

Covadonga-et-al-2022

f.u.

March 2022”

Algorithms and data description for Cell paper

Reference: Covadonga et al. ...XXXX

Main data structures

Here we load and describe the main data structures used in the rest of the document.

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.1.2
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(tidyr)
diffCounts <- read.csv("data/diffCounts.csv",
                      stringsAsFactors=TRUE)
diffCounts$percHuC <- 100*diffCounts$Huc / (diffCounts$Huc + diffCounts$noHuC)
summary(diffCounts)

##           id      bound      time      Huc      noHuC
## emb1e1:8  r2r3:10  30hpf:20  Min.   :  0.00  Min.   : 36.0
## emb2e5:8  r3r4:10  48hpf:20  1st Qu.:  0.00  1st Qu.: 59.0
## emb5e3:8  r4r5:10           Median : 16.50  Median : 86.5
## emb5e4:8  r5r6:10           Mean   : 34.23  Mean   : 86.4
## emb5e5:8           3rd Qu.: 69.50  3rd Qu.:108.0
##           Max.   :119.00  Max.   :171.0
##
##      percHuC
## Min.   :  0.00
## 1st Qu.:  0.00
## Median :17.18
## Mean   :21.36
## 3rd Qu.:41.84
## Max.   :48.82

diffCounts
```

##	id	bound	time	Huc	noHuC	percHuC
## 1	emb1e1	r2r3	30hpf	0	44	0.000000
## 2	emb2e5	r2r3	30hpf	0	45	0.000000
## 3	emb5e3	r2r3	30hpf	0	97	0.000000
## 4	emb5e4	r2r3	30hpf	4	149	2.614379
## 5	emb5e5	r2r3	30hpf	0	53	0.000000
## 6	emb1e1	r3r4	30hpf	0	37	0.000000
## 7	emb2e5	r3r4	30hpf	1	36	2.702703
## 8	emb5e3	r3r4	30hpf	0	111	0.000000
## 9	emb5e4	r3r4	30hpf	0	92	0.000000
## 10	emb5e5	r3r4	30hpf	0	39	0.000000
## 11	emb1e1	r4r5	30hpf	0	50	0.000000
## 12	emb2e5	r4r5	30hpf	3	53	5.357143
## 13	emb5e3	r4r5	30hpf	0	107	0.000000
## 14	emb5e4	r4r5	30hpf	0	108	0.000000
## 15	emb5e5	r4r5	30hpf	2	46	4.166667
## 16	emb1e1	r5r6	30hpf	2	86	2.272727
## 17	emb2e5	r5r6	30hpf	2	76	2.564103
## 18	emb5e3	r5r6	30hpf	0	129	0.000000
## 19	emb5e4	r5r6	30hpf	0	137	0.000000
## 20	emb5e5	r5r6	30hpf	2	63	3.076923
## 21	emb1e1	r2r3	48hpf	33	68	32.673267
## 22	emb2e5	r2r3	48hpf	29	71	29.000000
## 23	emb5e3	r2r3	48hpf	58	100	36.708861
## 24	emb5e4	r2r3	48hpf	71	104	40.571429
## 25	emb5e5	r2r3	48hpf	81	96	45.762712
## 26	emb1e1	r3r4	48hpf	52	61	46.017699
## 27	emb2e5	r3r4	48hpf	40	50	44.444444
## 28	emb5e3	r3r4	48hpf	68	108	38.636364
## 29	emb5e4	r3r4	48hpf	87	92	48.603352
## 30	emb5e5	r3r4	48hpf	60	66	47.619048
## 31	emb1e1	r4r5	48hpf	72	84	46.153846
## 32	emb2e5	r4r5	48hpf	83	87	48.823529
## 33	emb5e3	r4r5	48hpf	74	122	37.755102
## 34	emb5e4	r4r5	48hpf	85	122	41.062802
## 35	emb5e5	r4r5	48hpf	66	74	47.142857
## 36	emb1e1	r5r6	48hpf	47	69	40.517241
## 37	emb2e5	r5r6	48hpf	72	91	44.171779
## 38	emb5e3	r5r6	48hpf	87	171	33.720930
## 39	emb5e4	r5r6	48hpf	119	140	45.945946
## 40	emb5e5	r5r6	48hpf	69	122	36.125654

```
percBoundTime <- diffCounts %>% group_by(bound,time) %>%
  summarise(percHuC = mean(percHuC), .groups = 'keep')
percHuC_30 <- percBoundTime[percBoundTime$time=="30hpf", "percHuC"]$percHuC
percHuC_48 <- percBoundTime[percBoundTime$time=="48hpf", "percHuC"]$percHuC
boundaryColors = c("darkgray", "#5D9B90", "#48125E", "#c44b16")

percNotch <- read.csv("data/percNotch.csv", stringsAsFactors = TRUE)
percBoundTimeNotch <- percNotch %>% group_by(bound,time) %>%
  summarise(percNotch = mean(percNotch, na.rm = TRUE), .groups = 'keep')
summary(percNotch)
```

##	id	time	bound	percNotch
##	emb1:8	26hpf:20	r2r3:10	Min. : 0.000

```
## emb2:8 36hpf:20 r3r4:10 1st Qu.: 8.079
## emb3:8 r4r5:10 Median :15.849
## emb4:8 r5r6:10 Mean :31.978
## emb5:8 3rd Qu.:57.514
## Max. :73.786
## NA's :4
```

```
percNotch
```

```
##      id  time bound percNotch
## 1  emb1 26hpf r2r3 14.285714
## 2  emb2 26hpf r2r3 13.793103
## 3  emb3 26hpf r2r3 12.280702
## 4  emb4 26hpf r2r3  2.222222
## 5  emb5 26hpf r2r3  3.508772
## 6  emb1 26hpf r3r4 14.285714
## 7  emb2 26hpf r3r4 15.789474
## 8  emb3 26hpf r3r4 15.909091
## 9  emb4 26hpf r3r4  8.333333
## 10 emb5 26hpf r3r4  8.888889
## 11 emb1 26hpf r4r5 11.764706
## 12 emb2 26hpf r4r5  2.941176
## 13 emb3 26hpf r4r5 11.627907
## 14 emb4 26hpf r4r5  0.000000
## 15 emb5 26hpf r4r5  7.317073
## 16 emb1 26hpf r5r6 18.181818
## 17 emb2 26hpf r5r6  2.127660
## 18 emb3 26hpf r5r6  4.878049
## 19 emb4 26hpf r5r6  6.250000
## 20 emb5 26hpf r5r6  6.896552
## 21 emb1 36hpf r2r3 71.304348
## 22 emb2 36hpf r2r3 65.263158
## 23 emb3 36hpf r2r3 72.413793
## 24 emb4 36hpf r2r3 55.913978
## 25 emb5 36hpf r2r3      NA
## 26 emb1 36hpf r3r4 68.965517
## 27 emb2 36hpf r3r4 56.626506
## 28 emb3 36hpf r3r4 61.016949
## 29 emb4 36hpf r3r4 44.318182
## 30 emb5 36hpf r3r4      NA
## 31 emb1 36hpf r4r5 49.397590
## 32 emb2 36hpf r4r5 56.034483
## 33 emb3 36hpf r4r5 61.157025
## 34 emb4 36hpf r4r5 52.380952
## 35 emb5 36hpf r4r5      NA
## 36 emb1 36hpf r5r6 73.786408
## 37 emb2 36hpf r5r6 60.176991
## 38 emb3 36hpf r5r6 53.658537
## 39 emb4 36hpf r5r6 67.521368
## 40 emb5 36hpf r5r6      NA
```

For description of 2 groups

```

descr1 <- function(gr){
  gr <- na.omit(gr)
  return(list(N=length(gr), M=mean(gr, na.rm=TRUE),
             SD=sd(gr, na.rm = TRUE), SEM=sd(gr, na.rm = TRUE)/sqrt(length(gr))))
}

descr2groups <- function(gr1, gr2, nams=NULL, tit = NULL, not.paired=TRUE){
  # given array of two rows
  tab <- rbind(unlist(descr1(gr1)), unlist(descr1(gr2)))
  if (!is.null(nams)){
    rownames(tab) <- nams}
  pv <- t.test(gr1,gr2, paired= (!not.paired))$p.value

  if (!is.null(tit))
    print(paste(" ", tit))
  print(round(tab,2))
  print(paste(" t.test p-value:", pv, ifelse(pv<0.001, " *** ",
                                             ifelse(pv<0.01, " ** ",
                                             ifelse(pv<0.05, " * ", "ns")))))
}

descr2cols <- function(df, not.paired=TRUE, nams = NULL, tit = NULL){
  # given a 2 column data.frame or tibble
  descr2groups(unlist(df[,1]), unlist(df[,2]),
               not.paired = not.paired, nams = nams, tit = tit)
}

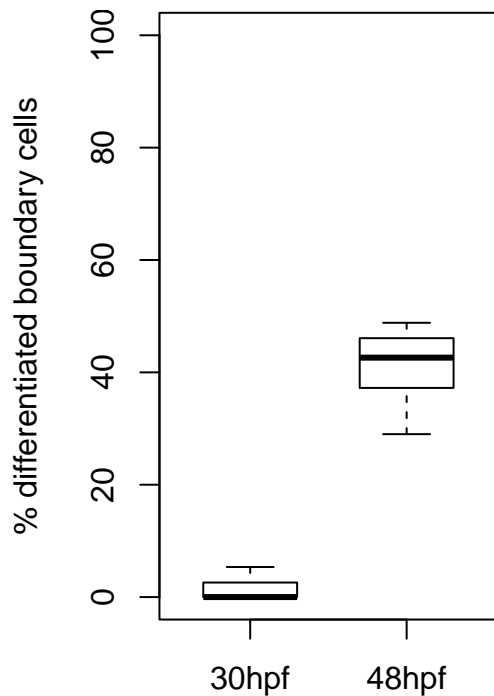
```

Figure 2C

Boxplot diff 30-48 hpf

```
b1 <- diffCounts[diffCounts$time=="30hpf","percHuC"]
b2 <- diffCounts[diffCounts$time=="48hpf","percHuC"]

boxplot(b1,b2, main = "", xlab = "", ylab = "% differentiated boundary cells",
        ylim=c(0,100),
        names = c("30hpf", "48hpf"), col = "white",
        boxwex=0.6)
```



```
descr2groups(b1,b2, nams=c("30hpf", "48hpf"))
```

```
##      N      M    SD  SEM
## 30hpf 20  1.14 1.71 0.38
## 48hpf 20 41.57 5.80 1.30
## [1] " t.test p-value: 1.79453265173272e-19 *** "
```

Figure 2E

Lineage tree

Figure 3B

Lineage 2

Figure 3E

number of cells along time

Figure 3H

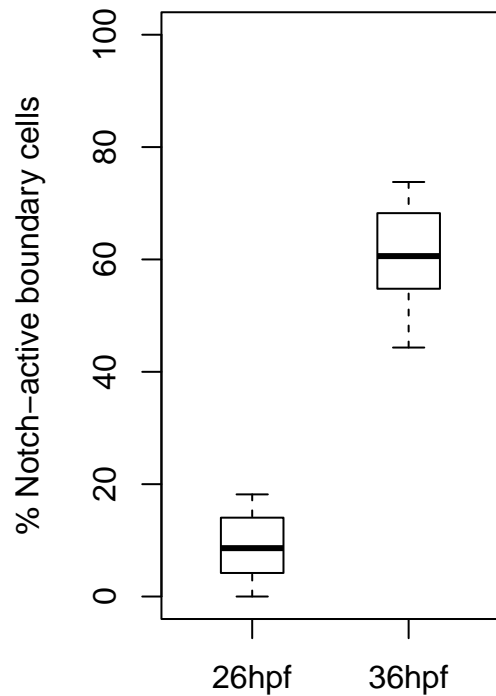
cell distance to ventr

Figure 4C

bp notch active 26-36

```
b1 <- percNotch$percNotch[percNotch$time=="26hpf"]
b2 <- percNotch$percNotch[percNotch$time=="36hpf"]

boxplot(b1, b2,
        ylim=c(0,100),
        main = "", xlab = "",
        ylab = "% Notch-active boundary cells",
        names = c("26hpf", "36hpf"), col = "white",
        boxwex=0.4, outpch='.', cex=2)
```



```
descr2groups(b1,b2, nams=c("26hpf", "36hpf"))
```

```
##          N      M   SD  SEM
## 26hpf  20   9.06 5.41 1.21
## 36hpf  16  60.62 8.69 2.17
## [1] " t.test p-value: 8.60491568290165e-17 *** "
```

Figure 4 I-J

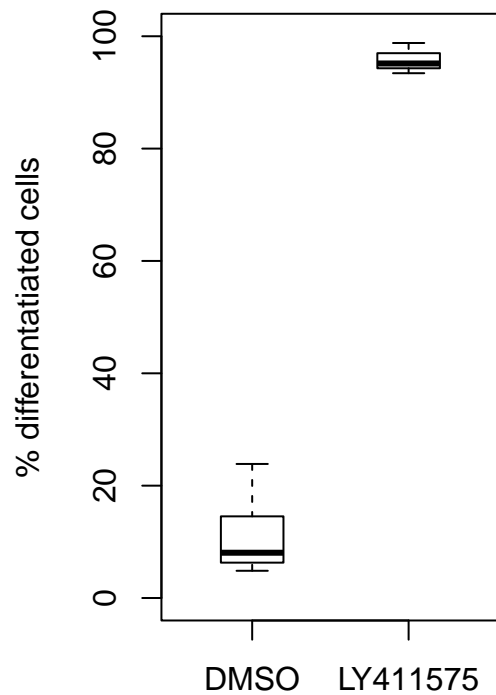
```
DMSO <- data.frame(  
  HuC = c(7L, 24L, 13L, 21L, 8L, 10L, 8L, 11L),  
  Total = c(113L, 132L, 151L, 88L, 125L, 133L, 165L, 101L))  
DMSO$percHuC <- DMSO$HuC/DMSO$Total*100  
DMSO
```

```
##   HuC Total  percHuC  
## 1    7   113  6.194690  
## 2   24   132 18.181818  
## 3   13   151  8.609272  
## 4   21    88 23.863636  
## 5    8   125  6.400000  
## 6   10   133  7.518797  
## 7    8   165  4.848485  
## 8   11   101 10.891089
```

```
LY <- data.frame(  
  HuC = c(71L, 122L, 82L, 99L, 79L, 142L, 104L, 112L),  
  Total = c(74L, 129L, 83L, 101L, 84L, 152L, 110L, 117L))  
LY$percHuC <- LY$HuC/LY$Total*100  
LY
```

```
##   HuC Total  percHuC  
## 1   71    74 95.94595  
## 2  122   129 94.57364  
## 3   82    83 98.79518  
## 4   99   101 98.01980  
## 5   79    84 94.04762  
## 6  142   152 93.42105  
## 7  104   110 94.54545  
## 8  112   117 95.72650
```

```
b1 <- DMSO$percHuC  
b2 <- DMSO$Total  
b3 <- LY$percHuC  
b4 <- LY$Total  
  
# fill colors  
rellcols <- c("white", "white")  
  
boxplot(b1,b3,  
  ylim = c(0,100),  
  ylab="% differentiated cells",  
  names = c("DMSO", "LY411575"),  
  col = rellcols,  
  boxwex=0.4, outpch='.', cex=2)
```

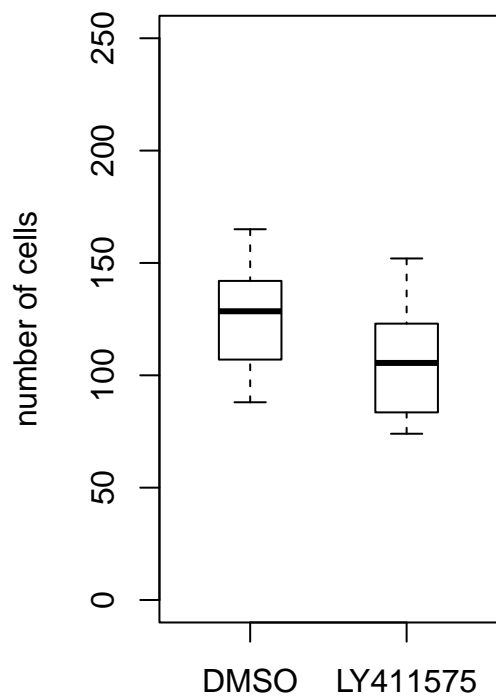


```
descr2groups(b1,b3, nams=c("DMSO", "LY411575"), not.paired = TRUE)
```

```
##          N      M   SD  SEM
## DMSO      8 10.81 6.73 2.38
## LY411575  8 95.63 1.91 0.67
## [1] " t.test p-value: 4.44043249694964e-10 *** "
```

```
maxRight <- 250
```

```
boxplot(b2, b4,
        ylim = c(0, maxRight),
        ylab="number of cells",
        names = c("DMSO", "LY411575"),
        col = relcols,
        boxwex=0.4, outpch='.', cex=2)
```

```
descr2groups(b2,b4, nams=c("DMSO", "LY411575"), not.paired = TRUE)
```

```
##          N      M   SD  SEM
## DMSO      8 126.00 25.27 8.93
## LY411575  8 106.25 26.30 9.30
## [1] " t.test p-value: 0.147972179934964 ns"
```

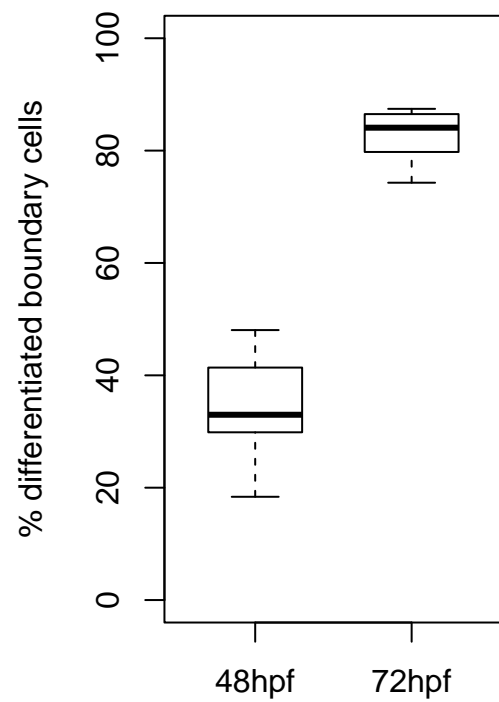
Figure 7C

```
diff4872 <- read.csv("data/diff4872.csv", stringsAsFactors=TRUE)
diff4872
```

##	emb	time	bound	percHuC	nCells
## 1	emb1e6	48hpf	r4r5	29.86111	144
## 2	emb2e3	48hpf	r4r5	48.05195	154
## 3	emb2e1	48hpf	r4r5	41.37931	174
## 4	emb3e7	48hpf	r4r5	35.77236	123
## 5	emb3e1	48hpf	r4r5	30.07519	133
## 6	emb3e6	72hpf	r4r5	86.50000	200
## 7	emb3e5	72hpf	r4r5	81.76101	159
## 8	emb2e1	72hpf	r4r5	75.00000	124
## 9	emb1e8	72hpf	r4r5	87.43169	183
## 10	emb1e2	72hpf	r4r5	74.28571	175
## 11	emb1e6	48hpf	r5r6	18.38235	136
## 12	emb2e3	48hpf	r5r6	30.18868	212
## 13	emb2e1	48hpf	r5r6	39.59732	149
## 14	emb3e7	48hpf	r5r6	47.28682	129
## 15	emb3e1	48hpf	r5r6	28.57143	154
## 16	emb3e6	72hpf	r5r6	79.75709	247
## 17	emb3e5	72hpf	r5r6	85.71429	154
## 18	emb2e1	72hpf	r5r6	82.44275	131
## 19	emb1e8	72hpf	r5r6	85.97561	164
## 20	emb1e2	72hpf	r5r6	87.09677	248

```
b1 <- diff4872$percHuC[diff4872$time=="48hpf"]
b2 <- diff4872$percHuC[diff4872$time=="72hpf"]

boxplot(b1, b2, ylim=c(0,100),
        main = "", xlab = "",
        ylab = "% differentiated boundary cells",
        names = c("48hpf", "72hpf"), col = "white",
        boxwex=0.6, outpch='.', cex=2)
```



```
descr2groups(b1,b2, nams=c("48hpf", "72hpf"), not.paired = TRUE)
```

```
##          N      M   SD  SEM
## 48hpf 10 34.92 9.28 2.93
## 72hpf 10 82.60 4.89 1.55
## [1] " t.test p-value: 1.25029141236881e-09 *** "
```

Figure Sup3 A-B

bp diff vs bound i 30-48

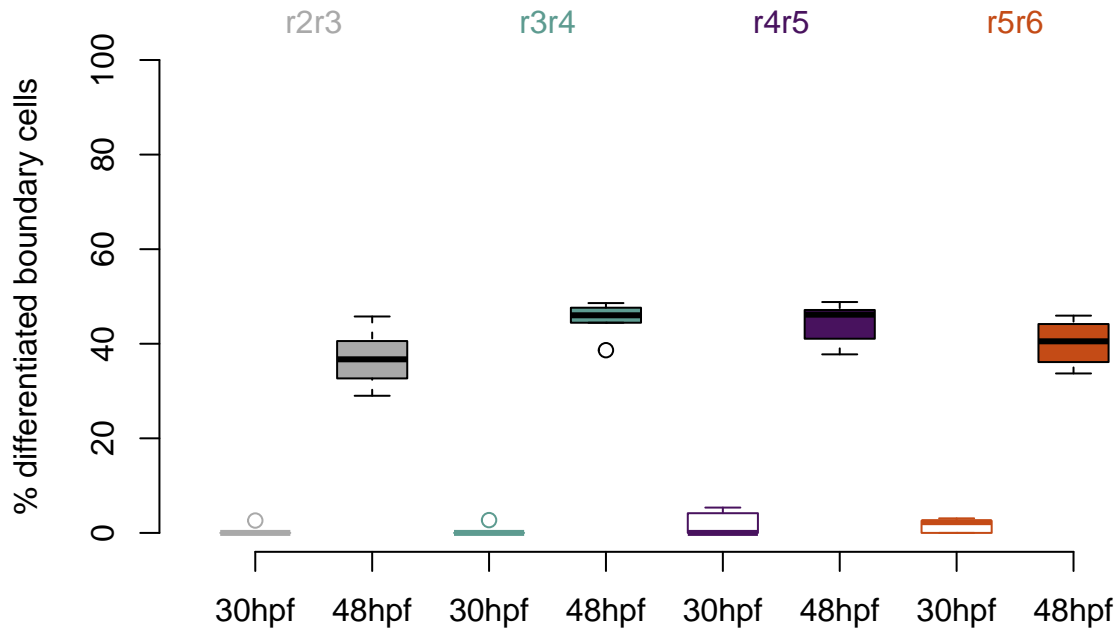
```
diffCounts$percHuC <- 100*diffCounts$Huc/(diffCounts$Huc+diffCounts$noHuC)

b1<- diffCounts %>% filter(time=="30hpf", bound=="r2r3")
b2 <- diffCounts %>% filter(time=="48hpf", bound=="r2r3")
b3<- diffCounts %>% filter(time=="30hpf", bound=="r3r4")
b4 <- diffCounts %>% filter(time=="48hpf", bound=="r3r4")
b5<- diffCounts %>% filter(time=="30hpf", bound=="r4r5")
b6 <- diffCounts %>% filter(time=="48hpf", bound=="r4r5")
b7<- diffCounts %>% filter(time=="30hpf", bound=="r5r6")
b8 <- diffCounts %>% filter(time=="48hpf", bound=="r5r6")

opar<-par(no.readonly = TRUE)
par(bty="n")

boundaryColors = c("darkgray", "#5D9B90", "#48125E", "#c44b16")
altcols <- c(boundaryColors[1], "black", boundaryColors[2],
             "black", boundaryColors[3], "black", boundaryColors[4], "black" )

boxplot(b1$percHuC, b2$percHuC, b3$percHuC, b4$percHuC,
        b5$percHuC, b6$percHuC, b7$percHuC, b8$percHuC,
        ylab = "% differentiated boundary cells",
        names = rep(c("30hpf", "48hpf"),4),
        col = c("white", boundaryColors[1], "white", boundaryColors[2],
                 "white", boundaryColors[3], "white", boundaryColors[4] ),
        boxcol= altcols, medcol = altcols, staplecol = altcols,
        whiskcol = altcols, outcol = altcols,
        ylim=c(0,100),
        boxwex=0.6
        )
mtext(text=c("r2r3", "r3r4", "r4r5", "r5r6"), side=3,
      at=seq(1.5, 7.5, by=2), col=boundaryColors)
```



```
par(opar)
```

```
descr2groups(b1$percHuC,b2$percHuC, tit = "r2r3", nams=c("30hpf", "48hpf"))
```

```
## [1] "    r2r3"
##      N      M   SD  SEM
## 30hpf 5   0.52 1.17 0.52
## 48hpf 5 36.94 6.56 2.94
## [1] " t.test p-value: 0.00017966409092005 *** "
```

```
descr2groups(b3$percHuC,b4$percHuC, tit = "r3r4", nams=c("30hpf", "48hpf"))
```

```
## [1] "    r3r4"
##      N      M   SD  SEM
## 30hpf 5   0.54 1.21 0.54
## 48hpf 5 45.06 3.93 1.76
## [1] " t.test p-value: 3.66182536599032e-06 *** "
```

```
descr2groups(b5$percHuC,b6$percHuC, tit = "r4r5", nams=c("30hpf", "48hpf"))
```

```
## [1] "    r4r5"
##      N      M   SD  SEM
## 30hpf 5   1.90 2.64 1.18
## 48hpf 5 44.19 4.62 2.06
## [1] " t.test p-value: 1.1495142048472e-06 *** "
```

```
descr2groups(b7$percHuC,b8$percHuC, tit = "r5r6", nams=c("30hpf", "48hpf"))
```

```
## [1] "    r5r6"
##      N      M   SD  SEM
## 30hpf 5   1.58 1.47 0.66
## 48hpf 5 40.10 5.18 2.32
## [1] " t.test p-value: 3.08634422148325e-05 *** "
```

```
plot(0,0, type="n", xlab="", ylab="",
     xlim=c(0.5,2.5), ylim=c(0,100),
     xaxt="n", yaxt="n")
```

```

points(x=rep(1,4), y=percHuC_30, pch=21,cex=2, bg=boundaryColors )
points(x=rep(2,4), y=percHuC_48, pch=21,cex=2, bg=boundaryColors)

axis(1, at=1:2, labels=c("30 hpf", "48 hpf"), cex.axis=1.6)
axis(2, at=seq(0,100,by=20), labels = seq(0,100,by=20), las=2, cex.axis=1.6)

legend(x="topleft",legend = levels(percBoundTime$bound),
      text.col=boundaryColors, cex=1.2, bty="n")

```

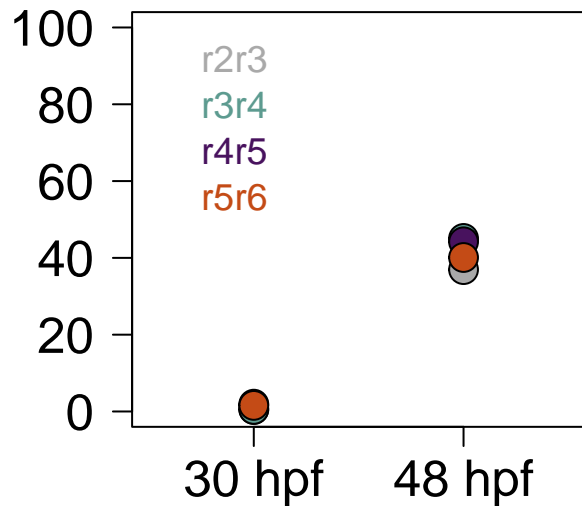


Figure Sup3 C-D

Ventric surfaces

Figure Sup3 E-F

position along DV and LM

Figure Sup 3 G

Histo proliferating

Figure Sup 4 A-B

notch active vs bound and 26-36 hpf

```

pw_percNotch <-pivot_wider(percNotch, names_from = c(time, bound), values_from = percNotch)

opar<-par(no.readonly = TRUE)
par(bty="n")

boxplot(pw_percNotch[,c(2,6,3,7,4,8,5,9)],
      ylab = "% Notch-active boundary cells",
      names = rep(c("26hpf", "36hpf"),4),
      boxlty=rep(1,8),
      col = c("white", boundaryColors[1], "white", boundaryColors[2],
              "white", boundaryColors[3], "white", boundaryColors[4] ),
      boxcol= altcols, medcol = altcols, staplecol = altcols,
      whiskcol = altcols, outcol = altcols,

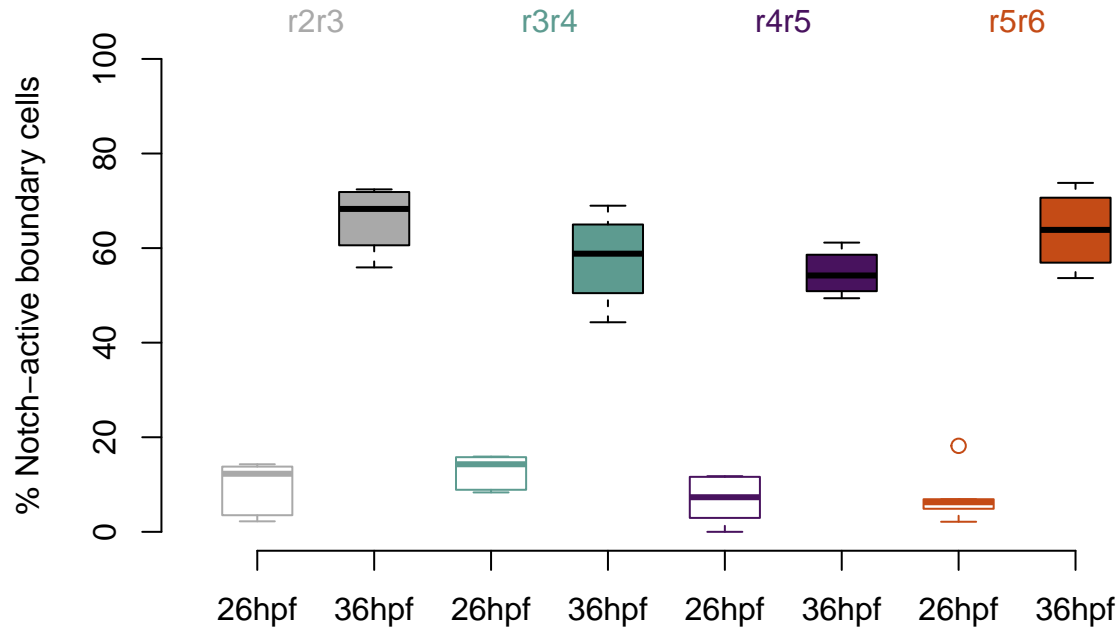
```

```

ylim=c(0,100),
boxwex=0.6
)

mtext(text=c("r2r3", "r3r4", "r4r5", "r5r6"), side=3,
      at=seq(1.5, 7.5, by=2), col=boundaryColors)

```



```

par(opar)

descr2cols(pw_percNotch[,c(2,6)], tit = "r2r3", nams=c("30hpf", "48hpf"))

```

```

## [1] "    r2r3"
##      N      M    SD  SEM
## 30hpf 5   9.22 5.86 2.62
## 48hpf 4  66.22 7.56 3.78
## [1] " t.test p-value: 2.74404565196448e-05 *** "

```

```

descr2cols(pw_percNotch[,c(3,7)], tit = "r3r4", nams=c("30hpf", "48hpf"))

```

```

## [1] "    r3r4"
##      N      M    SD  SEM
## 30hpf 5  12.64 3.74 1.67
## 48hpf 4  57.73 10.30 5.15
## [1] " t.test p-value: 0.00170158398092096 ** "

```

```

descr2cols(pw_percNotch[,c(4,8)], tit = "r4r5", nams=c("30hpf", "48hpf"))

```

```

## [1] "    r4r5"
##      N      M    SD  SEM
## 30hpf 5   6.73 5.23 2.34
## 48hpf 4  54.74 5.06 2.53
## [1] " t.test p-value: 3.56889512735125e-06 *** "

```

```

descr2cols(pw_percNotch[,c(5,9)], tit = "r5r6", nams=c("30hpf", "48hpf"))

```

```

## [1] "    r5r6"

```

```
##          N      M   SD  SEM
## 30hpf  5    7.67 6.16 2.75
## 48hpf  4   63.79 8.75 4.37
## [1] " t.test p-value: 8.70923321486027e-05 *** "

plot(NULL,NULL, type="n",
      xlim=c(0.5,2.5), ylim=c(0,100),
      xaxt="n", yaxt="n",
      main= "",
      xlab="", ylab="") #, ylab="% of Notch-active cells by boundary")
points(x=ifelse(percBoundTimeNotch$time=="26hpf",1,2),
       y=percBoundTimeNotch$percNotch,
       pch=21,cex=1.6,bg=rep(boundaryColors,each=2))

legend(x="topleft",legend = levels(percBoundTime$bound),
       text.col=boundaryColors, cex=1.2, bty="n")
axis(1,at=1:2,labels = c("26hpf","30hpf"), cex.axis=1.6)

axis(2, at=seq(0,100,by=20), labels = seq(0,100,by=20), las=2, cex.axis=1.6)
```

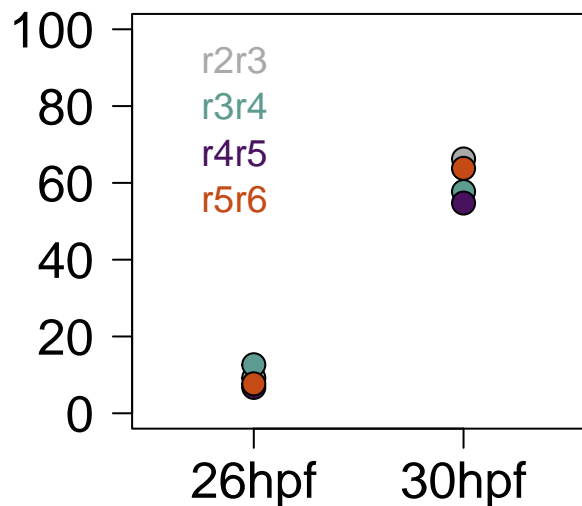
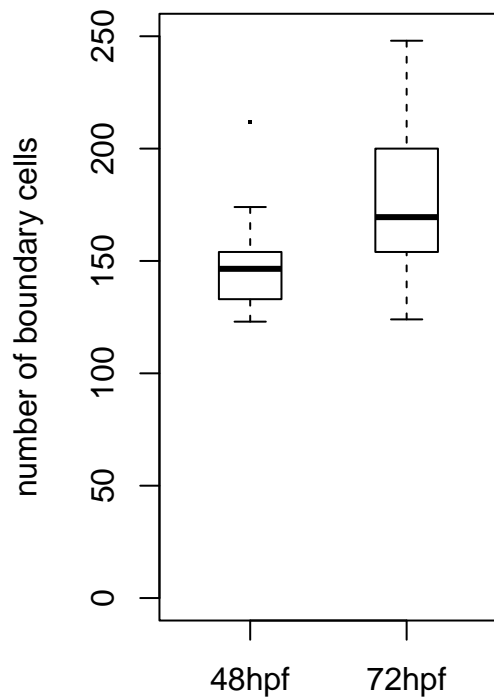


Figure Sup 5A

bp number vs 48-72

```
tot48 <- diff4872$nCells[diff4872$time=="48hpf"]
tot72 <- diff4872$nCells[diff4872$time=="72hpf"]

boxplot(tot48, tot72,
        ylim = c(0,maxRight),
        xlab="", ylab = "number of boundary cells",
        names = c("48hpf", "72hpf"),
        col = rep("white",2),
        boxwex=0.4, outpch='.', cex=2)
```

```
# descripcion total number of cells
descr2groups(tot48, tot72, nams = c("48hpf", "72hpf"), not.paired=TRUE)
```

```
##          N      M    SD   SEM
## 48hpf  10 150.8 26.11  8.26
## 72hpf  10 178.5 42.73 13.51
## [1] " t.test p-value: 0.100837106653853 ns"
```