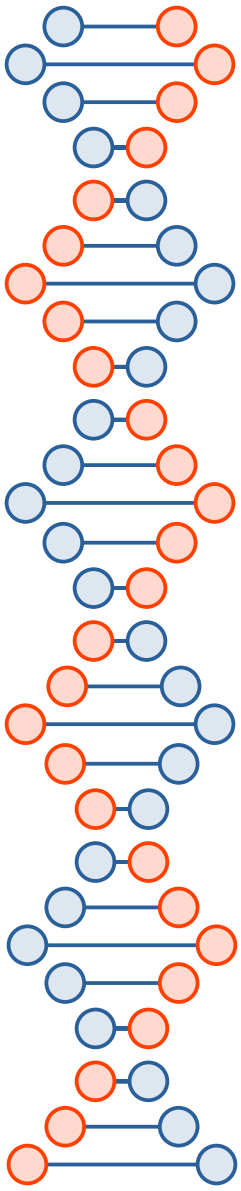
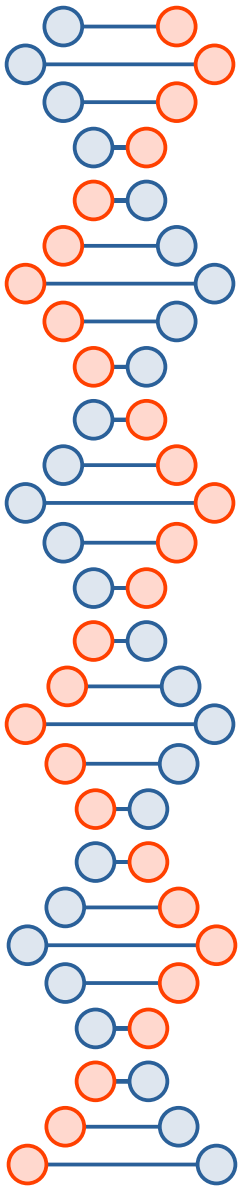


Disha Automated Y1H Assay Analysis system

Computer Vision Team

Mahir Patel
Cosmin Ciausu



Project Description and Goals

An automated tool to detect and analyze yeast-based assays to learn more about transcriptional regulation in immune responses and cancer.

Our steps to perform such analysis is as follows:

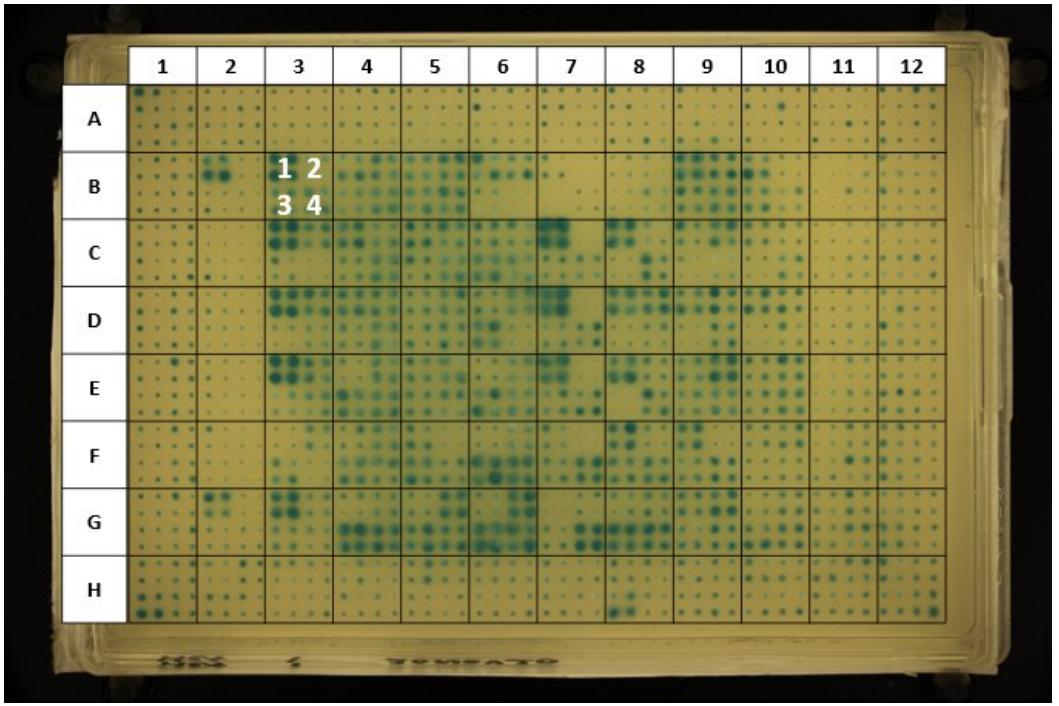
1. Crop the image such that only the assay is visible.
2. Generate grid lines based on our coordinate system.
3. Segment each colony and analyze its color and area.
4. Generate score using metric that uses the intensity and area of the colonies.
5. Generate HTML and excel files.

The Coordinate System

The rows are indexed on A-H and the columns are indexed on 1-12.

Each cell contains 4 quads where each quad contains 4 yeast colonies.

Our coordinate system is #Quad-<row><col>
3-B3 defines the quad marked with ‘3’

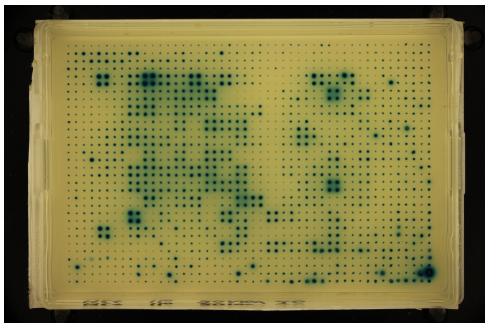


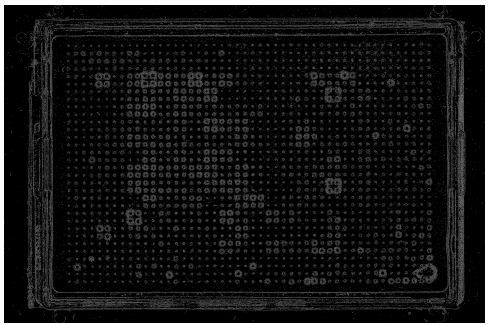
Border detection and cropping

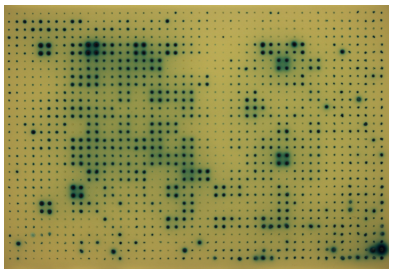
- Performed edge detection using Canny Edge detection.
- Extracted these two templates for top left and bottom right corner.
- Performed template matching using Normalized Cross Correlation

$$\frac{1}{n}\sum_{x,y}\frac{1}{\sigma_f\sigma_t}\left(f(x,y)-\mu_f\right)\left(t(x,y)-\mu_t\right)$$

- Found corners and cropped image accordingly.

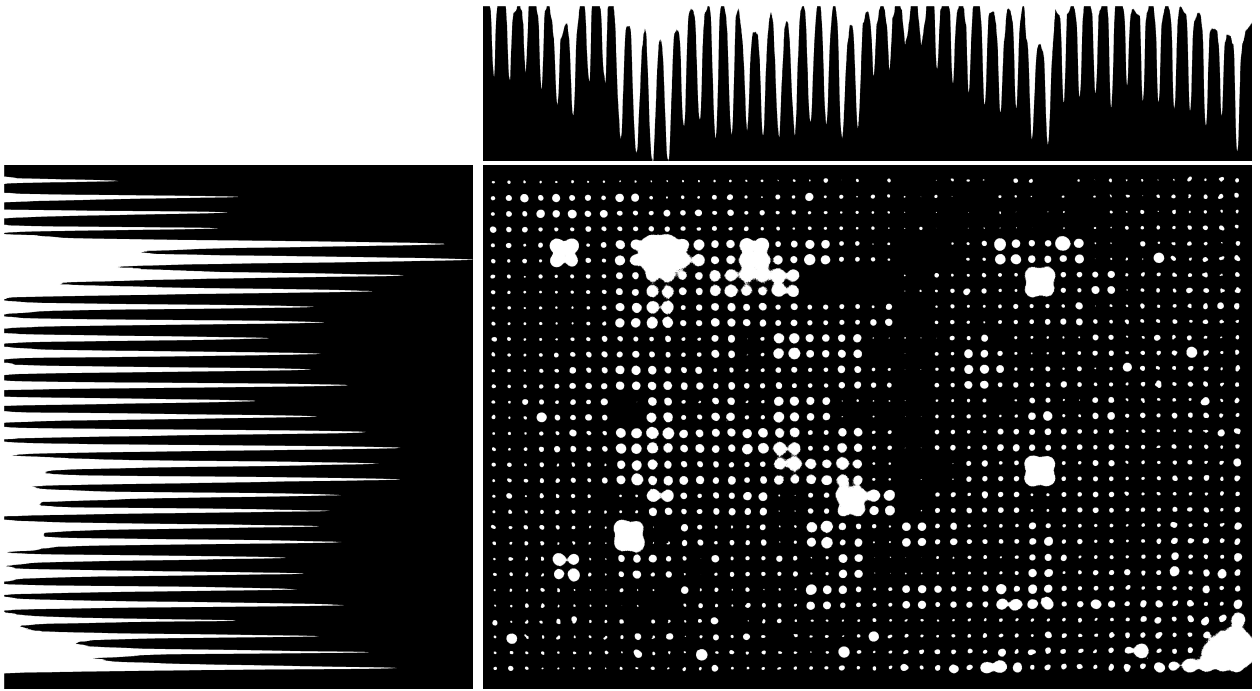






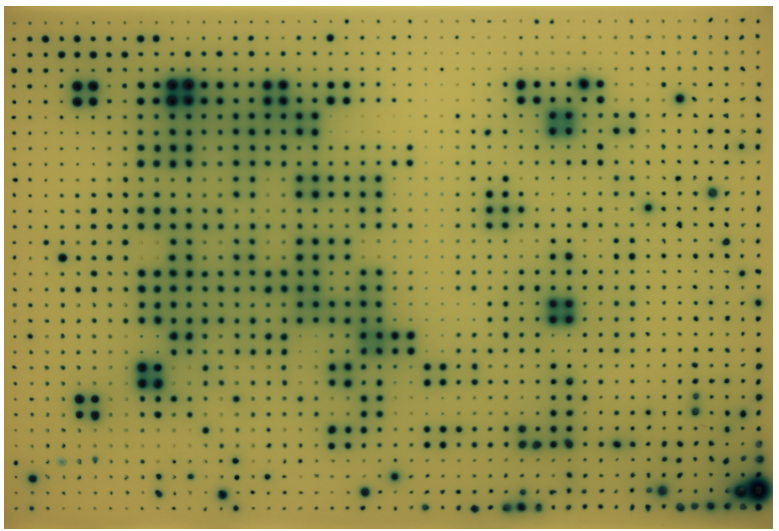
Further Automated Cropping

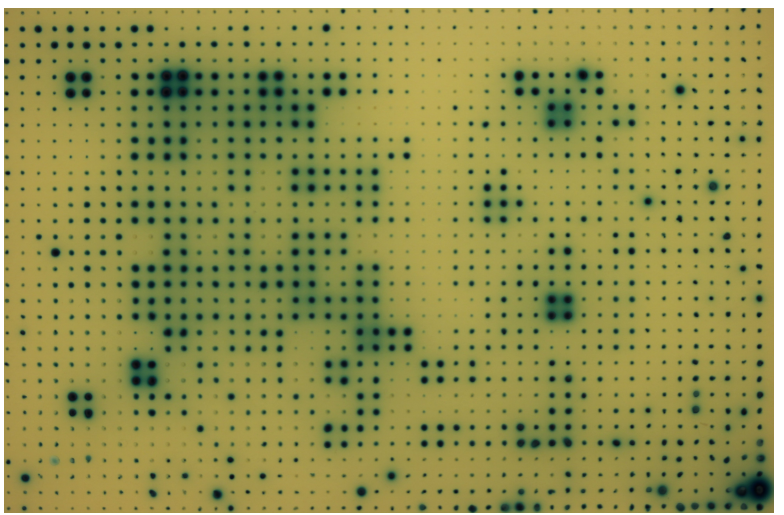
- Converted image to binary image and computed histogram projections.



Further Automated Cropping

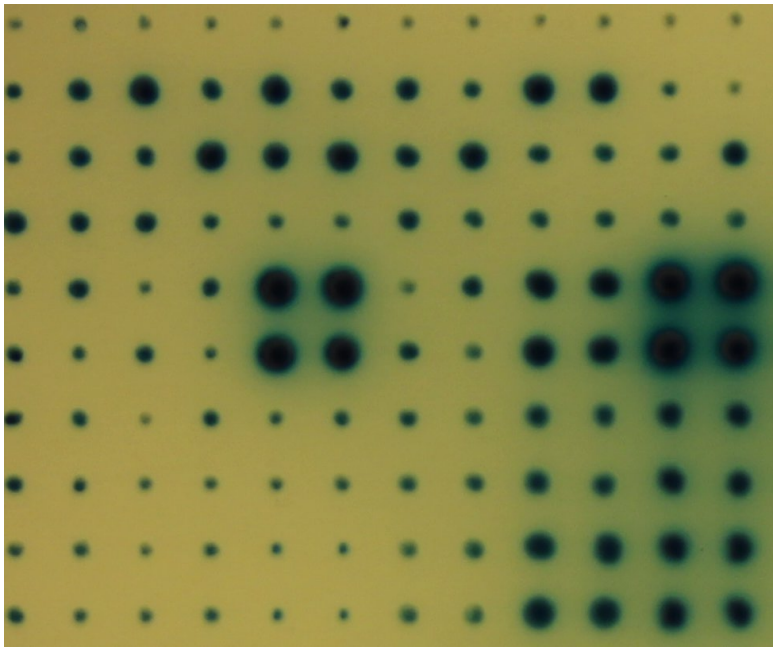
- Using initial peaks in vertical and horizontal histogram projections, we can detect where the assays begin.

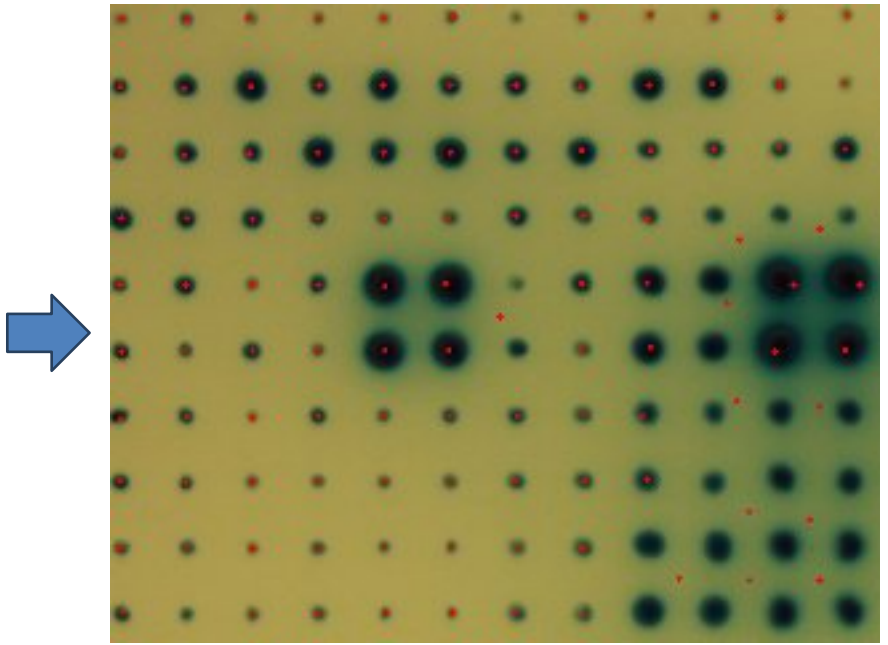




Locating Yeast Colonies

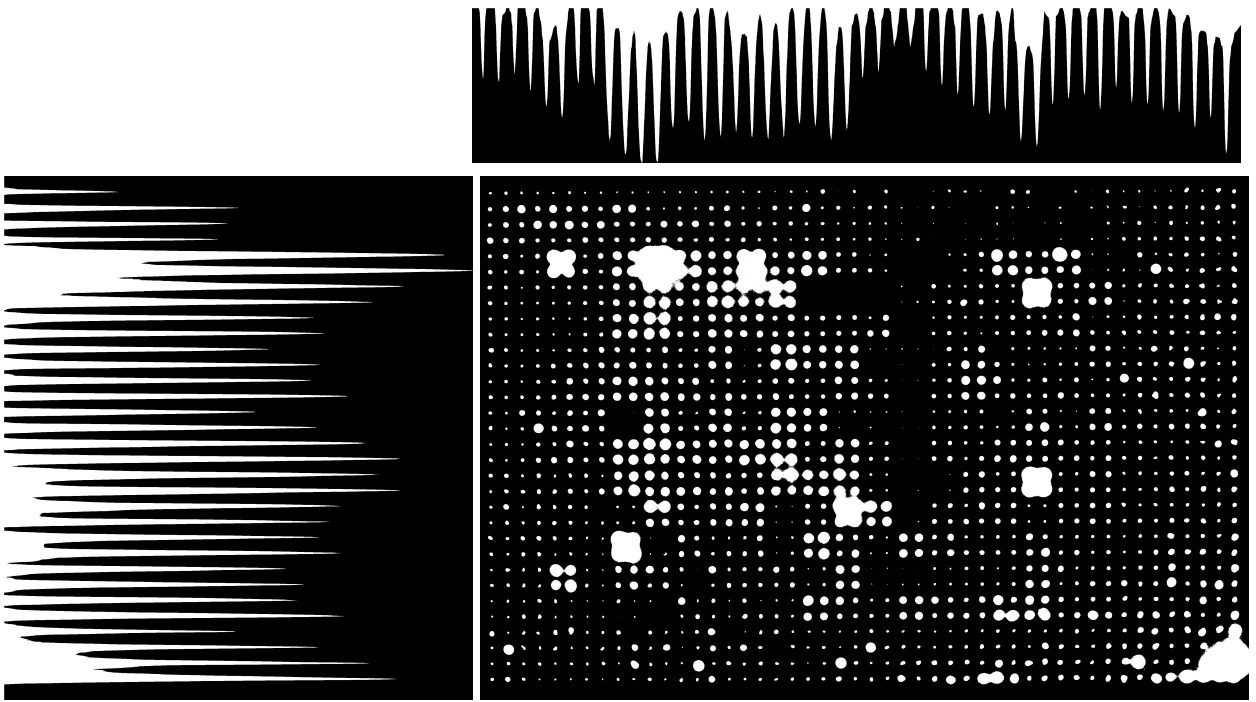
- We tried using Hough Circle Transform to locate yeast colonies.





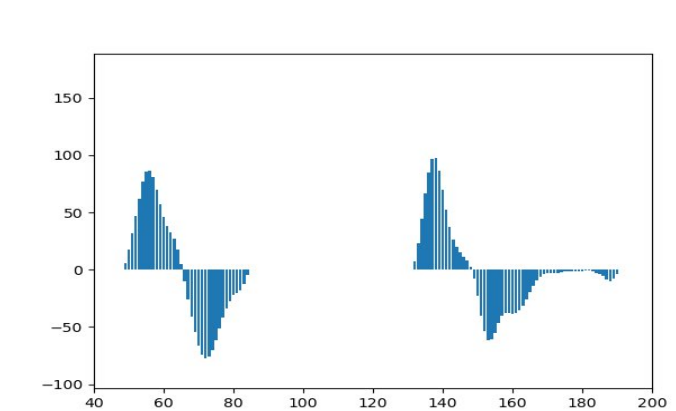
Grid Generation

- Instead of Hough circle transforms, we decided to come up with our own algorithm for more robustness. We noticed that the peaks in the histogram projections are our point of interest.

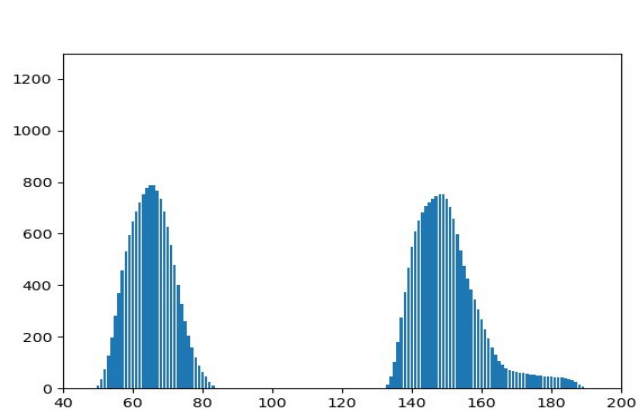


Peak Detection in Histogram

- First We smooth the projections using one dimensional convolution with kernel [0.2,0.2,0.2,0.2,0.2].
- Then we compute the gradient. Now all we have to do is find “Zero Crossings” to determine the peaks.
- We compute discrete difference of the peaks and take an average to find the approximate lengths of each cells.
- Using this data, we filter our list of peaks to remove false positives and add missing points.

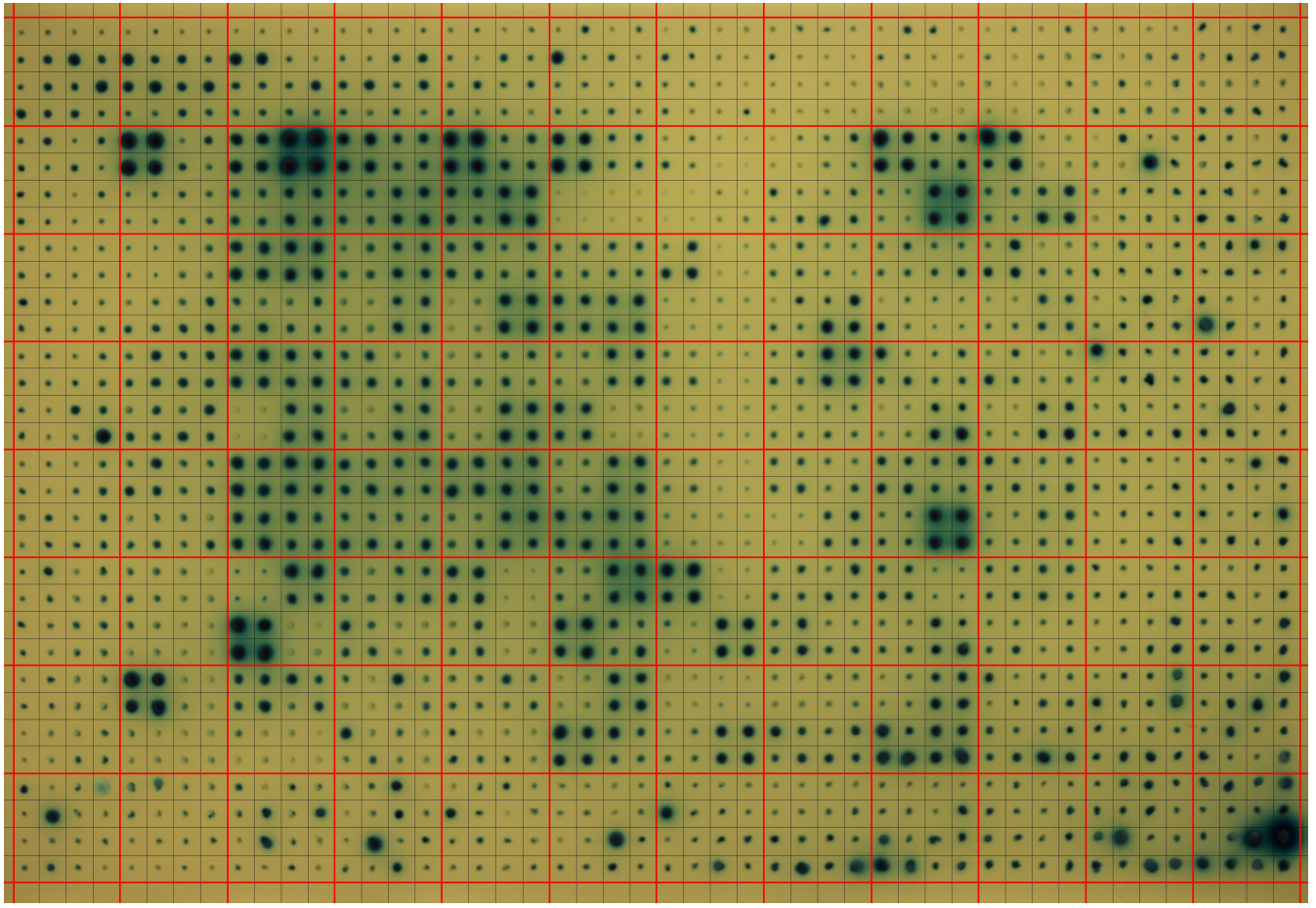


Gradient of Histogram projection

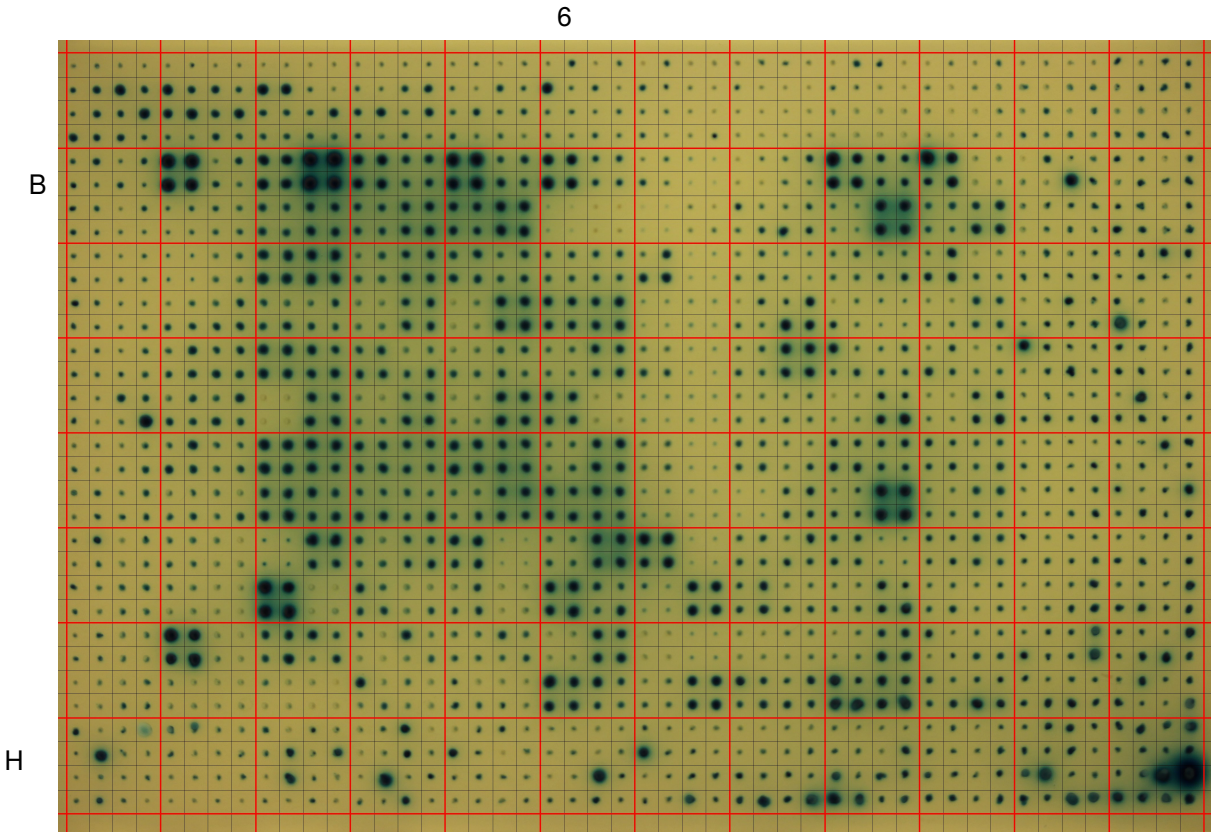


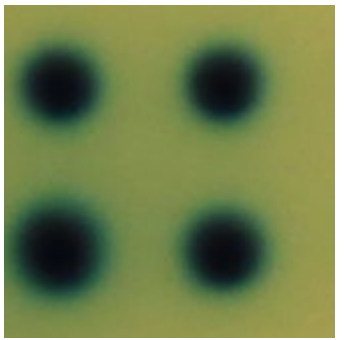
Histogram Projection

Example of Grid generation



- Using this data we can extract each colony and plates.
- Example: extract 1B-6,4-H12





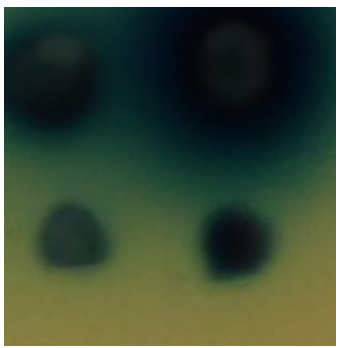
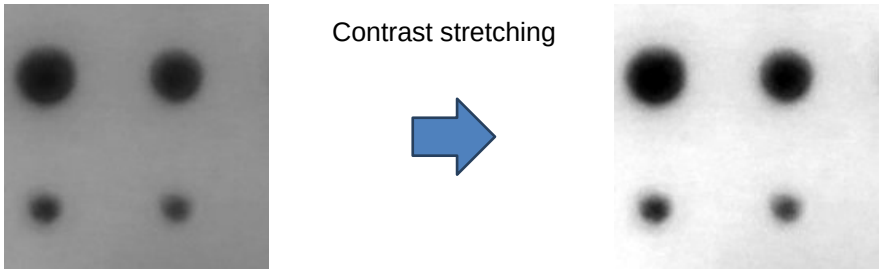


Image Processing Decisions

1. We decided to use **HSV** color space instead of RGB because, the colonies were more pronounced in the **value** channel than any of the RGB channels.

2. Contrast Stretching

We analyzed the histogram of the colony ROI to select lower and upper intensity value such that it covered 99% of the pixels. And we mapped [lower,upper] to [0,255] using interpolation to perform contrast stretching.



3. Histogram analysis to find threshold range to convert to binary image. We also perform contour detection and shape analysis to get rid of bleeding effect which the phenomena of bleed over noise of the surrounding reactive colonies.

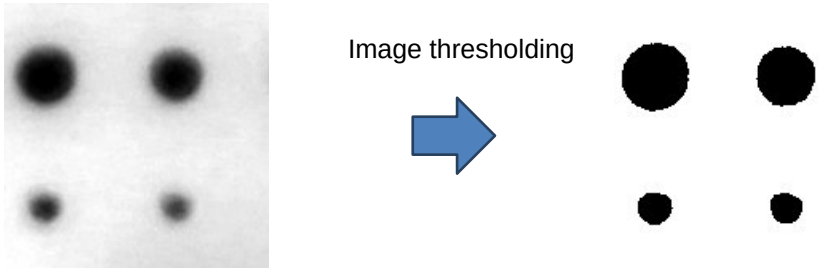
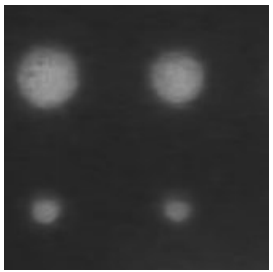


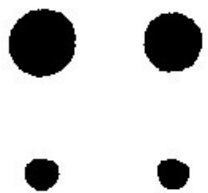
Image Processing Decisions

4. Non-reactive yeast assays have low intensity therefore to penalize that, we first performed contrast stretching on the **hue** channel where the lower bound is generated as explained before but the upper bound is set to 255. This penalizes brighter values. Our segmentation pipeline is very robust so it even detects the non-activated colonies that are barely visible. To solve that, we use these pre-mask images to filter out those images. It's application is explained better in the next slide.

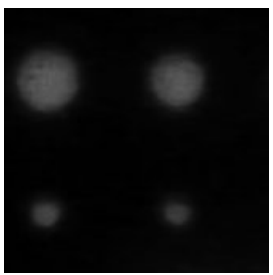


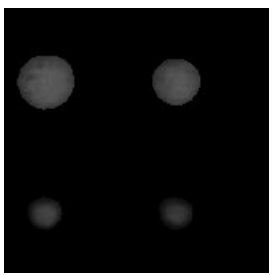
Pre-Mask Generation

5. We compute logical and of the pre-mask and our segmentation mask to generate intensity image.



&





Intensity Image

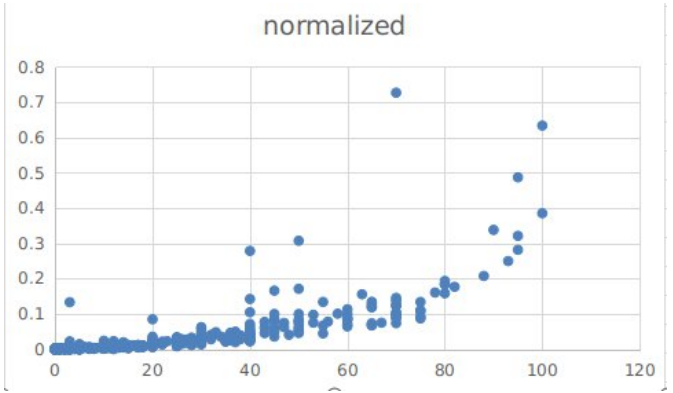
Logic behind pre-mask images

The following ROI is a section of non-reactive quad. The ideal score of this image should be zero. Since our segmentation pipeline was able to detect the colonies, just using it would have generated a really high score.
However, the pre-mask image for this ROI is blank. So our final intensity image will also be zero.

Current Metric

Note: Our goal is to define a greedy metric because false positives are allowed but there should not be any false negatives.

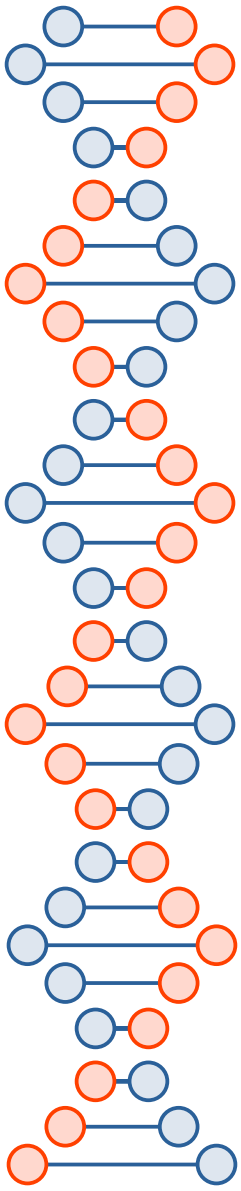
For now, Anna gave us ground truth scores generated by her. We calculate the areas of the colonies from the binary mask and the intensities from the intensity image.



We plotted Anna's score with our normalized intensity values and found that it resembled an exponential curve.
We fitted our metric to this curve and generated the score.

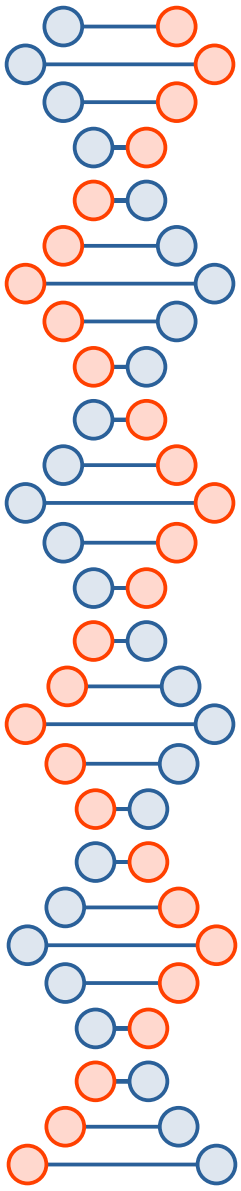
The following table contains Pearson correlation and Spearman correlation coefficients of our generated score with Anna's score.

Name	Pearson	Spearman
AB_HetY1Hpilot3-16_5mm_3AT_X-gal_7d	0.939617824	0.976503567
AB_HetY1Hpilot3-9_5mm_3AT_X-gal_7d	0.946093899	0.726469999
AB_HetY1Hpilot3-7_5mm_3AT_X-gal_5d	0.908813135	0.930786987
AB_HetY1Hpilot3-5_5mm_3AT_X-gal_7d	0.937676163	0.945422382
Average	**0.933050255**	**0.894795734**



Future Work

- We are waiting for quantified scores generated by a machine so that we can model a more accurate method.
- We are currently developing a Web-App that they can host on SCC so that anyone in the department can use it.
- We need to develop a dynamic way to select a threshold for solving outliers.
- Make our grid detection algorithm more robust.



Thank You

Git Repository

The project was named after my sister who was recently diagnosed with terminal cancer.