ethnicity" "tissue_source_site"
## [21] "birth_days_to" "last_contact_days_to"
## [23] "death_days_to" "vital_status"
## [25] "tumor_status" "ajcc_pathologic_tumor_stage"
## [27] "age_at_initial_pathologic_diagnosis" "noHM_PC1"
## [29] "noHM_PC2" "noHM_PC3"
## [31] "noHM_PC4" "methylation_barcode"
## [33] "methylation_file" "platform"
## [35] "array" "segment_count"
## [37] "abs_call" "abs_purity"
## [39] "abs_ploidy" "abs_doublings"
## [41] "hyperMeth" "CIMP.Status"
## [43] "SMASH_S_hg38" "SMASH_oE_hg38"
## [45] "SMASH_wE_hg38" "SMASH_oNE_hg38"
## [47] "SMASH_wNE_hg38" "AscatPurity"
## [49] "AscatPloidy" "raw_MB_hg38_SNV"
## [51] "raw_MB_hg38_INDEL" "IDH_CNV_status_hg38"
## [53] "tCN_burden" "tCN_burden_ap"
## [55] "tumor_type.x" "num_clonal"
## [57] "num_subclonal" "prop_clonal"
## [59] "gender.y" "age"
## [61] "tumor_type.y" "initial_pathologic_diagnosis_method"
## [63] "histological_type" "yr_of_tobacco_smoking_onset"
## [65] "pathologic_T" "stage"
## [67] "Time" "Delta"
## [69] "radiation_therapy"
```

```

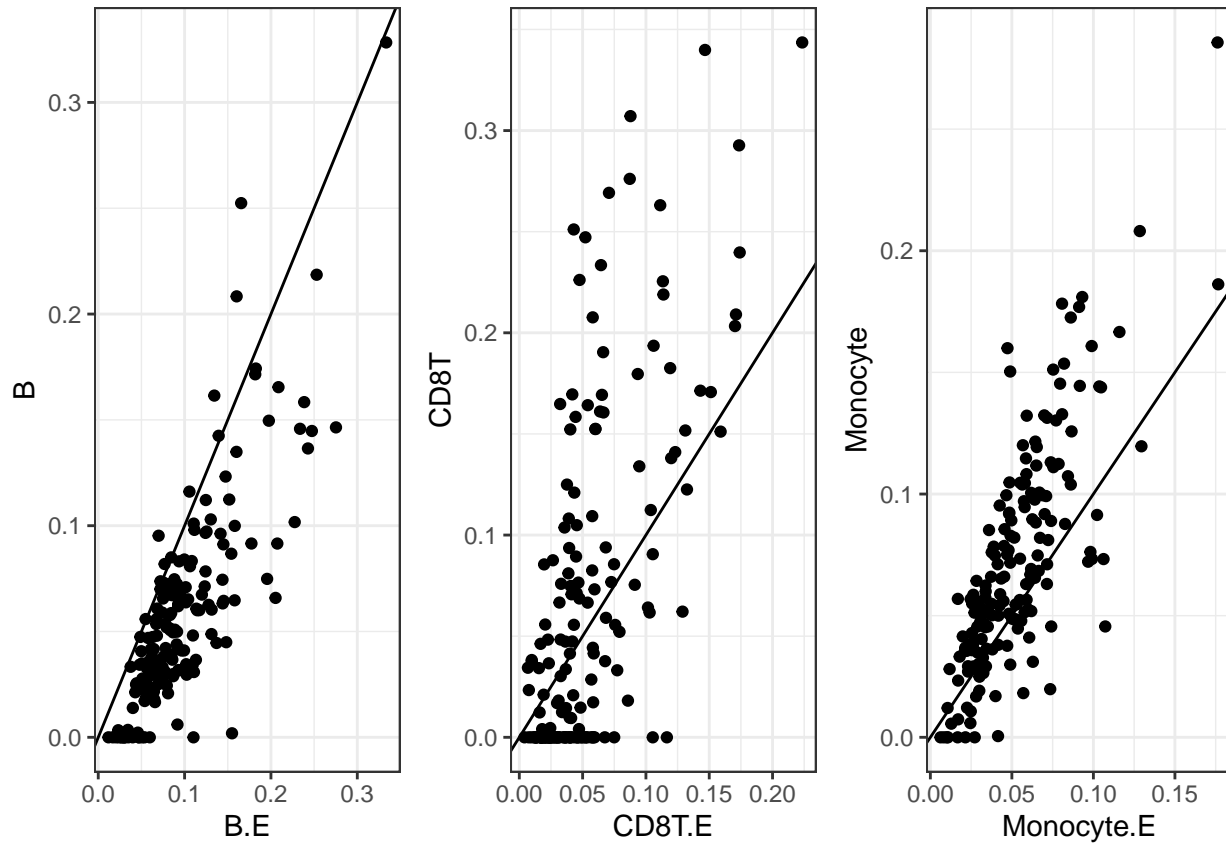
g1 = ggplot(ct, aes(x=B.E, y=B)) + geom_point() +
  geom_abline(intercept=0, slope=1)

g2 = ggplot(ct, aes(x=CD8T.E, y=CD8T)) + geom_point() +
  geom_abline(intercept=0, slope=1)

g3 = ggplot(ct, aes(x=Monocyte.E, y=Monocyte)) + geom_point() +
  geom_abline(intercept=0, slope=1)

ggarrange(g1, g2, g3, nrow = 1, ncol = 3)

```

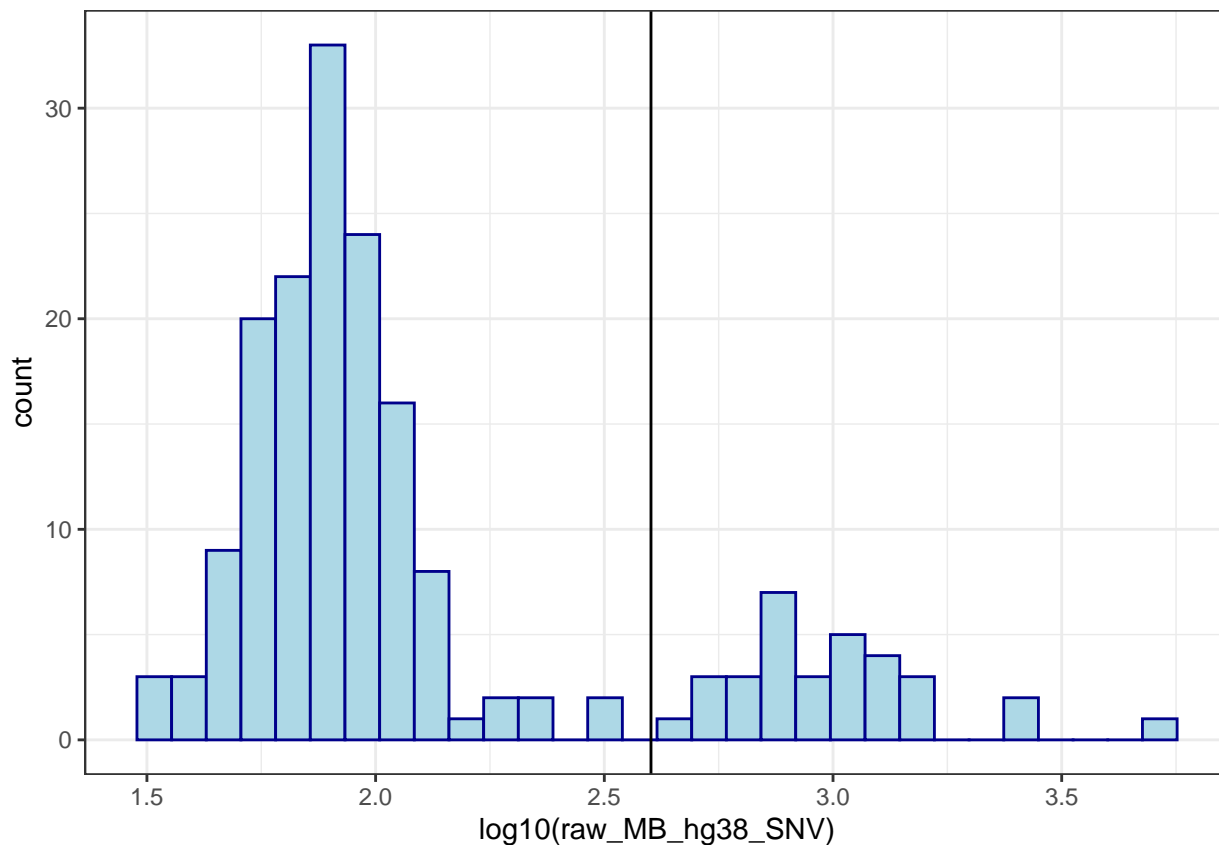


```

ggplot(ct, aes(x=log10(raw_MB_hg38_SNV)))+
  geom_histogram(color="darkblue", fill="lightblue") +
  geom_vline(xintercept = log10(400))

```

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



```
ct$hyper_mutation = ct[, "raw_MB_hg38_SNV"] > 400

g1 = ggplot(ct, aes(x=hyper_mutation, y=B.E)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(shape=16, position=position_jitter(0.2))

g2 = ggplot(ct, aes(x=hyper_mutation, y=B)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(shape=16, position=position_jitter(0.2))

g3 = ggplot(ct, aes(x=hyper_mutation, y=CD8T.E)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(shape=16, position=position_jitter(0.2))

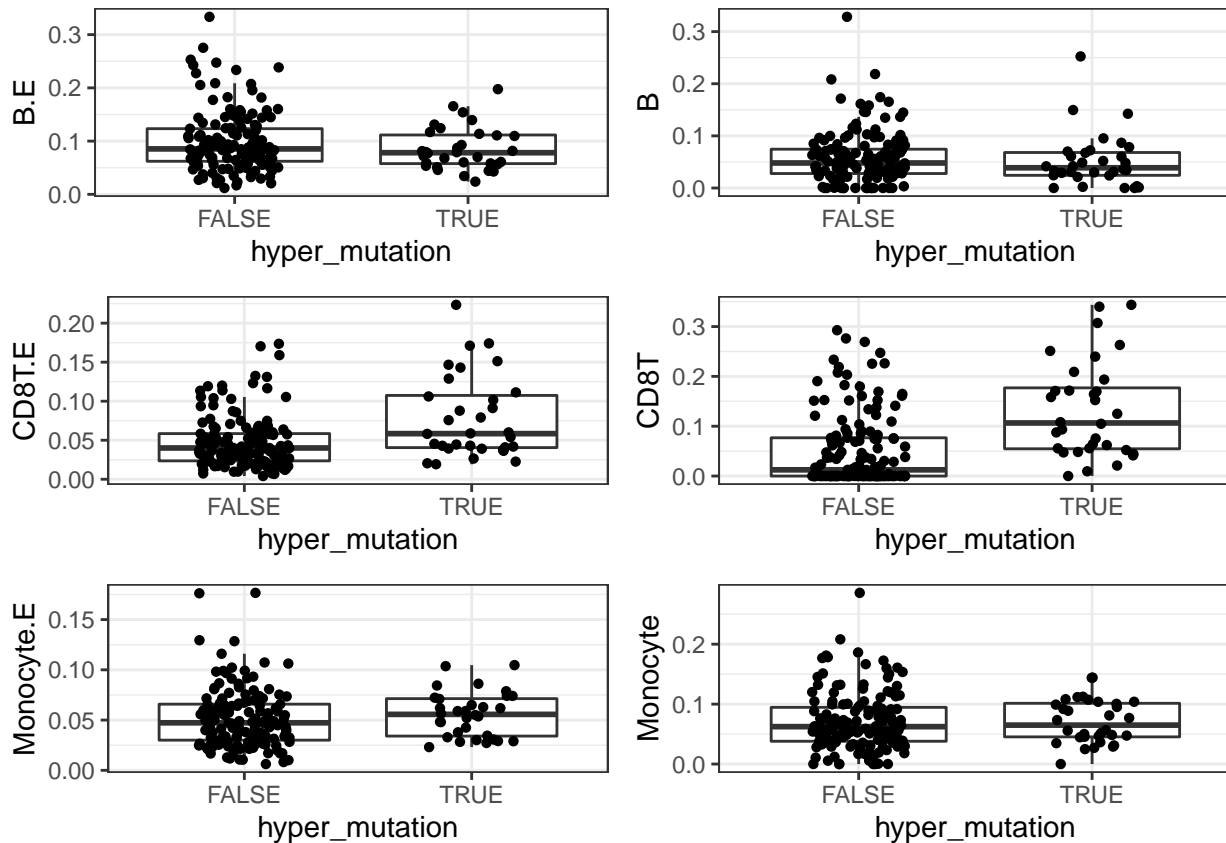
g4 = ggplot(ct, aes(x=hyper_mutation, y=CD8T)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(shape=16, position=position_jitter(0.2))

g5 = ggplot(ct, aes(x=hyper_mutation, y=Monocyte.E)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(shape=16, position=position_jitter(0.2))

g6 = ggplot(ct, aes(x=hyper_mutation, y=Monocyte)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(shape=16, position=position_jitter(0.2))

ggarrange(g1, g2, g3, g4, g5, g6, nrow = 3, ncol = 2)
```





```
wilcox.test(ct$CD8T~ ct$hyper_mutation)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: ct$CD8T by ct$hyper_mutation
## W = 1030, p-value = 8.867e-07
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(ct$CD8T.E~ ct$hyper_mutation)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: ct$CD8T.E by ct$hyper_mutation
## W = 1420, p-value = 0.0006066
## alternative hypothesis: true location shift is not equal to 0
```