Evaluating DECODE with Colorectal Cancer Genome Samples

Stephen DeBenedittis (Mentor: Dr. Khanh Dinh, 2026 IICD SRP)

DECODE (Deciphering Cancer Origin from DNA Evolution) is a novel subclonal reconstruction algorithm created in the Dinh Lab, characterized by its ability to accurately detect neutral mutation tails and reliably detect mutation clusters in the Site Frequency Spectra (SFS). Due to its novelty, DECODE's applications in distinct oncological contexts are currently untested, leaving gaps in our understanding of its utility.

We aim to elucidate the use of DECODE on paired colorectal cancer samples at diagnosis and metastasis, collected at Columbia University Irving Medical Center, with regard to its robustness against differing data quality. We investigate differences in DECODE's results on progressively filtered subsets of colorectal cancer data, and also assess the viability of using DECODE to glean insights about patient outcomes and tumor characterization based on predetermined factors.