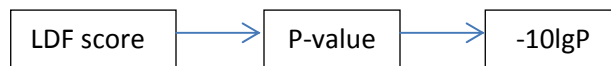


# PEAKS Score (-10lgP)

Bioinformatics Solutions Inc.

July 4, 2011

This document introduces PEAKS DB's -10lgP score. The score calculation in PEAKS DB can be illustrated in the following figure.

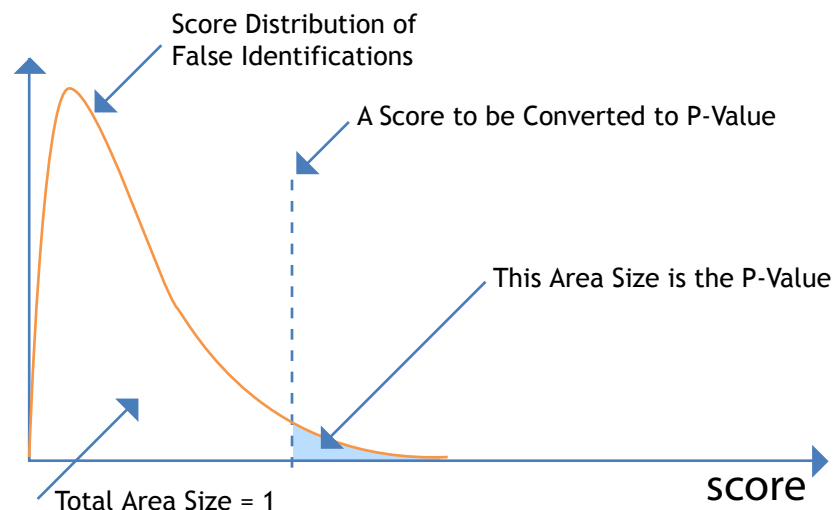


**First**, PEAKS DB internally uses an LDF (linear discriminative function) score to measure the quality of the peptide-spectrum match. The LDF score uses not only the matching between the fragment ions and the peaks in the spectrum, but also many other factors such as the similarity between the *de novo* sequencing peptide and the database peptide. This LDF score tries to achieve two goals: (1) for each spectrum in the MS/MS dataset, finds the most likely correct peptide from the database; (2) for the whole dataset, separates the true and false identifications as much as possible.

**Then**, the LDF score is converted to a P-value for better human interpretation.

**P-value:** for a given score  $x$ , its equivalent P-value is the probability that a false identification has score  $\geq x$

The smaller the P-value is, the less likely the peptide-spectrum match is due to a random match. The meaning of the P-value is better explained in the following figure.



Note that although many software packages use the same term “P-value”, their meanings can be different. Another popular definition of P-value is “the probability as that a random peptide matches the current spectrum with score  $\geq x$ ”. However, in the database search, a false identification is the result of many random peptides in the database, instead of just one random peptide. Thus, the definition in PEAKS DB is more useful for the controlling of result quality with P-value.

**Lastly**, the P-value is converted to  $-10 \cdot \log_{10}(\text{P-value})$  to make it more “human-friendly”. In PEAKS, this value is denoted by  $-10\lg P$  as  $\lg$  is the ISO reserved notation for  $\log_{10}$ . By this conversion, a more significant match will have a higher  $-10\lg P$  value. Additionally, a P-value of 1% is equivalent to  $-10\lg P$  of 20.

The following figure is a screenshot from a PEAKS DB’s search result. The x-axis is the  $-10\lg P$  score, and the y-axis is the number of peptide-spectrum matches with that score. Normally, a greater than 20 score is of relatively high confidence (as illustrated by many target but very few decoy matches above that threshold). For a large dataset, it is recommended to use the FDR (false discovery rate) to select the right  $-10\lg P$  score threshold (which is easy in PEAKS). However, when the dataset is small ( $\#$  spectra  $< 100$  or protein database contains only a few proteins), 20 is a good threshold for  $-10\lg P$ .

