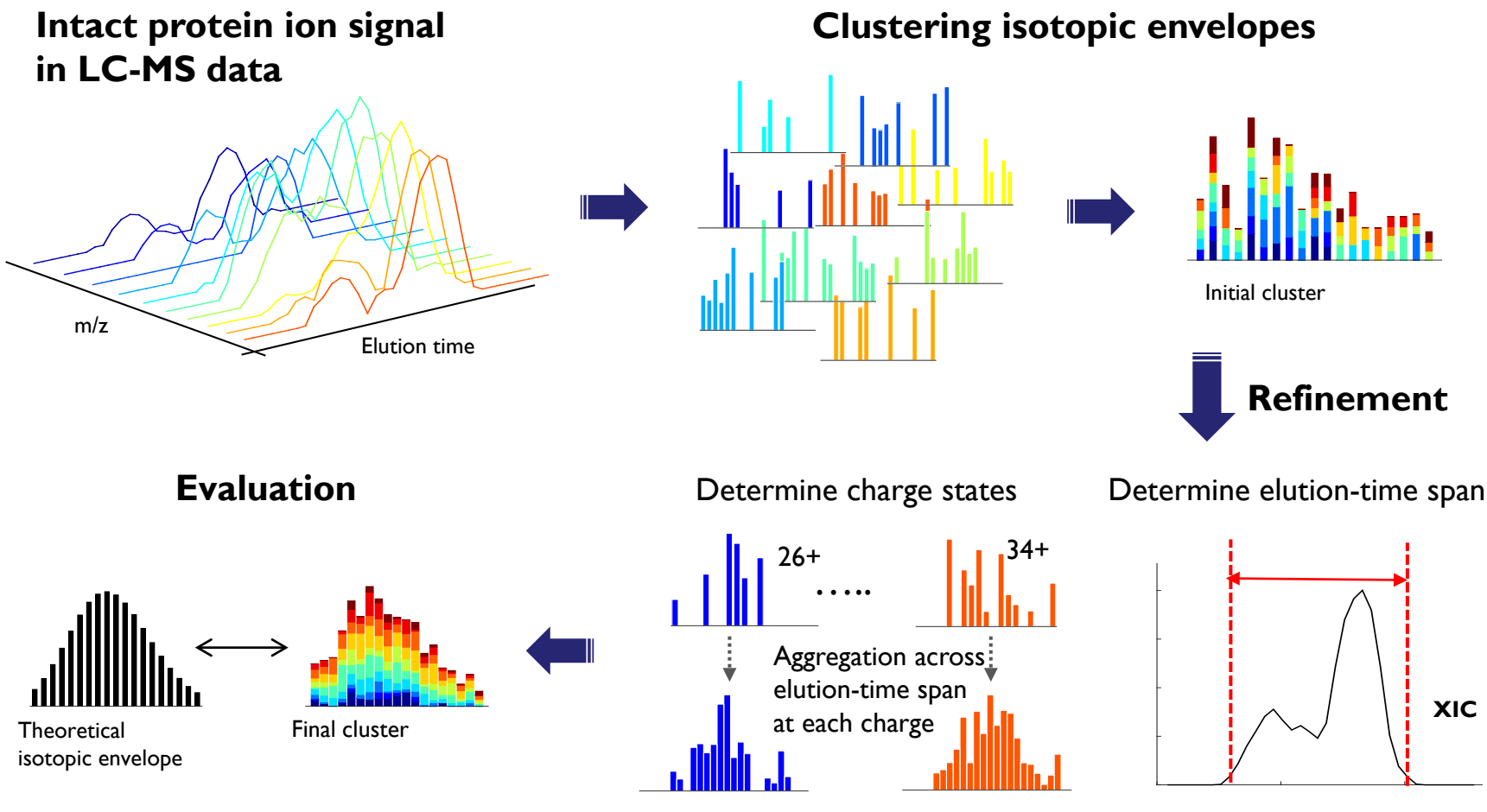
ProMex Scoring

Manuscript:

"Informed-Proteomics: open-source software package for top-down proteomics" Nature Methods (2017) doi:10.1038/nmeth.4388

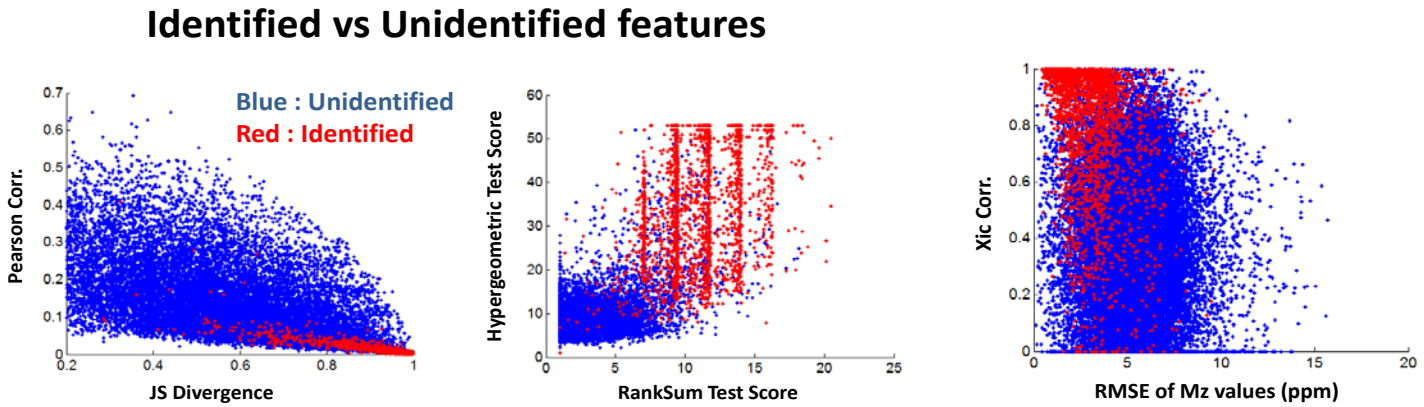
https://www.ncbi.nlm.nih.gov/pubmed/28783154

Clustering isotopic envelopes



Scoring LC-MS Features

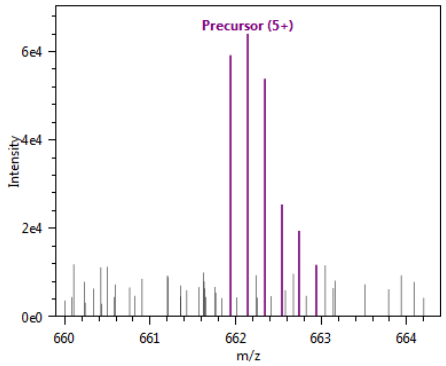
- **Various scoring metrics**



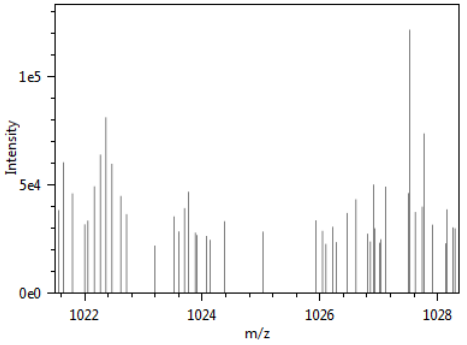
- **However, still too many false features!**
  - **Errors in the subsequent identification and quantification analysis**
  - **Slow down identification search engine**

Likelihood Ratio Score

It seems a true feature.
How confident are we?



Let's shuffle the peaks and see
how frequently such patterns occur?

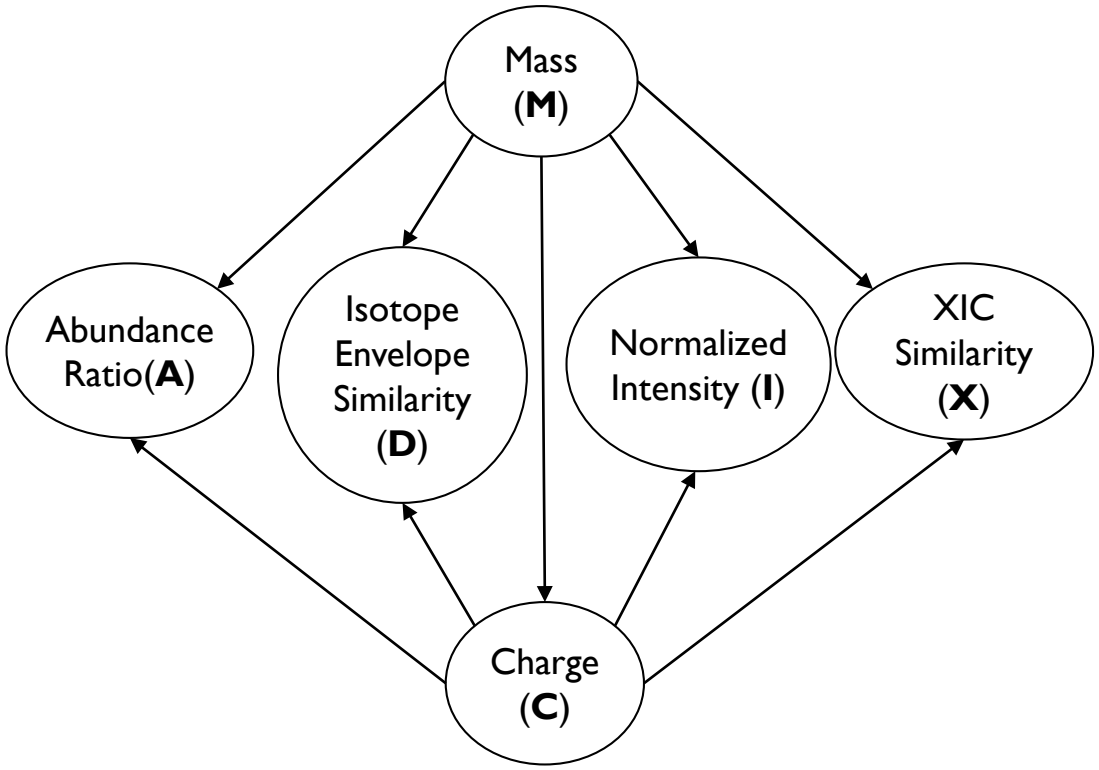


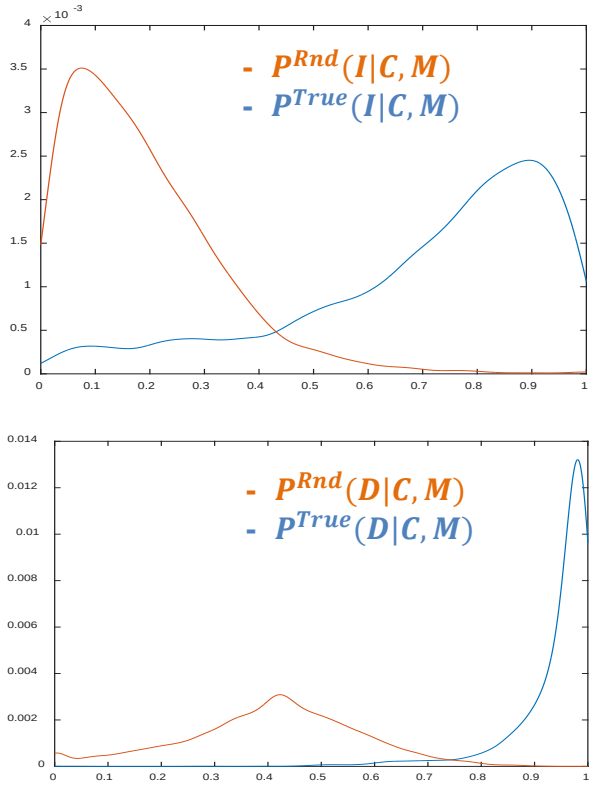
$$Score(M,S) = \log \frac{P^{True}(S|M)}{P^{Random}(S|M)}$$

M: Protein Mass

S: Spectra

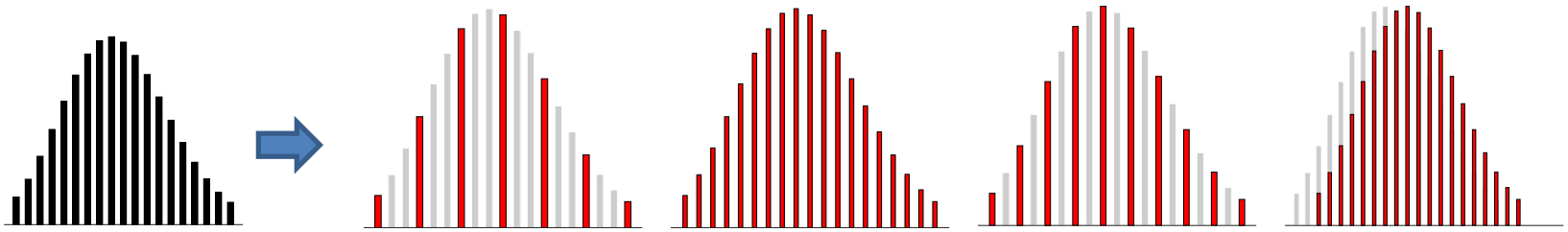
Scoring in ProMex





$$Score(M,S) = \sum_{C=MinCharge}^{MaxCharge} log \frac{P^{True}(C,A,I,D,X|M)}{P^{Random}(C,A,I,D,X|M)}$$

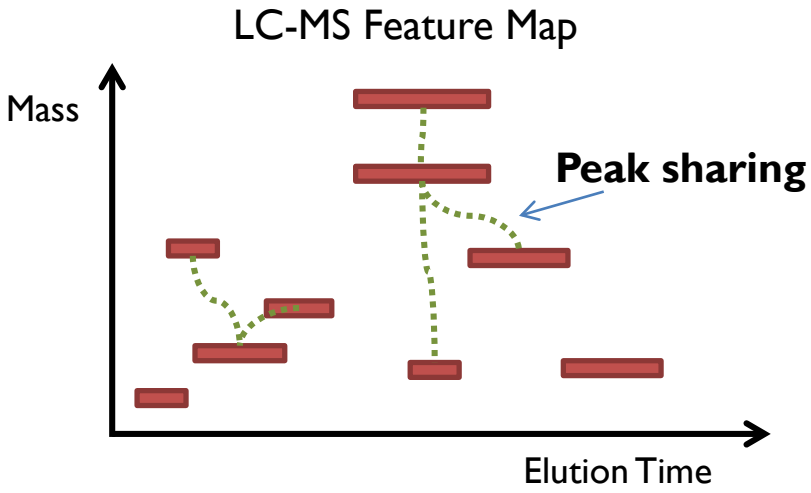
Choosing the Best Interpretation



Observed Peaks

Different Isotope Envelopes (red) can have good match

Filter features using likelihood score



Survival match in features sharing peaks

1) Make an edge between features if they share peaks
2) Find a set of connected features. If there is no set, stop
3) In a connected set, report the best and remove peaks, if the score > 0. Otherwise, go to 2)
4) Re-score remaining features in the set. Go to step 3).