Cancer results from an evolutionary process where somatic mutations occur and accumulate in a population of cells. There are many types of mutations that can cause cancer. A mutation that causes genetic variation at a single genomic site is called a SNV. The different lineages which comprise a tumor are known as clones, and the phenomenon of clonal admixture is known as intratumor heterogeneity. The sequencing and analysis of tumors has revealed extensive intratumor heterogeneity in cancers. The ability to genetically profile a tumor would improve physicians' ability to tailor treatments according to the subpopulations and mutations.

We view the problem of assigning mutations to clones as a general machine learning clustering problem. Our method models this process as an infinite binomial mixture model, with a nonparametric Dirichlet Process prior for the number of clusters. We learn the parameters of the model using variational inference, which is a fast, scalable, deterministic algorithm that has been shown to have advantages over classical MCMC techniques in terms of convergence and accuracy. The results obtained from implementing this model are better or on par with existing state of the art clonal inference techniques.