

## CNV-associated DEGs

```
library(infercnv)
library(plyr)
library(reshape2)

label <- read.delim("../merged.label.txt", header = F, row.names = 1)
label_mal <- subset(label, !V2 %in% c("NK", "CD8"))

ptlabel <- read.delim("../res_various/labels.txt", row.names = 1) # an output from the clustering
ptlabel <- ptlabel[, c(1:6, 10)]; head(ptlabel)
```

##	nCount_RNA	nFeature_RNA	Library
## AAACCTGCATACGCCG-1-PM-PS-0001-T-A1	35998	4823	PM-PS-0001-T-A1
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1	31383	5252	PM-PS-0001-T-A1
## AAACCTGTCCTTGCA-1-PM-PS-0001-T-A1	7302	1713	PM-PS-0001-T-A1
## AAACGGGAGGGAAACA-1-PM-PS-0001-T-A1	3759	1233	PM-PS-0001-T-A1
## AAACGGGTATAGGTA-1-PM-PS-0001-T-A1	23097	3874	PM-PS-0001-T-A1
## AAAGATGAGGCCGAAT-1-PM-PS-0001-T-A1	14860	3282	PM-PS-0001-T-A1

##	Patient	Sample	Cell_subtype	RNA_snn_res.0.3
## AAACCTGCATACGCCG-1-PM-PS-0001-T-A1	SMC01	SMC01-T	CMS2	2
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1	SMC01	SMC01-T	CMS2	0
## AAACCTGTCCTTGCA-1-PM-PS-0001-T-A1	SMC01	SMC01-T	CMS2	2
## AAACGGGAGGGAAACA-1-PM-PS-0001-T-A1	SMC01	SMC01-T	CMS2	1
## AAACGGGTATAGGTA-1-PM-PS-0001-T-A1	SMC01	SMC01-T	CMS2	1
## AAAGATGAGGCCGAAT-1-PM-PS-0001-T-A1	SMC01	SMC01-T	CMS2	2

```
label_mal$Patient <- ptlabel$Patient
```

```
### HMM states
# State 1: 0x: complete loss
# State 2: 0.5x: loss of one copy
# State 3: 1x: neutral
# State 4: 1.5x: addition of one copy
# State 5: 2x: addition of two copies
# State 6: 3x: essentially a placeholder for >2x copies but modeled as 3x
hmmres <- readRDS("17_HMM_predHMMi6.hmm_mode-samples.infercnv_obj")
hmmstates <- hmmres@expr.data[, rownames(label_mal)]
remove(hmmres)
usegenes <- rownames(hmmstates)
hmmstates[1:4, 1:4]
```

##	AAACCTGCATACGCCG-1-PM-PS-0001-T-A1
## MTND1P23-ENSG00000225972.1-4	3
## MTND2P28-ENSG00000225630.1-4	3
## MTC01P12-ENSG00000237973.1-5	3

```
## MTC03P12-ENSG00000198744.5-5 3
## AAACCTGGTCGCATAT-1-PM-PS-0001-T-A1
## MTND1P23-ENSG00000225972.1-4 3
## MTND2P28-ENSG00000225630.1-4 3
## MTC01P12-ENSG00000237973.1-5 3
## MTC03P12-ENSG00000198744.5-5 3
## AAACCTGTCCCTTGCA-1-PM-PS-0001-T-A1
## MTND1P23-ENSG00000225972.1-4 3
## MTND2P28-ENSG00000225630.1-4 3
## MTC01P12-ENSG00000237973.1-5 3
## MTC03P12-ENSG00000198744.5-5 3
## AACCGGAGGGAAACA-1-PM-PS-0001-T-A1
## MTND1P23-ENSG00000225972.1-4 3
## MTND2P28-ENSG00000225630.1-4 3
## MTC01P12-ENSG00000237973.1-5 3
## MTC03P12-ENSG00000198744.5-5 3
```

### ### PT markers w/ CNV profile

```
ptmarkers <- read.delim("../res_various/res_0.6/ptmarkers/markers.MAST.txt")
head(ptmarkers, n = 3) # 3287 unique genes
```

```
## p_val avg_logFC pct.1 pct.2 p_val_adj cluster gene
## 1 0 1.178095 0.835 0.454 0 SMC01 PRSS23-ENSG00000150687.12-6
## 2 0 1.109789 0.876 0.761 0 SMC01 STRAP-ENSG00000023734.11-5
## 3 0 1.040737 0.922 0.196 0 SMC01 AL450405.1-ENSG00000230202.1-6
## Symbol
## 1 PRSS23
## 2 STRAP
## 3 AL450405.1
```

```
inferredGenes <- intersect(usegenes, as.character(ptmarkers$gene))
length(inferredGenes) # 3027
```

```
## [1] 3027
```

```
hmmstates_markers <- as.data.frame(hmmstates[inferredGenes, ])
dim(hmmstates_markers) # 3027 17334
```

```
## [1] 3027 17334
```

```
ord <- read.delim('../gencode.v34lift37.annotation.geneorder.txt', header = F, row.names = 1)
ord <- ord[usegenes, ]; head(ord)
```

```
## V2 V3 V4
## MTND1P23-ENSG00000225972.1-4 chr1 564442 564813
## MTND2P28-ENSG00000225630.1-4 chr1 565020 566063
## MTC01P12-ENSG00000237973.1-5 chr1 566454 567996
## MTC03P12-ENSG00000198744.5-5 chr1 569756 570302
## LINC01128-ENSG00000228794.10-7 chr1 760518 794826
## NOC2L-ENSG00000188976.11-5 chr1 879583 894689
```

```
hmmstates_markers <- cbind(Chr = as.character(ord[inferredGenes, "V2"]), hmmstates_markers)
hmmstates_markers[1:4, 1:4]
```

```
##                               Chr AACCTGCATACGCCG-1-PM-PS-0001-T-A1
## MTND1P23-ENSG00000225972.1-4 chr1                                     3
## MTND2P28-ENSG00000225630.1-4 chr1                                     3
## MTC01P12-ENSG00000237973.1-5 chr1                                     3
## HES4-ENSG00000188290.11-5   chr1                                     3
##                               AACCTGGTCGCATAT-1-PM-PS-0001-T-A1
## MTND1P23-ENSG00000225972.1-4                                     3
## MTND2P28-ENSG00000225630.1-4                                     3
## MTC01P12-ENSG00000237973.1-5                                     3
## HES4-ENSG00000188290.11-5   3
##                               AACCTGTCCCTTGCA-1-PM-PS-0001-T-A1
## MTND1P23-ENSG00000225972.1-4                                     3
## MTND2P28-ENSG00000225630.1-4                                     3
## MTC01P12-ENSG00000237973.1-5                                     3
## HES4-ENSG00000188290.11-5   3
```

```
#saveRDS(hmmstates_markers, "hmmstates.markers.Rds")
```

```
### CNV scores
```

```
cnv_degs <- data.frame( matrix(nrow = length(inferredGenes), ncol = length(levels(label_mal$Patient)))
rownames(cnv_degs) <- inferredGenes
colnames(cnv_degs) <- levels(label_mal$Patient)
dim(cnv_degs)
```

```
## [1] 3027 22
```

```
for (pt in colnames(cnv_degs)) {
  pt_bcs <- rownames(subset(label_mal, Patient == pt))

  pt_degs <- as.character(subset(ptmarkers, cluster == pt)$gene)
  pt_degs <- intersect(inferredGenes, pt_degs)

  for (deg in pt_degs) {
    summ <- summary( as.factor(t(hmmstates_markers[deg, pt_bcs])) )/length(pt_bcs)*100
    cnv_degs[deg, pt] <- mean(as.numeric(names(summ[summ == max(summ)])))
  }
}

cnv_degs <- cbind(Chr = hmmstates_markers$Chr, cnv_degs)
cnv_degs$Chr <- factor(cnv_degs$Chr, levels = unique(cnv_degs$Chr))
head(cnv_degs)
```

```
##                               Chr SMC01 SMC02 SMC03 SMC04 SMC06 SMC07 SMC08
## MTND1P23-ENSG00000225972.1-4 chr1      NA      NA      NA      NA      NA      NA
## MTND2P28-ENSG00000225630.1-4 chr1      NA      NA      NA      NA      NA      NA
## MTC01P12-ENSG00000237973.1-5 chr1      NA      NA      NA      NA      NA      NA
## HES4-ENSG00000188290.11-5   chr1      NA      NA      NA      NA      NA      NA
## ISG15-ENSG00000187608.10-8  chr1      NA      3      3      3      NA      NA
```

```
## AGRN-ENSG00000188157.15-6 chr1 NA NA NA NA NA NA NA
## SMC09 SMC10 SMC11 SMC14 SMC15 SMC16 SMC17 SMC18
## MTND1P23-ENSG00000225972.1-4 3 NA NA NA NA NA NA NA
## MTND2P28-ENSG00000225630.1-4 3 NA NA NA NA NA NA NA
## MTC01P12-ENSG00000237973.1-5 NA NA NA NA NA NA 3 NA
## HES4-ENSG00000188290.11-5 NA 3 NA 3 NA 3 NA NA
## ISG15-ENSG00000187608.10-8 NA 3 NA 3 NA NA NA NA
## AGRN-ENSG00000188157.15-6 NA NA NA NA NA NA 3 NA
## SMC19 SMC20 SMC21 SMC22 SMC23 SMC24 SMC25
## MTND1P23-ENSG00000225972.1-4 NA NA NA NA NA NA NA
## MTND2P28-ENSG00000225630.1-4 3 NA NA NA NA NA NA
## MTC01P12-ENSG00000237973.1-5 NA NA NA NA NA NA NA
## HES4-ENSG00000188290.11-5 NA NA NA NA NA NA NA
## ISG15-ENSG00000187608.10-8 NA NA NA NA 3 NA NA
## AGRN-ENSG00000188157.15-6 NA 3 NA NA NA NA NA
```

```
#saveRDS(cnv_degs, "hmmstates.markers.CNVstates.Rds")
```

```
### plot DEGs per chromosome
```

```
library(ggplot2)
```

```
dir.create("degs_hmmStates")
```

```
for (pt in colnames(cnv_degs)[-1]) {
  tmp_cnv_degs <- na.omit(cnv_degs[, c("Chr", pt)])
  tmp_cnv_degs[, pt] <- mapvalues(tmp_cnv_degs[, pt], from = c(1,2,3,4,5,6), to = c("0x", "0.5x", "1x",
  tmp_cnv_degs[, pt] <- factor(tmp_cnv_degs[, pt], levels = c("0x", "0.5x", "1x", "1.5x", "2x", "3x"))

  ggplot(tmp_cnv_degs, aes(tmp_cnv_degs[, pt])) +
    geom_histogram(stat = "count") +
    geom_vline(xintercept = "1x") +
    labs(x = pt, y = "DEG counts") +
    facet_wrap(~Chr, ncol = 5) +
    theme_bw() +
    theme(axis.ticks = element_line(color = 'black'),
          axis.text = element_text(color = 'black'))
  ggsave(paste(c('degs_hmmStates/', pt, '.chr.pdf'), collapse = ''), units = "cm", width = 14, height =

  ggplot(tmp_cnv_degs, aes(tmp_cnv_degs[, pt])) +
    geom_histogram(stat = "count") +
    geom_vline(xintercept = "1x") +
    labs(x = pt, y = "DEG counts") +
    theme_bw() +
    theme(axis.ticks = element_line(color = 'black'),
          axis.text = element_text(color = 'black'))
  ggsave(paste(c('degs_hmmStates/', pt, '.pdf'), collapse = ''), units = "cm", width = 5, height = 5)
}
```

```
## The following 'from' values were not present in 'x': 1, 6
```

```
## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
```

```
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
```

```
## The following 'from' values were not present in 'x': 1, 6
```

```

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 5, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 5, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

```

```

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 5, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 5, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

## The following 'from' values were not present in 'x': 1, 6

## Warning in geom_histogram(stat = "count"): Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'
## Ignoring unknown parameters: 'binwidth', 'bins', and 'pad'

```

### ### Proportions of CNV associated DEGs

```
cnv_degs_summary <- data.frame( matrix(nrow = 3, ncol = length(levels(label_mal$Patient))) )
rownames(cnv_degs_summary) <- c("Gain", "Loss", "Neutral")
colnames(cnv_degs_summary) <- levels(label_mal$Patient)
cnv_degs_summary
```

```
##          SMC01 SMC02 SMC03 SMC04 SMC06 SMC07 SMC08 SMC09 SMC10 SMC11 SMC14 SMC15
## Gain      NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## Loss      NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## Neutral   NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##          SMC16 SMC17 SMC18 SMC19 SMC20 SMC21 SMC22 SMC23 SMC24 SMC25
## Gain      NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## Loss      NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
## Neutral   NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
```

```
for (pt in colnames(cnv_degs)[-1]) {
  tmp_cnv_degs <- na.omit(cnv_degs[, c("Chr", pt)])

  tmp_cnv_degs$cnv <- "Neutral"
  tmp_cnv_degs[rownames( subset(tmp_cnv_degs, tmp_cnv_degs[, pt] > 3) ), "cnv"] <- "Gain"
  tmp_cnv_degs[rownames( subset(tmp_cnv_degs, tmp_cnv_degs[, pt] < 3) ), "cnv"] <- "Loss"
  tmp_cnv_degs$cnv <- factor(tmp_cnv_degs$cnv, levels = c("Gain", "Loss", "Neutral"))

  print (pt)
  print (summary(as.factor(tmp_cnv_degs$cnv))/nrow(tmp_cnv_degs) * 100)
  cnv_degs_summary[, pt] <- summary(as.factor(tmp_cnv_degs$cnv))/nrow(tmp_cnv_degs) * 100
}
```

```
## [1] "SMC01"
##      Gain      Loss  Neutral
## 17.00508 12.18274 70.81218
## [1] "SMC02"
##      Gain      Loss  Neutral
## 53.191489  7.801418 39.007092
## [1] "SMC03"
##      Gain      Loss  Neutral
## 19.28687 11.66937 69.04376
## [1] "SMC04"
##      Gain      Loss  Neutral
## 23.07692 10.68376 66.23932
## [1] "SMC06"
##      Gain      Loss  Neutral
## 10.64516 17.09677 72.25806
## [1] "SMC07"
##      Gain      Loss  Neutral
## 25.33333 10.00000 64.66667
## [1] "SMC08"
##      Gain      Loss  Neutral
## 33.171913  6.779661 60.048426
## [1] "SMC09"
##      Gain      Loss  Neutral
## 39.534884  6.511628 53.953488
## [1] "SMC10"
```

```
##      Gain      Loss  Neutral
## 18.61472  9.52381 71.86147
## [1] "SMC11"
##      Gain      Loss  Neutral
## 34.88372 10.46512 54.65116
## [1] "SMC14"
##      Gain      Loss  Neutral
## 30.612245  6.377551 63.010204
## [1] "SMC15"
##      Gain      Loss  Neutral
## 20.96386 12.04819 66.98795
## [1] "SMC16"
##      Gain      Loss  Neutral
## 30.769231  8.846154 60.384615
## [1] "SMC17"
##      Gain      Loss  Neutral
## 28.486647  7.121662 64.391691
## [1] "SMC18"
##      Gain      Loss  Neutral
## 41.052632  5.789474 53.157895
## [1] "SMC19"
##      Gain      Loss  Neutral
## 22.01835 11.00917 66.97248
## [1] "SMC20"
##      Gain      Loss  Neutral
## 26.929134  8.661417 64.409449
## [1] "SMC21"
##      Gain      Loss  Neutral
## 31.79191 11.56069 56.64740
## [1] "SMC22"
##      Gain      Loss  Neutral
## 22.674419  8.139535 69.186047
## [1] "SMC23"
##      Gain      Loss  Neutral
## 29.80132  3.97351 66.22517
## [1] "SMC24"
##      Gain      Loss  Neutral
## 16.10942 14.89362 68.99696
## [1] "SMC25"
##      Gain      Loss  Neutral
## 28.717949  8.717949 62.564103
```

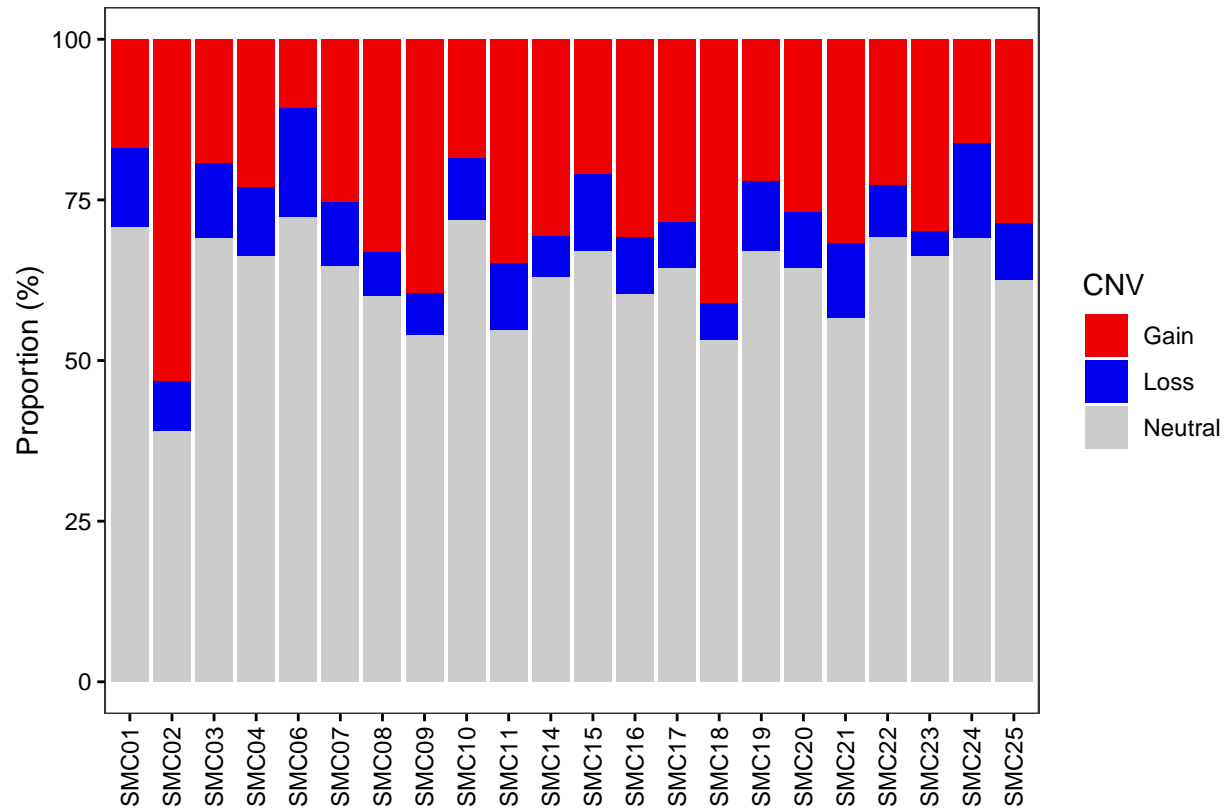
```
cnv_degs_summary <- cbind(CNV = rownames(cnv_degs_summary), cnv_degs_summary)
cnv_degs_summary_m <- melt(cnv_degs_summary)
```

```
## Using CNV as id variables
```

```
ggplot(cnv_degs_summary_m, aes(variable, value, fill = CNV)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c("red2", "blue2", "grey80")) +
  labs(x = "", y = "Proportion (%)") +
  theme_bw() +
  theme(panel.grid = element_blank(),
```



```
axis.ticks = element_line(color = 'black'),
axis.text = element_text(color = 'black'),
axis.text.x = element_text(angle = 90, hjust = 1, vjust = .5))
```



```
ggsave("deg_hmmStates/Proportion.pdf", units = "cm", width = 12, height = 5)
```