

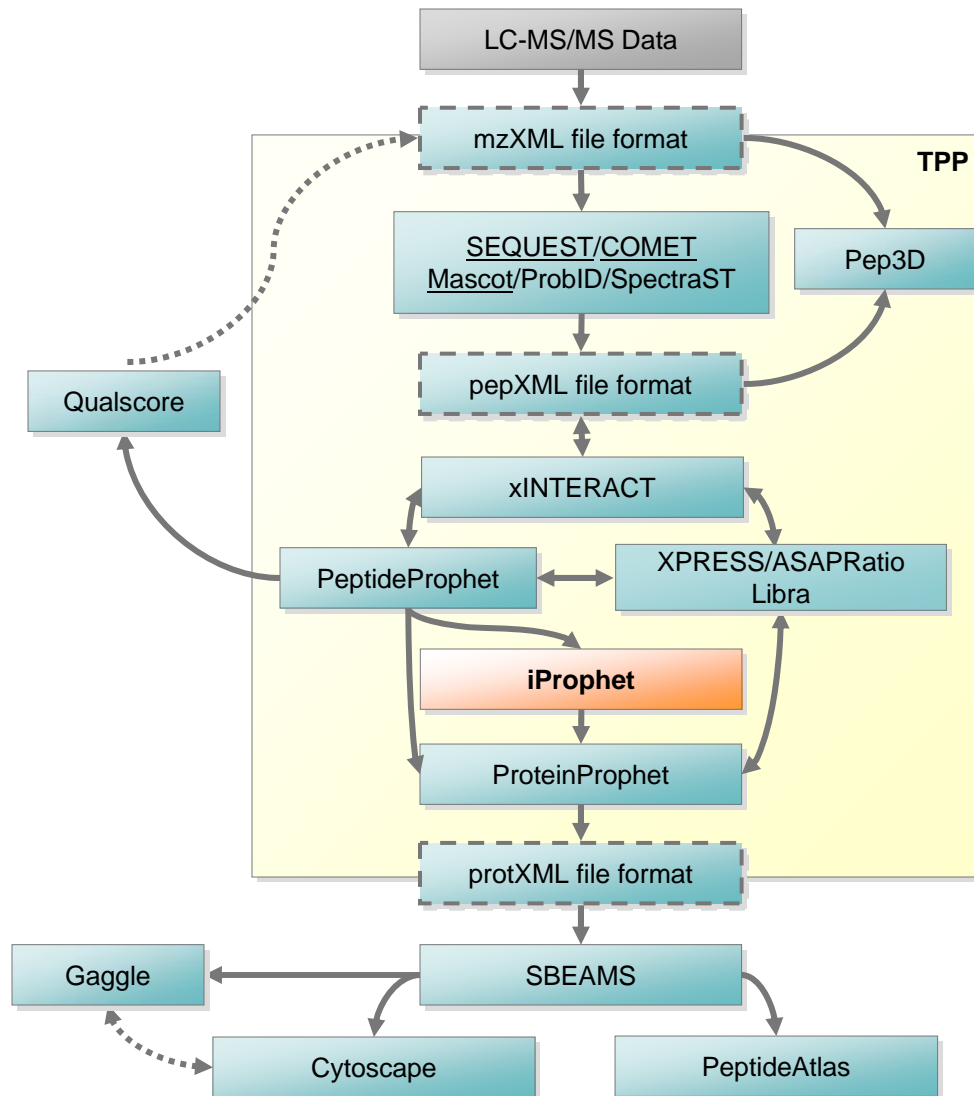
# iProphet: Statistical Refinement of PeptideProphet Results

David Shteynberg

Day 2

April 28, 2009

# Unique Peptide Sequence Validation



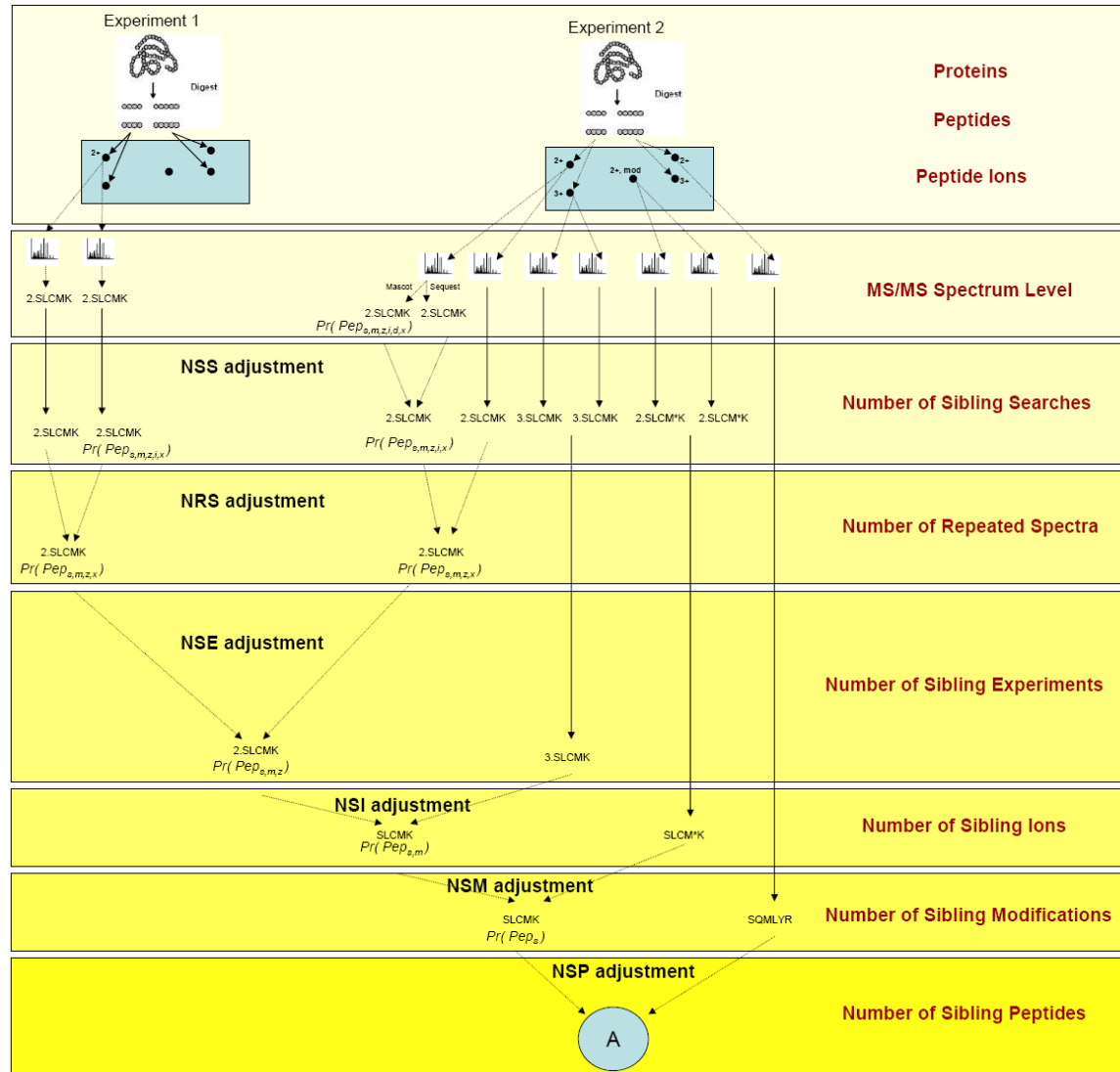
# iProphet: What is it?

---

- An intermediate between PeptideProphet and ProteinProphet
- A statistical tool that considers additional information about Peptide Spectrum Matches not currently considered by PeptideProphet or ProteinProphet

# iProphet Statistical Models

- **NSS**
  - Number Sibling Searches
- **NRS**
  - Number Replicate Spectra
- **NSE**
  - Number Sibling Experiments
- **NSI**
  - Number Sibling Ions
- **NSM**
  - Number Sibling Modifications



# NSS Model

- Number of Sibling Searches
  - Statistic used to represent parallel searches of the same spectrum with different search engines

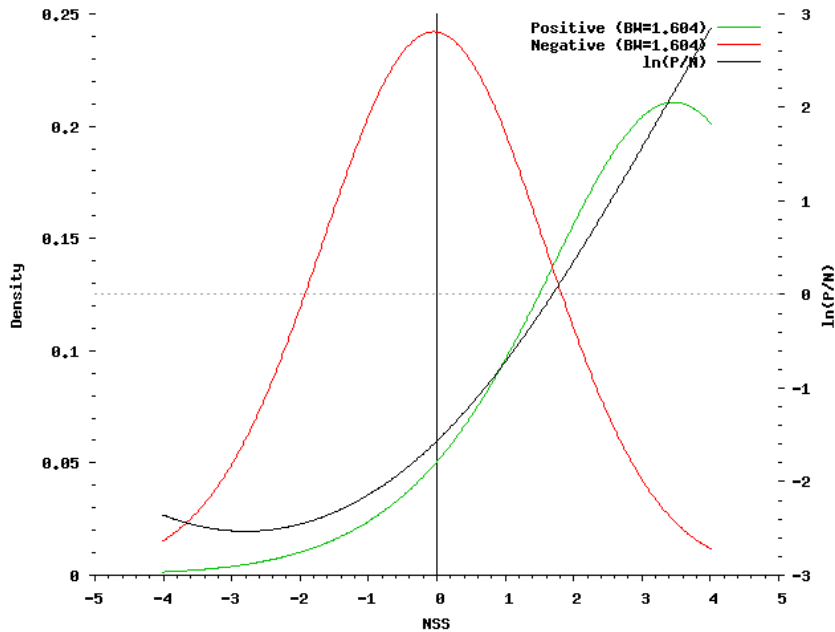
$$NSS_d = \sum_{\{d'|d \neq d' \wedge Pep_d = Pep_{d'}\}} \Pr(Pep_{d'}) - \sum_{\{d'|d \neq d' \wedge Pep_d \neq Pep_{d'}\}} \Pr(Pep_{d'})$$

$$\Pr(+ | NSS_d) = \frac{\Pr(NSS_d | +) \Pr(Pep_d)}{\Pr(NSS_d | +) \Pr(Pep_d) + \Pr(NSS_d | -) (1 - \Pr(Pep_d))}$$

$$\Pr(NSS_d | +) = \frac{\sum \Pr(Pep)}{\sum_{\forall Pep} \Pr(Pep)}$$

$$\Pr(NSS_d | -) = \frac{\sum (1 - \Pr(Pep))}{\sum_{\forall Pep} (1 - \Pr(Pep))}$$

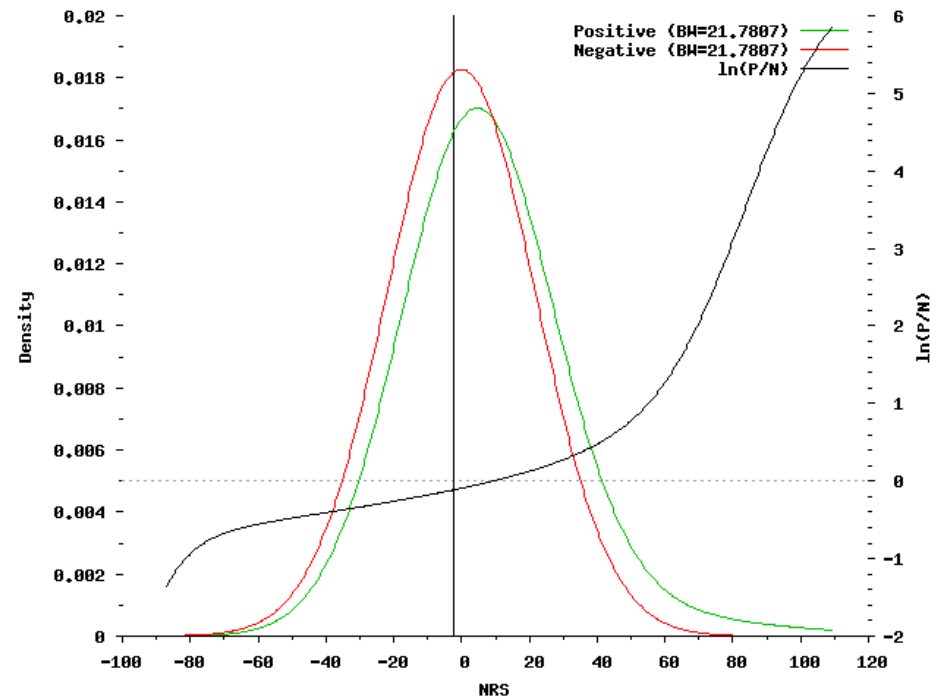
# NSS Model Example



- Dataset searched with 5 different search engines: OMSSA, Sequest, XTandem, Myrimatch and Inspect
- Range of NSS is [-4, 4]

# NRS Model

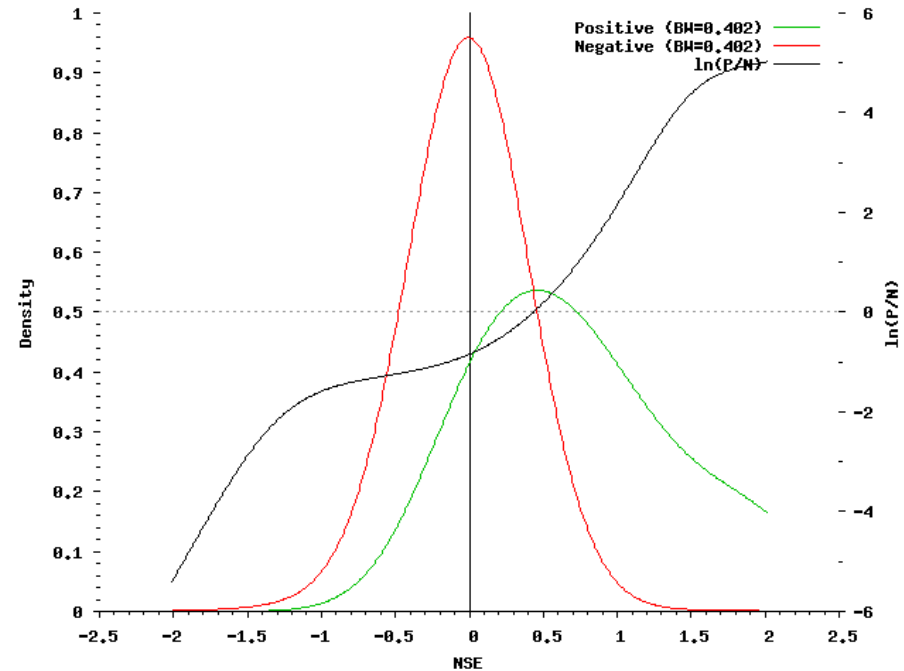
- Number of Replicate Spectra
  - Statistic used to represent repeated identifications of the same **peptide ion** in 1 experiment
  - Replicate spectra contribute a positive value to NRS if their PeptideProphet probability is above 0.5, and a negative value if their PeptideProphet probability is below 0.5
    - Prevent single spectrum wonders from getting punished by this model



$$NRS_i = \sum_{\{i' | i \neq i' \wedge Pep_i = Pep_{i'}\}} (\Pr(Pep_{i'}) - 0.5)$$

# NSE Model

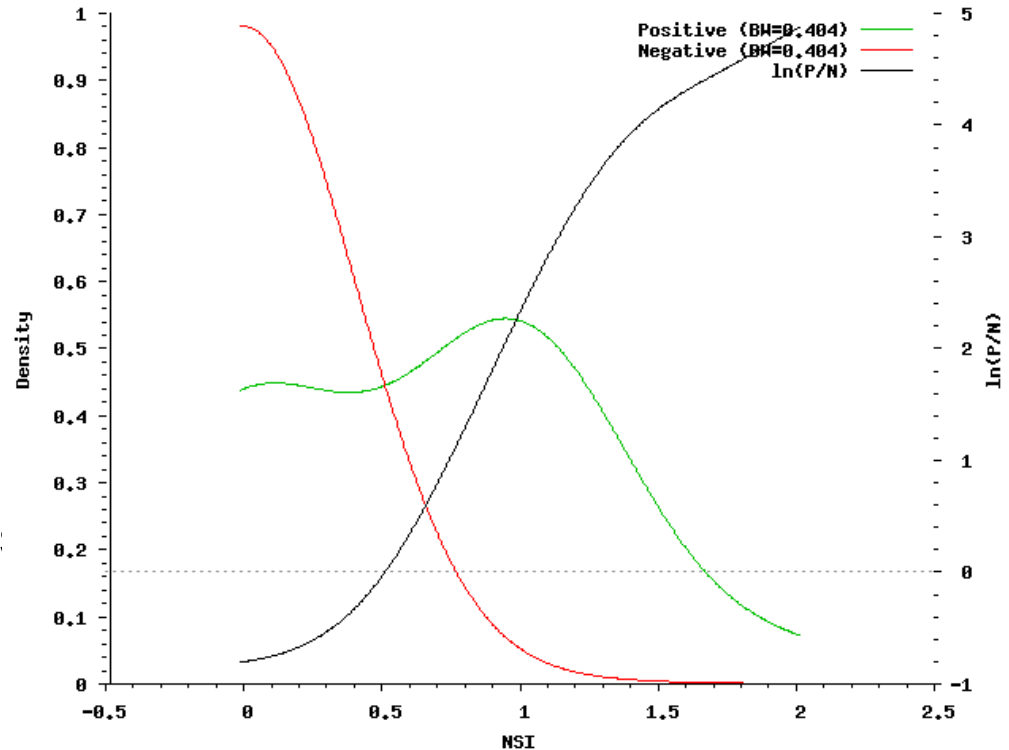
- Number of Sibling Experiments
  - Statistic used to represent repeated identifications of the same **peptide ion** across different experiment
  - Replicate spectra in a different experiment contribute a positive value to NSE if their PeptideProphet probability is above 0.5, and a negative value if their PeptideProphet probability is below 0.5
    - Prevents single experiment wonders from getting punished by this model



$$NSE_x = \sum_{\{x'|x \neq x' \wedge Pep_x = Pep_{x'}\}} (\Pr(Pep_{x'}) - 0.5)$$

# NSI Model

- Number of Sibling Ions
  - Statistic used to represent identifications of the same **peptide+mods** across different charge states
  - Same peptide sequence
  - Same peptide modification:
  - Different charge state

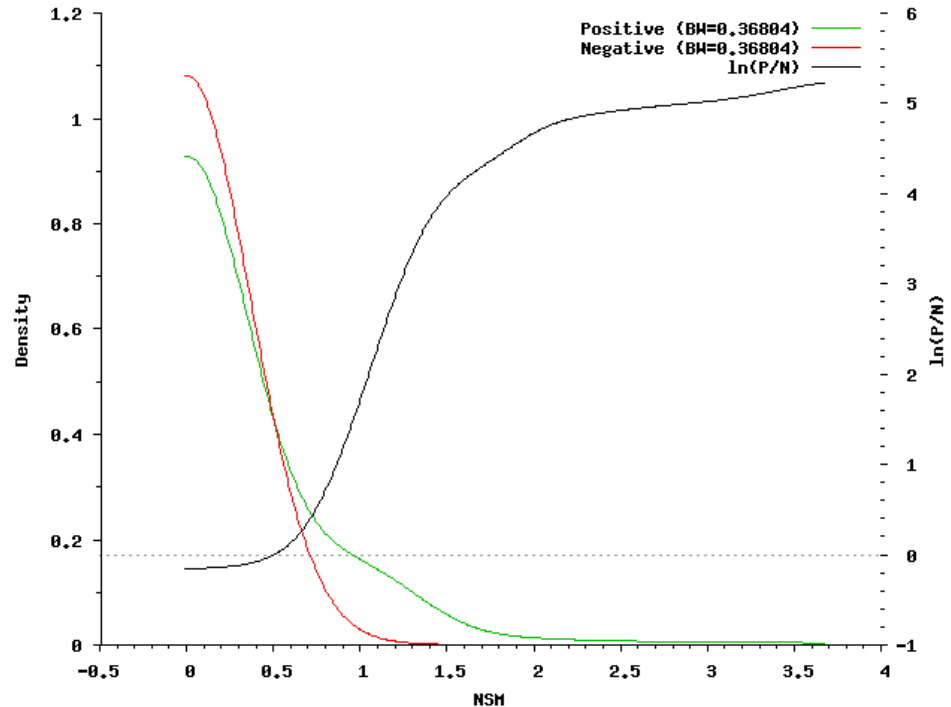


$$NSI_z = \sum_{\{z'|z \neq z'\}} \Pr(Pep_{z'})$$

# NSM Model

- Number of Sibling Modifications

- Statistic used to represent identifications of the same **peptide** sequence across different modified states
- Same peptide sequence
- Different peptide modifications



$$NSM_m = \sum_{\{m'|m \neq m'\}} \Pr(Pep_{m'})$$