APPENDIX | COVERAGE INTERPRETATION

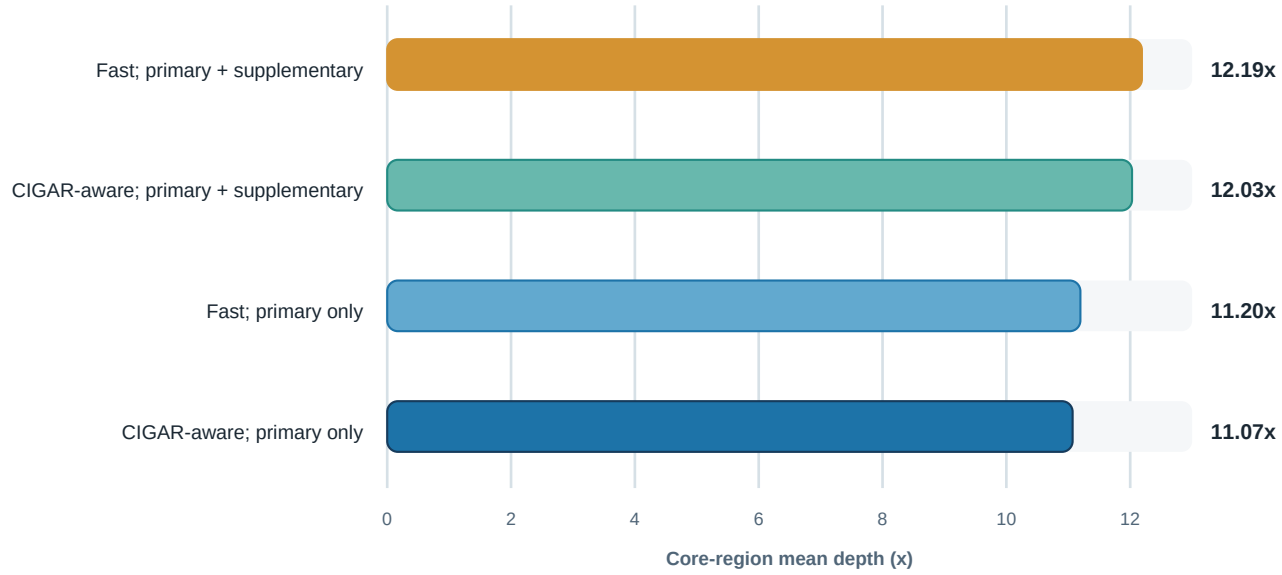
HG002 Take101 ONT mosdepth: what 12.19x includes

A bounded remeasurement of the already produced Bjuice-prevalidation long-read CRAM, separating two independent choices: mosdepth **fast mode versus CIGAR-aware depth**, and **supplementary-inclusive versus primary-only records**.

Direct answer. The released 12.19x core-region mosdepth value already includes supplementary records and uses fast mode. Removing fast mode gives **12.03x** while still including supplementaries. The conservative CIGAR-aware, primary-only value is **11.07x**. Supplementaries therefore contribute 0.96x of reference-alignment depth in the CIGAR-aware comparison, but that increment is not additional unique molecules.

RELEASED PROFILE	NO FAST MODE	CONSERVATIVE DEPTH
12.19x	**12.03x**	**11.07x**
fast; primary + supplementary	CIGAR-aware; primary + supplementary	CIGAR-aware; primary only

Same CRAM, four explicit measurement profiles



Decomposition: supplementary inclusion adds 0.96x (8.7% over primary-only) in the CIGAR-aware comparison. Fast mode adds 0.16x (1.3%) when supplementaries are retained.

MEASUREMENT DEFINITIONS

The reported number is a profile, not a property of the CRAM alone

Each result uses the identical CRAM, reference, core-region BED, MAPQ threshold, and thread count. Only the two variables in the table change. Means below are recomputed from integer covered-base totals divided by the shared 2,889,585,716-base denominator; the parenthetical value is mosdepth's two-decimal rendering.

Profile	Fast	Supplementary	-F mask	Core covered bases	Core mean
Released / reproduced	yes	included	1796	35,220,785,261	12.188870x (12.19x)
CIGAR-aware segment depth	no	included	1796	34,761,383,177	12.029885x (12.03x)
Fast primary-only	yes	excluded	3844	32,351,023,925	11.195731x (11.20x)
Recommended baseline	no	excluded	3844	31,987,811,792	11.070034x (11.07x)

What each switch does

Switch	Operational meaning	Observed HG002 effect
Remove --fast-mode	Honor internal CIGAR operations. In long reads, reference skips/deletions are no longer treated as continuously covered by the alignment envelope.	-0.16x with supplementaries; -0.13x primary-only
Add 0x800 to -F	Exclude supplementary alignment records as well as unmapped, secondary, QC-failed, and duplicate records (1796 + 2048 = 3844).	-0.96x in CIGAR-aware mode; -0.99x in fast mode

Interpretation guardrail. The 12.19x released value is valid for its declared profile: nonsecondary segment depth using fast-mode envelopes. It should not be labeled primary-only coverage, unique-molecule yield, or caller-effective coverage.

Why supplementary depth is not simply 'extra coverage'

- A supplementary record represents another alignment segment from the same original QNAME; it is not another sequenced molecule.
- Segments at distinct loci can add legitimate breakpoint or split-read evidence.
- Overlapping primary and supplementary reference segments can both contribute to reference depth, so segment depth may double-count a molecule locally.
- A QNAME-level union of query coordinates answers a different question: how much of each raw molecule is explained by one or more nonsecondary alignments.

Independent morphology cross-check. In the prior deterministic 2% QNAME sample, primary alignments explained 90.10% of mapped-read query bases and the primary-plus-supplementary union explained 97.96%. That 7.86-point recovery is 8.7% relative to the primary amount - essentially the same scale as the 8.67% reference-depth increment here. The data support modest recovery, not a 2x increase.

SENTIEON INTERPRETATION

The CRAM supplied to a caller is known; its internal evidence policy is not

The exact Take101 Hybrid CLI 1.7.1i log (Sentieon Genomics 202503.03; HybridIlluminaONT2.0.bundle) passes the complete long-read CRAM to multiple algorithms. No command-line supplementary or secondary filter is shown for the listed calls. That proves input binding, but not which records survive each algorithm's internal filtering, genotyping, or evidence aggregation.

Algorithm in exact run	Observed input	What can be concluded	What remains unknown
LongReadSV	Complete LR CRAM	Split alignment evidence is available to the process.	Exact primary/supplementary/secondary inclusion and weighting policy.
Hybrid DNAscope / pass 2	Complete LR CRAM plus SR evidence	No upstream DayOA record-mask removal in the logged command.	Which chimeric or supplementary records contribute to small-variant genotyping.
HybridStage1/2/3	Complete LR CRAM in logged stages where bound	The LR evidence stream is present.	Stage-specific record filtering, normalization, and contribution to final calls.

Therefore there is no single defensible 'coverage Sentieon is working with' number. For the exact input CRAM, report 11.07x CIGAR-aware primary depth and 12.03x CIGAR-aware nonsecondary segment depth as a bounded evidence interval. Do not market the upper value as caller-effective coverage until Sentieon confirms algorithm-specific record handling.

Why the distinction is material

- Small-variant calling usually needs independent molecule support and calibrated likelihoods; a split record is not an independent observation.
- Structural-variant calling specifically benefits from split/supplementary geometry at breakpoints; discarding it can remove valuable evidence.
- Secondary records are alternative placements, not extra molecule evidence, and should remain excluded from depth unless a specialized analysis explicitly models them.
- Sentieon's public materials document whole-CRAM inputs and model workflows, but they do not expose an auditable record-by-record contribution policy for every algorithm used here.

Vendor question that would close the remaining gap

For DNAscope LongRead/hybrid DNAscope, HybridStage1/2/3, and LongReadSV, please specify whether SAM secondary (0x100) and supplementary (0x800) records are ignored, retained, reconstructed from SA tags, de-duplicated by QNAME/query coordinates, or down-weighted; and whether this differs by SNV/indel versus SV evidence.

PRACTICE AND RECOMMENDATION

Ignoring every supplementary record is not accepted universal practice

The appropriate treatment depends on the scientific unit. Supplementaries are often excluded from one-number primary depth, intentionally retained for split-read/SV evidence, and unioned by QNAME for molecule-completeness analysis. The mistake is not choosing one policy; it is publishing an unlabeled number whose counting unit is ambiguous.

Question	Recommended unit/profile	Supplementary treatment	Secondary treatment
Routine mapped depth / coverage thresholds	CIGAR-aware primary-only reference depth	Exclude; publish mask 3844	Exclude
Available split-alignment reference evidence	CIGAR-aware nonsecondary segment depth	Include; publish mask 1796 and warn about same-QNAME overlap	Exclude
Raw molecule utilization / clipping recovery	QNAME-level primary + supplementary query-coordinate union	Union once per query coordinate	Exclude as alternative placements
SV breakpoint evidence	Caller-specific split-read and discordant-locus evidence	Retain or reconstruct according to caller specification	Caller-specific; not independent support
SNV/indel effective coverage	Caller-specific contributing-read or likelihood accounting	Do not infer from CRAM presence alone	Do not infer from CRAM presence alone

A compact production reporting stack

Layer	Metric	Purpose
1. Primary depth	mosdepth, no --fast-mode, -F 3844	Stable conservative reference coverage for cross-sample reporting.
2. Split-evidence depth	mosdepth, no --fast-mode, -F 1796	Quantifies the increment available in supplementary alignment segments.
3. Molecule completeness	Per-QNAME union of primary + supplementary query intervals	Explains raw-read versus aligned-query utilization without double-counting query bases.

Recommended wording for this HG002 release. 'ONT depth over the declared core regions is 11.07x using CIGAR-aware primary alignments. Including nonsecondary supplementary segments yields 12.03x of split-alignment evidence. The historical 12.19x mosdepth value used fast mode and included supplementaries.'

Is the field making a mistake?

A blanket claim that the field 'ignores ONT supplementaries' is not supported. Many general-purpose depth summaries are intentionally primary-centric, while long-read SV callers use split evidence. There *is* a recurring reporting problem: raw yield, primary reference depth, nonsecondary segment depth, and caller-effective evidence are often collapsed into one word - coverage. Separating those units improves both scientific interpretation and reproducible product claims.

IGV VISUALIZATION GUIDANCE

Configure IGV to expose the molecule, not just the primary segment

IGV Desktop has dedicated chimeric-read features for primary and supplementary segments linked by the SAM SA tag. These views are more informative than simply turning soft clips on: they show reference-order and read-order relationships, orientation, and remote loci. The features require a valid SA tag in the CRAM.

Recommended morphology-inspection preset

Setting / action	Recommendation	Reason
Experiment Type	Right-click track -> Experiment Type -> 3rd Gen	Activates long-read defaults and the Third Gen preference profile.
Supplementary filter	If exposed by the installed IGV version, leave 'Filter supplementary alignments' OFF	The supplementary records must remain visible for split-read morphology.
Secondary filter	Keep secondary alignments filtered for the default audit view	Secondary records are alternative placements, not additional molecule segments.
Soft-clipped bases	Turn ON while inspecting clipping and segment boundaries	Makes the sequence outside each displayed segment visible; turn off for uncluttered locus screenshots.
Mapping quality	Keep the display threshold at 0; Shade alignments by -> mapping quality low	Retains ambiguous split evidence while visually de-emphasizing low-MAPQ records. Make a separate MAPQ>=20 view if needed.
Downsampling	Leave Third Gen downsampling off where practical; otherwise zoom narrower before raising limits	Avoids hiding the linked segment of interest. IGV marks downsampled windows with a black bar.
SA tag visibility	Remove SA from Hidden SAM tags if the installed version hides it	Makes the segment-link evidence reviewable in read details; linking itself still depends on a valid SA tag.

Use the chimeric-read tools in this order

Step	IGV action	Question answered
1	Right-click alignment track -> Group by -> chimeric	Which visible reads carry linked split-alignment structure?
2	Sort by -> left clip or right clip	Are long terminal clips concentrated at this locus?
3	Right-click a read -> Supplementary Reads Diagram	How do reference-order segments map onto original read order and orientation?
4	Right-click -> View chimeric reads in split screen	Where do the linked segments land, including other chromosomes?
5	View -> Preferences -> Alignments -> Chr labels ON	Which chromosome is connected when a supplementary segment is interchromosomal?

Do not infer absence from a single locus pane. A supplementary segment can be on a different chromosome or outside the current visibility window. Use the Supplementary Reads Diagram or split-screen view before concluding that a soft-clipped portion lacks a corresponding alignment.

Snapshot convention: record IGV version, reference, CRAM, locus, MAPQ threshold, supplementary/secondary filters, soft-clip state, and downsampling. Pair the locus view with the Supplementary Reads Diagram for molecule-level claims. Soft-clip BLAT is exploratory, not a replacement for the production alignment.

SEQONE / SEQ1 CLOUD VISUALIZATION

Use the embedded viewer for triage; use IGV Desktop to prove the split chain

SeqOne's public Spring 2026 release describes both a purpose-built browser for large variants and upgraded IGV views. The public product name is **SeqOne**; 'Seq1' may be a local shorthand, but it is not the name used in the materials reviewed here.

Publicly documented capabilities

SeqOne capability	How to use it for this question	What it establishes
Breakpoint navigation: start, end, full span	Inspect both breakpoints and the complete event span; do not rely on one breakpoint pane.	Whether visible read/coverage morphology is consistent at both event boundaries.
Soft-clipped reads displayed by default for SVs	Confirm clipped sequence at each breakpoint and capture a screenshot with the event selected.	That clipping is visible in the SeqOne SV context, not that the clipped bases lack another alignment.
BAM tracks with nucleotide display	Zoom to base level for mismatch/indel context after the event-scale review.	Local base-level evidence rendered from the selected alignment track.
Family-member or other-sample tracks	Display proband and relatives/controls together when inheritance or recurrence matters.	Cross-sample visual support; track settings still need to be recorded per sample.
Quick-copy coordinates with 1 kbp padding	Copy both breakpoint windows and open the exact CRAM/BAM in IGV Desktop.	A reproducible handoff window for deeper SA/supplementary inspection.

Important limitation. SeqOne's public guidance does not state whether its embedded view includes SAM supplementary records (0x800), follows SA-tag links, filters secondary records, or de-duplicates multiple segments from one QNAME. A visible soft clip is therefore not proof that the rest of the molecule is unaligned.

Recommended SeqOne -> IGV review path

Gate	Action	Stop condition
1. Event triage	Use SeqOne start/end/full-span buttons; inspect soft clips and family tracks.	Both breakpoints reviewed; deployment/workset version recorded.
2. Coordinate handoff	Copy each breakpoint with 1 kbp padding and preserve the selected sample/BAM identity.	Exact two windows and evidence file identity captured.
3. Split-chain proof	Open the same BAM/CRAM in current IGV Desktop; supplementary filter off, secondary filter on, soft clips on.	Supplementary Reads Diagram or split-screen view reconciles SA-linked segments.
4. Escalation	If SeqOne and IGV disagree, inspect SAM flags/CIGAR/SA tags for the exact QNAME.	Difference explained as filtering, downsampling, tag absence, or rendering behavior.

Questions for SeqOne support

- Does the embedded IGV track retain or filter 0x800 supplementary records and 0x100 secondary records?
- Does it follow SA tags to linked loci, and is there a read-order/supplementary-chain view?
- Are coverage and pileup tracks primary-only, primary-plus-supplementary, or precomputed by another service?
- What downsampling, MAPQ, duplicate, and QC-fail filters apply to long-read BAM/CRAM tracks?
- Can it export the exact QNAME, flags, CIGAR, MAPQ, and SA tag used for an interpretation screenshot?

METHODS, PROVENANCE, AND LIMITATIONS

Bounded existing-data analysis

No ONT alignment, variant calling, packaging, or release workflow was rerun. The analysis only reread the existing full-coverage CRAM four times with mosdepth 0.3.14 under explicit record masks and CIGAR modes. All four commands completed with exit status 0.

Frozen input

Item	Identity
Analysis unit	HG002-ck2ky6p4wk6exh; Bjuice-prevalidation Take101
LR CRAM	HG002-ck2ky6p4wk6exh.sentdhiomr2.fastq.cram
S3 size / ETag	28,187,720,428 bytes; b5db8bd255d86d5c03caa4f27e4458ce-673
Reference	GRCh38_no_alt_analysis_set.fasta
Core BED	hg38_broad_male_diploid_no_alt.bed
Mosdepth denominator	2,889,585,716 core-region reference bases
Common options	--threads 8 --by --use-median -n --mapq 0 -f -T 0,10,20,30

Flag-mask accounting

Decimal / hex	Excluded SAM flags	Resulting record set
1796 / 0x704	UNMAP 0x4; SECONDARY 0x100; QCFAIL 0x200; DUP 0x400	Primary + supplementary mapped records
3844 / 0xF04	Mask 1796 plus SUPPLEMENTARY 0x800	Primary mapped records only

Reproducibility receipts

Comparison script SHA-256: `4e3fca6c77c9db7f715b165fc38e15fe5b7b3f19205e656c36606e05f69c63a3`. Each summary SHA-256 is retained in the companion TSV committed beside this appendix. Observed wall times were 1:14-1:21 per profile; runtime is reported only as an execution receipt, not a benchmark.

Limitations

- This is one HG002 full-coverage Bjuice-prevalidation CRAM; the numerical deltas are not assumed to generalize to every sample, chemistry, aligner, or reference.
- Reference-depth addition from supplementary segments does not de-duplicate by QNAME or overlapping reference interval.
- The experiment measures alignment evidence available in the CRAM, not which bases an opaque caller ultimately uses.
- A vendor response or auditable per-read caller trace is still required to claim Sentieon algorithm-effective coverage.

Primary references

1. mosdepth official README - fast mode and flag mask.
2. SAM format specification - alignment flag definitions.
3. Sentieon software manual - public algorithm context.
4. DNAscope LongRead app note - long-read workflow.
5. Sentieon model registry - hybrid bundle identity.
6. IGV chimeric-read guide - SA-linked views and navigation.
7. IGV alignment basics - Third Gen, MAPQ, downsampling.
8. SeqOne Spring 2026 release - viewer capabilities.
9. SeqOne instructions for use - local IGV handoff.