Universidad Nacional de San Agustín

SplitThreader: Exploration and analysis of rearrangements in cancer genomes

Review

MSc. Vicente Machaca Arceda

2021

Content

Overview

Genomics and Big Data

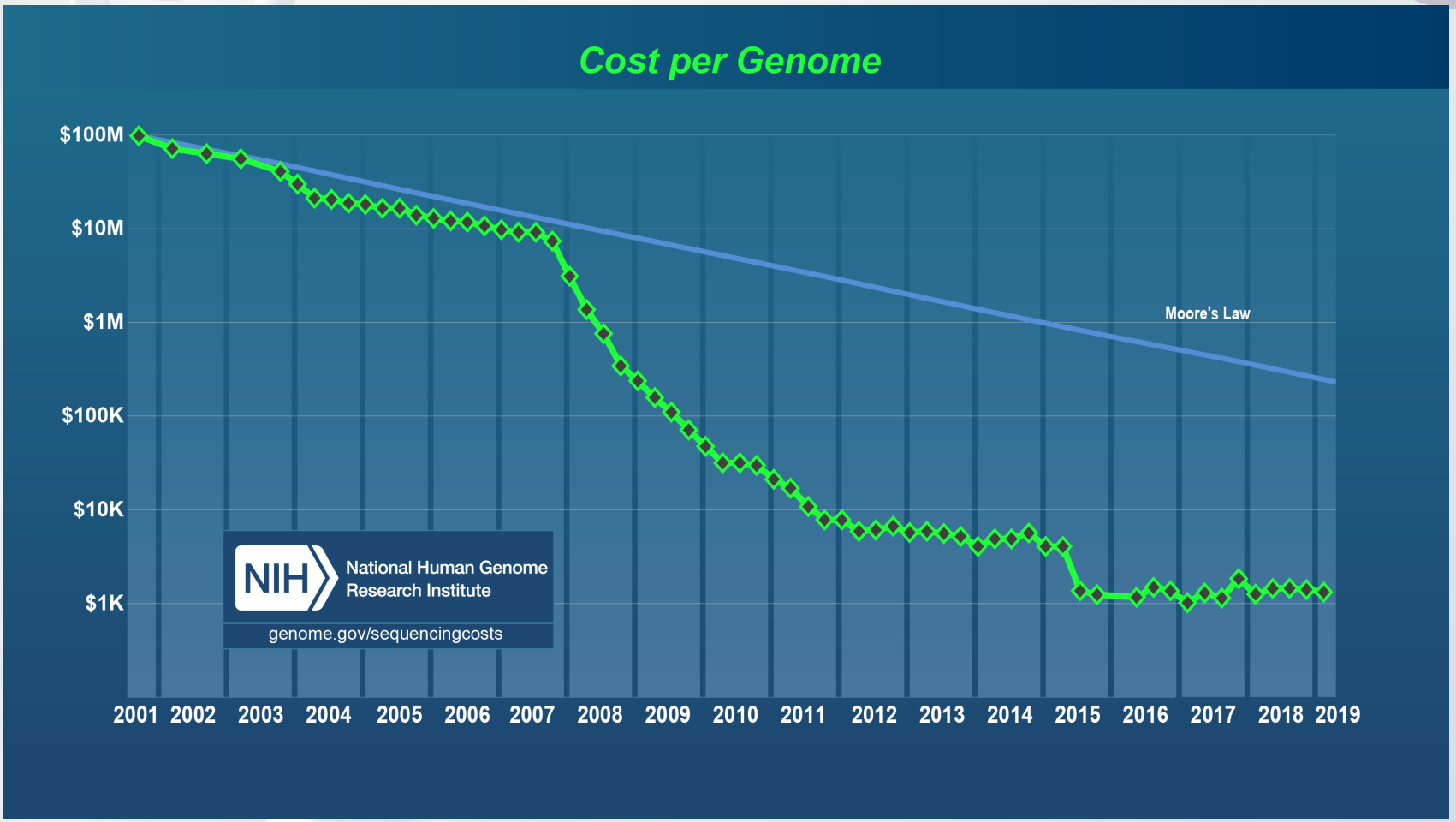


Figure: Cost per genome sequencing over the years.

Genomics and Big Data

- 6.4 billions of bases.
- 20k genes approximately.
- No technology exist that can read an entire chromosome from end to end.
- Some changes in the genome encode normal variation like hair color, other can cause diseases.

Overview

Problem

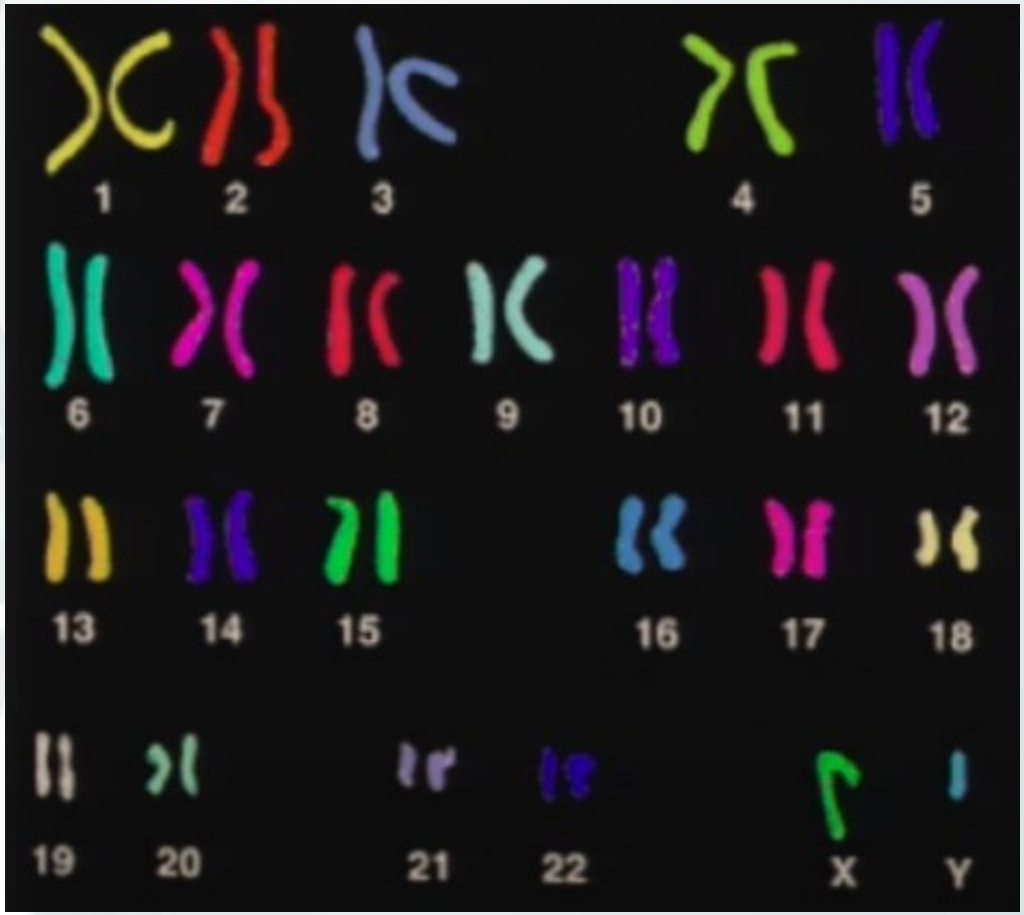


Figure: 46 Chromosomes presented in cells.

Problem

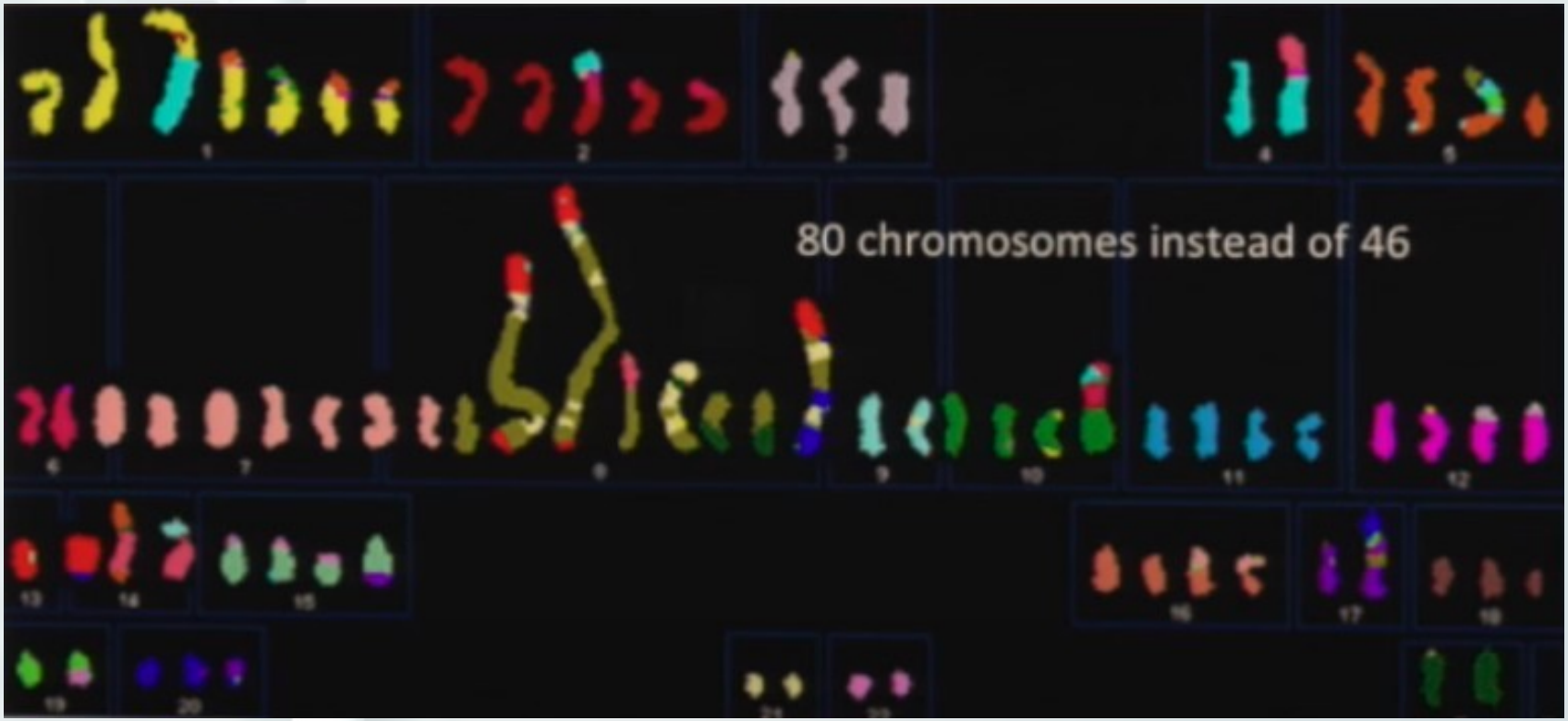


Figure: CII line from a woman with metastatic breast cancer in 1971. Tumor cells have been grown and studied in the lab ever since.

Problem

Genomic instability is one of the **hallmarks of cancer** [1, 2], resulting in:

- Widespread copy number changes.
- Structural variants.
- Chromosome-scale rearrangements.

Copy number variants and **gene fusions** are common drivers in cancer [3, 4].

Problem

The available algorithms for identifying gene fusions **do not have perfect specificity** (False positive rate). Require a joint analysis of genomic and transcriptomic data to correctly analyze.

Rearrangements variants are difficult to study, because of the sheer complexity of rearrangements, which often include adjacencies between distant regions of a chromosome or even between unrelated chromosomes

Overview

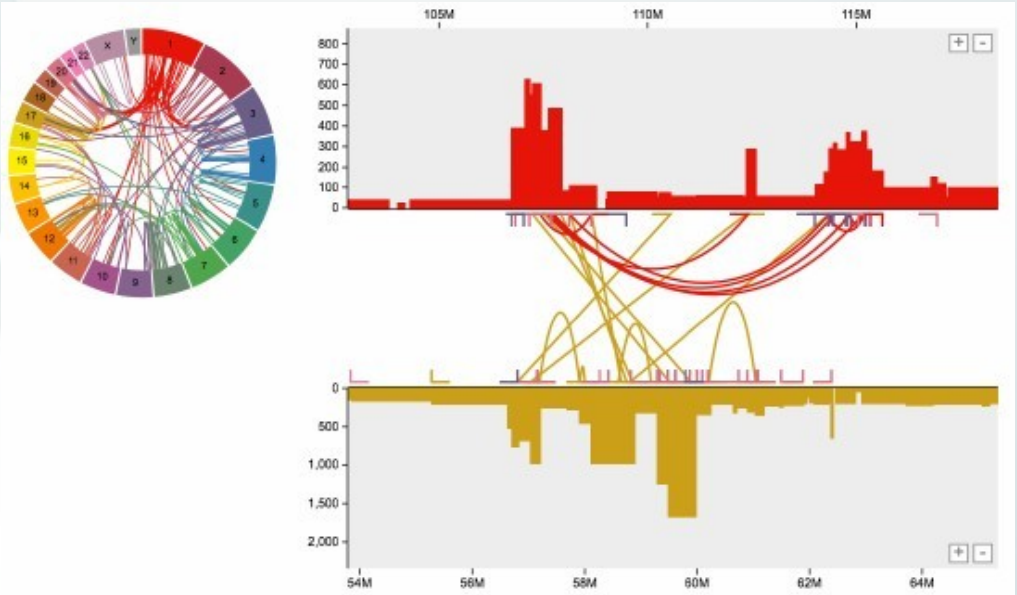
Research question

Exploration and analysis of rearrangements in cancer genomes could be performed with a web platform?

Overview

SplitThreader
Proposal

SplitThreader, an opensource interactive **web application** for analysis and visualization of genomic rearrangements and copy number variation in cancer genomes [5].



Overview

SplitThreader
Method

- Constructing the SplitThreader graph.
- Priority queue breadth-first search.
- Evaluating graph search and gene fusion matches.
- Copy number segmentation.
- Copy number concordance analysis.
- Variant neighborhood analysis.

Overview

SplitThreader
Results

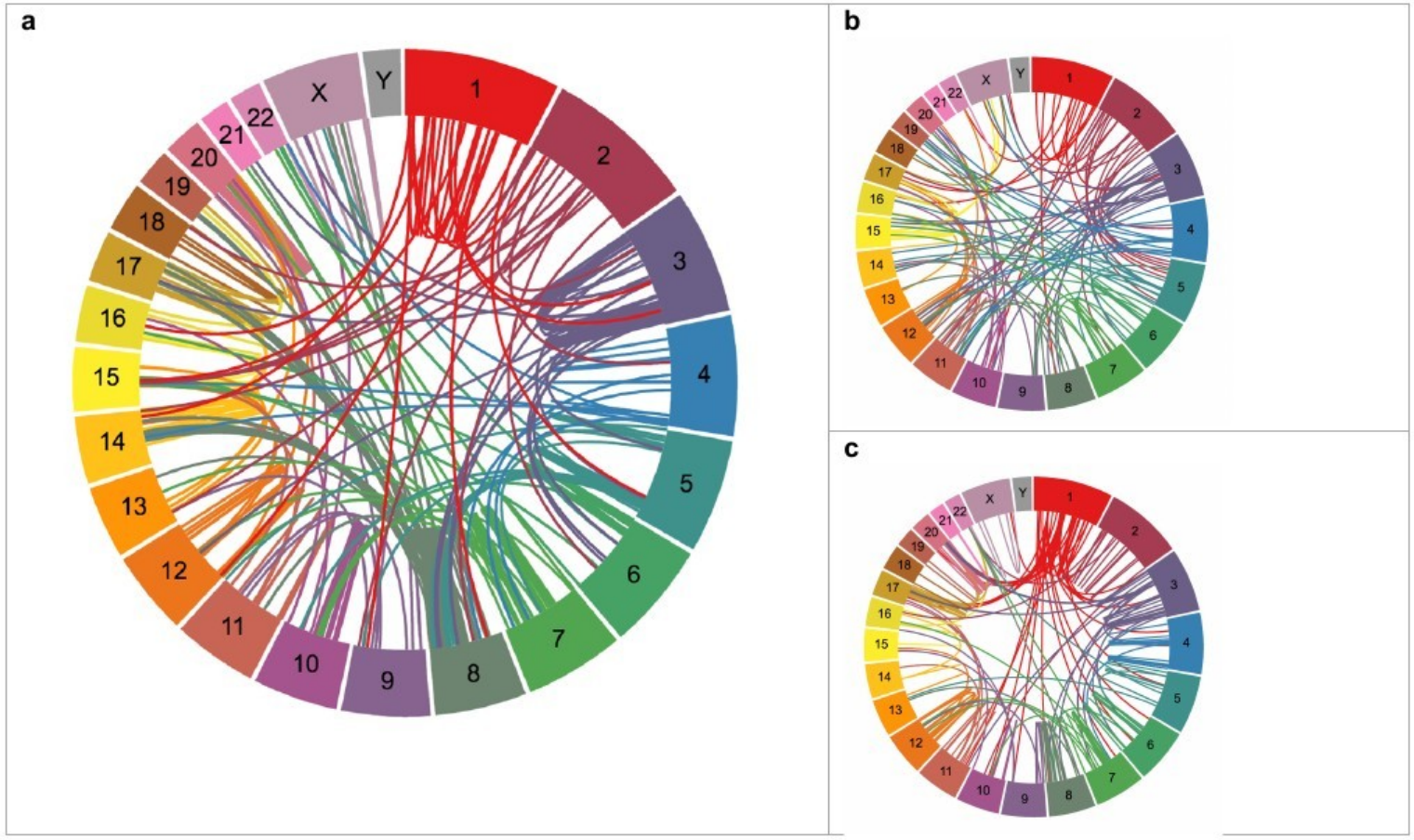


Figure: Circos plots showing genomic rearrangements in the cell lines SK-BR-3 (a), A549 (b), and MCF-7 (c).

SplitThreader
Results

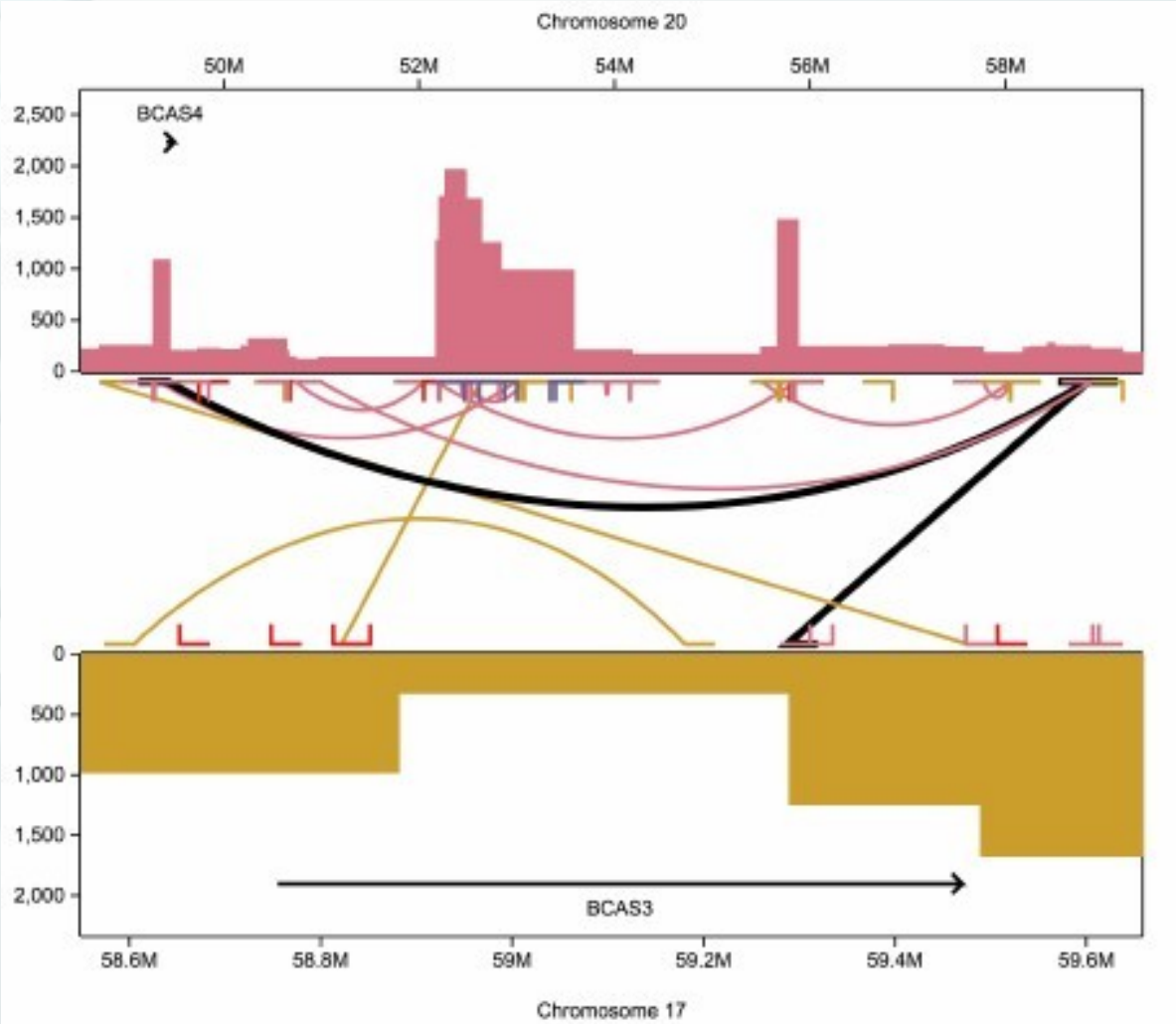


Figure: BCAS4-BCAS3 two-hop gene fusion gene fusion in MCF-7.

SplitThreader
Results

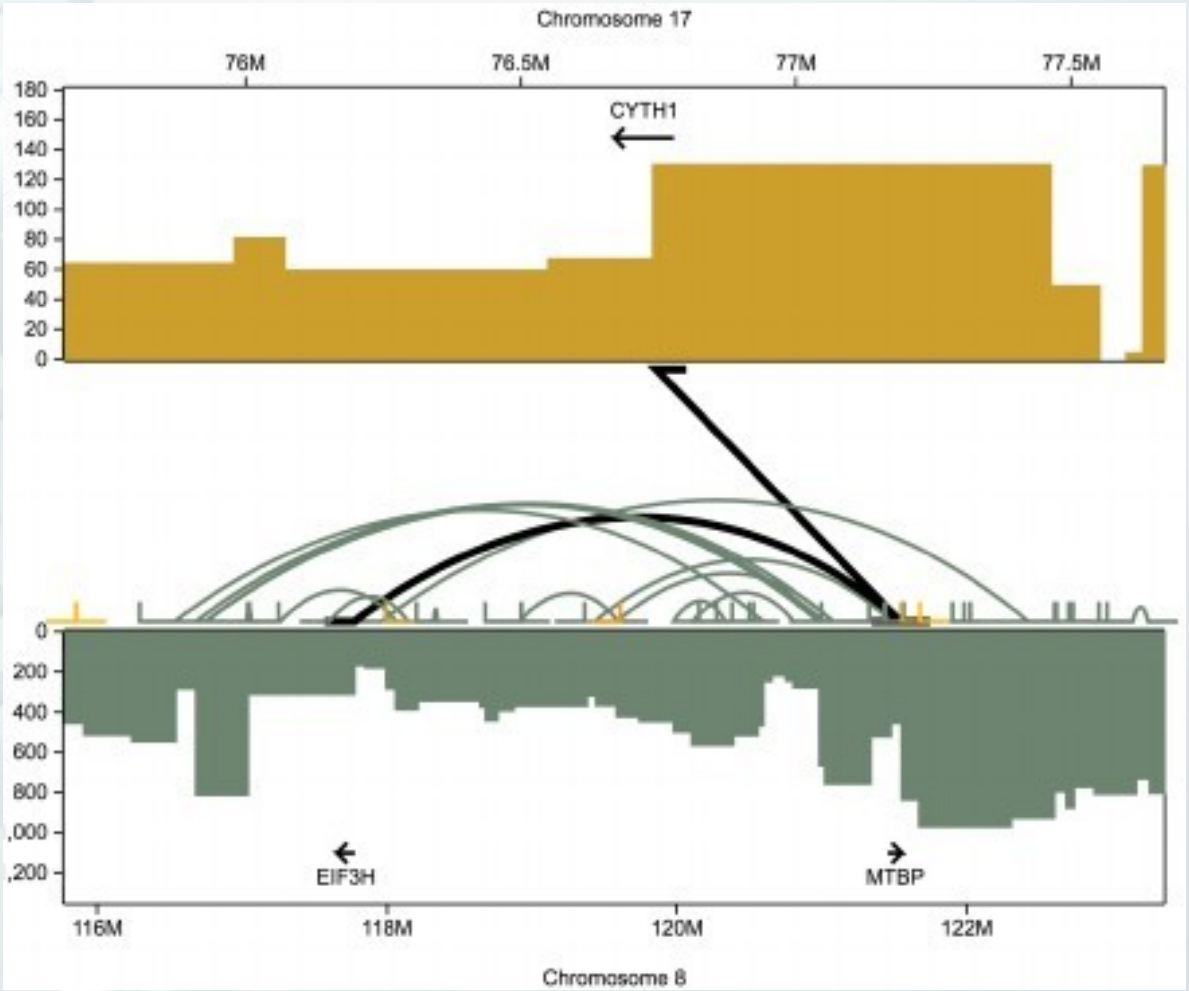


Figure: CPNE1-PHF20-PREX1 two-hop gene fusion in SK-BR-3.

References I

[1] D. Hanahan and R. A. Weinberg, “Hallmarks of cancer: the next generation,” *cell*, vol. 144, no. 5, pp. 646–674, 2011.

[2] P. J. Hastings, J. R. Lupski, S. M. Rosenberg, and G. Ira, “Mechanisms of change in gene copy number,” *Nature Reviews Genetics*, vol. 10, no. 8, pp. 551–564, 2009.

[3] A. Shlien and D. Malkin, “Copy number variations and cancer,” *Genome medicine*, vol. 1, no. 6, pp. 1–9, 2009.

[4] F. Mitelman, B. Johansson, and F. Mertens, “The impact of translocations and gene fusions on cancer causation,” *Nature Reviews Cancer*, vol. 7, no. 4, pp. 233–245, 2007.

[5] M. Nattestad, M. C. Alford, F. J. Sedlazeck, and M. C. Schatz, “Splitthreader: Exploration and analysis of rearrangements in cancer genomes,” *bioRxiv*, p. 087981, 2016.