```r
library(ggplot2)
library(ggpubr)
```

```
## Loading required package: magrittr
```

```r
library(ggpointdensity)
theme_set(theme_bw())

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

load("deconv_expr_COAD.RData")
ls()
```

```
## [1] "deconv_expr_COAD"
```

```r
load("deconv_methy_COAD.RData")
ls()
```

```
## [1] "deconv_expr_COAD" "rho_COAD"
```

```r
load("barcode/subsample_coad.RData")
ls()
```

```
## [1] "deconv_expr_COAD" "rho_COAD"           "subsample.coad"
```

```r
dim(deconv_expr_COAD)
```

```
## [1] 175    7
```

```r
dim(rho_COAD)
```

```
## [1] 175    7    5
```

```r
dim(subsample.coad)
```

```
## [1] 175    3
```

```r
deconv_expr_COAD[1:2,]
```

```
##          CD4T       CD8T   Monocyte          B         NK   Neutrophil
## 5656 0.10297062 0.02254813 0.03457699 0.0591461 0.02676888 0.008814001
## 6781 0.06966723 0.14954229 0.12023085 0.0921302 0.06659492 0.057396469
##          Treg
## 5656 0.01517529
## 6781 0.02443803
```

```r
rho_COAD[1,,]
```

```
##                   EMeth        svr         ls          rls            qp
## CD4T       4.522374e-02 0.08870032 0.00000000 0.0000000000  2.700000e-01
## CD8T       9.513481e-20 0.05843881 0.14527590 0.1433246771 -2.420395e-16
## Monocyte   6.179749e-02 0.05511302 0.04652121 0.0493472750 -9.639956e-17
## B          2.972052e-02 0.02804754 0.03134310 0.0336701679  0.000000e+00
## NK        -1.703507e-19 0.00707734 0.00000000 0.0004295355  1.299862e-16
## Neutrophil 5.421388e-02 0.01964418 0.04685979 0.0432283445 -1.481225e-16
## Treg       7.904436e-02 0.01297878 0.00000000 0.0000000000 -8.181769e-17
```

```
subsample.coad[1:2,]
```

```
##   num patient_id      barcode
## 1   1       2671 TCGA-A6-2671
## 2   2       2675 TCGA-A6-2675
```

```
dimnames(rho_COAD)
```

```
## [[1]]
##   [1] "1"   "2"   "3"   "4"   "5"   "6"   "7"   "8"   "9"   "10"  "11"  "12"
##  [13] "13"  "14"  "15"  "16"  "17"  "18"  "19"  "20"  "21"  "22"  "23"  "24"
##  [25] "25"  "26"  "27"  "28"  "29"  "30"  "31"  "32"  "33"  "34"  "35"  "36"
##  [37] "37"  "38"  "39"  "40"  "41"  "42"  "43"  "44"  "45"  "46"  "47"  "48"
##  [49] "49"  "50"  "51"  "52"  "53"  "54"  "55"  "56"  "57"  "58"  "59"  "60"
##  [61] "61"  "62"  "63"  "64"  "65"  "66"  "67"  "68"  "69"  "70"  "71"  "72"
##  [73] "73"  "74"  "75"  "76"  "77"  "78"  "79"  "80"  "81"  "82"  "83"  "84"
##  [85] "85"  "86"  "87"  "88"  "89"  "90"  "91"  "92"  "93"  "94"  "95"  "96"
##  [97] "97"  "98"  "99"  "100" "101" "102" "103" "104" "105" "106" "107" "108"
## [109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"
## [121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143" "144"
## [145] "145" "146" "147" "148" "149" "150" "151" "152" "153" "154" "155" "156"
## [157] "157" "158" "159" "160" "161" "162" "163" "164" "165" "166" "167" "168"
## [169] "169" "170" "171" "172" "173" "174" "175"
##
## [[2]]
## [1] "CD4T"      "CD8T"      "Monocyte"  "B"         "NK"
## [6] "Neutrophil" "Treg"
##
## [[3]]
## [1] "EMeth" "svr"   "ls"    "rls"   "qp"
```

```
table(rownames(deconv_expr_COAD) %in% subsample.coad$patient_id)
```
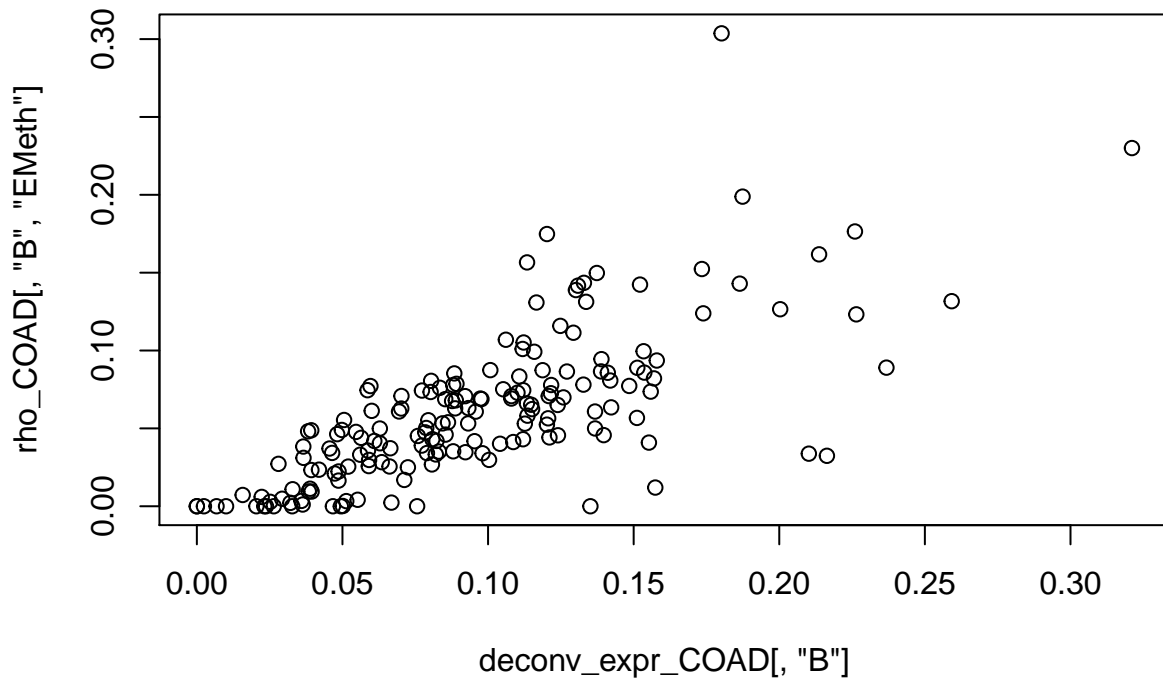
```
##
## TRUE
##  175
```

```
setequal(rownames(deconv_expr_COAD), subsample.coad$patient_id)
```

```
## [1] TRUE
```

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

```
## [1] 0.7301608
```

```
plot(deconv_expr_COAD[,"B"], rho_COAD[,"B","EMeth"])
```
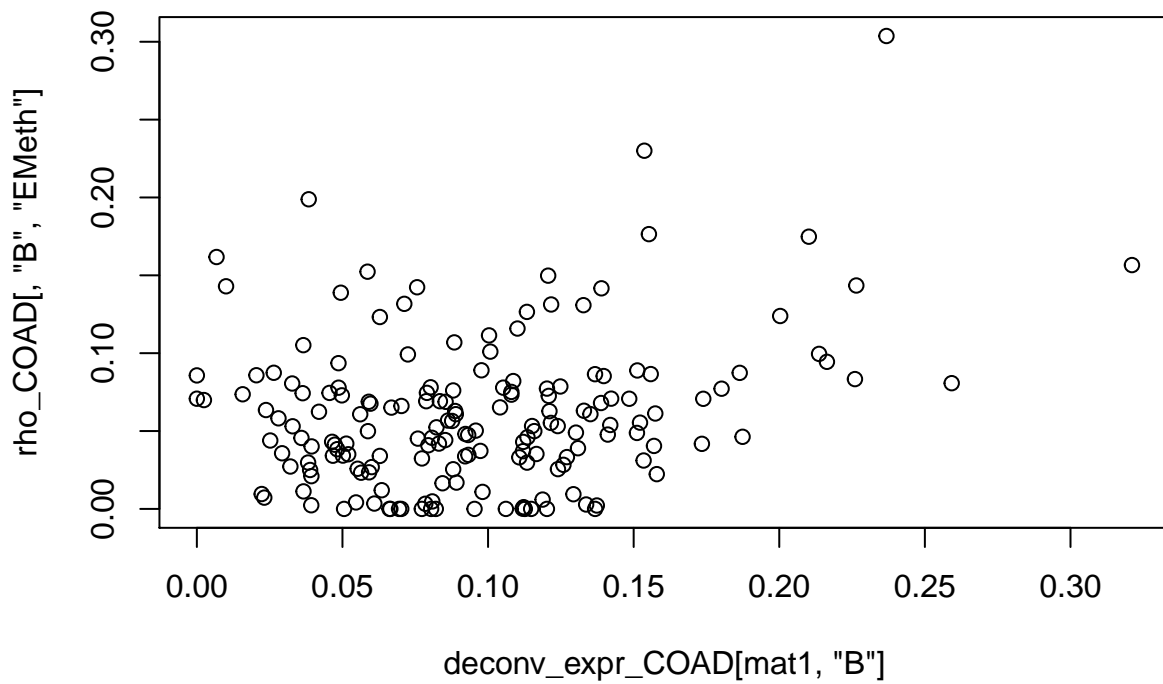
```
mat1 = match(subsample.coad$patient_id, rownames(deconv_expr_COAD))

cor(deconv_expr_COAD[mat1,"B"], rho_COAD[,"B","EMeth"])
```

```
## [1] 0.2677583
```

```
plot(deconv_expr_COAD[mat1,"B"], rho_COAD[,"B","EMeth"])
```



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

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

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

```r
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 pathologic_stage
## 1                                 <NA> Colon Adenocarcinoma         Stage I
## 2                                 <NA> Colon Adenocarcinoma      Stage IIIB
##   gleason_score psa_level hpv_status_by_ish_testing hpv_status_by_p16_testing
## 1          <NA>      <NA>                        NA                        NA
## 2          <NA>      <NA>                        NA                        NA
##   prior_glioma history_of_neoadjuvant_treatment
## 1           NA                               No
## 2           NA                               No
##   breast_carcinoma_estrogen_receptor_status
## 1                                        NA
## 2                                        NA
##   breast_carcinoma_progesterone_receptor_status neoplasm_histologic_grade
## 1                                            NA                        NA
## 2                                            NA                        NA
##   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
```

```r
# mat1 = match(rownames(deconv_expr_COAD), subsample.coad$patient_id)
# subsample.coad = subsample.coad[mat1,]

colnames(deconv_expr_COAD) = paste0(colnames(deconv_expr_COAD), ".E")
ct = cbind(subsample.coad, deconv_expr_COAD, rho_COAD[,,"EMeth"])
dim(ct)
```

```
## [1] 175  17
```

```r
ct[1:2,]
```

```
##      num patient_id      barcode      CD4T.E      CD8T.E Monocyte.E        B.E
## 5656   1       2671 TCGA-A6-2671 0.10297062 0.02254813 0.03457699 0.0591461
## 6781   2       2675 TCGA-A6-2675 0.06966723 0.14954229 0.12023085 0.0921302
##          NK.E Neutrophil.E     Treg.E        CD4T         CD8T    Monocyte
## 5656 0.02676888  0.008814001 0.01517529 0.04522374 9.513481e-20 0.06179749
## 6781 0.06659492  0.057396469 0.02443803 0.17210686 9.463496e-02 0.06969726
##            B          NK Neutrophil        Treg
```

```
## 5656 0.02972052 -1.703507e-19 0.05421388 0.07904436
## 6781 0.07070684  3.419027e-02 0.03432683 0.10433699
```

```r
rownames(ct) = 1:nrow(ct)

table(ct$barcode %in% sc$barcode)
```

```
##
## FALSE   TRUE
##     5    170
```

```r
ct = merge(ct, sc, by="barcode")
dim(ct)
```

```
## [1] 170  54
```

```r
ct[1:2,]
```

```
##        barcode num patient_id      CD4T.E      CD8T.E Monocyte.E       B.E
## 1 TCGA-A6-2671   1       2671 0.10297062 0.02254813 0.03457699 0.0591461
## 2 TCGA-A6-2675   2       2675 0.06966723 0.14954229 0.12023085 0.0921302
##         NK.E Neutrophil.E      Treg.E        CD4T         CD8T   Monocyte
## 1 0.02676888  0.008814001 0.01517529 0.04522374 9.513481e-20 0.06179749
## 2 0.06659492  0.057396469 0.02443803 0.17210686 9.463496e-02 0.06969726
##          B           NK Neutrophil       Treg SMASH_S_hg38 SMASH_oE_hg38
## 1 0.02972052 -1.703507e-19 0.05421388 0.07904436            3          1.06
## 2 0.07070684  3.419027e-02 0.03432683 0.10433699            2          0.68
##   SMASH_wE_hg38 SMASH_oNE_hg38 SMASH_wNE_hg38 AscatPurity AscatPloidy
## 1     0.9274599      0.5531430      0.4846527        0.68    3.609337
## 2     0.7213201      0.6600493      0.6585205        0.49    3.351309
##   raw_MB_hg38_SNV raw_MB_hg38_INDEL IDH_CNV_status_hg38 tCN_burden
## 1              49                 7       IDH wild type   2.228662
## 2              72                 2       IDH wild type   1.572784
##   tCN_burden_ap tumor_type.x num_clonal num_subclonal prop_clonal gender age
## 1     1.0712969         COAD         39            10   0.7959184   MALE  85
## 2     0.8041944         COAD         46            26   0.6388889   MALE  78
##   tumor_type.y initial_pathologic_diagnosis_method    histological_type
## 1        Colon                                <NA> Colon Adenocarcinoma
## 2        Colon                                <NA> Colon Adenocarcinoma
##   pathologic_stage gleason_score psa_level hpv_status_by_ish_testing
## 1        Stage IV          <NA>      <NA>                        NA
## 2        Stage IIA         <NA>      <NA>                        NA
##   hpv_status_by_p16_testing prior_glioma history_of_neoadjuvant_treatment
## 1                        NA           NA                               No
## 2                        NA           NA                               No
##   breast_carcinoma_estrogen_receptor_status
## 1                                        NA
## 2                                        NA
##   breast_carcinoma_progesterone_receptor_status neoplasm_histologic_grade
## 1                                            NA                        NA
## 2                                            NA                        NA
##   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
```
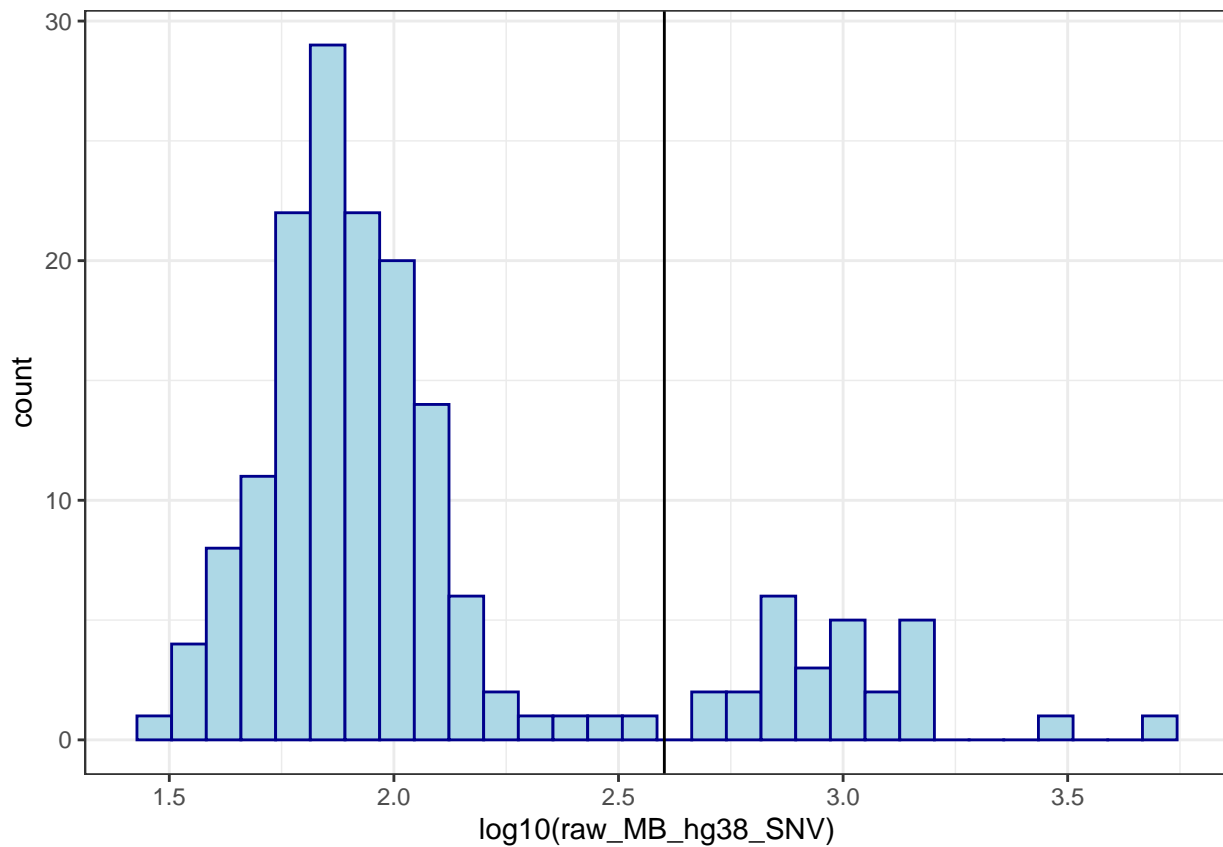
```
## 2                   NO
```

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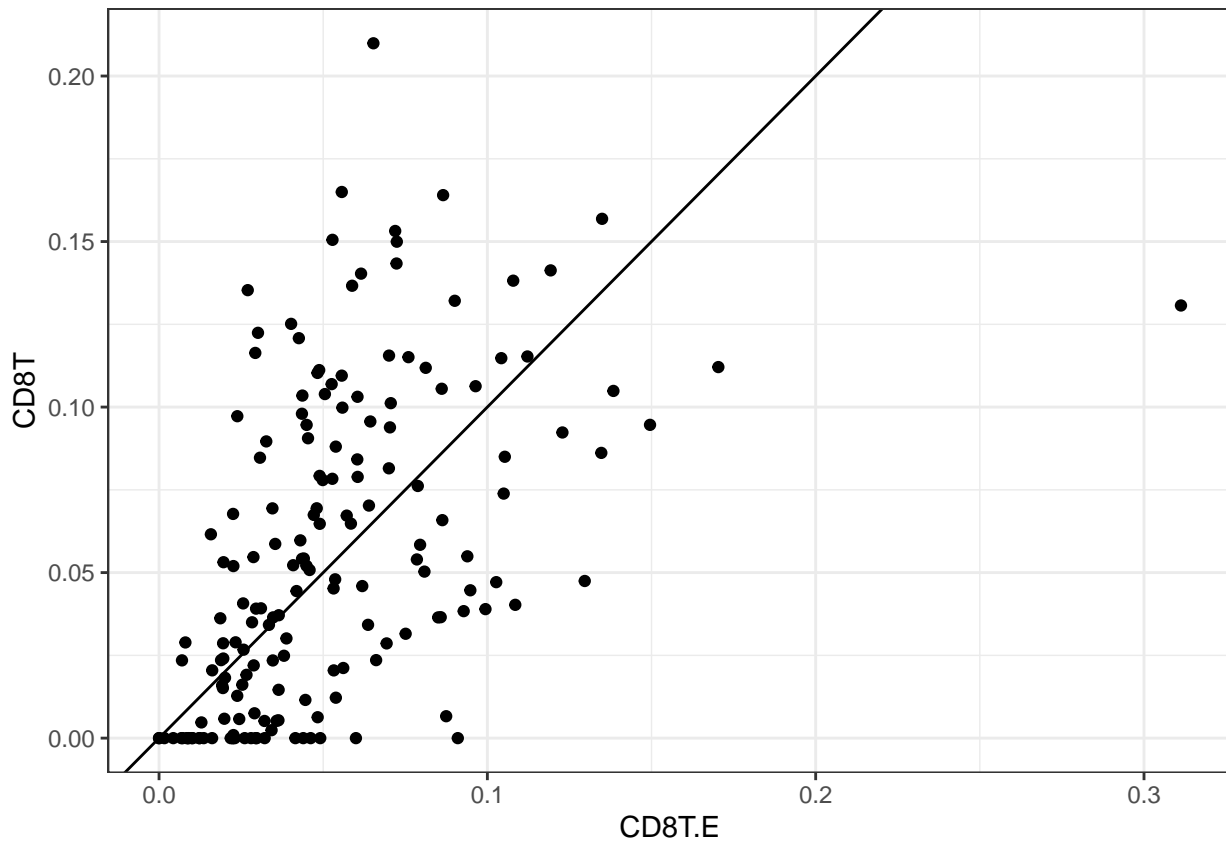


```r
ggplot(ct, aes(x=log10(raw_MB_hg38_SNV)))+
  geom_histogram(color="darkblue", fill="lightblue")  +
  geom_vline(xintercept = log10(400))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

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

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

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

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

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

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