

Long read only genome sequencing and assemblies

Weihong Qi, Andrea Patrignani, Ralph Schlapbach

Functional Genomics Center Zurich, ETH Zürich / University of Zurich

Summary

We sequenced a bacterial genome de novo using both MinION and PacBio sequencing technologies. Long read only assemblies were produced and polished using different pipelines. Assembly accuracy was benchmarked using Illumina paired-end reads from the same sample. We found that data analysis pipelines had significant influence on the amount of residual errors in long read only assemblies.

Table 1. Summary statistics of the reads

| | MinION reads | PacBio Subreads | Illumina PE reads |
|--------------|--------------|-----------------|-------------------|
| # reads | 197,560 | 161,250 | 5,264,891 |
| # bases (Mb) | 873 | 1,369 | 1,316 |
| N50 (Mb) | 6.11 | 11.269 | 2X125bp |
| GC% | 52.2% | 51.75% | 51.21% |
| Coverage | 175X | 274X | 263X |

Figure 1. Assembly and evaluation workflows

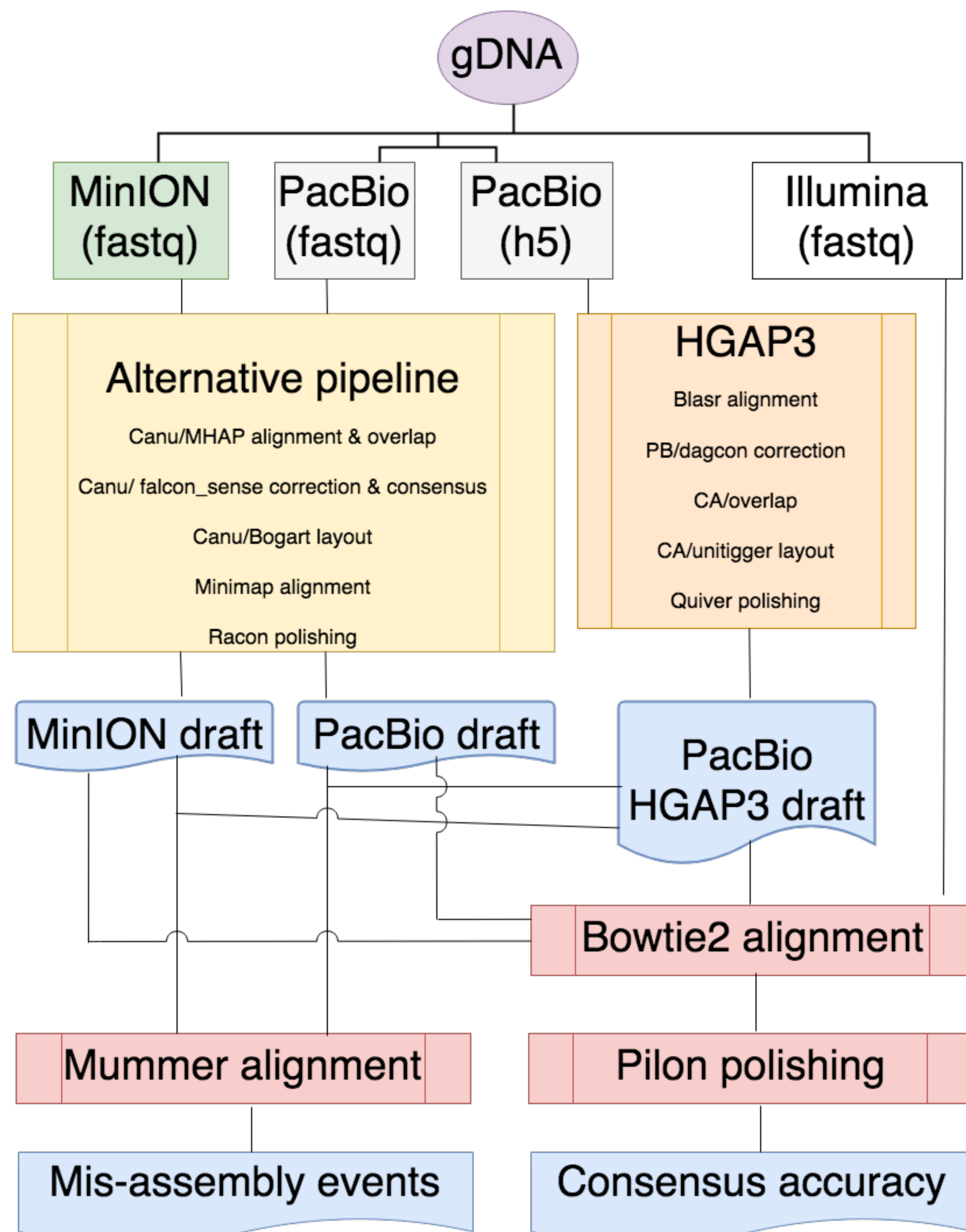


Table 2. Summary statistics of the assemblies

| | MinION draft | PacBio draft | PacBio HGAP3 draft |
|----------------------|--------------|--------------|--------------------|
| # contigs | 4 | 5 | 6 |
| Assembly size (Mb) | 5.150 | 5.161 | 5.179 |
| Longest contigs (Mb) | 4.909 | 4.905 | 4.905 |
| N50 (Mb) | 4.909 | 4.905 | 4.905 |
| L50 | 1 | 1 | 1 |
| GC% | 52.12 | 52.11 | 52.08 |

Figure 2. Dot plots of assemblies from the alternative pipeline as they are aligned to the HGAP3 pipeline. MinION and PacBio drafts from the alternative pipeline both contain similar relocation events.

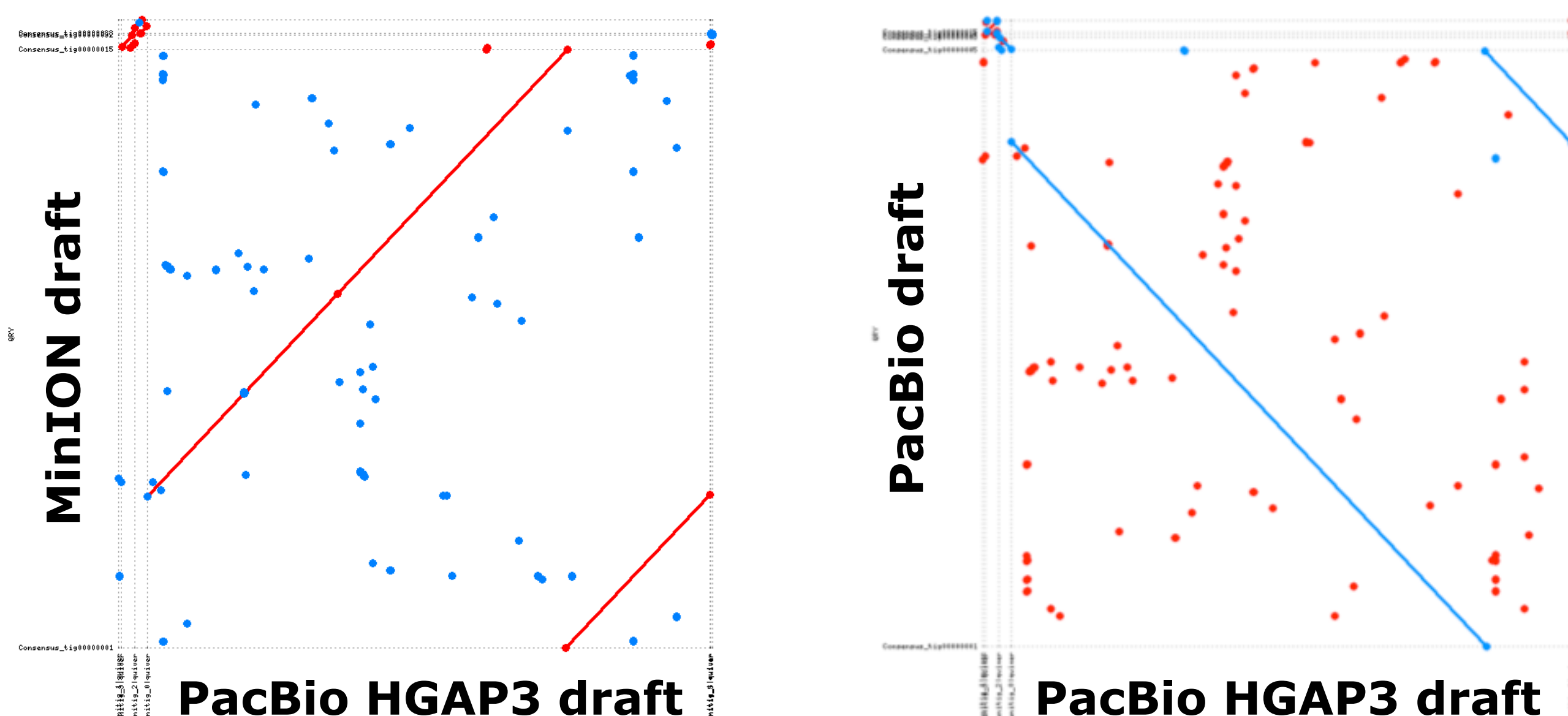
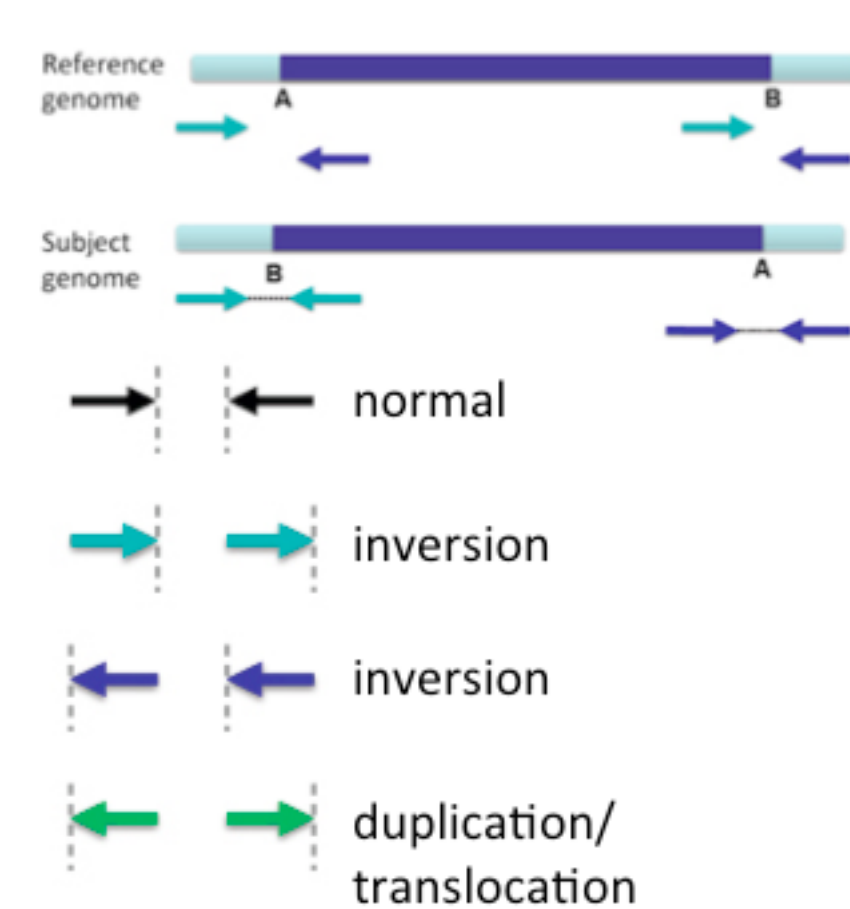
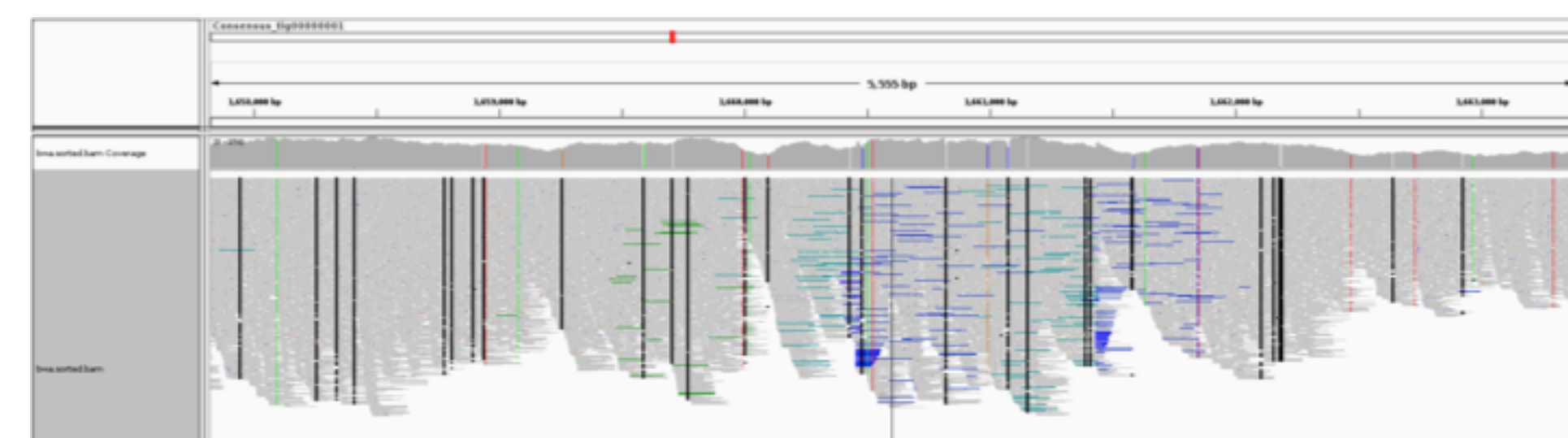


Figure 3. Alignment of Illumina Paire-end reads around the breakpoints of the relocation events in three different assemblies.

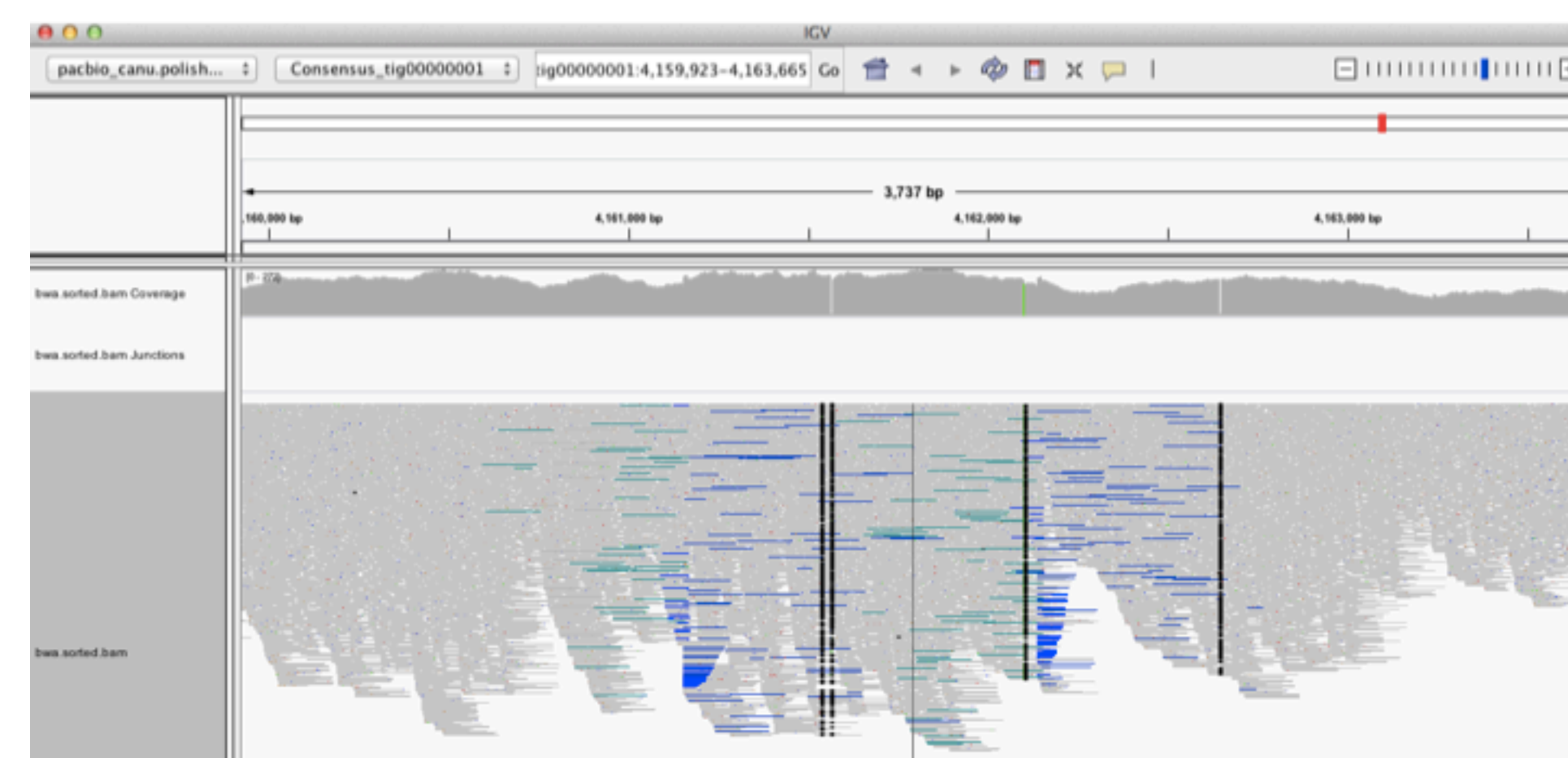


Alignment of Illumina PE reads showed expected read orientation and insert size in the assembly from the HGAP3 pipeline, but not in the two assemblies from the alternative pipeline. More sequence variations were also observed in these two assemblies.

MinION draft



PacBio draft



PacBio HGAP3 draft

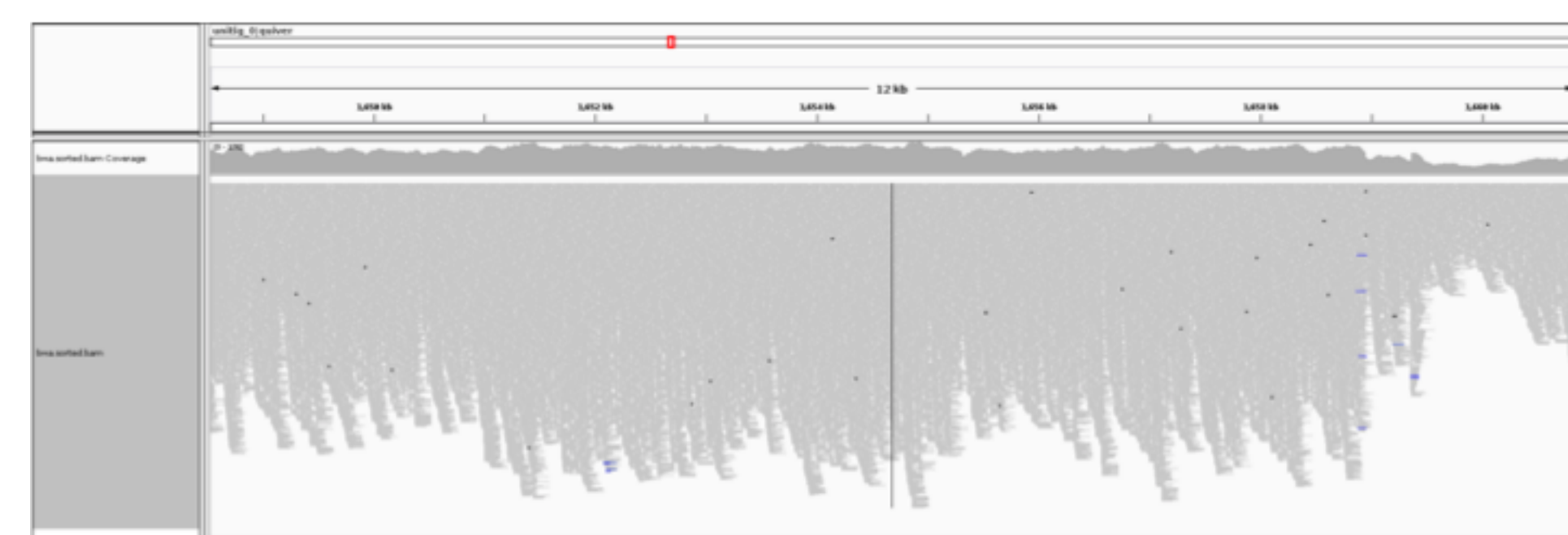


Table 4. Summary of pilon correction using Illumina data

| | # SNV (SNP + InDel) | # SV (>=2bp) |
|--------------------|-----------------------|--------------|
| MinION draft | 20,712 (8,603+12,109) | 5,624 |
| PacBio draft | 2,042 (132+1,910) | 976 |
| PacBio HGAP3 draft | 23 (1+22) | 1 |

