Abstract of "Computational detection of driver mutations in cancer genomes" by Hsin-Ta Wu, Ph.D., Brown University, May, 2016.

Cancer is caused largely by the accumulation of somatic mutations throughout an individual's life. Recent advances in next generation sequencing (NGS) enable measurement of somatic mutations in a cohort of samples. Cancer sequencing projects like The Cancer Genome Atlas (TCGA) have generated numerous somatic mutations in thousands of tumors. This thesis addresses two challenges. The first challenge is to distinguish the handful of *driver* mutations that are responsible for cancer development from the multitude of *passenger* mutations that play no role in cancer in a cohort of samples. The second challenge is to accurately identify larger genomic variants, also known as structural variants (SV). Although several computational methods have addressed this challenge using NGS technologies, they are limited by the underlying NGS data, resulting in a significant amount of SVs remaining undetectable, particularly in highly repetitive regions of the genome. To address the first challenge, we present two computational methods. First, we introduce a combinatorial approach for the problem of identifying independent and recurrent copy number aberrations, which are gains and losses of large genomic regions ranging in size from a few kilo-bases to whole chromosomes. We show that our method outperforms other methods on simulated data and performs well on TCGA cancer datasets. Second, we introduce a statistical model to identify combinations of *driver* mutations that are mutual exclusivity, a pattern expected for mutations in cancer pathways. Our model is more sensitive in detecting combinations of lower frequency mutations and outperforms other methods on simulated and real data. To address the second challenge, we present a method to identify SVs in a tumor more accurately by utilizing a new sequencing technology from 10X Genomics. 10X Genomics linked-reads technology uses barcodes to label reads originating from a longer DNA molecule, which enables the construction of synthetic long reads. Our method uses both signals from paired-end reads and synthetic long reads to identify SVs. We demonstrate that our method shows high sensitivity and specificity on simulated and real data. Together, these novel algorithms help address the challenge of characterizing and identifying *driver* mutations in cancer genomes.