

Tigmint

Correct Misassemblies Using Linked Reads From Large Molecules

Shaun Jackman [@sjackman](#)

Benjamin P Vandervalk, Rene L Warren, Hamid
Mohamadi, Justin Chu, Sarah Yeo, Lauren Coombe,
Joerg Bohlmann, Steven JM Jones, Inanc Birol

ISMB HiTSeq 2017-07-25

<https://sjackman.ca/tigmint-slides>



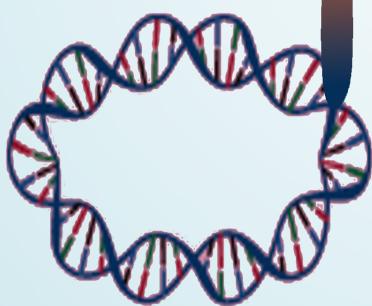
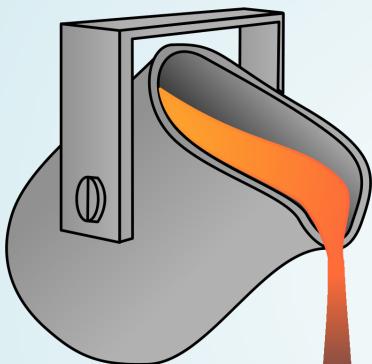
Shaun Jackman

BCCA Genome Sciences Centre
Vancouver, Canada

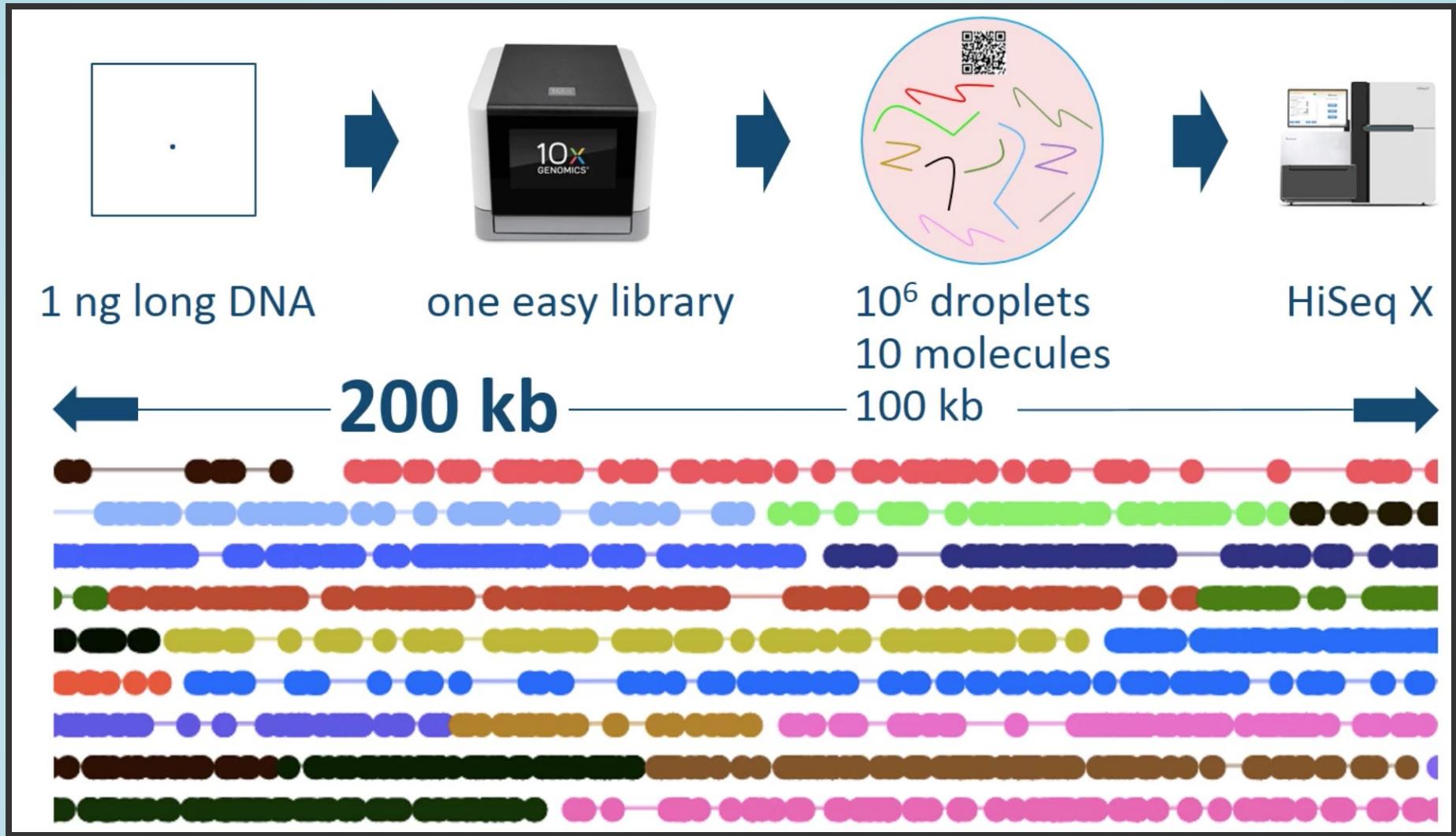
@sjackman · github.com/sjackman · sjackman.ca



Tigmint



<https://github.com/bcgsc/tigmint>



10x Genomics Chromium Linked Reads
<http://www.10xgenomics.com/assembly/>

Linked Reads

- Call variants in repetitive regions
- Identify structural variants
- Phase variants

Tools for Linked Reads

10x Genomics

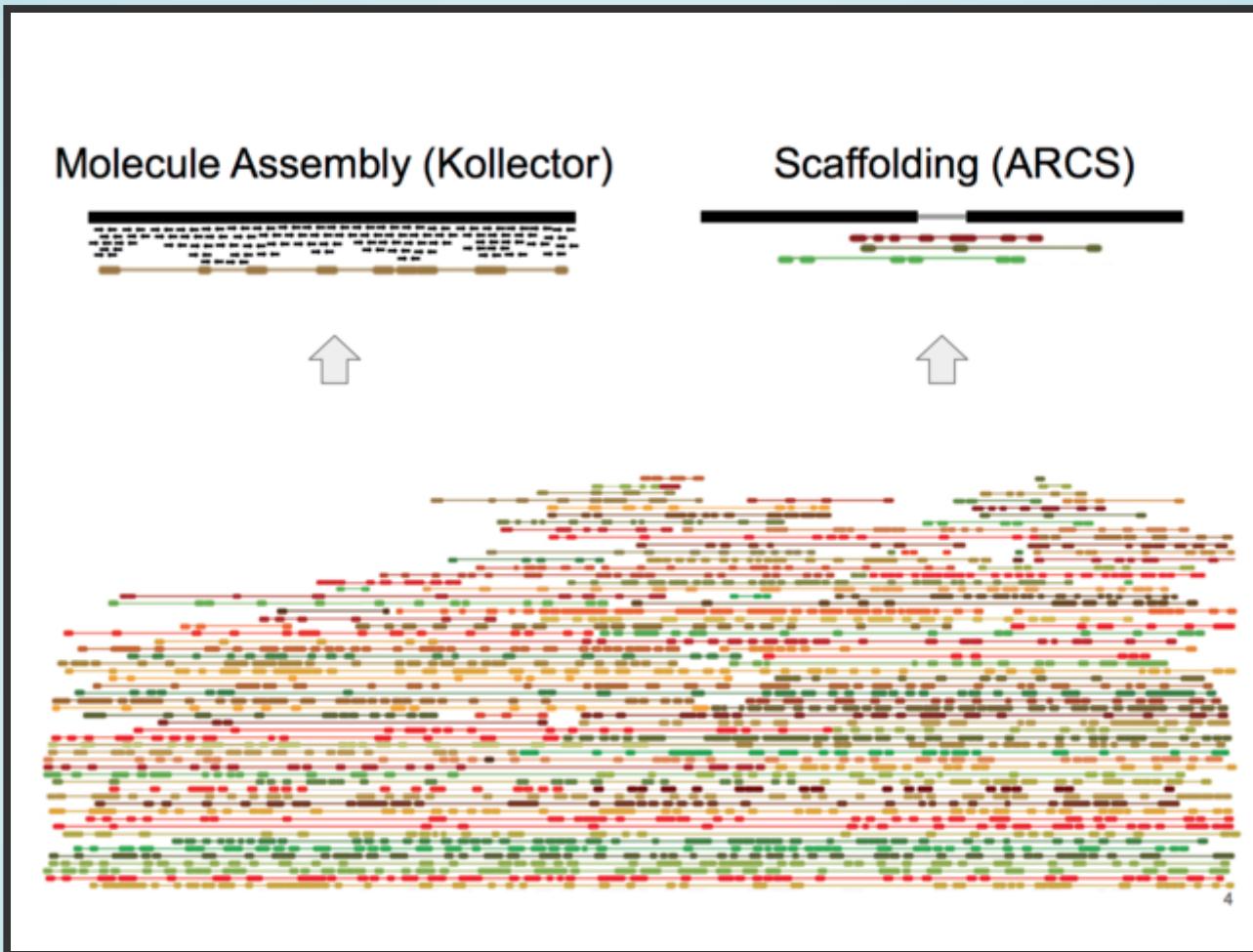
Long Ranger · Supernova

Structural Variants

GROC-SVs · Topsorter

Scaffolding

ARCS · Architect · Fragscuff



B-040 · Local assembly with Kollector · Justin Chu
B-048 · ABYSS 2.0 · Ben Vandervalk

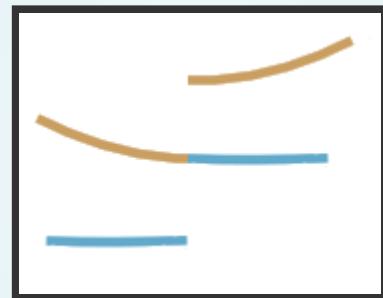
Contigs and scaffolds come to an end due to...

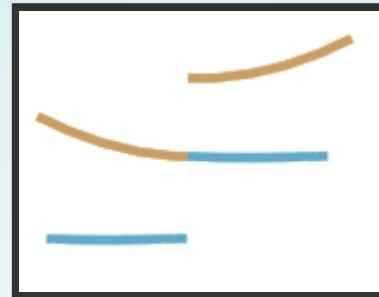
- repeats
- sequencing gaps
- heterozygous variation
- misassemblies

Misassemblies limit contiguity

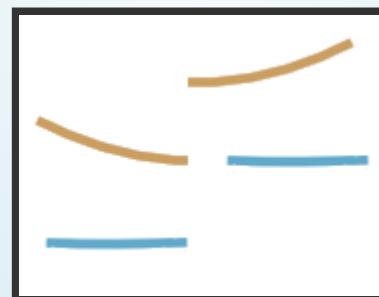
particularly for highly contiguous assemblies.

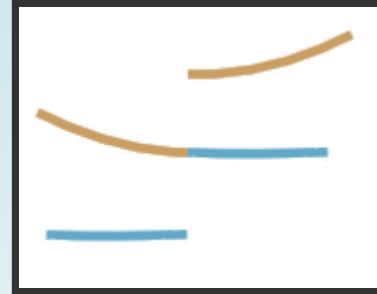
Most scaffolding algorithms address repeats and gaps, but not misassemblies.





Correct misassemblies





Correct misassemblies

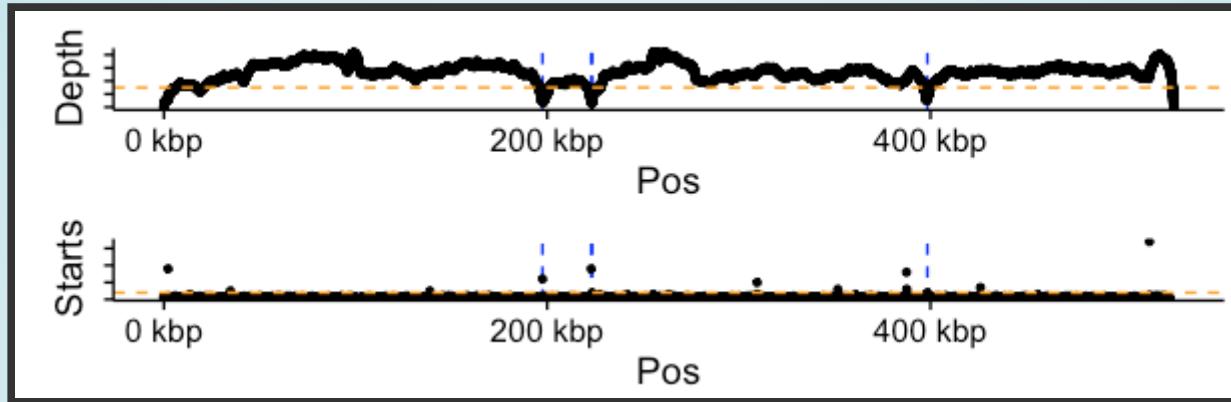


Scaffold

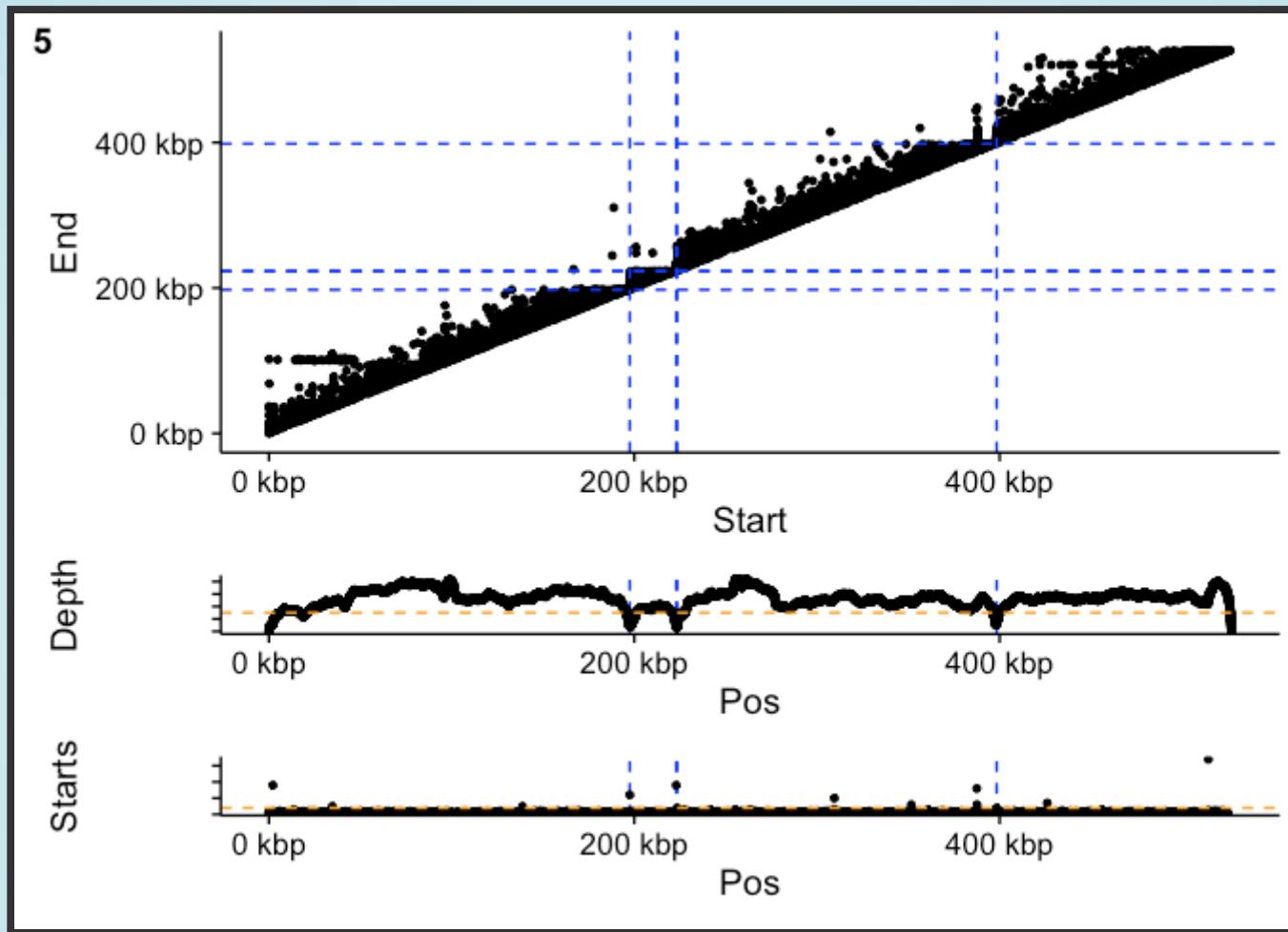


Tigmint

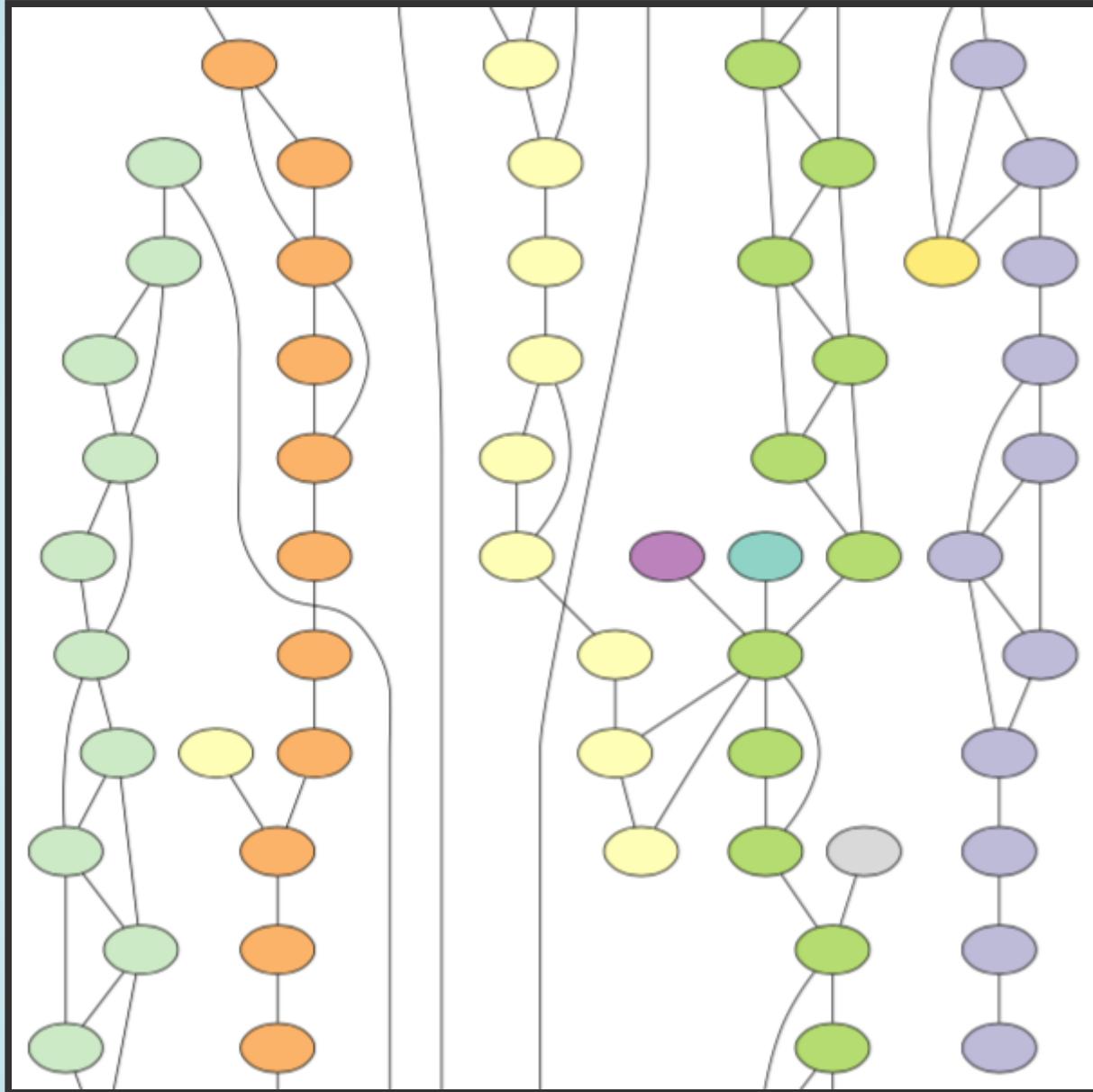
- Count molecules that support and refute each position of the assembly
- Output positions of possible misassemblies



- Regions with poor molecule depth are suspect
- Clipped read alignments at molecule ends pinpoint the position of the breakpoint
- Highlight possible misassemblies

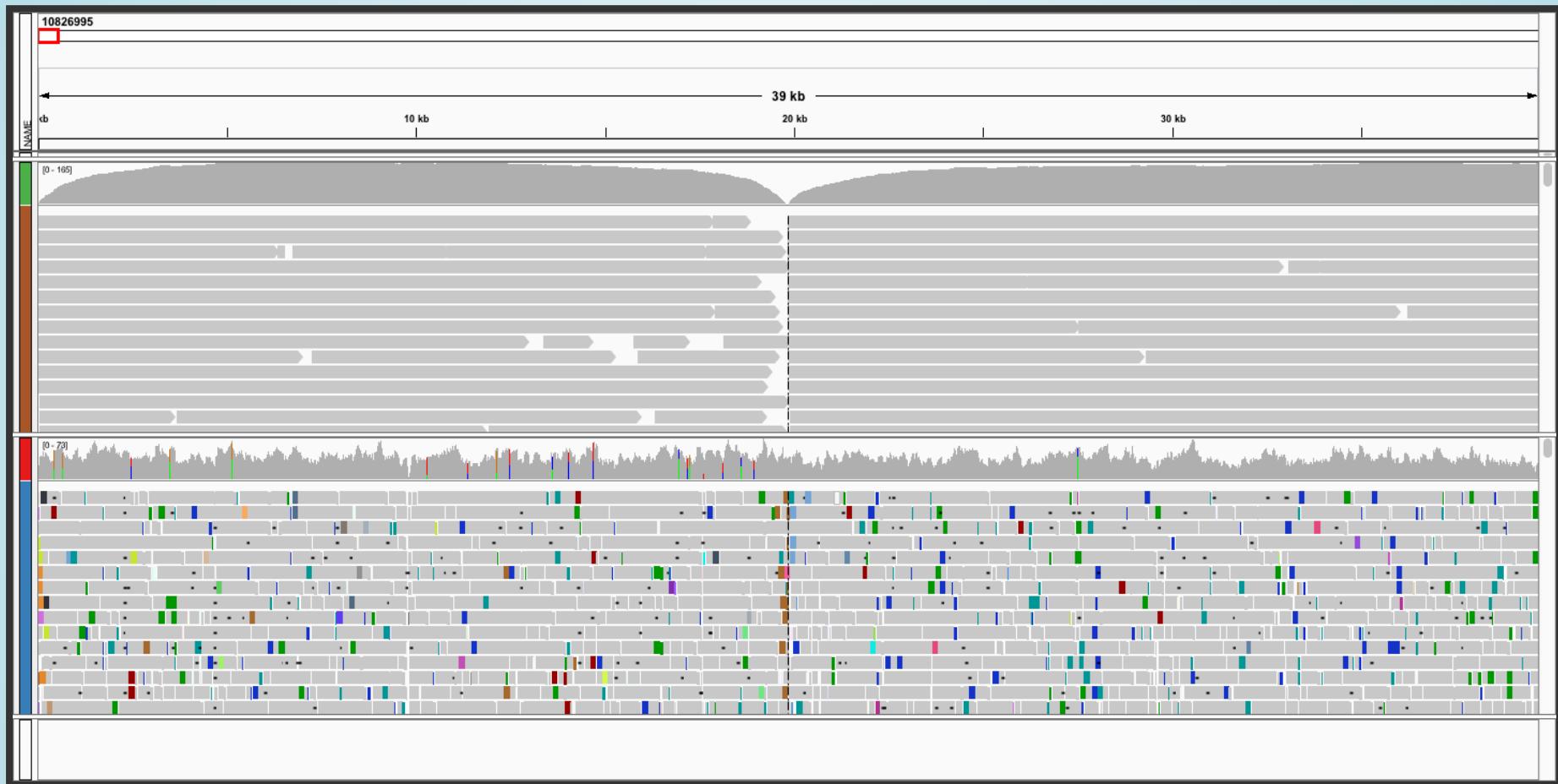


Scatter plot of molecule start and end coordinates

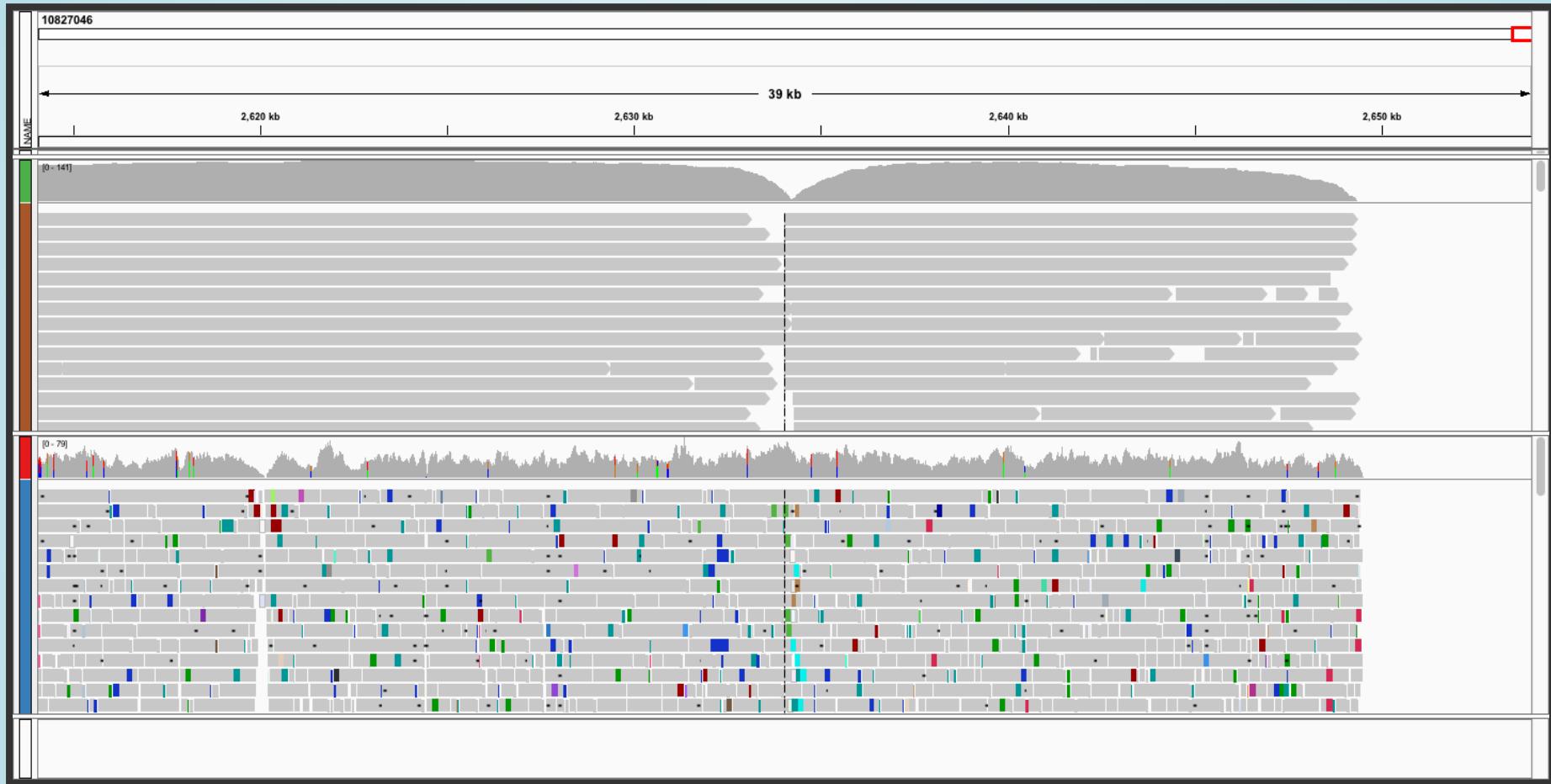


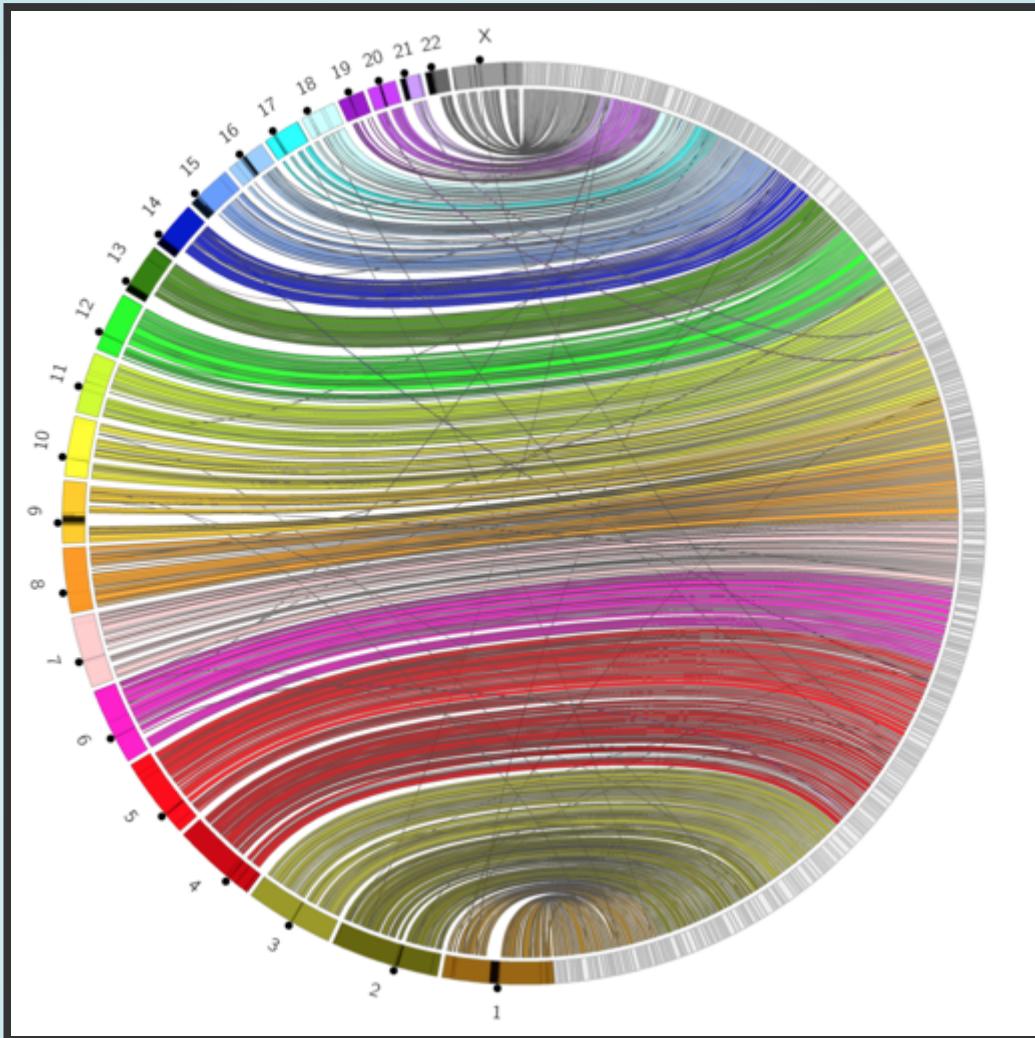
Menagerie of Misassemblies

- **Chimeric sequence**
- **Missing sequence** (deletion)
- Chimeric insertion
- Inversion
- Collapsed repeat

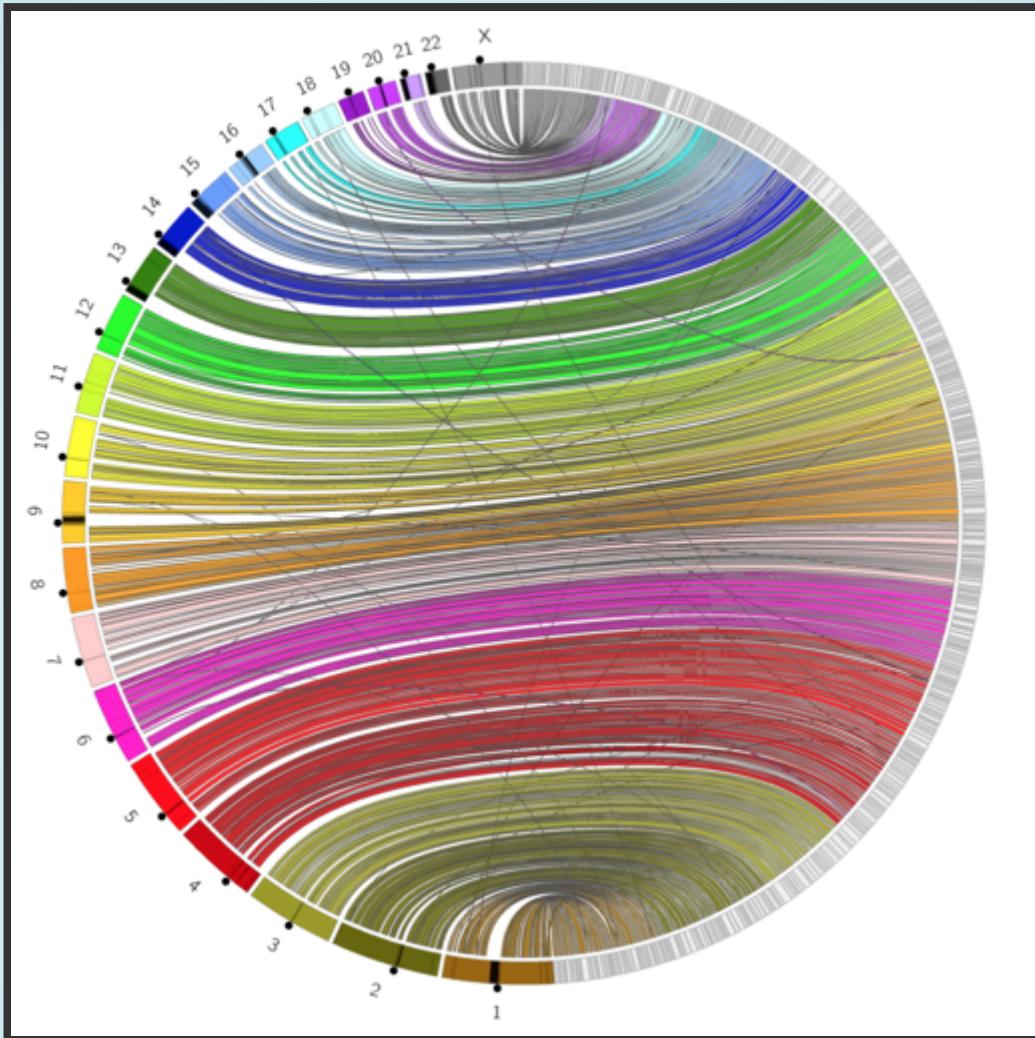




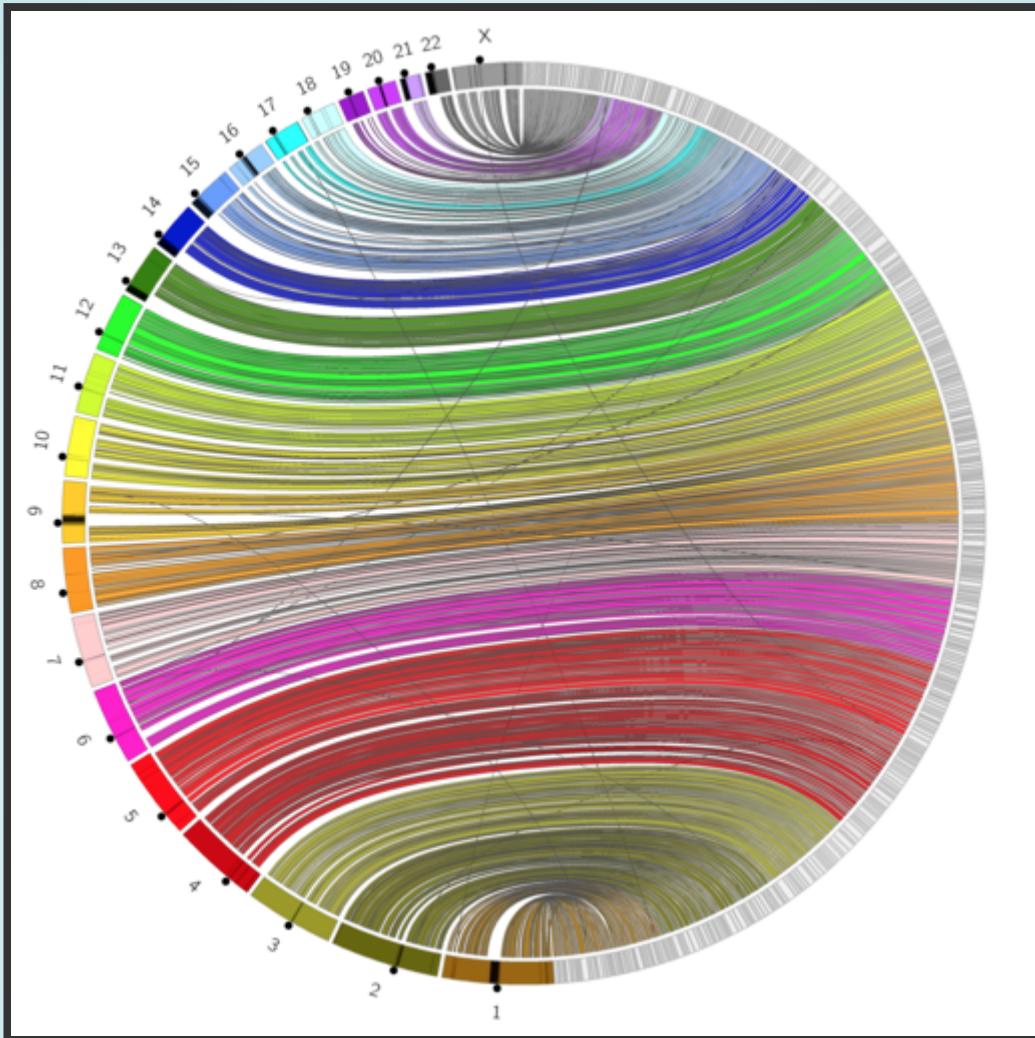




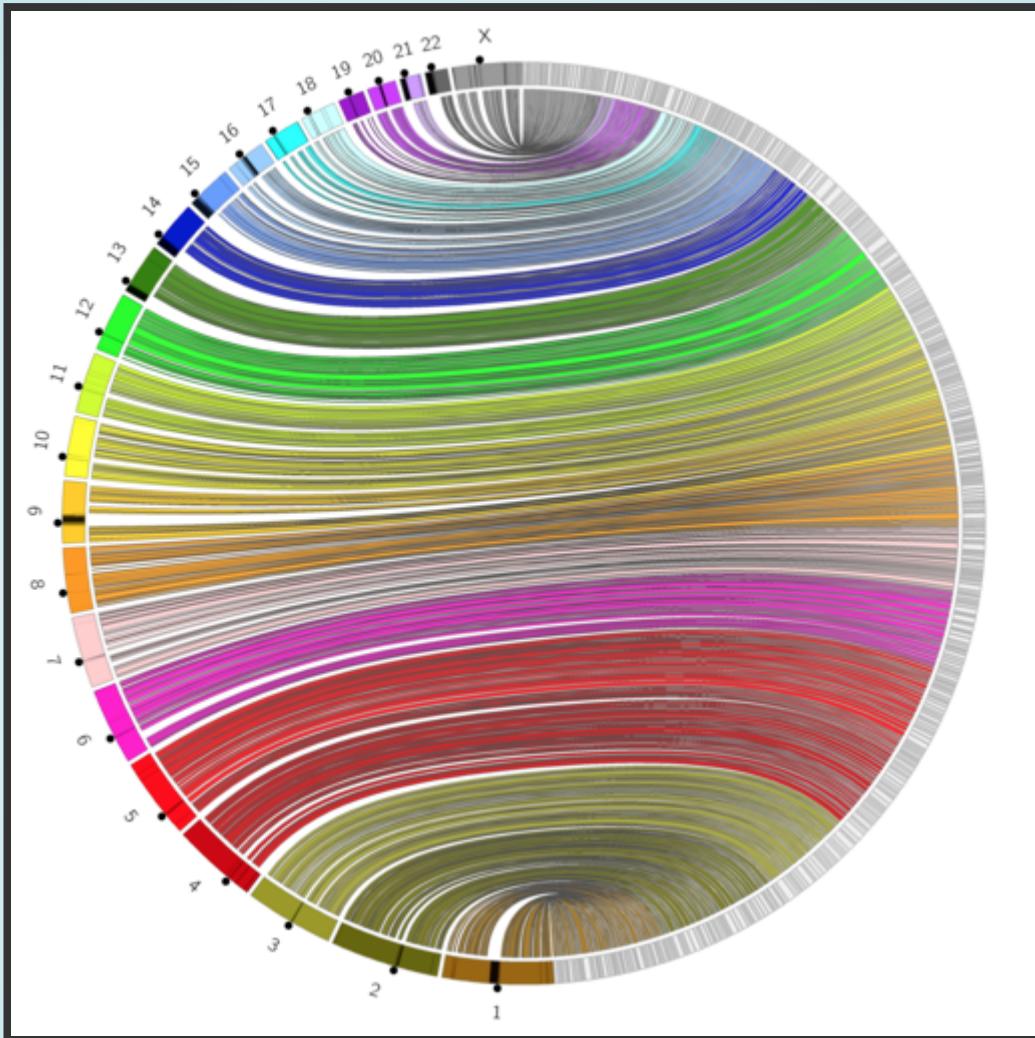
ABYSS 2.0 assembly of GIAB HG004



Tigmint, at least 4 clipped molecules

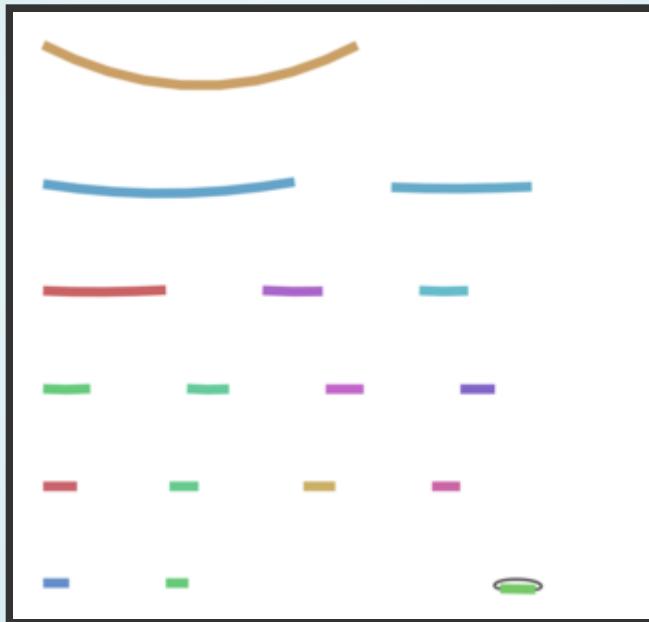


Tigmint, at least 3 clipped molecules



Tigmint, at least 2 clipped molecules

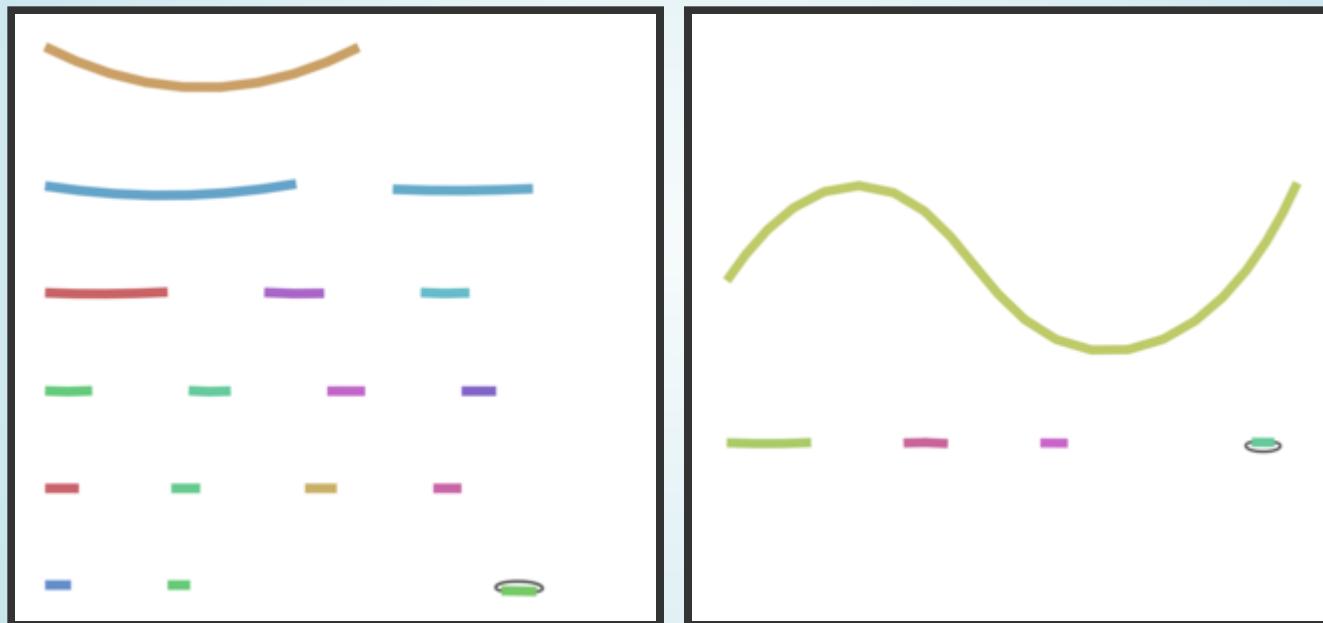
Sitka Spruce Mitochondrion



16 scaffolds, N50 is 0.46 Mbp

Sitka Spruce Mitochondrion

Before and After Tigmint



N50 improved nine fold from 0.46 Mbp to 4.2 Mbp
Poster B-066

fin

Shaun Jackman

BCCA Genome Sciences Centre
Vancouver, Canada

@sjackman · github.com/sjackman · sjackman.ca

Tigmint

<https://github.com/bcgsc/tigmint>

Slides

<https://sjackman.ca/tigmint-slides>

Markdown source code

<https://github.com/sjackman/tigmint-slides>

Links

ABySS · ARCS · Architect · Fragscaff · GROC-SVs ·
LINKS · Long Ranger · MAKER · Pilon · Prokka ·
Sealer · Supernova · Tigmint · Topsorter