About OMICS Group

OMICS Group International is an amalgamation of Open Access publications and worldwide international science conferences and events. Established in the year 2007 with the sole aim of making the information on Sciences and technology 'Open Access', OMICS Group publishes 400 online open access scholarly journals in all aspects of Science, Engineering, Management and Technology journals. OMICS Group has been instrumental in taking the knowledge on Science & technology to the doorsteps of ordinary men and women. Research Scholars, Students, Libraries, Educational Institutions, Research centers and the industry are main stakeholders that benefitted greatly from this knowledge dissemination. OMICS Group also organizes 300 International conferences annually across the globe, where knowledge transfer takes place through debates, round table discussions, poster presentations, workshops, symposia and exhibitions.

About OMICS Group Conferences

OMICS Group International is a pioneer and leading science event organizer, which publishes around 400 open access journals and conducts over 300 Medical, Clinical, Engineering, Life Sciences, Pharma scientific conferences all over the globe annually with the support of more than 1000 scientific associations and 30,000 editorial board members and 3.5 million followers to its credit.

OMICS Group has organized 500 conferences, workshops and national symposiums across the major cities including San Francisco, Las Vegas, San Antonio, Omaha, Orlando, Raleigh, Santa Clara, Chicago, Philadelphia, Baltimore, United Kingdom, Valencia, Dubai, Beijing, Hyderabad, Bengaluru and Mumbai.

St. Jude Children's Research Hospital

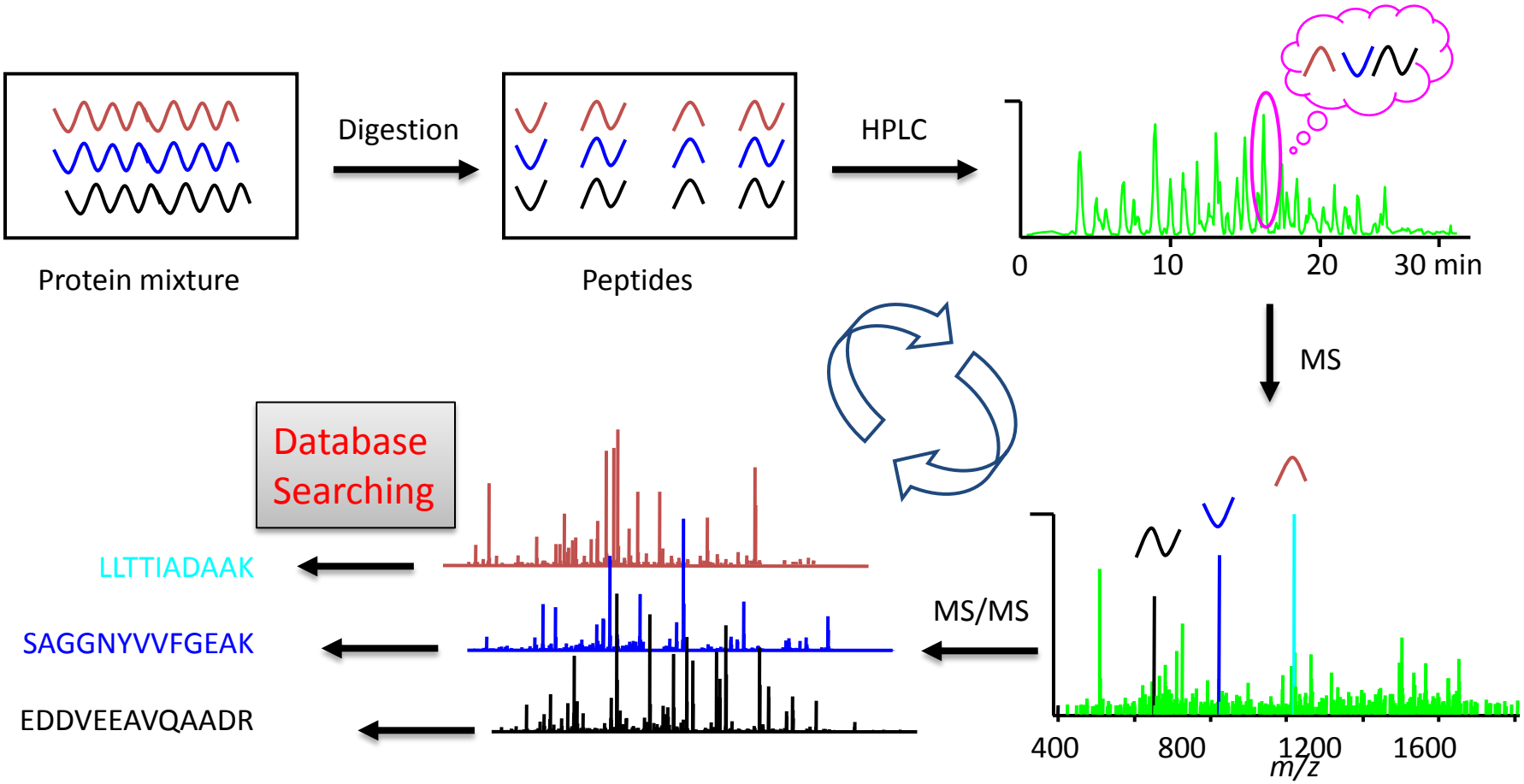
JUMP: a Tag-based Database Search Tool for Peptide Identification with High Sensitivity and Accuracy

Xusheng Wang

4th International Conference on Proteomics & Bioinformatics



Shot-gun proteomics



Current available search engines

Database search **Hybrid** **De novo**

SEQUEST
MASCOT
X!Tandem
OMSSA
MyriMatch
Andromeda
Morpheus

InspecT
PEAKS DB
GutenTag
ByOnics
Lookup Peaks
OpenSea

PEAKS
Lutefisk
PepNovo
Vonode



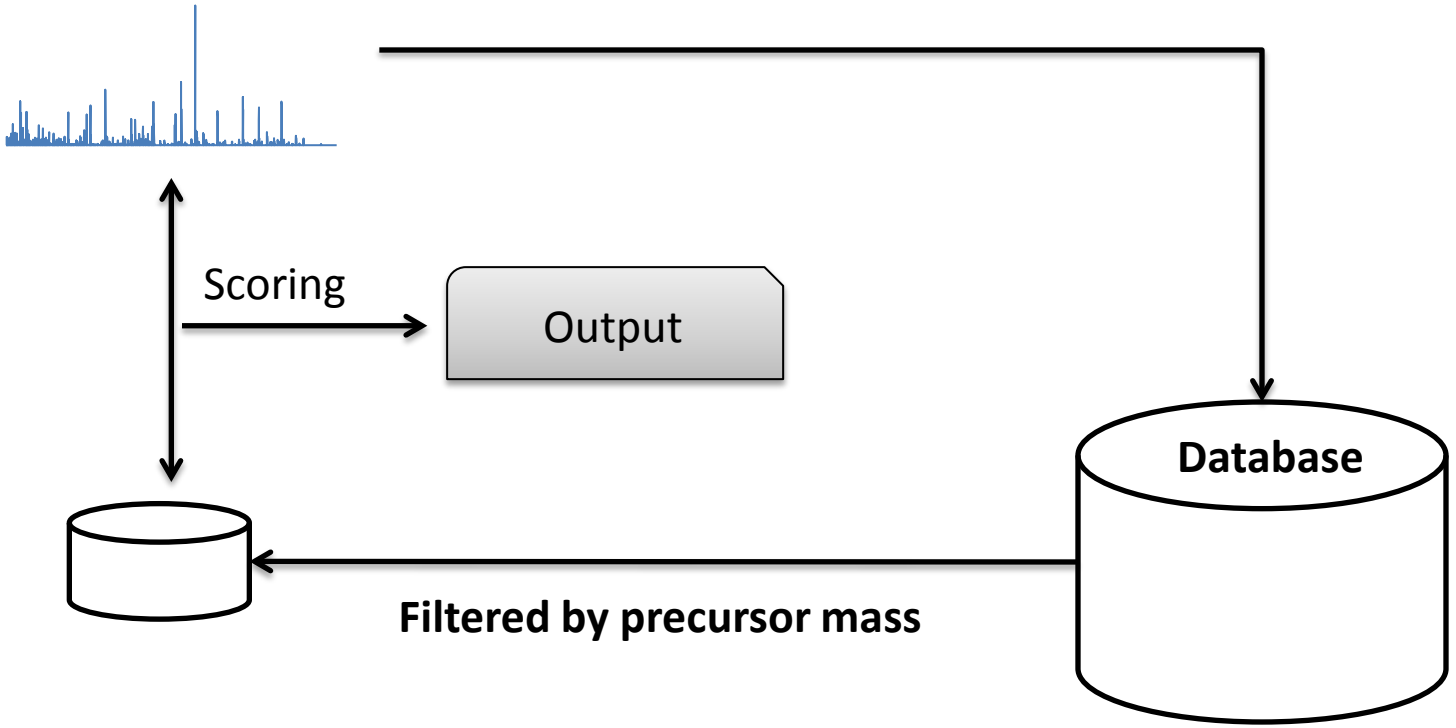
History of development

1994 1999 2001 2003 2004 2005 2007 2010 2011 2012 2013 2014

Pattern	**SEQUEST**	**MASCOT**	**X!Tandem**	**OMSSA**		**MyriMatch**		**Andromeda**	**Peaks DB**	**Morpheus**	
Hybrid	**Mann's Tag**		**GutenTag** **OpenSea**		**InspecT**	**Lookup Peaks** **ByOnics**					**JUMP**
De novo		**Lutefisk**	**PEAKS**		**PepNovo**		**Vonode**				

Traditional database search

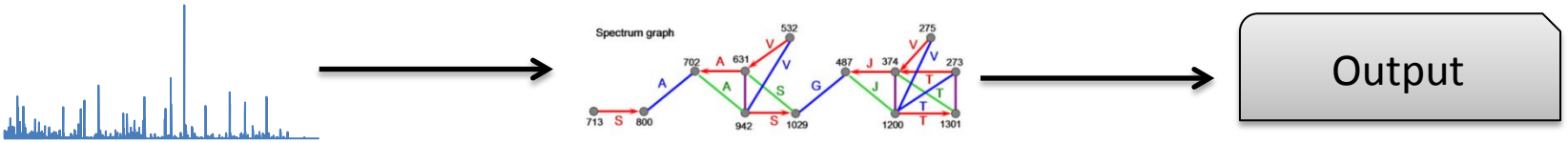
Traditional database search



Advantage: High sensitivity
Disadvantage: Low specificity; random matches; unknown modifications and mutations

De novo method

De novo method

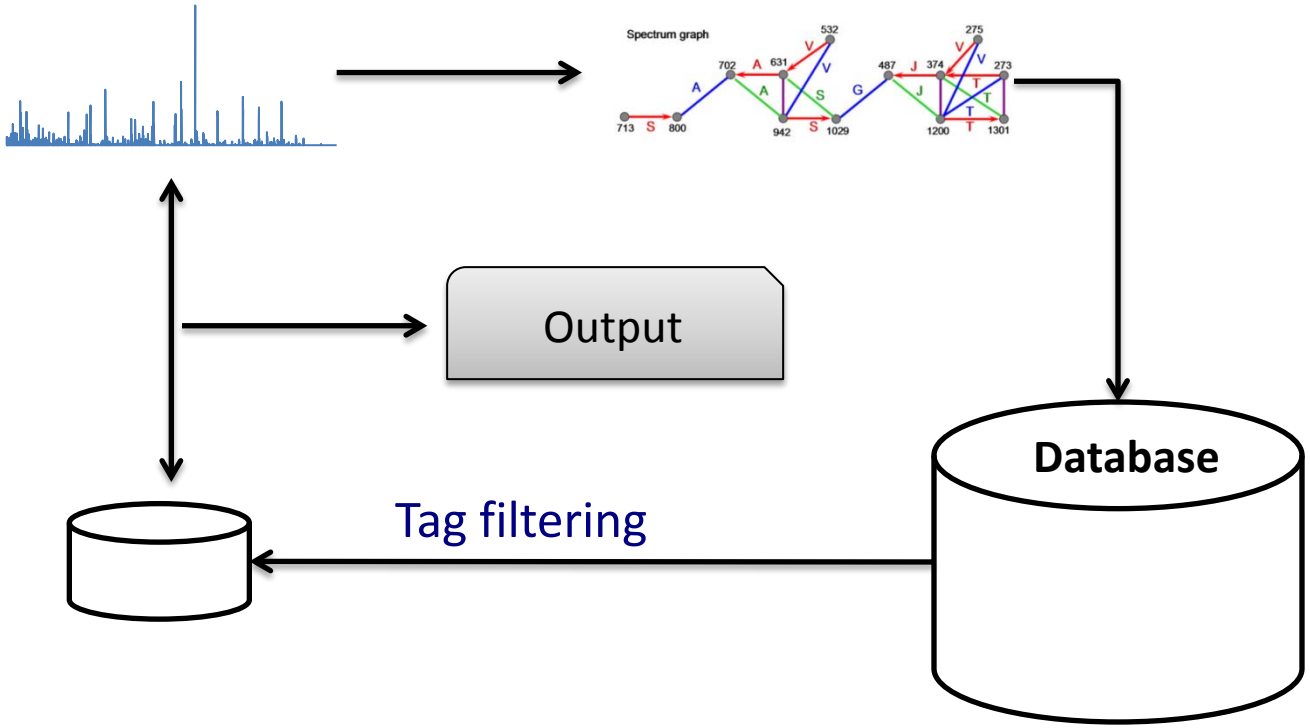


Advantage: Derive novel peptides and PTMs without database

Disadvantage: Low sensitivity; MS/MS fragmentation can not generate all predicted production ions

Hybrid method

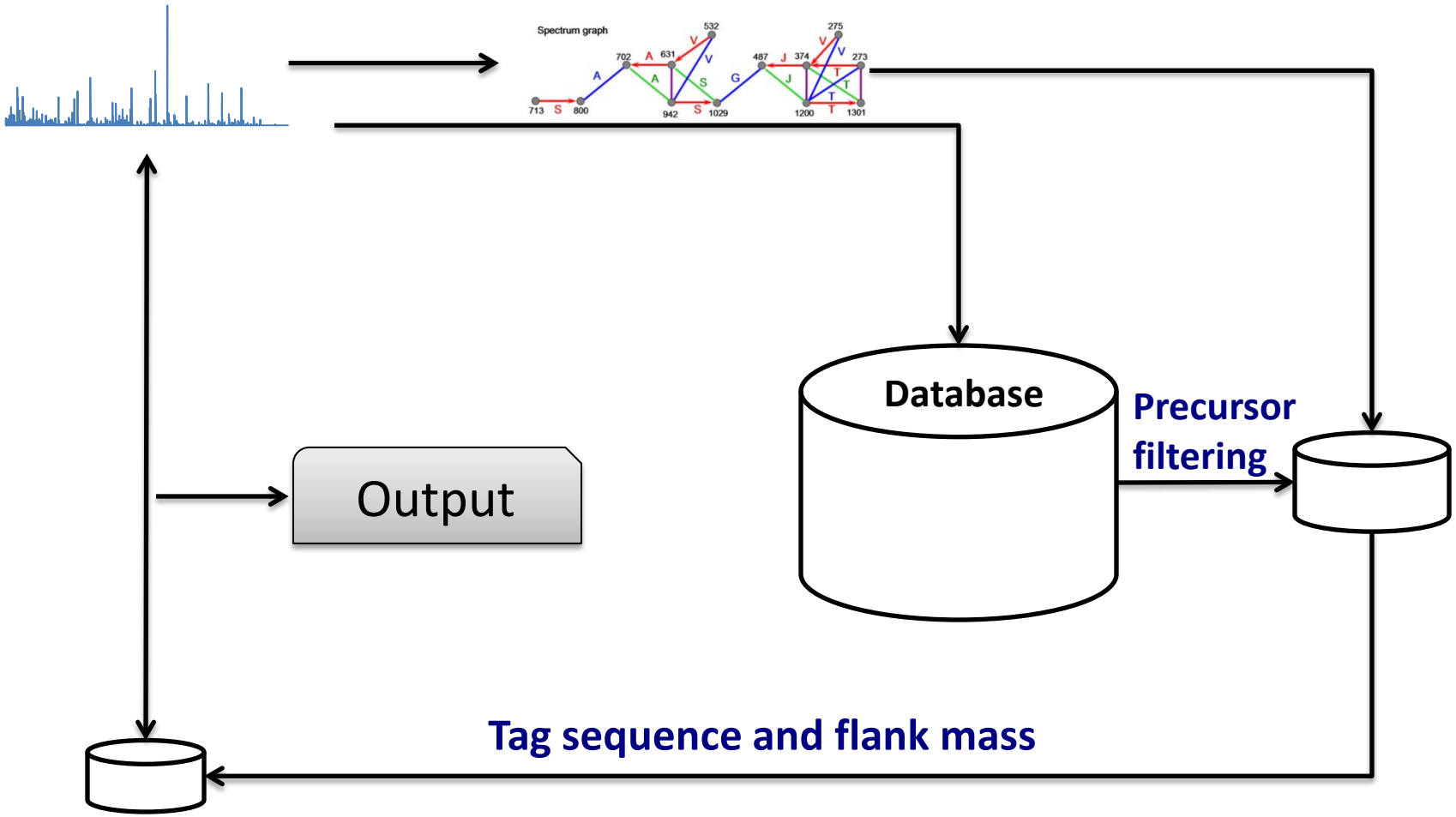
Hybrid method



Advantage: Improve specificity

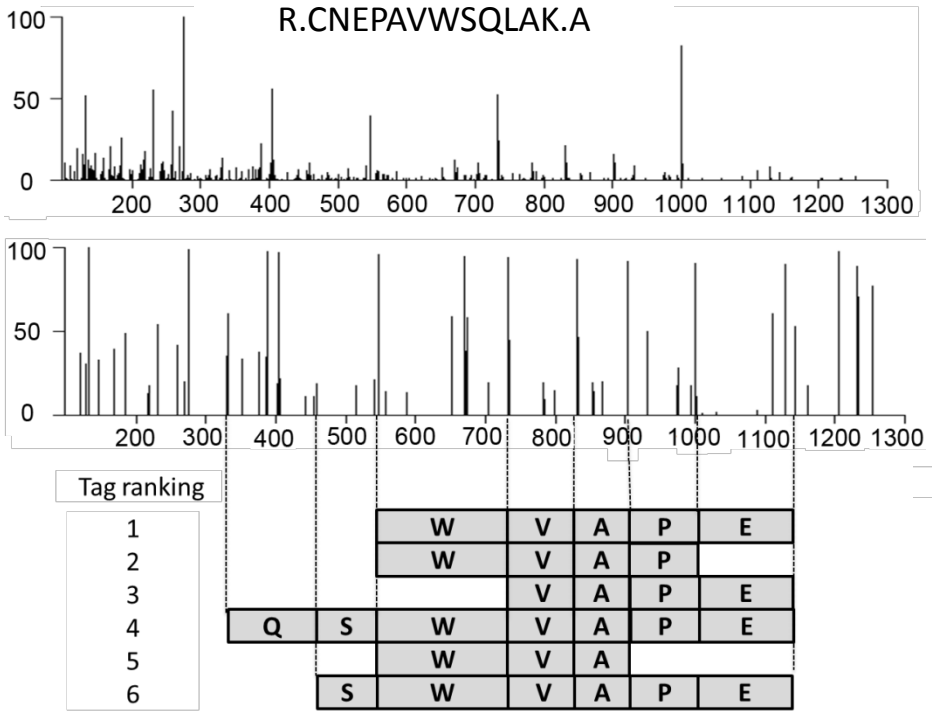
Disadvantage: requires a minimum tag length

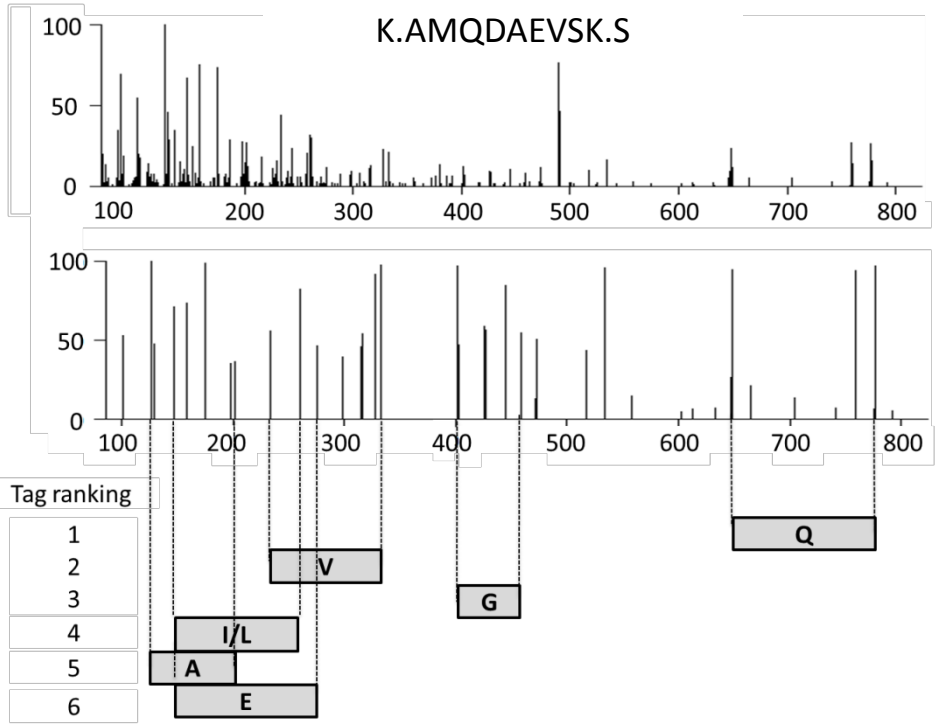
JUMP method



Advantage: High sensitivity and specificity

Tag sequence generation

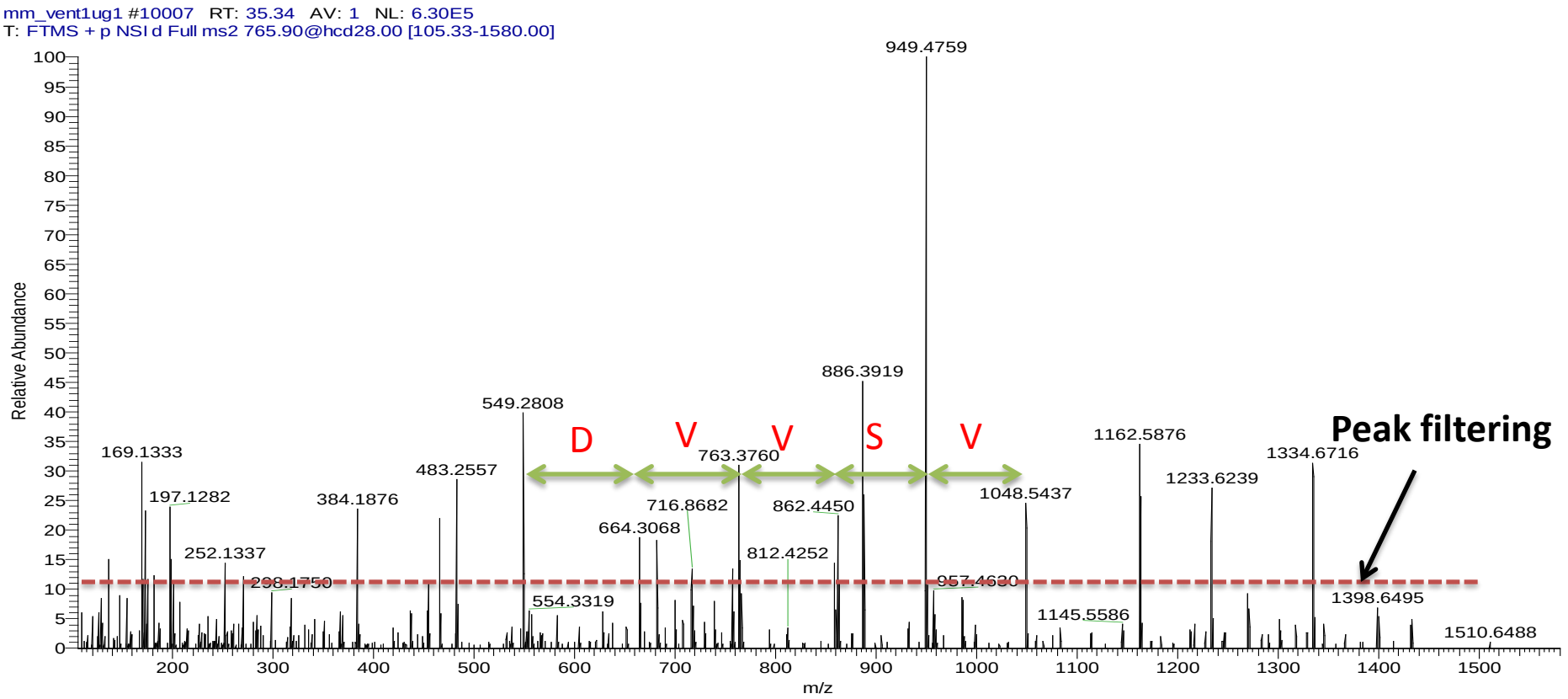




Good spectrum

Poor spectrum

Tag scoring



Wilcoxon Rank sum test

$$\omega_{NA} = \sum_{i=1}^{n=4} P_i$$

$$P - value = 2\,pr(W < \omega_{NA})$$

Tag=DVVSV={1,3,4,11,18}

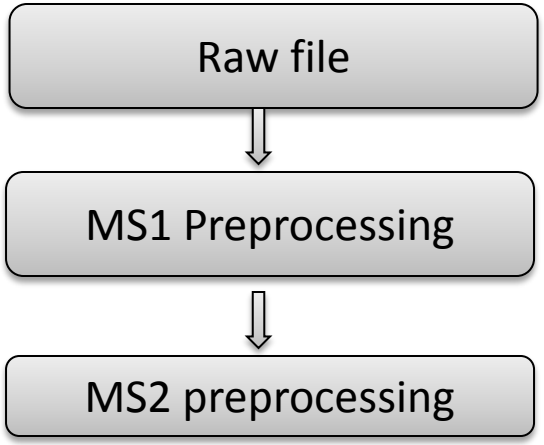
P-value = 0.001

Peptide scoring

Generating theoretical mass ladder

Series	conditions
y, b	Precursor charge = 1
y, b, y++, b++	Precursor charge =2
…	…

Fixed and dynamic modification will be used



Peptide score: hypergeometric test

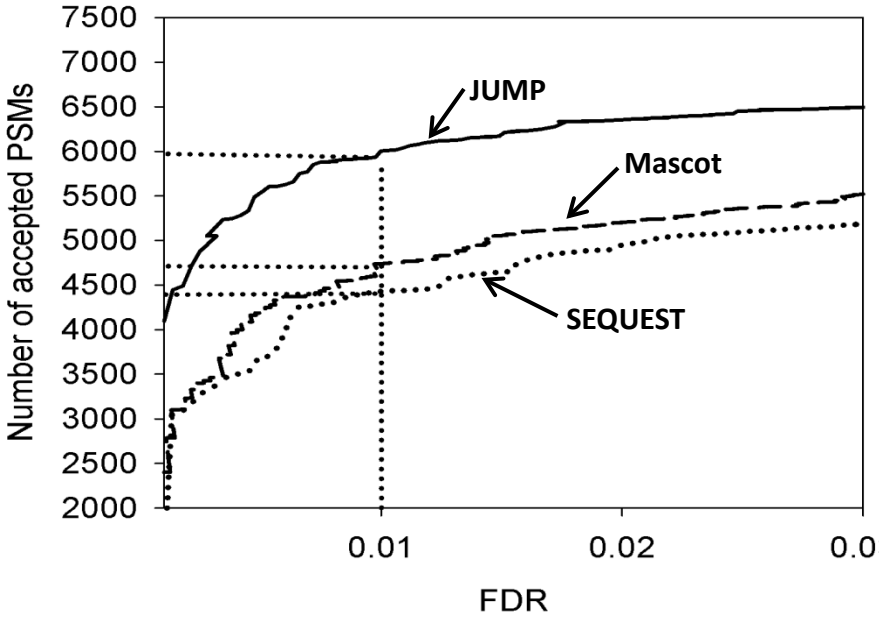
$$P(x = A) = \frac{\binom{n_1}{x}\binom{n_2}{r-x}}{\binom{n}{k}}$$

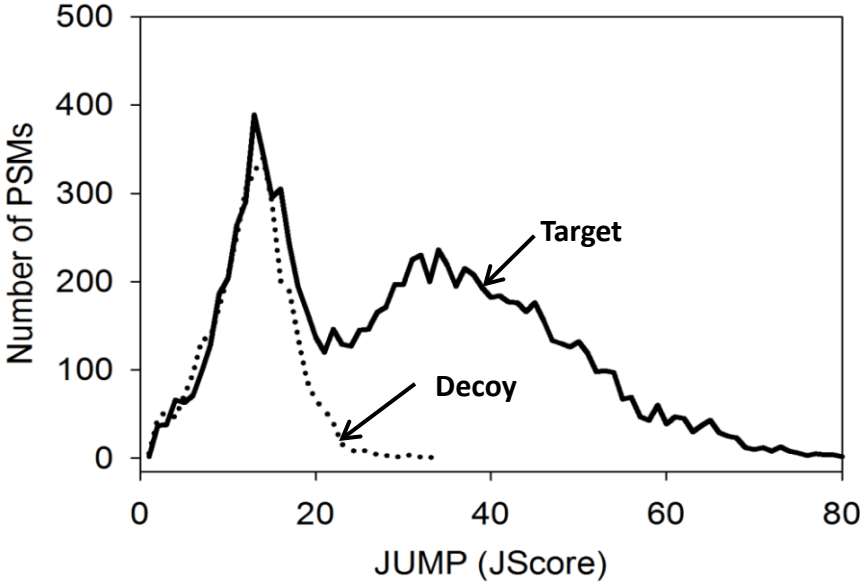
Final score

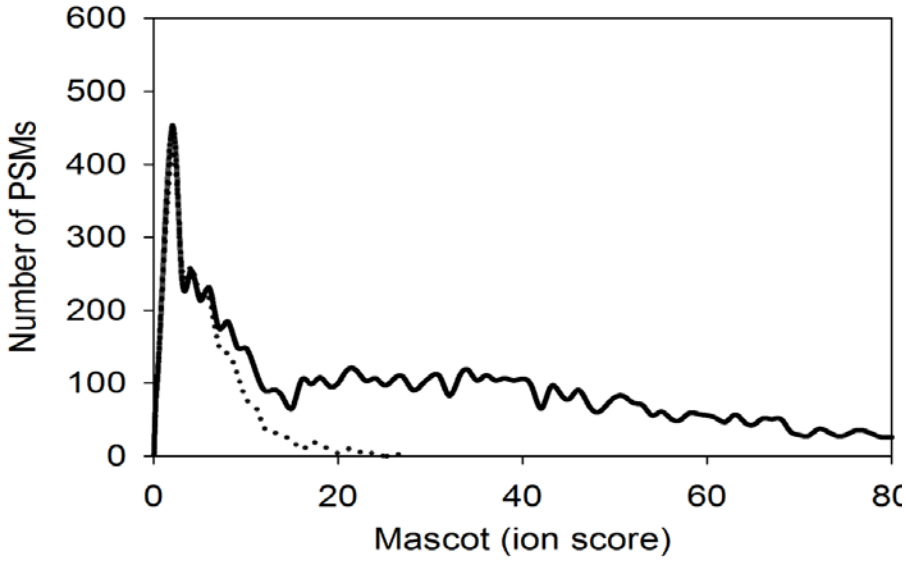
$$J_{score} = E_{tc} + E_{pc}$$

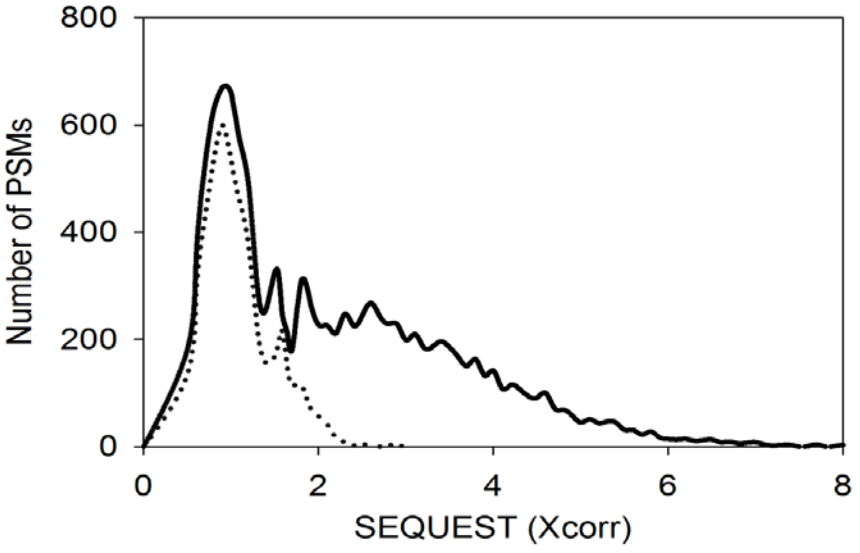
Performance comparison

Compared to Mascot and SEQUEST





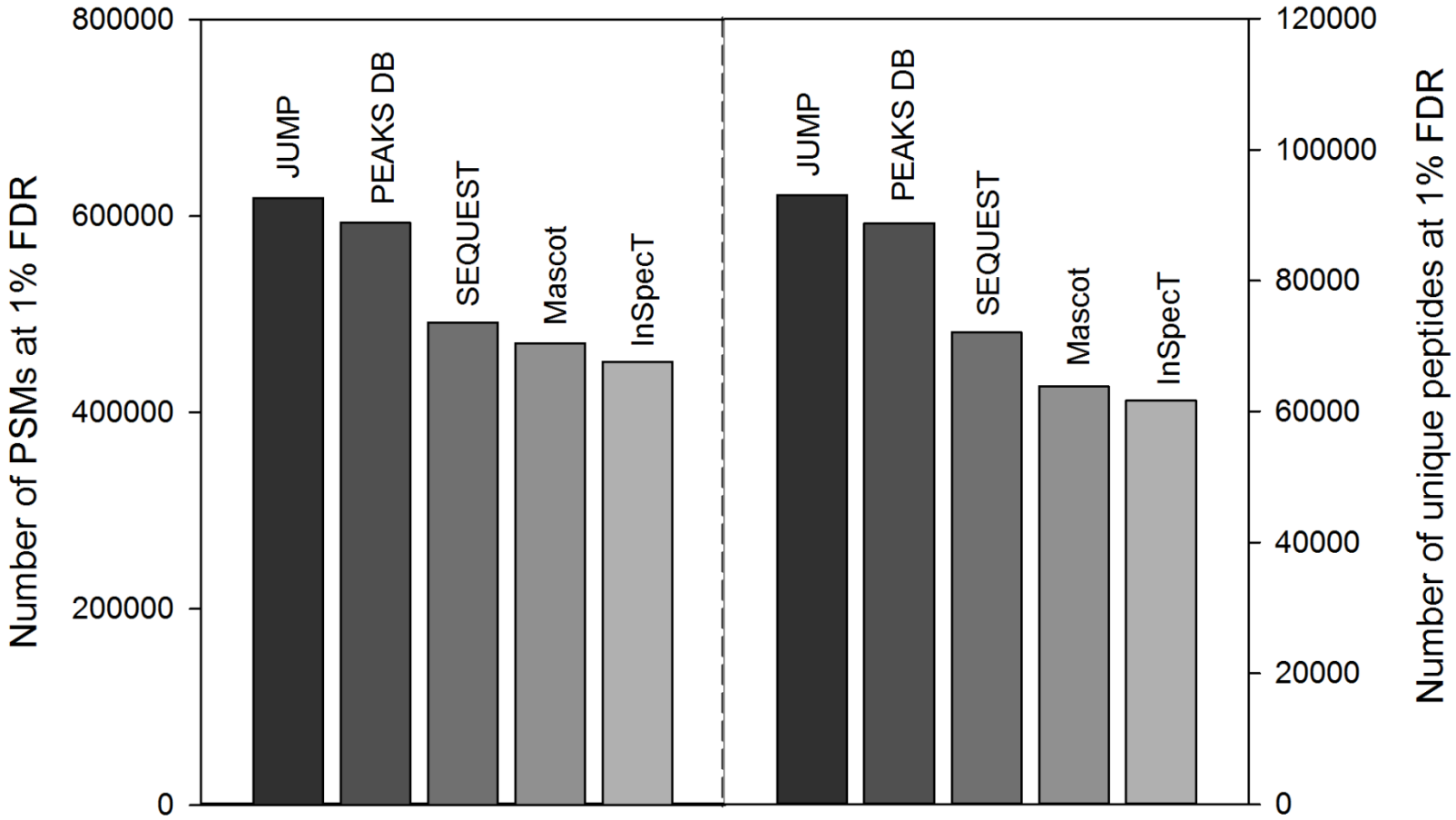




Performance comparison

Large data set:

contains 10 fractions, 1,718,768 MS/MS

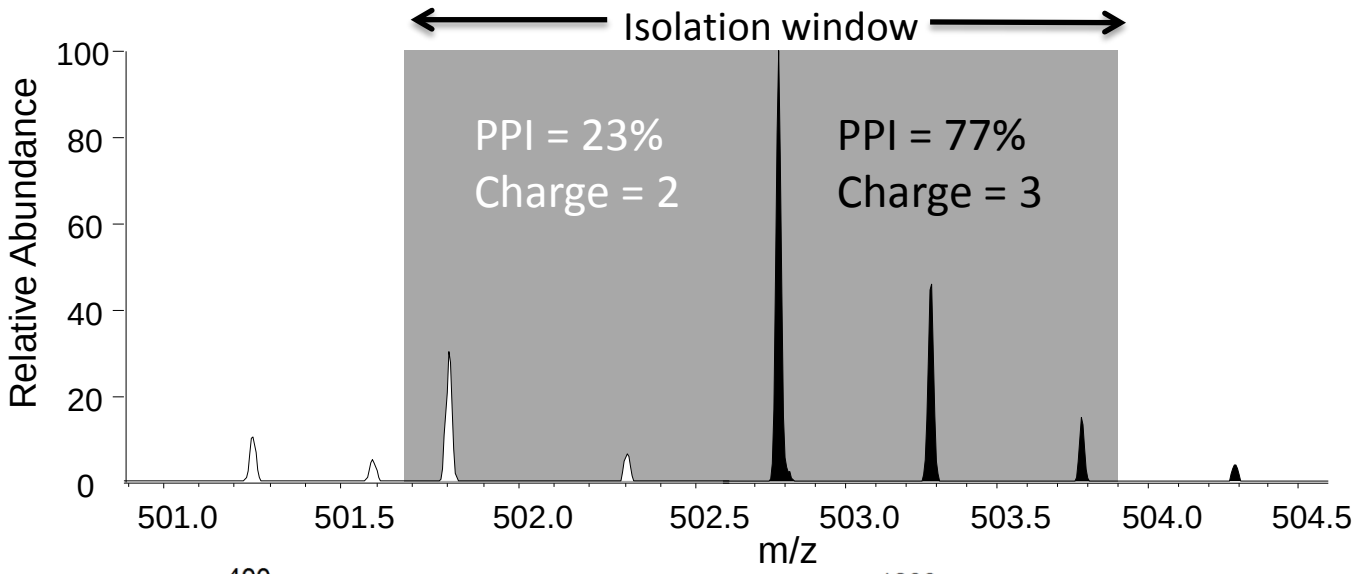


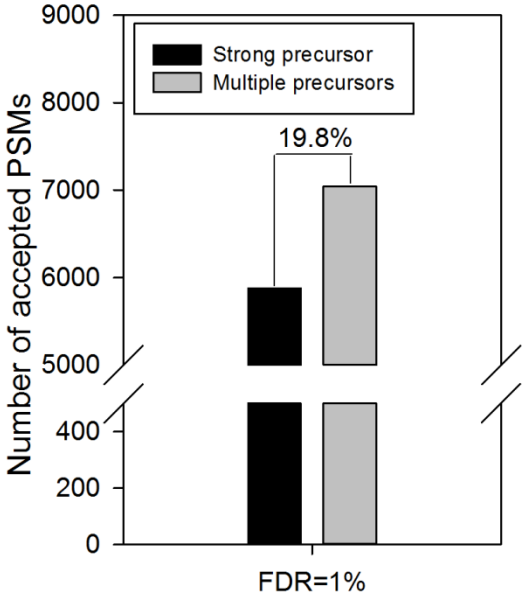
Identification of multiple precursors

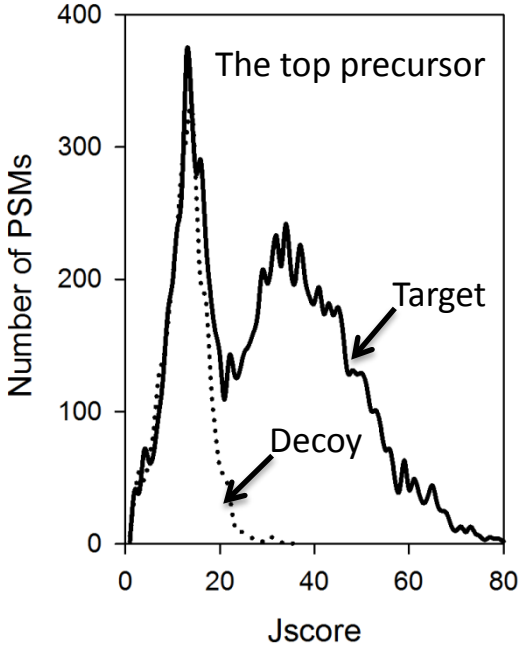
Two Method:

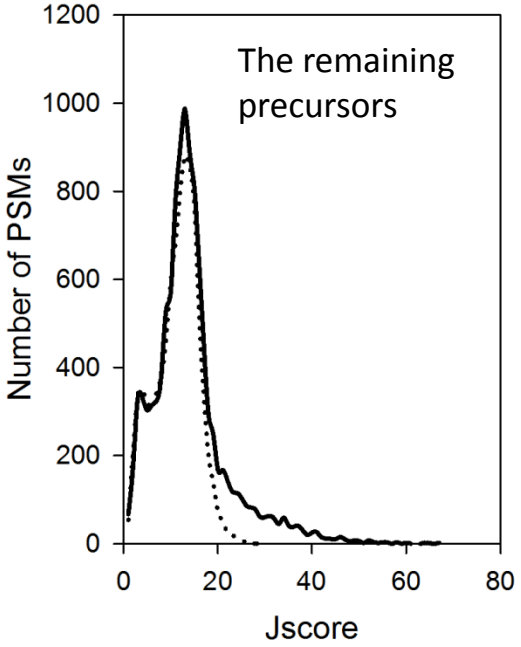
Andromeda

ProbIDtree



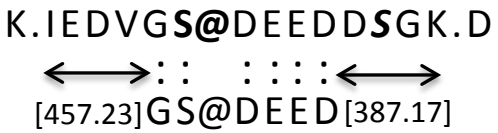




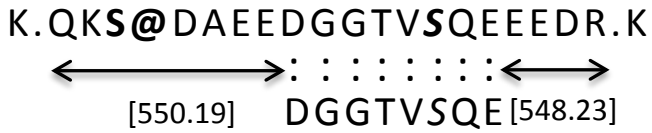


Assignment of PTM sites

- **1. PTM sites in the tag**



- **2. PTM site in the tag, but not modified**

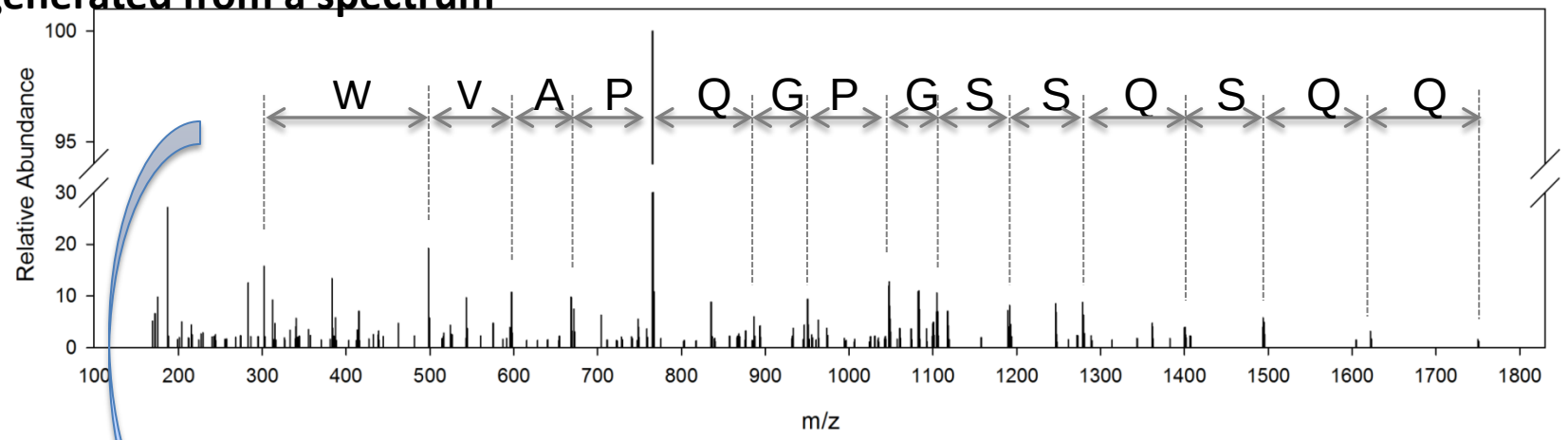


- **3. PTM site does not present in the tag, but site mass can determine the site assignment**



Partial de novo sequencing

Tag generated from a spectrum





Tag sequence QQSQ*S*SGPGQPAVW

: : : : : : : : : : : : :

Reference Sequence QQSQ*N*SGPGQPAVW

CAGCAGAGCCAGA***A***CTCCGGGCCCGGGCAGCCCGCTGTGTGG

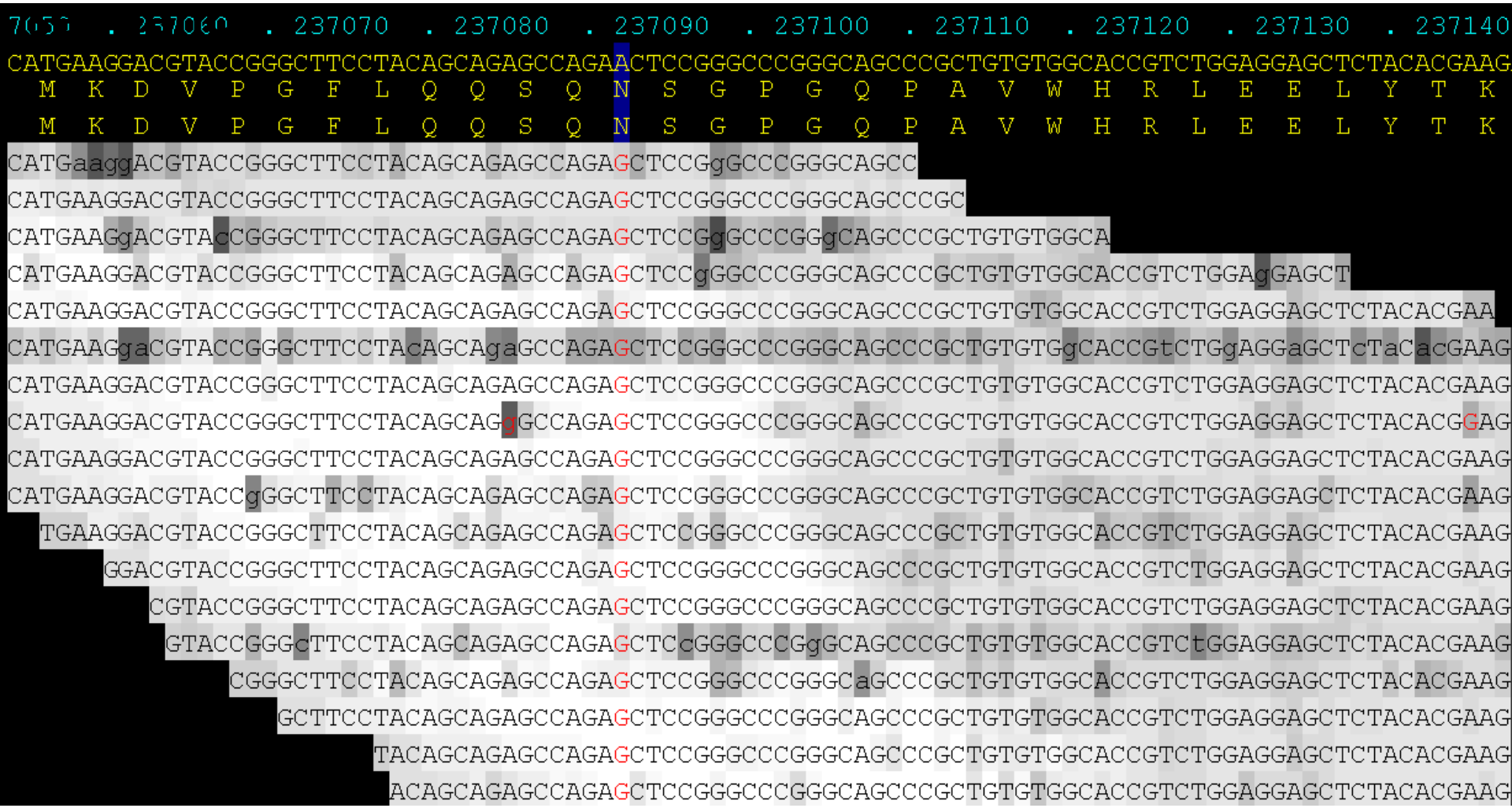
Q Q S Q ***N*** S G P G Q P A V W

Q Q S Q ***S*** S G P G Q P A V W

CAGCAGAGCCAGA***G***CTCCGGGCCCGGGCAGCCCGCTGTGTGG

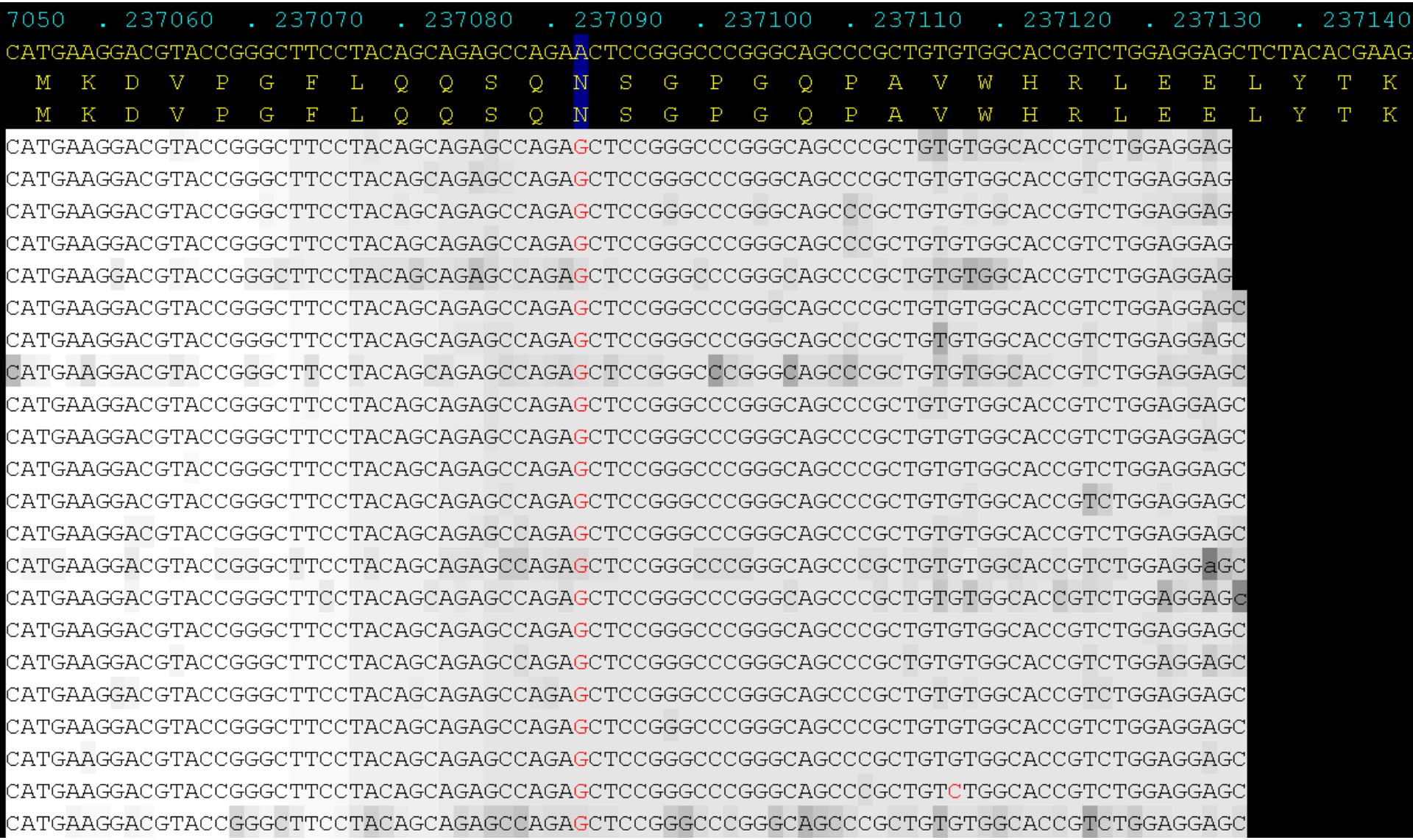
De novo sequencing of peptides

Whole genome sequencing confirmed the novel peptide



De novo sequencing of peptides

RNA-seq confirmed the novel peptide



JUMP is designed for HPC

- JUMP was written in *perl* by a modular approach
- designed for high performance parallel computing systems
- JUMP source codes together with detailed documents are freely available at http://stjuderesearch.org/site/lab/peng



Summary

- JUMP provides an effective, specific, and sensitive algorithm for peptide identification
- JUMP uses a novel algorithm for tag scoring, which combines the peak intensity-based score and the peak position-based score
- JUMP uses tags of as short as one amino acid
- JUMP can identify multiple candidate peptides for one spectrum
- JUMP is designed as a modular program for high performance computing systems

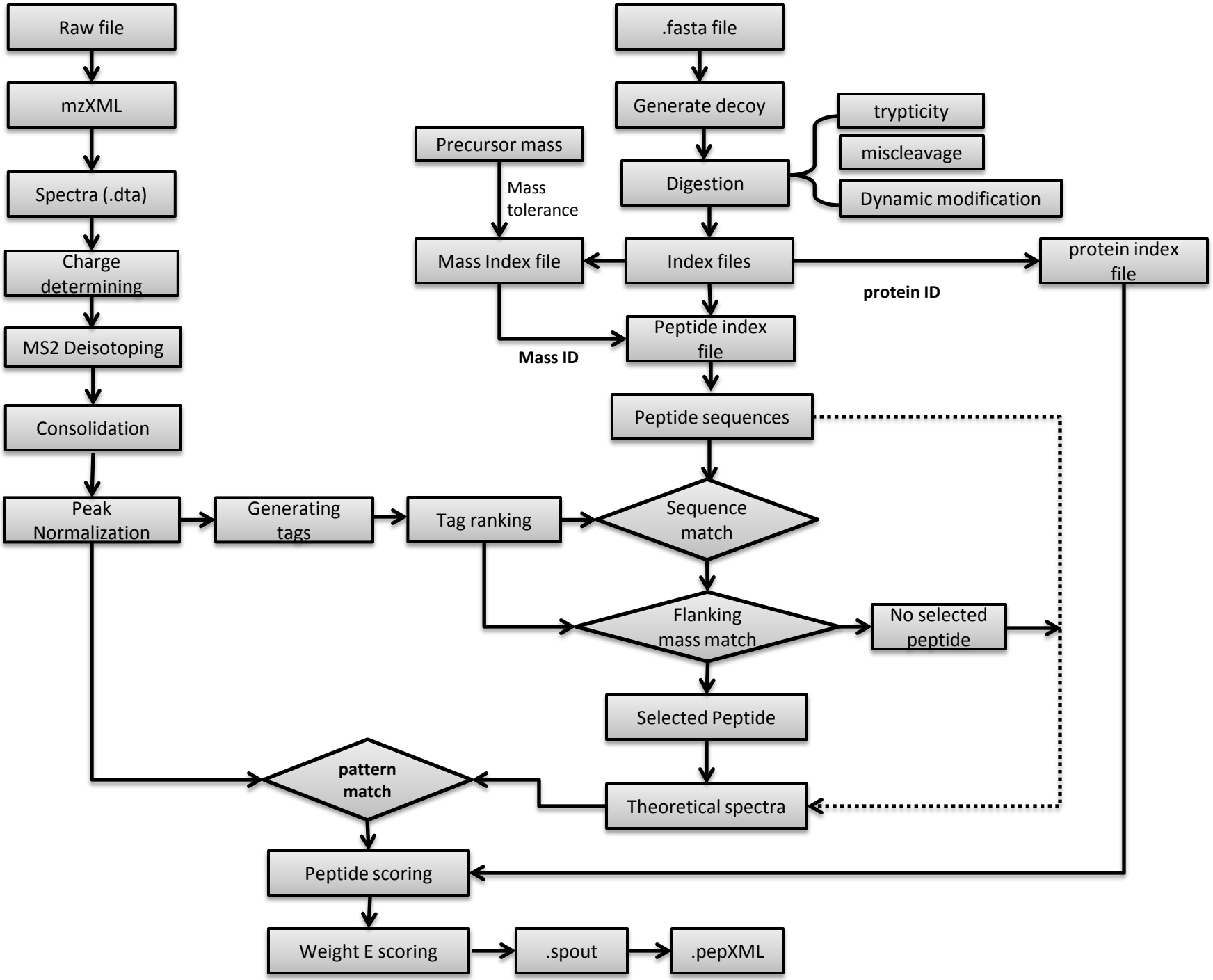
Acknowledgement



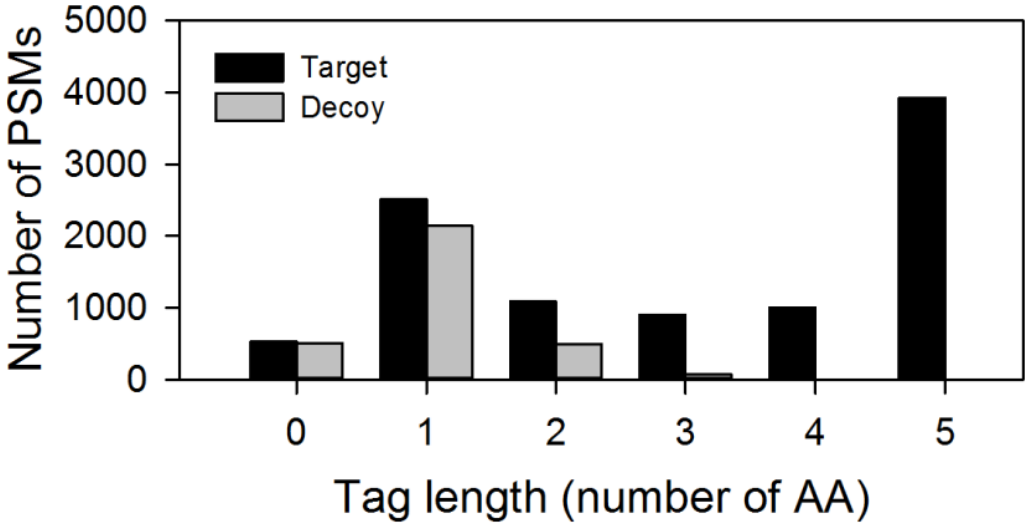
Junmin Peng
Yuxin Li
Ji-Hoon Cho
Timothy Thaw
Anthony High
Vishwajeeth Pagala
Haiyan Tan
Ashutosh Mishra
Kiran Kodali
Kanisha Kavdia
Zhiping Wu
Yanling Yang
Ping-Chung Chen
Drew Jones
Bing Bai
Joseph Mertz

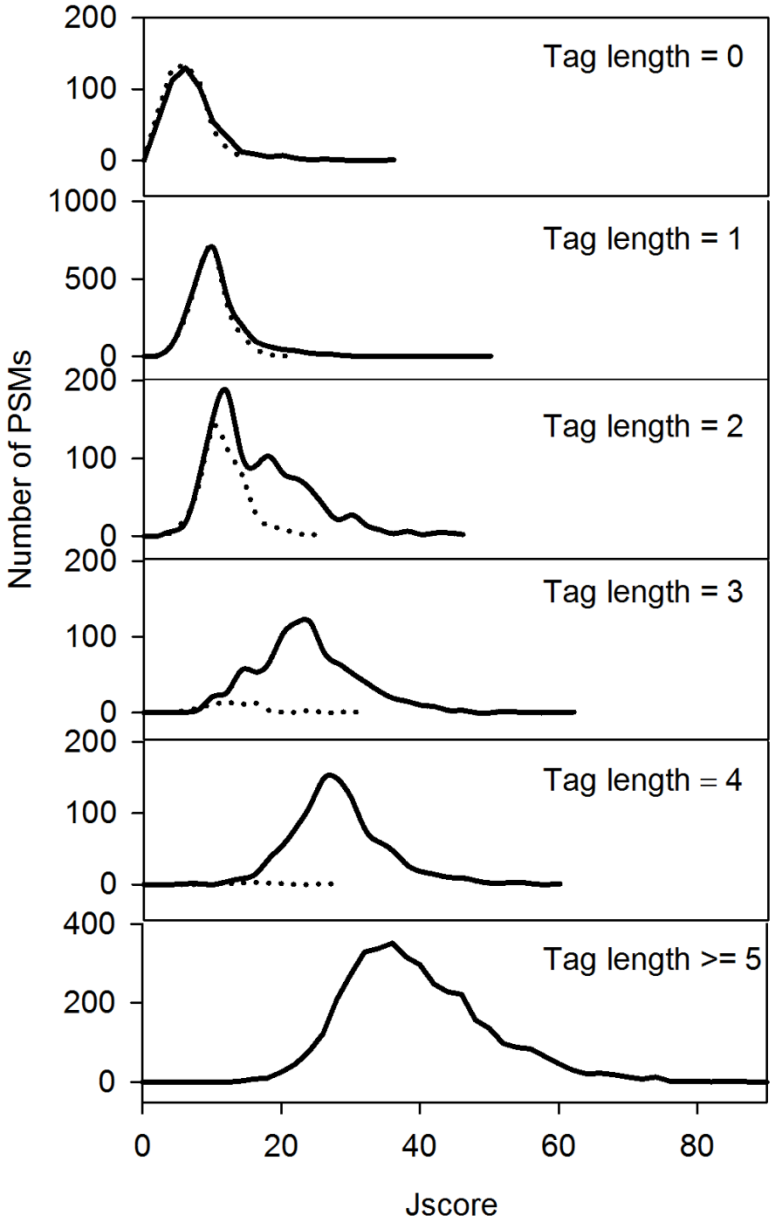
Financial Support:

NIH
American Cancer Society
St. Jude Children's Research Hospital (ALSAC)

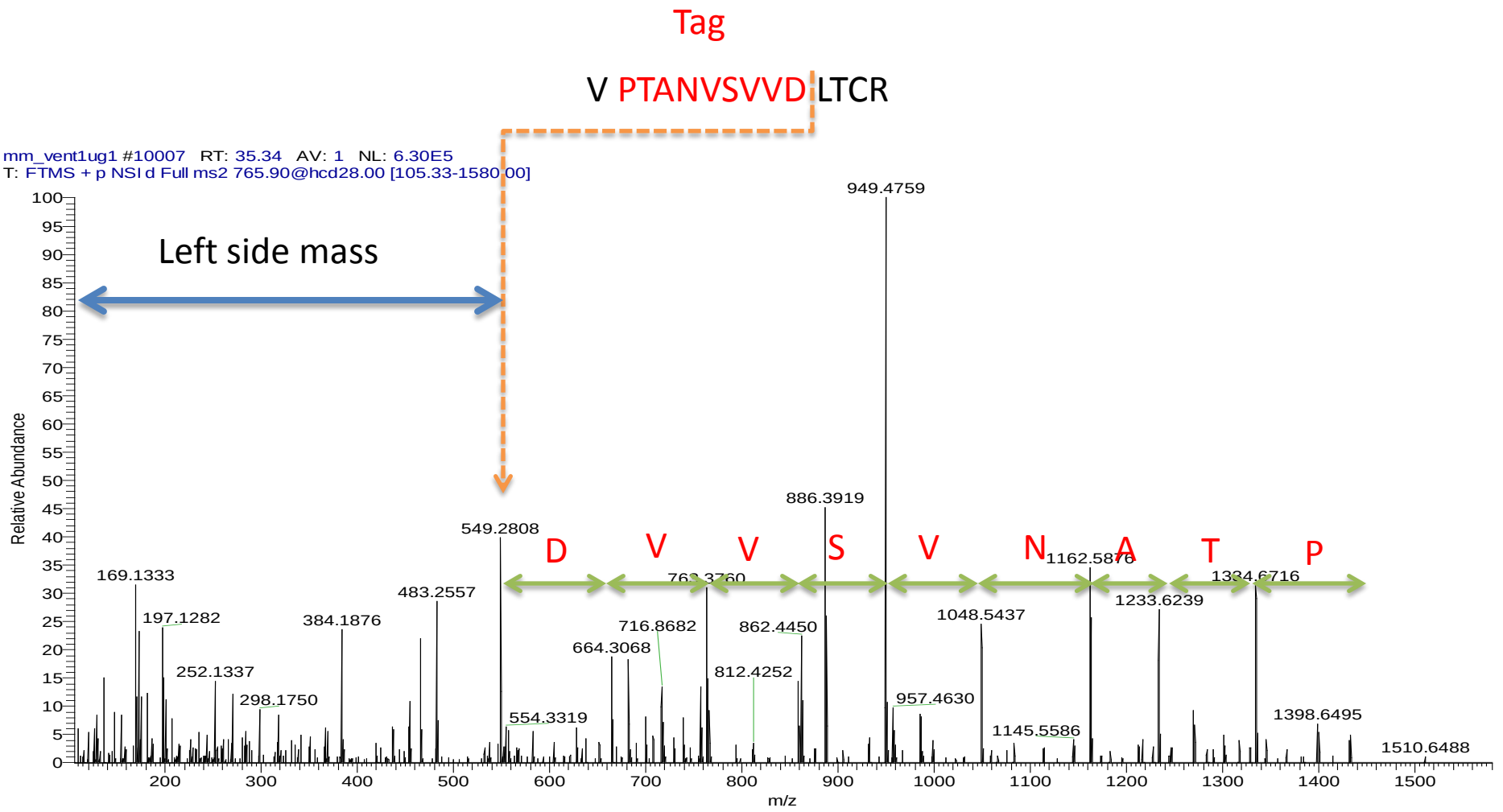

Raw file
mzXML
Spectra (.dta)
Charge determining
MS2 Deisotoping
Consolidation
Peak Normalization
Generating tags
Tag ranking
.fasta file
Generate decoy
Digestion
trypticity
miscleavage
Dynamic modification
Precursor mass
Mass tolerance
Mass Index file
Index files
protein index file
protein ID
Mass ID
Peptide index file
Peptide sequences
Sequence match
Flanking mass match
No selected peptide
Selected Peptide
pattern match
Theoretical spectra
Peptide scoring
Weight E scoring
.spout
.pepXML

De novo tag generation from MS/MS





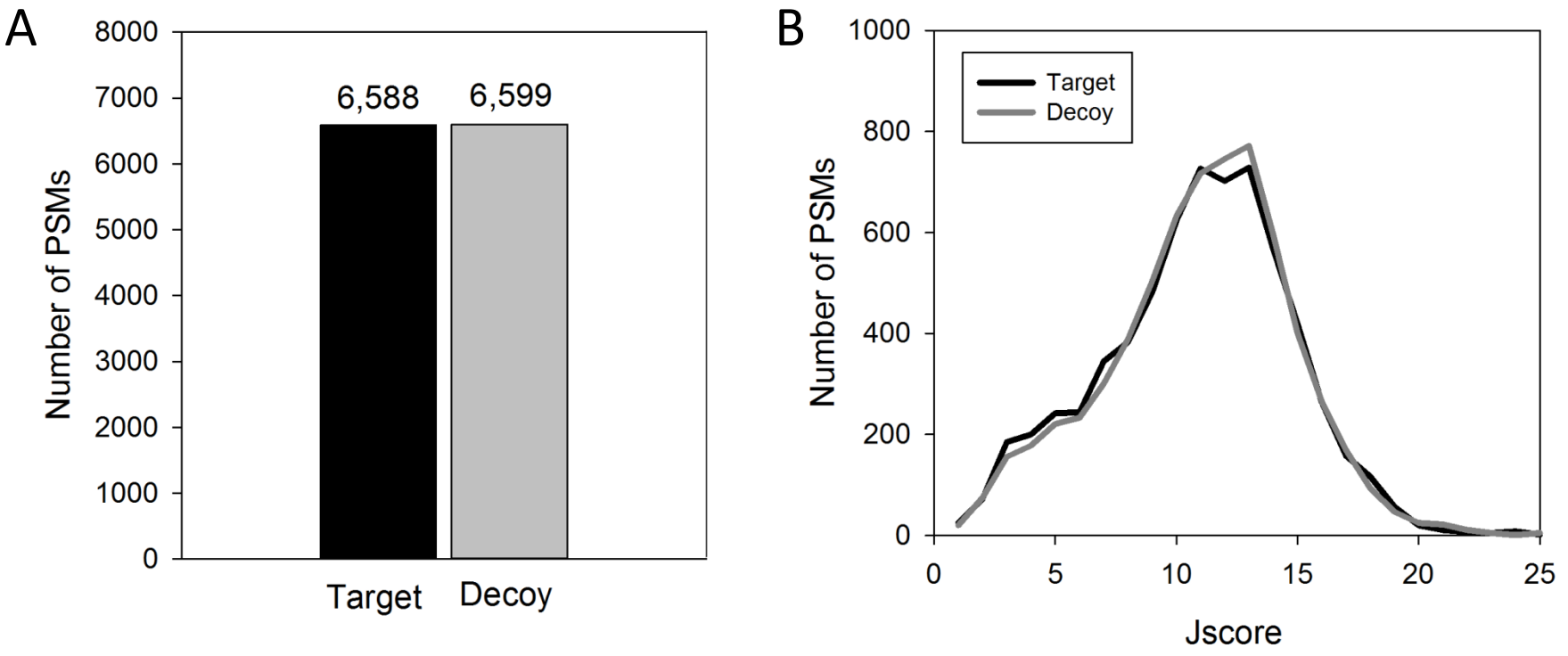
Flanking mass filtering

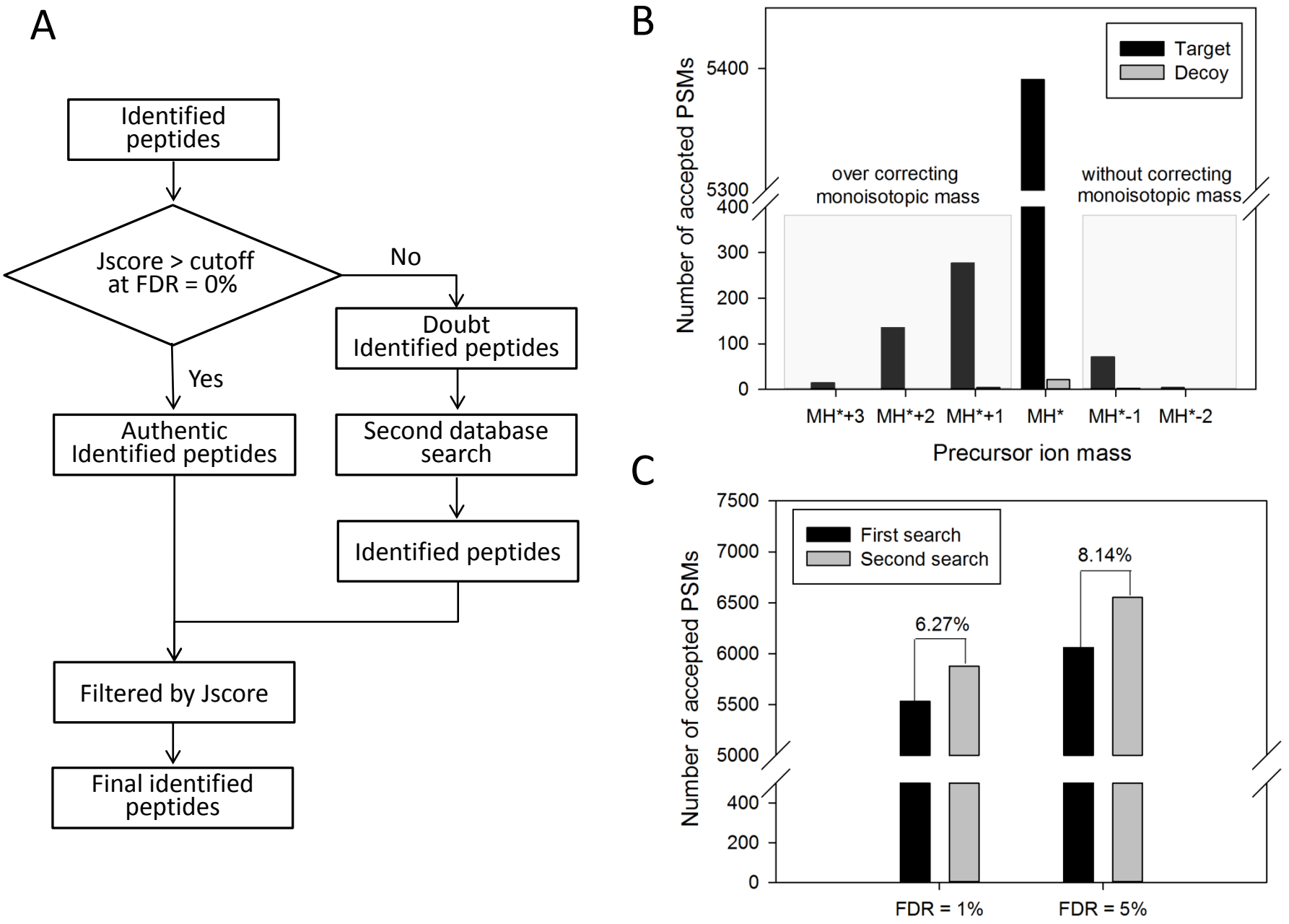


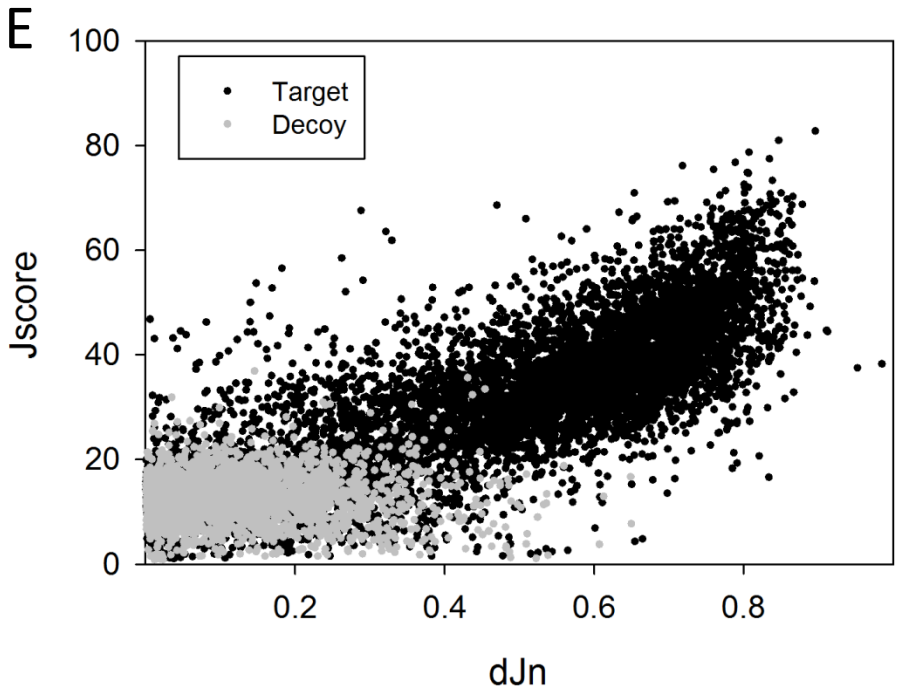
Evaluation of FDR by target-decoy

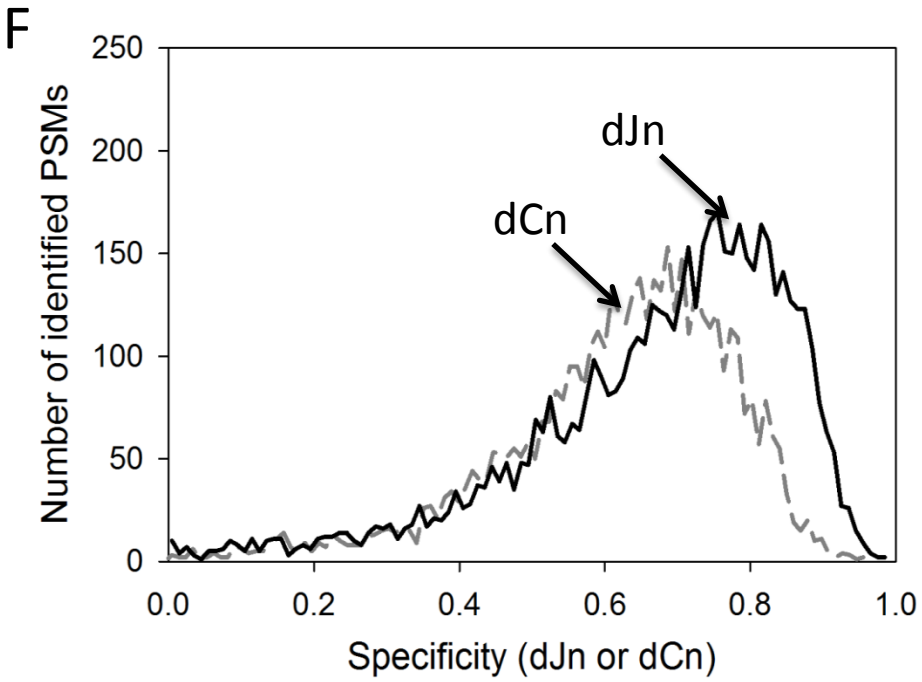
A simulated data set based on a real data:

The null dataset was falsified by increasing precursor ion mass of each MS/MS spectrum by 100 Da.




A
Identified peptides
Jscore > cutoff at FDR = 0%
No
Yes
Doubt Identified peptides
Authentic Identified peptides
Second database search
Identified peptides
Filtered by Jscore
Final identified peptides
B
Number of accepted PSMs
5400
5300
400
300
200
100
0
Target
Decoy
over correcting monoisotopic mass
without correcting monoisotopic mass
MH*+3
MH*+2
MH*+1
MH*
MH*-1
MH*-2
Precursor ion mass
C
Number of accepted PSMs
7500
7000
6500
6000
5500
5000
400
200
0
First search
Second search
6.27%
8.14%
FDR = 1%
FDR = 5%


E
Target
Decoy
Jscore
dJn
0
20
40
60
80
100
0.2
0.4
0.6
0.8


F
dJn
dCn
Number of identified PSMs
Specificity (dJn or dCn)
0
50
100
150
200
250
0.0
0.2
0.4
0.6
0.8
1.0

Let Us Meet Again

We welcome you all to our future conferences of OMICS Group International

Please Visit:

www.omicsgroup.com

www.conferenceseries.com

www.proteomicsconference.com