Endometrial carcinoma is a type of uterine cancer that begins in the endometrium lining within the uterus. It accounts for 90% of total uterine cancer patients. It is also the 4th most common cancer in women in the US. Approximately 66,000 cases per year with a fatality rate of ~20% (13,000). Endometrial carcinoma can be easy to detect if there are signs of abnormal bleeding, otherwise it is undetectable without specific testing. Endometrial carcinoma occurs as multiple subtypes, which have different pathology and should have different treatment plans, but targeted molecular medicines for each subtype are not the standard of care yet. We previously found that two subtypes of tumors, endometroid tumors and serous tumors which have different DNA methylation landscapes that could provide new approaches for therapies.

DNA methylation landscapes are often altered in cancers, and when there is extensive gain of DNA methylation, it is coined CIMP, “CpG island methylator phenotype”. There are three subtypes of CIMP: CIMP positive, CIMP negative, and CIMP intermediate. If a sample’s DNA is high in CpG island hypermethylation then the subject is considered CIMP+ while if a subject’s DNA has low or no CpG island hypermethylation the sample is considered CIMP-. CIMPi indicates a moderate level of hypermethylation. Tumor suppressor and DNA repair genes are frequently silenced in CIMP+ cases. Being able to correctly classify a patient’s tumor sample could drastically shift the recovery process. It has previously been shown that it is possible to determine CIMP tumors based on a limited set of 89 DNA methylation sites, which is useful when tumor methylation data is available.

To test our hypothesis, we are examining whether DNA mutations in endometrial tumors could be used to predict CIMP, if so, they will be useful when DNA methylation data is not available for analysis. We broke down our hypothesis into three aims, the first being to find the important mutations relating to CIMP. The evidence is a result of mining data from The Cancer Genome Atlas, which is the most comprehensive cancer genomics study to date. The data used consists of 250 unique samples. 108 samples are CIMP+ while the other 142 samples are CIMP-. There are 8085 total mutations. Of the 8085 mutations, 739 demonstrated a strong correlation to either the CIMP+ or the CIMP- grouping. With these mutations we can, on average, correctly classify a sample with 87% accuracy. We achieve this by randomly separating the samples into two groups. Then analyzing how well the first group can correctly classify, just with mutations in the samples, the second group of samples. This process is repeated ten times then averaged to remove any particularly unique combinations that would act as outliers. With this procedure we were able to show a strong relationship between the mutations in a sample with the CIMP phenotype in endometrial carcinoma.

There are two other aims to my research, finding the important mutations, as described above, was the first aim, next is to find the relationships between mutations. This can take form in many different statistical methods, but the goal will be to build strong correlations between any two mutations; This leads to the last aim. The last aim is to interpret the findings from a biological perspective. This would take form in various biological analyses to determine a linkage between biological pathways, cancer, and CIMP.