

# Studying the Formation of Biofilms and the Relationship of Genotype to Phenotype

## Discovering Quantitative Features

The biofilms below were subjected to different mutants and allowed to grow for five days. We want to understand the differences in expressed phenotypes (structure) to learn more about the underlying process that drives their formation. Historically, the research community uses qualitative descriptions to compare features. The goal of this project is to develop methods to extract quantitative features.

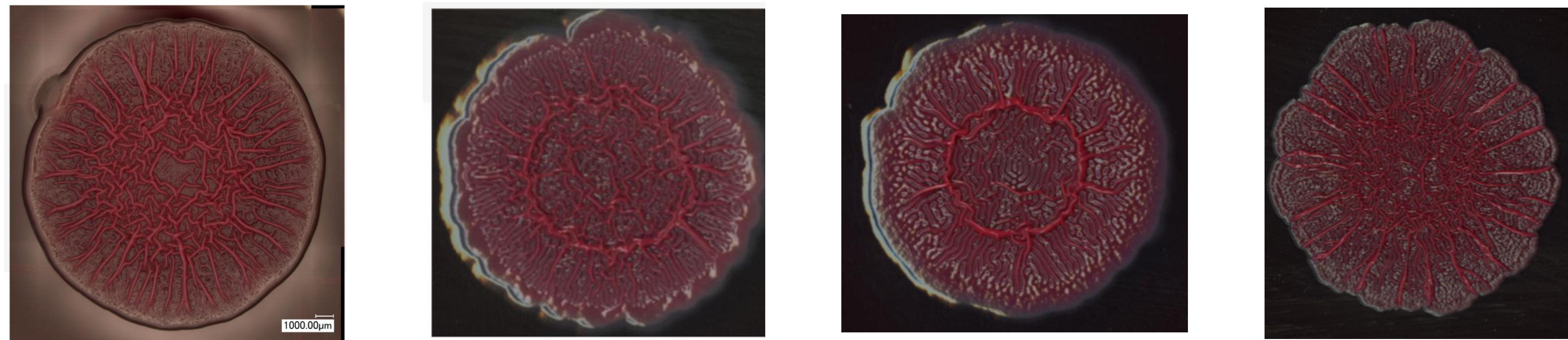


Figure 1. Example biofilms with varying characteristics

## Known Feature Extraction

This method involves quantifying features that are known to the biology team to be important. Namely, quantifying the phenotype by highlighting the structure within the biofilm and calculating “percent wrinkled”.

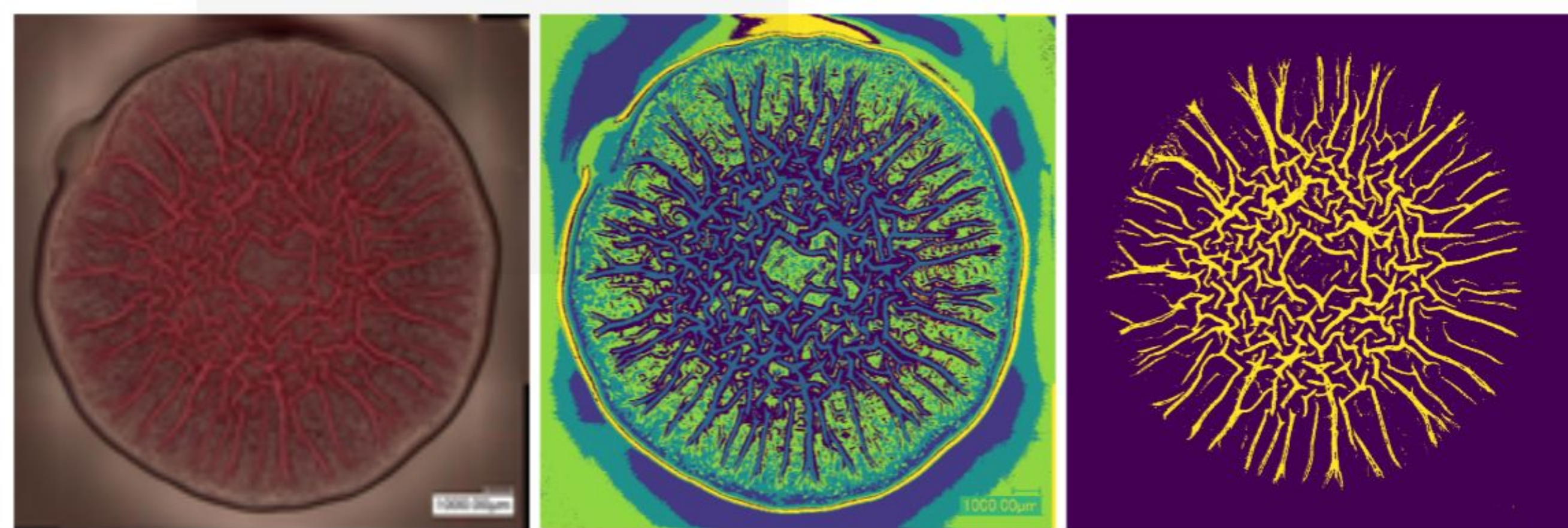


Figure 2. K-means clustering of pixel color value to identify wrinkles

Line detection is used to identify the existence and prominence of “spokes” in the biofilm structure, and to measure the size of the spokes.

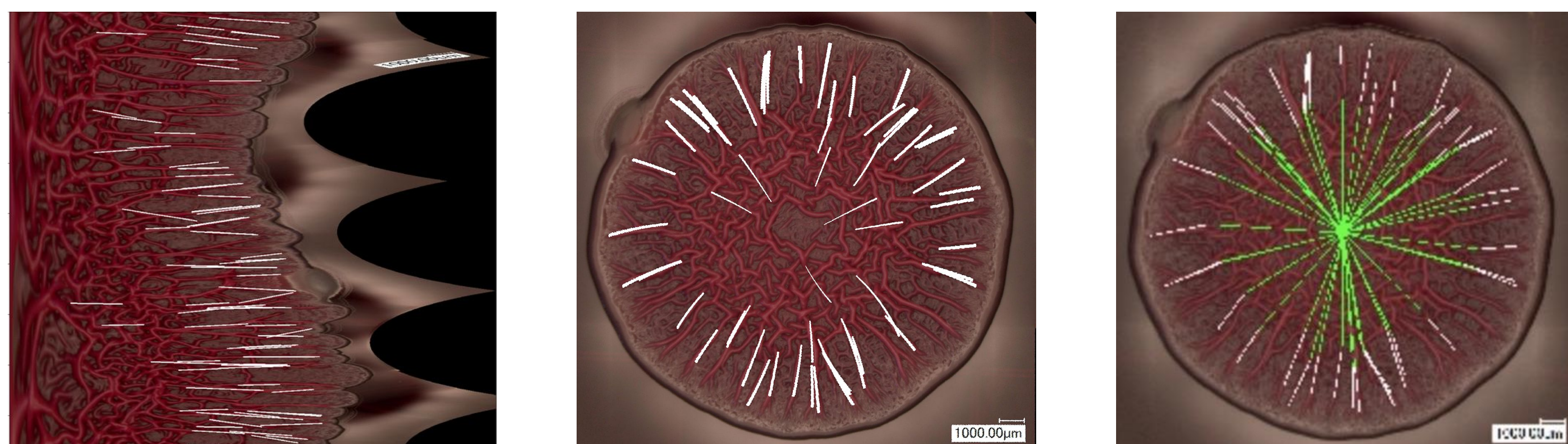


Figure 3. Number of spokes and ratio of spoke to clustered wrinkle

## Growth of a Biofilm

In addition to measuring the wrinkle after growing for five days, the growth pattern over time can be used to identify how different mutants trigger different growth patterns.

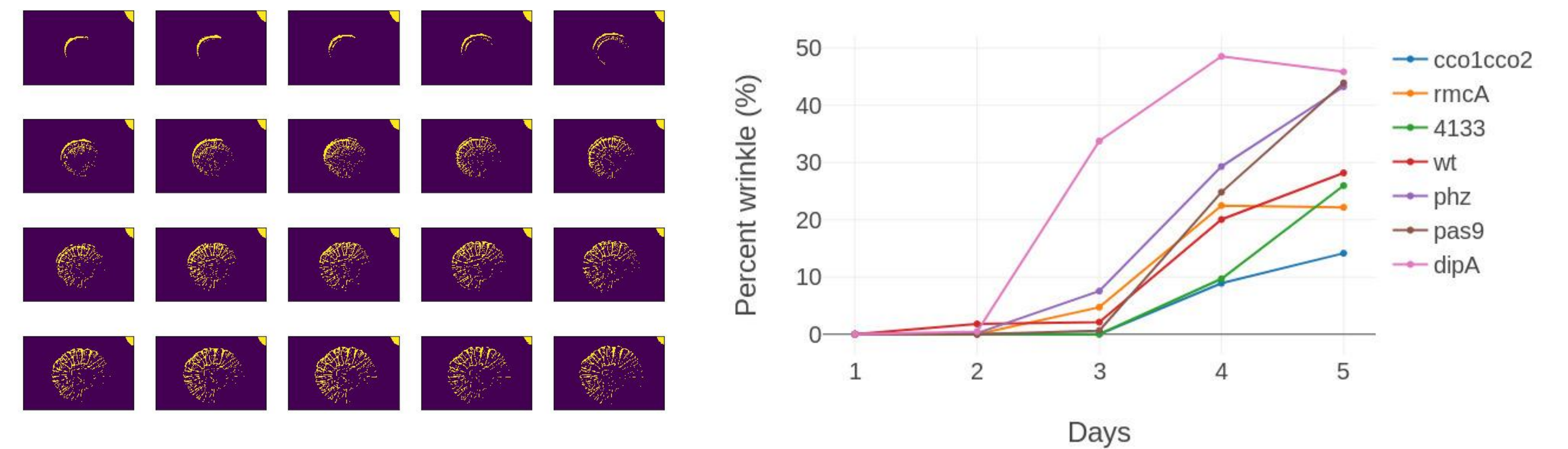


Figure 4. Wrinkle detected over time

## Data Representation and Dimension Reduction

Each image was resized to 128x128x3 pixel/color values. Representing all these images in a single visualization is valuable to understand similarities between images, and informative when developing image processing techniques for this data set.

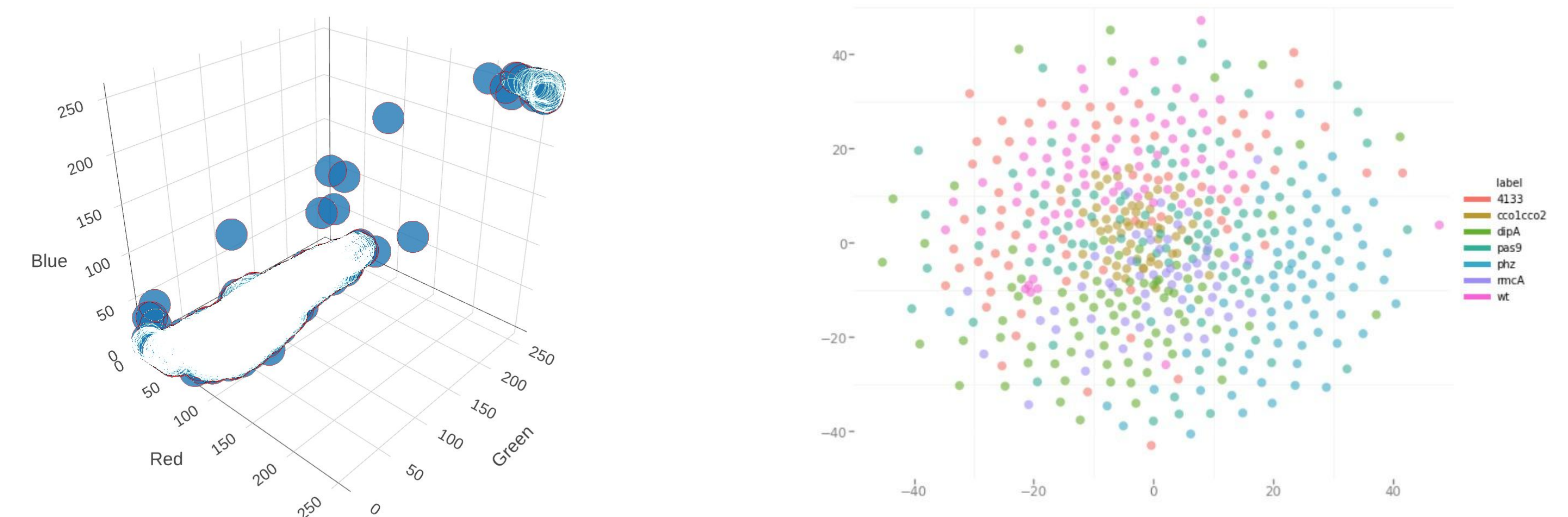


Figure 5. (left) RGB representation of each image (right) t-SNE dimension reduction colored by mutant

## Classification

As a final implementation step, we implemented an SVM on the pre-processed images. PCA was used to reduce dimension prior to training the model with 500 images in 7 classes. We achieve ~84% classification accuracy. Future projects may explore ways to improve this result.

## Outcomes

- Robust algorithms to process any biofilm image and quantify features
- Public code for researchers to build on
- Web application for lab researchers to use the algorithms on new images

## Acknowledgments

Thank you to Professor Lars Dietrich, Lisa Dahl, and Hannah Dayton for their support and guidance throughout the project. Thank you to Professor Andreas Mueller for feedback on progress reports