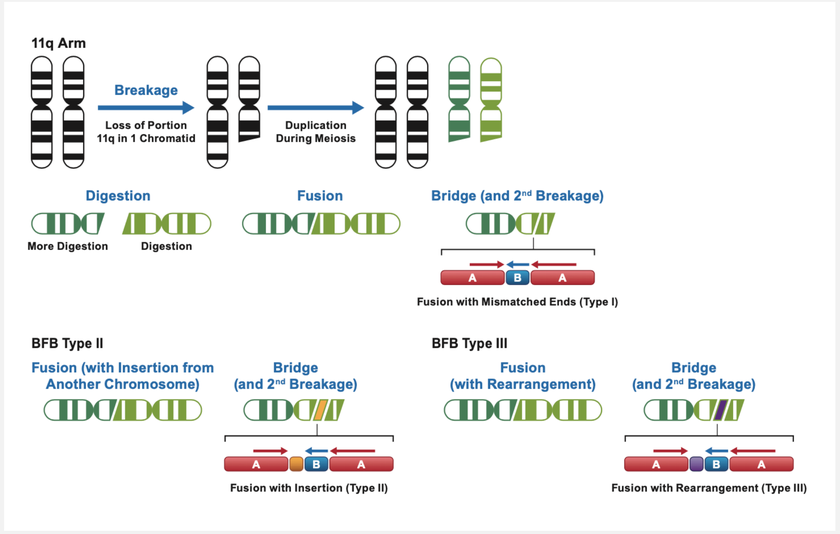
Long-read sequencing of cervical cancer cell lines reveals digestion of DNA ends during Breakage-Fusion-Bridge events in cancer

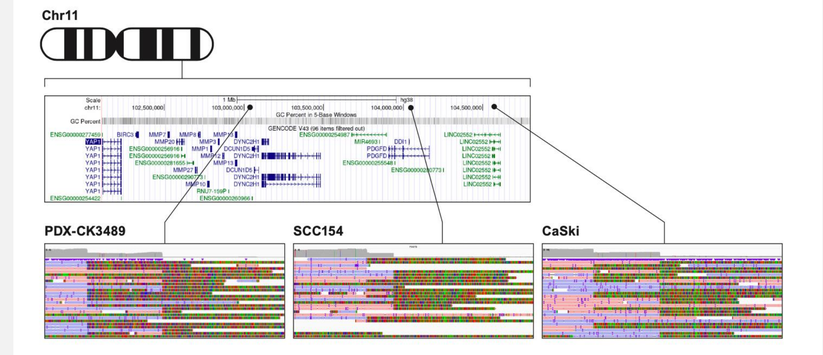
Michael Dean[1], Isabel Rodriguez[1]*, Nicole M. Rossi[1]*, Yi Xie[1]*, Ayse Keskus[2], Tanveer Ahmad[2], Hong Lou[3], Jesica Godinez Paredes[1], Rose Milano[1], Nina Rao[1], Joseph F. Boland[3], Wen Luo[3], Jia Liu[3], Tim O'Hanlan[2], Yuko Yuhki[3], Laksh Malik[4] , Kimberley Billingsley[4], Cornelis Blauwendraat[4], Mary Carrington[3], Meredith Yeager[3], Lisa Mirabello[4], Mikhail Kolmogorov[2]

Introduction:

Cervical cancer ranks fourth in cancer prevalence and mortality worldwide, with significant disparities in incidence and mortality associated with poverty. Currently there are no targeted therapeutic approaches for treatment. To provide improved models for cervical cancer therapeutics we characterized 22 cell lines using long-read DNA and RNA sequencing. Cervical cancer cell lines have several recurrent chromosomal alterations, and structural variation analysis of long-read sequences revealed telomeric deletions associated with DNA inversions consistent with breakage-fusion-bridge cycles. Analysis of the DNA sequences at the inversion site revealed staggered ends consistent with exonuclease deletion of the ends after breakage and before fusion (Breakage-Digestion-Fusion-Bridge, BDFB). Events involved the insertion of sequences from another chromosome or local rearrangement, indicating that BFB is complex. The chr11q BFB event, with *YAP1/BIRC2/BIRC3* amplification, is more common in cervical cancer than other cancers.

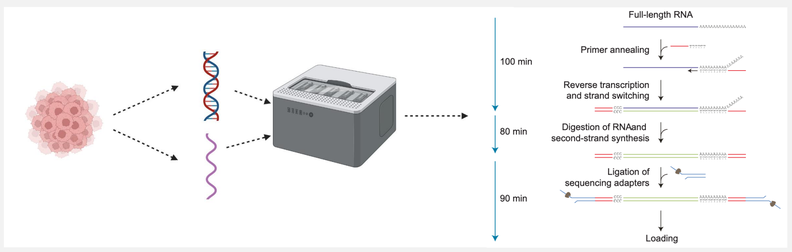


Model for BDFB types.
Model for a BDFB Type I event on chromosome 11q. A telomere deletion results in a pair of deleted chromatids. The free ends are subject to exonuclease digestion, and uneven digestion generates staggered ends. Fusion results in a lower copy number (one half) of the B segment. Type II events are formed when a segment from another chromosome (orange) is inserted at the junction, and Type III events involve insertion of a sequence derived from the flanking sequence (purple)



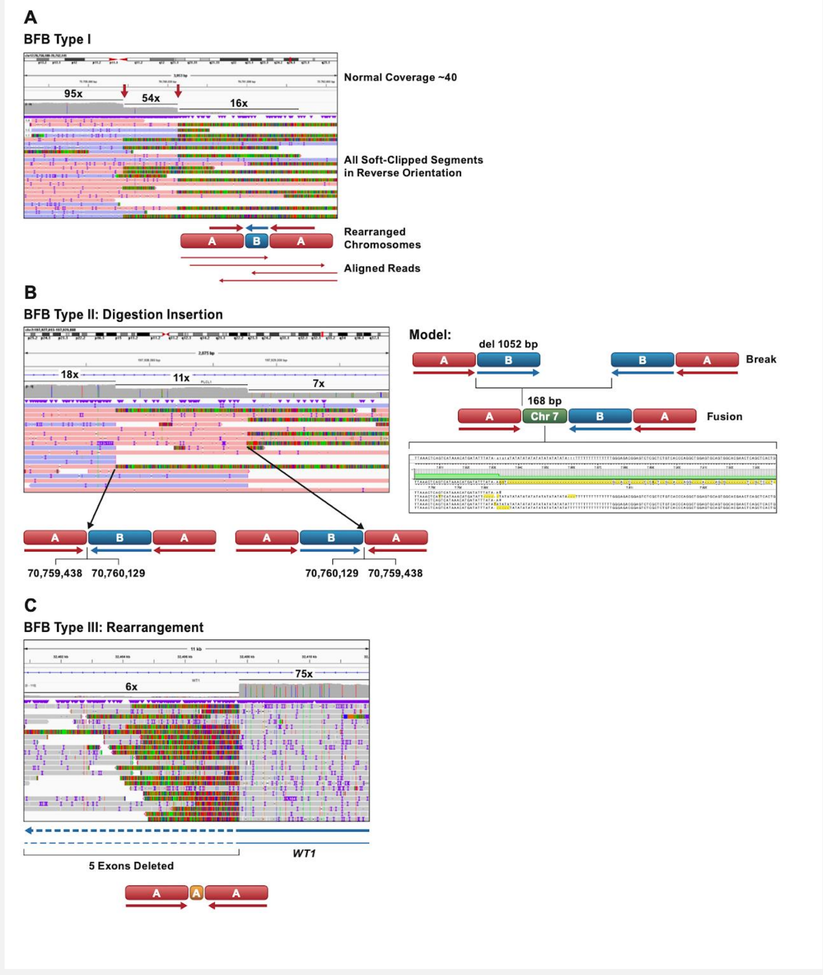
Map of the region of chromosome 11q.
Displaying a cluster of 3 BFB events associated with amplification of the *YAP1*, *BIRC2*, and *BIRC3* genes.

Methods:



Whole genome sequencing
Ultra High molecular weight sequencing
Adaptive Sampling (Human/HPV)
cDNA Transcriptome
Direct RNA sequencing

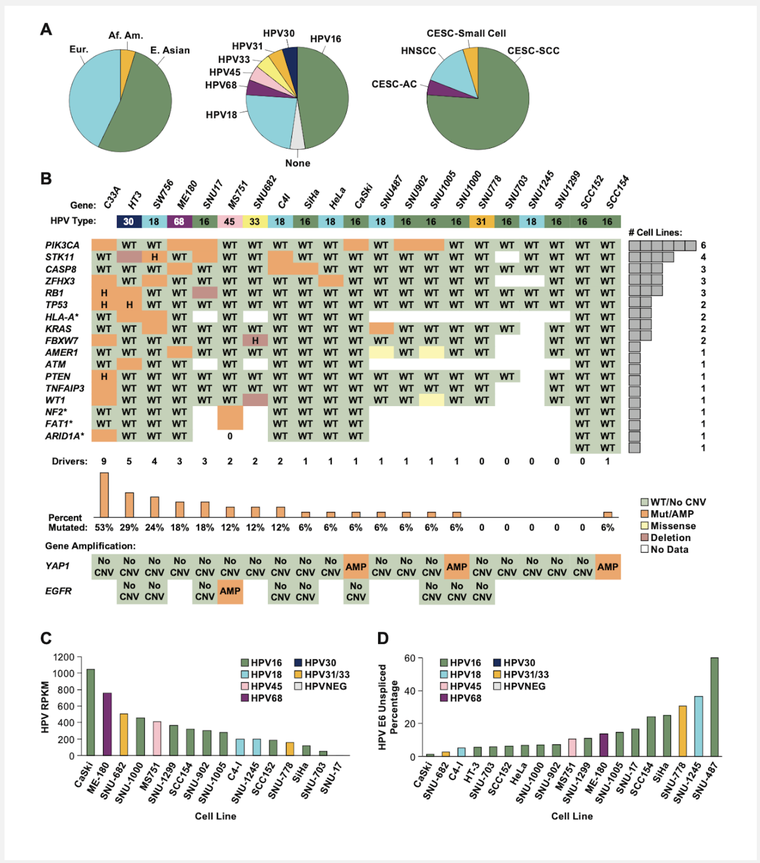
For each cell line, split alignments were used to cluster and identify breakpoints. Then clusters with foldback inversions and change in coverage were selected as BFB candidates. Identified breakpoints were further manually investigated using IGV.



Structure of BFB events.
A. The structure of the reads at the inversion site on chr17q in HT-3 cells is shown. The coverage drops from 95 to 54 to 16X, and all soft-clipped portion of reads are inverted in relation to the aligned portion of the reads. Arrows mark the start and end of the segment shown as B in the diagram below. All the reads crossing the junction align to the structure shown consistent with staggered ends of the sequence as the fusion of inverted chromatids. **B.** Model for the formation of the BFB junction of chromosome 2 in SNU-1000 cells, a Type 2 BDFB event. Following the deletion of one of the ends, a segment of chromosome 7 is inserted **in between the joined chromosomes**. **C.** A BFB event on chromosome 11p in SNU-682 cells, inside the *WT1* tumor suppressor gene. Inserted into the junction is a complex sequence likely derived from sequences at the junction.

***YAP1, BIRC 2, BIRC3* cohort data.**
A. Copy number data for genes centromeric and telomeric to *YAP1* in all TCGA cervical tumors with a log score for *YAP1* copy number >1. **B.** Comparison of the age of diagnosis and **C.** Self-identified race of TCGA cervical cancer patients with and without *YAP1* amplification.

Results:



Cell line patient demographics and HPV expression data.
A. The distribution of the ancestry of the subjects, HPV types, and cancer and histological types among the 22 cell lines is shown. **B.** The cancer driver gene mutations found mutated in at least one cell line is shown. Orange highlights are pathogenic variants and in yellow are variants of unknown impact homozygous variants **C.** The cDNA or RNA reads per million reads mapped (RPM) to the HPV genome is shown. **D.** The percentage of HPV E6/E7 containing reads that are unspliced in the E6 gene.

Conclusion:

The major molecular subtypes of cervical cancer are represented in this panel and will allow the development of targeted therapies.
Using this panel of 22 cervical cell lines and long-read sequencing we were able to fully characterize the sequence of HPV integrations and the consequences on gene expression. We show that cervical cell lines that are HPV-negative or integrated by a non-high-risk-type have a higher level of driver mutations, including in *TP53* and *RB1* to compensate for low or absent activity of HPV E6 and E7. We characterized the sequences at the inversion junctions of breakage-fusion-bridge events and provide new insight into the formation of these critical chromosome rearrangements in cancer. These cell lines can serve as models for specialized treatments of cervical cancer.

Disclaimer: