

Comprehensive Genome and Transcriptome Structural Analysis of a Breast Cancer Cell Line using PacBio Long Read Sequencing

Maria Nattestad

Schatz + McCombie + Hicks at Cold Spring Harbor Laboratory

McPherson + Beck at the Ontario Institute for Cancer Research

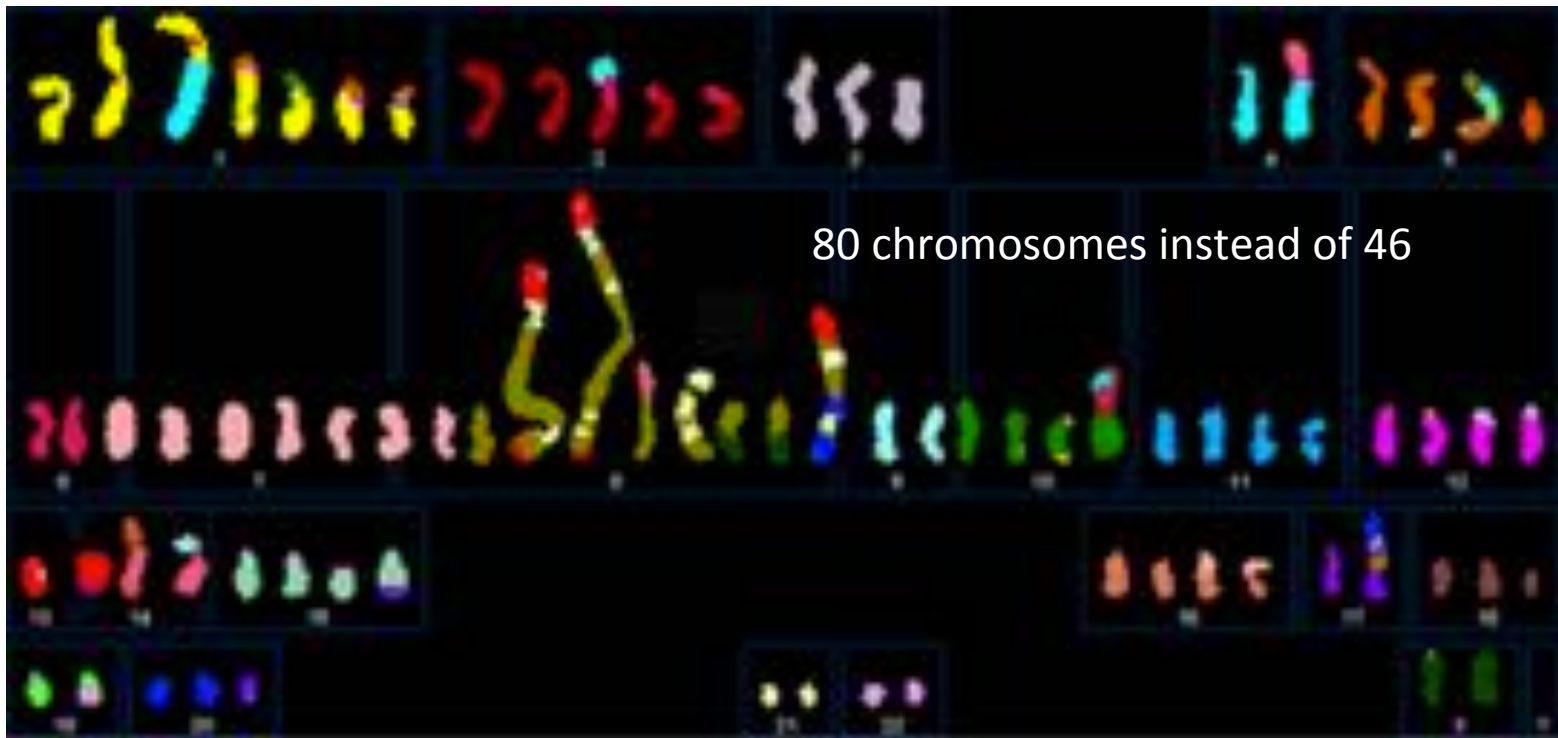
Pacific Biosciences

DNAexus



SK-BR-3

Most commonly used Her2-amplified breast cancer cell line



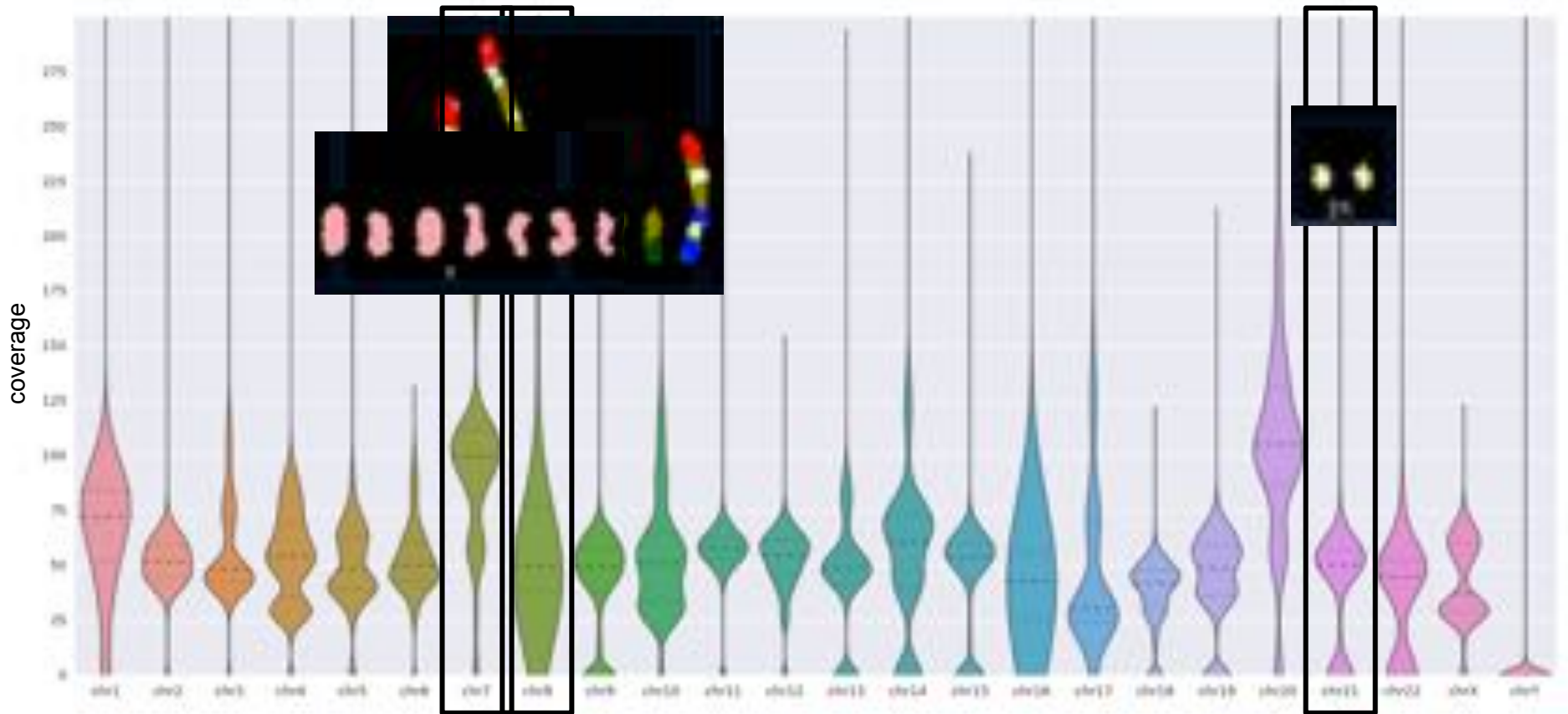
Often used for pre-clinical research on Her2-targeting therapeutics such as Herceptin (Trastuzumab) and resistance to these therapies.

(Davidson et al, 2000)

PacBio long-read DNA sequencing

mean read length: 9 kb
max read length: 71 kb

72X coverage

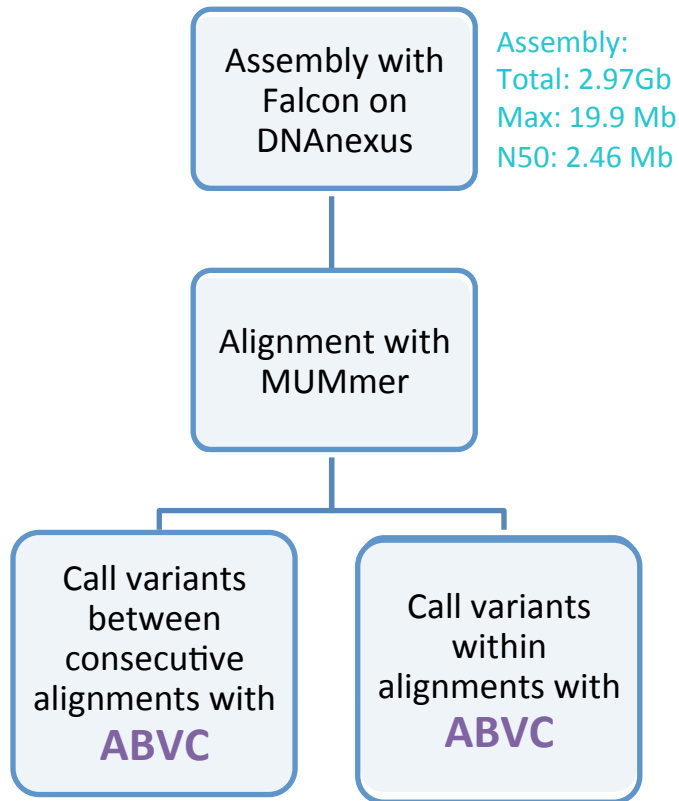


Genome-wide coverage averages around 54X

Coverage per chromosome varies greatly as expected from previous karyotyping results

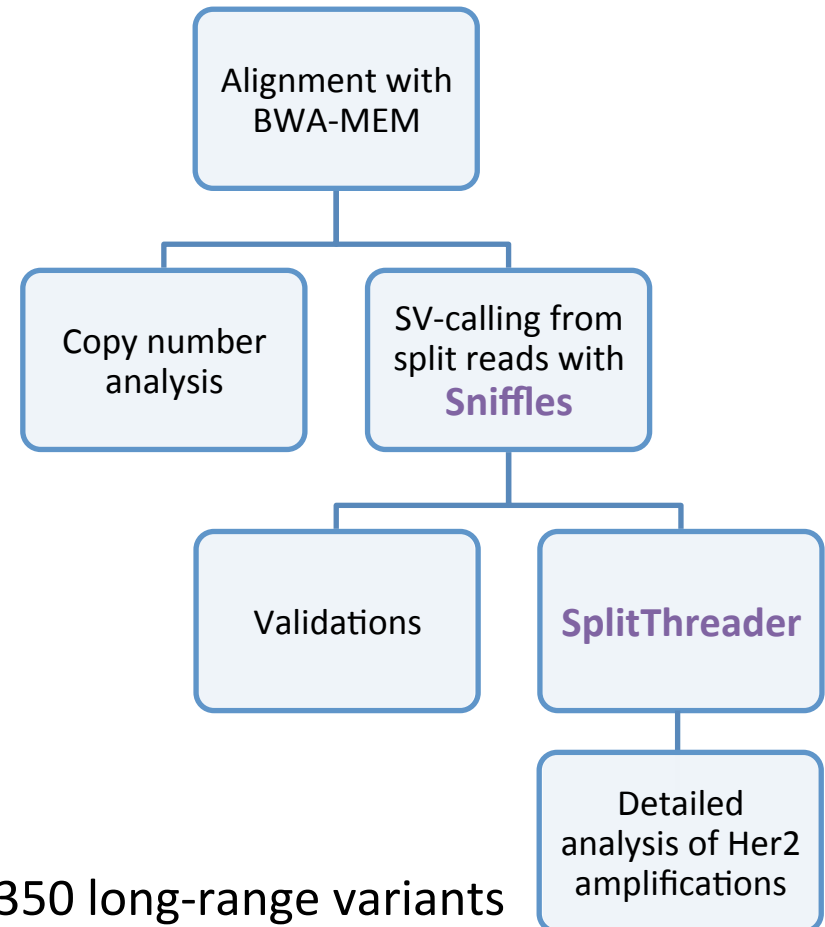
Structural variant detection

Assembly-based



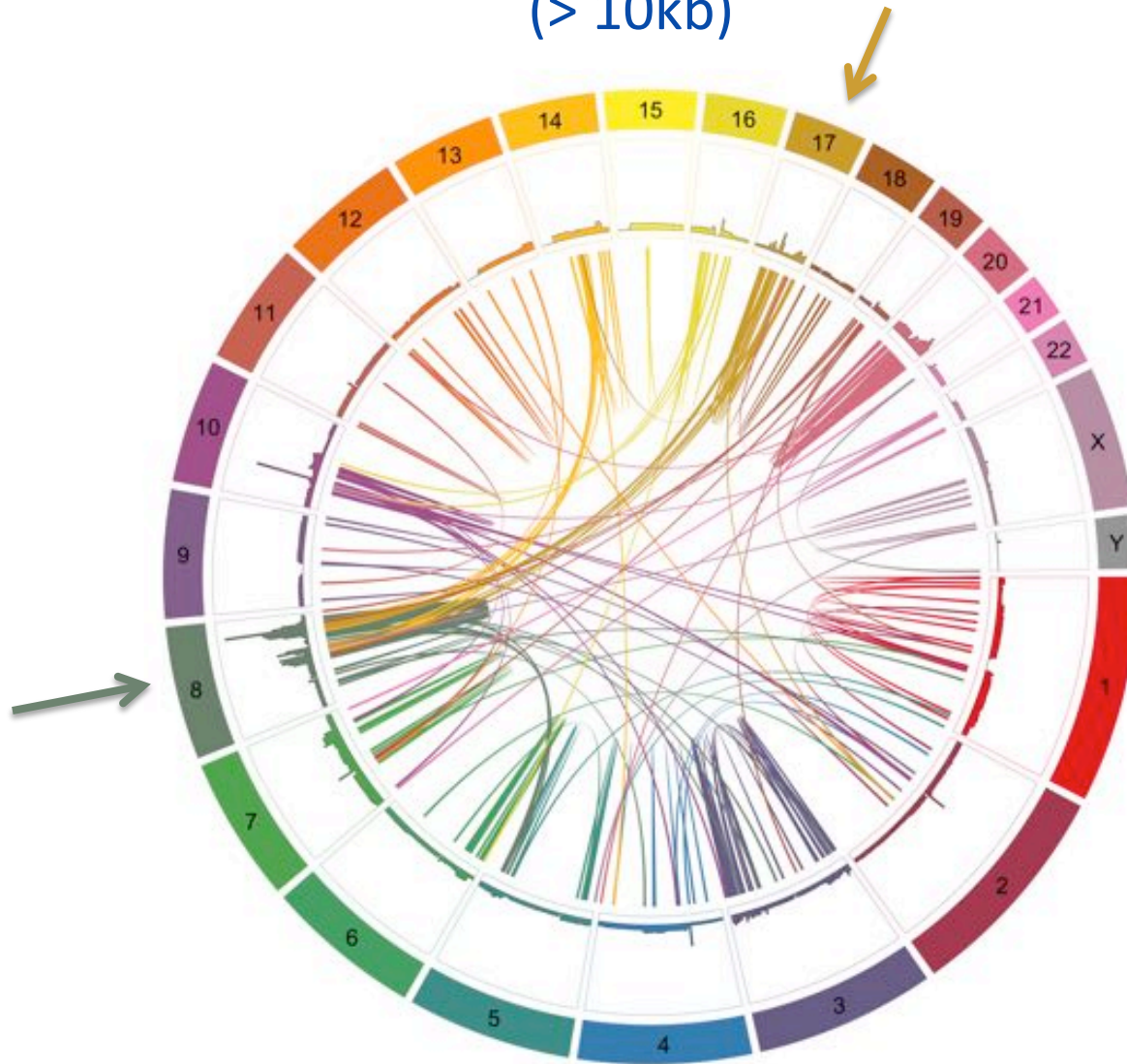
~ 11,000 local variants
50 bp < size < 10 kbp

Alignment-based

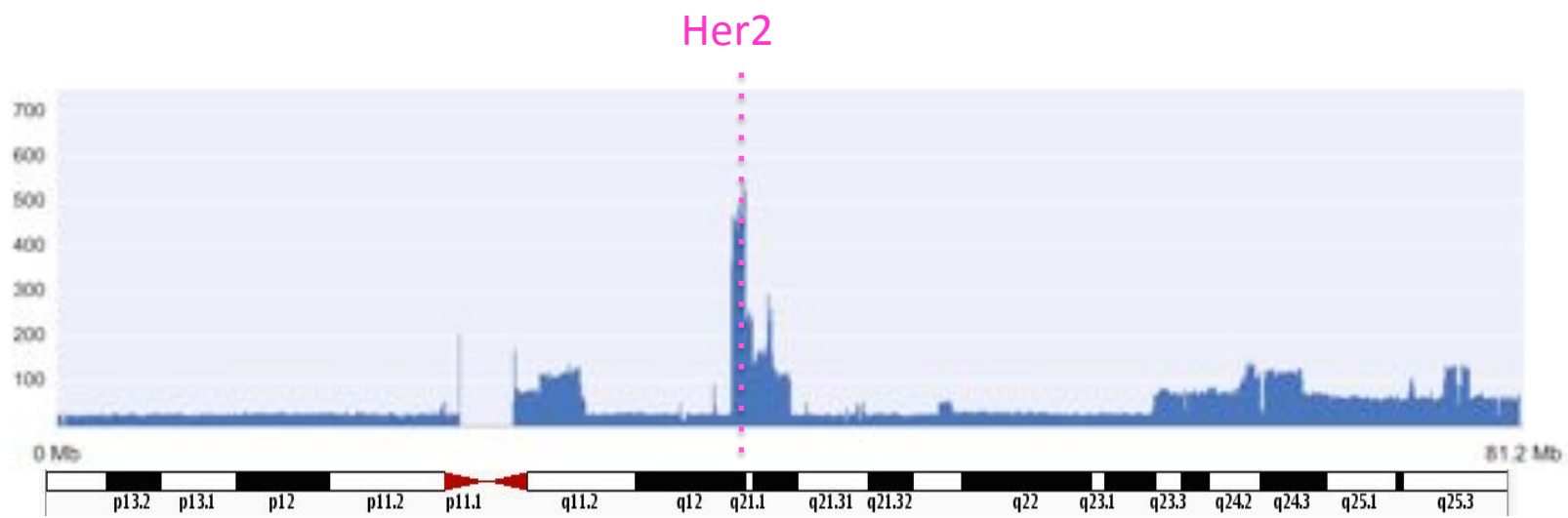


350 long-range variants
(>10kb distance)

Long-range structural variants found by Sniffles (> 10kb)



Threshold: 10 split reads broken within a 200 bp interval on both sides of the translocation

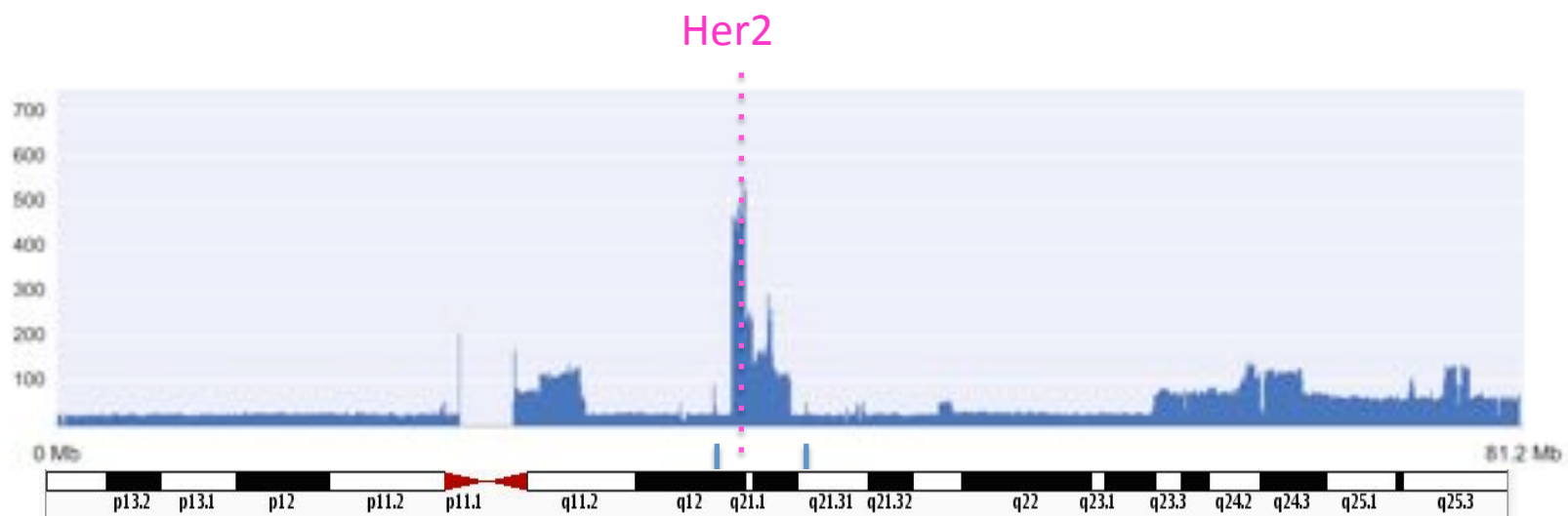


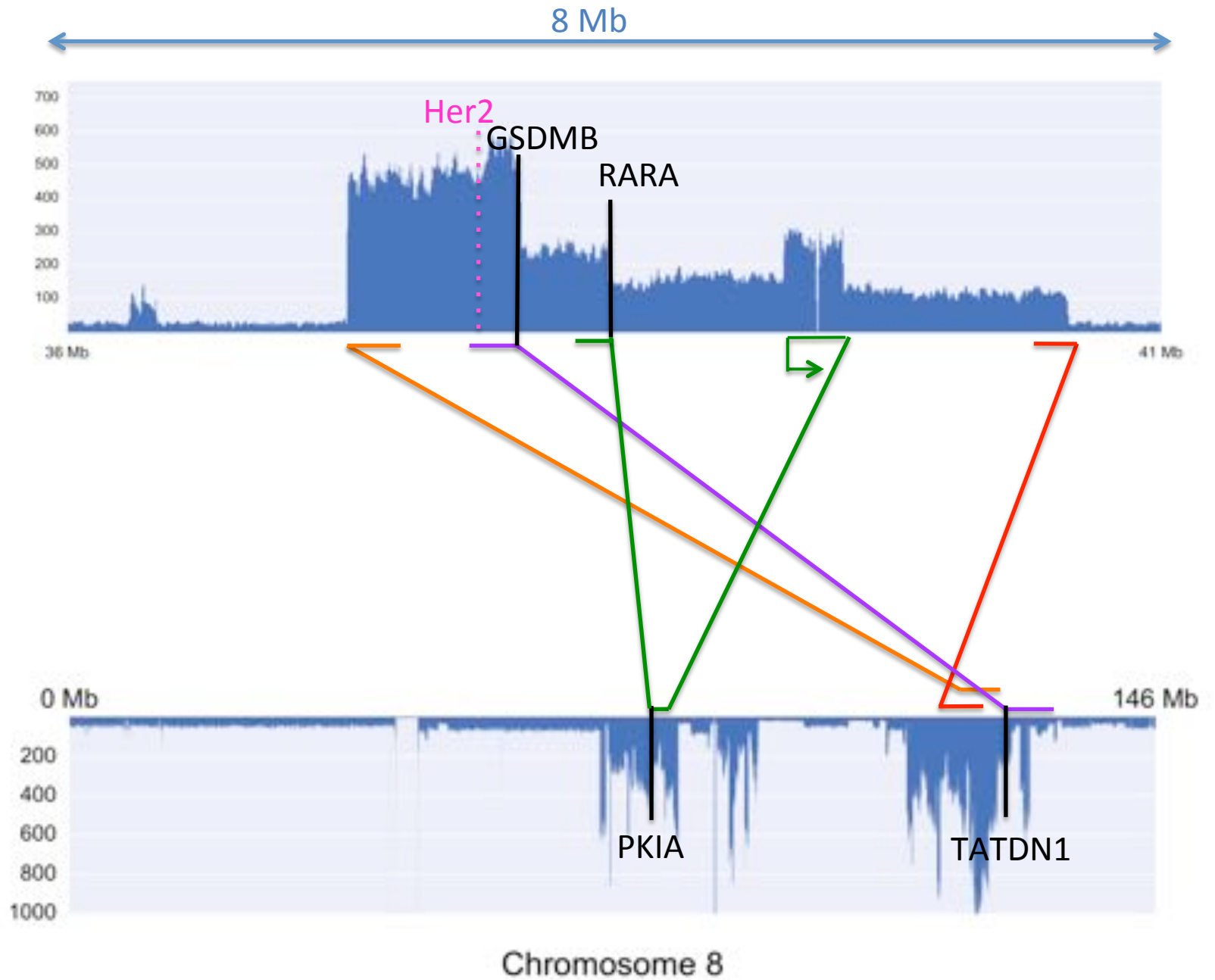
Chr 17: 83 Mb



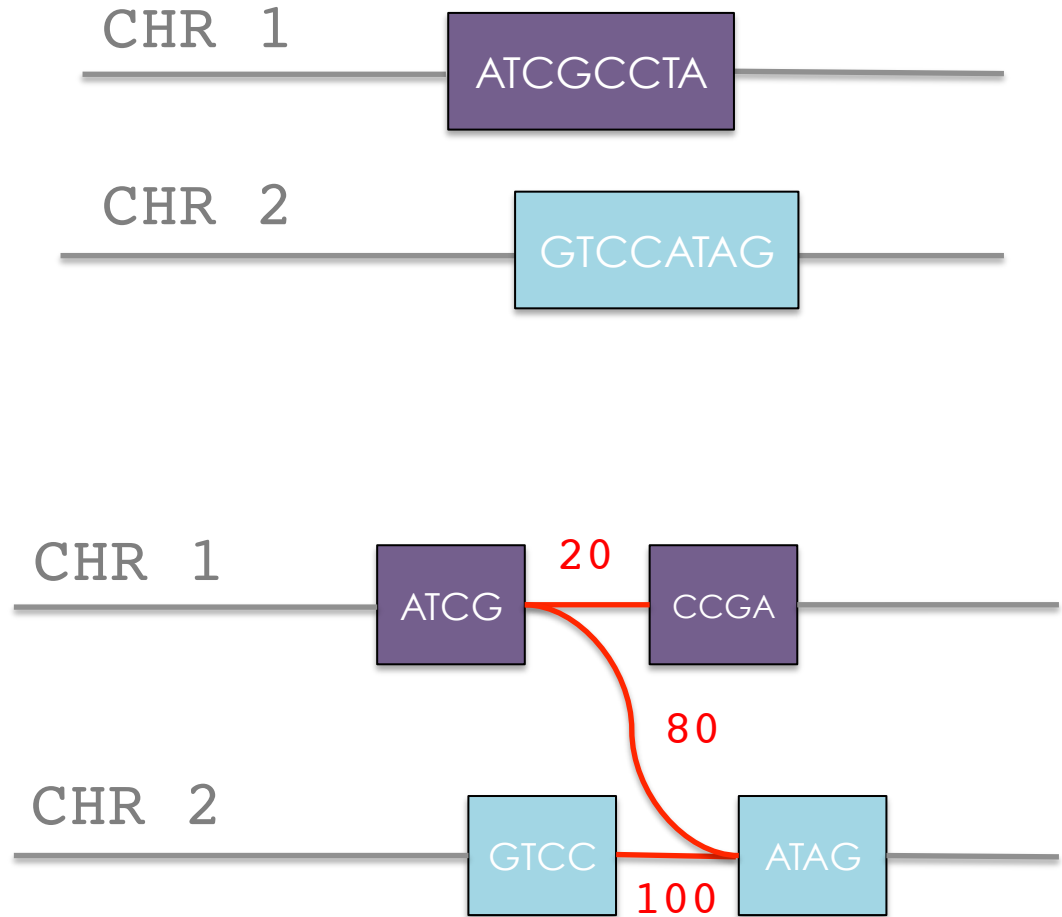
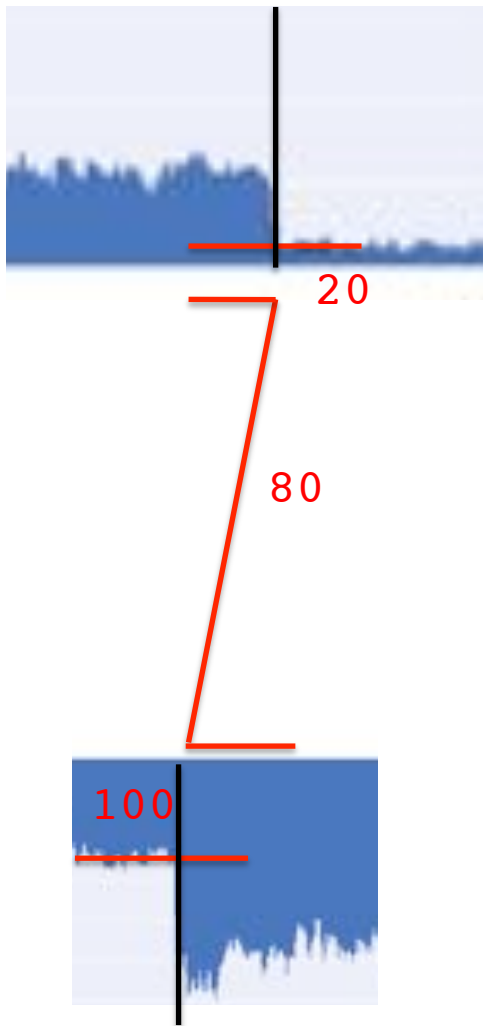
8 Mb

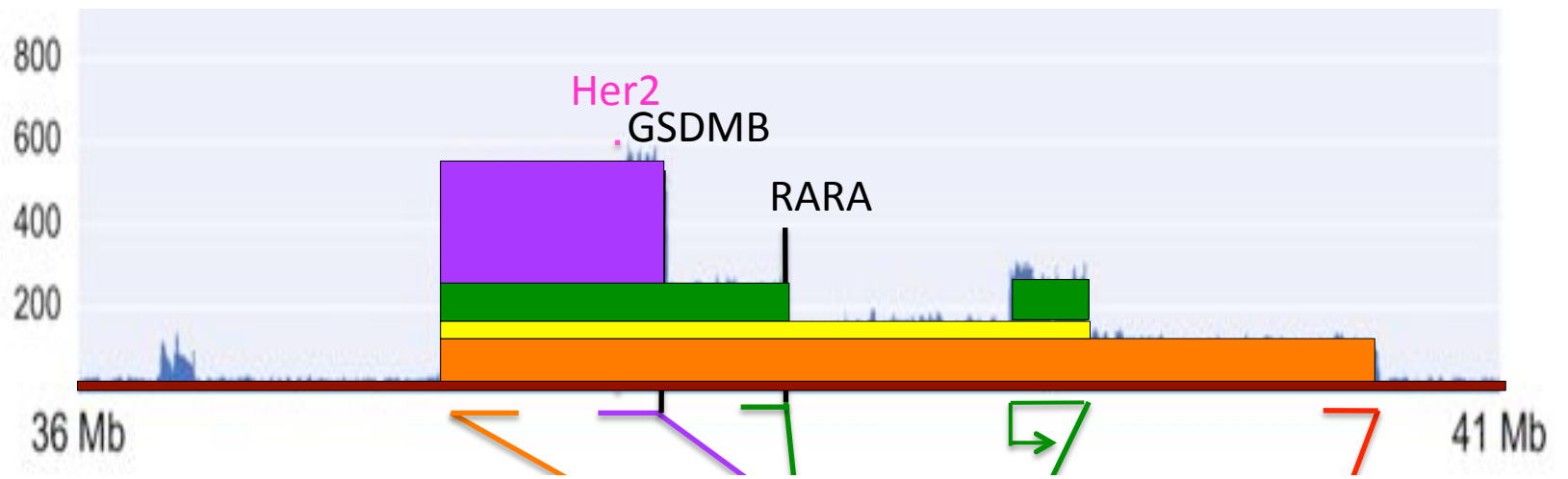






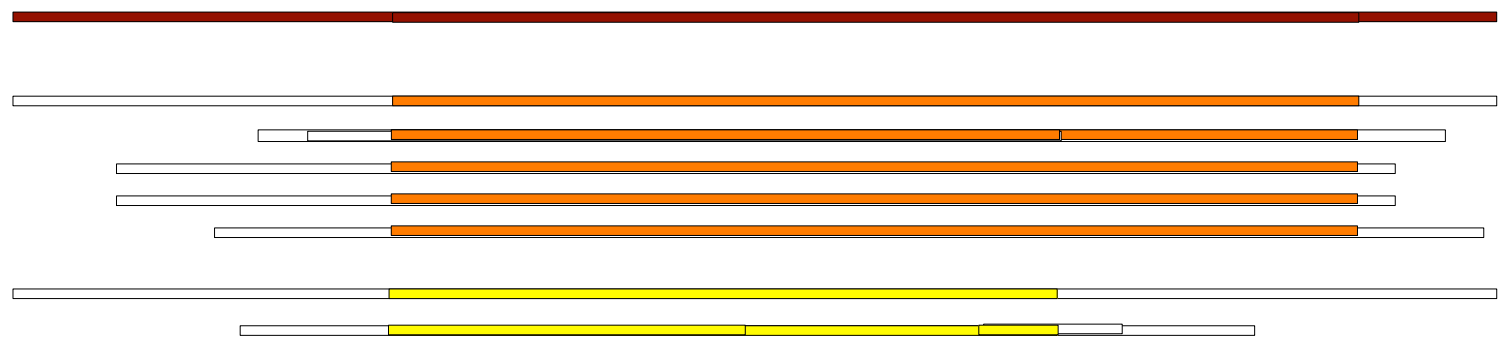
SplitThreader: Graphical threading to retrace complex history of rearrangements in cancer genomes





Chr 17

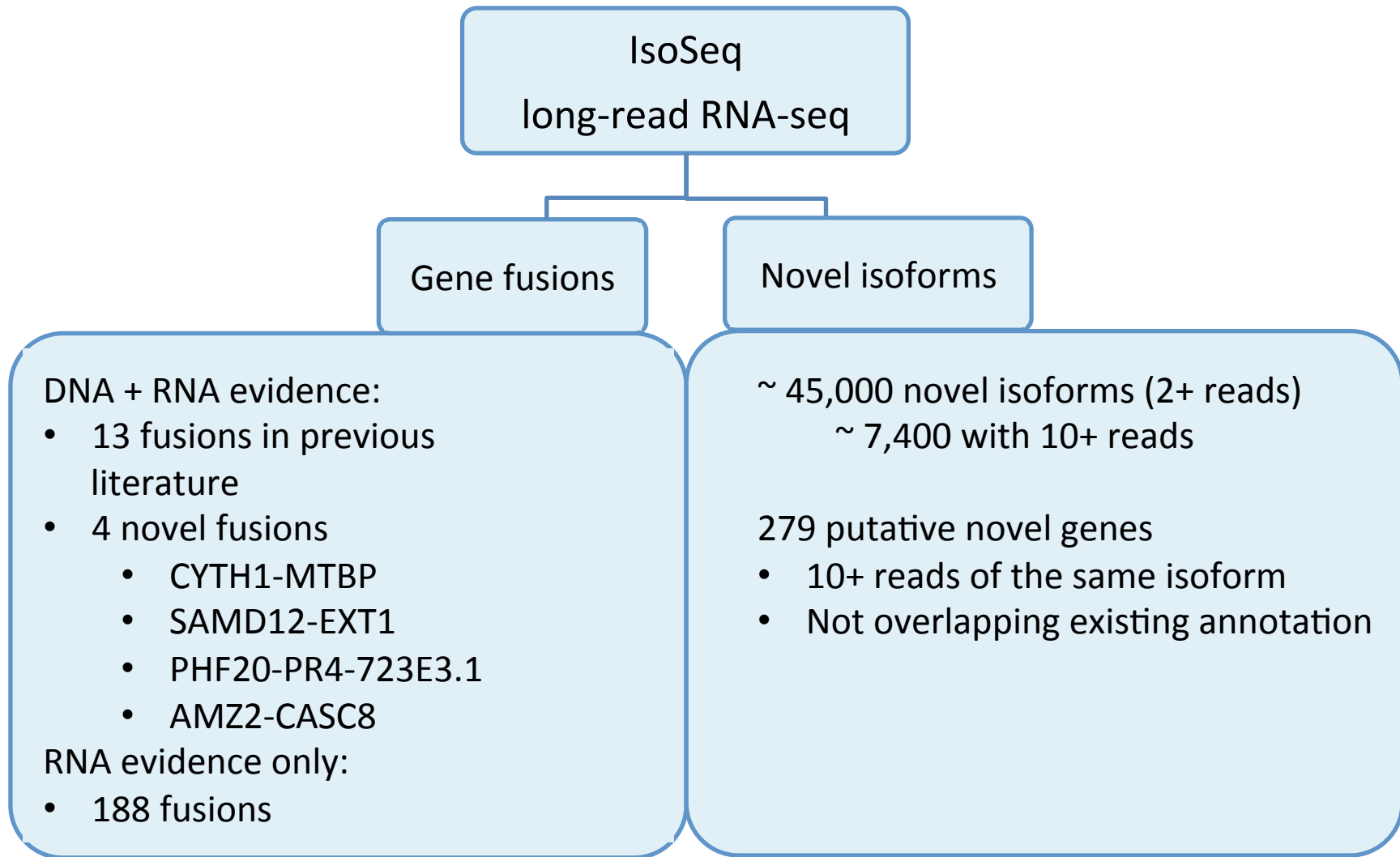
Chr 8



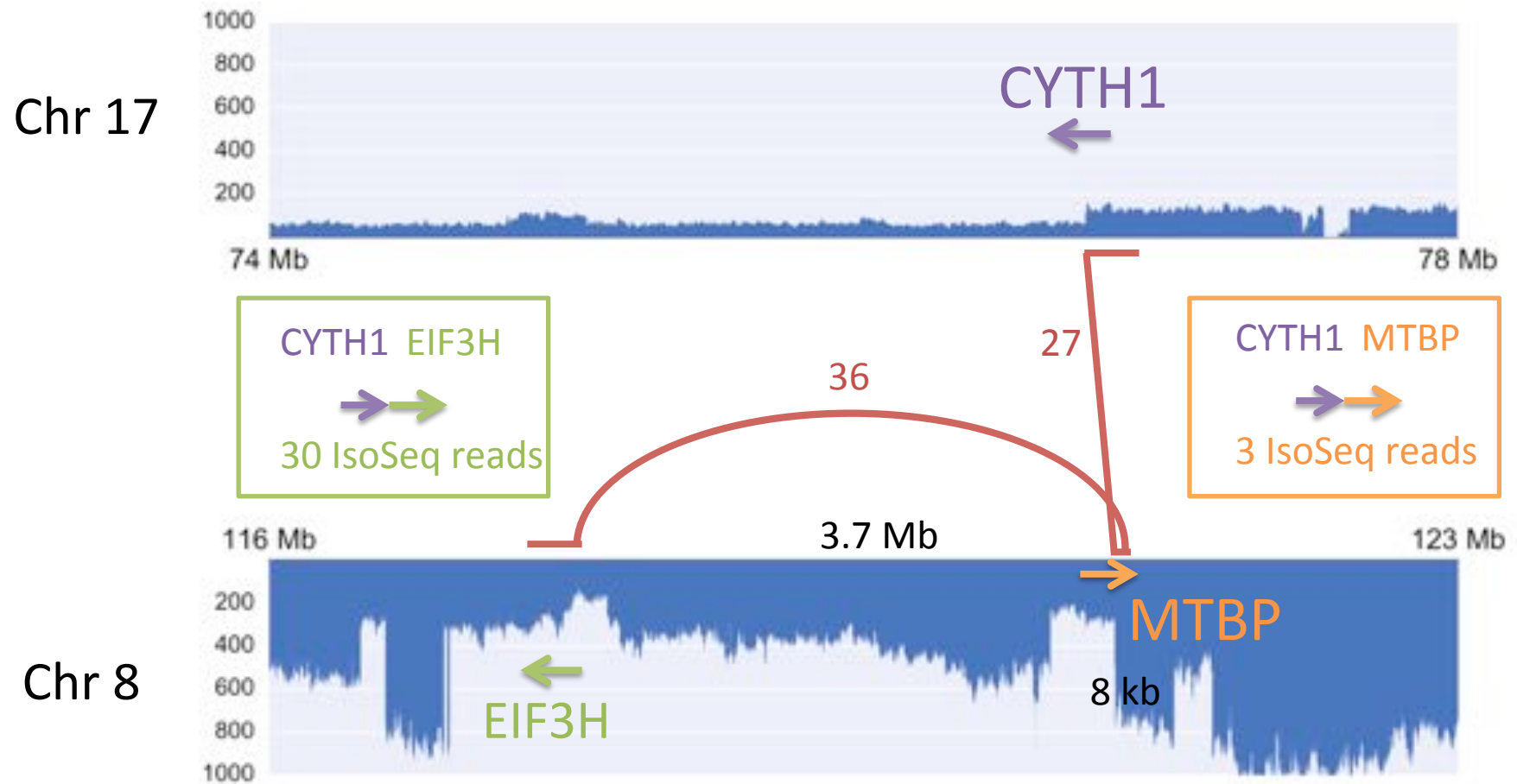
1. Healthy chromosome 17
2. Translocation into chromosome 8
3. Translocation within chromosome 8
4. Complex variant and inverted duplication within chromosome 8
5. Translocation within chromosome 8



Transcriptome analysis with IsoSeq



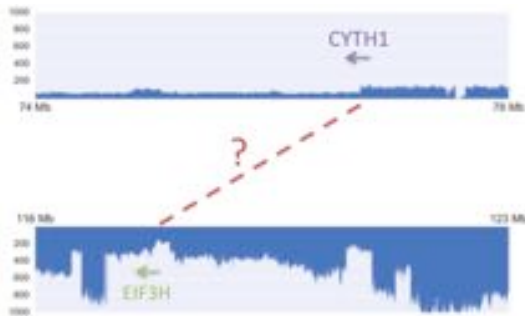
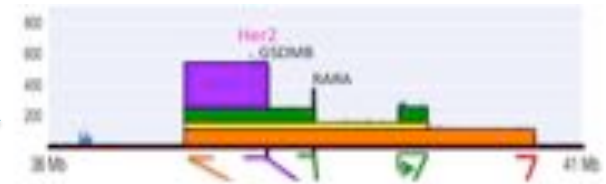
CYTH1-EIF3H gene fusion



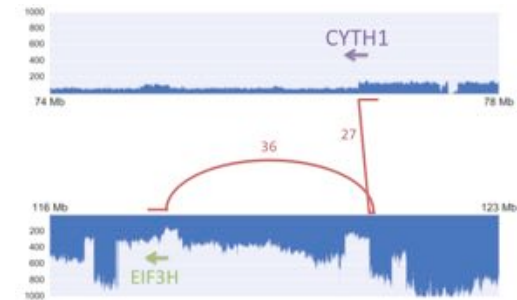
The genome informs the transcriptome



Explain amplifications



Trace gene fusions



Data and additional results: <http://schatzlab.cshl.edu/data/skbr3/>

Acknowledgments



Cold
Spring
Harbor
Laboratory

Sara Goodwin
Timour Baslan
Fritz Sedlazeck
Tyler Garvin
Han Fang
James Gurtowski
Elizabeth Hutton
Marley Alford
Melissa Kramer
Eric Antoniou
James Hicks
Michael Schatz
W. Richard McCombie



Karen Ng
Timothy Beck
Yogi Sundaravadanam
John McPherson

DNAexus

Pacific Biosciences

Elizabeth Tseng
Jason Chin

Our software in development: Sniffles by Fritz, ABVC and SplitThreader by Maria