

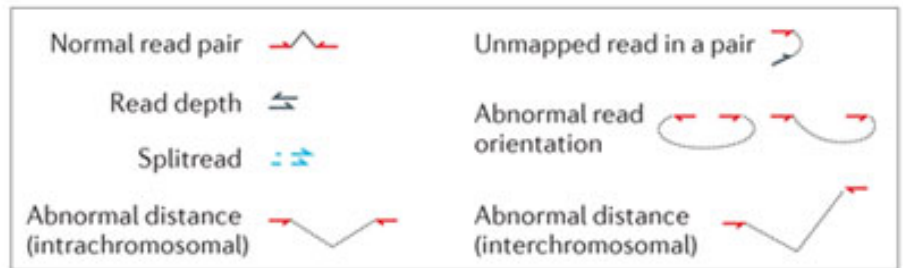
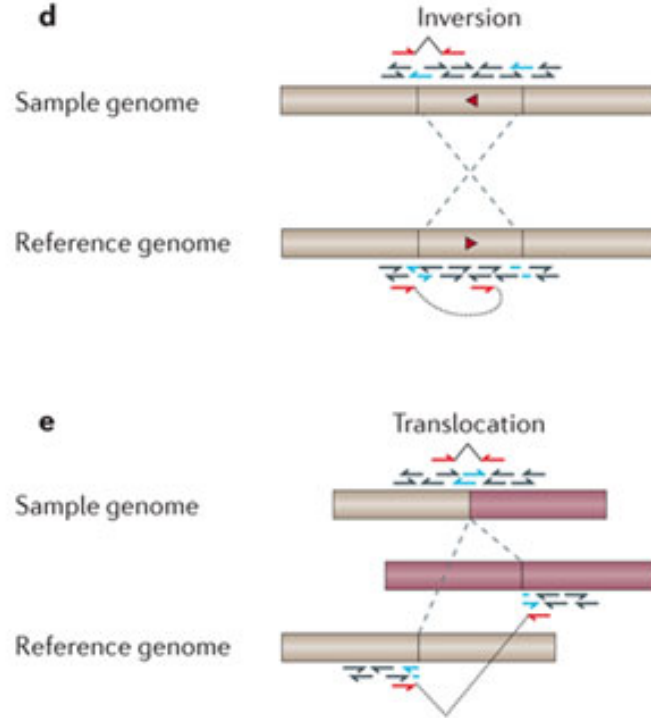
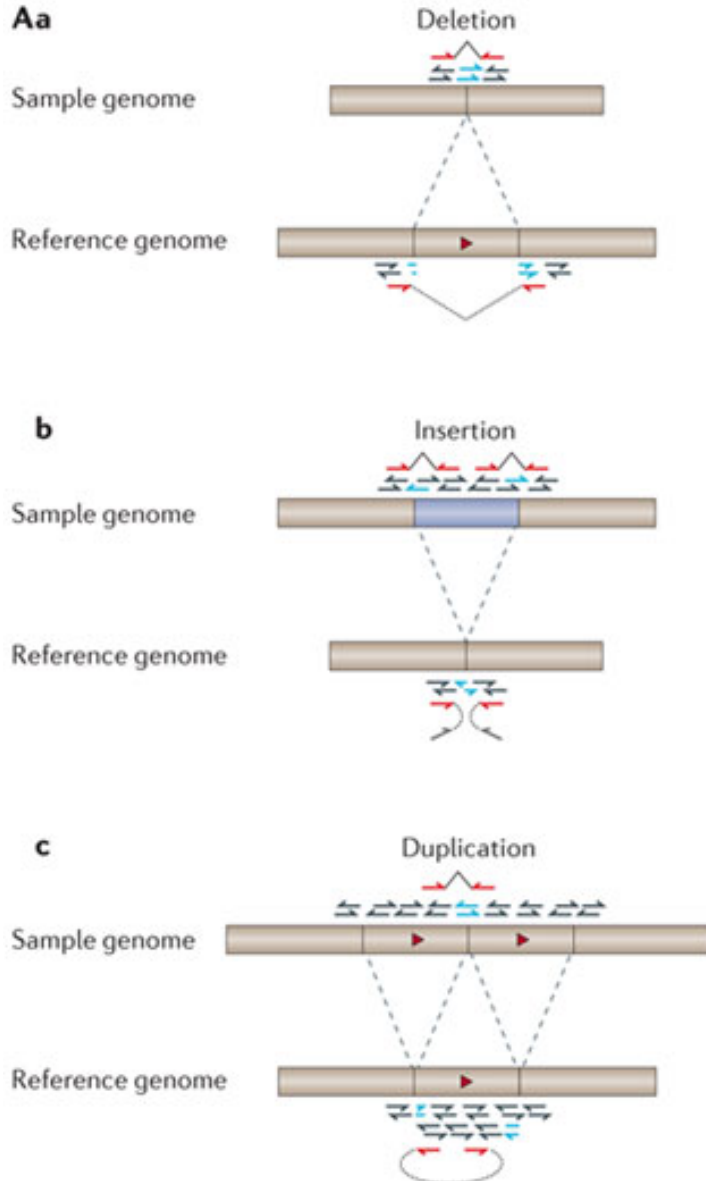
Accurate and fast detection of complex and nested structural variations using long read technologies

Fritz Sedlazeck

Friday, Oct 28
CSHL



Structural variation



Long Read Technologies

- (+) SVs in repetitive regions
- (+) Can identify nested SVs

- (-) Higher error rate
- (-) Hard to align



PACBIO®

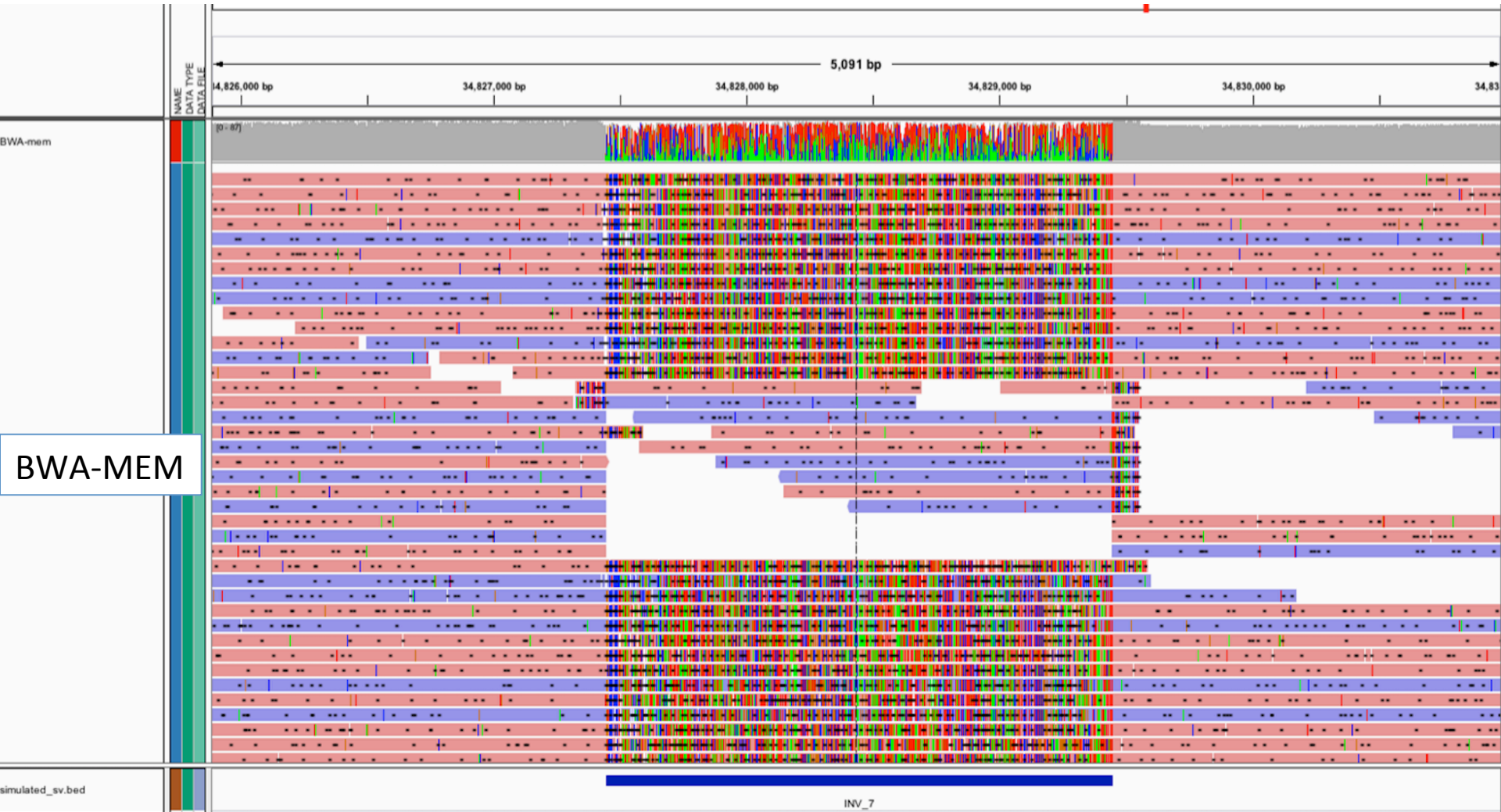


Oxford

NANOPORE

Technologies

Hard to align



Human genome: 1kb Inversion

Improving long read alignment



Philipp
Rescheneder

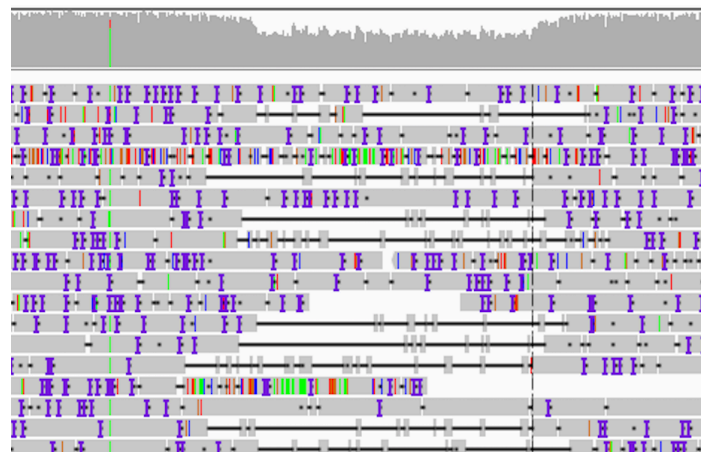
1. Split the reads:

- Translocations
- Inversions
- Duplications

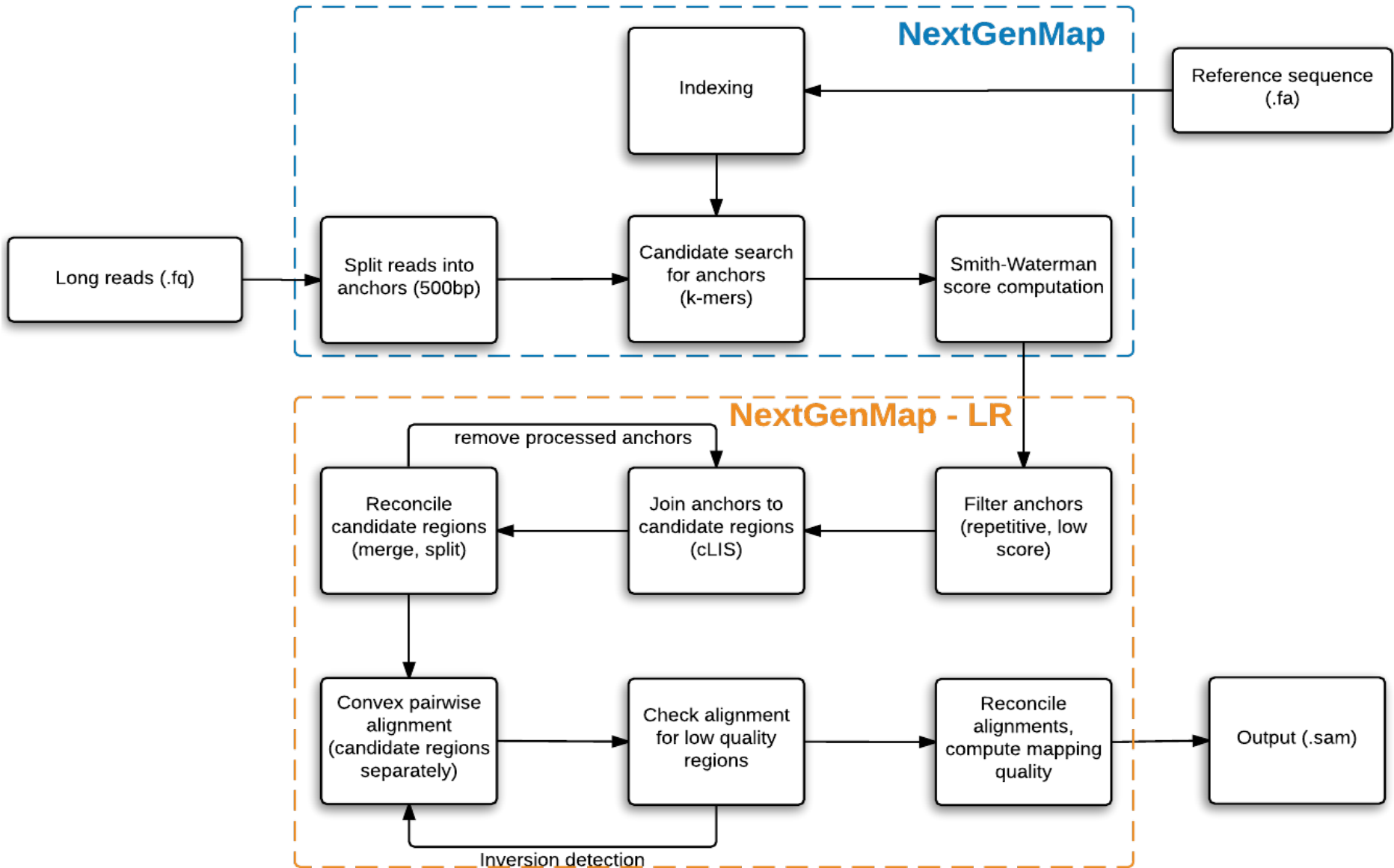


2. Improve alignment:

- Insertions
- Deletions



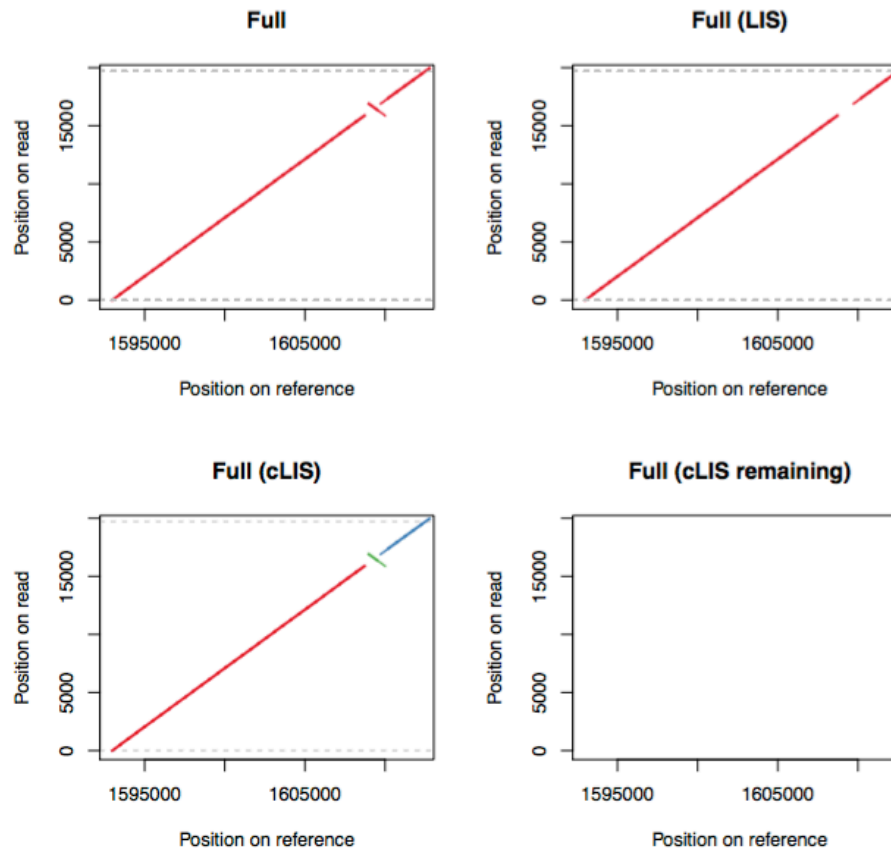
NGM-LR workflow



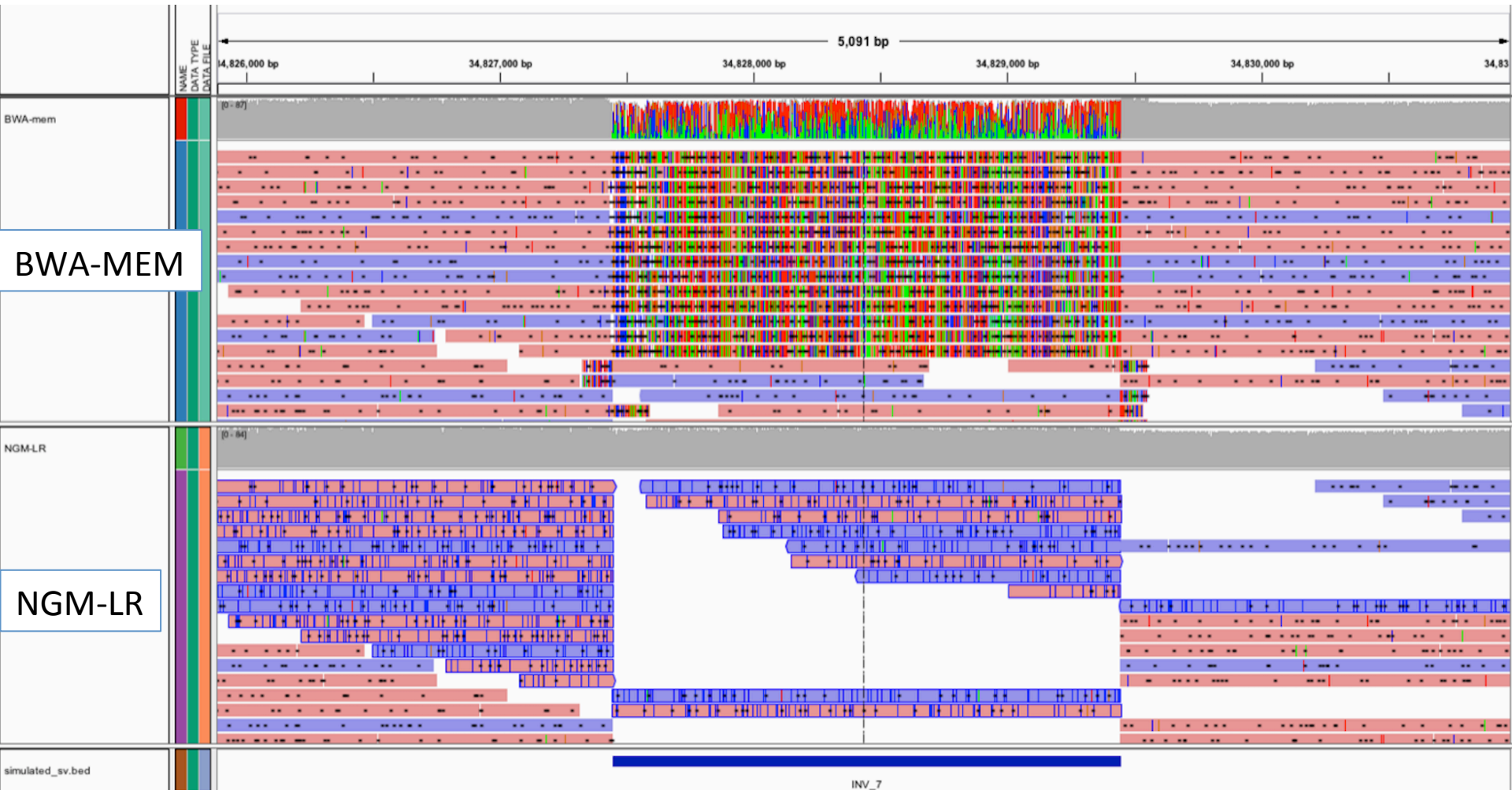
NGM-LR reconcile

Dot plot of 500bp segments from a read spanning a inversion

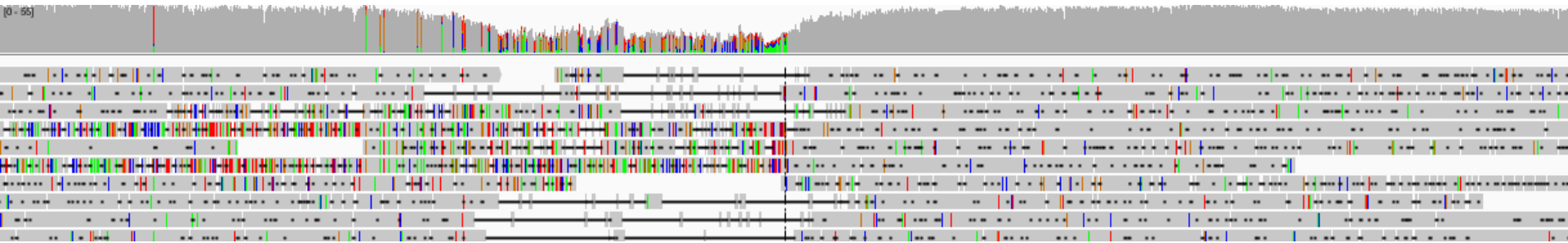
925758/52



NGM-LR inversion



Convex Pairwise Alignment

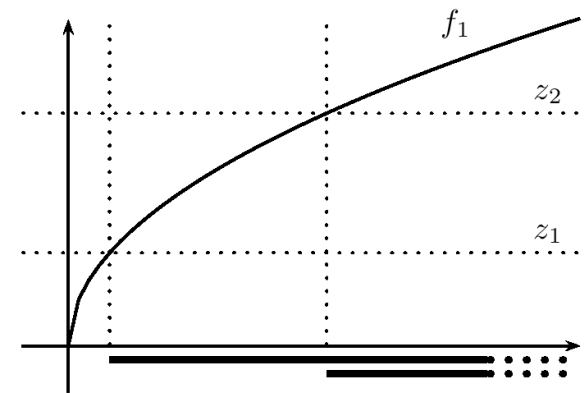


AAAGAATTCA
A-A-A-T-CA

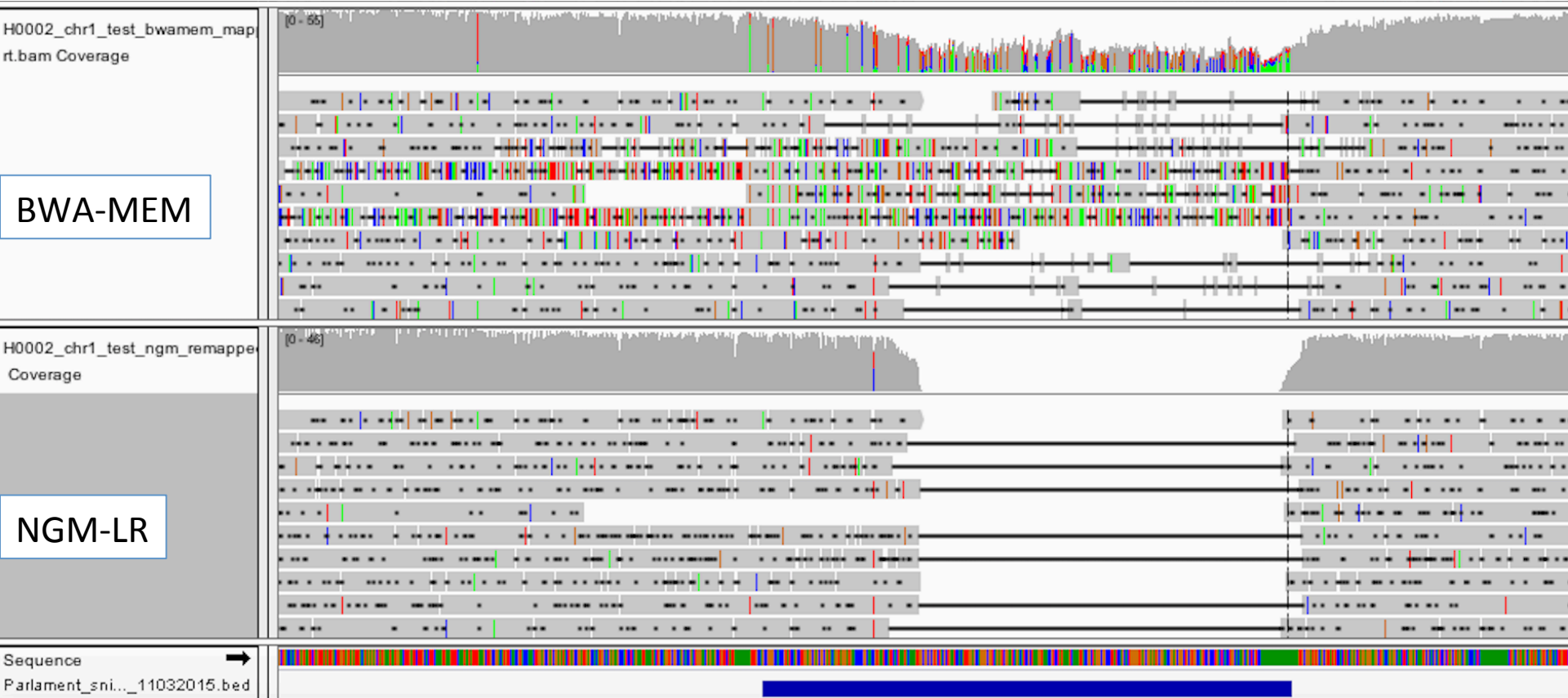
vs.

AAAGAATTCA
AAA-----TCA

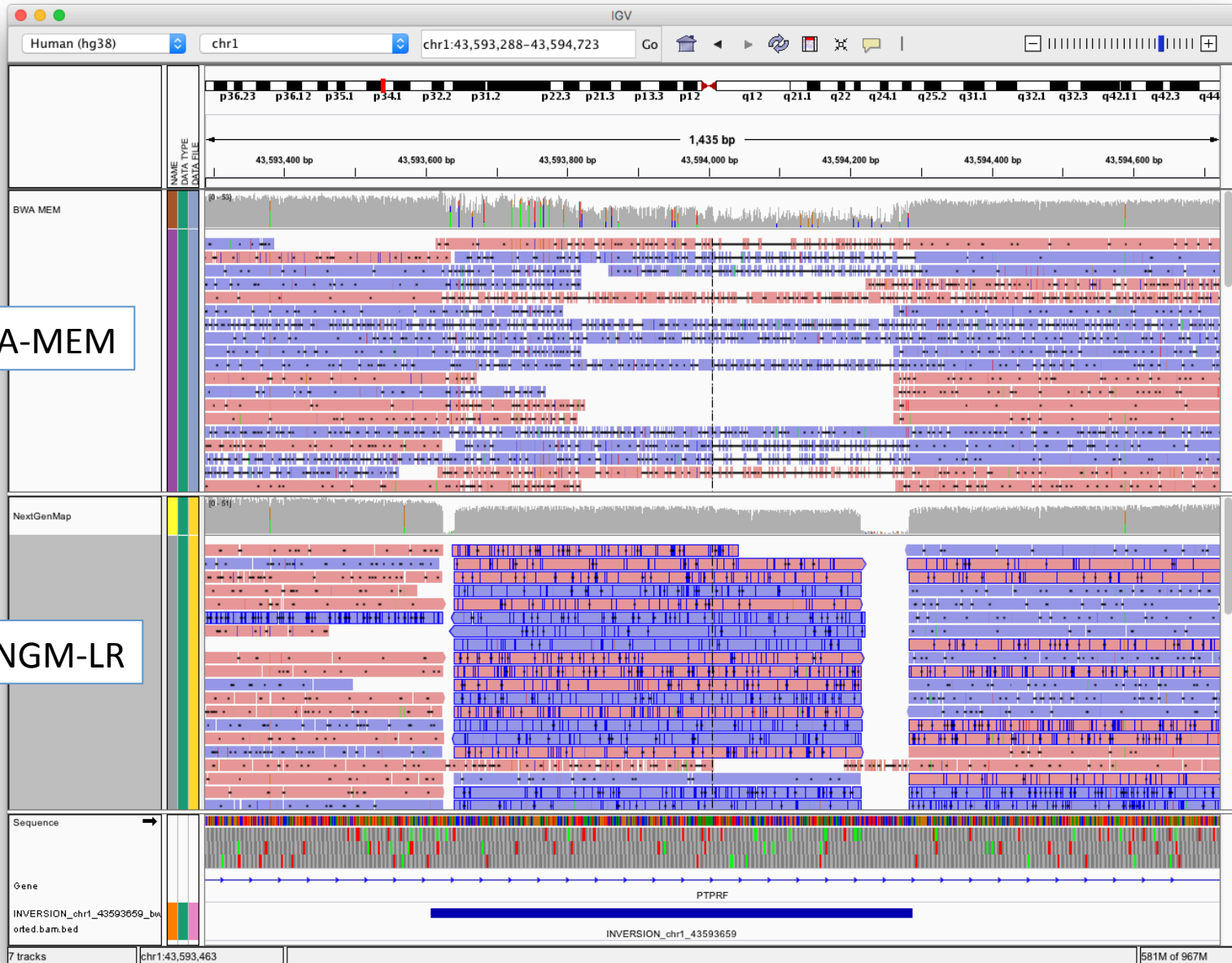
- **Linear:** gap cost always the same
- **Affine:** separate penalties for opening and extending a gap
- **Convex:** initially similar to affine, but becomes proportionally less costly for larger gaps



NGM-LR deletion



NGM-LR complex SV



Sniffles

- Analyzing split reads, alignment events and noisy regions.
- Parameter estimation
- Optional: Genotype estimation
- Optional: Clustering of SVs




Analyzing noisy regions

- Extract the differences in the alignment
- Detect the noisy regions:
Plane sweep algorithm
- Store potential regions in a self balancing binary tree.



Simulation/Evaluation

1. Simulate 20 SVs of each type using SURVIVOR
2. Simulate Pacbio and illumina paired end reads
3. Evaluation using SURVIVOR

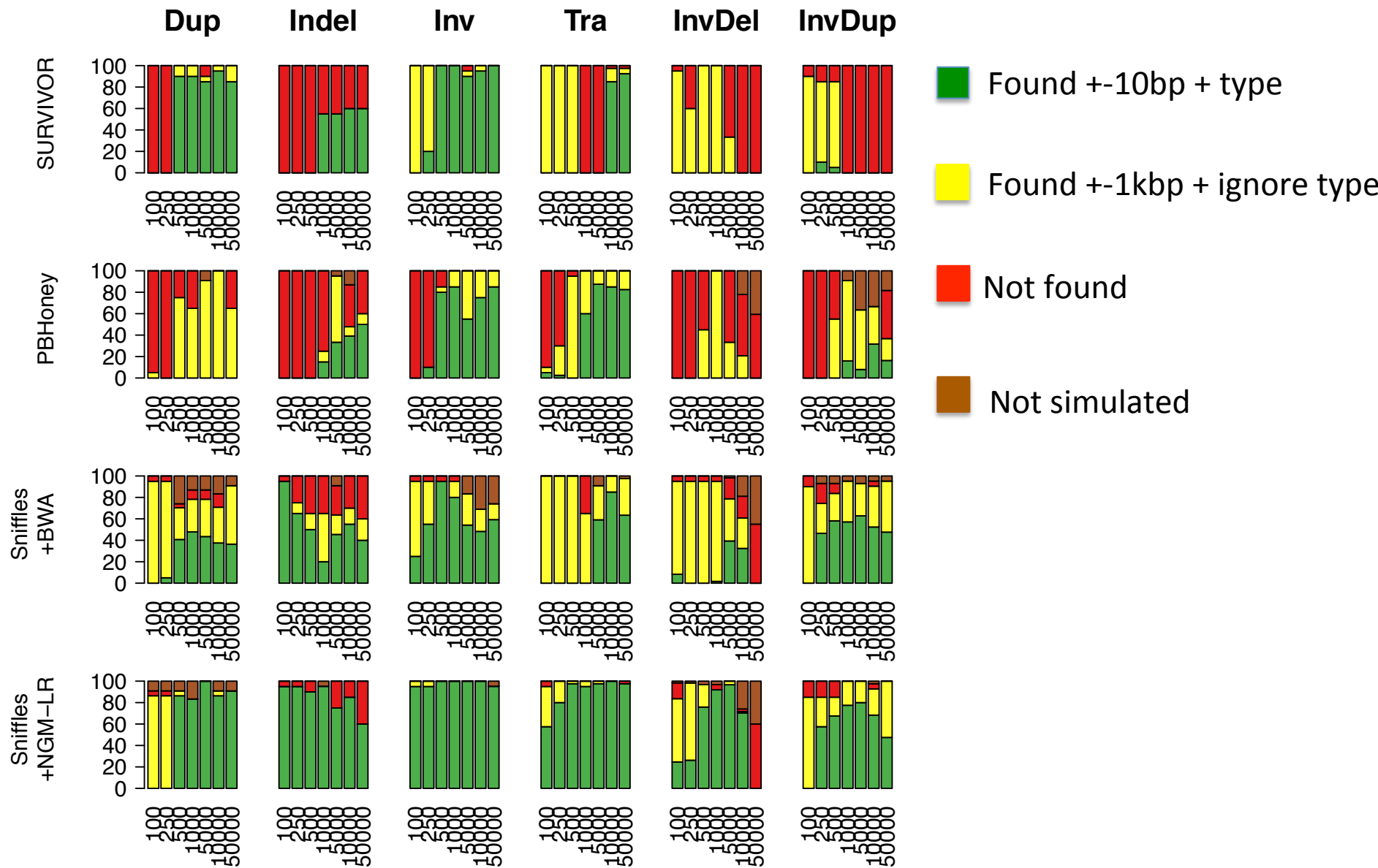
 Found ± 10 bp + type

 Found ± 1 kbp + ignore type

 Not found

 Not simulated

Evaluation of Sniffles



Summary

NextGenMap:

- Short read version:
github.com/cibiv/NextGenMap
- Long read mapper:
github.com/philres/nextgenmap-lr
 - Self detection of SVs
 - Manuscript in preparation

Sniffles:

- SVs detection for long reads
- Nested SV
- Manuscript in preparation
- Available:
github.com/fritzsedlazeck/Sniffles

SURVIVOR:

- Toolkit for SV detection on short reads
- Simulation/Evaluation of current methods
- Consensus approach
- Accepted Nature Communications
- Available:
github.com/fritzsedlazeck/SURVIVOR

Future work:

- How much coverage is needed?
- Nanopore support
- Analysis of nested SVs
- Application to Cancer Genomes (See **Maria Nattestad's Poster #79**)

Acknowledgments



Cold
Spring
Harbor
Laboratory

Maria Nattestad

Han Fang

Srividya Ramakrishnan



Daniel Jeffares

Jürg Bähler

Christophe Dessimoz



universität
wien

Philipp Rescheneder

Moritz Smolka

Arndt von Haeseler



JOHNS HOPKINS
UNIVERSITY

Michael Schatz

NGM-LR nanopore



Evaluation of Sniffles: SKBR3

