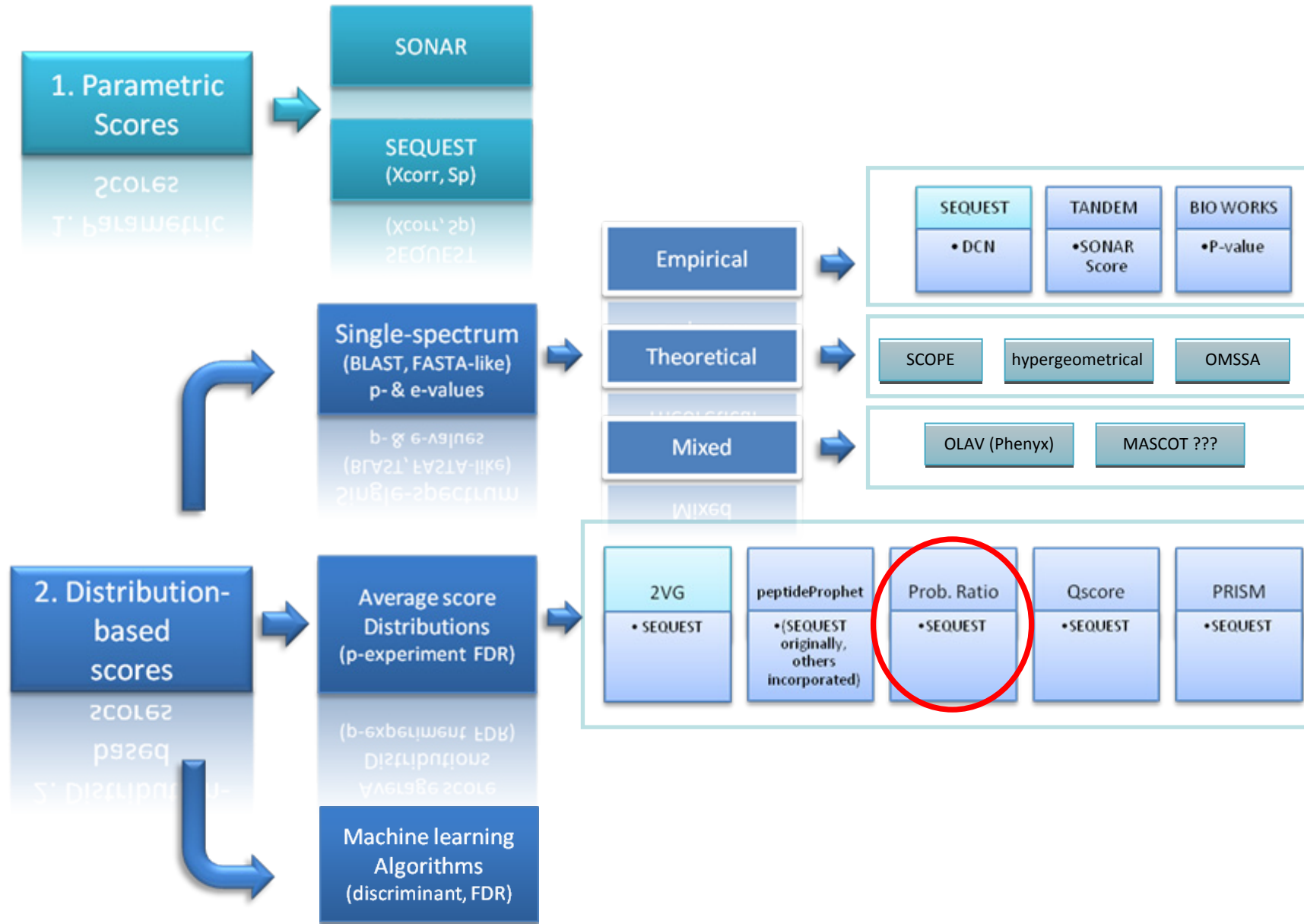


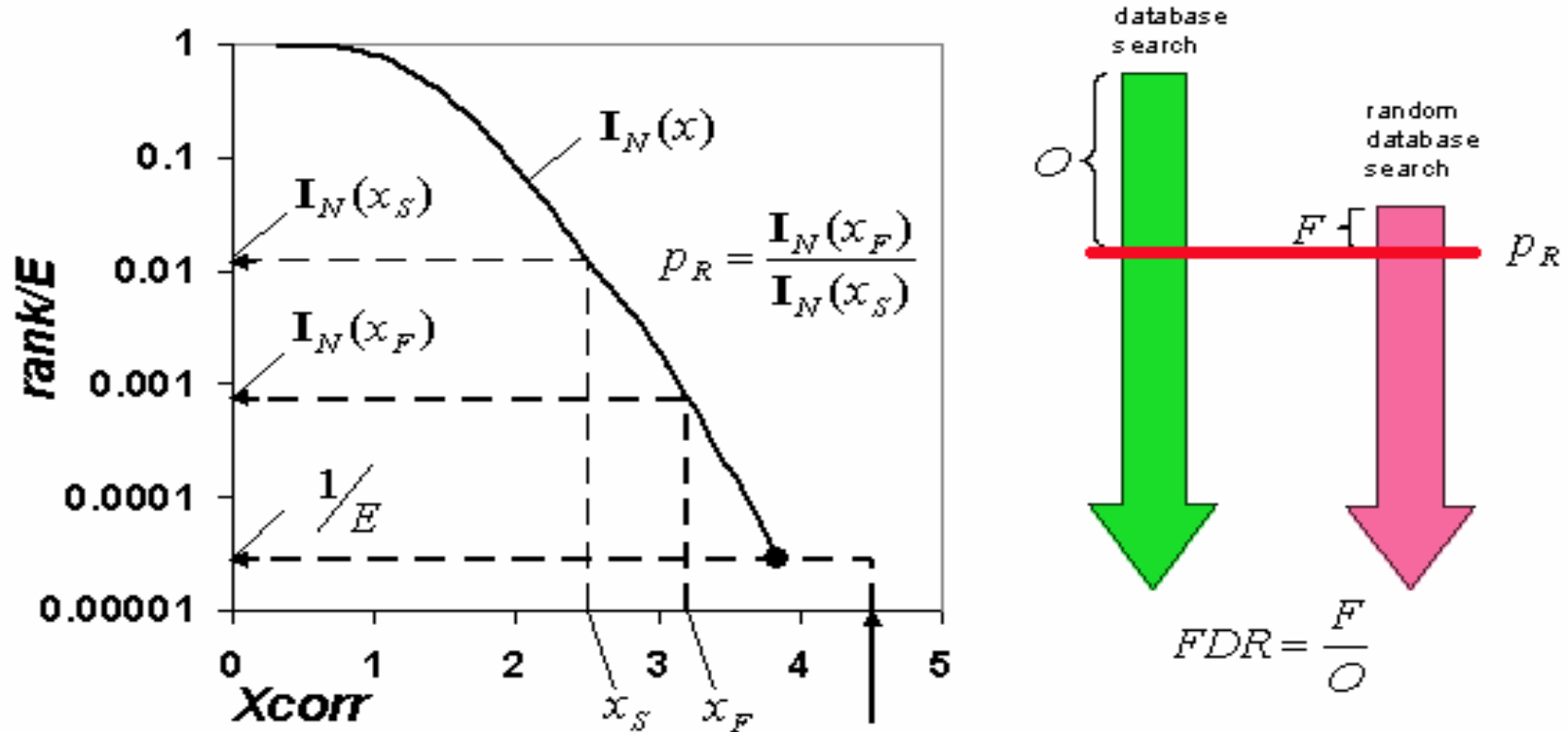
MS/MS searches & validation methods

The Probability Ratio Method

SCORES



A robust and non-parametric method: the probability ratio



Scans are corrected by a quality factor dependent on all the scans.

Particularly appropriated for fully automated, unattended peptide identification.

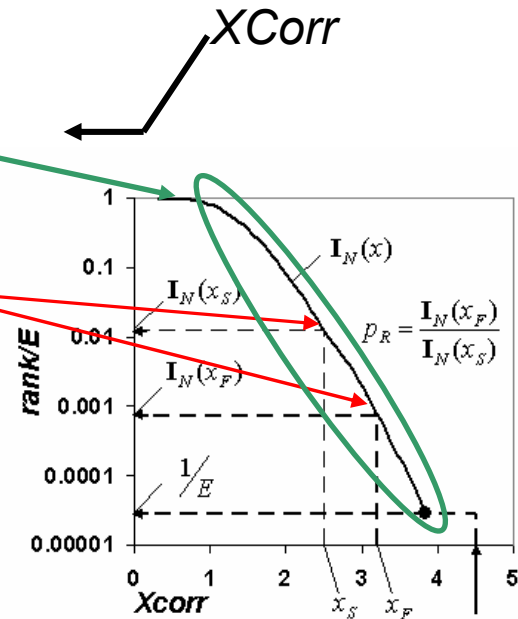
Probability ratio:

Combines characteristics from single-spectrum & average-score distributions

Spectrum #

	1	2	3	4	5	6	7	8	...	10.000
	6.71	6.01	5.64	5.31	3.20	3.18	3.13	1.65	1.60	1.35
	2.58	2.37	2.52	2.14	1.43	1.56	3.09	1.49	1.44	1.32
	2.51	2.19	2.39	2.07	1.34	1.47	2.92	1.41	1.38	1.22
	2.41	2.15	2.37	2.04	1.27	1.31	2.81	1.37	1.37	1.12
	2.33	2.09	2.30	1.96	1.25	1.30	2.80	1.33	1.33	1.11
	2.31	2.00	2.28	1.95	1.22	1.30	2.79	1.31	1.33	1.10
	2.29	2.00	2.24	1.94	1.20	1.28	2.78	1.28	1.26	1.07
	2.23	1.98	2.23	1.92	1.17	1.27	2.76	1.26	1.17	1.06
	2.21	1.97	2.12	1.90	1.17	1.27	2.74	1.25	1.12	1.03
	2.18	1.95	2.11	1.85	1.12	1.25	2.73	1.21	1.09	0.97

Scores



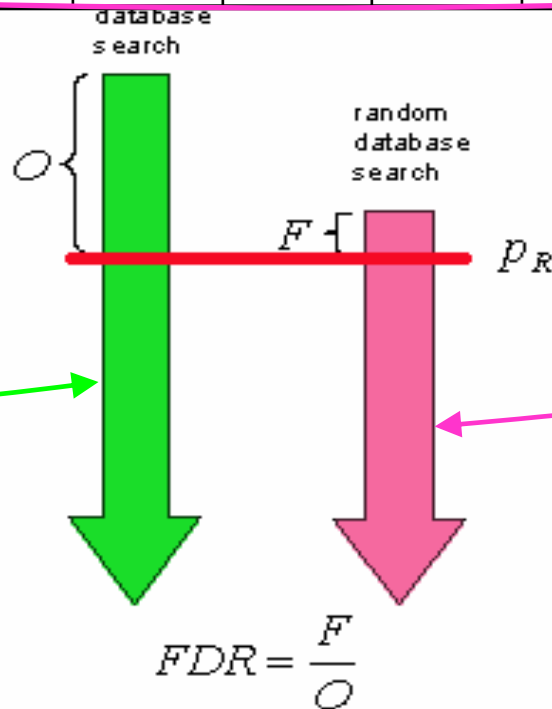
Target database search

Probability ratios

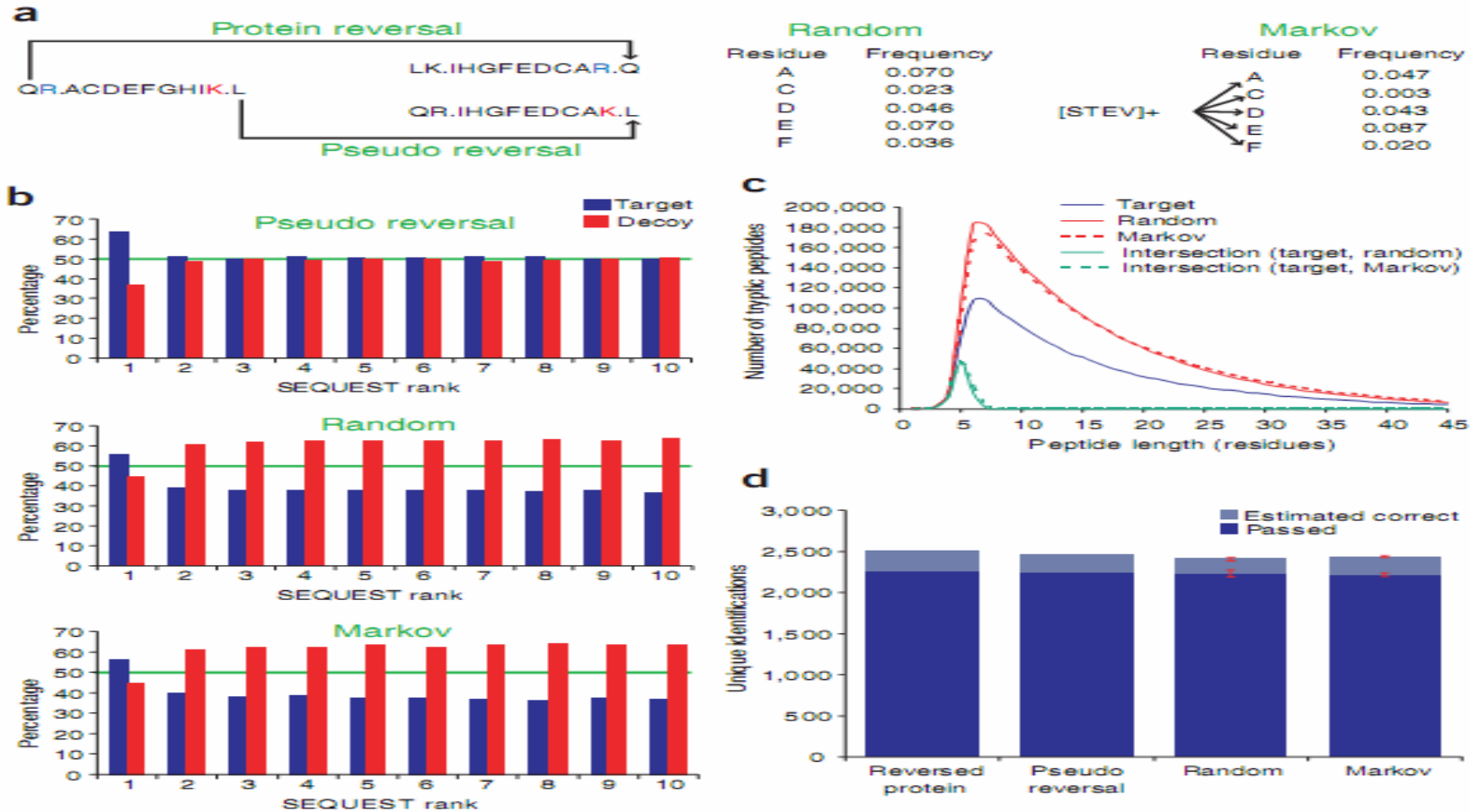
1	2	3	4	5	6	7	8	...	10.000
0.0001	0.0005	0.0007	0.2	0.3	0.5	0.57	0.63	0.96	1

Decoy (random) database search

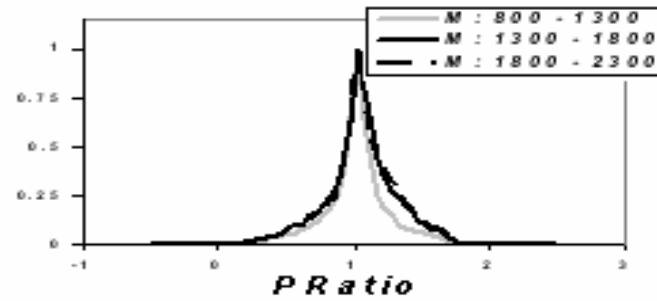
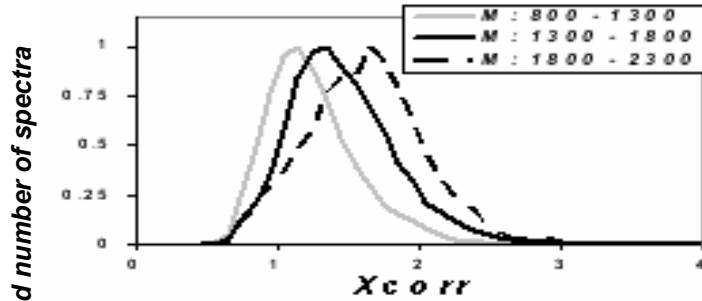
1	2	3	4	5	6	7	8	...	10.000
0.0005	0.05	0.1	0.2	0.32	0.54	0.61	0.73	0.97	1



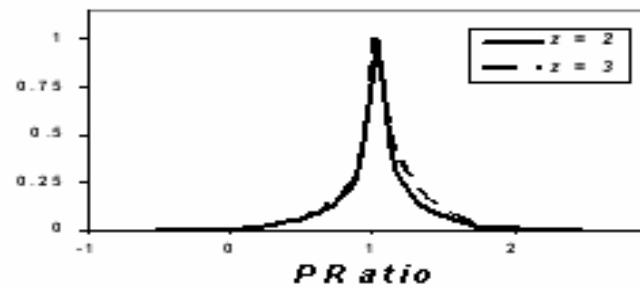
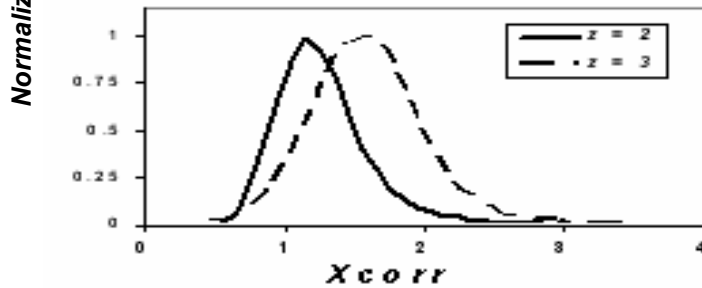
Decoy databases must be correctly chosen: the protein sequence reversal method



Pseudo reversal databases are demonstrated as the best decoy databases, compared with others



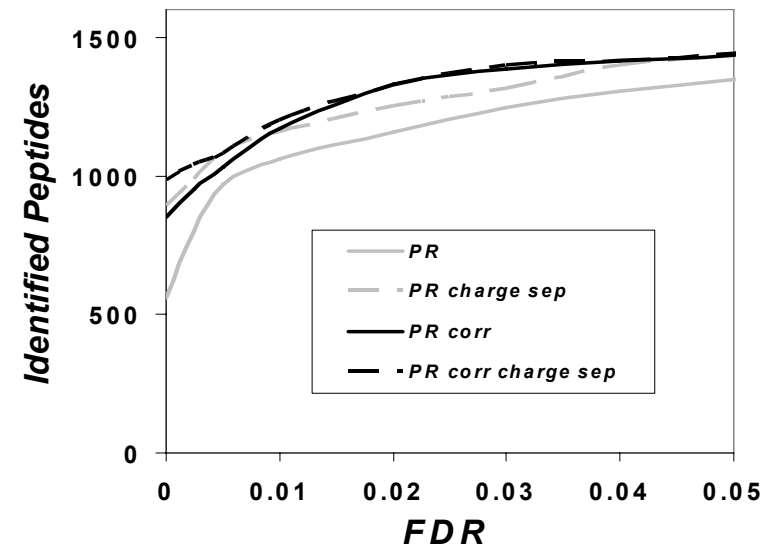
Dispersion due to peptide mass



Dispersion due to charge

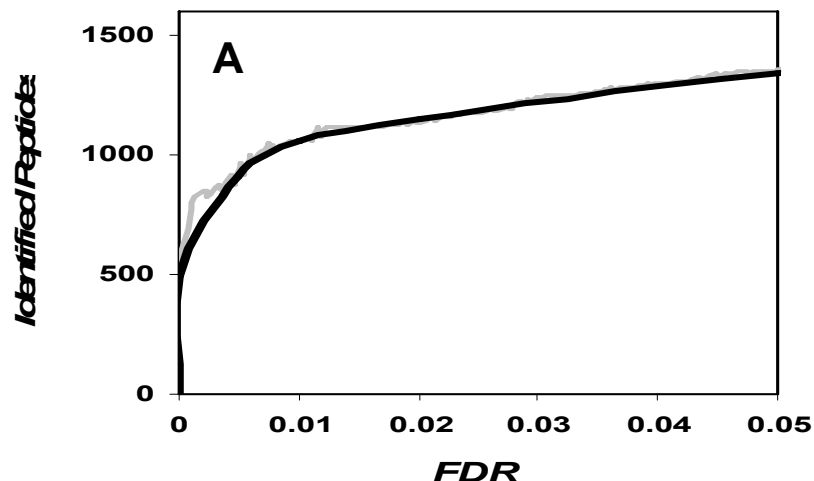
SEQUEST correlation values are clearly affected on its distribution by peptide mass and charge, therefore statistics must be done separately according to charge state and the peptide mass influences peptide identifications.

The probability ratios distribution seems to be centered, and does not depend on charge or peptide mass.

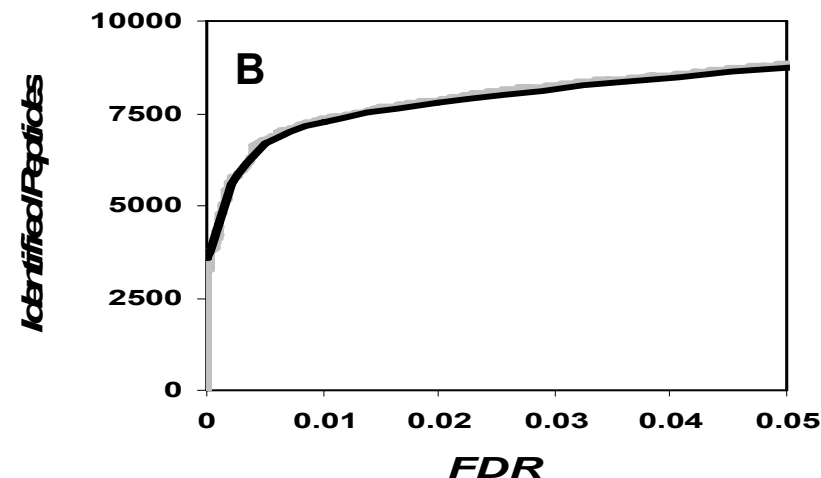


What about with non high-throughput analysis?

Sample A



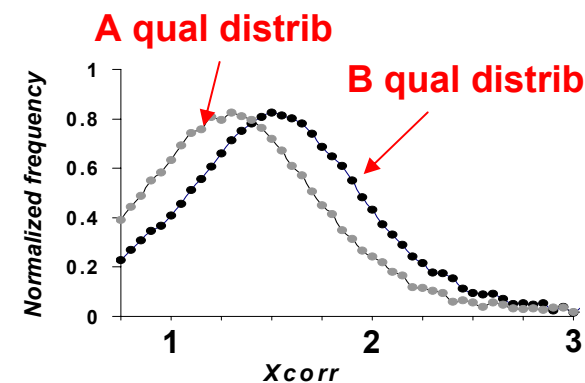
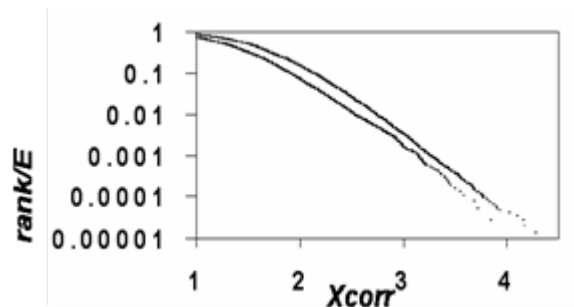
Sample B



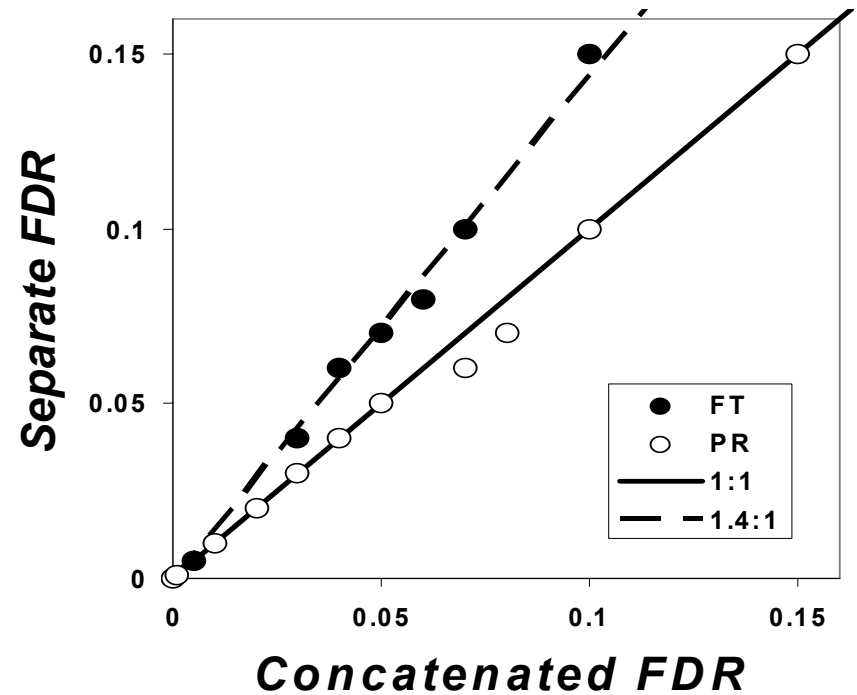
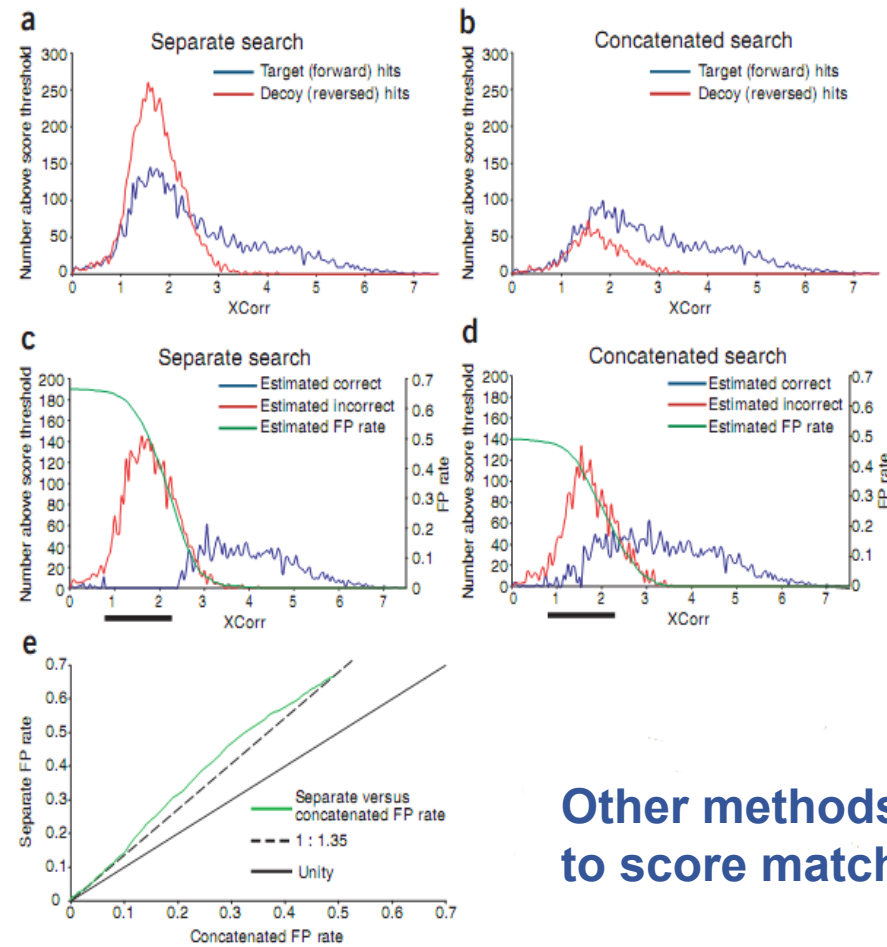
Two samples are evaluated using its own quality distribution...

...and swaping them (sample A evaluated with B generated quality distribution and viceversa). **Same results are obtained!**

We may use a different (greater!) sample as quality distribution



Concatenated or separated DDBB may be used with the probability ratio method

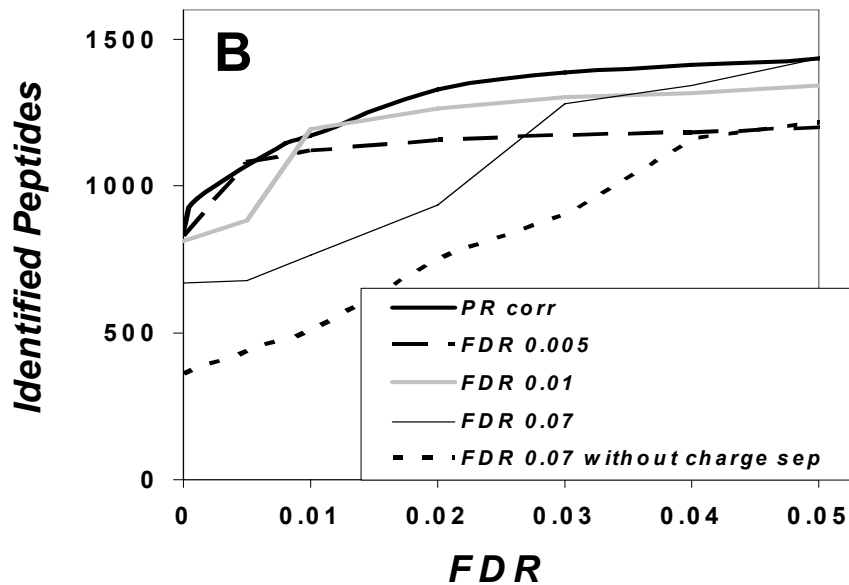
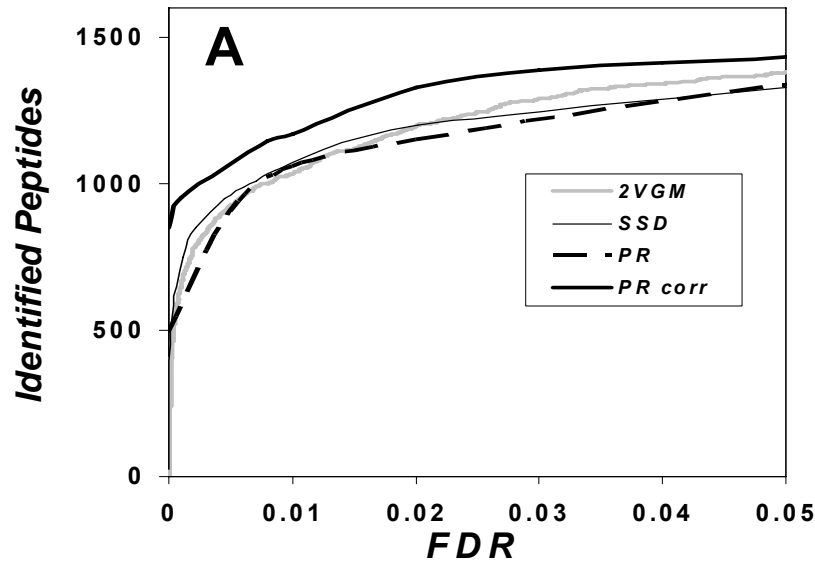


Other methods work better with concatenated DDBB due to score matches competition

Elias, J. E. and Gygi, S. P. (2007). *Nat Methods* 4(3): 207-14.

Martínez-Bartolomé, Navarro P, ... Vázquez, J. *Mol. Cellular Proteomics*, 2008

Comparative performance of the probability ratio method



The **probability ratio** gives a **better performance (or at least the same)** than:

- PeptideProphet (A. Keller, A.I. Nesvizhskii, E.Kolker, R.Aebersold, *Anal. Chem*, 2002),
- The Two-variable gaussian method (D.López, M.Villar, F.Martín-Maroto and J.Vázquez. *Anal.Chem*, 2004)
- The Iteratively fixed method (*Elias and Gygi, Nature Methods*, 2007)