Inappropriate data filtering of **LABEL modules** for IRMA

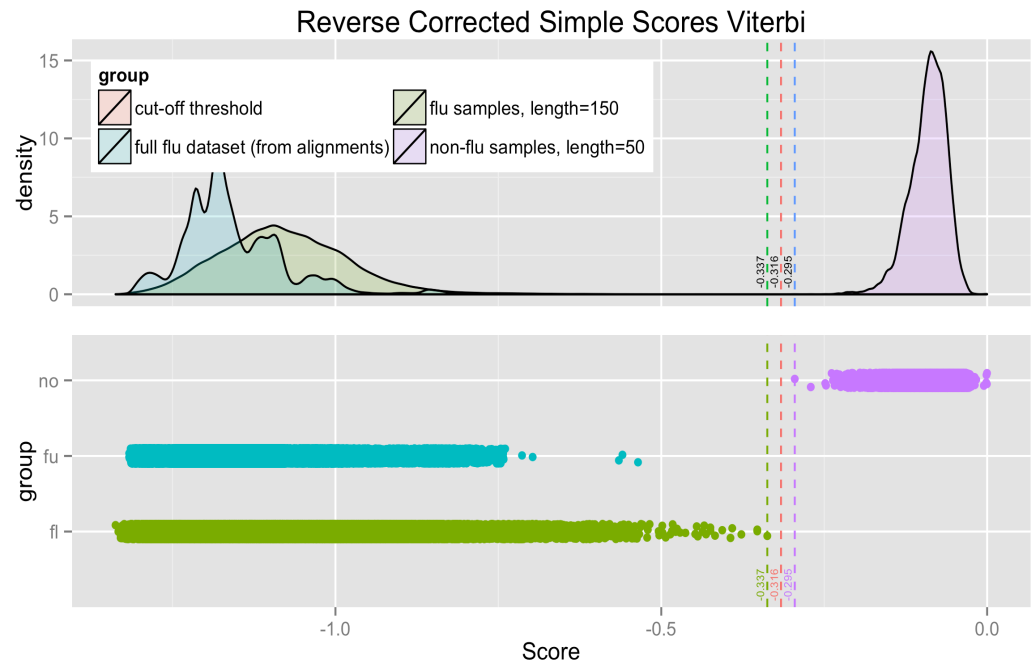# ADDITIONAL FILE 2

LABEL module
irma-FLU

Classifies influenza A and B gene segments as well as 16 HA subtypes and 9 NA subtypes.

Density and jitter-plots show separation of full influenza gene sequences and subsequences versus non-flu subsequence control. Threshold used to separate non-flu-like reads.

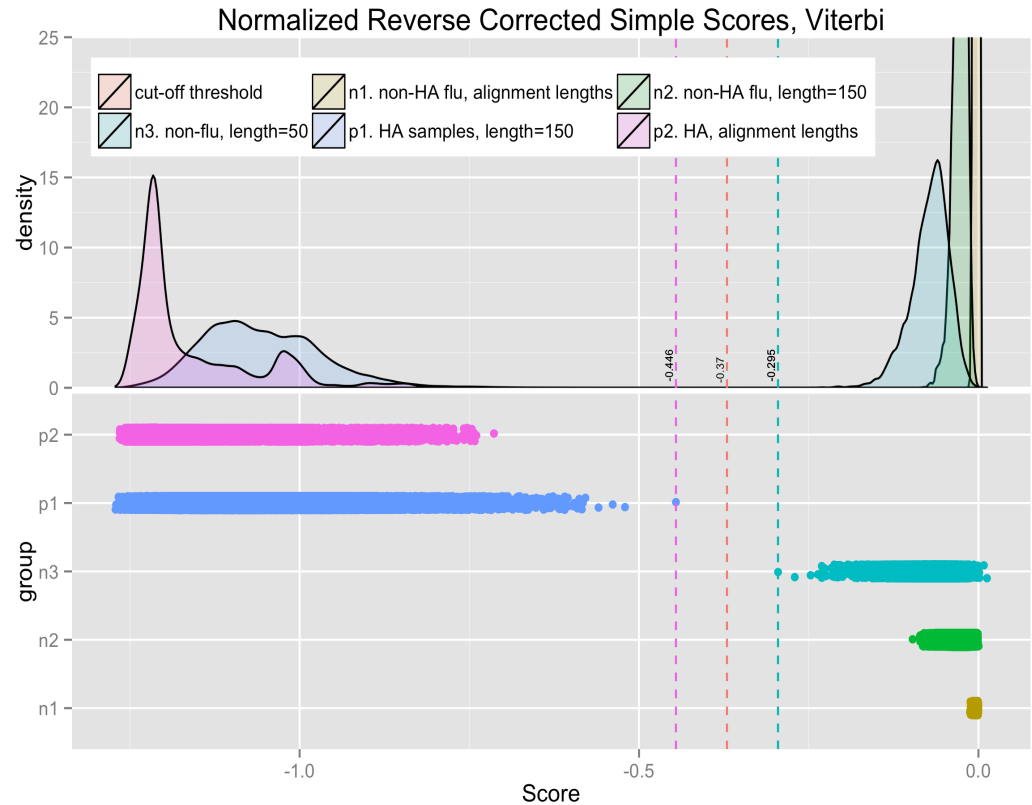Not used by default.

See Figure 3 and Supplementary Figure S1

Reverse Corrected Simple Scores Viterbi

group
- cut-off threshold
- full flu dataset (from alignments)
- flu samples, length=150
- non-flu samples, length=50

LABEL module
irma-FLU-HA

Classifies influenza A HA subtypes (1-16) and influenza B HA, given an HA.

Density and jitter-plots show separation of full influenza HA sequences and subsequences versus non-HA subsequence control. Threshold used to separate non-HA-like reads.

See Figure 3 (LABEL trio).

Normalized Reverse Corrected Simple Scores, Viterbi

Legend:
- cut-off threshold
- n1. non-HA flu, alignment lengths
- n2. non-HA flu, length=150
- n3. non-flu, length=50
- p1. HA samples, length=150
- p2. HA, alignment lengths

Threshold lines: -0.446, -0.37, -0.295

LABEL module
irma-FLU-NA

Classifies influenza A NA subtypes (1-9) and influenza B NA, given an NA.

Density and jitter-plots show separation of full influenza NA sequences and subsequences versus non-NA subsequence control. Threshold used to separate non-NA-like reads.
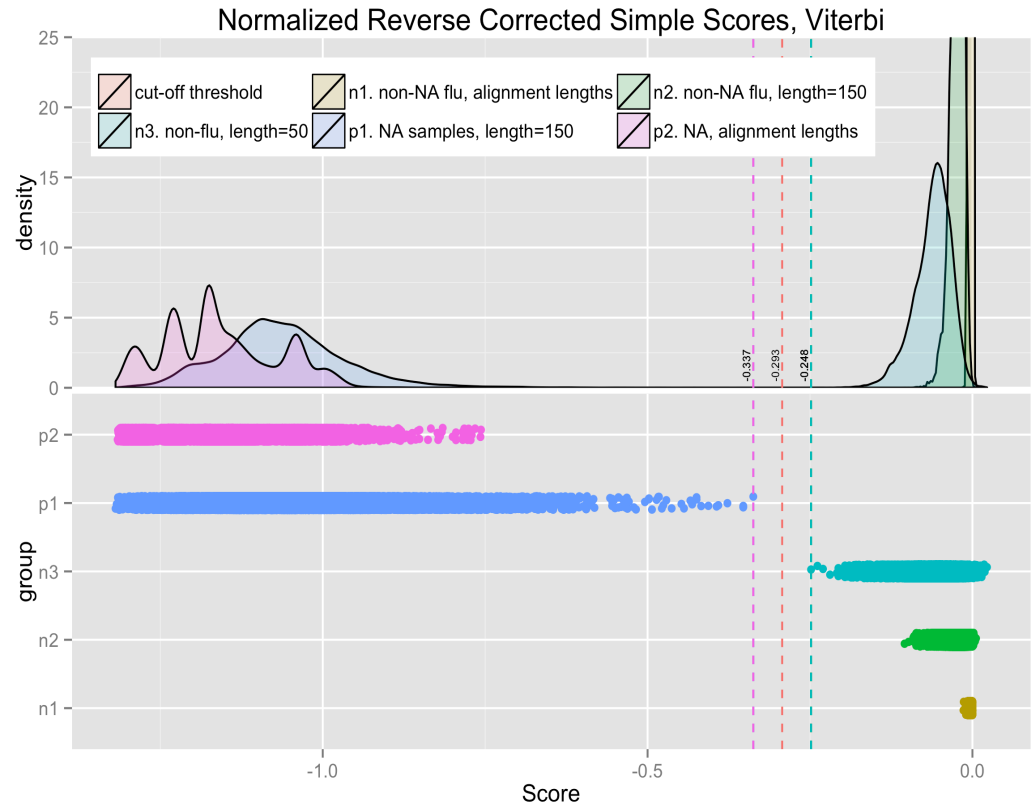
See Figure 3 (LABEL trio).

Normalized Reverse Corrected Simple Scores, Viterbi

LABEL module
irma-FLU-OG

Classifies internal (PB2, PB1, PA, NP, MP, NS) gene segments for influenza A and B. Termed "other gene" or OG for short.

Density and jitter-plots show separation of full influenza OG sequences and subsequences versus non-OG (HA/NA) subsequence control. Threshold used to filter non-internal-gene-like reads.

See Figure 3 (LABEL trio).

Normalized Reverse Corrected Simple Scores, Viterbi

Legend:
- cut-off threshold
- n1. HA/NA flu, alignment lengths
- n2. HA/NA flu, length=150
- n3. non-flu, length=50
- p1. int. gene samples, length=150
- p2. int. genes, alignment lengths