

MixtureAceMT Technical Note

Strand Direction Odds Ratios

March 2026

1. Background

The strand direction odds ratio (aka strand odds ratio, SOR) is a metric used to identify strand bias, a type of systematic error where sequencing reads supporting a specific variant map preferentially to either the forward or reverse DNA strand. Extreme strand bias can indicate unreliable variant calls, and possibly false-positive variant calls. SOR is an alternative to Fisher's exact test that is more accurate at high read coverage and less sensitive to coverage asymmetry. Either of these conditions can drive Fisher's exact test to identify statistically significant differences even if the differences are not practically meaningful in variant calling. Like Fisher's exact test, SOR is based on the analysis of a contingency table of read counts for reference (*ref*) and alternate (*alt*) alleles in the forward (*Fw*) and reverse (*Rv*) direction (Figure 1).

Throughout this monograph, the terms allele and base are used interchangeably. Reference base refers to the base at a position in the revised Cambridge reference sequence (rCRS). The alternate base refers to a non-reference base. When the alternate base is clearly the major base at a position it constitutes a variant (relative to rCRS) potentially reportable to CODIS or EMPOP. Significant counts of both the reference base and an alternate base may be present at a position indicating either a point heteroplasmy, or a mixed sample. The SOR metric is used to help filter out non reference allele noise that could be misinterpreted as a variant.

	Forward Strand	Reverse Strand
Ref Allele	refFw	refRv
Alt Allele	altFw	altRv

Figure 1. Illustration of contingency table used in SOR calculations. The values entered in the table cells are read counts.

2. Calculation of Strand Odds Ratio (SOR)

MixtureAceMT reports a SOR metric that parallels the SOR metric reported by GATK *HaplotypeCaller* or *VariantAnnotator* ([StrandOddsRatio – GATK](#)). Calculations relating to the SOR and component metrics that comprise the SOR are described in this section.

2.1. Strand Ratios

Strand balance underlies the SOR calculations and is quantified by the strand ratio metric which refers to the forward to reverse reads for reference and alternative bases (i.e., **refFw**, **altRv**). The smaller of the two values is placed in the numerator and the larger in the denominator.

$$refRatio = \frac{\min (refRv, refFw)}{\max (refRv, refFw)}$$

$$altRatio = \frac{\min (altRv, altFw)}{\max (altRv, altFw)}$$

2.2. Odds Ratios (OR)

The odds ratio is a concept applied in different fields in different ways. The group-wise odds ratio is the version applied here and describes the ratio of the odds of an event occurring in one group to the odds of it occurring in another group. In the present case the groups are forward and reverse reads, and the events are reference and alternate bases at a position. The odds ratio is a ratio of ratios that can be algebraically reduced to a single ratio as shown in the equation below.

$$Odds\ Ratio = \frac{ref\ ratio}{alt\ ratio} = \frac{refFw / refRv}{altFw / altRv} = \frac{refFw \times altRv}{refRv \times altFw}$$

An odds ratio of 1 indicates that the condition or event under study is equally likely to occur in both groups (i.e., the groups and events are uncorrelated). An odds ratio greater than 1 indicates that the condition or event is more likely to occur in the first group. An odds ratio less than 1 indicates that the condition or event is less likely to occur in the first group (https://en.wikipedia.org/wiki/Odds_ratio).

2.3. Symmetrical Odds Ratios

The symmetrical odds ratio is designed to account for strand imbalance in either the reference or alternate base.

$$\text{symmetrical OR} = \left(OR + \frac{1}{OR}\right)$$

2.4. Strand Odds Ratio (SOR)

MixtureAce expresses the SOR on a log scale and with a normalization adjustment. The normalization adjustment addresses the situation where total forward coverage \neq total reverse coverage. High coverage asymmetry can inflate bias and indicate a falsely high SOR simply because one strand has more reads overall.

$$SOR = \ln(\text{symmetrical ratio}) + \ln(\text{refRatio}) - \ln(\text{altRatio})$$

3. Implementation of SOR in MixtureAceMT

The magnitude of SOR metrics provide support (or not) to variant base calls and is a component of noise filtering in MixtureAceMT. Generalized interpretations of SOR values are shown in **Table 1**. The SOR is used as a filter threshold combined with minimum read depth thresholds (**Figure 2**). The read depth threshold(s) (**Figure 2**) protect against stochastic SOR values that can occur at low read coverage and can be the same as the overall analytical threshold.

Table 1. Interpretation of SOR values.

SOR Value	Interpretation
<1	No strand bias
1-3	Mild bias
>3	Bias

FAIL if: SOR > 3.5 AND (altFw < 10 OR altRv < 10)

Figure 2. Implementation of SOR as a filter in MixtureAceMT

4. Numerical Example

Raw Counts

$$\text{refFw} = 1450$$

$$\text{refRv} = 345$$

$$\text{altFw} = 160$$

$$\text{altRv} = 212$$

One is routinely added to each read count to avoid possible division by zero.

Add Pseudo Count of 1:

$$\text{refFw} = 1450 + 1 = 1451$$

$$\text{refRv} = 345 + 1 = 346$$

$$\text{altFw} = 160 + 1 = 161$$

$$\text{altRv} = 212 + 1 = 213$$

$$\text{refRatio} = \frac{\min(\text{refRv}, \text{refFw})}{\max(\text{refRv}, \text{refFw})} = \frac{346}{1451} = 0.2384$$

$$\text{altRatio} = \frac{\min(\text{altRv}, \text{altFw})}{\max(\text{altRv}, \text{altFw})} = \frac{161}{213} = 0.7559$$

$$\text{Odds Ratio} = \frac{\text{ref ratio}}{\text{alt ratio}} = \frac{\text{refFw}/\text{refRv}}{\text{altFw}/\text{altRv}} = \frac{1451 \times 213}{161 \times 346} = 5.548$$

$$\text{Symmetrical Odds Ratio}_{\text{raw}} = \text{OR} + \frac{1}{\text{OR}} = 5.548 + 0.1802 = 5.7282$$

$$\begin{aligned} \text{SOR}_{\text{normalized}} &= \ln(\text{SOR}_{\text{raw}}) + \ln(\text{refRatio}) - \ln(\text{altRatio}) \\ &= \ln(5.7282) + \ln(0.2384) - \ln(0.7559) = 0.5914 \end{aligned}$$