

Eric, R, Espinal
Undergraduate
The University of Alabama, Tuscaloosa, Alabama

Mentor: Chindo Hicks, Ph.D.
Louisiana State University Health Sciences Center, Department of Genetics

“Integrative Genomics Approach to Biomarker Discovery in Colorectal Cancer”

Background: Despite remarkable progress in screening and patient management, colorectal cancer (CC) remains a major public health problem. CC is the third most commonly occurring cancer in men and the second most commonly occurring cancer in women. World-wide there were over 1.8 million new cases of CC in 2018. In United States it is estimated that there will be 147,950 newly diagnosed cases of CC and an estimated 53,200 individuals will die from the disease in 2020. Therefore, a critical unmet and urgent medical need is discovery of molecular markers for early detection of the disease. The recent surge of next generation sequencing technology have enabled generation of vast amounts of gene expression and somatic mutation data on CC. These advances have enabled molecular classification of subtypes and increased our understanding of the molecular taxonomy of CC. However, gene expression has not been optimally leveraged and integrated with somatic mutation information for the discovery of diagnostic markers. The objective of this investigation was to discover clinically actionable biomarkers for diagnosis and prognosis of CC using gene expression and somatic mutation data. Our working hypothesis was that genomic alterations in individuals diagnosed with CC and control samples could lead to measurable changes distinguishing patients diagnosed with CC from controls. **Material and Methods:** We addressed this hypothesis using gene expression and somatic mutation data derived from a total of 523 samples (481 CC samples and 42 control samples) from the Cancer Genome Atlas (TCGA). The data was partitioned into two data sets tumor samples and control normal. We performed analysis comparing gene expression levels between the two sample groups to discover a signature of significantly differentially expressed genes distinguishing tumors from controls. Significantly differentially expressed genes were evaluated for the presence of somatic mutations to identify a signature of significantly differentially expressed genes which were also significantly differentially mutated distinguishing the two sample groups. **Results:** The analysis revealed a signature of 100 highly significantly ($p < 1.00 \times 10^{-7}$) differentially expressed genes distinguishing individuals with CC from controls. Evaluation of these genes for the presence of somatic mutations revealed a signature of 80 significantly differentially expressed genes which were also differentially mutated distinguishing the two sample groups. Among the top somatic mutated differentially expressed genes distinguishing the two samples groups included the genes ATP1A1, PIGR, FCGBP, MYH11, PTPRF, CDH17, MYH14, AHNAK, FLNB, and CSDE1. **Conclusion:** We discovered a signature of somatic mutated differentially expressed genes distinguishing patients diagnosed with CC from controls. Our investigation demonstrates that integrative analysis combining gene expression with somatic mutation data is a powerful approach to discovery of molecular diagnostic and prognostic markers in CC.