Colorectal adenocarcinoma is the 3rd most diagnosed cancer and accounts for 9% of all cancer deaths; it is encoded by complex interactions between the genome, transcriptome, and epigenome. To enable more effective treatment and diagnosis, these interactions must be characterized. The Cancer Genome Atlas (TCGA) identified 24 sources of mutation (some novel). Here, a computational model mapping interactions between the “omic” components of colorectal cancer to understand different pathogenetic scenarios in colorectal cancer was built.

TCGA data from COSMIC (genome) and Gene Expression Omnibus (transcriptome and epigenome) was used. After processing, functional enrichment and pathway analysis was conducted using Gene Ontology and KEGG. Dimensionality Reduction and Multivariable Regression Analysis were employed to construct the networks with the help of Arboretum (Python), visualized in Cytoscape, based on the multifactorial expression data underlying mucous and sigmoid adenocarcinomas and carcinoid tumors. An algorithm analogous to the Hyperlink-Induced Topic Search was employed to elucidate expression patterns within the network.

After this model was built ($R^2 = 0.85$), several hub genes were screened out. CEACAM5, is the primary hub gene, present before colorectal cancer differentiates into adenocarcinoma or sigmoid tumors, and has 95 connections. Other hub genes, with 46 connections on average, included MLH1, MSH2, KRAS, HRAS, CTACNB1, CTNNB1, ERCC1, RUNX3, and SNAI1. The average disease score for these hub genes was 2.24. The mean neighborhood connectivity degree and topological coefficient were 3054.43 and 0.485, respectively. Understanding the key genetic players and interactions in colorectal cancer is important to detect and treat colorectal cancer early.