Supporting Information

Ranking Fragment Ions Based on Outlier Detection for Improved Label-Free Quantification in Data-Independent Acquisition LC-MS/MS

Aivett Bilbao[1,2§], Ying Zhang[1§], Emmanuel Varesio[1§], Jeremy Luban[3], Caterina Strambio-De-Castillia[3], Frédérique Lisacek[2,4], Gérard Hopfgartner[1*]

1. Life Sciences Mass Spectrometry, School of Pharmaceutical Sciences, University of Geneva, University of Lausanne, Geneva, Switzerland.
2. Proteome Informatics Group, SIB Swiss Institute of Bioinformatics, Geneva, Switzerland.
3. Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA, USA.
4. Faculty of Sciences, University of Geneva, Geneva, Switzerland.

§ These authors contributed equally to this work.

* Corresponding author.

Table of contents

Table S1. Q1 windows with variable width used for SWATH acquisition of the DC dataset.

Figure S1. Calibration curves for all three technical replicates of selected example peptides from the SGS dataset.

Figure S2. Box-plots illustrating the distribution of intercepts for the SGS dataset.

Figure S3. Cumulative fraction of peptides as a function of the mean absolute error for the SGS dataset.

Figure S4. Box-plots illustrating the distributions of individual fragment ion intensities.

Figure S5. Distribution of the scores obtained for each attribute.

Figure S6. Spectronaut detection rate and NOFI rank distribution of the 806 fragment ions manually annotated as affected by interferences in the DC dataset.

Figure S7. Example of closely eluting peaks from the DC dataset.

Table S1. Q1 windows with variable width used for SWATH acquisition of the DC dataset.

#	*m/z* start	*m/z* stop	#	*m/z* start	*m/z* stop	#	*m/z* start	*m/z* stop	#	*m/z* start	*m/z* stop
1	350	402	10	575	591	19	685	703	28	875	915
2	401	433	11	590	603	20	702	720	29	914	951
3	432	459	12	602	614	21	719	738	30	950	987
4	458	481	13	613	624	22	737	758	31	986	1023
5	480	503	14	623	636	23	757	777	32	1022	1059
6	502	523	15	635	647	24	776	799	33	1058	1100
7	522	543	16	646	659	25	798	822	34	1099	1150
8	542	560	17	658	672	26	821	846	35	1149	1200
9	559	576	18	671	686	27	845	876	36	1199	1250

For SWATH acquisition, a set of 36 sequential Q1 variable isolation windows was used to cover the precursor m/z range of 350-1250 Da. The accumulation time for each SWATH experiment was 69 ms for a total cycle time of 2.5 sec.

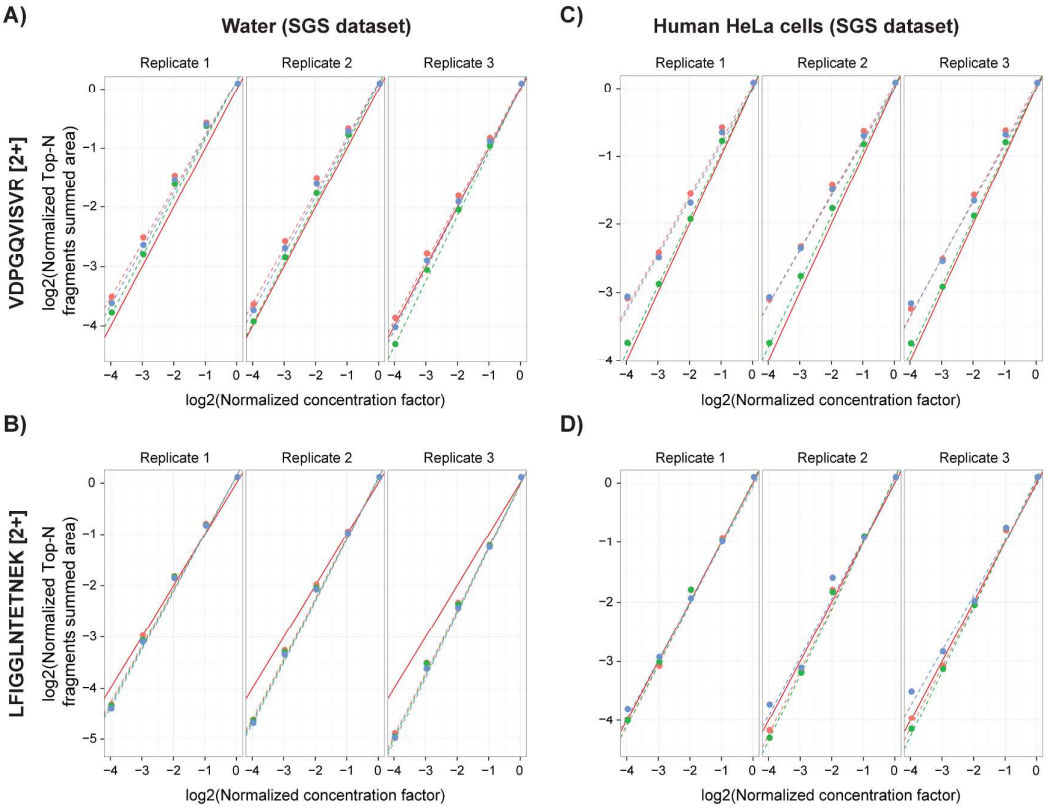


Figure S1. Calibration curves for all three technical replicates of selected example peptides illustrating the background effect and impact of interferences for quantification of peptides from the SGS dataset. Top panels (A and C) show peptide VDPGQVISVR [2+] and bottom panels (B and D) peptide LFIGGLNTETNEK [2+] spiked respectively in water (A and B) or in human HeLa cell lysate digests (C and D). Linear regressions were computed using the Top-3 most intense fragment ions obtained from the library (red dashed line), from the NOFI ranking algorithm (green dashed line), as well as for all ten-fragment ions from the library (blue dashed line). *Dots show the summed areas of the selected fragment ions with their corresponding regression lines (dashed lines). The solid red line shows the theoretical 1:1 dilution curve.*

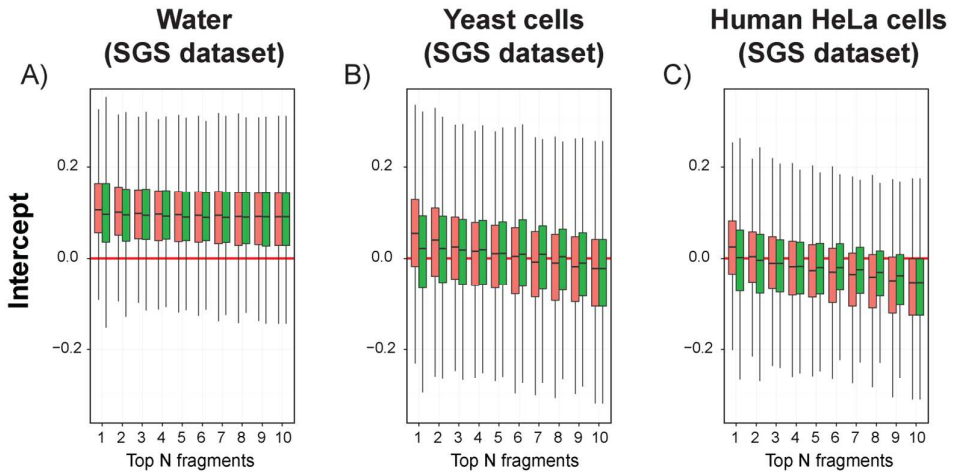
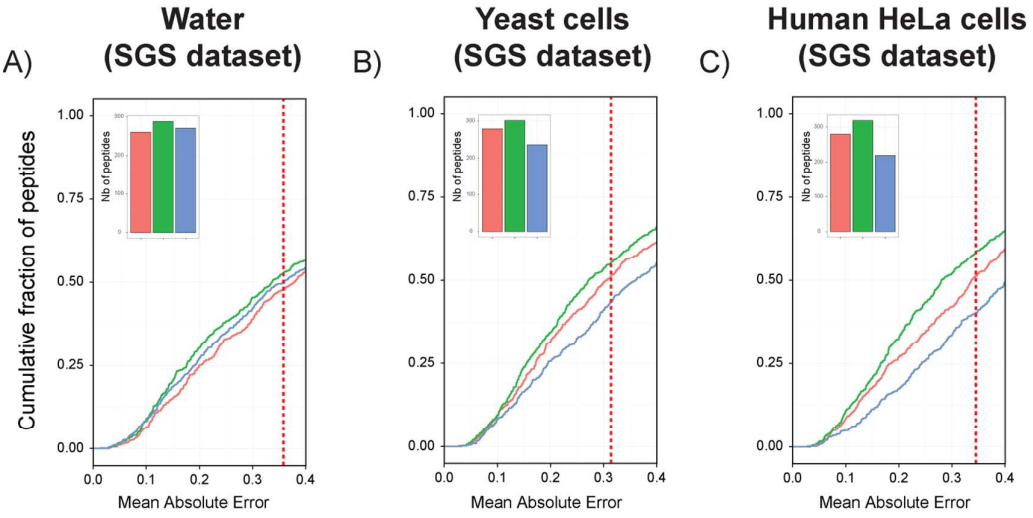


Figure S2. Box-plots illustrating the distribution of intercepts calculated with the library intensity (red) or the NOFI ranking methods (green). Results are obtained from 182 peptides selected from the SGS dataset spiked in water (A), yeast cells (B) or human HeLa cells (C) lysate digests. *Solid horizontal red lines indicate expected values according to the theoretical 1:1 dilution curve.*



NUMBER OF PEPTIDES QUANTIFIED WITH AN AVERAGE ERROR BELOW CUTOFF

A) WATER

Top-N	Library	NOFI	10Frags	Error cutoff	Diff. NOFI-Library	NOFI vs Library	NOFI vs 10Frags	Library vs 10Frags
1	246	295	278	0.367	49	20%	6%	-12%
2	254	290	275	0.361	36	14%	5%	-8%
3	260	288	271	0.358	28	11%	6%	-4%
4	266	283	270	0.355	17	6%	5%	-1%
5	270	279	270	0.353	9	3%	3%	0%
6	272	277	270	0.352	5	2%	3%	1%
7	270	279	270	0.352	9	3%	3%	0%
8	274	274	271	0.356	0	0%	1%	1%
9	272	275	271	0.356	3	1%	1%	0%
10	273	273	273	0.360	0	0%	0%	0%

B) YEAST DIGEST

Top-N	Library	NOFI	10Frags	Error cutoff	Diff. NOFI-Library	NOFI vs Library	NOFI vs 10Frags	Library vs 10Frags
1	260	317	242	0.318	57	22%	31%	7%
2	276	309	234	0.311	33	12%	32%	18%
3	280	302	237	0.314	22	8%	27%	18%
4	282	299	238	0.314	17	6%	26%	18%
5	268	303	248	0.322	35	13%	22%	8%
6	271	300	248	0.324	29	11%	21%	9%
7	266	304	249	0.328	38	14%	22%	7%
8	266	298	255	0.333	32	12%	17%	4%
9	272	285	262	0.343	13	5%	9%	4%
10	273	273	273	0.361	0	0%	0%	0%

C) HUMAN HELA CELL DIGEST

Top-N	Library	NOFI	10Frags	Error cutoff	Diff. NOFI-Library	NOFI vs Library	NOFI vs 10Frags	Library vs 10Frags
1	266	332	221	0.348	66	25%	50%	20%
2	274	325	220	0.347	51	19%	48%	25%
3	280	319	220	0.345	39	14%	45%	27%
4	266	328	225	0.349	62	23%	46%	18%
5	269	324	226	0.353	55	20%	43%	19%
6	268	322	229	0.357	54	20%	41%	17%
7	271	313	235	0.365	42	15%	33%	15%
8	265	308	246	0.373	43	16%	25%	8%
9	260	300	259	0.390	40	15%	16%	0%
10	273	273	273	0.404	0	0%	0%	0%

Figure S3. Cumulative fraction of peptides as a function of the mean absolute error for the SGS dataset peptides spiked in water (A), yeast cell lysate (B) or human HeLa cell lysate (C) digests. Plots show results for the Top-3 ranked fragment ions using either the library intensity ranking method (red) or the NOFI ranking algorithm (green). In blue are indicated the results with all the 10 fragment ions. Dashed red lines in the plots show the error cutoff and the corresponding numbers of peptides (selected rows in tables) are shown in the bar charts (insets). Tables show the number of peptides quantified within the error cutoff for N ranging from 1 to 10, as well as the comparison between the different ranking methods. Each error cutoff

corresponds to the median value of the respective Top-N overall distribution, regardless of the ranking method.

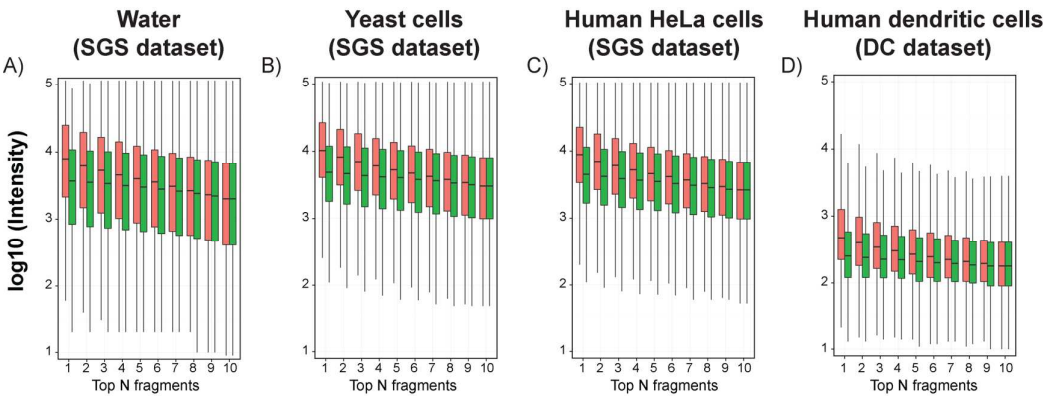


Figure S4. Box-plots illustrating the distributions of individual fragment ion intensities ranked with the library intensity (red) or the NOFI ranking methods (green) for 182 SGS dataset selected peptides spiked in water (A), yeast cell lysate (B) or human HeLa cell lysate (C) digests, as well as for the 2284 DC dataset peptides (D).

Note: The TOF accumulation time was 100ms for the SGS dataset and 69ms for the DC dataset.

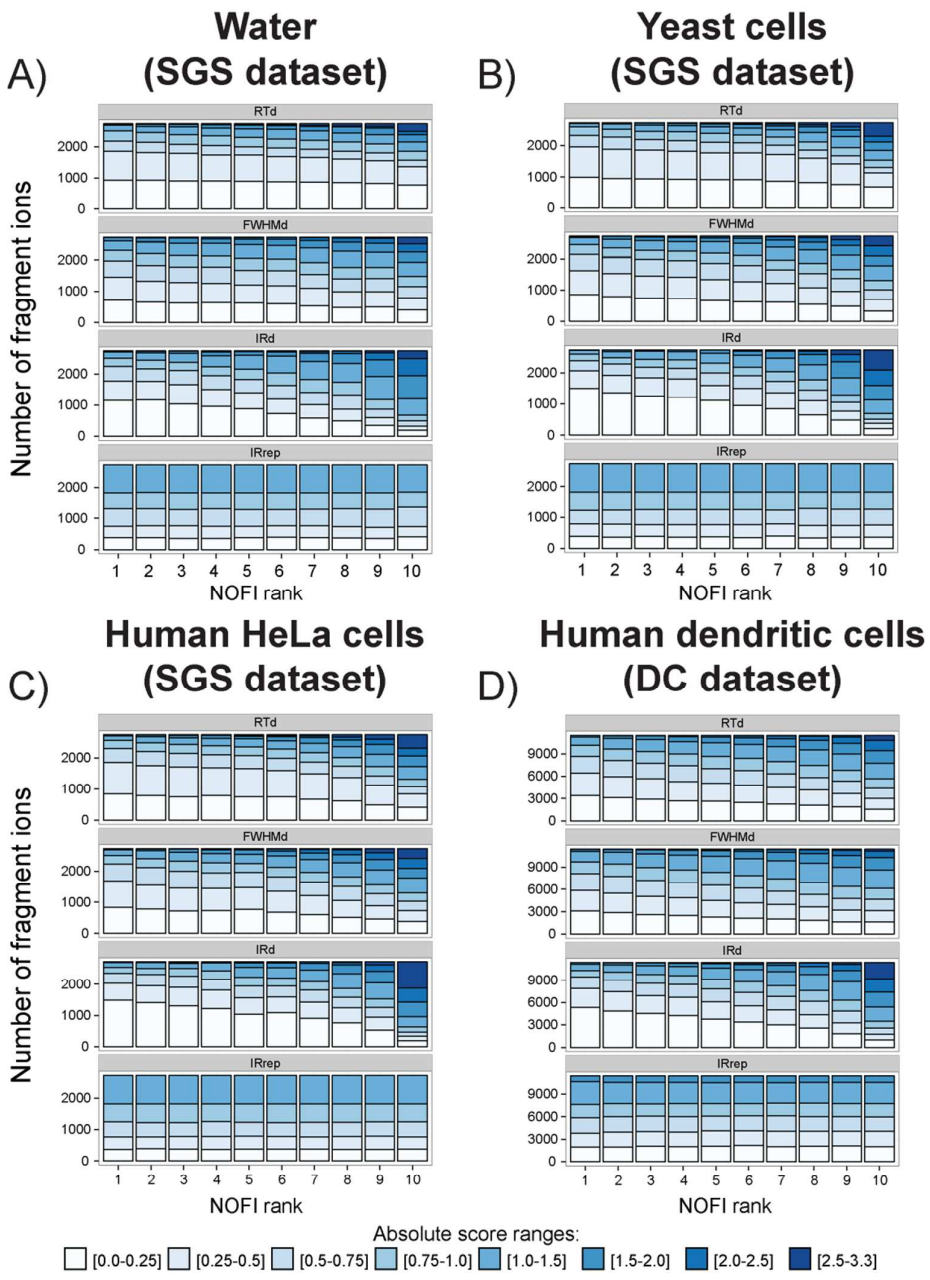


Figure S5. Distribution of the scores obtained for each attribute used in the NOFI ranking algorithm applied to the different backgrounds and datasets: i.e. water (A), yeast cell lysate digest (B) and human HeLa cell lysate digest (C) from the SGS dataset, as well as for the DC dataset (D). *The color intensity scale corresponds to eight absolute score ranges, dark blue being the range with a higher likelihood of outlier detection.*

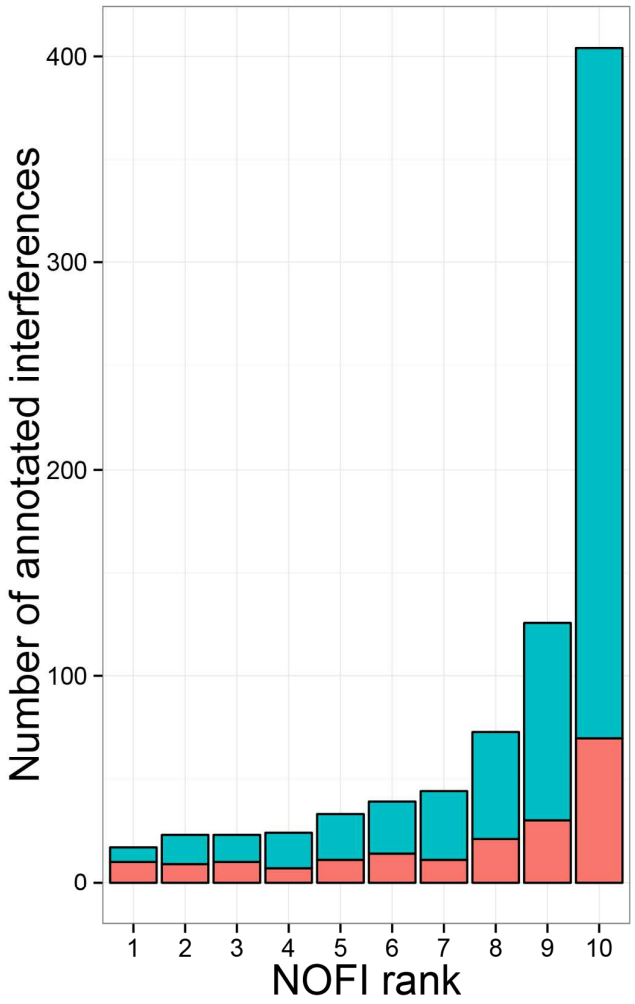
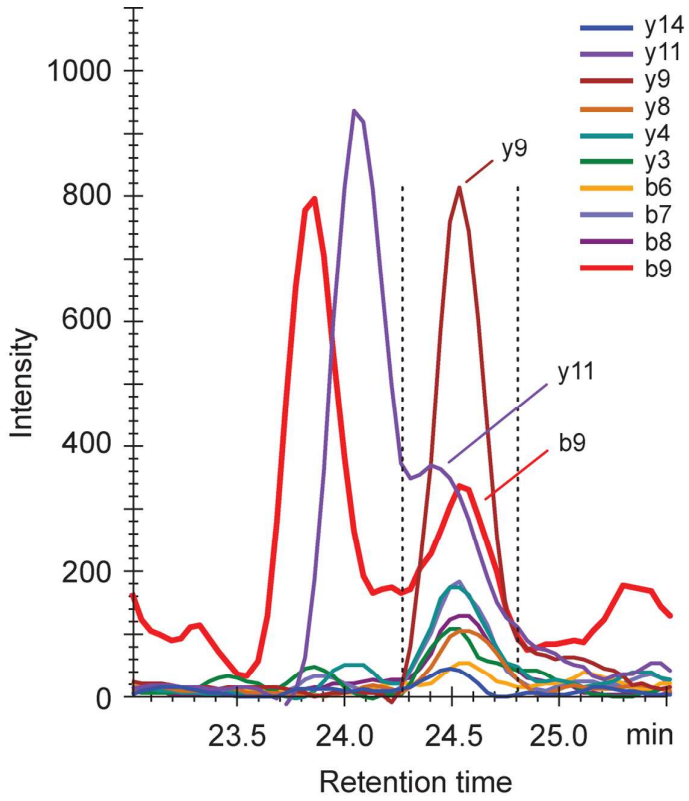


Figure S6. Spectronaut detection rate and NOFI rank distribution of the 806 fragment ions manually annotated as affected by interferences in the DC dataset. The colors of the bars show the detection rate by Spectronaut: green = detected and red = undetected. The complete DC dataset of 2284 peptides was processed by both algorithms independently.



Fragment ion	NOFI rank	Library rank	RTd	FWHMd	IRd	IRrep
b9	1	5	-0.12	0.52	0.54	-1.48
y4	2	4	-0.57	-0.30	-0.23	-1.51
y3	3	9	-0.57	-0.16	0.06	1.03
b8	4	7	-0.12	-0.16	-0.06	-0.26
b7	5	10	0.88	0.38	1.39	0.35
y14	6	6	-0.12	-1.80	-0.93	-0.34
y9*	7	1	**0.88**	-0.16	**1.87**	-0.05
y8	8	2	1.44	1.47	-0.83	-1.40
b6	9	8	0.43	-1.12	-1.16	-0.97
y11	10	3	**-2.13**	**1.33**	0.73	-1.20

Figure S7. Example of closely eluting peaks from the DC dataset (i.e. shoulder peaks and partially resolved neighboring peaks) that were manually annotated as potential interferences for the y11 and b9 fragment ions of peptide LSLEGDHSTPPSAYGSVK [2+]. The table shows the fragment ranks obtained by the library intensity method and by the NOFI ranking algorithm, as well as the scores of the individual attributes. *The two vertical lines represent the retention time boundaries applied by the Skyline software.*

**Note: The y9 fragment is not ranked first because its IRd attribute shows the presence of interference although the peak shape looks acceptable.*