Understanding Breast Cancer Disparities in African American Women

Hector Cruz¹, Dr. Zeynep Madak-Erdogan²,³, Ashlie Santaliz-Casiano²
¹REU INCLUSION, ²Division of Nutritional Sciences, ³Food Science and Human Nutrition Department, University of Illinois, Urbana-Champaign

ABSTRACT
The survival rate for breast cancer (BC) is around 80% in developed countries, this drops to 60% in middle-income countries and to 40% in low-income countries. This is because of how expensive it is to diagnose it. BC mortality rate for African American (AA) woman is about 40% higher when compared to Non-Hispanic White (Caucasian) woman. This could be due to socioeconomic status differences or genetic and other non-genetic factors. We suspect that we can find differences in molecules when trying to predict BC in AA compared to Caucasians. To do this, we are using Python to be able to pass the data through some machine learning models and from there get an accurate prediction.

METHOD
Some of the data is missing, so the first thing we did was refill the data that was missing with the minimum. Next is to verify which molecules to use for the prediction we opted for the ExtraTreesClassifier model.

After getting those correlated molecules we run them through another model Logistical regression to get a probability on the prediction.

RESULTS
After running the model a couple of times, we got a somewhat acceptable accuracy.

Biomarkers
This are six of the 20 biomarkers from this specific run.

Accuracy

Future Work
• Change the algorithm of the ExtraTreesClassifier for one that is not random.
• Experiment different ways to fill up the data.
• Improve accuracy.

ACKNOWLEDGEMENTS
• Dr. Zeynep Madak-Erdogan, REU INCLUSION mentor
• Ashlie Santaliz-Casiano, REU INCLUSION co-mentor
• Ms. Olena Kindratenko, REU INCLUSION co-PI and program coordinator

Aim
Identify Biomarkers to predict BC utilizing machine learning models.