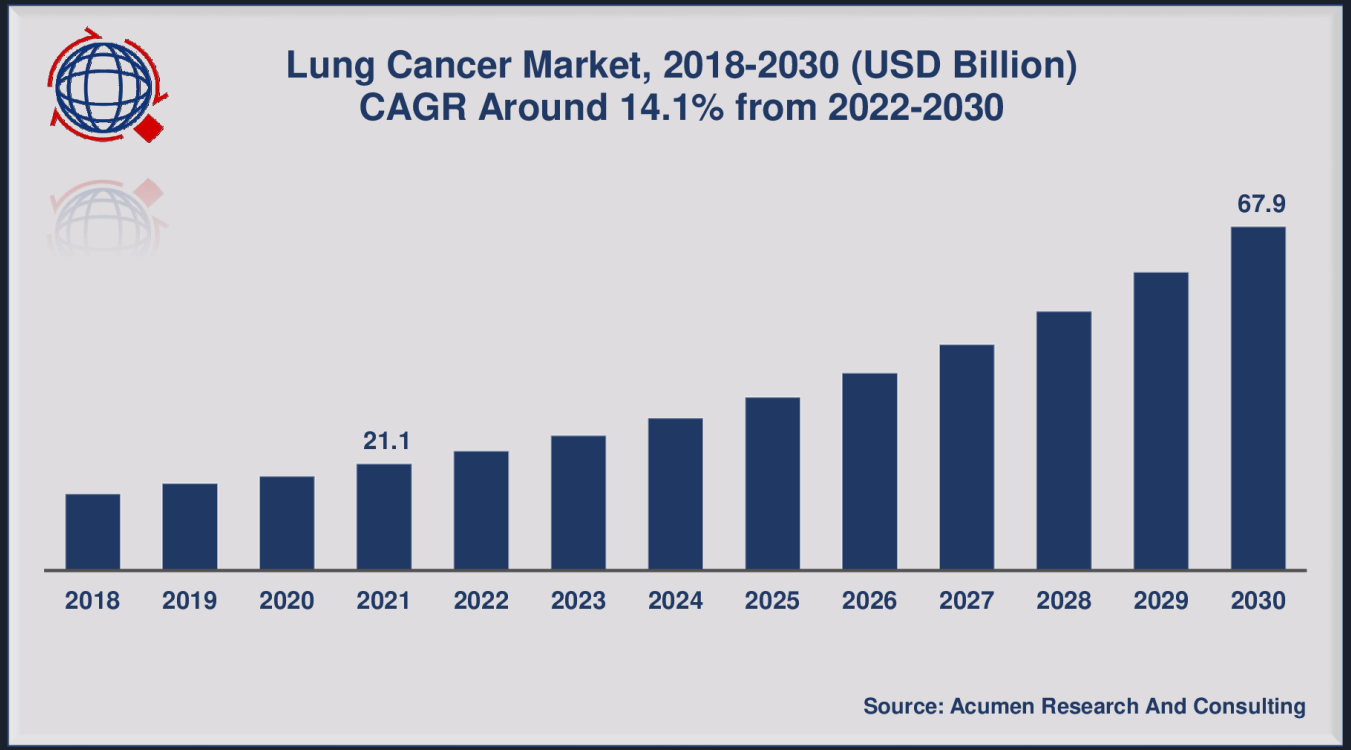
Classification and Tracking of Small Cell Lung Cancer Cell Populations

Matt Anikiej, Shreyasi Periketi

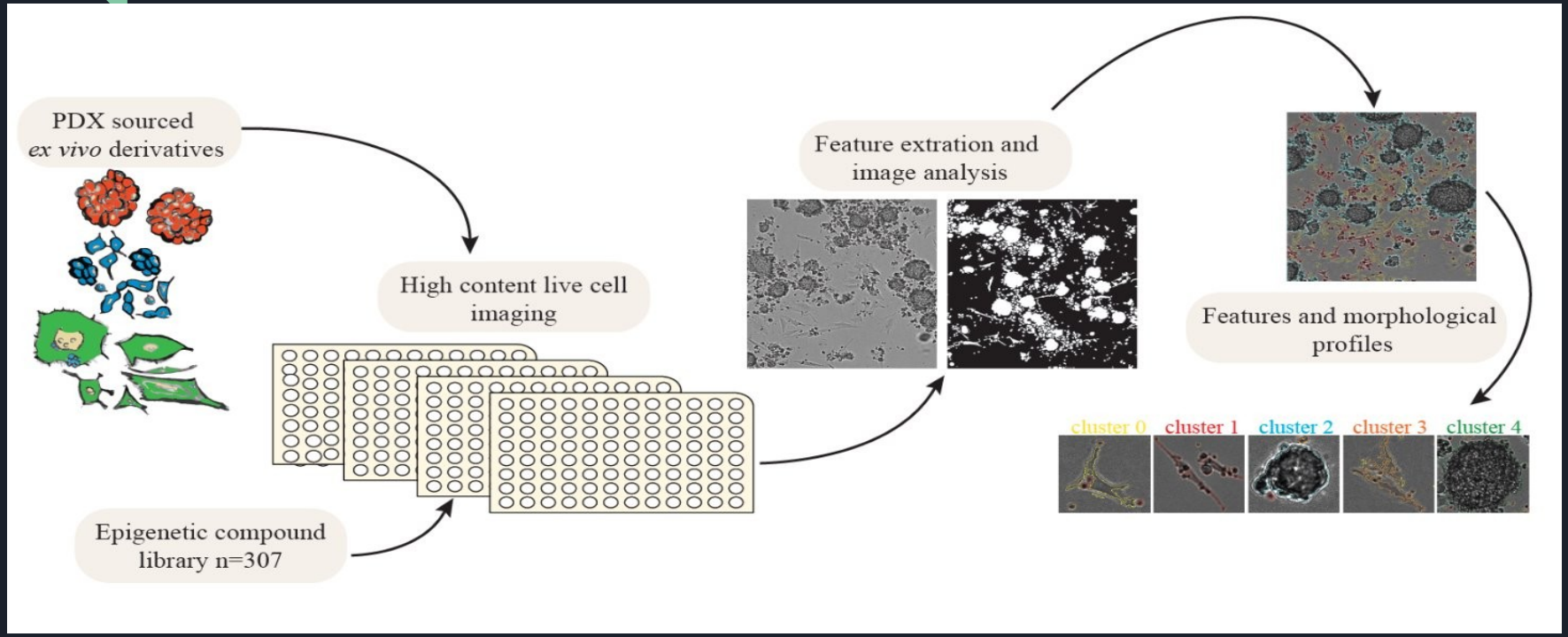
Problem

- Cancer is a devastating disease for anyone to get.
- Intratumoral heterogeneity of small cell lung carcinoma (SCLC) is a very aggressive tumor that has not been able to be studied well.
- Currently there is a cell clustering pipeline that exists with poor results, but it is promising and we can improve it further.

Incentive

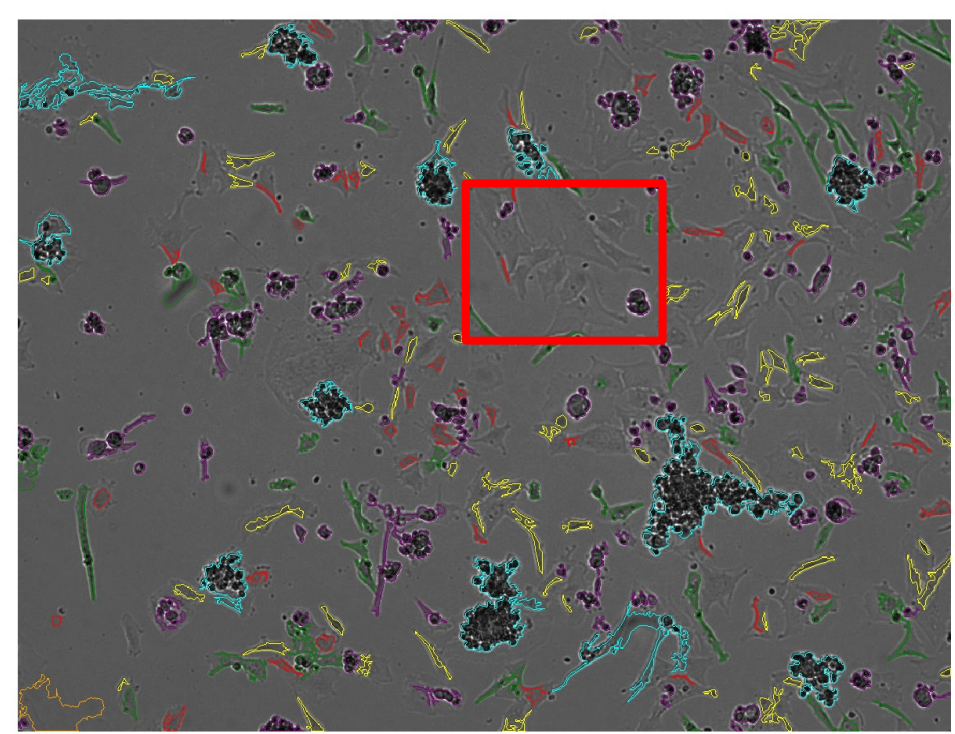

Lung Cancer Market, 2018-2030 (USD Billion)
CAGR Around 14.1% from 2022-2030
21.1
67.9
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
Source: Acumen Research And Consulting

Pipeline Architecture



Issues with Current Clustering

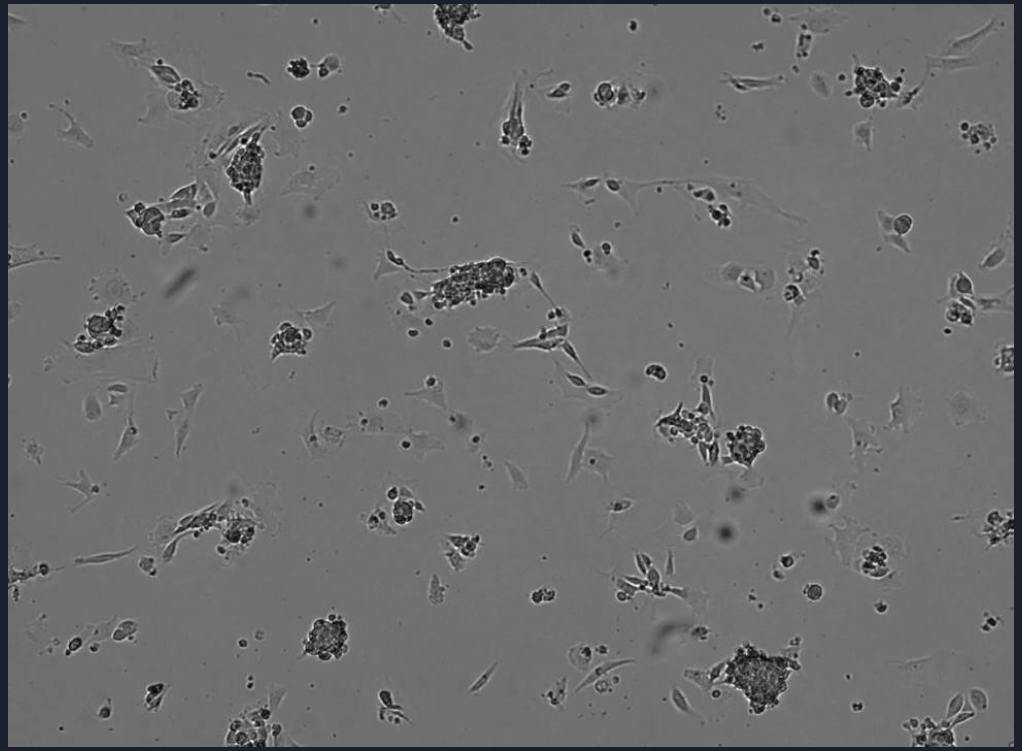
- Current masks filter out “flat” cells along with the background
- These cells are some of the fastest growing populations but become hard for the models to track because of this
- This leads to immense limitations in analysis and model performance



Impact

- Enhanced classification and tracking of SCLC cell populations can lead to better understanding and treatment of SCLC.
- Potential to improve patient outcomes and contribute to the development of targeted therapies.

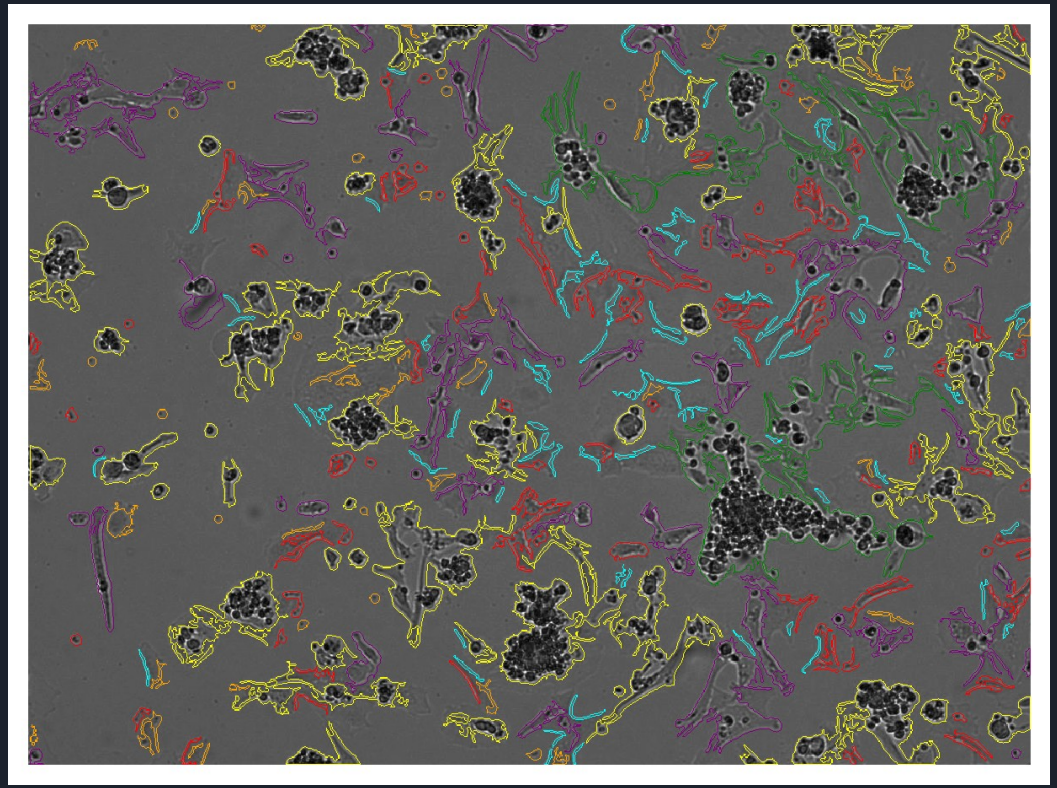
Data



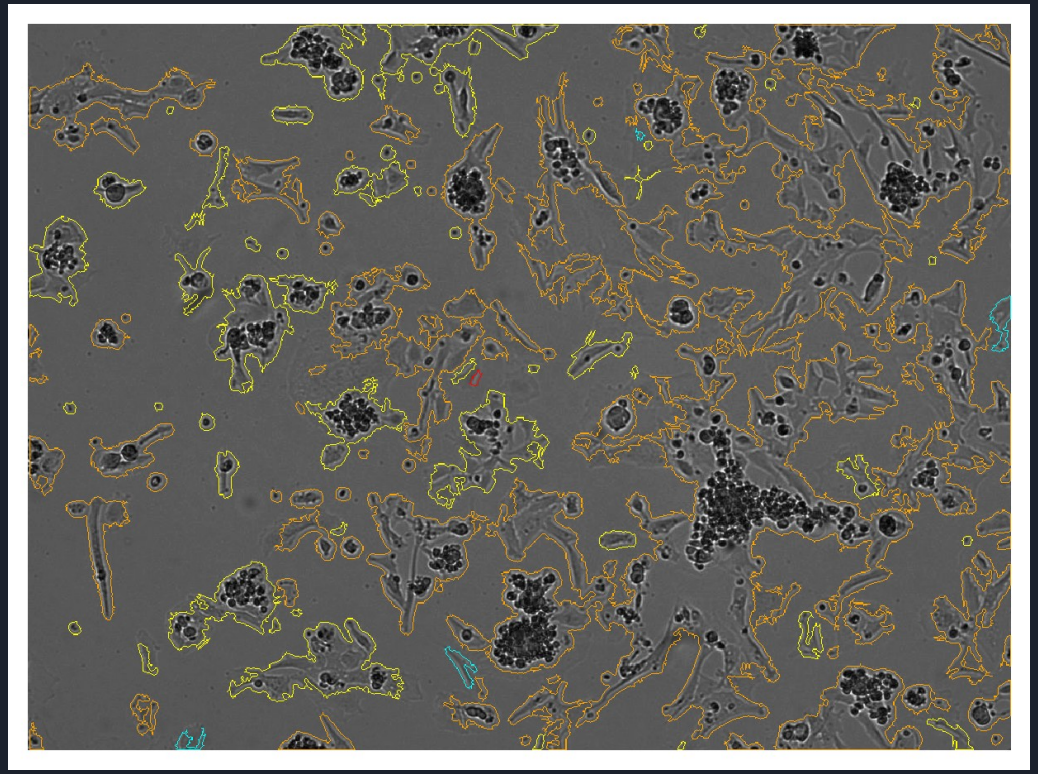
Solutions

- Median Filter - CV2
  - Will smooth out the edges and stop the capture of debris
- Rolling Ball Filter - SkImage
  - Detects edges much better than contrasting the image during background removal
- High Pass Filter - CV2
  - Buffs up the edges of the image so the background removal process is able to better detect them

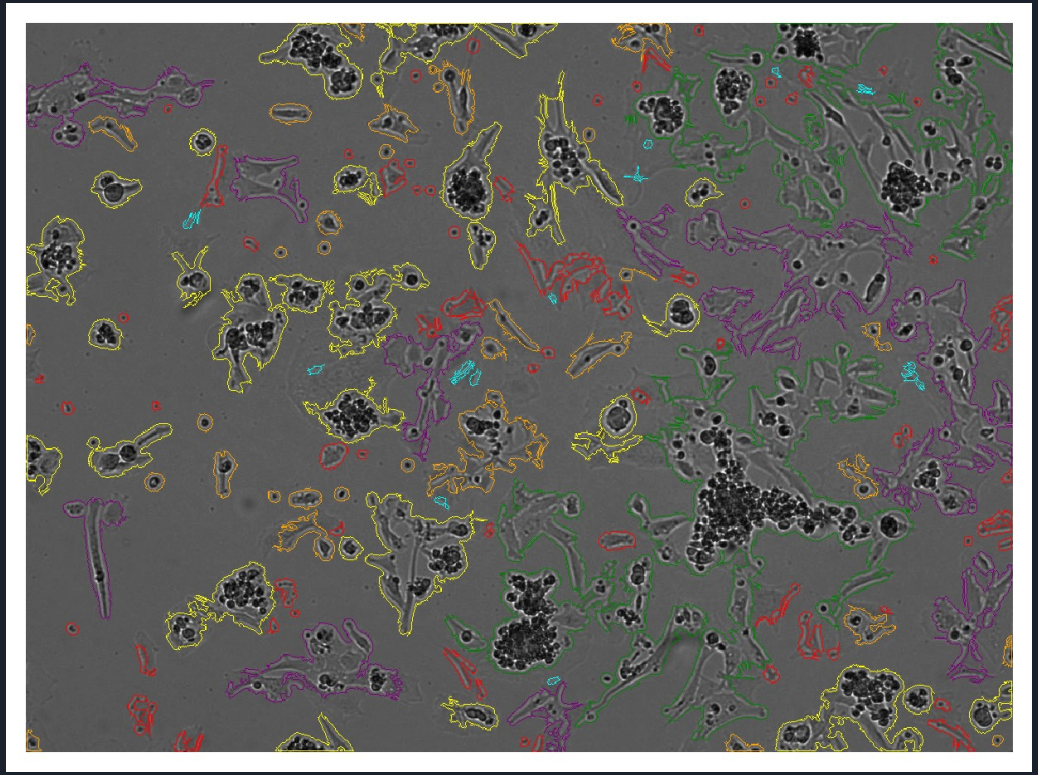
High Pass Filter Only



High Pass + Rolling Ball Filters

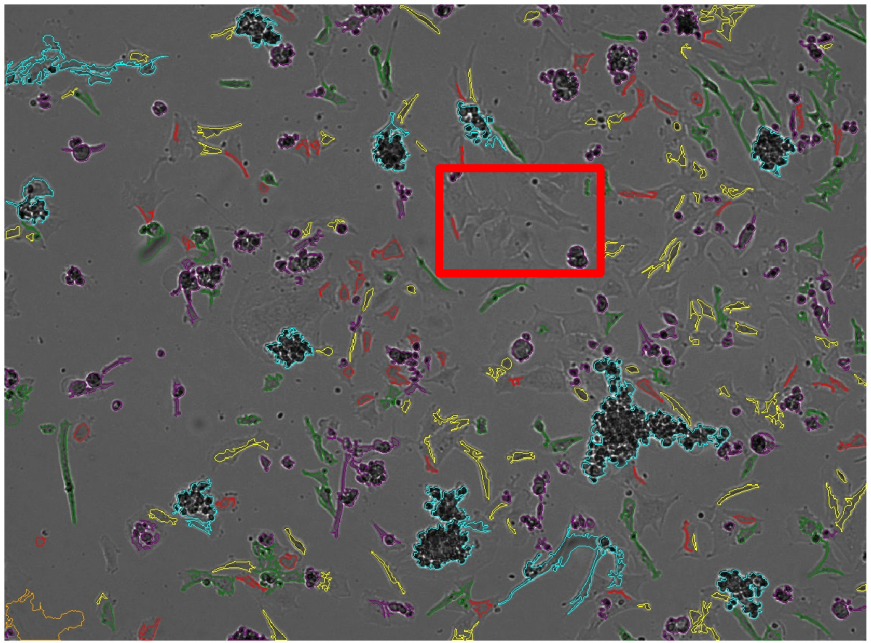


High Pass + Rolling Ball Filters + Smaller Kernel Size + Median Filter

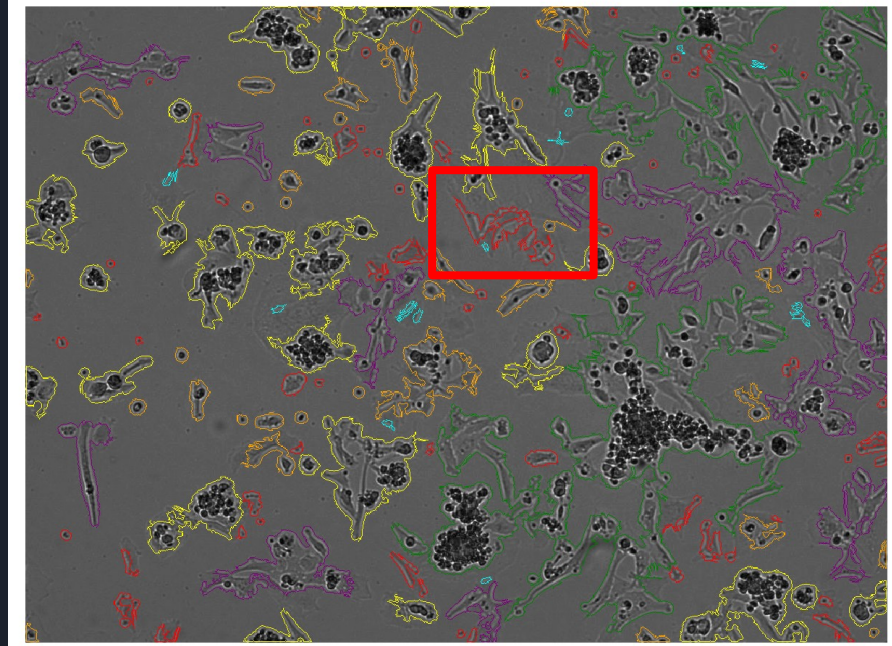


Solution Comparison

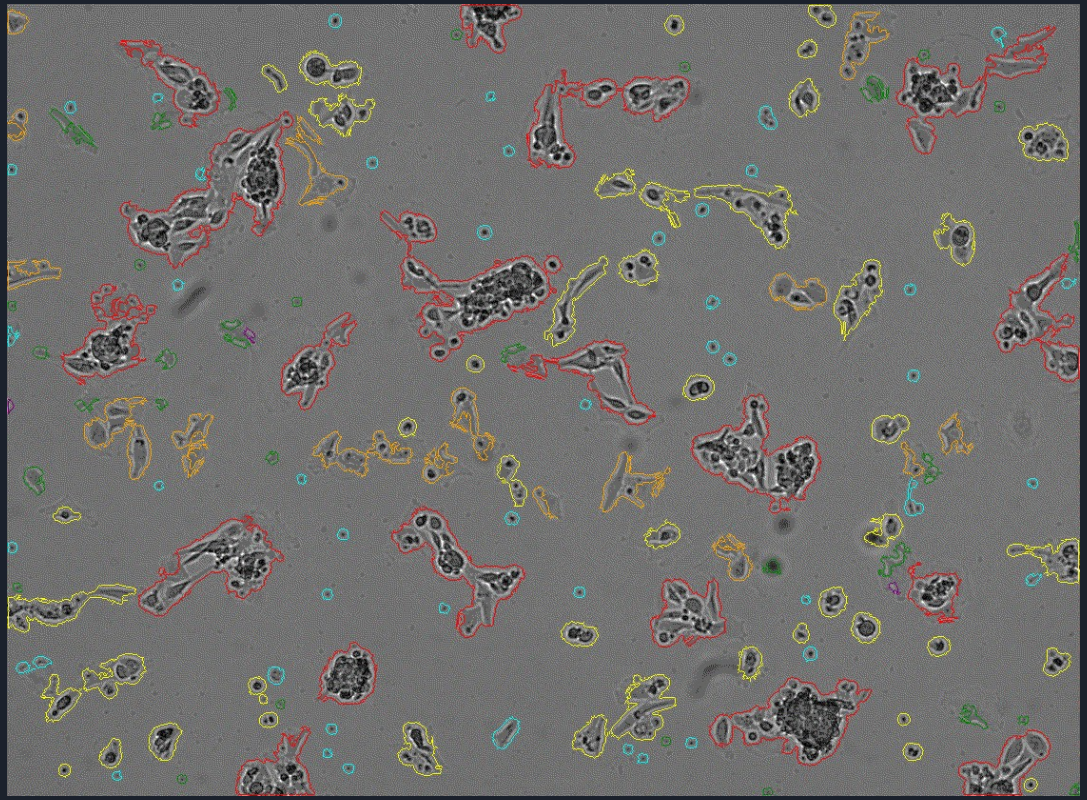
Previous



Current



Demo



Conclusion

- Flat cells are segmented at a much higher rate than previously
- The clustering performs much better than previously as well across all images in a stack
  - All features were deemed statistically significant compared to only the principle components initially used
- Kmeans was the best clustering algorithm for this data set
  - DBScan was also tested to no success
- Tends to fail as cell populations become very large and mixed
  - Masks overlap a lot of cells -> expected issue in cell research
- Deep learning is a future consideration
  - Labeling data is very expensive

Q/A