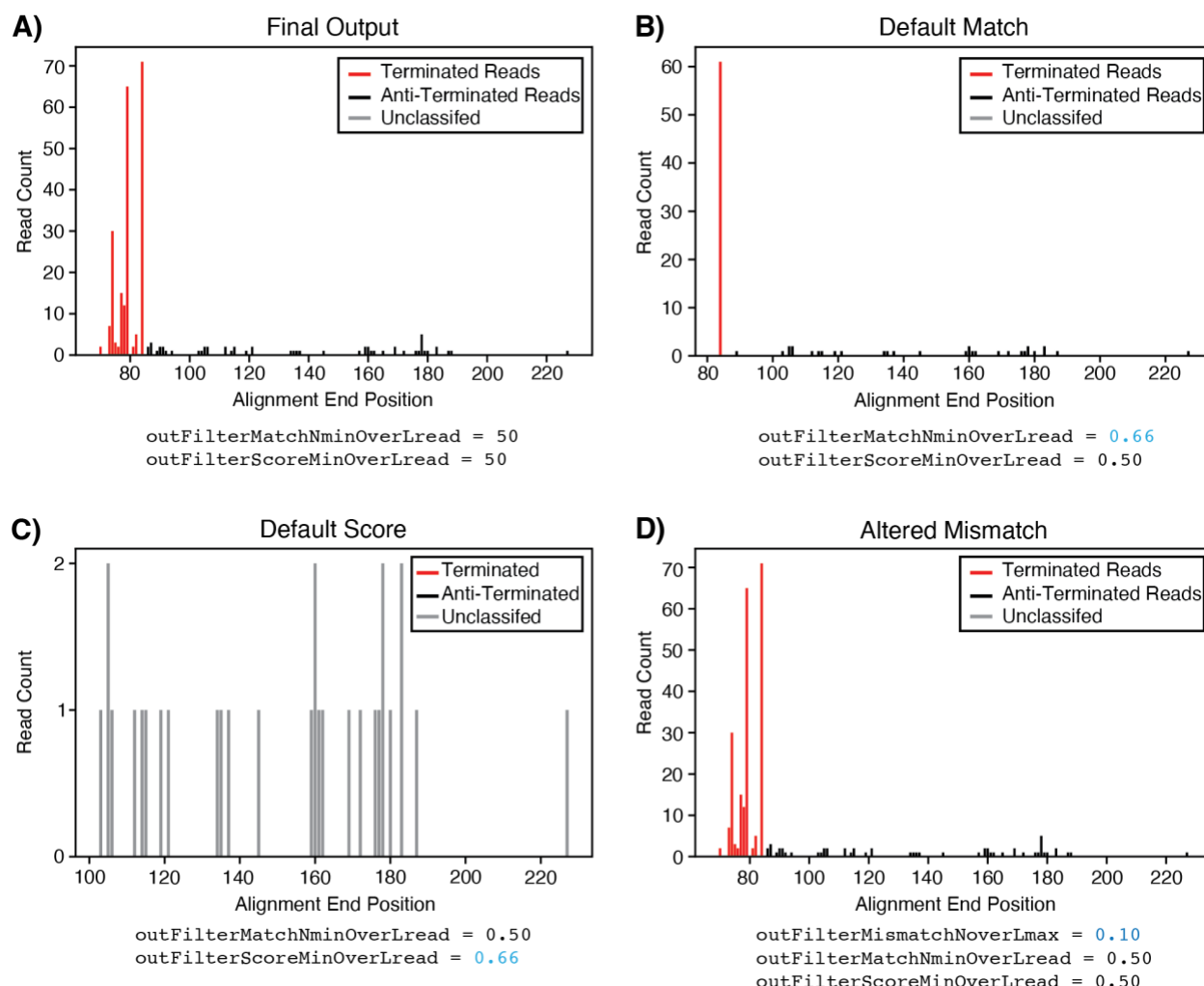




STAR parameter adjustments. Using *Bacillus cereus* (CP000227.1/4763720-4763779) as a positive control, we altered several STAR parameters. We adjusted the STAR parameters since the reference genes are 235 nucleotides long, while the terminated *B. ce* reads are around 80 nucleotides. Without proper parameter adjustment, the shorter *B. ce* reads were being filtered out during alignment. A) STAR parameters used to align reads to reference genes. B) The same parameters as (A), but with the outFilterMatchNminOverLread set to its default value of 0.66. Reads where the ratio of matched bases to reference bases is below this threshold are discarded. For instance, a 100 nt read with 70 matching bases is retained, but a read with only 60 matching bases is discarded. Without lowering this value, terminated reads were filtered out. C) The same parameters as (A), but with the outFilterScoreMinOverLread set to its default value of 0.66. Reads with an alignment score normalized to the read length below this ratio are discarded. For example, a 100 nt read with an alignment score of 0.68 is retained, while a 100 nt read with a score of 0.54 is discarded. Without lowering this value, terminated reads were filtered out. D) Addition of the outFilterMismatchNoverLmax parameter lowered from the default of 0.3 to 0.1. This filter

discards reads where the ratio of mismatches to read length exceeds 0.3. For example, a 100 nt read with 32 mismatches would be discarded, while a 100 nt read with 28 mismatches would be retained. This filter was applied to account for potential overlap in aptamer sequences between fluoride riboswitch variants. Default parameters and detailed explanations are found in the STAR manual at: https://physiology.med.cornell.edu/faculty/skrabanek/lab/angsd/lecture_notes/STARmanual.pdf. This figure is adapted from Hertz, et. al. (2025) *bioRxiv*.

Further explanation of STAR parameters. To generate an encompassing response, the following prompt was given to ChatGPT:

Please explain the following:

The reads were aligned to their accession number and genome positions using STAR (v2.7.9a) (PMID: 23104886). STAR was run with the following specified parameters:

```
STAR --runMode alignReads \
--alignIntronMax 1 \
--genomDir /path/ \
--readFilesIn Folder/path/.fastq \
--outFileNamePrefix Folder/name_ \
--alignEndsType Local \
--alignSoftClipAtReferenceEnds Yes \
--outFilterScoreMinOverLread 0.5 \
--outFilterMatchNminOverLread 0.5 \
--outSAMmultNmax 1
```

The following output was human edited:

STAR (Spliced Transcripts Alignment to a Reference): STAR is a software tool used for aligning RNA-seq reads to a reference genome. It is specifically designed for speed and accuracy.

--runMode alignReads: Specifies the mode of operation for STAR. In this case, it is aligning reads to a reference genome.

--alignIntronMax 1: Sets the maximum intron size for spliced alignments. In this case, introns longer than 1 base will not be considered during the alignment process. We set this as the parameter as we expect no introns in bacteria and any intron results would be artificial results.

--genomDir /path/: Specifies the path to the directory containing the reference genome index for alignment.

--readFilesIn Folder/path/.fastq: Specifies the input RNA-seq data files

--outFileNamePrefix Folder/name_: Sets the prefix for all output files

--alignEndsType Local: This parameter specifies how STAR will handle the alignment of the read ends. With Local, STAR will perform local alignment, which allows for the alignment of fragments that may not align perfectly across the entire length (e.g., when the read spans an exon-exon junction). We used this parameter to compensate for terminated reads.

--alignSoftClipAtReferenceEnds Yes: If this is set to "Yes", STAR will allow soft-clipping (trimming of bases at the ends of the read) at the reference genome ends. This is useful for aligning reads where part of the read may be unaligned at the ends, often seen in RNA-seq due to partial or incomplete transcript coverage. We used this parameter to compensate for terminated reads.

--outFilterScoreMinOverLread 0.5: This filter ensures that only reads with a score (based on the alignment quality) of at least 50% of the read length are kept. Essentially, it keeps alignments where the quality of the match is sufficiently good. The default value is 0.66. We set this parameter to prevent read filtering of terminated reads. See Figure at top this document.

--outFilterMatchNminOverLread 0.5: This sets the minimum fraction of matched bases over the read length to 50%. In other words, the read must have at least 50% of its bases matching the reference genome to pass the filter. The default value is 0.66. We set this parameter to prevent read filtering of terminated reads. See Figure at top this document.

--outSAMmultNmax 1: This parameter limits the number of alignments allowed for each read. If the read has more than one valid alignment, STAR will only keep the first one. This can reduce the number of ambiguous or multi-mapped reads in the output. We set this parameter to avoid reads being assigned to multiple variants given that there is high sequence similarity between some variants.