Reconstructing Cancer Genome Organization from Paired End Sequencing

Layla Oesper*†, Sarah J. Aerni‡, Ryan Drebin§and Benjamin J. Raphael¶

Abstract

Motivation: A cancer genome is derived from the normal germline genome through a series of somatic mutations. Somatic structural aberrations – including duplications, deletions, inversions, translocations, and other rearrangements – result in a cancer genome that is a scrambling of intervals, or "blocks" of the normal genome. We present an efficient algorithm for the reconstruction of a cancer genome using alignments of DNA sequence reads from the cancer genome to a normal reference genome.

Method: We assume that a cancer genome has been sequenced using a paired end, or mate-pair approach, and that the resulting paired reads are aligned to the reference genome. From these alignments, we derive a partition of the reference genome into intervals. We also derive an estimated copy number for each interval using depth of coverage by aligned reads, and determine adjacencies between intervals from clusters of discordant paired reads. From this information we formulate the Copy Number and Adjacency Genome Reconstruction Problem of determining the cancer genome whose organization is most consistent with the derived intervals, adjacencies, and copy numbers. We design an efficient algorithm to solve this problem by reducing it to a network flow problem on a corresponding interval-adjacency graph. The solution to the network flow problem results in an Eulerian (multi) graph, containing an Eulerian tour that corresponds to a possible cancer genome that is consistent with the sequencing data.

Results: We apply our algorithm to ovarian cancer genomes that were sequenced as part of The Cancer Genome Atlas. We identify numerous structural aberrations in these genomes and reconstruct rearrangements that suggest the occurrence of breakage/fusion/bridge (B/F/B) cycles, a particular mechanism of DNA repair and duplication.

*Brown University, Providence, RI, USA. E-mail: layla@cs.brown.edu
†Corresponding Author
‡Stanford University, Palo Alto, CA, USA. E-mail: saerni@stanford.edu
§Brown University, Providence, RI, USA. E-mail: ryan@drebin.com
¶Brown University, Providence, RI, USA. E-mail: braphael@cs.brown.edu