2018-11-15 qbic test 2a

Theresa Swayne

11/16/2018

## Testing nuclear detection and DNA content analysis

The background-corrected integrated intensity (“total intensity”) of nuclei should be proportional to DNA content.

The DNA content in a mixed population of actively cycling cells should form a roughly bimodal distribution representing 1N and 2N populations, with some intermediate cells in S phase.

CellProfiler pipelines were used to perform background correction, identify nuclei using DAPI, and measure intensity of DAPI and other channels in a set of 100 images collected on 11-5-2018.

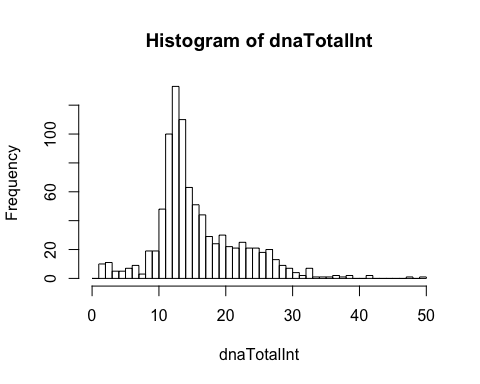
## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## ImageNumber = col\_integer(),  
## ObjectNumber = col\_integer(),  
## Metadata\_FileLocation = col\_character(),  
## Metadata\_Frame = col\_integer(),  
## Metadata\_Series = col\_integer(),  
## Metadata\_XY = col\_integer(),  
## Metadata\_basename = col\_character(),  
## AreaShape\_Area = col\_integer()  
## )

## See spec(...) for full column specifications.

## 922 cells detected.

This is a histogram of the DNA content over all nuclei:

# use echo=FALSE to prevent printing code  
hist(dnaTotalInt, breaks = (0:50))



The histogram shows that there is a roughly bimodal distribution with peaks ~ 12 and ~ 24, and a long upper tail containing few cells.

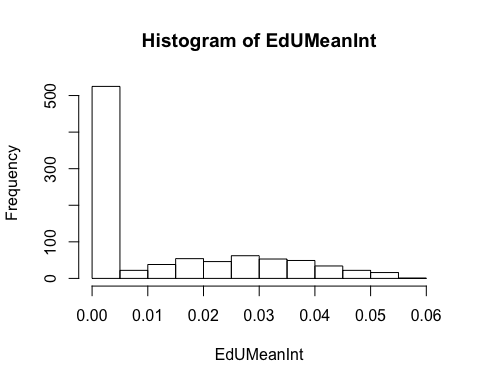
low\_vals <- mean(dnaTotalInt < 1) \* 100  
high\_vals <- mean(dnaTotalInt > 30) \* 100  
central\_vals <- mean(dnaTotalInt > 1 & dnaTotalInt < 30) \* 100  
  
cat(central\_vals, "% of the values are between 1 and 30; ",low\_vals,"% <1; and",high\_vals,"% >30.")

## 97.2885 % of the values are between 1 and 30; 0 % <1; and 2.711497 % >30.

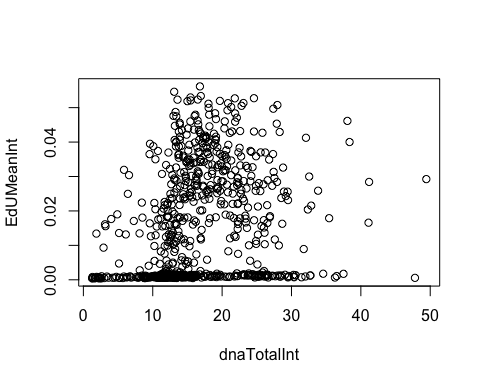
Ultimately we would like to see an inverted horseshoe-type plot of EdU (showing S phase cells) vs. DNA content where the cells incorporating EdU are primarily those between the 2 peaks.

Generating the 2nd dimension of the plot from the corrected mean intensity from Channel 2, EdU:

#EdUMeanInt <- qbic\_C1DNANuclei$Intensity\_MeanIntensity\_Channel2EdU  
EdUMeanInt <- qbic\_C1DNANuclei$Intensity\_MeanIntensity\_Channel2EdUCorr  
hist(EdUMeanInt)



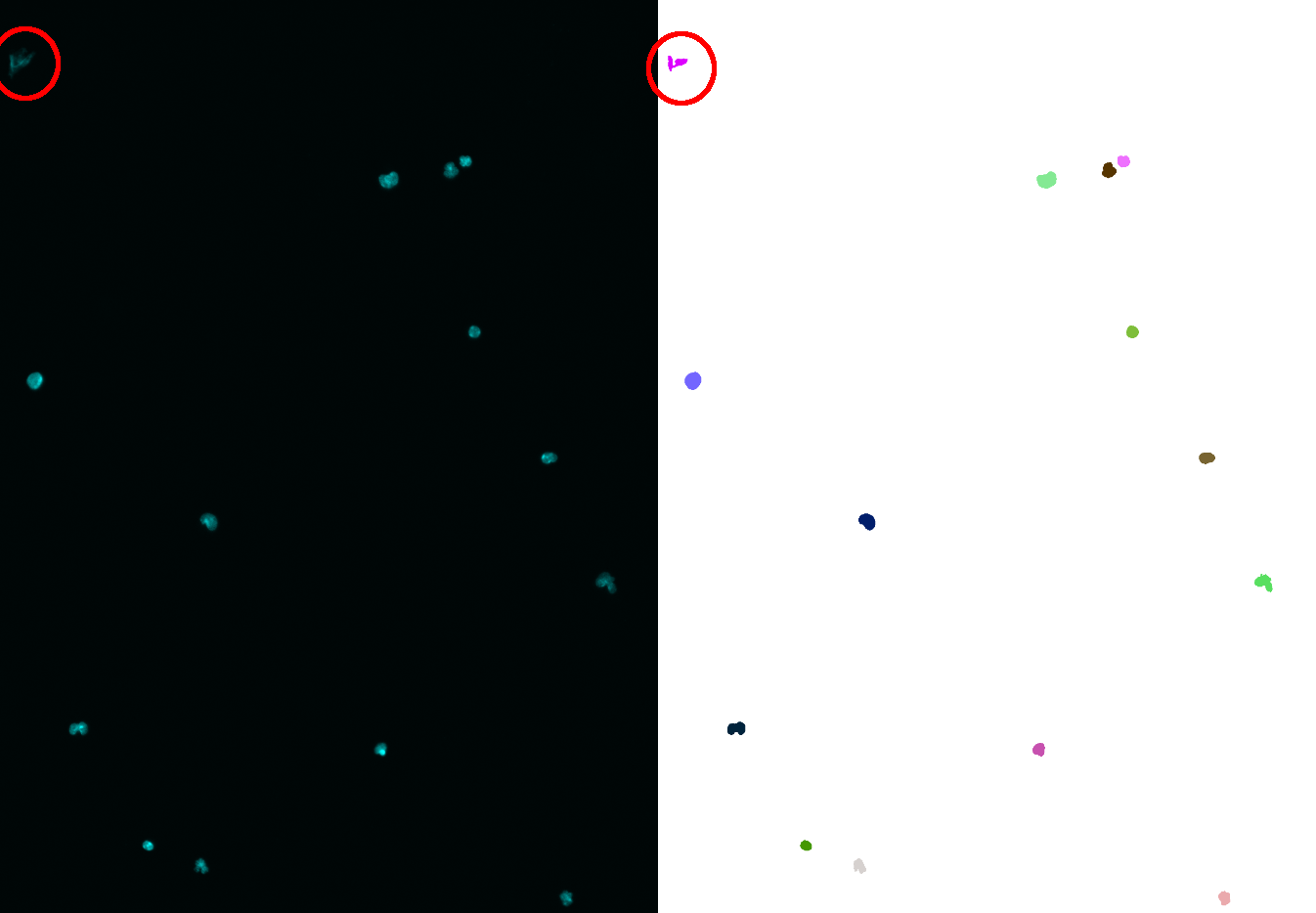
plot(x = dnaTotalInt, y = EdUMeanInt) # basic scatter plot



### Are nuclei detected properly?

Several images were selected at random, and the image of detected objects was compared with the original data.

False positives (non-cells counted as objects) and false negatives (cells not included in the object map) were scored manually.



Detection of nuclei in XY point 30, 11/5/18. Left, original DAPI channel. Right, detected objects, each in a different color. Red circle shows a false positive.

qbic\_check <- read\_csv("~/github\_theresaswayne/cellprofiler/2018-11-15 qbic test 2a.csv")

## Parsed with column specification:  
## cols(  
## Image = col\_integer(),  
## `Cells detected by CP` = col\_integer(),  
## `False positives` = col\_integer(),  
## `False negatives` = col\_integer(),  
## Notes = col\_character()  
## )

kable(qbic\_check[,(1:5)])

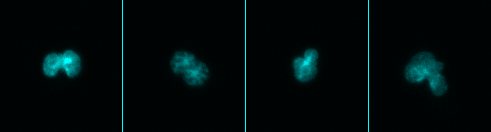
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Image | Cells detected by CP | False positives | False negatives | Notes |
| 89 | 9 | 0 | 0 | NA |
| 51 | 11 | 0 | 0 | NA |
| 31 | 12 | 0 | 0 | NA |
| 36 | 9 | 2 | 0 | NA |
| 45 | 0 | 0 | 0 | NA |
| 13 | 11 | 0 | 0 | NA |
| 2 | 7 | 0 | 0 | NA |
| 12 | 9 | 0 | 0 | NA |
| 7 | 13 | 0 | 0 | NA |
| 45 | 0 | 0 | 0 | NA |
| 43 | 16 | 0 | 0 | 1 possible false pos (out of focus) lower right |
| 31 | 12 | 0 | 0 | Edge objects inconsistent – here a cell was detected that is at edge (partial nucleus) but in #43 a cell barely on edge was eliminated |
| 32 | 15 | 0 | 0 | NA |
| 29 | 10 | 0 | 0 | NA |
| 30 | 14 | 1 | 0 | false pos = elongated junk |
| 39 | 9 | 0 | 0 | NA |
| 68 | 15 | 0 | 0 | NA |
| 85 | 8 | 0 | 0 | NA |
| 65 | 3 | 1 | 0 | NA |
| 86 | 8 | 0 | 0 | one cell near top edge was segmented smaller than in reality |

total\_cells <- sum(qbic\_check$`Cells detected by CP`)  
false\_pos <- 100 \* sum(qbic\_check$`False positives`)/total\_cells  
false\_neg <- 100 \* sum(qbic\_check$`False negatives`)/total\_cells  
cat("Out of",total\_cells,"cells checked, there were",false\_pos,"% false positives and",false\_neg,"% false negatives.")

Out of 191 cells checked, there were 2.094241 % false positives and 0 % false negatives.

## Question: What to do with apparent clumps?

Some DAPI structures appear to have multiple lobes. Are these multi-lobed nuclei, mitotic cells, or actually multiple cells close together? If needed, we can modify the object detection to split objects by shape.



Examples of clumped nuclei from XY points 21 and 30, 11/5/18.

## Summary

* Time of processing with CellProfiler is < 10 min per 100 images for illumination correction and the same for object detection.
* Nuclei are detected well in this system. There are virtually no false negatives.
* Some bright debris results in false positives, about 2%.
* We need to decide if some objects should be split or if they are actually single nuclei.