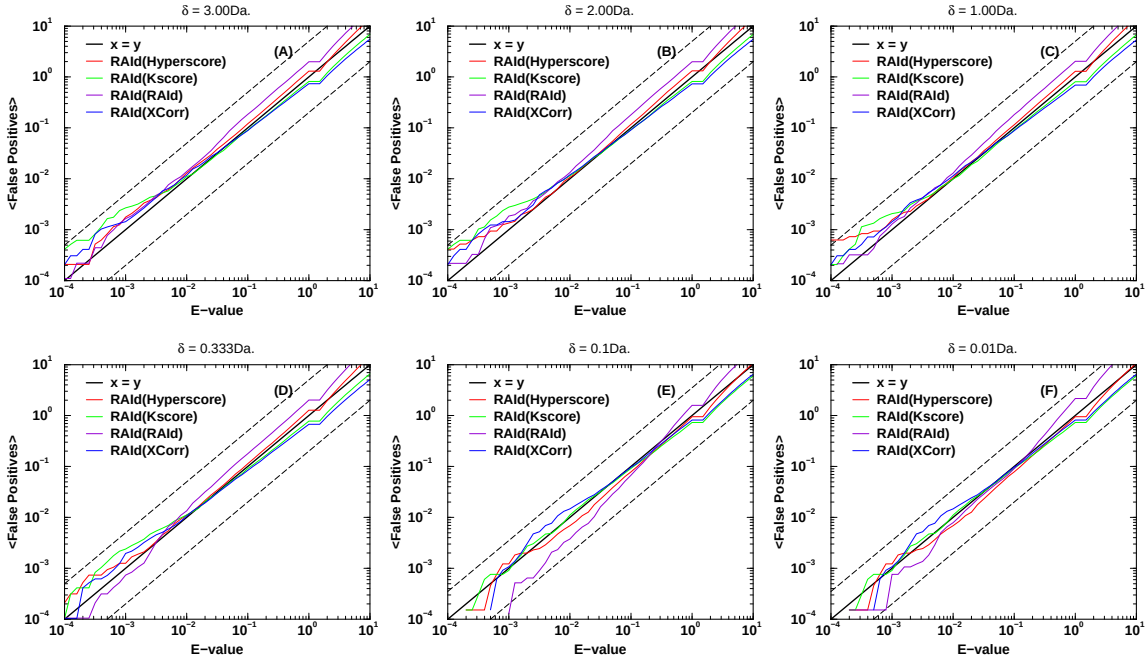

Figure S5. E-value accuracy assessment. The agreement between the reported E-value and the textbook definition is examined using profile data (the NHLBI data set: $10,000$ spectra). The random database size used is 500 MB. The molecular weight range considered while searching the database is $[MW - \delta, MW + \delta]$. In each panel, the dashed lines, corresponding to $x = 5y$ and $x = y/5$, are used to provide a visual guide regarding how close/off the experimental curves are from the theoretical curve.