VALIDATION SUMMARY – RNA fusion engine v1.0

GENOVESA I Bioxsys s.r.o.

Field	Value
Document	GP-VAL-RNA-2026-001-S, version 1.0
Engine	Genovesa RNA fusion engine (fusions, exon skipping), version 1.0
Applies to	all GENOVESA workflows built on this engine
Reference workflow	Genovesa RNA FusionPlus, version 1
Valid from	29 October 2026
Based on	internal validation protocol GP-VAL-RNA-2026-001 v1.0 (available for audit)

1. Scope

Bioinformatic processing of RNA-seq data (FASTQ, paired-end) from targeted hybridisation capture:

- **gene fusions** (translocations, inversions, deletions, duplications) with HGVS (r.) and VICC nomenclature, reading-frame assessment and confirmation by a second independent algorithm,
- **exon skipping** for defined events: MET ex14, EGFRvIII/vIIIb/vII/vIIb/vIVa/vIVb, ERBB2 Δ16.

Out of scope: SNV/indel variants from RNA and RNA QC (separate engine, validation in progress), cryptic exons (e.g. AR-V7), expression quantification, fusions without chimeric reads in the target regions, clinical interpretation. The validation covers the bioinformatic part; pre-analytics, library preparation and sequencing are verified by the laboratory.

2. Components and versions

Component	Version
Engine	RNA fusion 1.0
Reference genome and annotation	NCBI RefSeq GRCh38.p14, annotation RS_2025_08; HGNC gene symbols
Third-party tools (open source)	fastp 0.20.1, STAR 2.7.11b, Arriba 2.5.1 (primary fusion detection), STAR-Fusion 1.9.0 (confirmation), samtools 1.21
Bioxsys components	RefSeq reference data for fusion detection, detection configuration and filtering, fusion annotation 1.7.2 (HGVS r., VICC, exon numbers), exon-skipping detection, caller agreement and confidence, TSV / XLSX / PDF outputs

Third-party tools form only the detection core. The reference data, detection configuration and all downstream processing were developed by Bioxsys s.r.o.; their details are Bioxsys know-how and are not part of this document.

3. Test set

Type	Count	Source of truth
Clinical samples with a fusion (EQA)	3	EMQN external quality assessment
Clinical samples with MET ex14 skipping	6	laboratory PCR

Type	Count	Source of truth
Clinical samples without fusion and MET ex14	10 (+ 3 EQA without MET ex14)	laboratory PCR
Public STAR-Fusion benchmark (Haas et al., Genome Biology 2019), 2 × 50 bp	5 samples, 2,500 fusions	simulation with known truth and expression
Public STAR-Fusion benchmark, 2 × 101 bp	5 samples, 2,500 fusions	simulation with known truth and expression

4. Performance

Parameter	Result	95% CI	Criterion	Met
Sensitivity – fusions (EQA)	3/3 (100%)	29.2–100%	100%	YES
Sensitivity – MET ex14	6/6 (100%)	54.1–100%	100%	YES
Specificity – MET ex14	13/13 (100%)	75.3–100%	100%	YES
Sensitivity – benchmark 50 bp, fusions ≥ 10 TPM	1,389/1,528 (90.9%)	89.3–92.3%	≥ 90%	YES
Sensitivity – benchmark 101 bp, fusions ≥ 10 TPM	1,449/1,573 (92.1%)	90.7–93.4%	≥ 90%	YES
PPV – benchmark 50 bp, confidence high [1]	99.3%	98.8–99.7%	≥ 90%	YES
PPV – benchmark 101 bp, confidence high [1]	99.5%	99.1–99.8%	≥ 90%	YES
Reproducibility (different server and OS)	identical fusion results	–	identical	YES
Nomenclature of verified fusions	3/3 HGVS + VICC	–	filled in	YES

Findings with confidence high and medium are evaluated (low is informative). The confidence intervals for clinical samples are wide because of the small number of samples (3/3 cannot statistically show more); the precision of the performance estimate is shown by the benchmark with 1,528 and 1,573 fusions. [1] False positive = gene pair unrelated to any simulated fusion. If a correct breakpoint reported with a paralog symbol is also counted as false (see 6.3), PPV is 81.6% (50 bp) and 82.1% (101 bp).

Benchmark sensitivity depends on fusion expression: 5–10 TPM 83% (50 bp) and 93% (101 bp); below 5 TPM 31% and 65%. Fusions below 5 TPM are below the reliable detection limit of RNA-seq in general, not only of this engine.

5. Validated range

Workflows on this engine may change settings only within this range; otherwise a verification (VER) is required.

Area	Validated
Reported fusions	confidence high and medium; sensitivity demonstrated for fusions with expression ≥ 10 TPM
Exon skipping	thresholds high ≥ 20 reads and ≥ 10%, medium ≥ 5 reads and ≥ 5% (fixed in the engine)
Read length	50–160 bp, paired-end (benchmark 50 and 101 bp, clinical samples up to 160 bp)
Reads per sample	see QC limits below; limit of detection not established (see 6.1)

QC limits for input data. Derived from 19 clinical validation samples (lowest value with a correct result). The values of each examination are in the QC report (MultiQC) of the workflow output.

Metric	Range in validation	Limit
Read pairs (before filtering)	2.7–19.3 million	≥ 2.7 million
Reads passing quality filtering	91.5–100%	≥ 90%
Bases ≥ Q30 (after filtering)	82.8–98.2%	≥ 80%
Uniquely mapped reads	80.9–93.1%	≥ 80%

Metric	Range in validation	Limit
Read length (mean after filtering)	118–147 bp	50–160 bp

A sample that fails a limit can be processed, but **a negative result is not validated** (only positive findings may be reported, with a recommendation for confirmation). Quality and amount of input RNA (e.g. DV200, RIN) and library preparation are set by the laboratory according to the kit manufacturer.

6. Limitations

1. **Limit of detection (LoD) has not been established.** Weakest verified finding: MET ex14 with 106 reads (33%). Fusions with low expression or low read support may be missed; a negative result in a sample with poor RNA quality must be interpreted with caution.
2. **EGFRvIII/vII/vIV:** detection has not been verified on a positive control.
3. **Paralogs:** for fusions with a partner from a multi-copy gene family (e.g. NUTM2*, TBC1D3*, CGB*, SPANX*), a paralog may be reported instead of the actual gene; interpret the partner name as the gene family.
4. **Findings outside the targets of the orthogonal method:** a high-confidence fusion that was not covered by the laboratory's confirmation test should be confirmed by another method before the result is released.
5. Performance was verified on one type of targeted panel and on simulated whole-transcriptome data. **The laboratory verifies performance with its own kit, sequencer and QC limits on site** (user verification); Bioxsys provides QC metrics from the first runs.

7. Approval

Role	Name	Date
Prepared by	Ing. Petr Brož, Ing. Filip Zembol, Bioxsys s.r.o.	29 October 2026
Approved by	Ing. Petr Brož, Bioxsys s.r.o.	29 October 2026