Profiling Complex Somatic Rearrangements in Cancer Genomes using Nanopore Sequencing

Ayse Keskus[1], Tanveer Ahmad[1], Ataberk Donmez[1], Yi Xie[2], Isabel Rodriguez[2] , Rose Milano[2], Nicole M. Rossi[2], Hong Lou[3], Michael Dean[2], **Mikhail Kolmogorov**[1]

[1]Center for Cancer Research, NCI, Bethesda, MD, USA; [2]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA; [3]Leidos Biomedical Research, Inc., National Laboratory for Cancer Research, Frederick, MD, USA;

ABSTRACT

Structural variations (SV) are large genomic alterations from simpler indels to complex events involving multiple breakpoints and sequence gain/loss with established importance in various diseases including neurodevelopment diseases and cancer. Recent pan-cancer studies with large cohorts revealed the rich landscape of structural variants (SV) some of which shown to be involved in tumorigenesis and prognosis through direct modification of coding sequence or deregulation from copy number alterations, enhancer hijacking. However, a substantial part of the SV in the cancer genome is yet to be discovered since even the best short-read-based SV calling methods can reach 30-70% sensitivity partially due to mapping ambiguities. Long-read sequencing can overcome these limitations of short reads, however the current methods were not designed for the analysis of rearranged cancer genomes with complex copy number profiles.

We developed **B**reakpoint **G**raph **A**ssembler (BGA), utilizes long-read assembly and breakpoint graph frameworks to increase the detection capacity to cover SVs including complex events. BGA detects split reads from previously aligned bam files then builds a breakpoint graph that characterizes the structure of derived cancer karyotypes. Complex events with multiple breakpoints form connectivity clusters and are classified based on the subgraph properties. BGA also takes advantage of phased haplotypes and can incorporate multiple related datasets (such as in tumor-normal comparison or multi-site tumor sampling).

We first analyzed three cancer cell lines (HCC1954, H2009, and COLO829) and corresponding matching normals for the somatic SV-calling. In each cell line, we identified 8-56 breakpoint sub-graphs with more than one breakpoints. In HCC1954, a BRCA1 mutant cell line, exhibits the highest number of SVs including a large inverted-insertion with a duplication event in chr1q arm hosting FCRL4 gene. COLO829 showed the lowest number of somatic rearrangement clusters (n=8), including insertion and inversion events between chr3, chr10, and chr12 within RARB, BICC1, and TRHDE genes. H2009, we identified a inverted insertion of chr7 fragment to chr16 region hosting CDH13. We also analyzed three other cell lines with known HPV integration (CaSki, SCC152, SNU1000). In each of them, we observed complex clusters of HPV-HPV and HPV-human breakpoints that formed cycles, suggesting extrachromosomal amplification. The HPV fragments had many interactions with chromosomal DNA in CaSki and SCC152, but not in mostly episomal SNU1000. We also observed karyotype-scale changes that did not involve HPV sequence, such as the simultaneous exchange of six chromosome arms of chr2, chr7, and chr17 in CaSki.

BGA is freely available at: *https://github.com/KolmogorovLab/BGA*.

Breakpoint Graph Assembler (BGA)

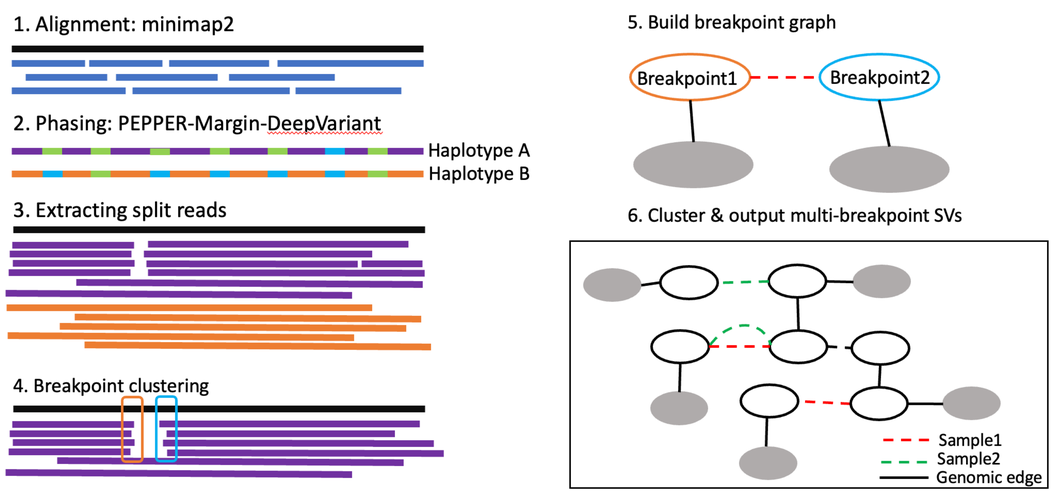

Figure 1: Workflow of Breakpoint Graph Assembler.

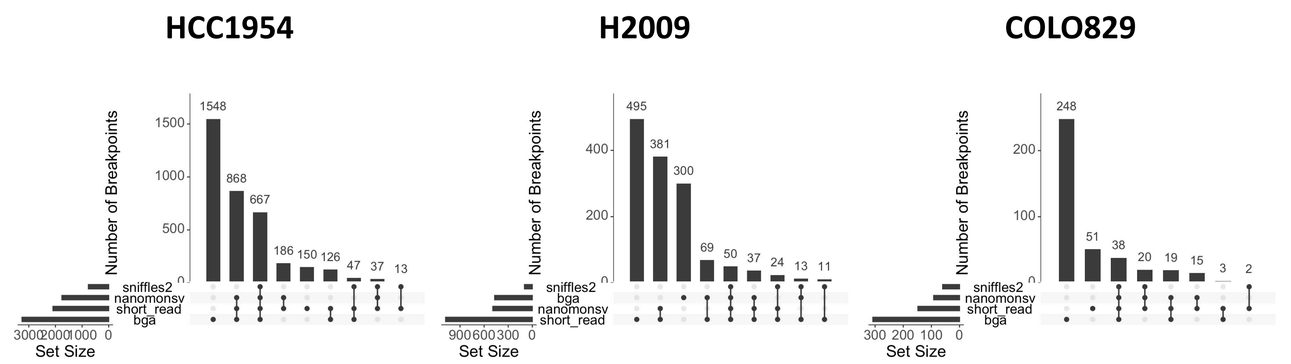

Figure 2: Overview of somatic breakpoints identified by BGA and their comparison with long-read; sniffles2 and nanomonsv and short-read SV calling methods; Manta, SVaBA, Delly, breakdancer in HCC1954, H2009, and COLO829. Breakpoints supported by at least three short-read based method are in short-read consensus.

Inverted-Insertion + Duplication – HCC1954

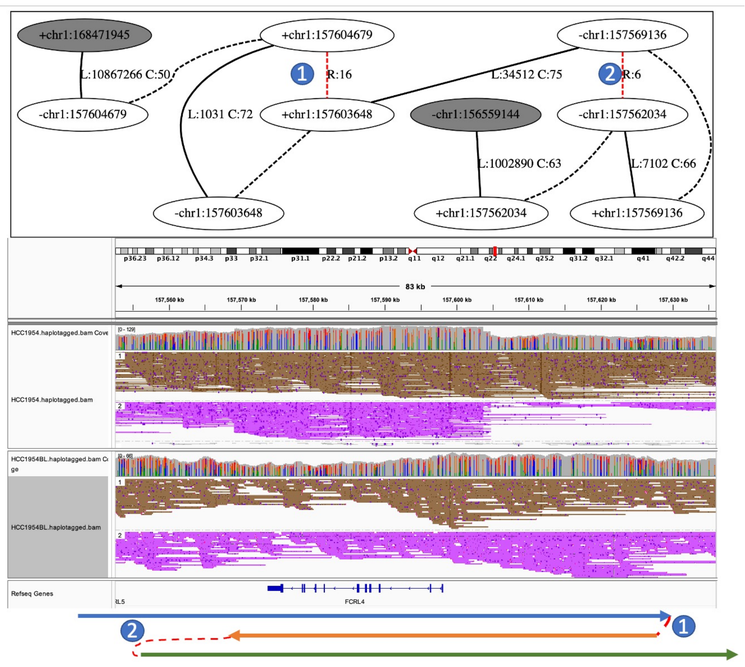

Figure 3: BGA graph for haplotype-specific inverted-insertion and duplication event in HCC1954 cell line (A). IGV representation of the SV shown in the BGA graph (chr1:157,556,743-157,609,968) in HCC1954 and HCC1954BL cell lines . Haplotypes are shown in purple and orange (B). Schematic representation of the SV(C).

Insertion + Inversion– COLO829

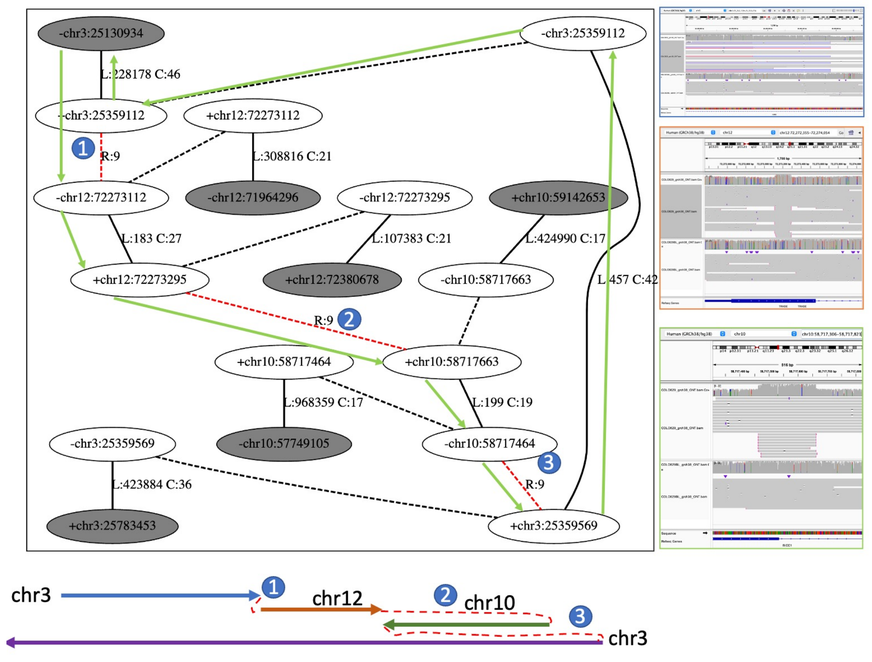

Figure 4: BGA graph for intrachromosomal insertion with inversion with chr3, chr 12 and chr 10 (A). IGV representation of the SV shown in the BGA graph in COLO829 and COLO829BL cell lines. (B). Schematic representation of the SV. Each breakpoint is labelled in accord with the BGA graph (C).

Insertion + Inversion– H2009

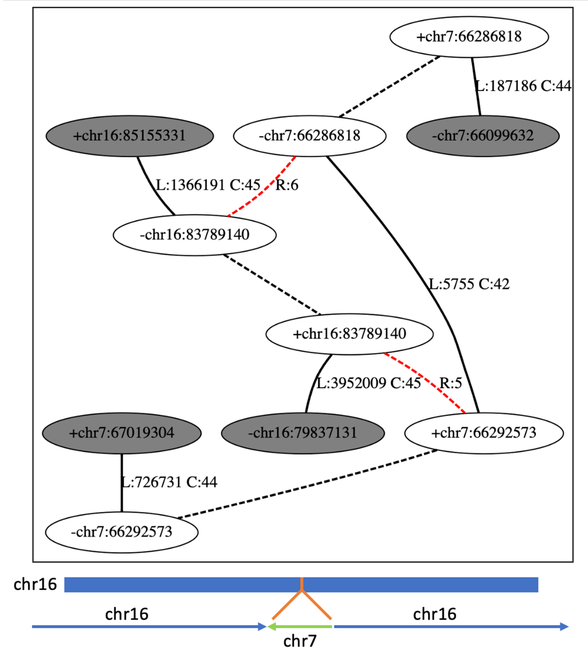

Figure 5: BGA graph for inverted-insertion event in H2009 cell line (A). Schematic representation of the SV. Each breakpoint is labelled in accord with the BGA graph (B).

Nonreciprocal translocation in CaSki

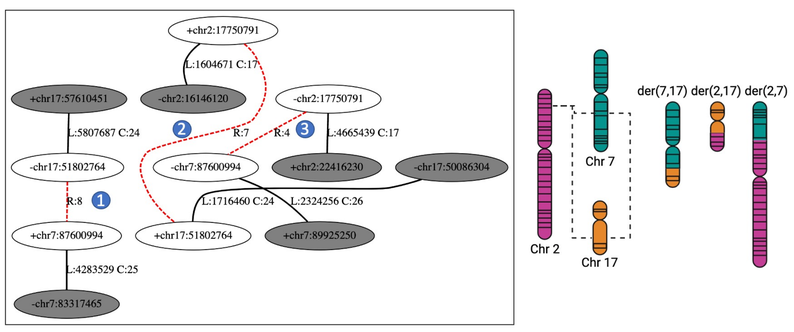

Figure 6: BGA graph for non-reciprocal translocation events in CaSki cell line (A). Schematic representation of the SV(B).

HPV integration in SCC152

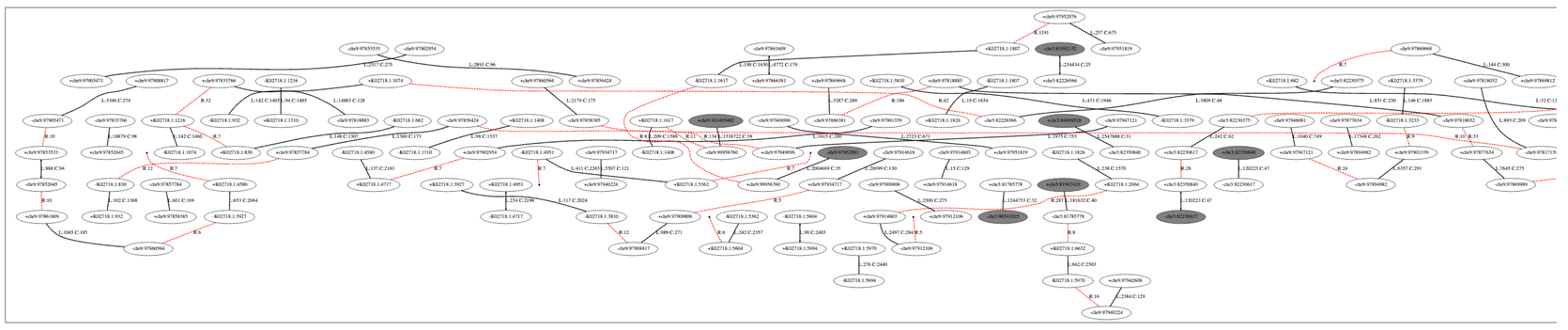

Figure 7: A partial BGA graph for HPV integration events in SCC152 cell line forming complex clusters of HPV-HPV and HPV-human breakpoints that formed cycles, suggesting extrachromosomal amplification.