

# PeptideProphet: Validation of Peptide Assignments to MS/MS Spectra

Andrew Keller  
Day 2  
October 17, 2006

Andrew Keller  
Rosetta Bioinformatics, Seattle

# Outline

---

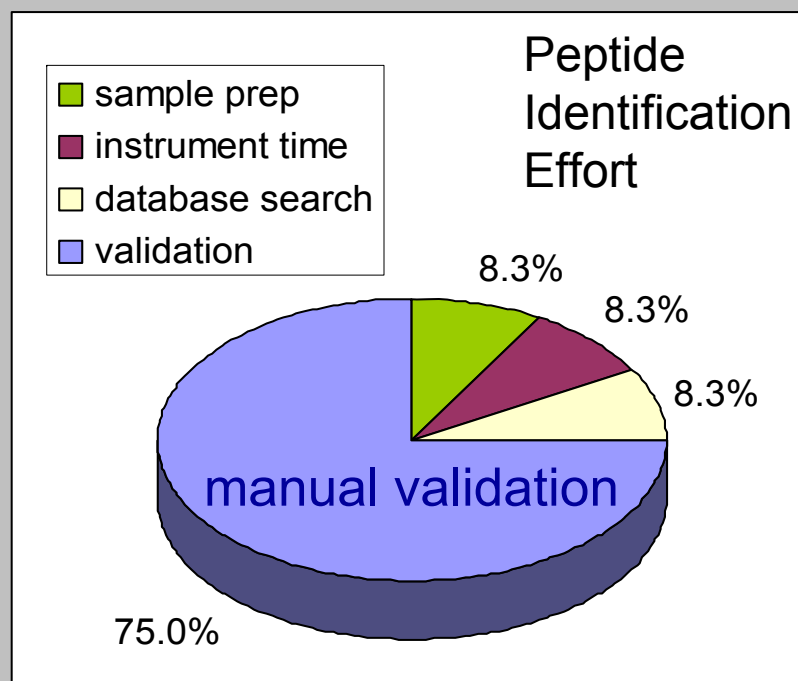
- Need to validate peptide assignments to MS/MS spectra
- Statistical approach to validation
- Running PeptideProphet software
- Interpreting results of PeptideProphet
- Exercises

# Most search results are wrong

- $[M+2H]^{2+}/[M+3H]^{3+}$  uncertainty (LCQ)
- Non-peptide noise
- Incomplete database
  - *e.g.* post-translational modifications
- Multiple precursors
- Limitation of database search algorithm

# Validation of Peptide Assignments

- In the past, a majority of analysis time was devoted to identifying the minority of correct search results from the majority of incorrect results
- Required manual judgment





# Need for Objective Criteria

---

- Manual scrutiny of search results is not practical for large datasets common to high throughput proteomics
- As an alternative to relying on human judgment, many research groups employ search scores and properties of the assigned peptides to discriminate between correct and incorrect results

# Traditional Filtering Criteria

- Each Mascot search result has:
  - Ionscore, Identityscore, Homologyscore, NTT (number of tryptic termini)
- Accept all results that satisfy:  
Ionscore > Identityscore

# Traditional Filtering Criteria

- Each Mascot search result has:
  - Ionscore, Identityscore, Homologyscore, NTT (number of tryptic termini)
- Accept all results that satisfy:  
Ionscore > Identityscore (NTT = 2)



# Traditional Filtering Criteria

- Each Mascot search result has:
  - Ionscore, Identityscore, Homologyscore, NTT (number of tryptic termini)
- Accept all results that satisfy:
  - Ionscore > Identityscore (NTT = 2)
  - Ionscore > Homologyscore

# Traditional Filtering Criteria

- Each Mascot search result has:
  - Ionscore, Identityscore, Homologyscore, NTT (number of tryptic termini)
- Accept all results that satisfy:
  - Ionscore > Identityscore (NTT = 2)
  - Ionscore > Homologyscore (NTT = 2)

# Traditional Filtering Criteria

- Each Mascot search result has:
  - Ionscore, Identityscore, Homologyscore, NTT (number of tryptic termini)
- Accept all results that satisfy:
  - Ionscore > Identityscore (NTT = 2)
  - Ionscore > Homologyscore (NTT = 2)
  - Ionscore  $\geq$  30

# Traditional Filtering Criteria

- Each Mascot search result has:
  - Ionscore, Identityscore, Homologyscore, NTT (number of tryptic termini)
- Accept all results that satisfy:
  - Ionscore > Identityscore (NTT = 2)
  - Ionscore > Homologyscore (NTT = 2)
  - Ionscore  $\geq$  30 (NTT = 2)

# Problems with Traditional Filtering

- Different research groups use different thresholds
- Divides data into correct and incorrect- no in between
- Unknown error rates (fraction of data passing filter that are incorrect)
- Unknown sensitivity (fraction of correct results passing filter)
- Appropriate threshold may depend on database, mass spectrometer type, sample, etc.

# Use of Forward/Reverse Database to Estimate False Positive Error Rates

- Do search against single Forward/Reverse database containing usual entries along with their sequence-reversed counterparts
- Forward and Reverse protein sequences each comprise 50% of the database peptides
- Incorrect results, taken at random from the database, are predicted to correspond with Reverse protein sequences on average 50% of the time
- Number of incorrect results passing any score filter calculated as twice the number of accepted results corresponding to Reverse proteins
- Search takes twice as long

# Use of Separate Forward and Reverse Database Searches

- Do searches against Forward and Reverse databases separately
- Number of incorrect results in Forward search passing any score filter calculated as the number of results passing the same filter applied to the Reverse search
- Gives an overestimate of the number of incorrect results passing a filter since compares the Reverse search which has no correct results with the Forward search which may have up to 100% correct results
- Results of 2 searches must be analyzed in parallel

# Statistical Approach

---

- Use search scores and properties of the assigned peptides to compute a probability that each search result is correct
- Desirable model properties:
  - Accurate
  - High power to discriminate correct and incorrect results
  - Robust



# Training Dataset

- Want dataset of Mascot search results for which the true correct and incorrect peptide assignments are known
- Sample of 18 control proteins (bovine, yeast, bacterial)
- Collect ~40,000 MS/MS spectra, and search using Mascot vs. a *Drosophila* database appended with sequences of 18 control proteins and common sample contaminants

# Training Dataset

Sort/Restore: Apply filtering below ... no sort or restore  [Help](#)

FILE: /data2/search/akeller/MASCOT\_TR/ORIG\_DTA\_UN tryptic termini:  1  2  MaxMissed: 9 DelRows:

Ion Score:  +1  +2  +3  InclAA:  MarkAA:  NXS/T:  Prob:  XPRESS:

Exclude charge: +1  +2  +3  Txt1:  Txt2:  J.Eng 04/2000

<a href="#">111 07.1321.1325.2</a>	1014.6 (+0.9)	56.24	61.70	48.73	<a href="#">8/ 16</a>	<a href="#">sp P02666 CASE_BOVIN</a>	F.LLYQEPVLGH
<a href="#">111 08.0780.0784.2</a>	1130.5 (+0.6)	60.89	61.63	34.49	<a href="#">11/ 18</a>	<a href="#">SW:AMY_BACLI</a>	K.GTSQADVGYK
<a href="#">111 07.1085.1087.2</a>	1229.7 (+0.1)	72.31	61.59	48.52	<a href="#">17/ 22</a>	<a href="#">sp P02754 LACE_BOVIN</a>	Y.VEELKPTPEK
<a href="#">111 05.1073.1075.2</a>	1229.7 (+0.6)	71.16	61.58	49.13	<a href="#">9/ 22</a>	<a href="#">sp P00921 CAH2_BOVIN</a>	K.YGDFGTAAGK
<a href="#">111 08.1079.1081.2</a>	1229.7 (+0.2)	71.62	61.58	48.99	<a href="#">9/ 22</a>	<a href="#">GP:AE003454_32</a>	M.SLLSNKSTKQ
<a href="#">111 01.0579.0581.2</a>	1245.6 (+0.9)	60.08	61.57	46.36	<a href="#">6/ 20</a>	<a href="#">sp P00921 CAH2_BOVIN</a>	R.TLNFNAEGEI
<a href="#">111 04.0607.0609.2</a>	1245.6 (+0.4)	53.44	61.57	40.53	<a href="#">18/ 20</a>	<a href="#">GP:AE003810_10</a>	S.MSQPKSKTEI
<a href="#">111 05.0610.0612.2</a>	1245.6 (+0.3)	56.95	61.57	43.10	<a href="#">8/ 20</a>	<a href="#">sp P00489 PHS2_RABIT</a>	K.VHINPNSLFI
<a href="#">111 06.0611.0613.2</a>	1245.6 (+0.9)	57.05	61.57	40.61	<a href="#">9/ 20</a>	<a href="#">sp P00921 CAH2_BOVIN</a>	R.MVNNGHSFNI
<a href="#">111 09.0615.0619.2</a>	1245.6 (+0.3)	55.94	61.57	41.05	<a href="#">7/ 20</a>	<a href="#">GP:AE003549_12</a>	I.LPIFSPSEKY
<a href="#">111 10.0621.0623.2</a>	1245.6 (+0.3)	59.76	61.57	40.41	<a href="#">6/ 20</a>	<a href="#">GP:AE003824_13</a>	H.VKTDEVNQNK
<a href="#">111 11.0628.0630.2</a>	1245.6 (+0.3)	56.25	61.57	42.70	<a href="#">16/ 20</a>	<a href="#">sp P46406 G3P_RABIT</a>	K.WGDAGAEYVY
<a href="#">111 14.0628.0630.2</a>	1245.6 (+0.3)	73.05	61.57	43.50	<a href="#">9/ 20</a>	<a href="#">GP:AE003671_25</a>	A.PNGDLYAQPI
<a href="#">111 06.0622.0622.2</a>	1245.6 (+0.5)	68.14	61.57	50.28	<a href="#">9/ 20</a>	<a href="#">sp P00432 CATA_BOVIN</a>	K.DAQLFIQK.I
<a href="#">111 08.0618.0622.2</a>	1245.6 (+0.8)	70.90	61.57	50.29	<a href="#">8/ 20</a>	<a href="#">sp Q29443 TRFE_BOVIN</a>	K.TYDSYLGDDY
<a href="#">111 08.0860.0862.2</a>	1357.8 (+1.1)	59.29	61.51	37.54	<a href="#">9/ 24</a>	<a href="#">sp P02666 CASE_BOVIN</a>	K.IHPFAQTQSI
<a href="#">111 07.0685.0687.2</a>	1113.6 (+0.7)	64.47	61.49	36.28	<a href="#">9/ 20</a>	<a href="#">sp P00921 CAH2_BOVIN</a>	K.VGDANPALQI
<a href="#">111 07.2180.2182.2</a>	1473.9 (+0.9)	59.82	61.48	34.99	<a href="#">9/ 22</a>	<a href="#">sp P00489 PHS2_RABIT</a>	K.ARPEFTLPLVH
<a href="#">111 02.1156.1158.2</a>	1262.6 (+0.8)	57.86	61.47	57.04	<a href="#">11/ 18</a>	<a href="#">sp P02769 ALBU_BOVIN</a>	R.HPYFYAPELI
<a href="#">111 05.1211.1213.2</a>	1262.6 (+0.4)	57.80	61.47	53.96	<a href="#">16/ 18</a>	<a href="#">sp P02754 LACE_BOVIN</a>	R.VYVEELKPTI
<a href="#">111 08.1043.1045.2</a>	1389.7 (+1.0)	58.35	61.47	53.88	<a href="#">8/ 22</a>	<a href="#">GP:AE003473_27</a>	L.KKTKKPKLTY
<a href="#">111 08.1019.1023.2</a>	1346.7 (+0.9)	54.52	61.46	46.05	<a href="#">9/ 22</a>	<a href="#">sp P00921 CAH2_BOVIN</a>	R.TLNFNAEGEI
<a href="#">111 02.0855.0867.2</a>	1475.8 (+0.5)	57.81	61.42	46.40	<a href="#">8/ 22</a>	<a href="#">SWN:VAF2_DRONE</a>	T.VDAHNLAVPV
<a href="#">111 10.1587.1589.2</a>	1440.7 (+1.1)	53.77	61.42	40.42	<a href="#">9/ 22</a>	<a href="#">GP:AE003820_20</a>	Q.KLVNGGQSQI

Internet

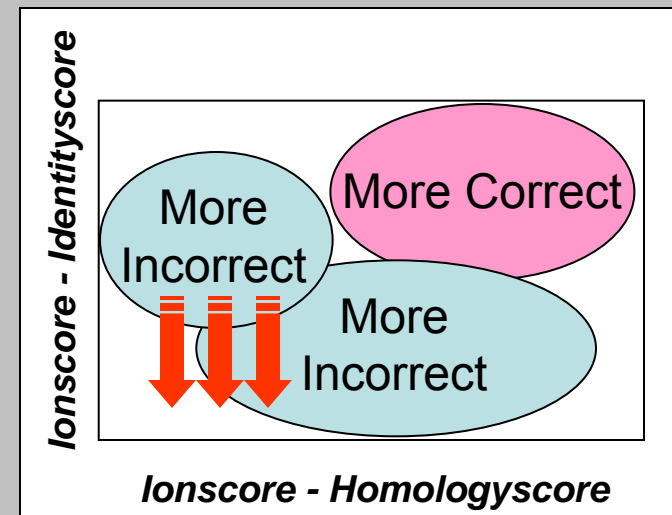
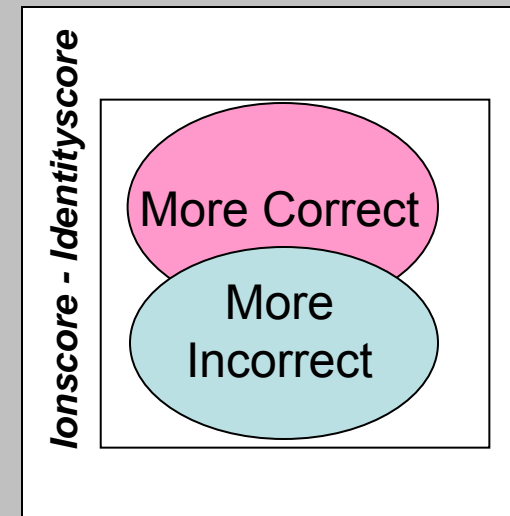
- Peptides corresponding to *Drosophila* proteins are incorrect
- Peptides corresponding to 18 control proteins or contaminants are correct\*

# Derive Discriminant Function

- Derive single search score best at discriminating correct from incorrect search results
  - Generally, can combine together multiple search engine scores, when available, into single linear combination score using Linear Discriminant Function Analysis (*e.g.* SEQUEST's Xcorr, DeltaCn, and SpRank)
  - Use search engine score directly if only one
- Derive separately for search results of each parent ion charge (1+, 2+, and 3+)

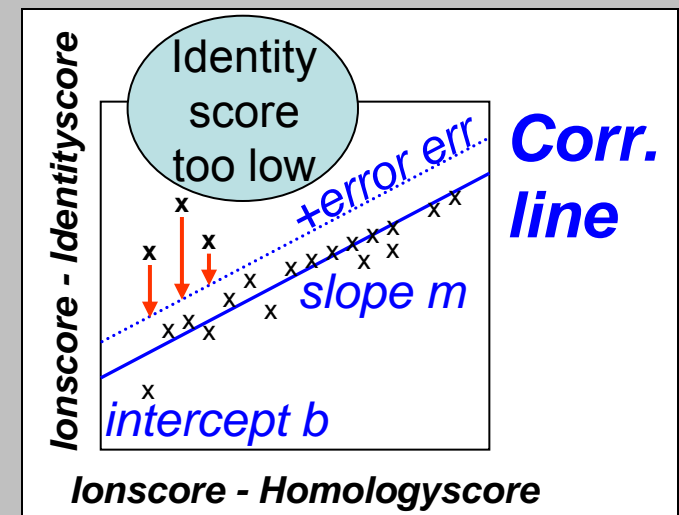
# Mascot Discriminant Function

- Use (Ionscore – Identityscore) difference
- Secondly, use (Ionscore – Homologyscore) difference to penalize some predominantly incorrect results and improve discrimination



# Mascot Discriminant Function

- In particular, use the (Ionscore – Identityscore) difference adjusted for the Average Identityscore in the dataset for given parent ion charge
- Require (Ionscore – Identityscore) not exceed  $m^*(\text{Ionscore} - \text{Homologyscore}) + b + \text{err}$ , where  $m$ ,  $b$ , and  $\text{err}$  are correlation parameters learned from the data for each parent ion charge
- Discriminant Function,  $F = 0.1 * \{(\text{Ionscore} - \text{Identityscore}) + \text{Average Identityscore}\} - 3.0$



# Compute Discriminant Score

*Example:*

Peptide = **LSISGTYDLK**

Precursor Ion Charge = 2

Ionscore = 50.91

Identityscore = 46

Homologyscore = 37

Ave. Identityscore = 47

Corr. Slope = 0.53, Intercept = -6.99, Error = 10

$(\text{Ionscore} - \text{Identityscore}) = 50.91 - 46 = 4.91$

$(\text{Ionscore} - \text{Identityscore})$  not allowed to exceed

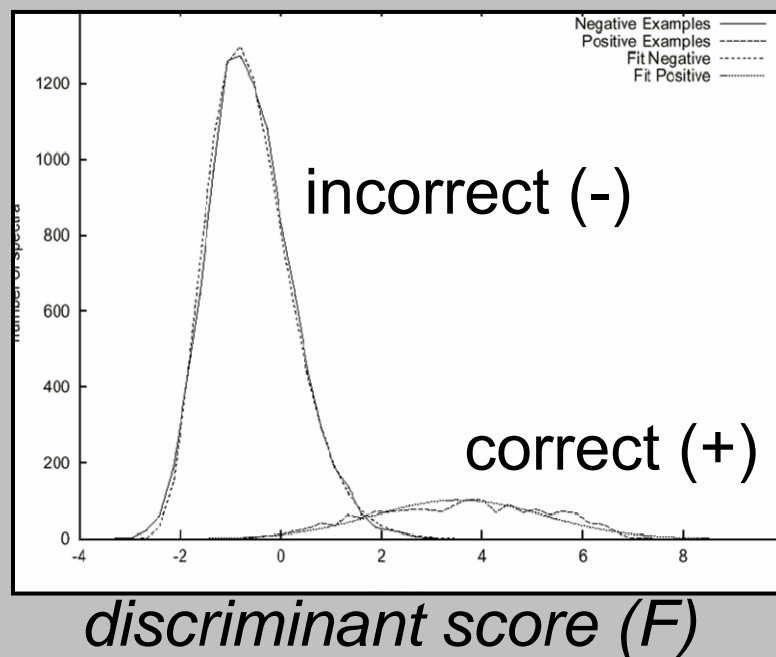
$0.53 * (\text{Ionscore} - \text{Homologyscore}) - 6.99 + 10,$

or  $0.53 * (50.91 - 37) - 6.99 + 10 = 10.38$

$F = 0.1 * \{4.91 + 47\} - 3.0 = 2.19$

# Discriminant Score Distributions

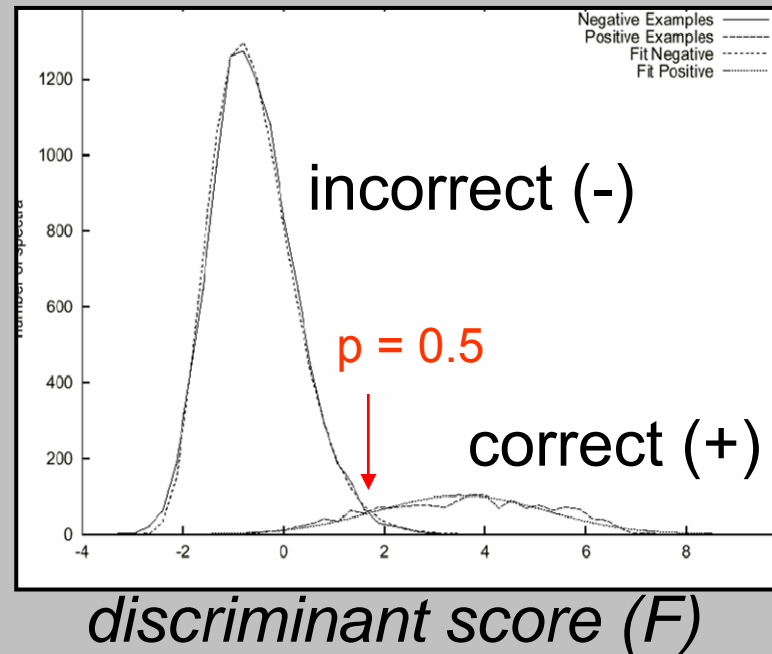
*no.  
of  
spectra*



Training dataset  $[M+2H]^{2+}$  spectra

# Computing probabilities from discriminant score distributions

*no.  
of  
spectra*



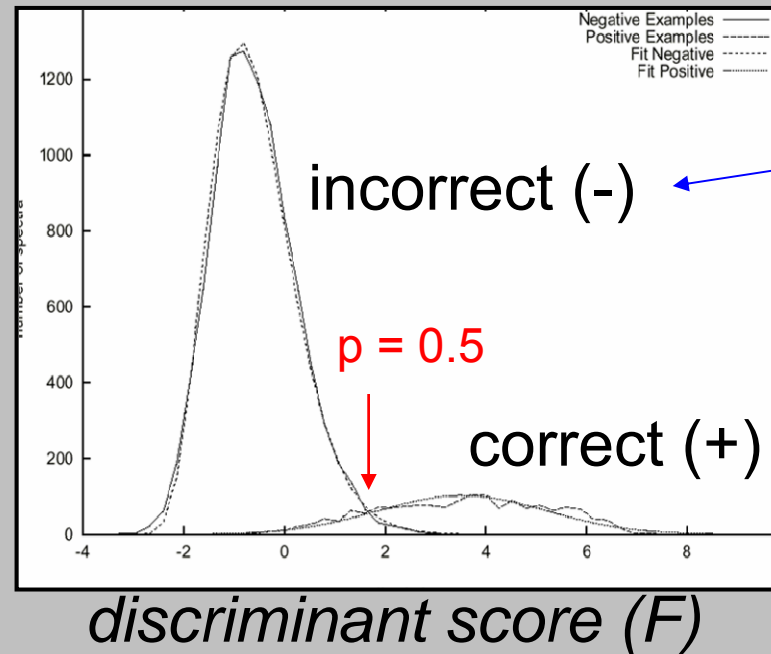
Probability of being correct, given discriminant score  $F_{\text{obs}}$ , is:

$$p = \frac{\text{Number of correct search results with } F_{\text{obs}}}{\text{Total number of search results with } F_{\text{obs}}}$$



# Computing probabilities from discriminant score distributions

*no.  
of  
spectra*



Model Incorrect  
results as  
**Extreme Value  
Distribution  
(EVD)**

Model Correct  
results as  
**Normal  
Distribution**

Probability of being correct, given discriminant score  $F_{\text{obs}}$ , is:

$$p = \frac{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total correct}}{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total correct} + \text{EVD}_{\beta,\mu}(F_{\text{obs}}) * \text{Total incorrect}}$$

# Employing peptide properties

---

- Properties of the assigned peptides, in addition to search scores, are useful information for distinguishing correct and incorrect results.
- For example, in unconstrained Mascot searches with MS/MS spectra collected from trypsinized samples, a majority of correct assigned peptides have 2 tryptic termini (preceded by K,R), whereas a majority of incorrect assigned peptides have 0 tryptic termini.

# Number of Tryptic Termini (NTT)

NTT can equal 0, 1, or 2:

**G**.HVEQLDSS**S**.D NTT = 0

**K**.HVEQLDSS**S**.D NTT = 1

**G**.HVEQLDSS**R**.D NTT = 1

**K**.HVEQLDSS**R**.D NTT = 2

# Number of Tryptic Termini (NTT)

For the same value of  $F$ , assigned peptides with *higher* NTT values are *more* likely to be correct

Example: training dataset

Correct: 0.03 NTT=0, 0.28 NTT=1, **0.69** NTT=2

Incorrect: 0.80 NTT=0, 0.19 NTT=1, **0.01** NTT=2

Probability of being correct, given discriminant score  $F_{\text{obs}}$  with **NTT=2** is:

$$p = \frac{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total corr} * \mathbf{0.69}}{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total corr} * \mathbf{0.69} + \text{EVD}_{\beta,\mu}(F_{\text{obs}}) * \text{Total incorr} * \mathbf{0.01}}$$

$F_{\text{obs}}$ :  $p = 0.5$  without NTT becomes  $p=0.99$  using NTT

# Number of Tryptic Termini (NTT)

For the same value of  $F$ , assigned peptides with *lower* NTT values are *less* likely to be correct

Example: training dataset

Correct: 0.03 NTT=0, 0.28 NTT=1, 0.69 NTT=2

Incorrect: 0.80 NTT=0, 0.19 NTT=1, 0.01 NTT=2

Probability of being correct, given discriminant score  $F_{\text{obs}}$  with **NTT=0** is:

$$p = \frac{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total corr} * 0.03}{\text{Normal}_{\mu,\sigma}(F_{\text{obs}}) * \text{Total corr} * 0.03 + \text{EVD}_{\beta,\mu}(F_{\text{obs}}) * \text{Total incorr} * 0.80}$$

$F_{\text{obs}}$ :  $p = 0.5$  without NTT becomes  $p=0.04$  using NTT

# Additional Peptide Properties

- Number of missed tryptic cleavages (NMC)
- Mass difference between precursor ion and peptide
- Presence of light or heavy cysteine (ICAT)
- Presence of N-glyc motif (N-glycosylation capture)
- Calculated pI (FFE)

Incorporate similar to NTT above, assuming independence of peptide properties and search scores among correct and incorrect results

# Computed Probabilities

Given training dataset distributions of F, NTT, NMC, Massdiff, ICAT, N-glyc, and pl among correct and incorrect search results,...

...then the probability of any search result with  $F_{\text{obs}}$ ,  $\text{NTT}_{\text{obs}}$ ,  $\text{NMC}_{\text{obs}}$ ,  $\text{Massdiff}_{\text{obs}}$ ,  $\text{ICAT}_{\text{obs}}$ ,  $\text{N-glyc}_{\text{obs}}$ , and  $\text{pl}_{\text{obs}}$  can be computed as described above, with terms for each piece of information

- Accurate
- Discriminating

# Robust Model

One cannot rely on the *training dataset* distributions of F, NTT, NMC, Massdiff, ICAT, N-glyc, and pl among correct and incorrect search results

These distributions are expected to vary depending upon:

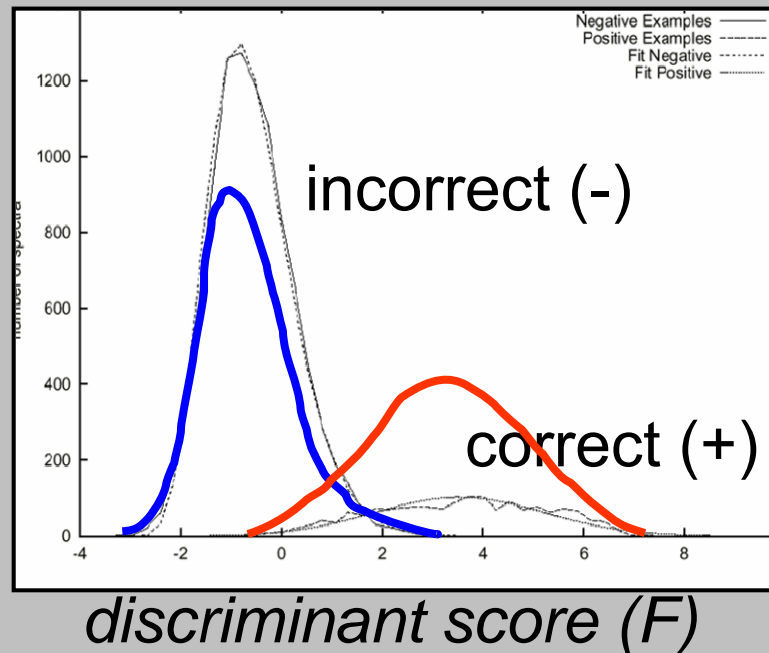
- Database used for search
- Mass spectrometer
- Spectrum quality
- Sample preparation and purity



# Variations in Discriminant Score Distributions

*Different proportion of correct results in dataset*

*no  
of  
spectra*

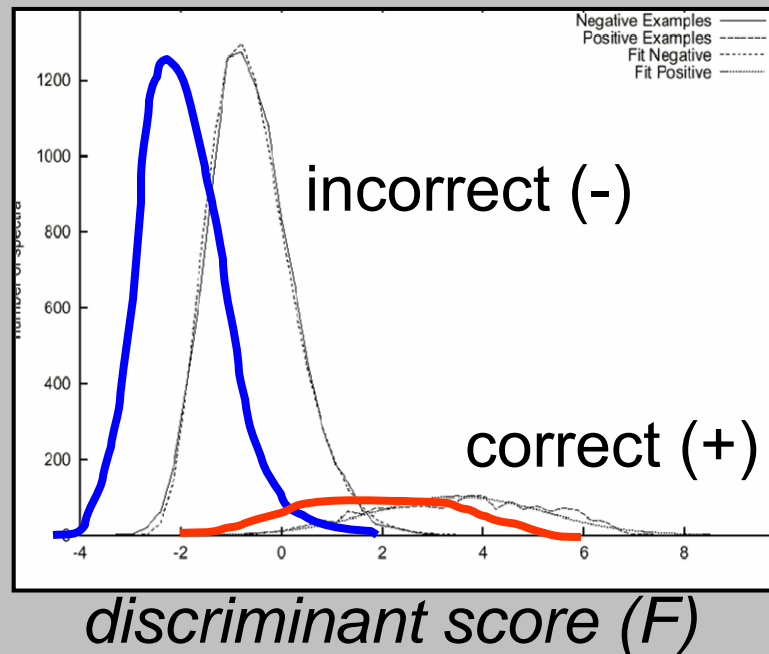


vs. training dataset  $[M+2H]^{2+}$  spectra

# Variations in Discriminant Score Distributions

*Different distribution means*

*no  
of  
spectra*



vs. training dataset  $[M+2H]^{2+}$  spectra

# EM Algorithm

- PeptideProphet learns the distributions of F and peptide properties among correct and incorrect search results in each dataset
- It then uses the learned distributions to compute probabilities that each search result is correct
- Expectation-Maximization (EM) algorithm: unsupervised learning method that *iteratively* estimates the distributions given probabilities that each search result is correct, and then computes those probabilities given the distributions
- Initial settings help guide algorithm to good solution

# EM Algorithm Details

## 1. Initial estimates of result probabilities

<i>Search Result</i>	<i>F</i>	<i>NTT</i>	<i>prob</i>	<i>1-prob</i>
A	3.0	2	1.0	0.0
B	2.0	1	0.5	0.5
C	1.0	1	0.5	0.5
D	0.0	0	0.0	1.0

## 2. Update F value distributions among correct and incorrect results

$$P(F|+): m = \frac{(3.0)(1.0) + (2.0)(0.5) + (1.0)(0.5) + (0.0)(0.0)}{1.0 + 0.5 + 0.5 + 0.0} = 2.25$$

$$P(F|-): m = \frac{(3.0)(0.0) + (2.0)(0.5) + (1.0)(0.5) + (0.0)(1.0)}{0.0 + 0.5 + 0.5 + 1.0} = 0.75$$

## 3. Update NTT distributions among correct and incorrect results

$$P(NTT=0|+) = (0.0) / (1.0 + 0.5 + 0.5 + 0.0) = 0.0$$

$$P(NTT=1|+) = (0.5 + 0.5) / (1.0 + 0.5 + 0.5 + 0.0) = 0.5$$

$$P(NTT=2|+) = (1.0) / (1.0 + 0.5 + 0.5 + 0.0) = 0.5$$

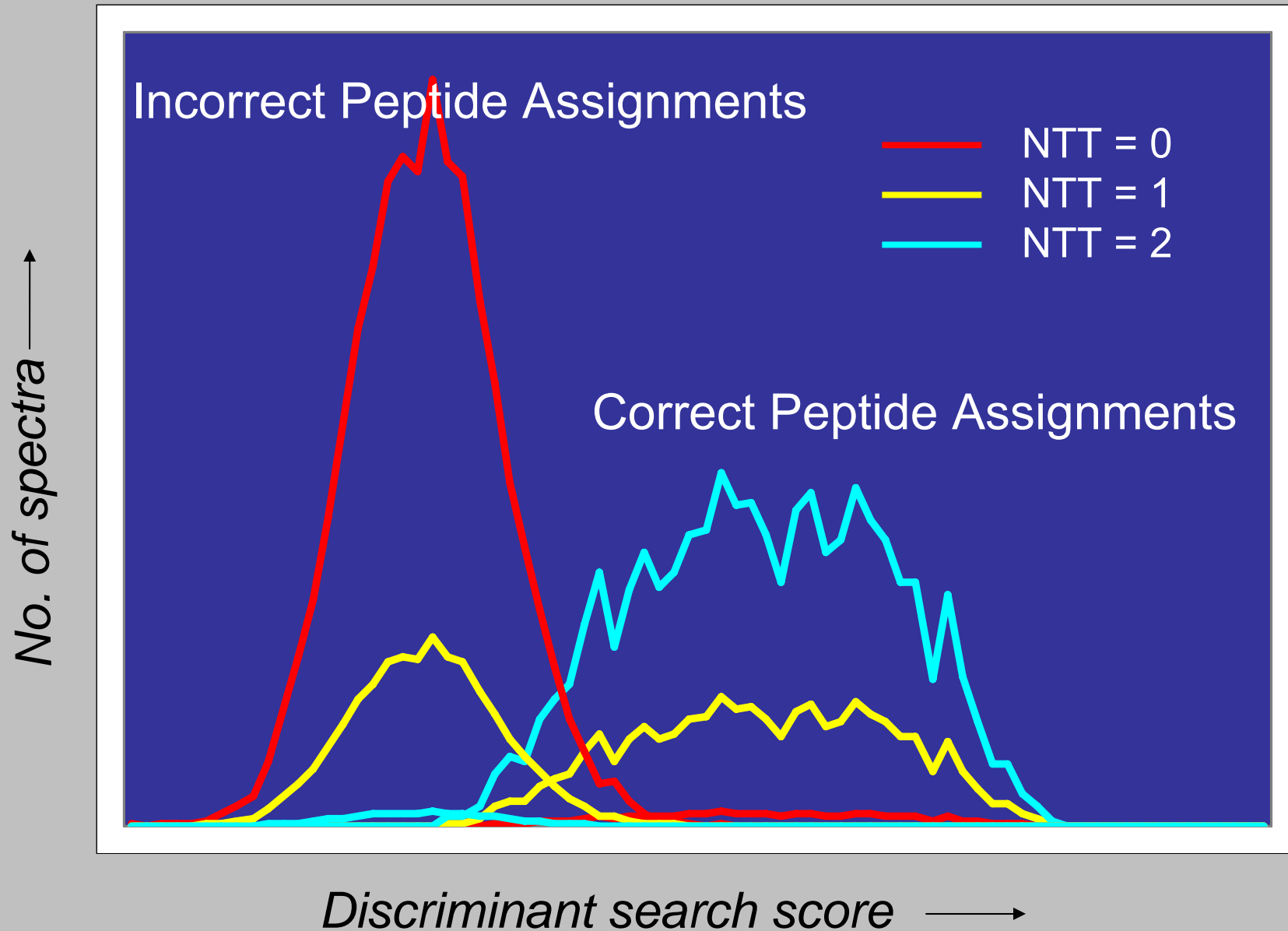
$$P(NTT=0|-) = (1.0) / (0.0 + 0.5 + 0.5 + 1.0) = 0.5$$

$$P(NTT=1|-) = (0.5 + 0.5) / (0.0 + 0.5 + 0.5 + 1.0) = 0.5$$

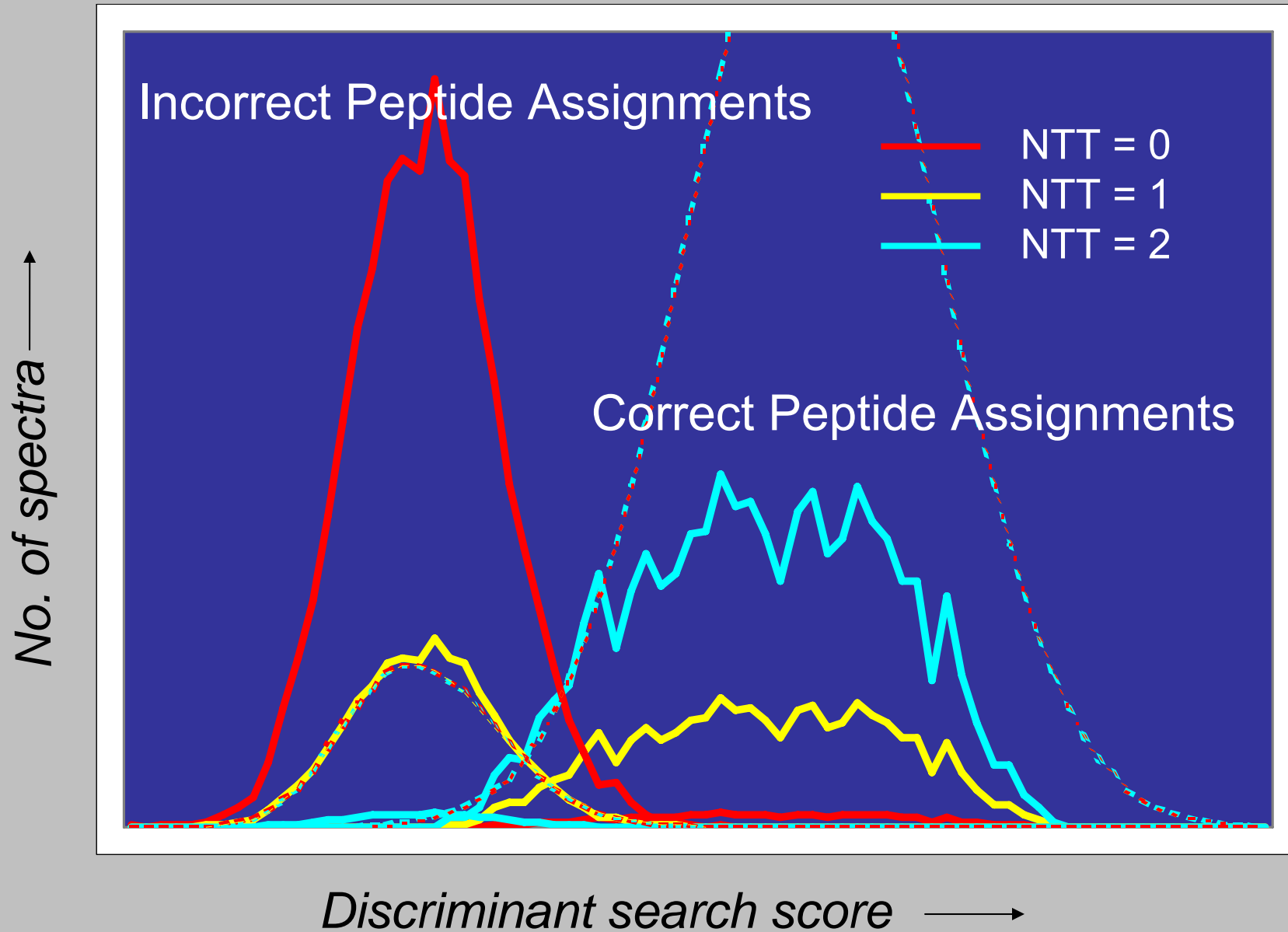
$$P(NTT=2|-) = (0.0) / (0.0 + 0.5 + 0.5 + 1.0) = 0.0$$

## 4. Recompute result probabilities using updated distributions, and iterate

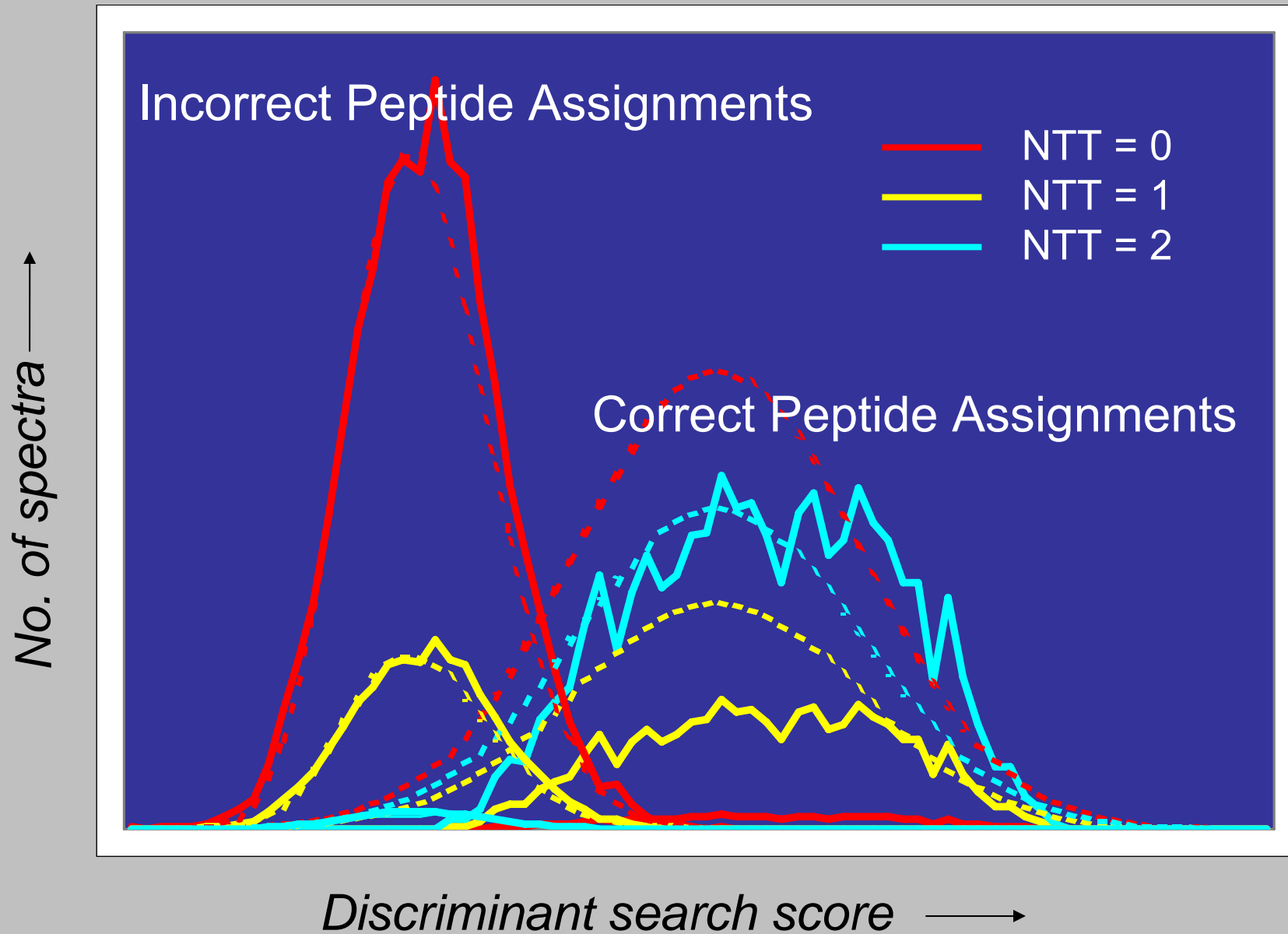
# EM Algorithm learns test data score distributions



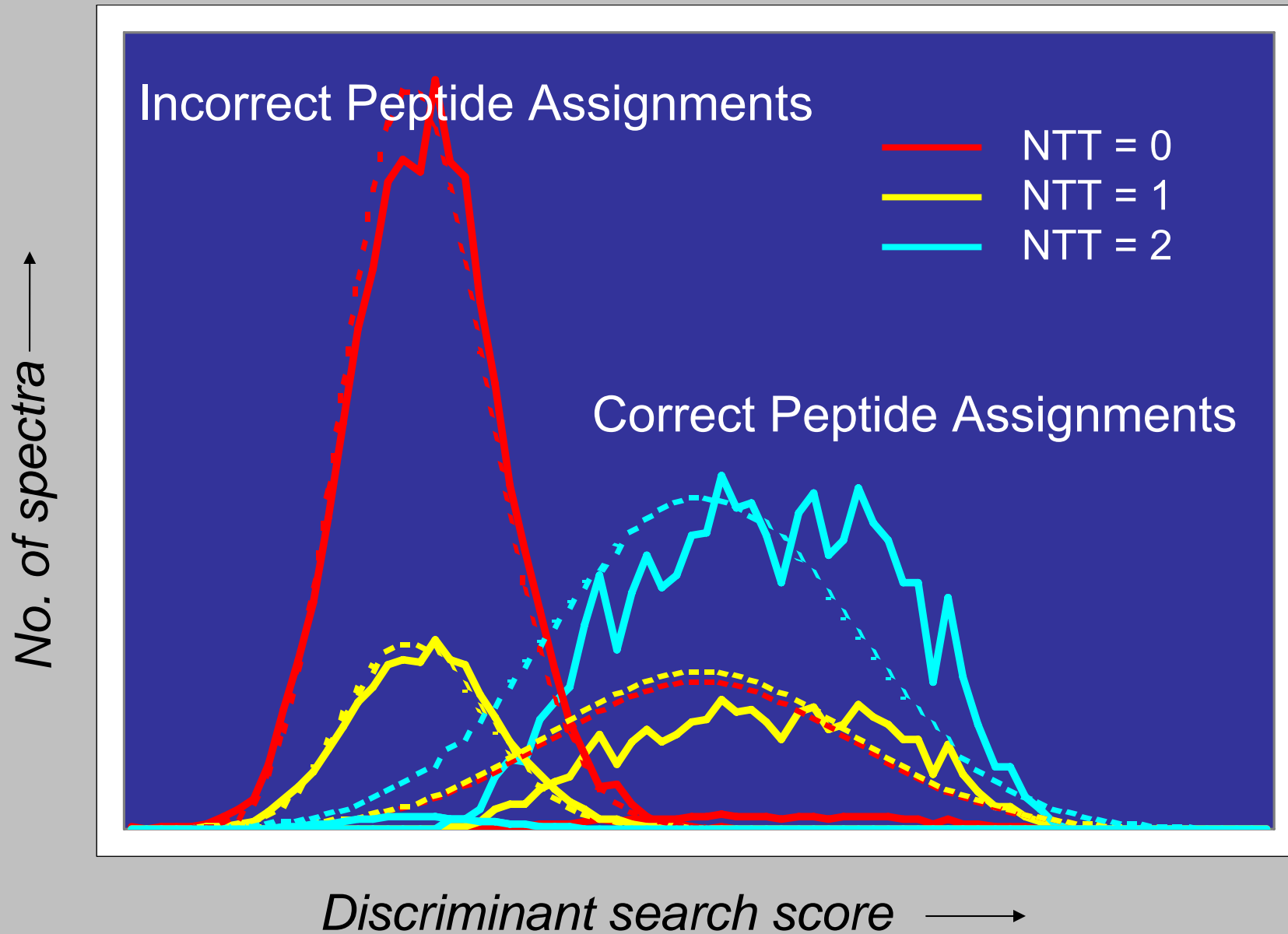
# EM Iteration 0



# EM Iteration I

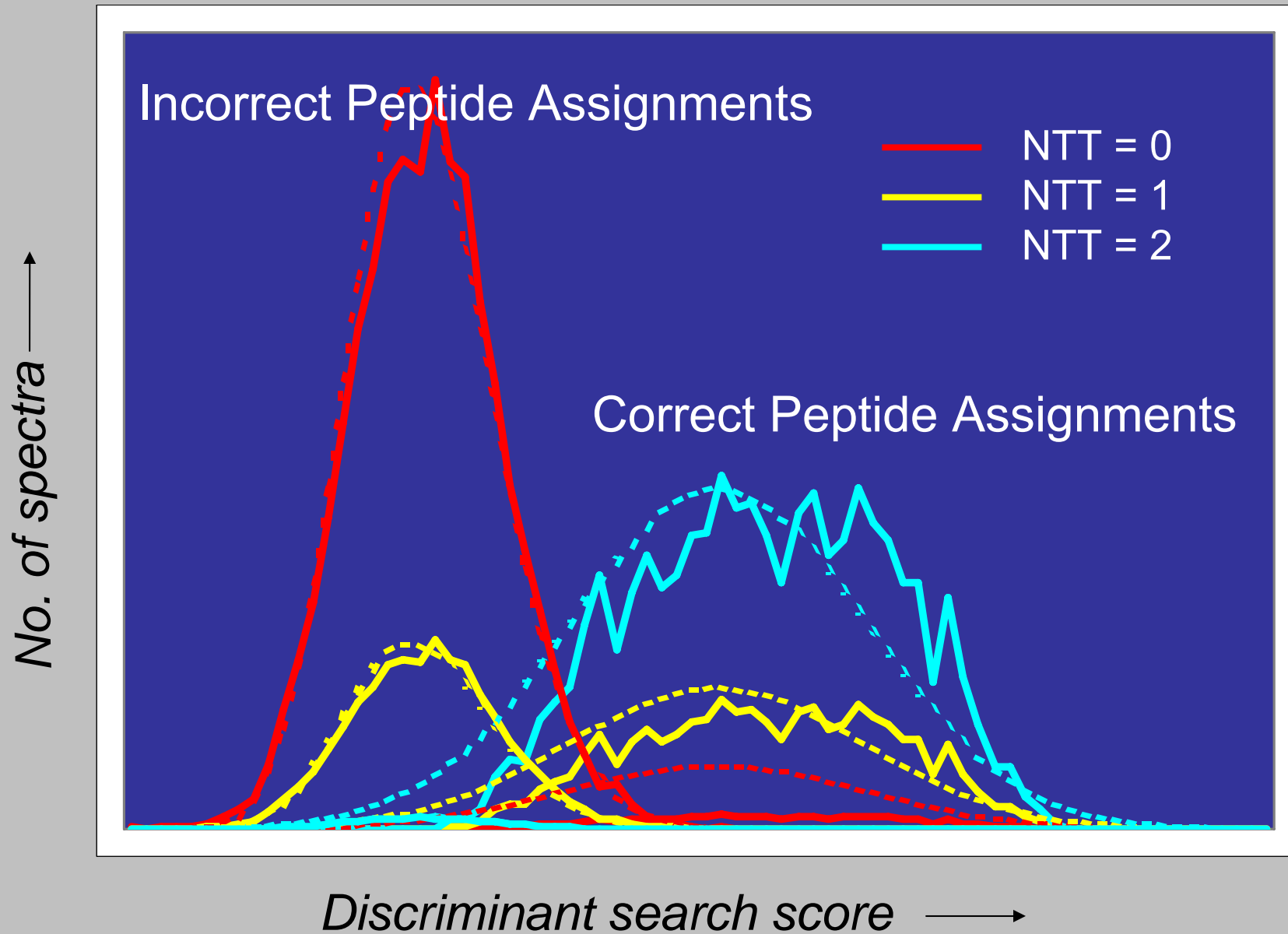


# EM Iteration 2

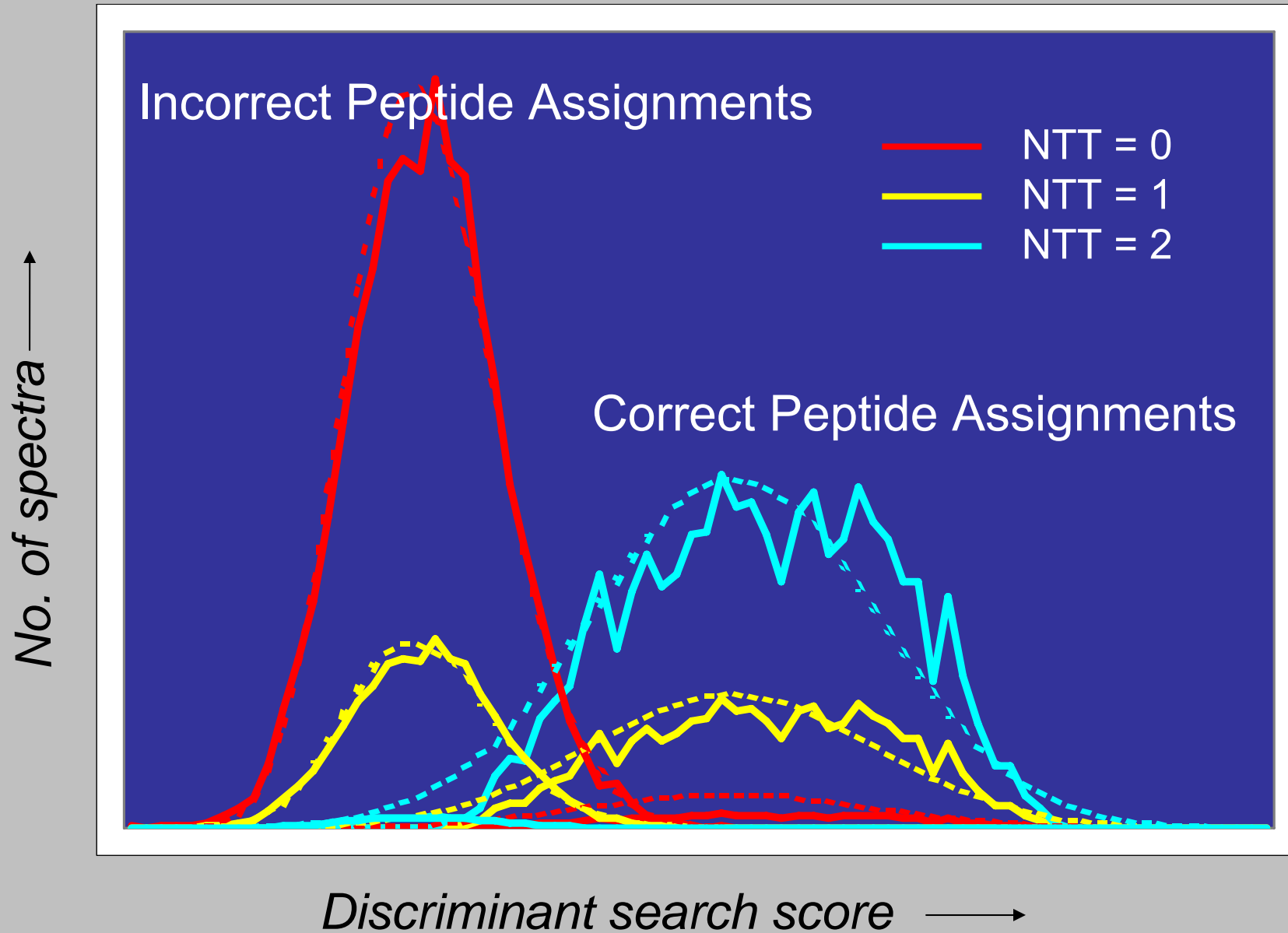




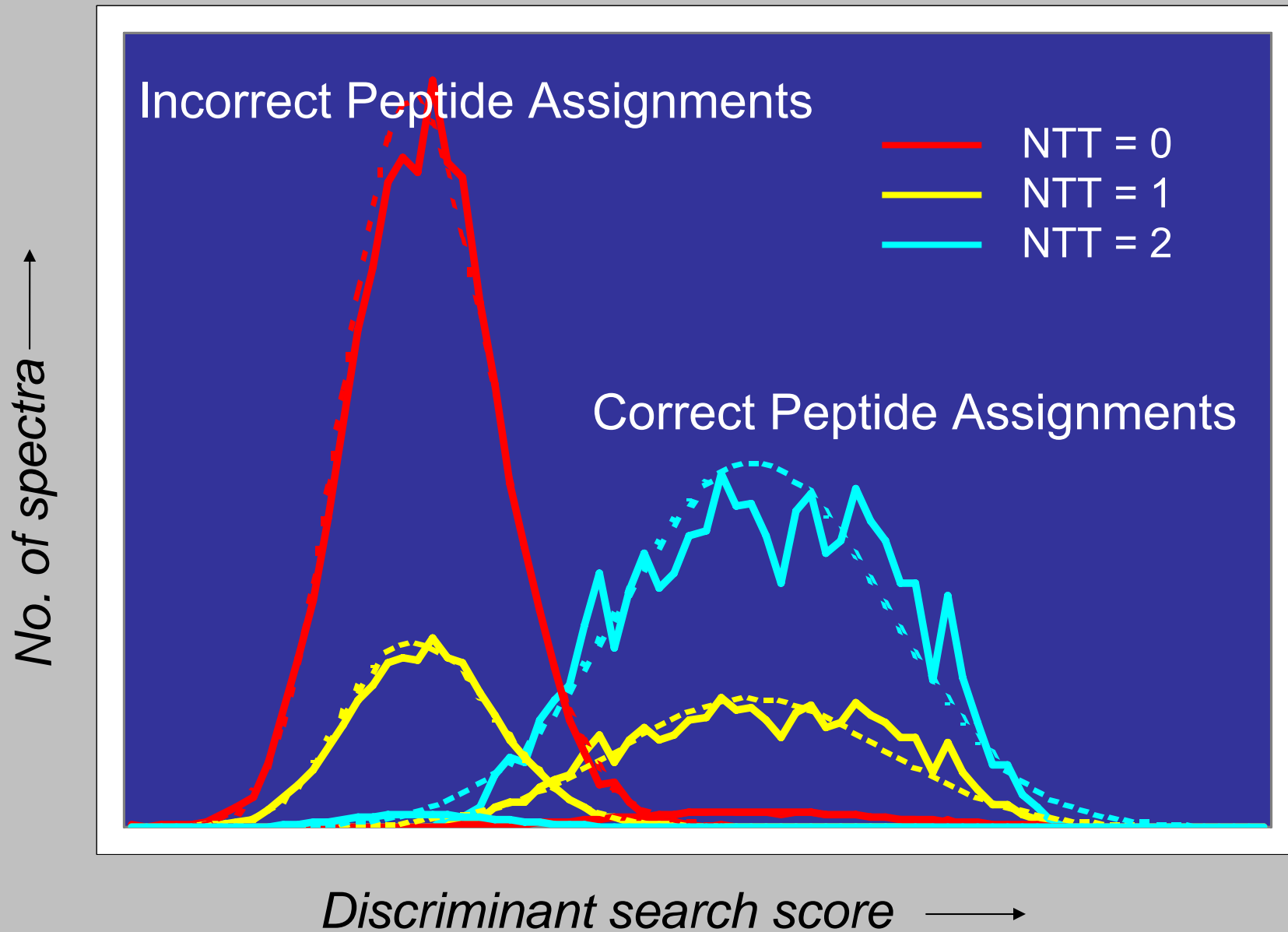
# EM Iteration 3



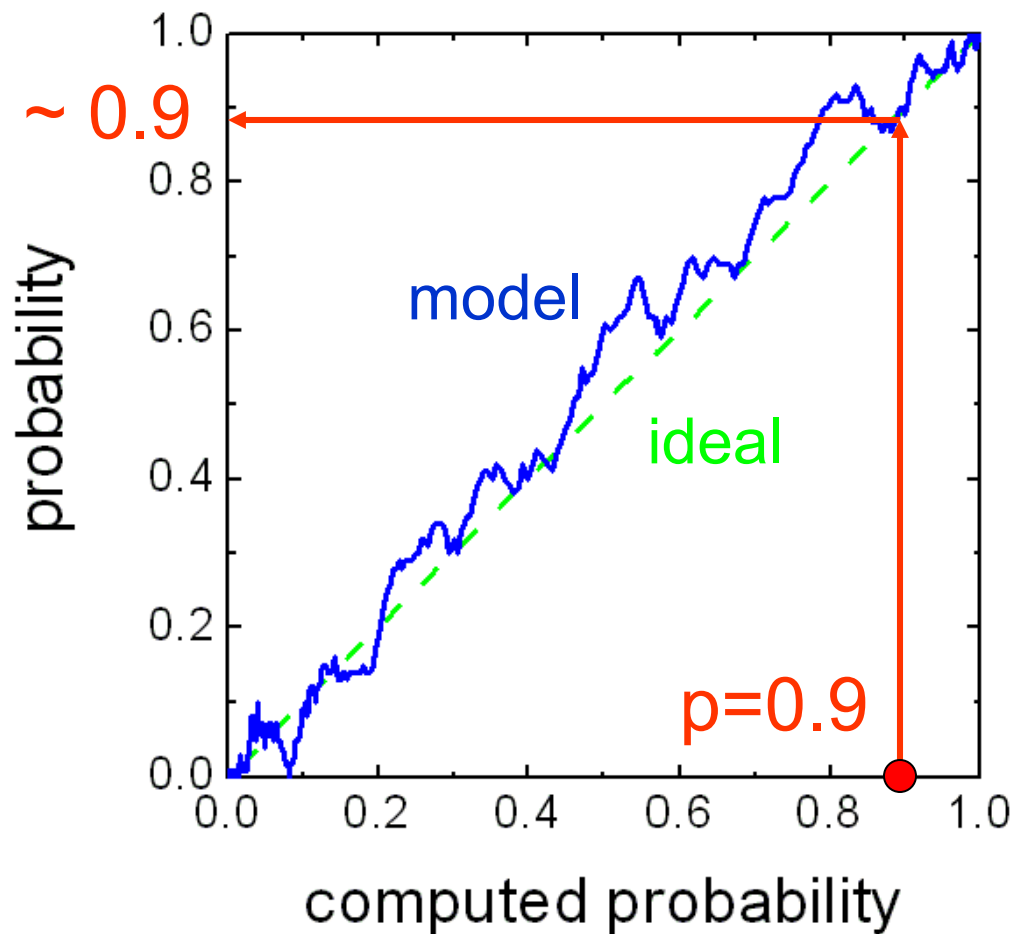
# EM Iteration 4



# EM Iteration 7



# Accuracy of the Model



test data: A. Keller *et al.* OMICS 6, 207 (2002)

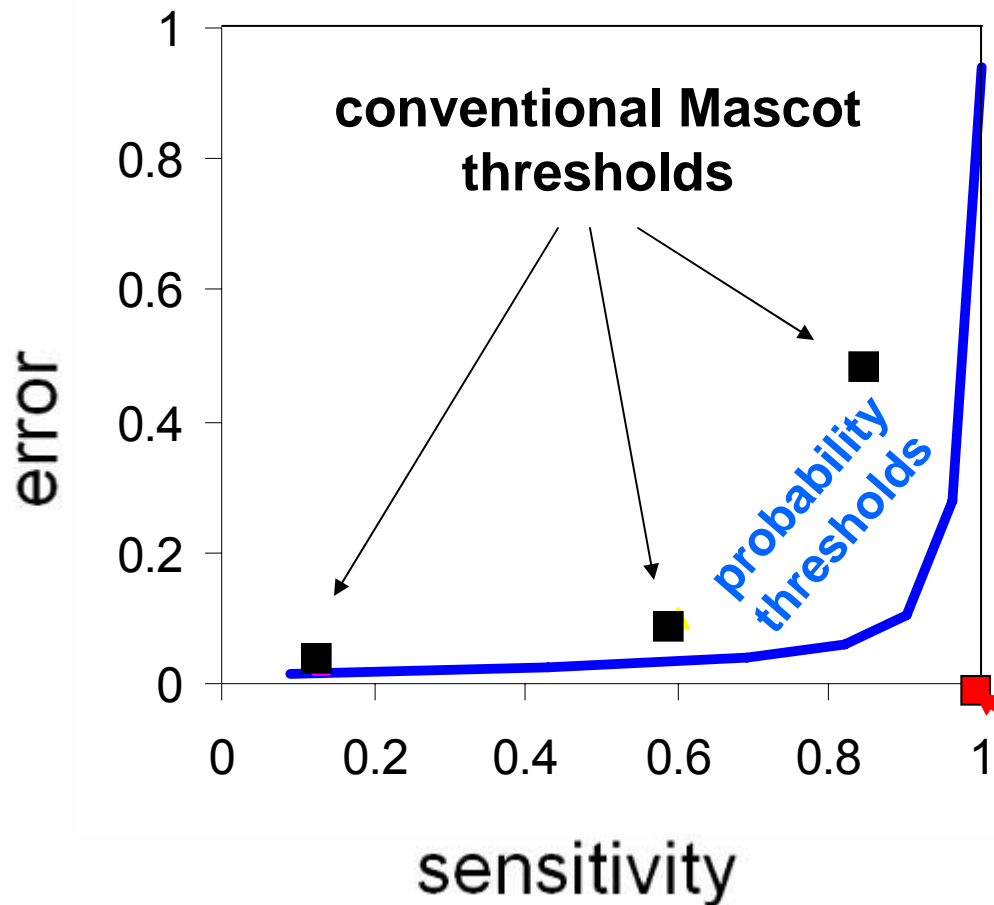
100 spectra with  
computed  $p \sim 0.9$

90% of them (90)  
should be correct

Observed  
probability is  
around 0.9

Model is accurate

# Discriminating Power of Computed Probabilities



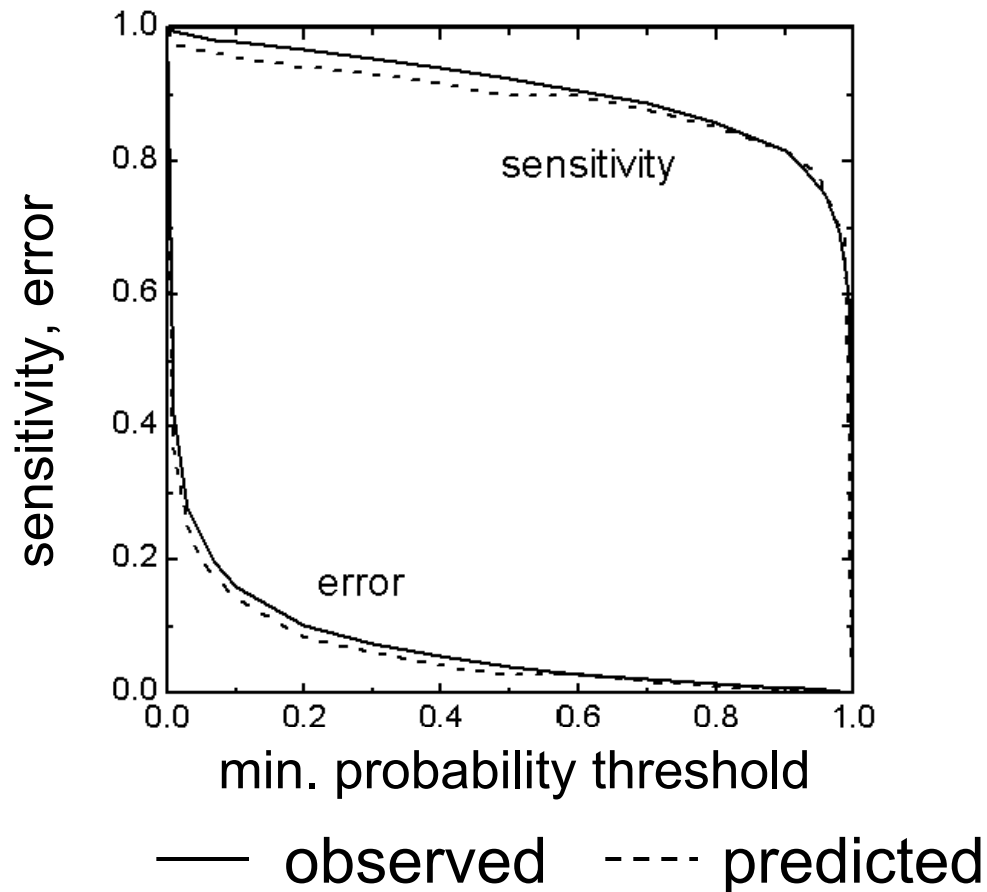
test data: A. Keller *et al.* OMICS 6, 207 (2002)

**Sensitivity:**  
fraction of all correct results passing filter

**Error:**  
fraction of all results passing filter that are incorrect

Ideal Spot

# Discriminating Power of Computed Probabilities



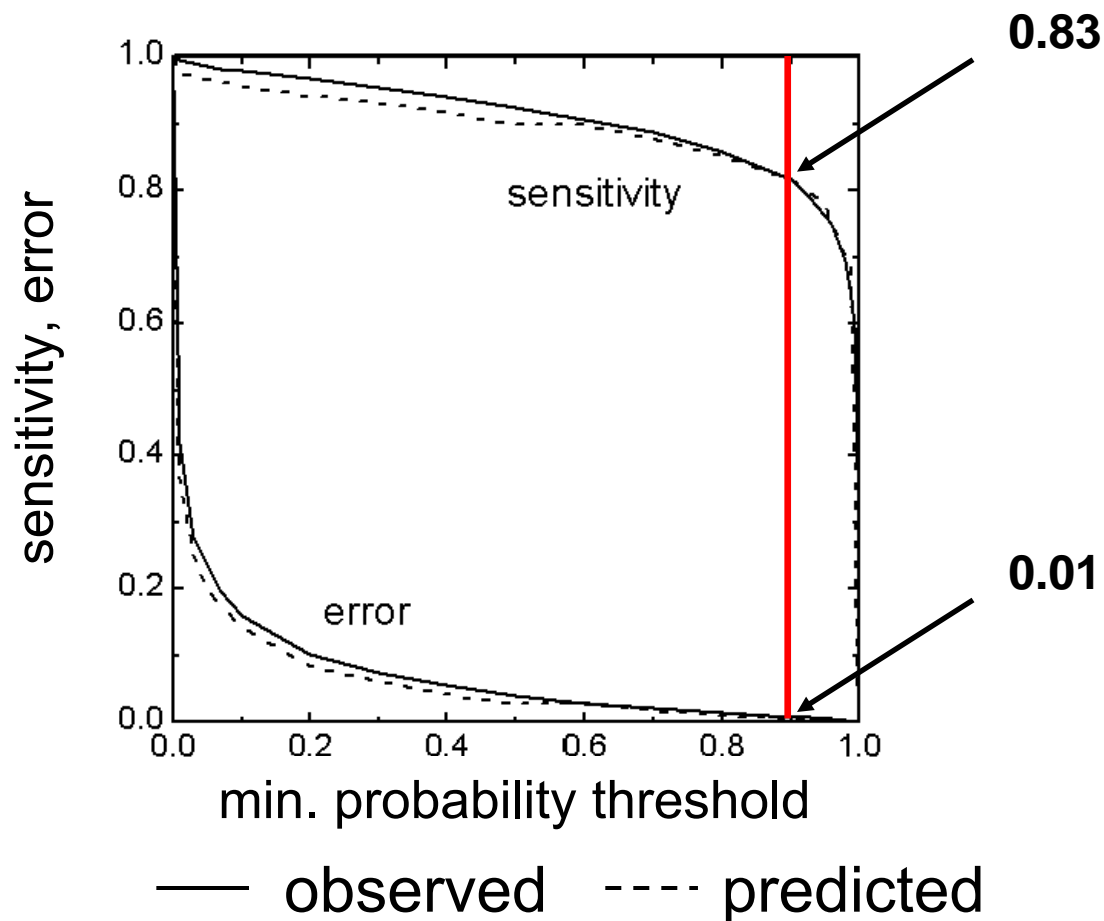
test data: A. Keller *et al.* OMICS 6, 207 (2002)

***Sensitivity:***  
fraction of all correct results passing filter

***Error:***  
fraction of all results passing filter that are incorrect

# Discriminating Power

## Example: $p \geq 0.9$



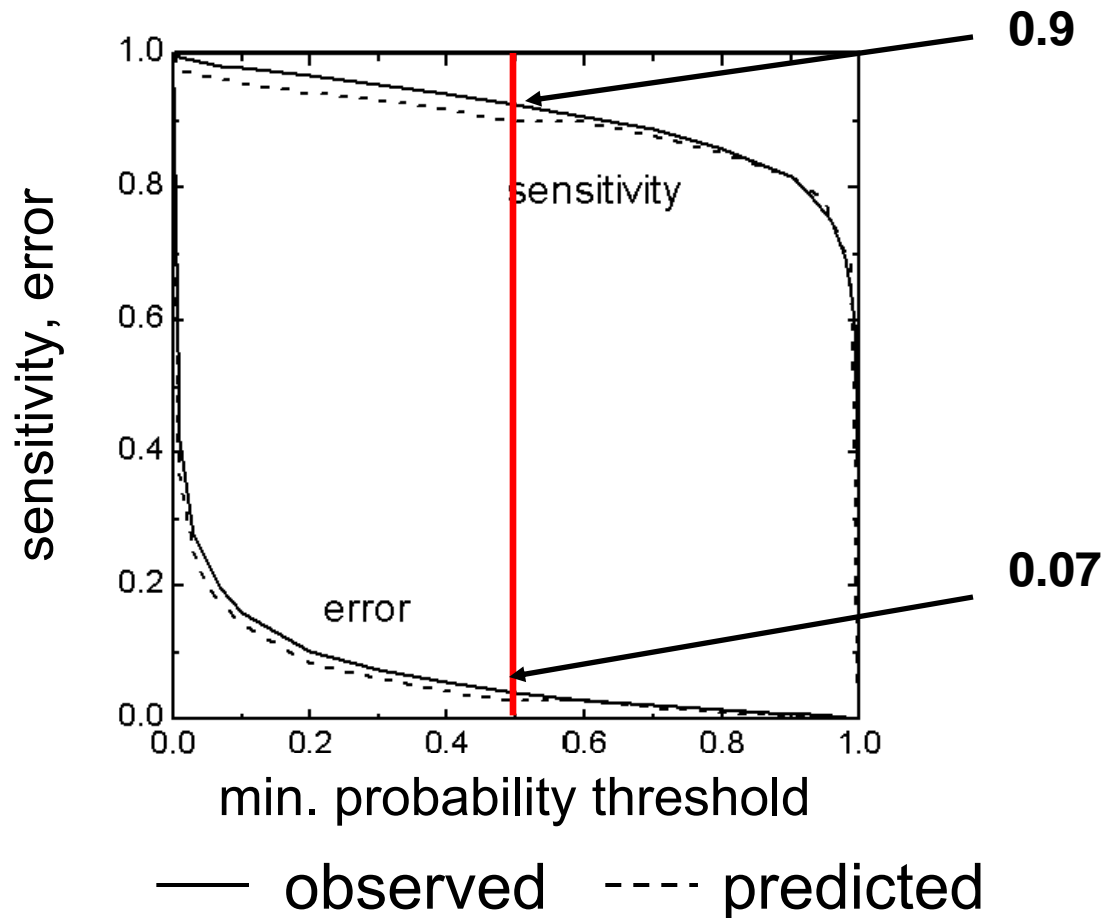
test data: A. Keller *et al.* OMICS 6, 207 (2002)

***Sensitivity:***  
fraction of all correct results passing filter

***Error:***  
fraction of all results passing filter that are incorrect

# Discriminating Power

## Example: $p \geq 0.5$



test data: A. Keller *et al.* OMICS 6, 207 (2002)

***Sensitivity:***  
fraction of all  
correct results  
passing filter

***Error:***  
fraction of all  
results passing  
filter that are  
incorrect



# Use of PeptideProphet Probabilities to Compare Searches

- False positive error rate predicted by PeptideProphet is an objective criterion for comparing different searches
  - Sample preparation and LC/MS/MS
  - Search conditions
  - Search engine
- Compare the number of results of each search passing its minimum probability threshold to achieve a fixed predicted false positive error rate
  - Reflects both search engine and PeptideProphet performance

# From Peptide to Protein Level Analysis

- When the identification of proteins rather than peptides is of interest, it is unnecessary in practice to filter search results based on probabilities
- Instead, all search results and their computed probabilities are passed to the **ProteinProphet** program which infers sample proteins by combining together the peptide evidence for each protein
  - Initially adjusts the PeptideProphet probabilities based on whether a peptide corresponds to a single-hit or multi-hit protein
  - Then apportions shared peptides among all their corresponding proteins in such a way to derive the simplest list of proteins that explain the observed peptides
  - Computes accurate protein probabilities

# PeptideProphet Software Tutorial

---

- How to run PeptideProphet through the TPP Web Interface
- Interpretation of analysis results
- User options

# Getting started with PeptideProphet

- Input: pepXML files (`file1.xml`, `file2.xml...`)
- XInteract program first merges files together into single file `interact.xml`, then PeptideProphet runs model, computes probabilities, and writes probabilities as first column
- Combine together runs that are similar (sample, database, search constraints, mass spectrometer)

# Getting started with PeptideProphet

*Specify search engine and select Analysis Pipeline*

SPCTools UI: ISB/SPC Proteomics Tools - TPP Web Interface - Microsoft Internet Explorer provided by Merck & Co., Inc.

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Print Mail New Tab

Address http: Go Links

## ISB/SPC Proteomics Tools - TPP Web Interface

You are logged in as **andy** [Log Out](#)

[Home](#) | [Account](#) | [Analysis Pipeline \(Mascot\)](#) | [Tools](#)

Home	ACCOUNT	<b>ANALYSIS PIPELINE</b>	TOOLS
------	---------	--------------------------	-------

### Welcome

Welcome to the Trans-Proteomic Pipeline (TPP) web interface. These tools and interfaces were developed at the [Institute for Systems Biology \(ISB\)](#) under a grant from [NHLEI](#). Please visit [www.proteomecenter.org](#) and [tools.proteomecenter.org](#) for more information.

Please select analysis pipeline you want to use:

### Analysis Pipeline

Follow these steps to convert, search, and analyze your data:

- 1. RAW to mzXML Conversion**  
Convert original .RAW files to the standard mzXML input format used by the tools
- 2. mzXML to mgf Conversion**  
Convert mzXML file to format that Mascot can search
- 3. Peptide Database Search and Identification**  
Upload file to Mascot server for searching.  
Transfer the .dat output file once the search is finished.
- 4. Conversion to pepXML**  
Convert original search results to the pepXML input format used by xinteract
- 5. Data Curation and (optional) Peptide validation and Quantification**

Done Internet

# Getting started with PeptideProphet

*Select peptide level analysis*

SPCTools UI: ISB/SPC Proteomics Tools - TPP Web Interface - Microsoft Internet Explorer provided by Merck & Co., Inc.

File Edit View Favorites Tools Help

Address <http:///> Go Links

## ISB/SPC Proteomics Tools - TPP Web Interface

You are logged in as **andy** [Log Out](#)

[Home](#) | [Account](#) | [Analysis Pipeline \(Mascot\)](#) | [Tools](#)

[Home](#) | [mzXML](#) | [Database Search](#) | [pepXML](#) | **Analyze Peptides** | [Analyze Proteins](#)

**1. Specify RAW Input File(s) to convert to mzXML** [Show / Hide]

No files selected yet.

[Add Files](#)

**2. Conversion Options** [Show / Hide]

Centroid  
 Profile

**3. Convert!**

No files selected yet.

Internet

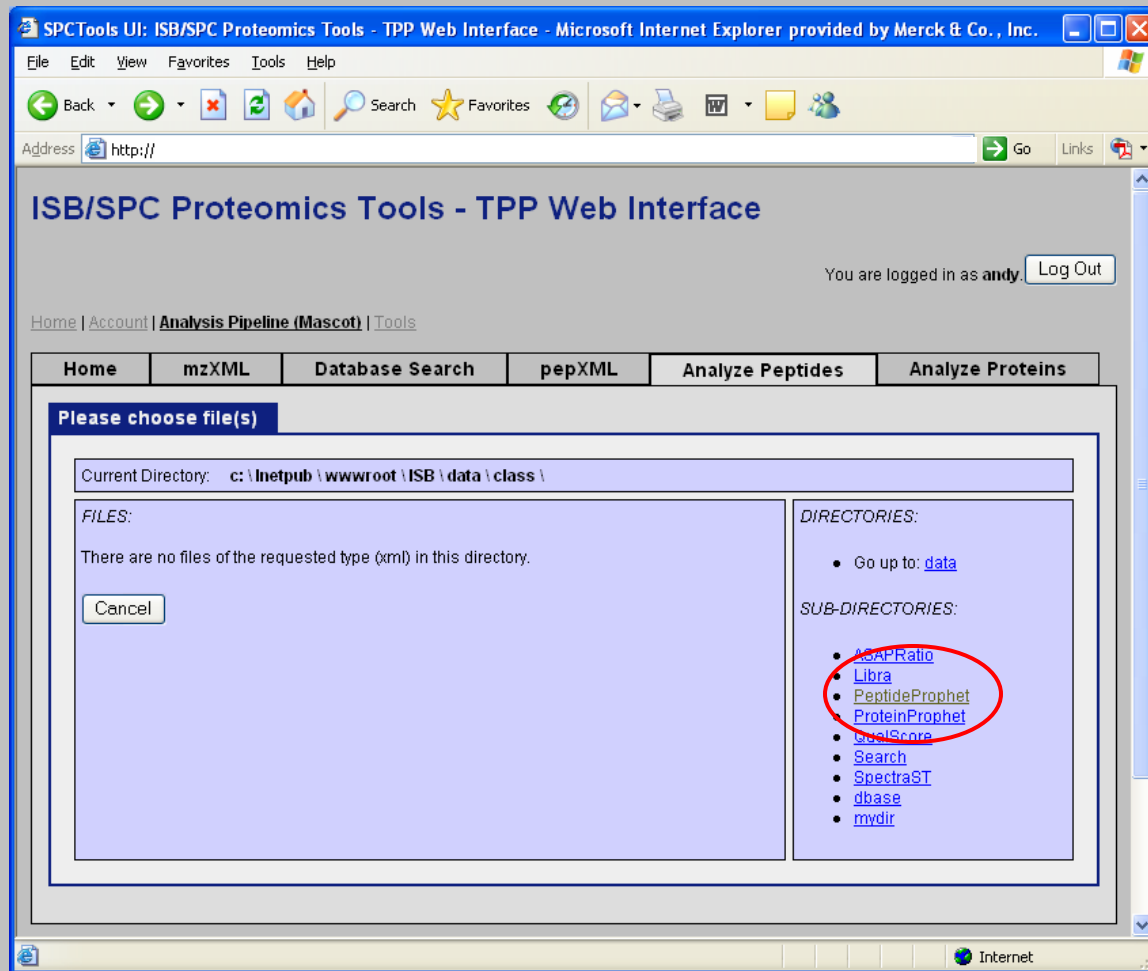
# Getting started with PeptideProphet

*Specify search results to analyze*

The screenshot displays the SPCTools UI web interface in a Microsoft Internet Explorer browser window. The browser title is "SPCTools UI: ISB/SPC Proteomics Tools - TPP Web Interface - Microsoft Internet Explorer provided by Merck & Co., Inc.". The address bar shows "http://". The page title is "ISB/SPC Proteomics Tools - TPP Web Interface". The user is logged in as "andy" and can click "Log Out". The navigation menu includes "Home", "Account", "Analysis Pipeline (Mascot)", and "Tools". The main content area has tabs for "Home", "mzXML", "Database Search", "pepXML", "Analyze Peptides", and "Analyze Proteins". The "Analyze Peptides" tab is active, showing a section titled "Select File(s) to Analyze" with a "[ Show / Hide ]" link. Inside this section, the text "No files selected yet." is circled in red, with an "Add Files" button below it. Below this section is the "Output File and Filter Options" section, which contains the text "Please select input files first." At the bottom is the "PeptideProphet Options" section with a "[ Show / Hide ]" link, containing two checkboxes: "RUN PeptideProphet" (checked) and "Use icat information" (unchecked). The browser status bar at the bottom shows "Internet".

# Getting started with PeptideProphet

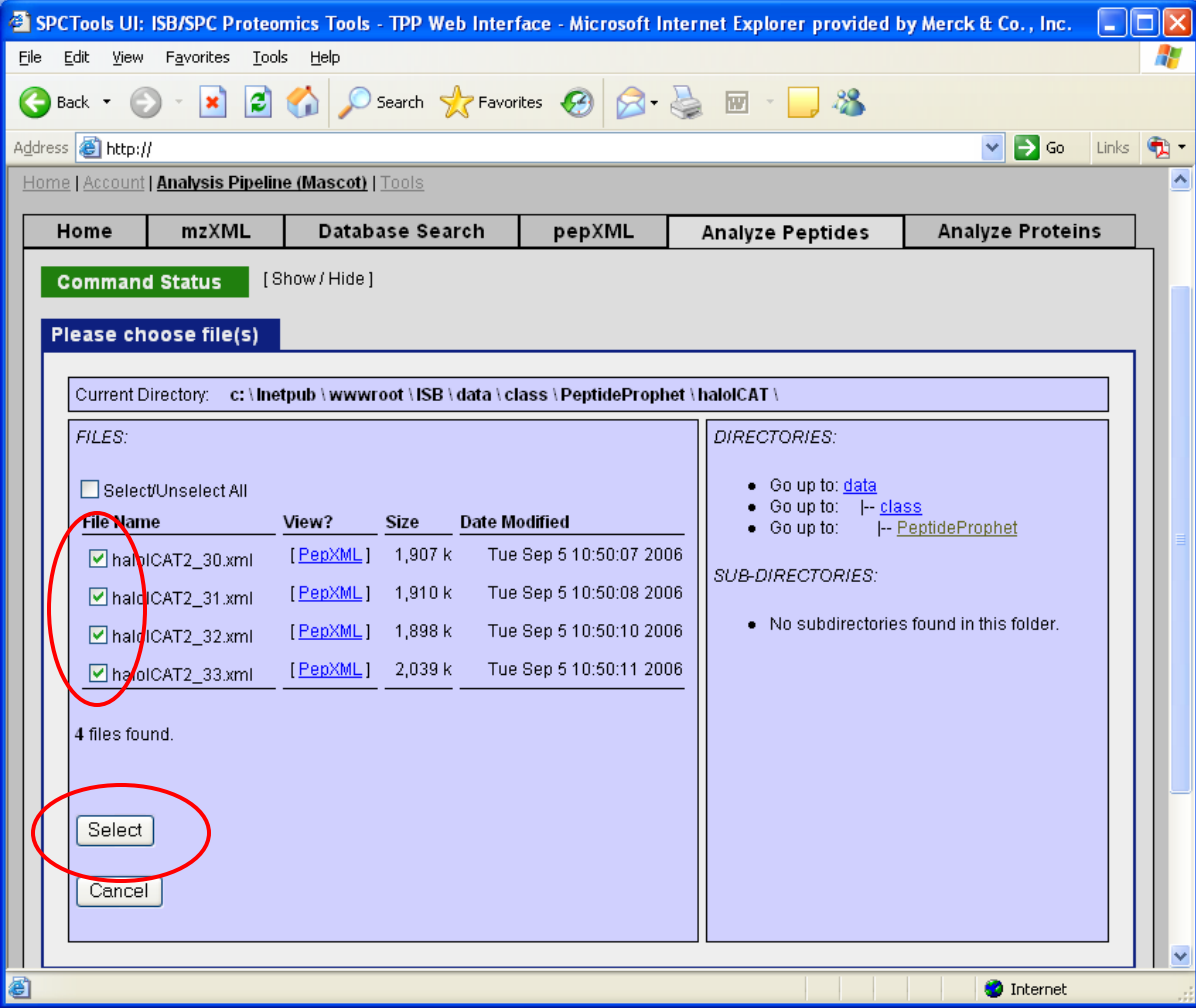
*Navigate data directories*





# Getting started with PeptideProphet

*Add each search run pepXML included in analysis*



The screenshot shows the SPCTools UI web interface in Microsoft Internet Explorer. The browser title is "SPCTools UI: ISB/SPC Proteomics Tools - TPP Web Interface - Microsoft Internet Explorer provided by Merck & Co., Inc.". The address bar shows "http://". The navigation menu includes "Home", "Account", "Analysis Pipeline (Mascot)", and "Tools". The main content area has tabs for "Home", "mzXML", "Database Search", "pepXML", "Analyze Peptides", and "Analyze Proteins". The "Command Status" section is expanded, showing a "Please choose file(s)" dialog. The current directory is "c:\inetpub\wwwroot\ISB\data\class\PeptideProphet\haloCAT\". The "FILES:" section contains a table with four files selected, each with a "[PepXML]" link. The "DIRECTORIES:" section shows navigation options for "data", "class", and "PeptideProphet". The "SUB-DIRECTORIES:" section indicates no subdirectories were found. The "Select" button is highlighted with a red circle.

Current Directory: c:\inetpub\wwwroot\ISB\data\class\PeptideProphet\haloCAT\

FILES:

File Name	View?	Size	Date Modified
<input checked="" type="checkbox"/> haloCAT2_30.xml	[PepXML]	1,907 k	Tue Sep 5 10:50:07 2006
<input checked="" type="checkbox"/> haloCAT2_31.xml	[PepXML]	1,910 k	Tue Sep 5 10:50:08 2006
<input checked="" type="checkbox"/> haloCAT2_32.xml	[PepXML]	1,898 k	Tue Sep 5 10:50:10 2006
<input checked="" type="checkbox"/> haloCAT2_33.xml	[PepXML]	2,039 k	Tue Sep 5 10:50:11 2006

4 files found.

DIRECTORIES:

- Go up to: [data](#)
- Go up to: [|-- class](#)
- Go up to: [|-- PeptideProphet](#)

SUB-DIRECTORIES:

- No subdirectories found in this folder.

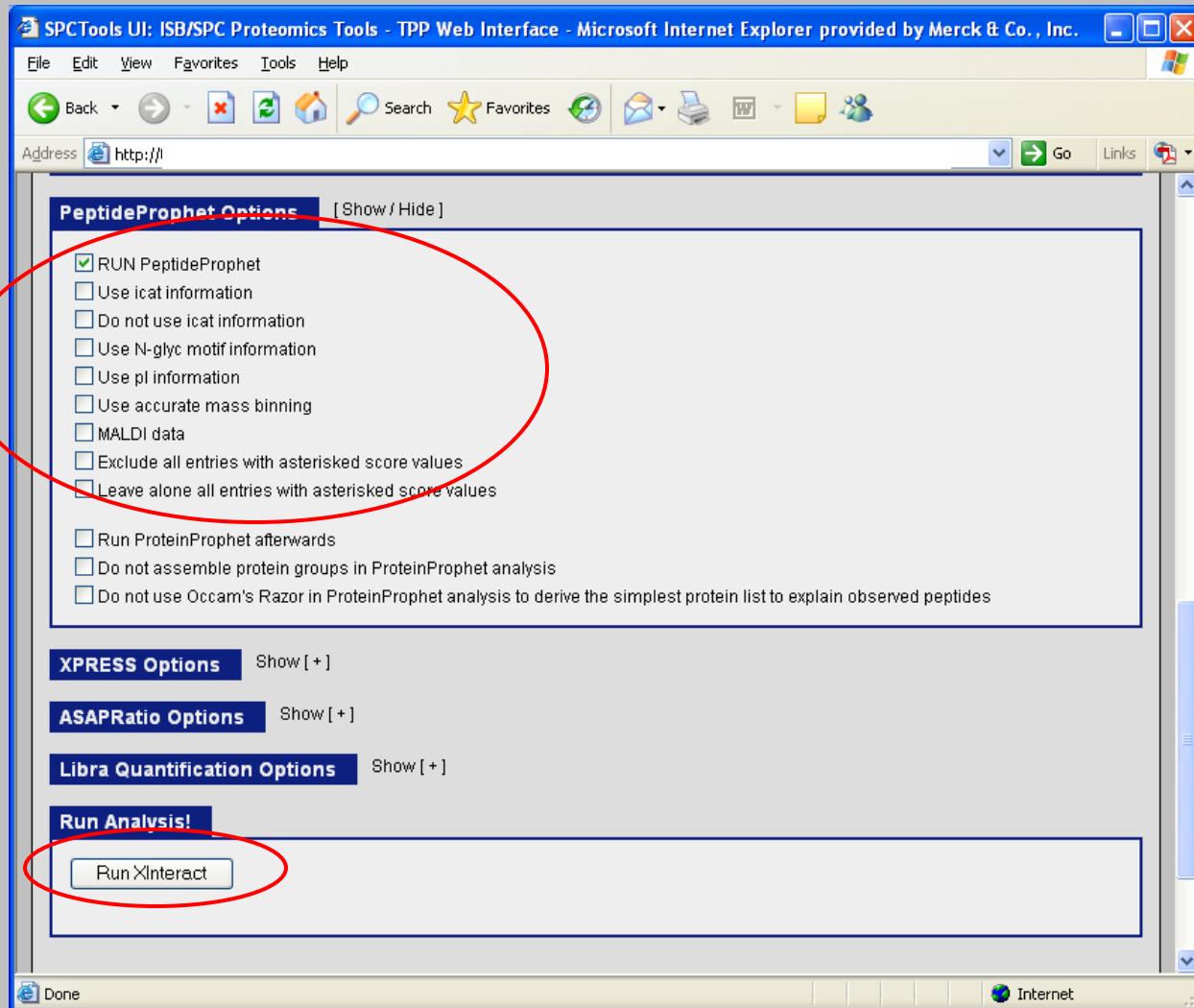
Select

Cancel



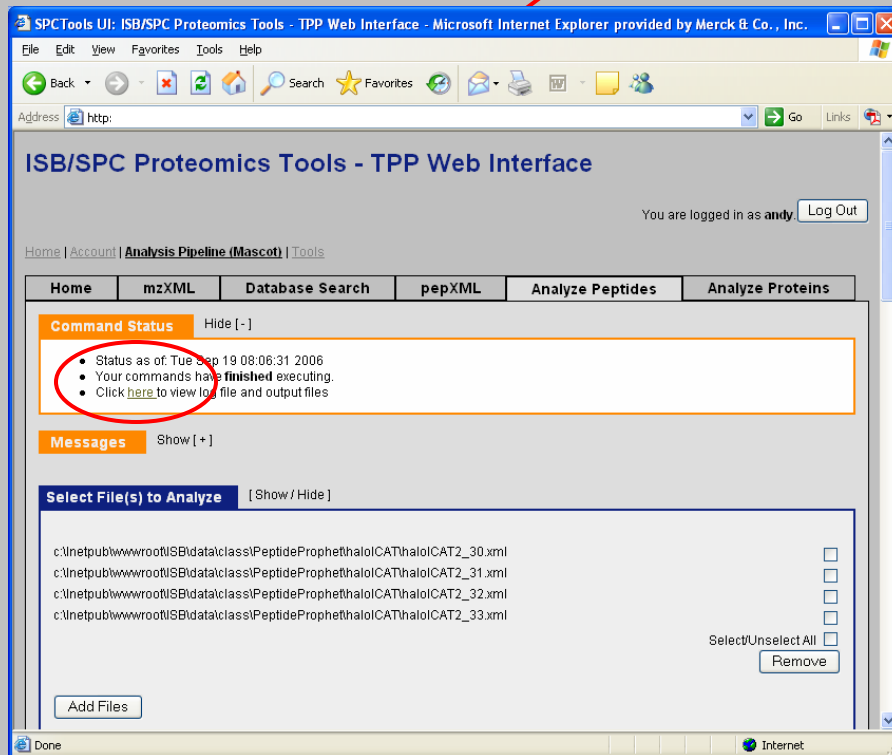
# Getting started with PeptideProphet

*Specify PeptideProphet optional parameters and run analysis*



# Getting started with PeptideProphet

*Click on links to view results of analysis*



SPCTools UI: ISB/SPC Proteomics Tools - TPP Web Interface - Microsoft Internet Explorer provided by Merck & Co., Inc.

Address: http:

### ISB/SPC Proteomics Tools - TPP Web Interface

You are logged in as **andy**. [Log Out](#)

Home | [Account](#) | [Analysis Pipeline \(Mascot\)](#) | [Tools](#)

Home | [mzXML](#) | [Database Search](#) | [pepXML](#) | [Analyze Peptides](#) | [Analyze Proteins](#)

**Command Status** [Hide \[-\]](#)

- Status as of: Tue Sep 19 08:06:31 2006
- Your commands have **finished** executing.
- Click [here](#) to view log file and output files

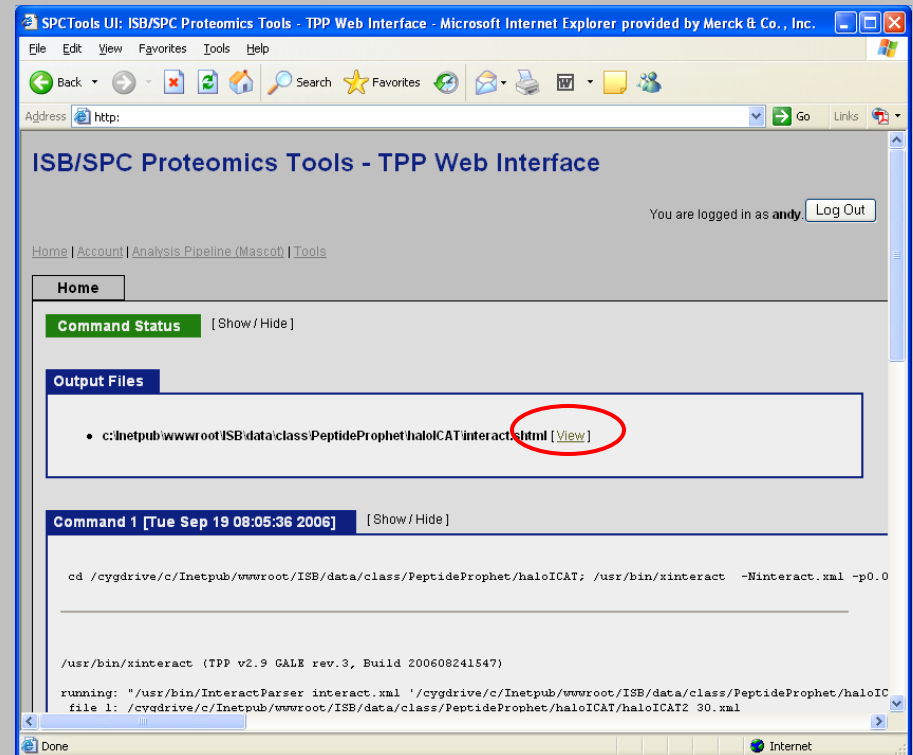
**Messages** [Show \[+\]](#)

**Select File(s) to Analyze** [\[ Show / Hide \]](#)

- c:\inetpub\wwwroot\ISB\data\class\PeptideProphet\haloICAT\haloICAT2\_30.xml
- c:\inetpub\wwwroot\ISB\data\class\PeptideProphet\haloICAT\haloICAT2\_31.xml
- c:\inetpub\wwwroot\ISB\data\class\PeptideProphet\haloICAT\haloICAT2\_32.xml
- c:\inetpub\wwwroot\ISB\data\class\PeptideProphet\haloICAT\haloICAT2\_33.xml

Select/Unselect All  [Remove](#)

[Add Files](#)



SPCTools UI: ISB/SPC Proteomics Tools - TPP Web Interface - Microsoft Internet Explorer provided by Merck & Co., Inc.

Address: http:

### ISB/SPC Proteomics Tools - TPP Web Interface

You are logged in as **andy**. [Log Out](#)

Home | [Account](#) | [Analysis Pipeline \(Mascot\)](#) | [Tools](#)

Home

**Command Status** [\[ Show / Hide \]](#)

**Output Files**

- c:\inetpub\wwwroot\ISB\data\class\PeptideProphet\haloICAT\interact.html [\[ View \]](#)

**Command 1 [Tue Sep 19 08:05:36 2006]** [\[ Show / Hide \]](#)

```
cd /cygdrive/c/inetpub/wwwroot/ISB/data/class/PeptideProphet/haloICAT; /usr/bin/xinteract -Ninteract.xml -p0.0

/usr/bin/xinteract (TPP v2.9 GALE rev.3, Build 200608241547)

running: "/usr/bin/InteractParser interact.xml '/cygdrive/c/inetpub/wwwroot/ISB/data/class/PeptideProphet/haloICAT
file 1: /cygdrive/c/inetpub/wwwroot/ISB/data/class/PeptideProphet/haloICAT/haloICAT2_30.xml
```

# PeptideProphet Results

PepXML Viewer: /cygdrive/c/Inetpub/wwwroot/ISB/data/class/PeptideProphet/haloCAT/Semirypic/i - Microsoft Inter...

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Mail Print W Links

Address http://

UPDATE EXPORT SPREADSHEET PEP 3D ADDITIONAL ANALYSIS INFO HELP restore original

**Display Options** [Show | Hide]

**Column Order and Selection** [Show | Hide]

**Filtering Options** [Show | Hide]

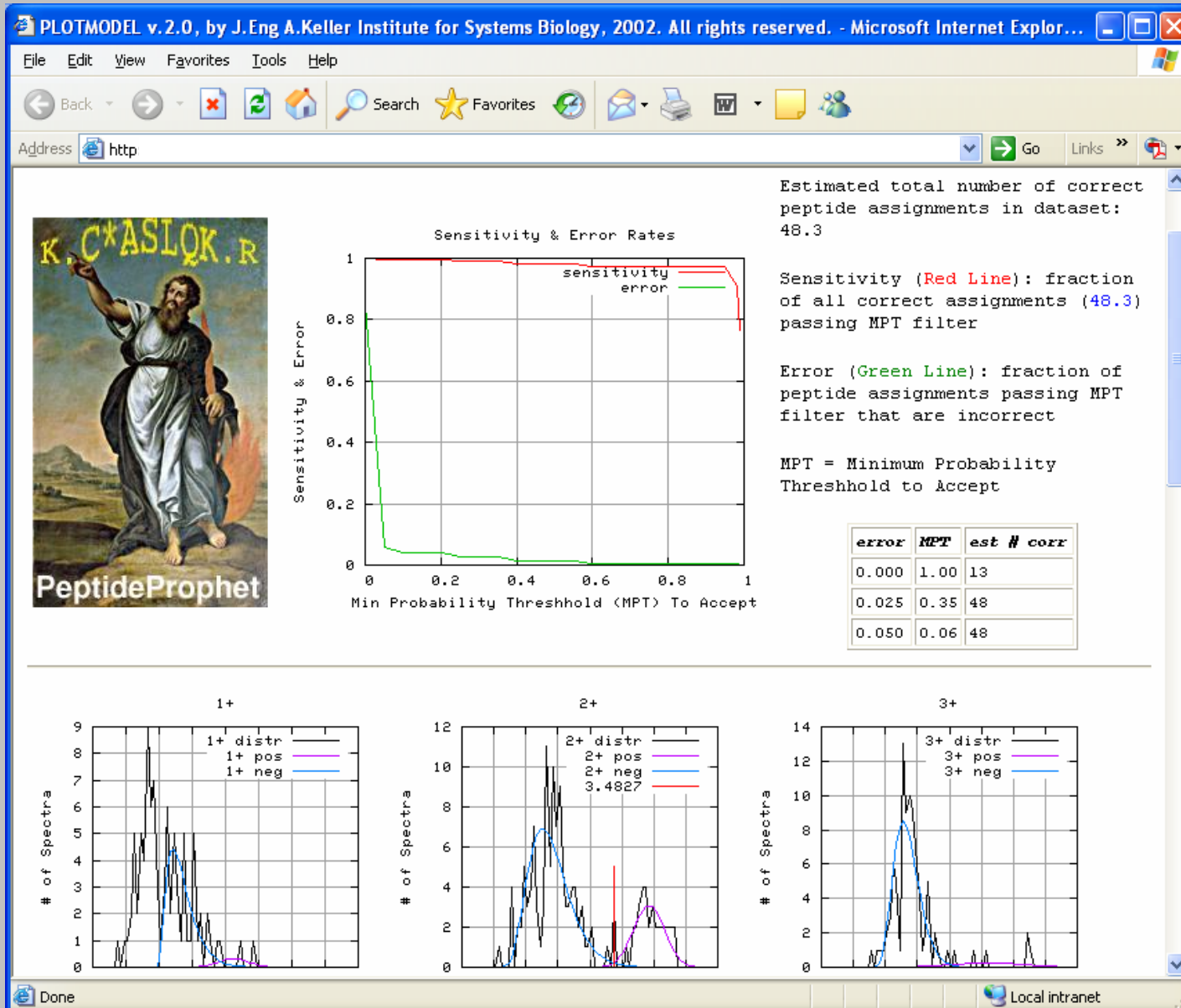
Page 1 of 1

jump to FIRST 1 LAST

INDEX	PROBABILITY	SPECTRUM	IONSORE	IDENTITYSCORE	HOMOLOGYSORE	IONS	PEP
247	1.0000	<a href="#">haloCAT2_32.1358.1358.2</a>	51.43	51.90	28.47	8/ 26	
116	1.0000	<a href="#">haloCAT2_31.1068.1068.1</a>	32.87	52.16	28.75	8/ 22	
114	1.0000	<a href="#">haloCAT2_31.1072.1072.2</a>	55.86	52.06	41.24	10/ 22	
253	1.0000	<a href="#">haloCAT2_32.1778.1778.2</a>	70.26	51.87	33.12	8/ 26	
112	1.0000	<a href="#">haloCAT2_31.1070.1070.1</a>	41.14	52.06	34.57	6/ 22	
259	1.0000	<a href="#">haloCAT2_32.1546.1546.2</a>	48.85	51.68	31.17	8/ 28	
262	1.0000	<a href="#">haloCAT2_32.1530.1530.2</a>	62.93	51.55	28.19	9/ 28	
269	1.0000	<a href="#">haloCAT2_32.1414.1414.2</a>	66.01	51.43	27.49	9/ 32	
98	1.0000	<a href="#">haloCAT2_30.1901.1901.2</a>	38.43	50.09	26.19	17/ 52	
97	1.0000	<a href="#">haloCAT2_30.1905.1905.2</a>	44.19	50.16	22.12	14/ 52	
96	1.0000	<a href="#">haloCAT2_30.1211.1211.2</a>	48.30	50.74	28.06	16/ 46	

Internet

# PeptideProphet Results: Model Summary



PepXML Viewer: /cygdrive/k

File Edit View Favorites Tool

Address http://

UPDATE EXPORT SPE

Display Options

Column Order and Se

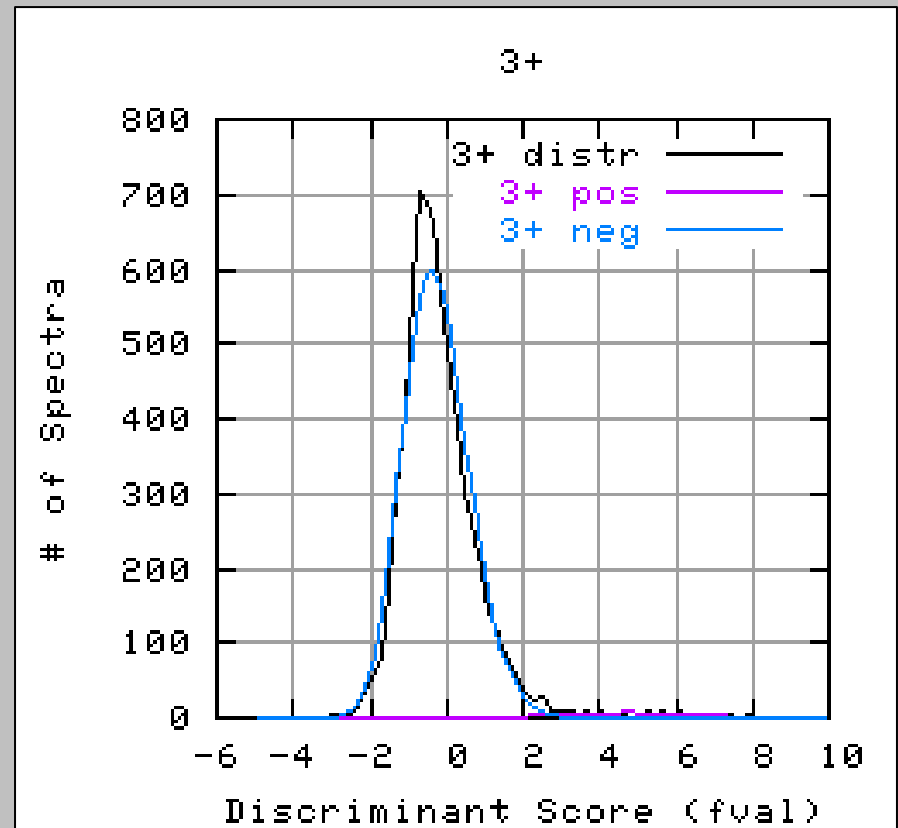
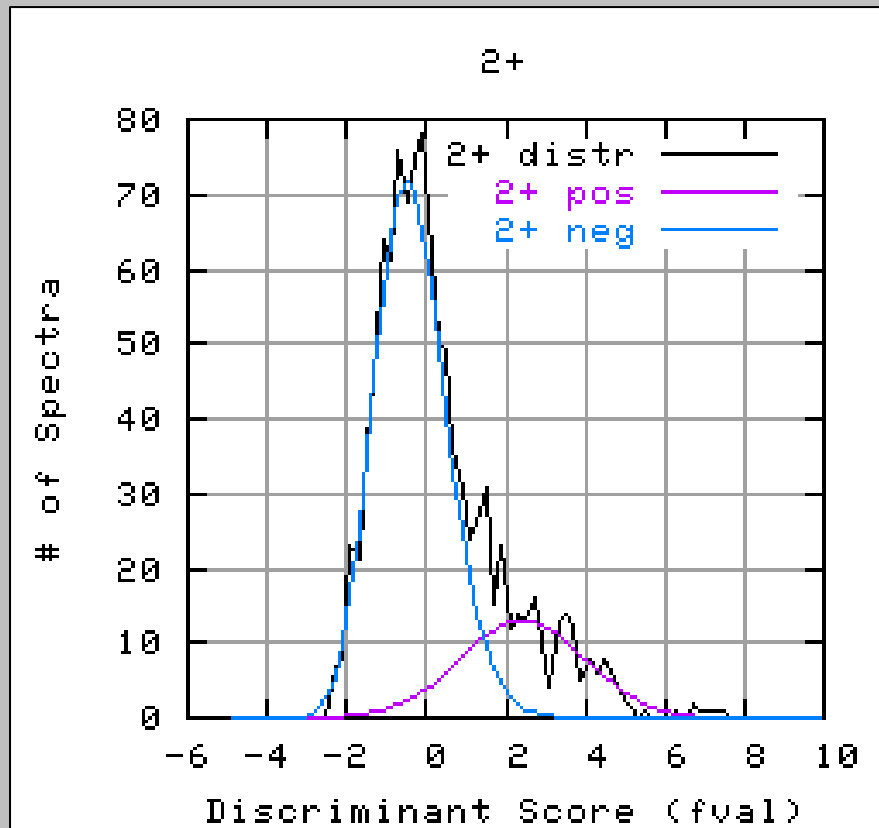
Filtering Options

Page 1 of 1

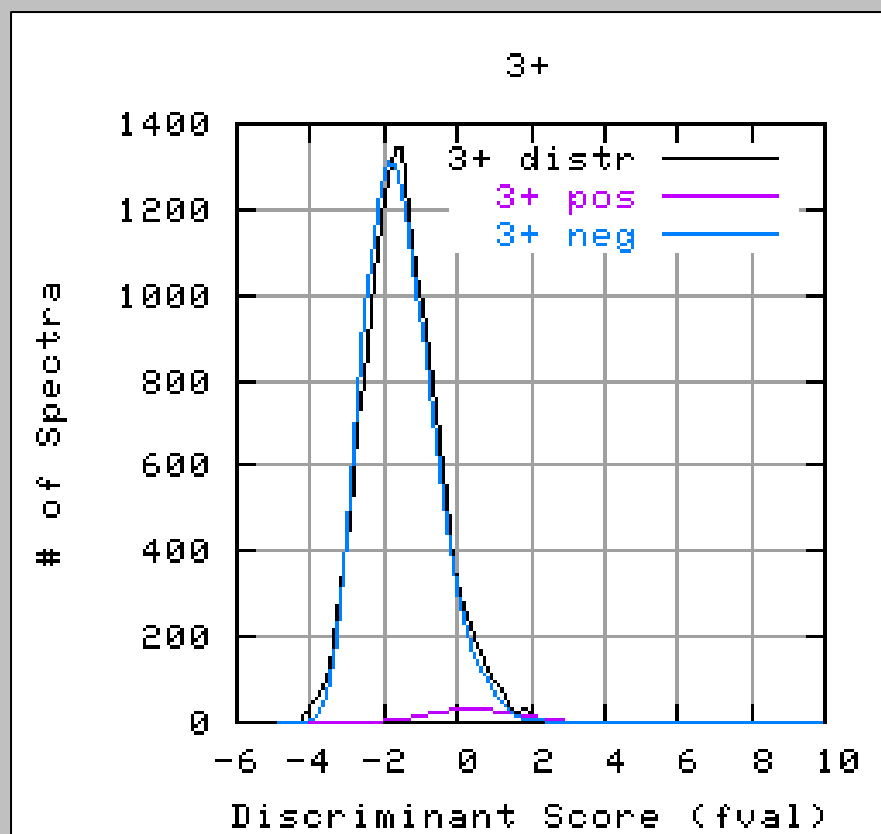
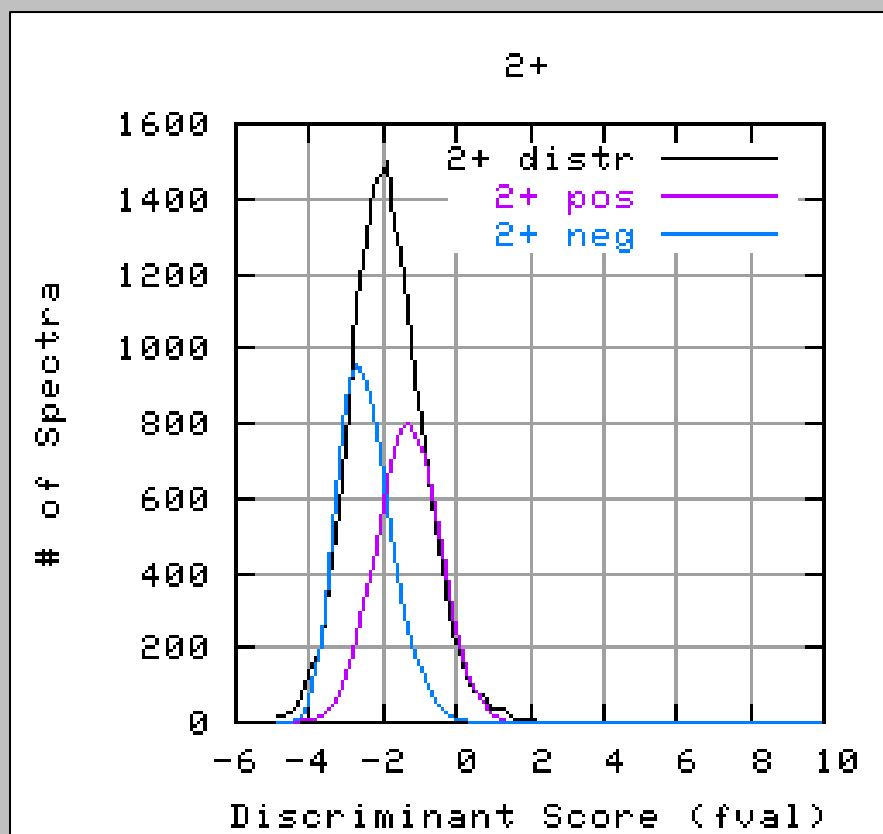
jump to FIRST 1 LAST

INDEX	PROBABILITY
247	<a href="#">1.0000</a>
116	<a href="#">1.0000</a>
114	<a href="#">1.0000</a>
253	<a href="#">1.0000</a>
112	<a href="#">1.0000</a>
259	<a href="#">1.0000</a>
262	<a href="#">1.0000</a>
269	<a href="#">1.0000</a>
98	<a href="#">1.0000</a>
97	<a href="#">1.0000</a>
96	<a href="#">1.0000</a>

# Reasonable Learned Discriminant Score Distributions

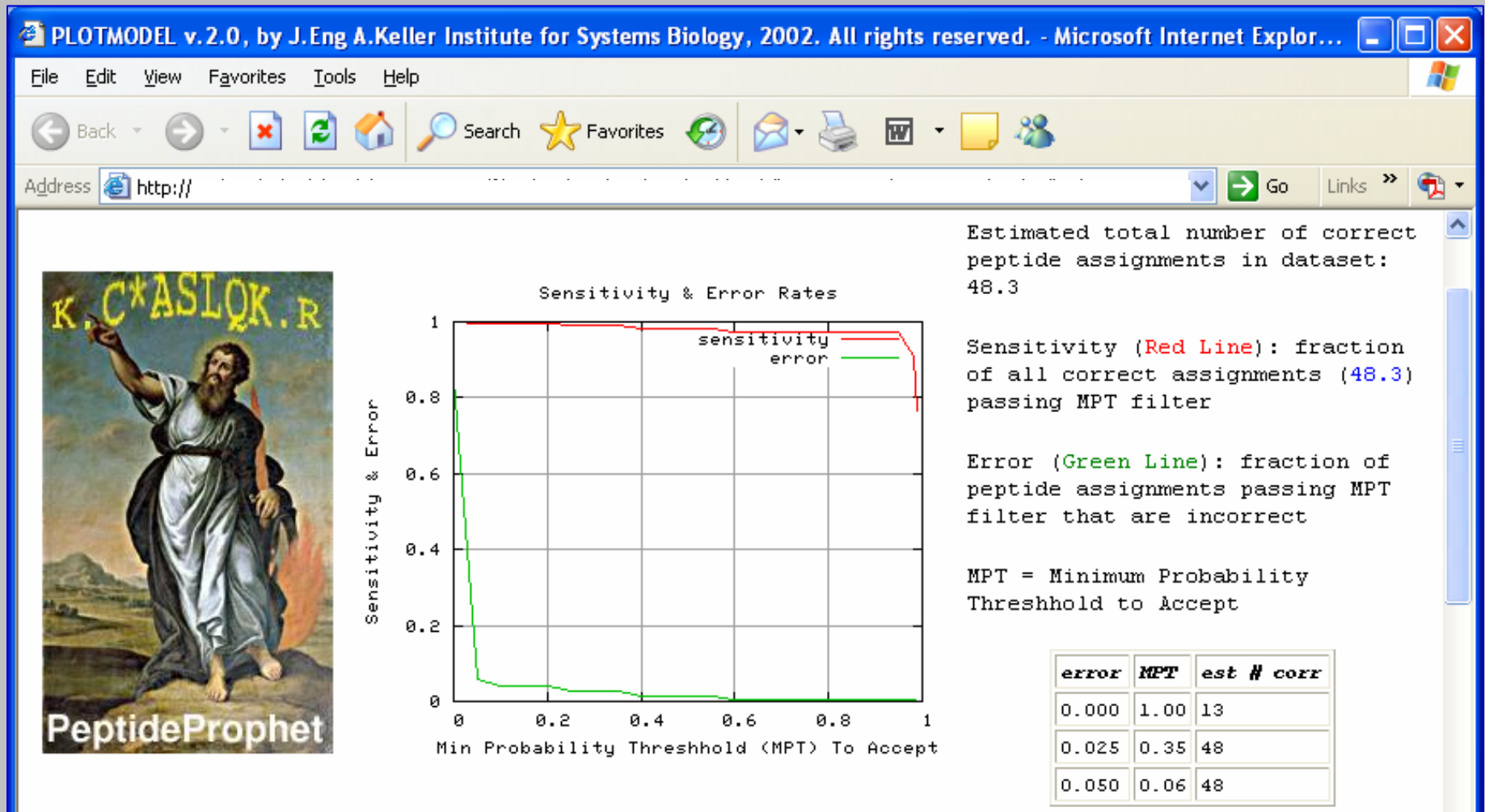


# Suspicious Looking Learned Discriminant Score Distributions

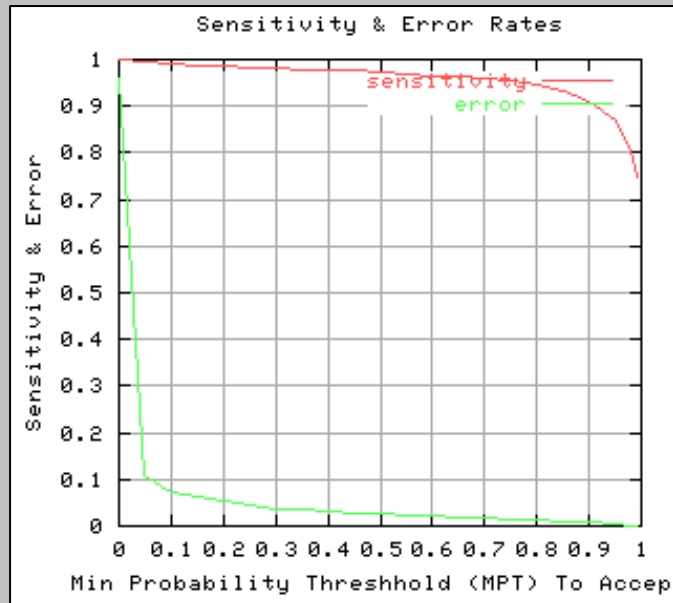




# PeptideProphet Results: Model Summary

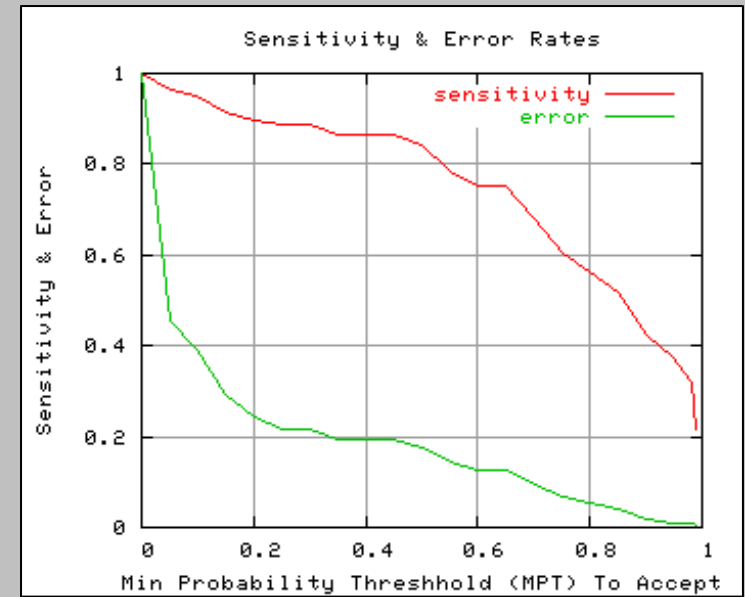


# PeptideProphet Results: Model Summary



Good

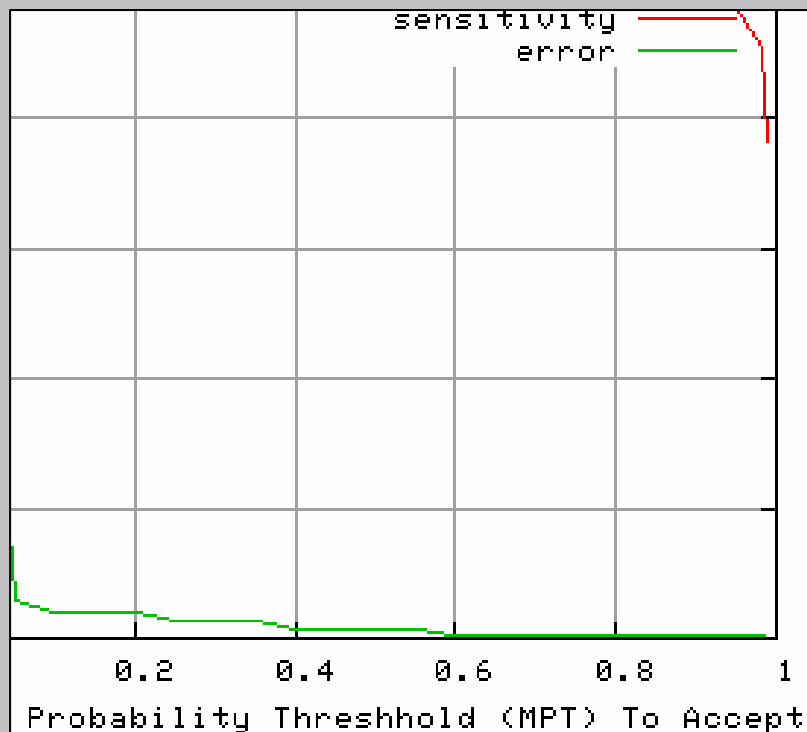
Not so good



# PeptideProphet Results: Model Summary

```
FINAL 2+ MODEL after 6 iterations;
number of spectra: 18495
using no. tolerable trypsin term. [ntt] 0 data as pseudonegatives
  prior: 0.087, est. total correct: 1612.2
MASCOT discrim score [fval]   slope: 1.04 intercept: -22.88
  regression error: 6.64 negmean: -1.26
  pos: (gaussian mean 1.43, stdev 1.52)
  neg: (evd mean -1.29, stdev 0.52, mu -1.52, beta 0.41)
no. tolerable trypsin term. [ntt]
  pos: (ntt=0 0.093, ntt=1 0.278, ntt=2 0.628)
  neg: (ntt=0 0.796, ntt=1 0.190, ntt=2 0.015)
no. missed enz. cleavages [nmc]
  pos: (nmc=0 0.949, 1<=nmc<=2 0.051, nmc>=3 0.000)
  neg: (nmc=0 0.385, 1<=nmc<=2 0.537, nmc>=3 0.077)
var offset mass diff [massd] (offset: 0.70)
  pos: (massd=-4.0 0.00, massd=-3.0 0.00, massd=-2.0 0.01, massd=-1.0 0.04,
        massd=0.0 0.80, massd=1.0 0.14, massd=2.0 0.02)
  neg: (massd=-4.0 0.03, massd=-3.0 0.15, massd=-2.0 0.16, massd=-1.0 0.17,
        massd=0.0 0.18, massd=1.0, 0.17, massd=2.0 0.13)
icat cys [icat]
  pos: (0.022 icat=0 (incompatible), 0.978 icat=1 (compatible))
  neg: (0.927 icat=0 (incompatible), 0.073 icat=1 (compatible))
```

# PeptideProphet Results: Predicted Numbers of Correct Peptides



Sensitivity (Red Line): fraction of all correct assignments (48.3) passing MPT filter

Error (Green Line): fraction of peptide assignments passing MPT filter that are incorrect

MPT = Minimum Probability Threshold to Accept

<i>error</i>	<i>MPT</i>	<i>est # corr</i>
0.000	1.00	13
0.025	0.35	48
0.050	0.06	48



# PeptideProphet $[M+2H]^{2+}$ vs $[M+3H]^{3+}$ Precursor Ions

334	<a href="#">1.0000</a>	<a href="#">haloCAT2_33.1062.1062.3</a>	51.06	51.94	36.29	<a href="#">10/44</a>	
338	<a href="#">0.9984</a>	<a href="#">haloCAT2_33.1042.1042.2</a>	33.62	51.98	33.58	<a href="#">6/22</a>	
357	<a href="#">0.9996</a>	<a href="#">haloCAT2_33.1034.1034.2</a>	27.70	51.85	26.77	<a href="#">7/18</a>	
331	<a href="#">0.9596</a>	<a href="#">haloCAT2_33.1024.1024.3</a>	24.96	52.07	32.23	<a href="#">12/44</a>	
386	<a href="#">0.4275</a>	<a href="#">haloCAT2_33.1014.1014.3</a>	17.19	49.67	27.34	<a href="#">6/96</a>	
372	<a href="#">0.5725</a>	<a href="#">haloCAT2_33.1014.1014.2</a>	41.36	51.37	27.11	<a href="#">15/36</a>	
373	<a href="#">0.9992</a>	<a href="#">haloCAT2_33.1012.1012.2</a>	34.49	51.39	35.29	<a href="#">11/36</a>	
312	<a href="#">0.6340</a>	<a href="#">haloCAT2_33.1004.1004.1</a>	27.62	52.49	33.39	<a href="#">7/22</a>	
330	<a href="#">0.9992</a>	<a href="#">haloCAT2_33.1002.1002.2</a>	26.84	52.06	29.37	<a href="#">8/22</a>	

Spectrum searched as both 2+ and 3+ precursor received significant probability

# PeptideProphet Results: Incomplete Analysis

The screenshot shows a web browser window titled "PepXML Viewer: /cygdrive/c:/inetpub/wwwroot/ISB/data/class/PeptideProphet/haloCAT/Semityptic/i - Microsoft Internet Explorer p". The browser address bar shows "http://". The main content is a table with the following columns: Index, Charge, Peptide Link, Score, and other numerical values. A red circle highlights the "Charge" column, which contains values of -1 and 0. The table data is as follows:

Index	Charge	Peptide Link	Score	Other Values
418	0.0697	<a href="#">haloCAT2_33.1274.1274.2</a>	14.70	51.90, 27.13, 5/24
384	0.0685	<a href="#">haloCAT2_33.0884.0884.3</a>	29.73	52.31, 38.33, 7/68
310	0.0677	<a href="#">haloCAT2_32.2534.2534.2</a>	10.85	51.48, 21.21, 7/38
339	0.0677	<a href="#">haloCAT2_32.1812.1812.2</a>	11.78	51.04, 20.39, 6/28
398	0.0660	<a href="#">haloCAT2_33.2382.2382.2</a>	11.60	52.01, 24.54, 9/34
461	0.0644	<a href="#">haloCAT2_33.2244.2244.3</a>	27.15	48.10, 39.66, 10/128, L. PYQVSLNSGSH
7	0.0644	<a href="#">haloCAT2_30.1851.1851.2</a>	23.05	52.42, 25.92, 9/24
441	0.0642	<a href="#">haloCAT2_33.2088.2088.2</a>	10.53	51.03, 21.11, 10/48
9	0.0634	<a href="#">haloCAT2_30.0895.0895.2</a>	24.31	52.38, 29.97, 4/24
163	0.0626	<a href="#">haloCAT2_31.1266.1266.2</a>	17.15	51.66, 30.10, 6/34
244	0.0626	<a href="#">haloCAT2_32.1656.1656.2</a>	11.73	52.64, 23.75, 5/22
317	0.0612	<a href="#">haloCAT2_32.2406.2406.3</a>	16.30	51.40, 25.23, 19/92
101	0.0591	<a href="#">haloCAT2_30.2307.2307.3</a>	15.64	50.85, 24.50, 6/108
430	-1	<a href="#">haloCAT2_33.0784.0784.1</a>	16.65	51.62, 28.23, 13/76
100	0	<a href="#">haloCAT2_32.2216.2216.1</a>	11.13	50.85, 23.25, 21/52
364	0	<a href="#">haloCAT2_33.0790.0790.1</a>	12.49	52.31, 23.89, 7/22
268	-1	<a href="#">haloCAT2_33.0116.0116.1</a>	12.36	52.19, 23.59, 4/30
272	0	<a href="#">haloCAT2_31.1348.1348.1</a>	12.27	52.15, 21.68, 4/26
70	0	<a href="#">haloCAT2_32.1794.1794.1</a>	13.89	51.49, 17.63, 10/34
394	0	<a href="#">haloCAT2_32.1892.1892.1</a>	17.13	52.06, 28.69, 12/60
437	0	<a href="#">haloCAT2_32.1052.1052.1</a>	32.60	51.42, 35.20, 6/88
11	0	<a href="#">haloCAT2_31.1314.1314.1</a>	25.08	52.31, 32.39, 8/52

Model incomplete for results of 1+ precursor ions

# PeptideProphet Results: Incomplete Analysis

Charge	Probability	Peptide
418	0.0697	haloCAT2_33
384	0.0685	haloCAT2_33
310	0.0677	haloCAT2_32
339	0.0677	haloCAT2_32
398	0.0660	haloCAT2_33
461	0.0644	haloCAT2_33
7	0.0644	haloCAT2_30
441	0.0642	haloCAT2_33
9	0.0634	haloCAT2_30
163	0.0626	haloCAT2_31
244	0.0626	haloCAT2_32
317	0.0612	haloCAT2_32
101	0.0594	haloCAT2_30
430	-1	haloCAT2_33
100	0	haloCAT2_32
364	0	haloCAT2_33
268	-1	haloCAT2_33
272	0	haloCAT2_31
70	0	haloCAT2_32
394	0	haloCAT2_32
437	0	haloCAT2_32
11	0	haloCAT2_31

In general, if analysis of results of precursor ion charge  $N$  is incomplete, results are partitioned into those **unlikely to be correct** (assigned probability '0'), and those **possibly correct** (assigned probability ' $-N$ '). These estimates are made using learned distributions for an adjacent charge when available, otherwise using training dataset distributions

Model incomplete for results of 1+ precursor ions



# Sort Data by Computed Probability

jump to new file:  
/cygdrive/c/Inetpub/wwwroot/ISB/data/class/PeptideProphet/Myster

**trypsin** digest, **MASCOT** search engine, quantitation: [none]  
displaying **2045** of 2045 total spectra  
viewing page 1 of 82  
**2020 unique peptides, 2018 unique stripped peptides**  
**1873 unique proteins, 1734 single hits**

[UPDATE](#) [EXPORT SPREADSHEET](#) [PEP 3D](#) [ADDITIONAL ANALYSIS INFO](#) [HELP](#) [restore original](#)

**Display Options** [\[ Show | Hide \]](#)

currently displaying: 25 rows per page, sorting by index (ascending), highlighted peptide text: [none], highlighted protein text: (Chr\_ORF)(pNRC), multiple protein hits as top hit .

- rows per page:  rows per page
- Sorting:  descending  ascending
- highlight peptide text (regex):
- highlight protein text (regex):
- multiple protein hits: top hit only  list of all hits
- column headers: regular  condensed

**Column Order and Selection** [\[ Show | Hide \]](#)

**Filtering Options** [\[ Show | Hide \]](#)

Internet

# Filter Data by Mascot Ionscore

The screenshot shows a web browser window titled "PepXML Viewer: /cygdrive/c:/inetpub/wwwroot/ISB/data/class/PeptideProphet/haloICAT/interact.xml - Microsoft Internet ...". The browser's address bar is empty, and the page content includes navigation links like "UPDATE", "EXPORT SPREADSHEET", "PEP 3D", "ADDITIONAL ANALYSIS INFO", "HELP", and "restore original".

The interface features three main sections, each with a "Show | Hide" link:

- Display Options** [Show | Hide]
- Column Order and Selection** [Show | Hide]
- Filtering Options** [Show | Hide]

The "Filtering Options" section is highlighted with a red oval and contains the following controls:

- require peptide aa's:
- require glycosylation motif (NxS/T) in peptides:
- required peptide text (regex allowed):
- exclude charges: 1+  2+  3+  Others
- Search results (for MASCOT):
  - min ionScore:
  - require ionscore > identityscore
  - require ionscore > homologyscore
- Peptide Prophet:
  - min probability:
  - max probability:

At the bottom of the page, a green bar displays "Page 1 of 16". The browser's status bar at the bottom right shows "Internet".

# Select and Color Specified AA's

The screenshot shows the PepXML Viewer web application interface. The browser window title is "PepXML Viewer: /cygdrive/c/inetpub/wwwroot/ISB/data/class/PeptideProphet/Mystery/interact\_new.x - Microsoft Inte...". The address bar shows "http:". The interface is divided into three main sections, each with a "Show | Hide" toggle button circled in red:

- Display Options**:
  - currently displaying: 25 rows per page, sorting by index (ascending), highlighted peptide text: [none], highlighted protein text: (Chr\_ORF)|(pNRC), multiple protein hits as top hit .
  - rows per page: all (dropdown) rows per page
  - Sorting: probability (dropdown) descending (radio) ascending (radio)
  - highlight peptide text (regex): C (input field)
  - highlight protein text (regex): (input field)
  - multiple protein hits: top hit only (radio) list of all hits (radio)
  - column headers: regular (radio) condensed (radio)
- Column Order and Selection**: [Show | Hide]
- Filtering Options**: [Show | Hide]
  - require peptide aa's: (input field)
  - require glycosylation motif (NxS/T) in peptides: (checkbox)
  - required peptide text (regex allowed): C (input field)
  - exclude charges: 1+ (checkbox) 2+ (checkbox) 3+ (checkbox) Others (checkbox)
  - Search results (for MASCOT): min ionScore: (input field) require ionscore > (checkbox)

# Pep3D and Analysis Summary Links

170 unique proteins, 105 single hits

[UPDATE](#) [EXPORT SPREADSHEET](#) [PEP 3D](#) [ADDITIONAL ANALYSIS INFO](#) [HELP](#) [restore original](#)

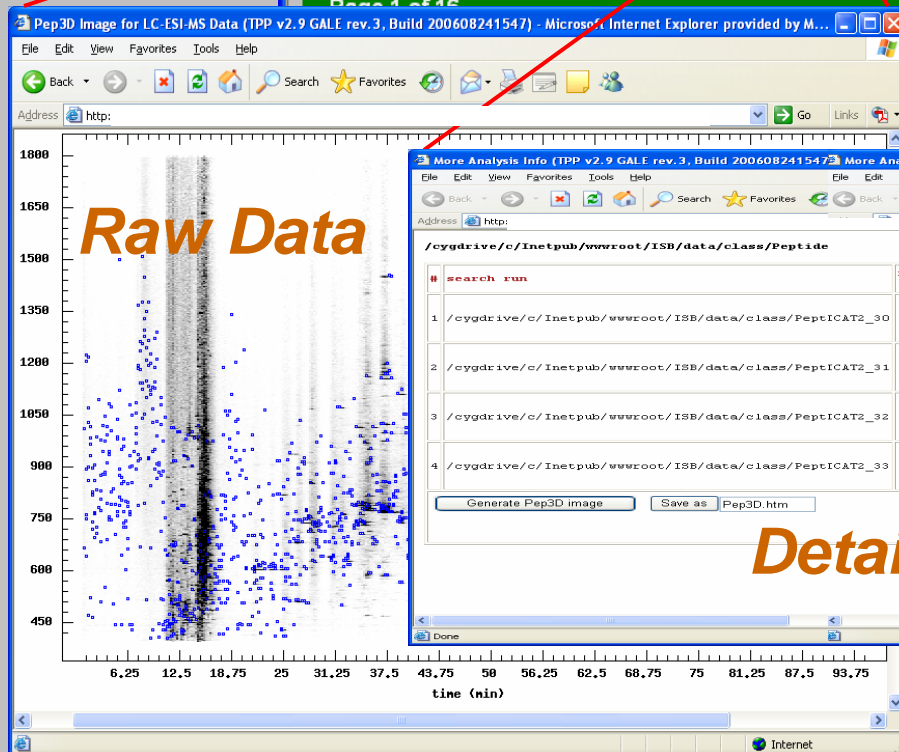
[Display Options](#) [Show | Hide]

[Column Order and Selection](#) [Show | Hide]

[Filtering Options](#) [Show | Hide]

Page 4 of 16

CORE	IDENTITYSCORE	HOMOLOGYScore	IONS	PEPTID
3.54	54.40	29.47	3/10	



More Analysis Info (TPP v2.9 GALE rev.3, Build 200608241547)

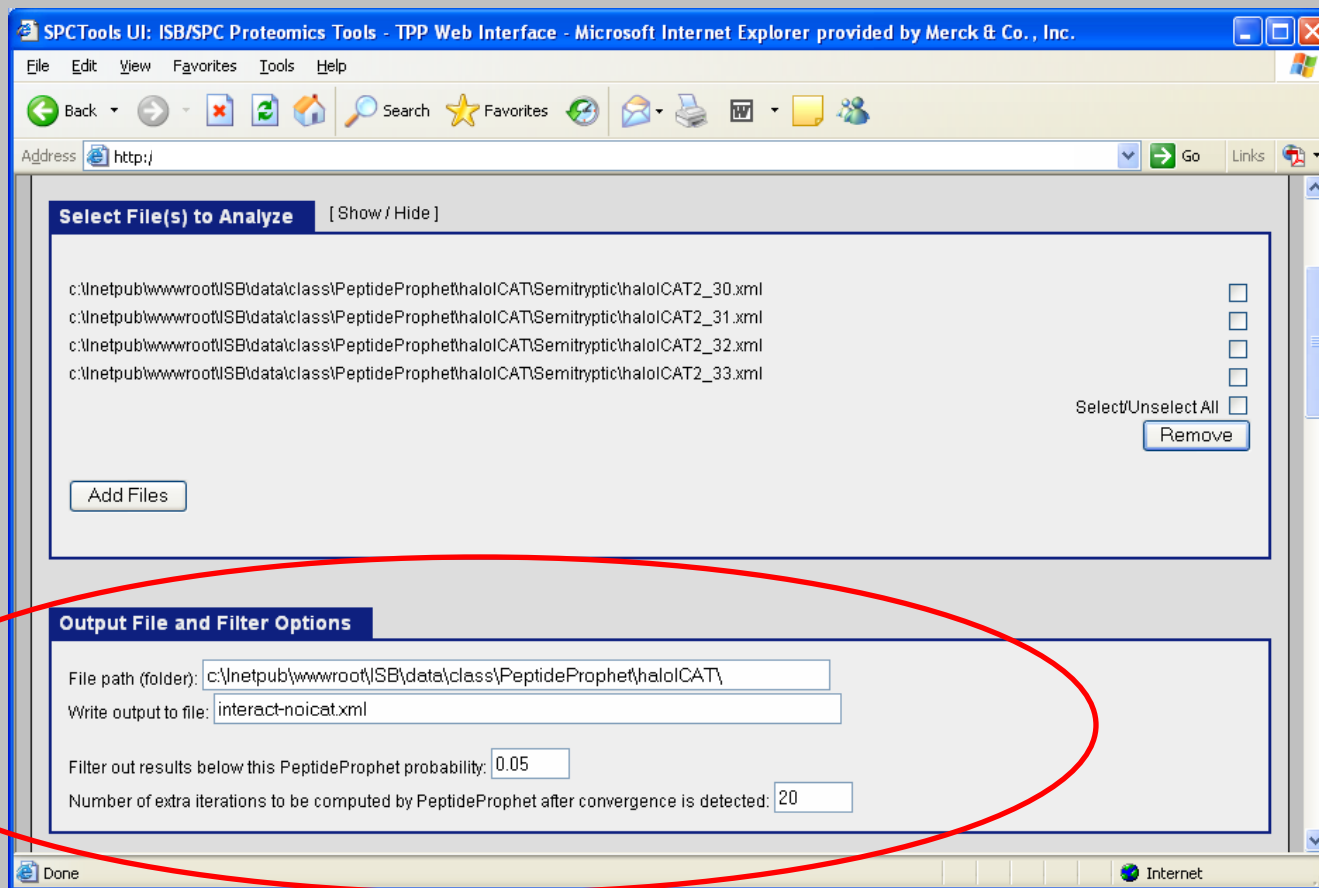
Address http: /cygdrive/c/Inetpub/wwwroot/ISB/data/class/Peptide

#	search run	no. results	mass spectrometer	sample enzyme	database search	refresh database	PeptideProphet™ analysis
1	/cygdrive/c/Inetpub/wwwroot/ISB/data/class/PeptICAT2_30	108	ThermoFinnigan LCQ Classic Ion Trap UNKNOWN	trypsin	MASCOT	/cygdrive/c/Inetpub/wwwroot/ISB/data/class/dbase/haIobac 2006-09-19T08:06:19	2006-09-19T08:06:19
2	/cygdrive/c/Inetpub/wwwroot/ISB/data/class/PeptICAT2_31	95	ThermoFinnigan LCQ Classic Ion Trap UNKNOWN	trypsin	MASCOT	/cygdrive/c/Inetpub/wwwroot/ISB/data/class/dbase/haIobac 2006-09-19T08:06:19	2006-09-19T08:06:19
3	/cygdrive/c/Inetpub/wwwroot/ISB/data/class/PeptICAT2_32	101	ThermoFinnigan LCQ Classic Ion Trap UNKNOWN	trypsin	MASCOT	/cygdrive/c/Inetpub/wwwroot/ISB/data/class/dbase/haIobac 2006-09-19T08:06:19	2006-09-19T08:06:19
4	/cygdrive/c/Inetpub/wwwroot/ISB/data/class/PeptICAT2_33	90	ThermoFinnigan LCQ Classic Ion Trap UNKNOWN	trypsin	MASCOT	/cygdrive/c/Inetpub/wwwroot/ISB/data/class/dbase/haIobac 2006-09-19T08:06:19	2006-09-19T08:06:19

Details of Peptide Analysis

# User Options for PeptideProphet

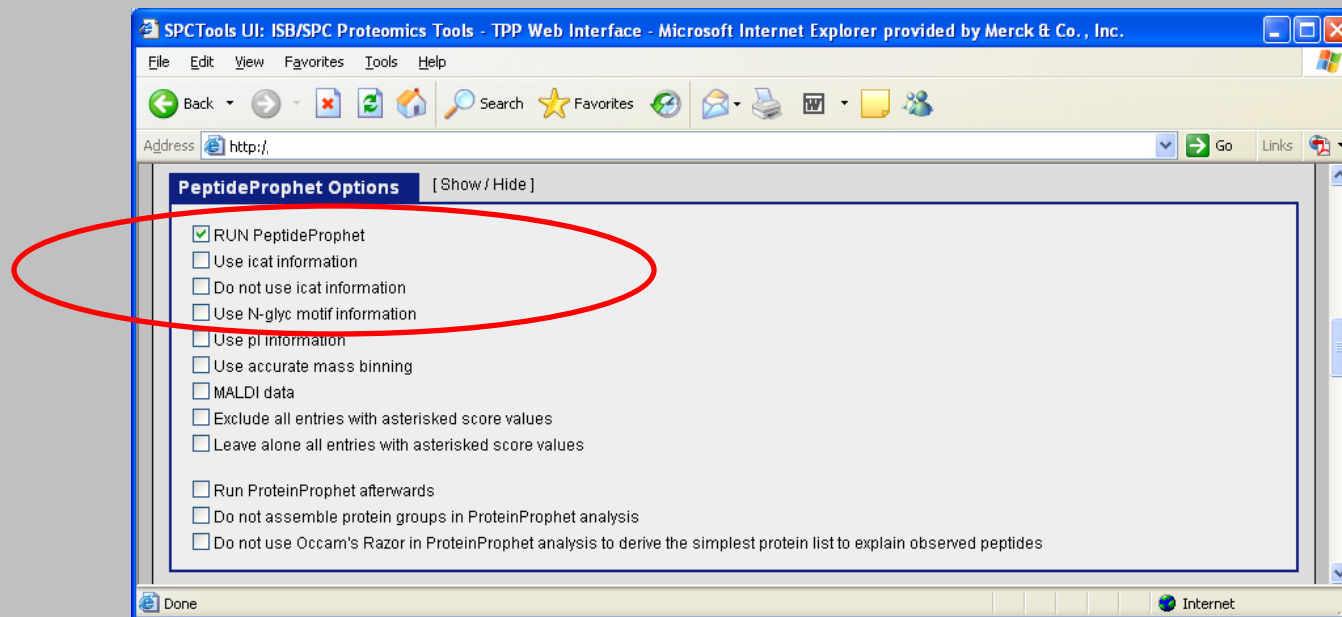
Rename Output File (e.g. to `interact-noicat.xml`):



# Use of Supplemental Discriminating Information

Use additional discriminating information, including ICAT or N-glyc, when relevant

- PeptideProphet automatically uses ICAT information when it thinks appropriate
- Nevertheless, you can explicitly set whether or not ICAT information is utilized



# Ionscore\* Example

- Search results are marked with asterisked Ionscore when runner up peptide(s) share at least 75% sequence identity with top peptide

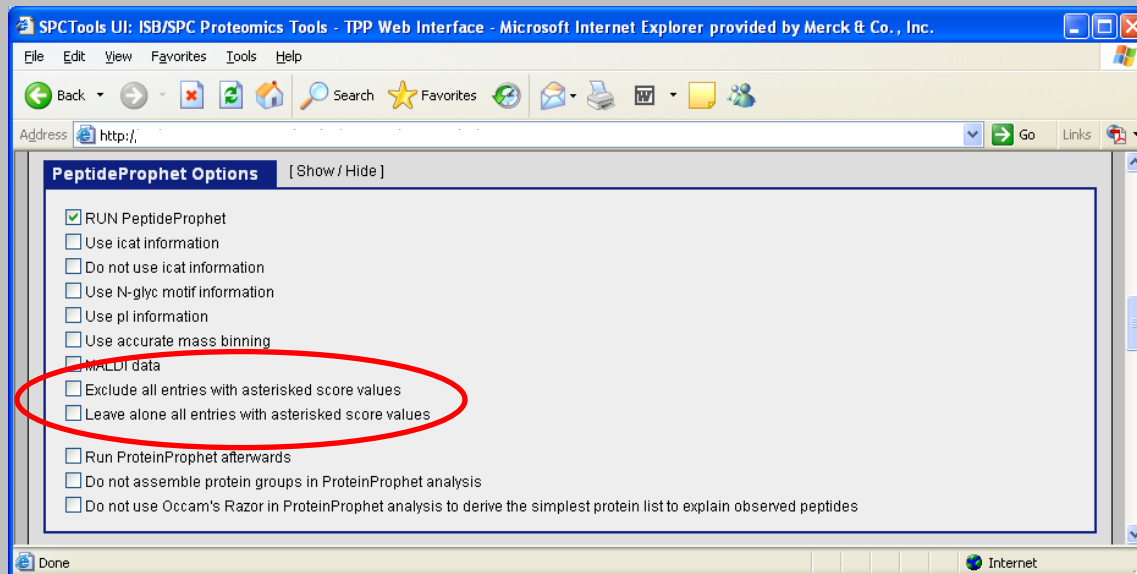
6779	0.0000	<a href="#">halolCAT2_32.1630.1630.3</a>	17.17	62.53	29.81	<a href="#">32/116</a>
6778	0.0000	<a href="#">halolCAT2_32.1282.1282.3</a>	16.15*	62.53	27.83	<a href="#">8/88</a>
6777	0.0000	<a href="#">halolCAT2_32.2442.2442.3</a>	18.98	62.53	30.02	<a href="#">29/112</a>
7349	0.0000	<a href="#">halolCAT2_32.0013.0013.3</a>	19.66	63.49	30.44	<a href="#">17/160</a>
6776	0.0000	<a href="#">halolCAT2_32.1122.1122.3</a>	18.83	62.55	29.99	<a href="#">8/100</a>
6759	0.0000	<a href="#">halolCAT2_32.2730.2730.3</a>	19.43*	62.53	31.44	<a href="#">10/108</a>
6775	0.0000	<a href="#">halolCAT2_32.1194.1194.3</a>	14.71	62.55	26.12	<a href="#">8/104</a>
6774	0.0000	<a href="#">halolCAT2_32.2836.2836.3</a>	22.91	62.55	29.73	<a href="#">18/108</a>
6773	0.0000	<a href="#">halolCAT2_32.2562.2562.3</a>	23.14	62.55	35.85	<a href="#">6/100</a>
7350	0.0000	<a href="#">halolCAT2_32.3348.3348.3</a>	16.86	63.49	29.49	<a href="#">10/152</a>
6830	0.0000	<a href="#">halolCAT2_32.2410.2410.3</a>	19.88	62.64	32.38	<a href="#">22/120</a>

#	MH+	IonSc	Ions	Ref	Sequence
1	2956.4 (-0.4)	19.43	<a href="#">10/108</a>	<a href="#">SWN:RN19 HUMAN</a>	<a href="#">F.STNTSSDNGLTSISKQIGDFIECPLCLL.R</a>
2	2956.4 (-0.4)	17.93	<a href="#">10/ 92</a>	<a href="#">SWN:RN19 HUMAN</a>	<a href="#">F.STNTSSDNGLTSISKQIGDFIEC<sub>345</sub>P.L</a>
3	2954.6 (+1.4)	17.79	<a href="#">14/100</a>	<a href="#">GP:AY038599 2</a>	<a href="#">T.AQSEVALLRFVNPDTGRVLFESKLHK.Q</a>
4	2955.5 (+0.4)	17.73	<a href="#">11/104</a>	<a href="#">SW:ELS HUMAN</a>	<a href="#">+6 F.PLGGVAARPGFGLSPIFPGGACLKAC<sub>345</sub>.G</a>
5	2954.6 (+1.4)	16.24	<a href="#">5/108</a>	<a href="#">SWN:Y450 HUMAN</a>	<a href="#">G.LFLRGPKPGSLDASHAAGRPPARPSVSQR.I</a>
6	2955.6 (+0.4)	15.79	<a href="#">12/ 92</a>	<a href="#">pNRC100 ORF5058</a>	<a href="#">+3 T.DVQPWRLLVGGVVFVGIGTRVGKGC<sub>345</sub>.T</a>
7	2954.6 (+1.4)	15.44	<a href="#">22/124</a>	<a href="#">GP:M24766 1</a>	<a href="#">P.KGDPGFPGAPGTGAPGIAGIPQKIAVQPGTV.G</a>
8	2955.4 (+0.5)	15.19	<a href="#">8/112</a>	<a href="#">SW:GGT5 HUMAN</a>	<a href="#">+2 P.CGPOAFAHAAVAADSKVCSDIGRAILQQQ.G</a>
9	2956.3 (-0.3)	14.84	<a href="#">19/112</a>	<a href="#">SW:VWF HUMAN</a>	<a href="#">+2 E.CCGRCLPSACEVVTGSPRGDSQSSWKSVG.S</a>
10	2954.4 (+1.6)	14.78	<a href="#">22/124</a>	<a href="#">GP:S79774 1</a>	<a href="#">P.PTGDSGPPVPPTGDSGAPPVTPTGDSETAPV.P</a>

# Ionscore\* Options

There are three ways asterisked Ionscores can be treated by PeptideProphet:

- Penalize (the default option, halves Ionscore values)
- Leave alone (suitable for the context of homologues)
- Exclude (the most conservative, assigns probability 0)





# Run/Don't Run PeptideProphet

SPCTools UI: ISB/SPC Proteomics Tools - TPP Web Interface - Microsoft Internet Explorer provided by Merck & Co., Inc.

File Edit View Favorites Tools Help

Address <http://> Go Links

### Output File and Filter Options

File path (folder):   
Write output to file:   
Filter out results below this PeptideProphet probability:   
Number of extra iterations to be computed by PeptideProphet after convergence is detected:

### PeptideProphet Options [ Show / Hide ]

- RUN PeptideProphet
- Use icat information
- Do not use icat information
- Use N-glyc motif information
- Use pI information
- Use accurate mass binning
- MALDI data
- Exclude all entries with asterisked score values
- Leave alone all entries with asterisked score values
- Run ProteinProphet afterwards
- Do not assemble protein groups in ProteinProphet analysis
- Do not use Occam's Razor in ProteinProphet analysis to derive the simplest protein list to explain observed peptides

Done Internet

# Ongoing Developments for PeptideProphet

- Optimize for various additional mass spectrometers
  - New discriminant function
- Adapt to additional methods for assigning peptides to tandem mass spectra
  - SEQUEST
  - COMET
  - Probid
  - SpectraST
  - X!Tandem
  - Others

# Exercises with PeptideProphet

---

- Accuracy of computed probabilities
- Utility of conventional Mascot score thresholds and PeptideProphet analysis
- Model results for ICAT data analyzed with and without ICAT information
- Model results for Mystery dataset

# Exercise Datasets

Many of the exercises utilize Mascot search results of ***HalolCAT*** datasets for which correct results are independently known:

- MS/MS spectra generated from *Halobacterium* ICAT sample searched against a halobacterium\_plus\_human protein sequence database

The pepXML Viewer is pre-configured for this class to automatically color all ***HalolCAT*** correct corresponding proteins **red!**