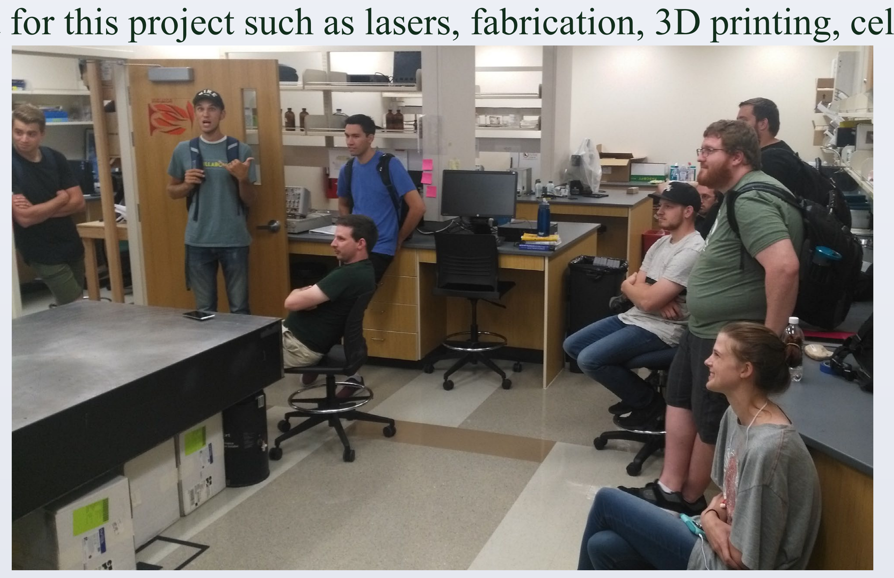
Use of Machine Learning for Non-Invasive Identification of Tumors

Mason Acree, Gunnar Jensen, Ryan Rainey, Vern Hart
Utah Valley University Physics Department |NASA Space Grant Consortium
Center for Imaging and Biophotonics Experiments Advancing Medicine (CIBEAM)

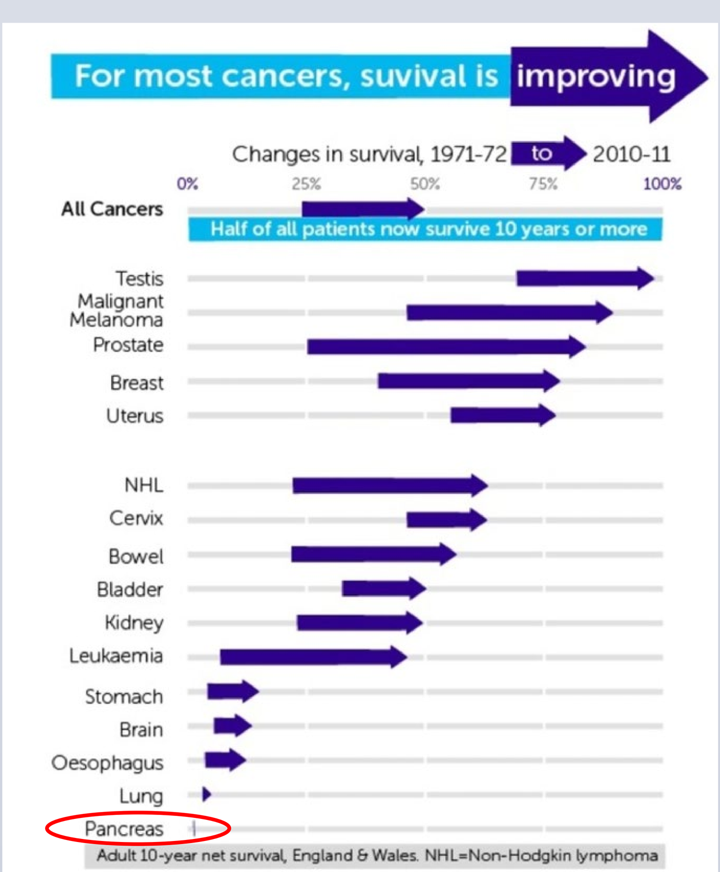
Introduction

Under Utah Valley University (UVU) Physics Department the Center for Imaging and Biophotonic Experiments Advancing Medicine. (CIBEAM) is focused on techniques for early cancer detection. The undergraduate based research project with the instruction of Dr. Vern Hart has developed a methodology using the scattering profile of a laser through a specimen to classify between different cancer cell types with machine learning algorithms. This technique works by passing a near-infrared (NIR) laser through a cell monolayer and collecting the scattering profile. Thousands of images are collected for the purpose of training a artificially intelligent convolution neural network as well as to build a testing dataset for verification. Additionally, this group has focused on involving many students in various research thanks to the many facets required for this project such as lasers, fabrication, 3D printing, cell growth, and machine learning.

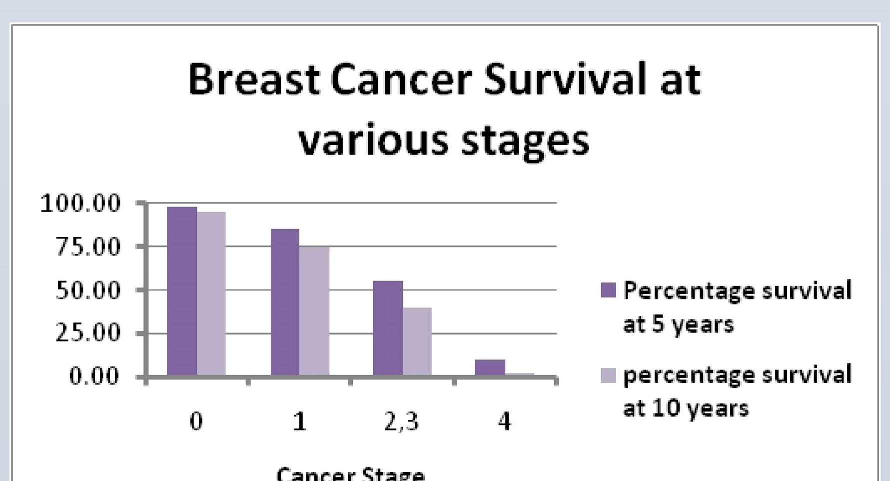


Background: Cancer Treatments

For many decades there has been a large amount of rigorous research towards cancer treatment. There are many different techniques for killing cancer once it's developed in the body. A few examples would be surgery, brachytherapy, radiotherapy, and chemotherapy. With these different techniques and many layers of sophistication for each one, the advanced medical expertise of modern cancer treatments has significantly increased survival rates for many different kinds of cancer. This can be shown in the second figure at the bottom, this demonstrates how survival rates have greatly increased. Although to help ensure cancer survival rates, there will need to be new techniques developed to detect cancer earlier in its development. The third figure demonstrates this idea, and the pancreatic cancer on the bottom of the second figure also has almost no change in survival due to the fact that the symptoms onset long after the initial development of the cancer, making it harder to find and treat earlier.



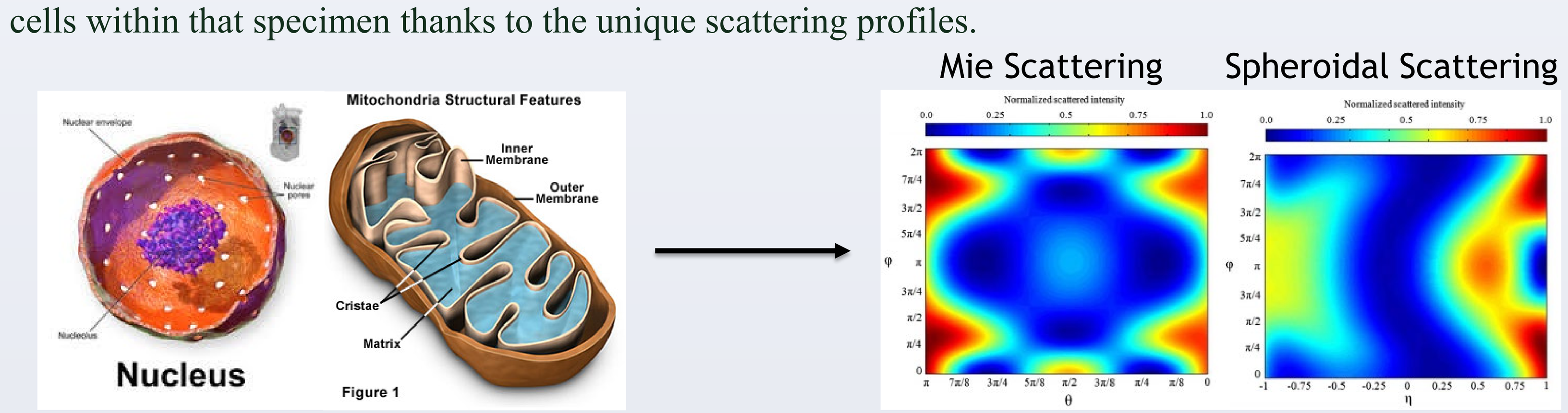
Adult 10-year net survival, England & Wales Credit: Cancer Research UK

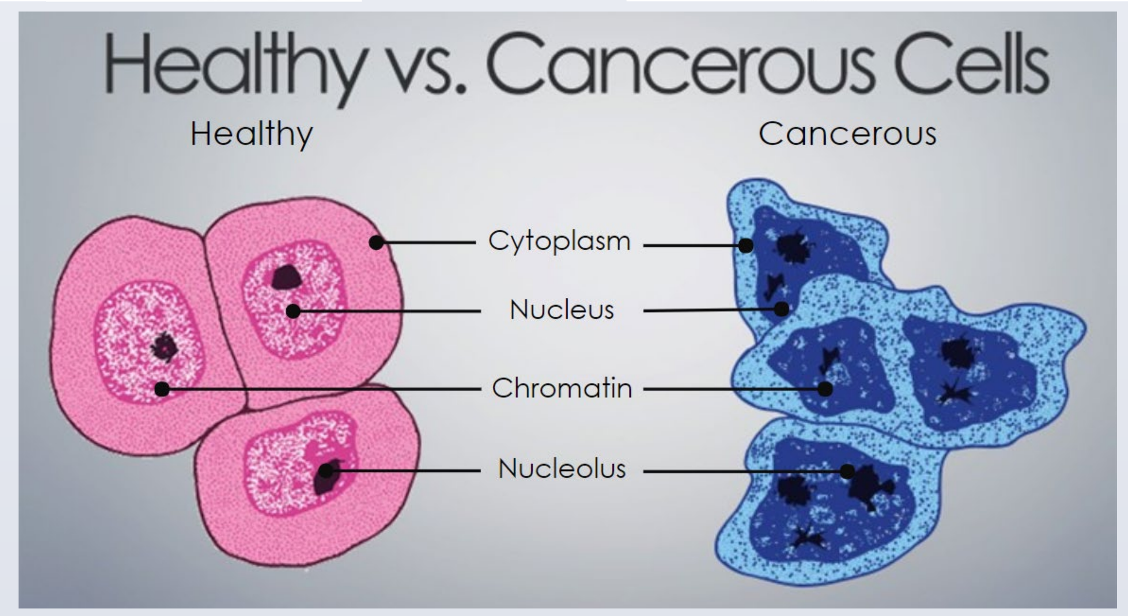


Credit: National Cancer Institute – Cancer.gov

Scattering and Interference

An important aspect of this technique is the physical properties behind it. Considering the understanding that light will scatter and diffract off of objects of proportional size we can use the understanding of how the scattering patterns should look for specific shapes of objects. Now considering the light passing through a collection of cells, the light is going to scatter off of primarily the Nucleus (spherical), Mitochondria (Spheroidal), and Cell Wall. Cancerous cells in contrast to healthy cells take up uniquely different structural shapes depending on the type, and are much less homogeneous than healthy cells. Thus, when a specimen has more or less of various scattering patterns we should be able to determine the type of cells within that specimen thanks to the unique scattering profiles.

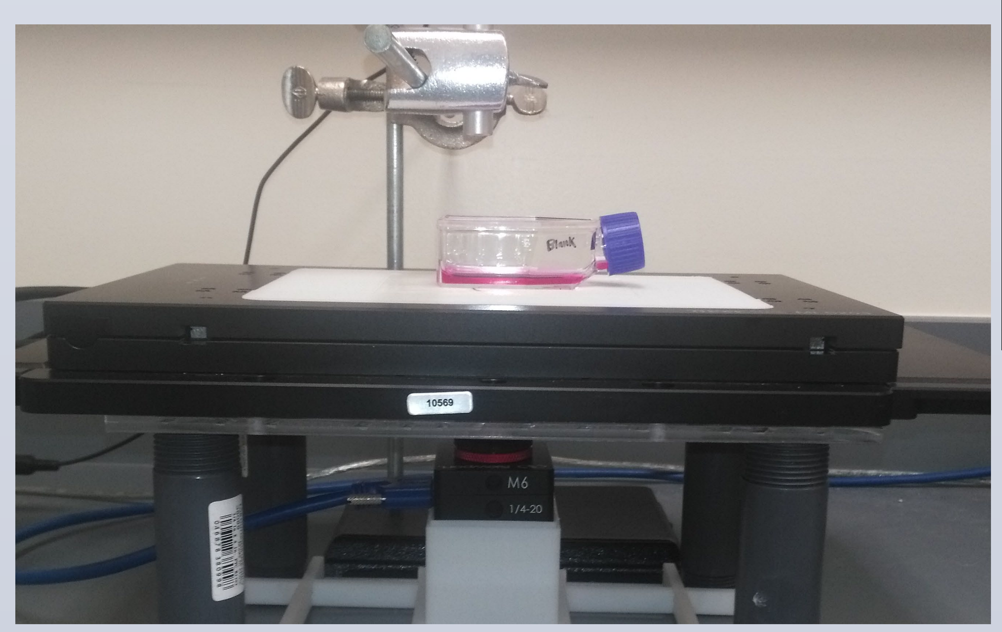




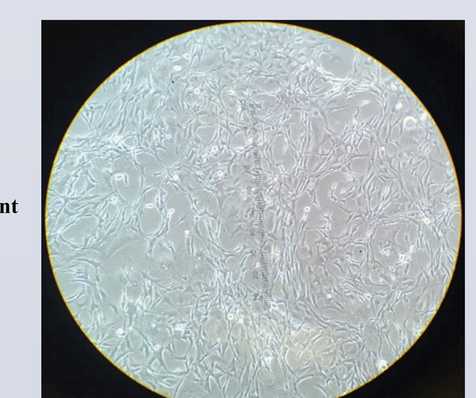
Methodology: Biology and Data Collection

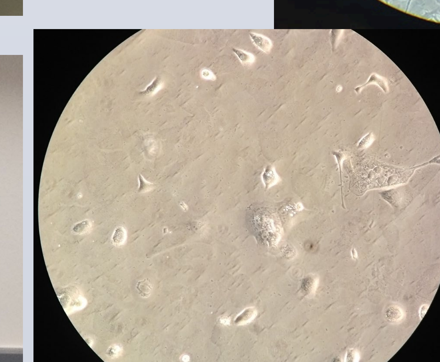
Based on the premise that we should be able to identify cells using the scattering patterns, the next step would then be to develop a model of the scattering as light interacts and scatters with a few hundred cells in a specimen. However, taking the complicated formulas that are used to model scattering with these different shapes and applying them to non perfect spheres and other various shapes with hundreds of interactions would be excessively complicated. Therefore we have used an end-to-end approach with machine learning. One problem: Machine learning requires DATA!

The biology students grow cells in flasks with a nutrient solution called Dulbecco's Modified Eagle Medium. The cells are fed and maintained until a clean cell monolayer has formed on the bottom, then they will image with our apparatus shown below.

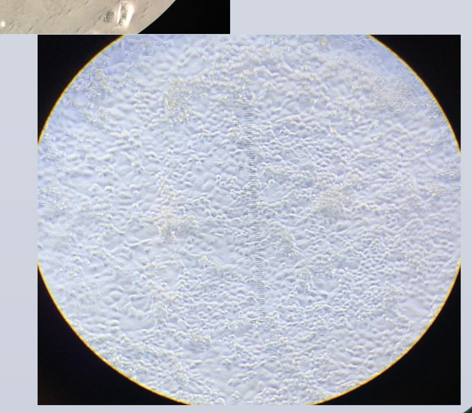


Uppsala 87 Malignant Glioma (U87)





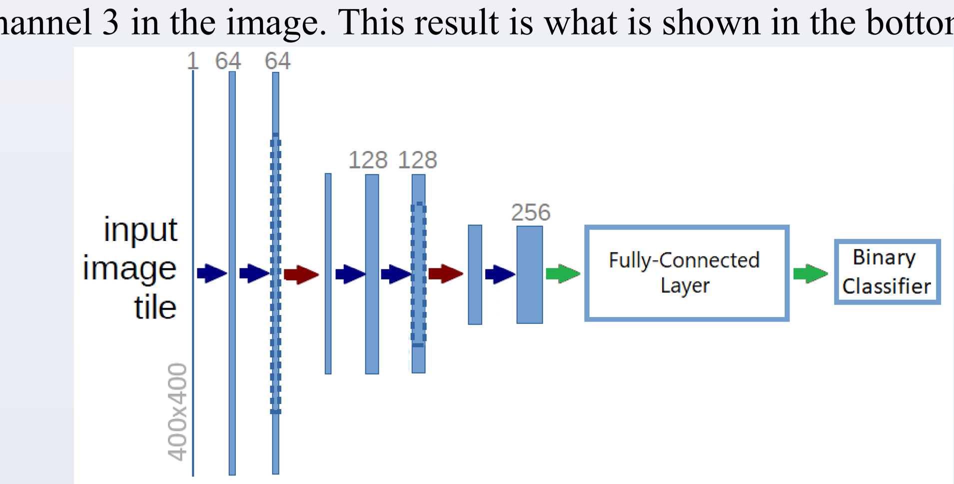
Primary Ductal Carcinoma (HCC70)

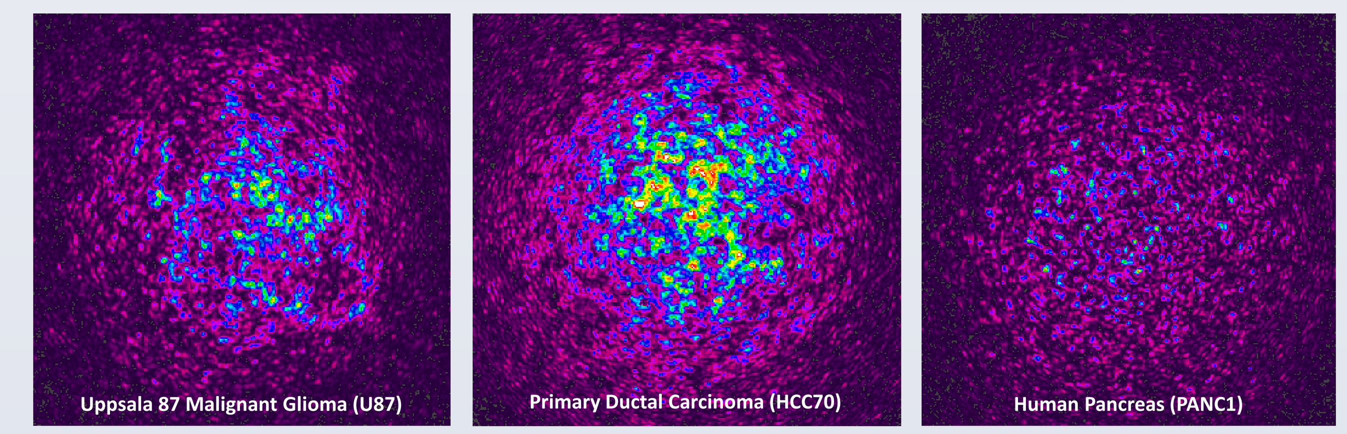


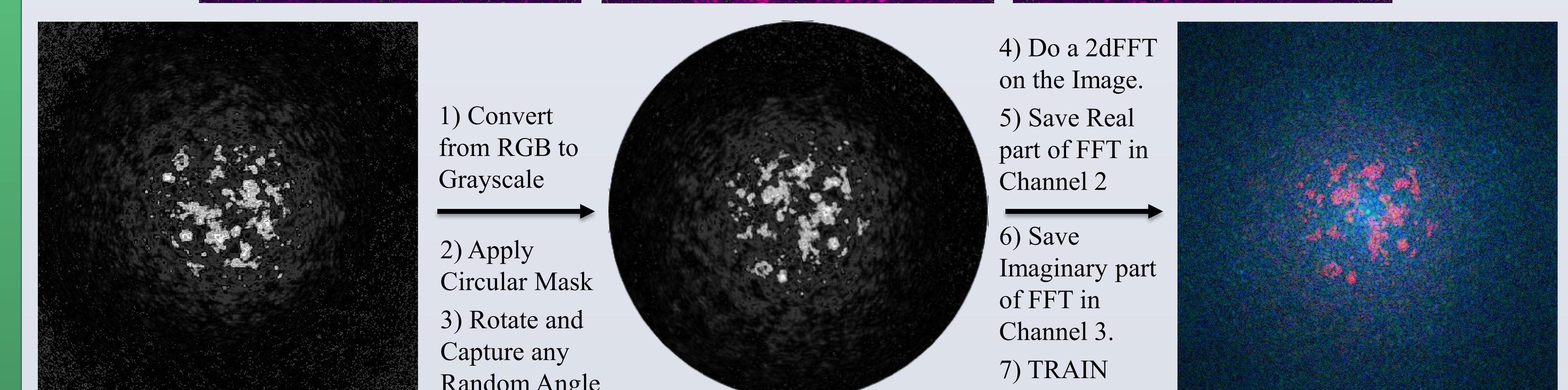
Human Pancreas (PANC1)

Convolution Neural Network & Augmentation

CNN's are a machine learning technique that is extremely effective at image classification. It uses several layers of different parameters to draw connections and make decisions based on different features within images. Another key aspect of our training process was by taking the images and augmenting our dataset. By applying a circular mask, and rotating a circular image at any random angle, we are able to augment any single image up to 30-60 or more depending on how rigorously we augment. The last step to preprocess our data is taking the FFT, then staking the real part of the FFT as channel 2 in the image, and the imaginary part of the FFT as channel 3 in the image. This result is what is shown in the bottom right colored image.



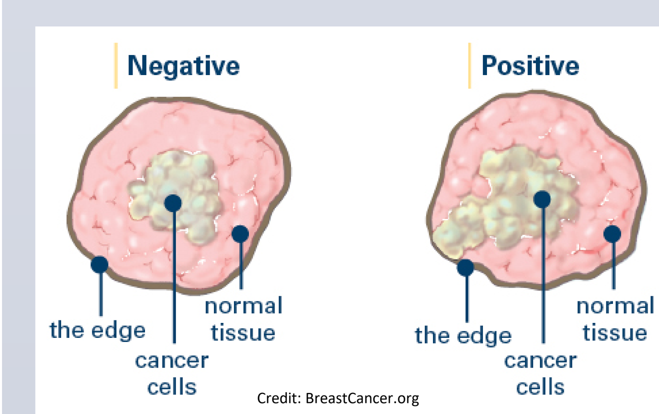




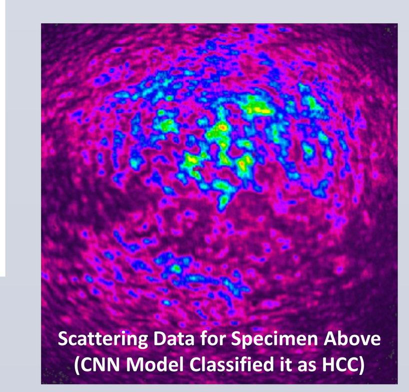
Results & Applications

Using our current methodology, we have been able to regularly train models that are able to predict the cell type in a specimen at high accuracies ranging from 87% to 94% depending on CNN, training data, and augmentation. However, when training on data when we've controlled for biological factors such as density, we have been able to acquire an accuracy of up to 98.5%! These results allow us to confidently say that there are deterministic features in the scattering patterns with this approach that should allow us to move onto more novel techniques. Such as using back scattering to detect cancer in tissues which can be used in margin mapping before pathology, or detecting at low concentrations (early detection). As well as possible clinical collaborations now that we've shown a proof of concept.

HUNTSMAN CANCER INSTITUTE UNIVERSITY OF UTAH



Credit: BreastCancer.org



Acknowledgements

I would like to thank Vern Hart for all the opportunities with research he has provided me and education given to get me here. As well as a thanks to the many handfuls of biologists and their help with cell growth and data collection. Finally, a thanks to the UVU-NSF S-Stem Scholarship that supports me as a student so that I have the ability and time to put the many hundreds of hours into this project.