1.1 Brucella Locations (J101-2C)

Nicolas Bennett

2015-05-25

The following investigation should shed a light on the quesion whether the location of an object (such as a cell) within an image or well has an influence on its features. If such an influence is visible, it may have to be considered to remove some objects (e.g. in image or well corners) or possibly even entire images (out of 9 image wells, maybe only the center image is usable).

```
features <- c("^Cells.Location_Center_X$",
              "^Cells.Location_Center_Y$",
              "^Cells.AreaShape_Area$",
              "^Cells.AreaShape_Perimeter$",
              "^Cells.Texture_Contrast_CorrDNA_3$",
              "^Cells.Texture_Contrast_CorrActin_3$",
              "^Cells.Texture_Entropy_CorrDNA_3$",
              "^Cells.Intensity_MeanIntensity_CorrActin$",
              "^Cells.Intensity_IntegratedIntensity_CorrPathogen$")
plate <- PlateLocation("J101-2C")
data  <- PlateData(plate, features)
```

```
## reading plate chache file.
## assuming 9 images per well:
## max legnth: 3456, fraction of full length features: 0.995
## removing 599 unmatched features.
```

```
augm <- augmentCordinateFeatures(data, ellipse=1, facet=c(14, 10),
                                 center.dist=TRUE)
```

```
## using a single ellipse, 50px dist from borders.
## facet size: 99 (x), 104 (y)
## including distance to image center.
```

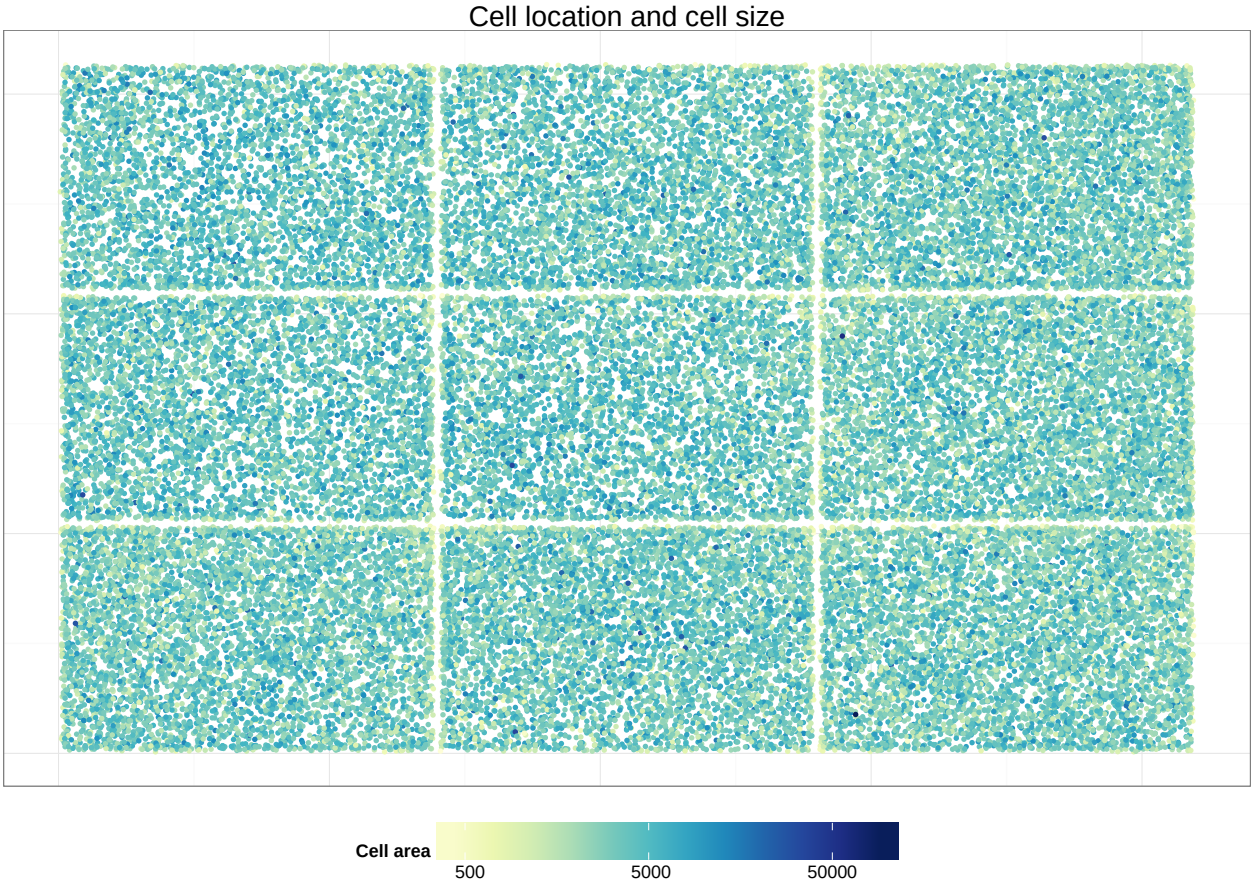
```
augm <- augmentImageLocation(augm)
melt <- meltData(augm)
cells <- melt$mat$Cells
colnames(cells)
```

```
##  [1] "Cells.AreaShape_Area"
##  [2] "Cells.AreaShape_Perimeter"
##  [3] "Cells.Intensity_IntegratedIntensity_CorrPathogen"
##  [4] "Cells.Intensity_MeanIntensity_CorrActin"
##  [5] "Cells.Location_Center_X"
##  [6] "Cells.Location_Center_Y"
##  [7] "Cells.Texture_Contrast_CorrActin_3"
##  [8] "Cells.Texture_Contrast_CorrDNA_3"
##  [9] "Cells.Texture_Entropy_CorrDNA_3"
## [10] "Cells.Location_In_Ellipse"
## [11] "Cells.Location_Facet_X"
```

```
## [12] "Cells.Location_Facet_Y"
## [13] "Cells.Location_Dist_Center"
## [14] "Cells.Location_Shifted_X"
## [15] "Cells.Location_Shifted_Y"
## [16] "Image.Index"
## [17] "Well.Index"
## [18] "Well.Name"
## [19] "Plate.Barcode"
## [20] "Image.Group"
```

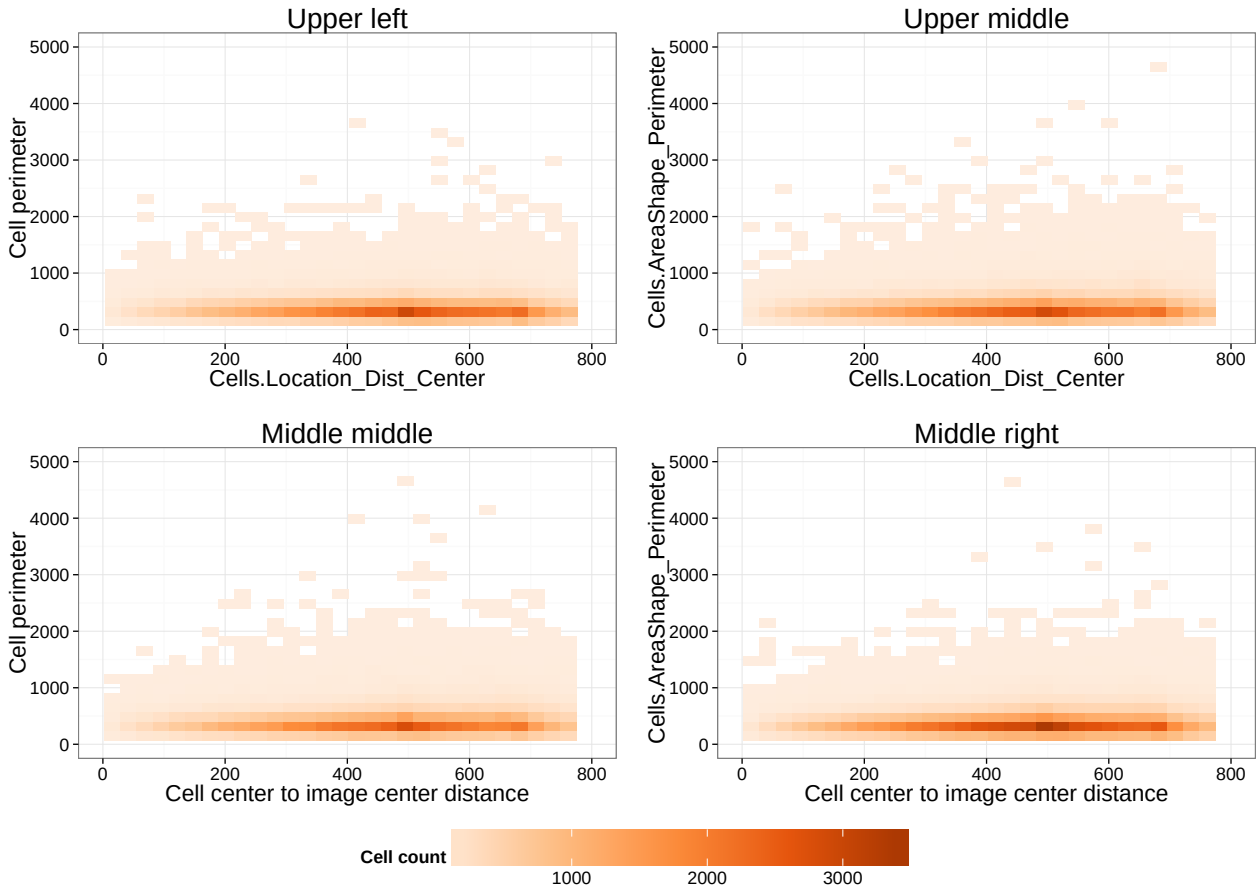
```
rm(features, data, augm, melt)
```

A subset of the cell features for the plate J101-2C (BRUCELLA-DU-K1) is loaded. The data is augmented with additional features: `augmentCordinateFeatures` is capable to add membership to concentric ellipses and tiles, as well as distance to image center for any kind of `Location_Center` features, while `augmentImageLocation` adds information on where the image is located within a well. FInally the data is *melted* into a single `data.frame`.

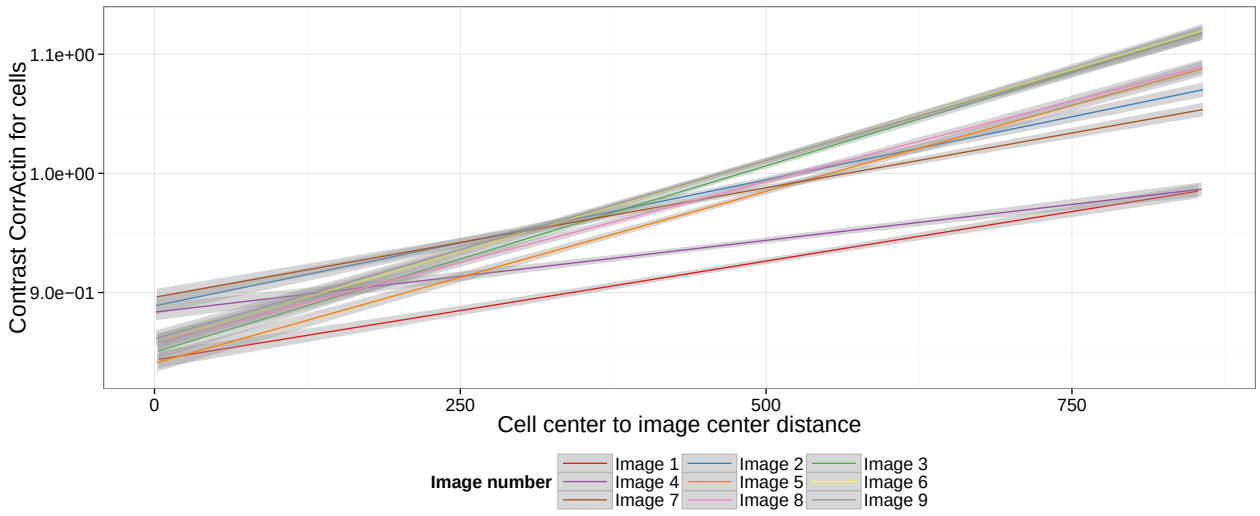


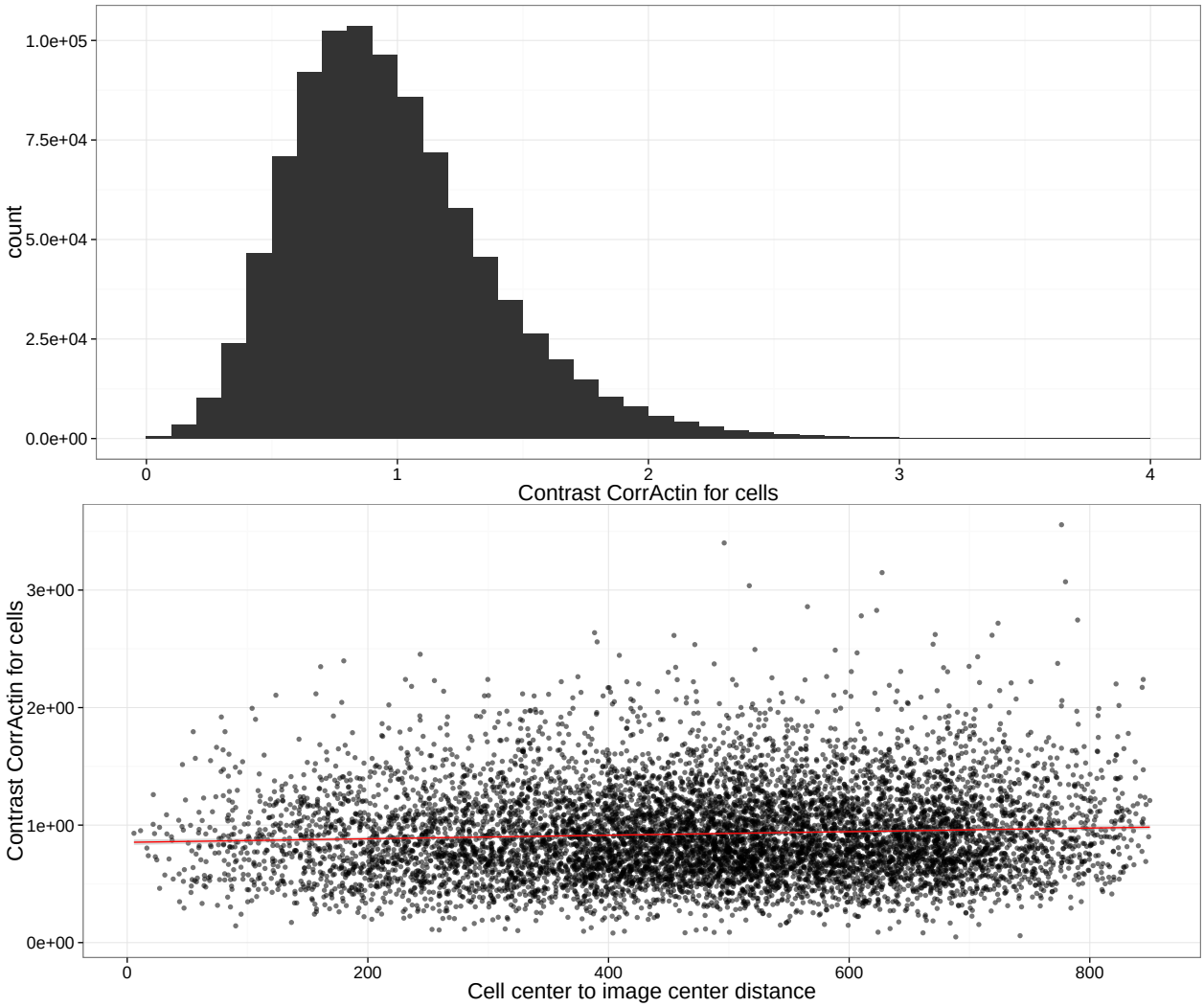
A first visualization approach: a scatterplot for cell center location with color values for cell size. Cell size varies from 315 to 3.68×10^5 with $\mu = 3987.48$ and $\sigma = 2624.99$. The color scale is logarithmic to achieve sensitivity at lower values while still showing the largest cells. There are 945851 datapoints and because of overlapping, only a randomly chosen subset of length 47292 is plotted.

For the next set of visualizations, for each cell, the distance (of the cell center) from the image center is plotted against the cell perimeter for the images in the upper left corner and the center of the well. Due to the large sample size (87540 for upper right, 96745 for upper middle, 91313 for middle and 116111 for middle right), the scatterplots are 2D-binned.

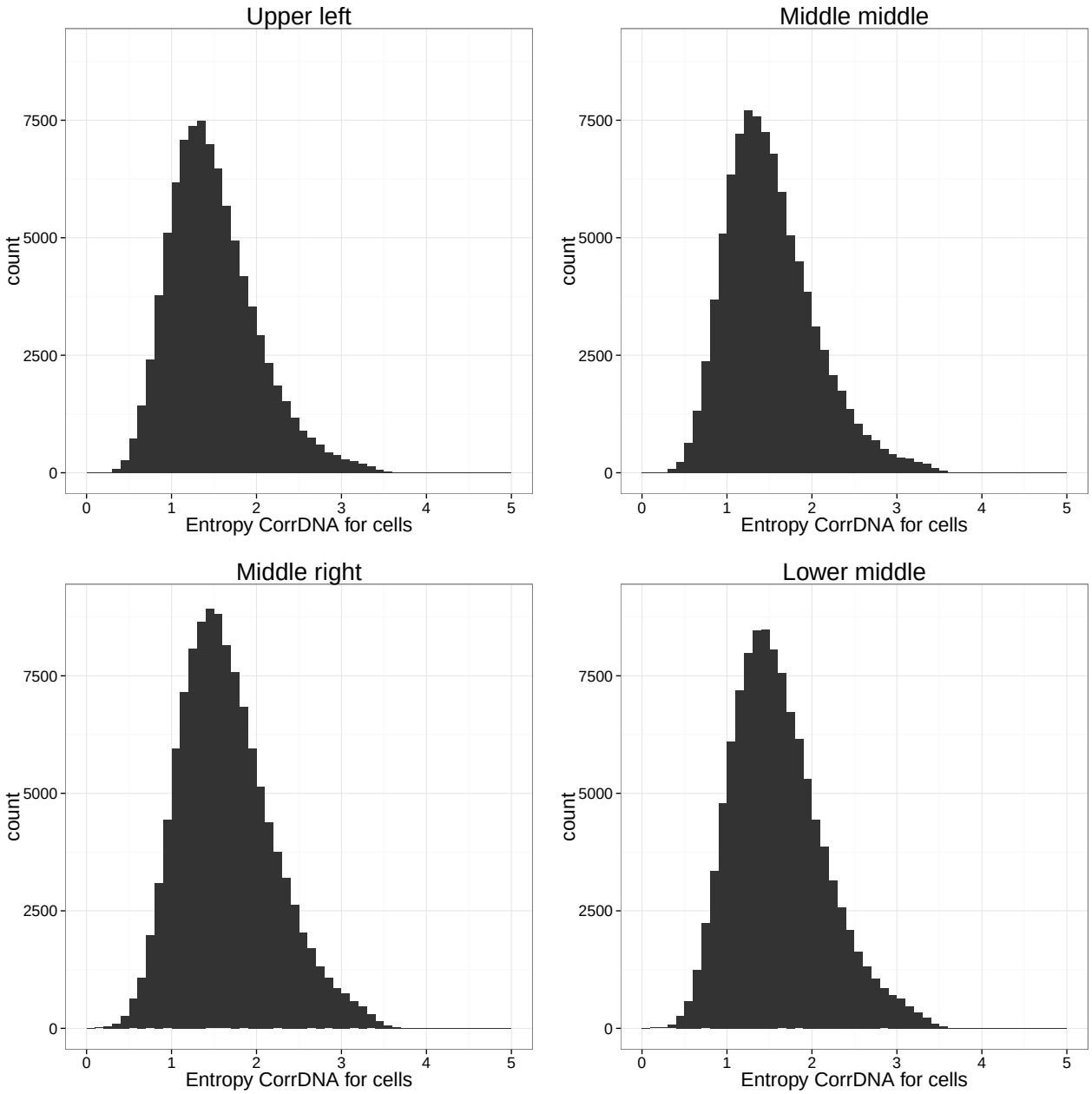


The feature `Cells.Texture_Contrast_CorrActin_3` is visualized in the following three plots. Trends for how the feature depends on image center to cell center distance are plotted for all nine images (image 1: upper left, image 2: upper middle, etc.) While a clear, approximately linear trend moving outwards is visible, the slopes are small. This is especially visible in the third plot, showing only the trend for image 1 but while keeping the datapoints visible (only a randonly selected subset of 1/10 of the datapoints are shown). The histrogram shows how the contrast values are distributed (whole plate).

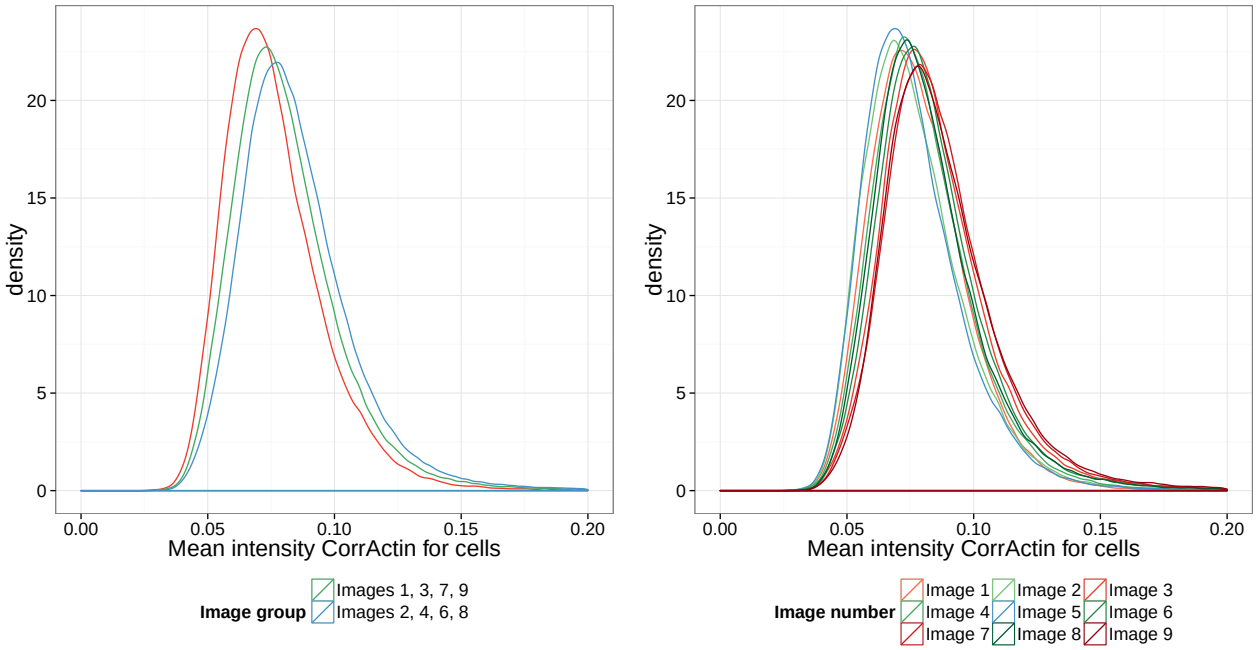




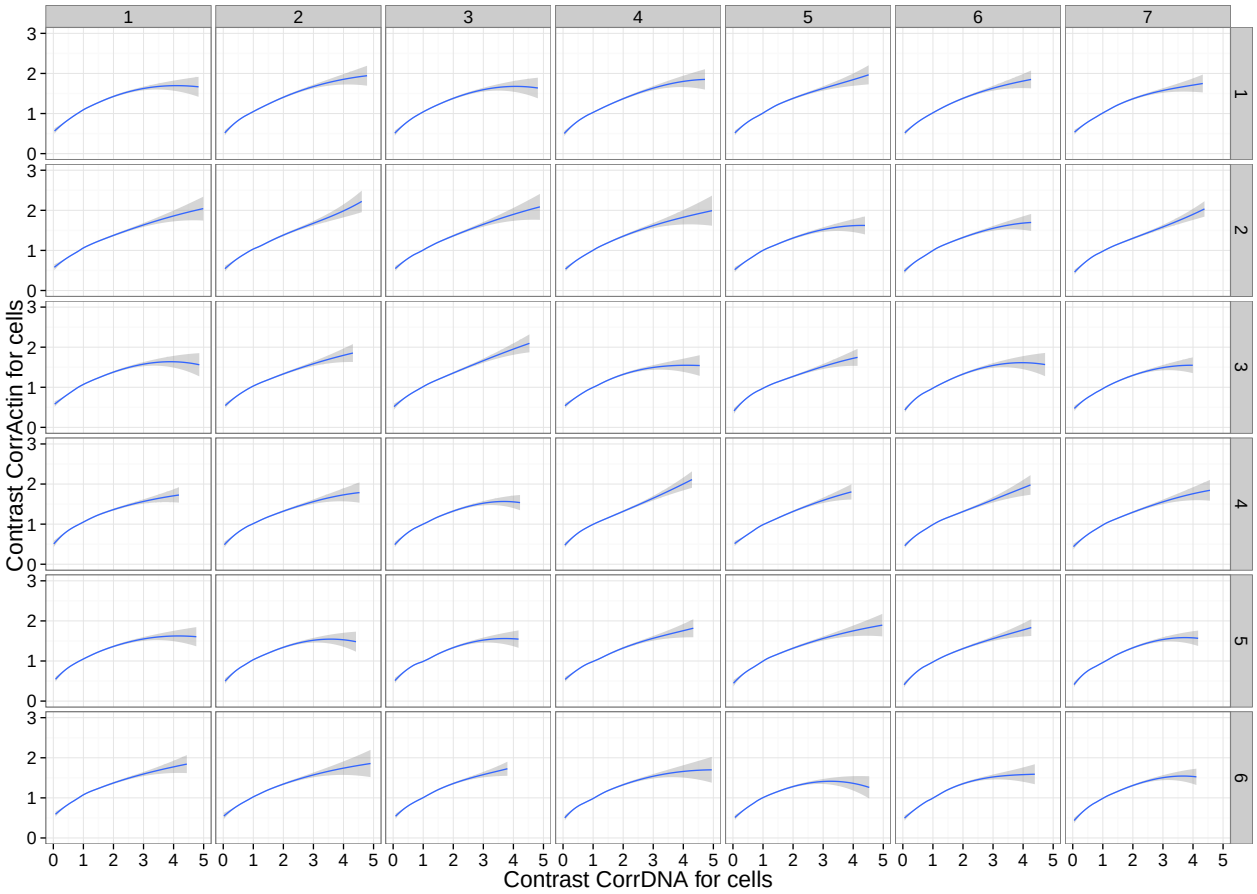
The following 4 histograms show how the feature `Cells.Texture_Entropy_CorrDNA_3` is distributed in four of the nine images (fig. a shows the upper left image, fig. b, the central image, fig. c the rightmost image in the middle row and fig. d, the middle image in the bottom row).



To compare feature distribution among different images, density estimated are plottes in the following two images. First, for visual clarity, the nine images are grouped according to well border length: images 1, 3, 7 and 9 are corer images, while images 2, 4, 6 and 8 have a well border along one side. Image 5 is located in the well center.



Next up is a ggsubplot-inspired view of how the relationship between `Cells.Texture_ Contrast_CorrActin_3` and `Cells.Texture_Contrast_CorrDNA_3` depends on cell location within an image. All nine images are superimposed, rendering the resulting view radially symmetric. Exploiting this symmetry, only the upper left quadrant of the image is shown. Each box corresonds to an area of 100 by 100 pixels.



Finally, a facet plot conditioned on whether the cell center lies within the ellipse bounded by a box moved 50px inwards from every image border shows how integrated intensity of the CorrPathogen channel for cells behaves within the ellipse, as well as outside.