J. Proteome Res.

Expanding Proteome Coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) Combined with Broad Specificity Proteolysis

Simon Davis[1#], Philip D. Charles[1#], Lin He[2], Peter Mowlds[3], Benedikt M. Kessler[1], Roman Fischer[1]*

[1]Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Roosevelt Drive, Oxford, OX3 7FZ, UK
[2]Bioinformatics Solutions Inc., 470 Weber St. N. Suite 204 Waterloo, ON Canada N2L 6J2
[3]Thermo Fisher Inc., Hemel Hampstead, London, UK

[#]equal contributions
*correspondence: roman.fischer@ndm.ox.ac.uk
Phone: +44 (0) 1865 612935

Running title: Parallel ion analysis MS enhances global proteome coverage
Key Words: Deep Proteome, LC-MS/MS, Sequence coverage, Isoform profiling, Protein sequence coverage

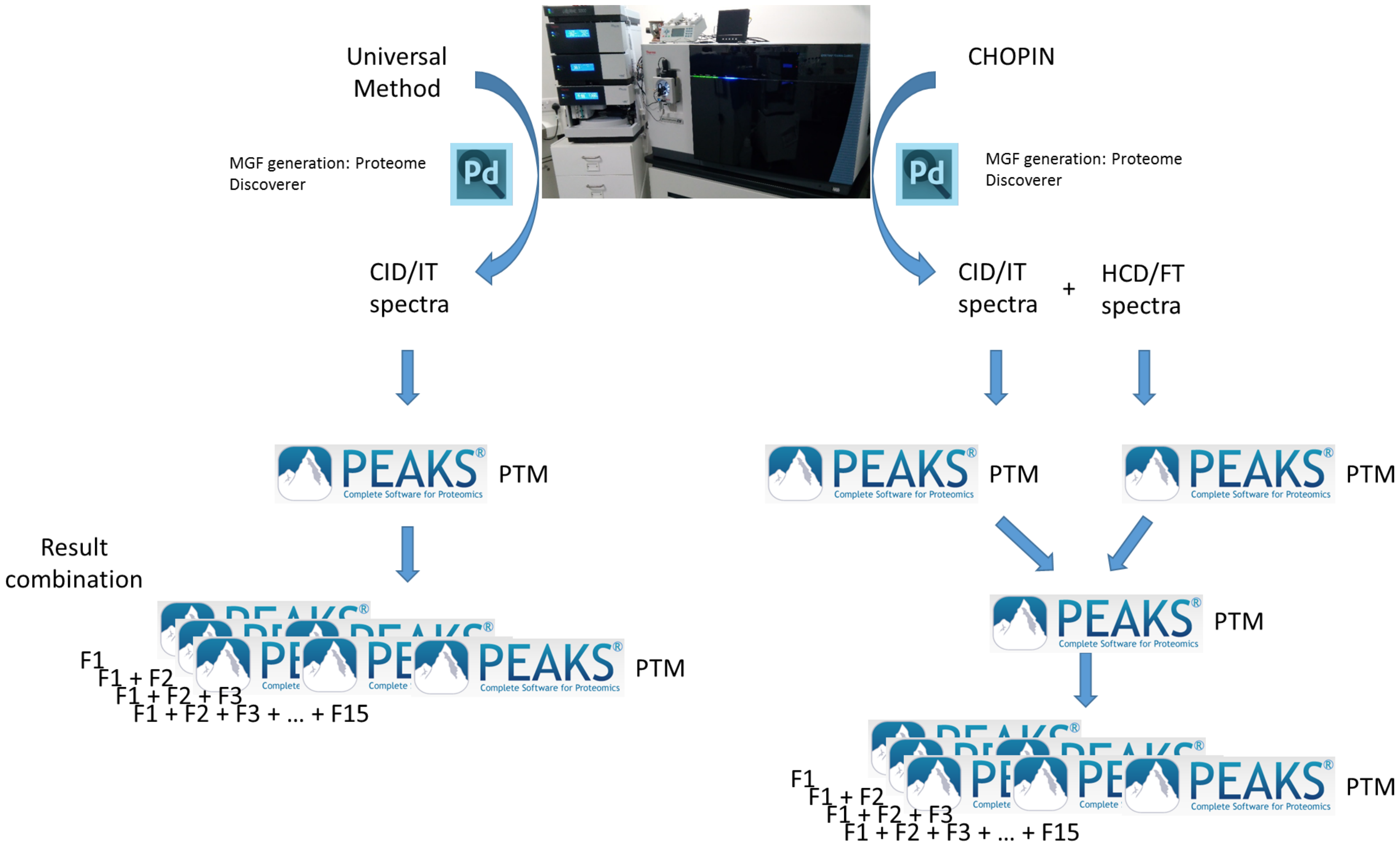
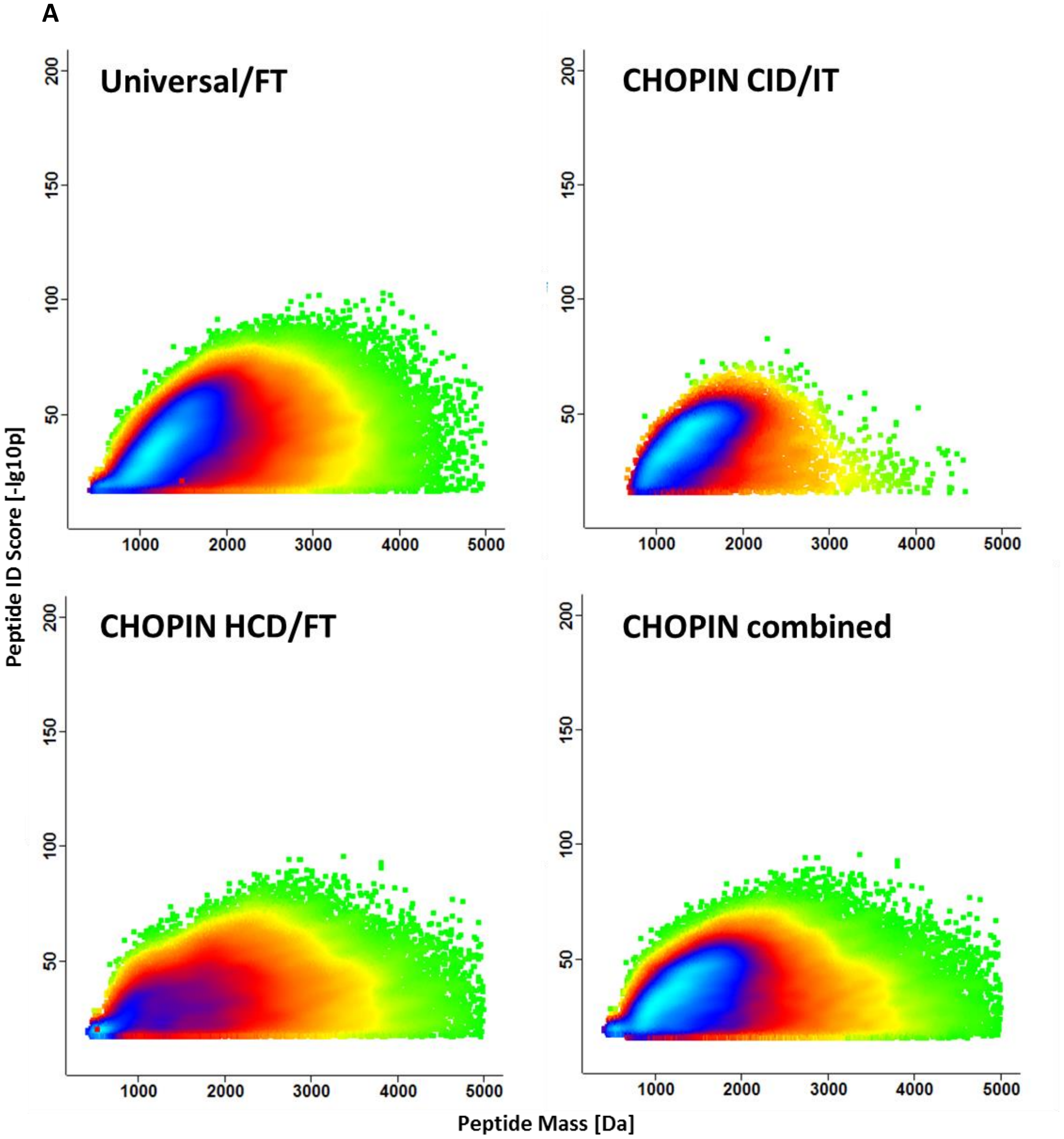
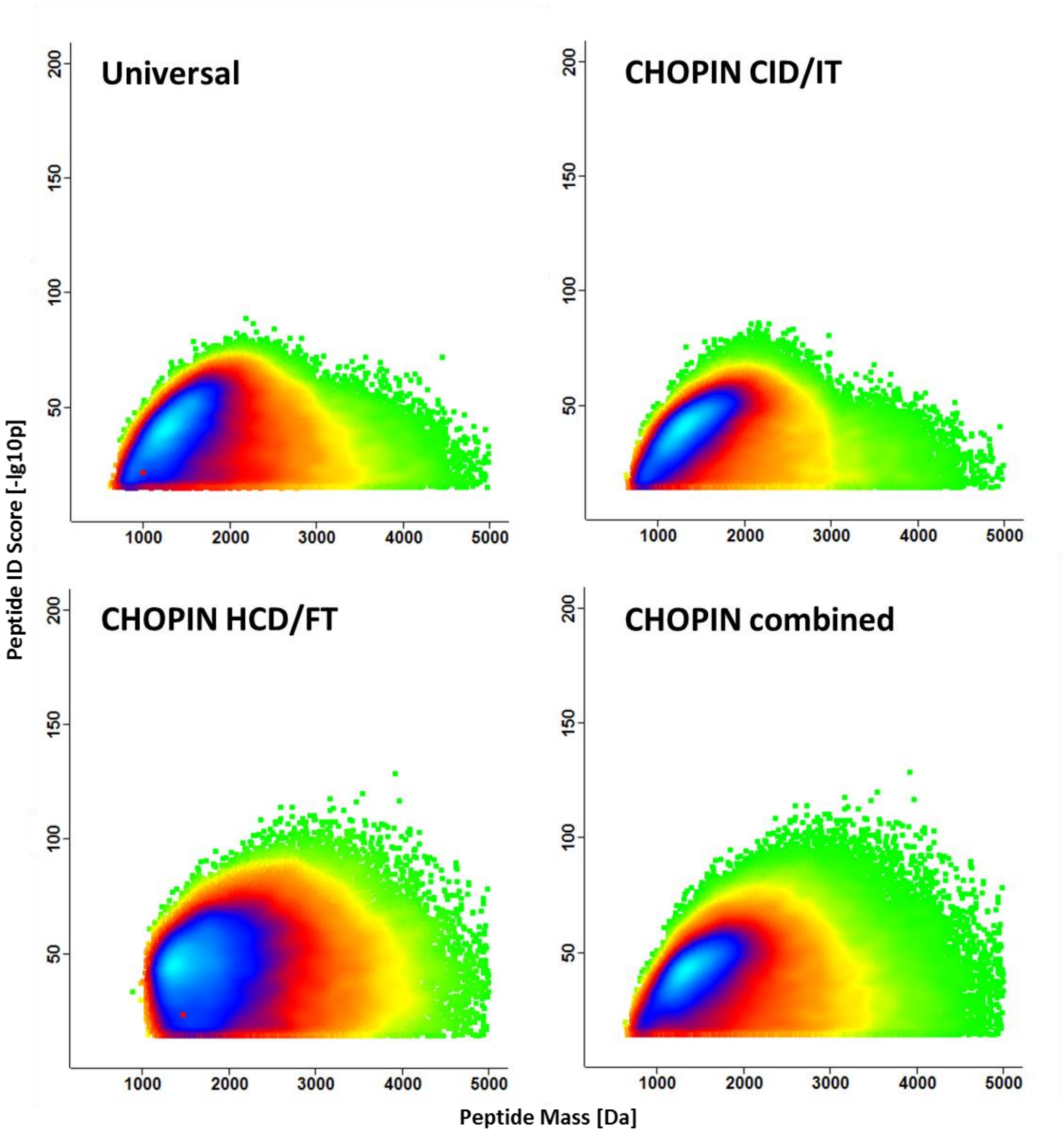

Universal Method
CHOPIN
MGF generation: Proteome Discoverer
Pd
MGF generation: Proteome Discoverer
Pd
CID/IT spectra
CID/IT spectra
+
HCD/FT spectra
PEAKS Complete Software for Proteomics PTM
PEAKS Complete Software for Proteomics PTM
PEAKS Complete Software for Proteomics PTM
Result combination
PEAKS Complete Software for Proteomics PTM
F1
F1 + F2
F1 + F2 + F3
F1 + F2 + F3 + … + F15
PEAKS Complete Software for Proteomics PTM
PEAKS Complete Software for Proteomics PTM
F1
F1 + F2
F1 + F2 + F3
F1 + F2 + F3 + … + F15

Figure S1 Data analysis strategy for the deconvolution of CID/IT HCD/FT hybrid data generated by CHOPIN. *MGF files were generated from data acquired using the CHOPIN and Universal method, and data sets processed using Proteome Discoverer V.2.1. MS/MS generated by CHOPIN were separated into HCD/FT and CID/IT spectra to allow analysis appreciating the different mass accuracies of Orbitrap and Ion Trap detectors. Data was analyzed in PEAKS 7.5 and results from prefractionated samples were further combined as described in the Methods section (Orbitrap Fusion Lumos photo by RF).*


A
Universal/FT
CHOPIN CID/IT
CHOPIN HCD/FT
CHOPIN combined
Peptide ID Score [-lg10p]
Peptide Mass [Da]
50
100
150
200
1000
2000
3000
4000
5000

38



B B

Figure S2 Score distribution of peptides identified in Elastase and Post Digest Mix samples with Universal and CHOPIN methods. *Analogous to Fig. 2A, we plotted the peptide score distributions for the elastase digested sample (**A**) and the Post Digest Mix (**B**) following MS analysis with the Universal/FT Method and CHOPIN. The CHOPIN HCD/FT spectra in the elastase digest also include singly charged ions while the Universal/FT method in this case was modified only to use HCD/FT fragmentation (see Methods section). Scores and success rates are generally lower with CHOPIN, but are compensated by the number of spectra (compare Fig. S3). The Post Digest Mix shows a very similar peptide score distribution as compared to the tryptic sample (Fig. 2A). CHOPIN benefits from the high scoring HCD/FT spectra yielding results of higher quality as compared to the Universal method. The separately analyzed unlinked Post Digest Mix fractions (**C**) result in a further increased number of peptide identifications with a similar score distribution at the expense of increased sample analysis time.*

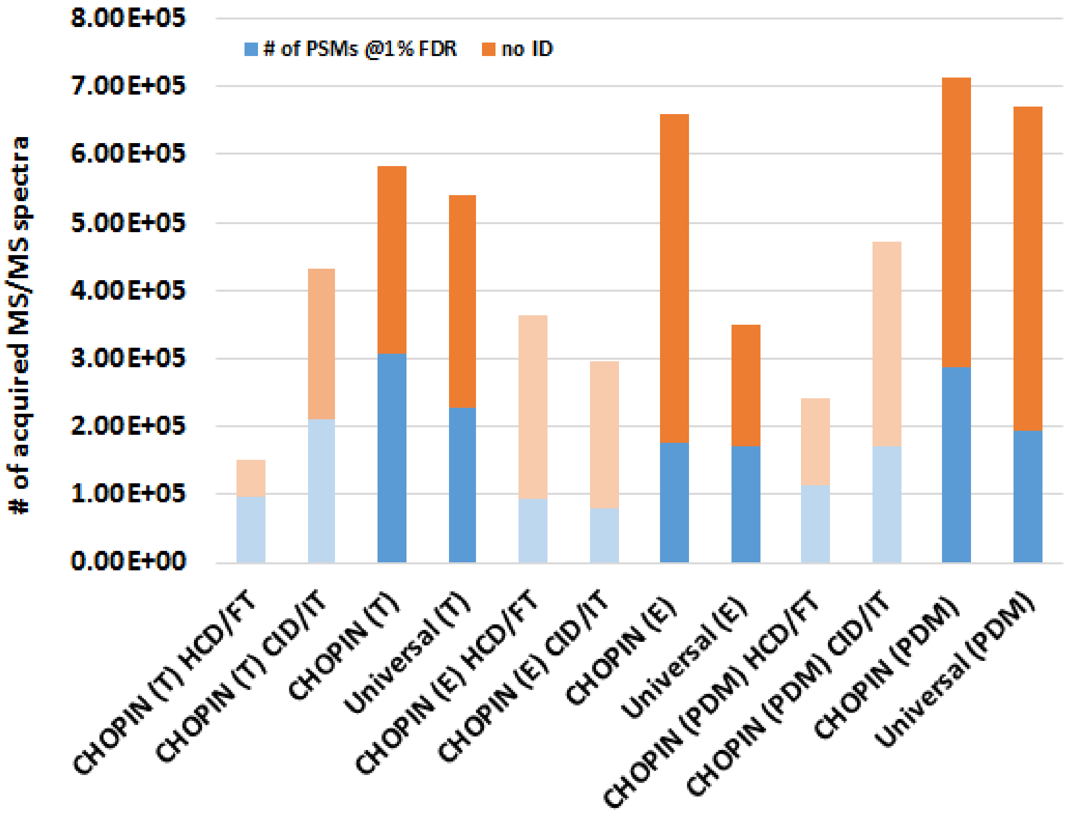


Figure S3 MS/MS analysis statistics using CHOPIN or the Universal method. *CHOPIN spectra were separated according to the used detector (pale colors) and split dependent on whether a peptide was identified (blue) or not (orange). The comparison between CHOPIN and the Universal method (intense colors) shows different success rates for Trypsin (T), Elastase (E) and Post Digest Mix (PDM). In the tryptic samples we generally see more acquired MS/MS spectra and a significantly higher success rate when CHOPIN is used. The "free" HCD/FT scans improve the results not only by generating more IDs but also by providing better quality MS/MS spectra. In the vastly more complex elastase sample the modified Universal method using only HCD/FT MS/MS scans benefits from the high mass accuracy when data is searched in PEAKS. However, CHOPIN leads to a similar number of identified spectra and a comparable number of identified proteins (compare to Tables 1 and S3) due to higher sampling frequency, demonstrating the benefit of high resolution MS/MS spectra for no-enzyme database searches.*

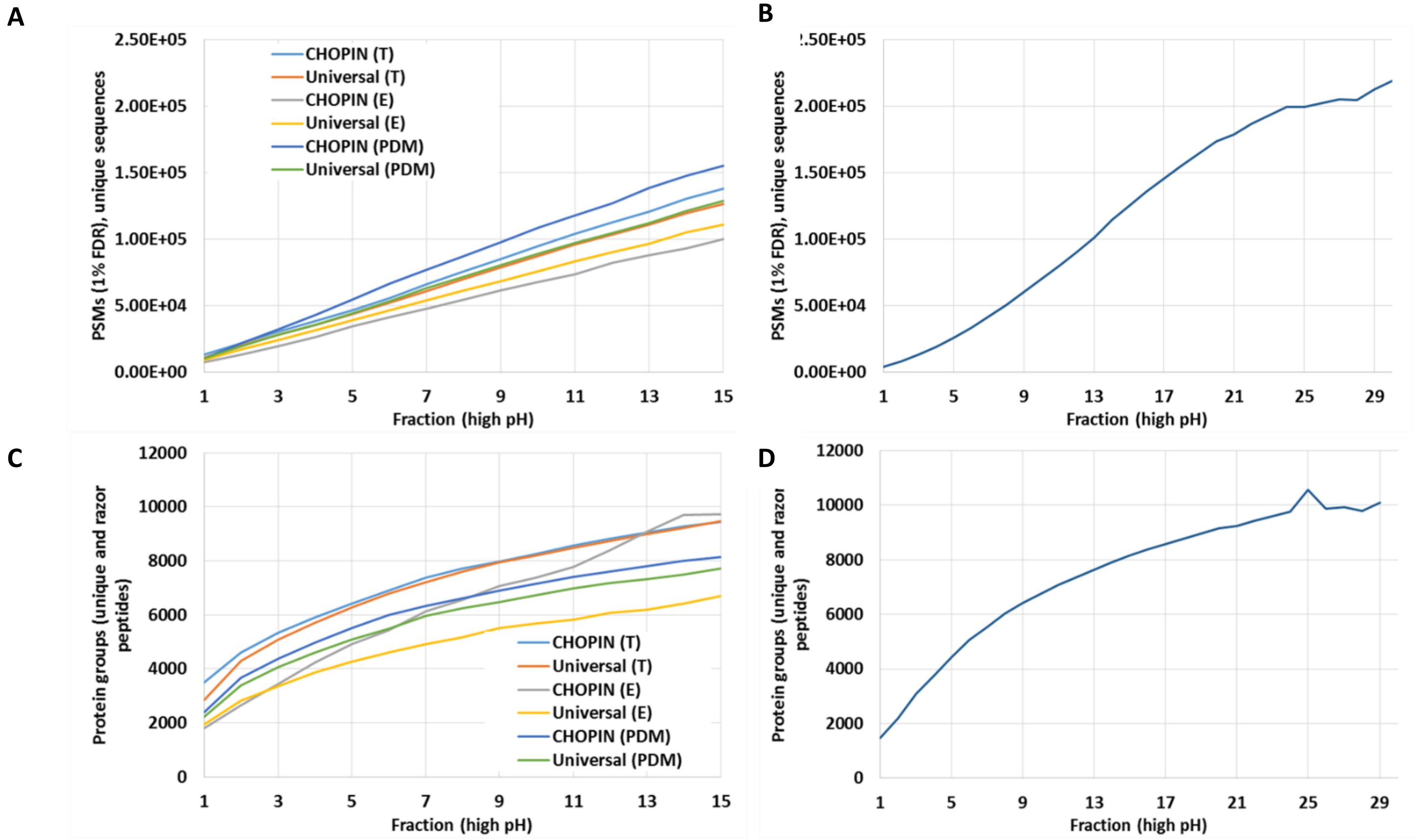


Figure S4 ***Accumulated peptide and protein identifications after high pH pre-fractionation.*** *(**A**) We plotted the accumulating unique peptide sequences following peptide pre-fractionation by high pH reversed phase chromatography and concatenation of a total of 30 fractions into 15 as shown in figure 1B. The mostly linear increase of unique peptide sequences demonstrates the orthogonal and effective strategy to maximise peptide identification. Without concatenation (**B**) we observe a sigmoidal curve demonstrating suboptimal use of MS acquisition time. However, the total number of peptides identified is further increased. (**C**) The tryptic digest appears to deliver the highest ID numbers with low fraction numbers. Protein ID numbers in this plot have not been corrected for Protein FDR, resulting in the elastase result to be too optimistic. However, the number of accumulating protein groups reaches a maximum between 25 and 30 fractions with the extremely complex Post Digest Mix sample (**D**). After adding more data to the database search, the protein group results appears to become unstable in the individual fractions/ Post Digest Mix sample set when reaching a plateau at about 10000 protein groups. This plateau effect in very large data sets has been reported before[1].*

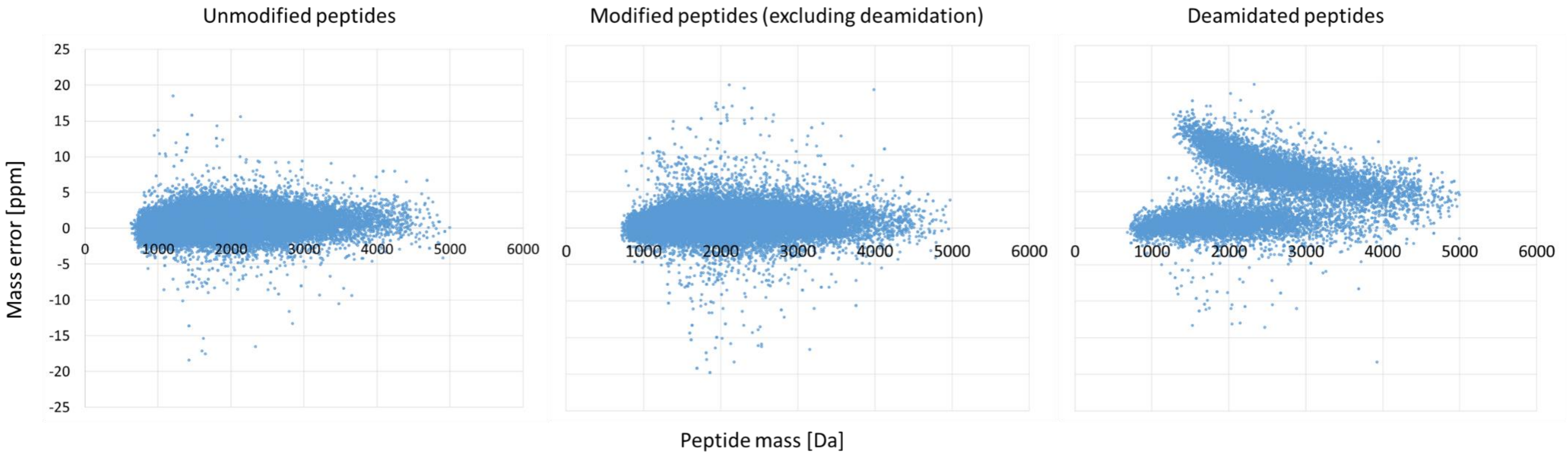


Figure S5 Mass error distribution of peptides (Tryptic digest) identified with and without modifications. *We analyzed the mass error of peptides identified in the tryptic samples to evaluate the PTM assignment of the PEAKS algorithm. The mass error of 205 detected modification types (excluding deamidation, middle panel) matches the distribution of unmodified peptides (left panel). The distribution of deamidated peptides shows 2 clusters. The upper cluster represents peptides in which the monoisotopic precursor mass was incorrectly assigned by selecting the first ^{13}C isotope as the precursor ion mass in the mass spectrometer while the lower cluster shows correctly assigned deamidated peptides, indicating that high mass accuracy and precision is required to distinguish real deamidations from incorrectly assigned precursor masses.*

	different types	counts
PTMs	91	81905
Chemical derivatives	87	104910
Artefacts	28	6733
	206	193548

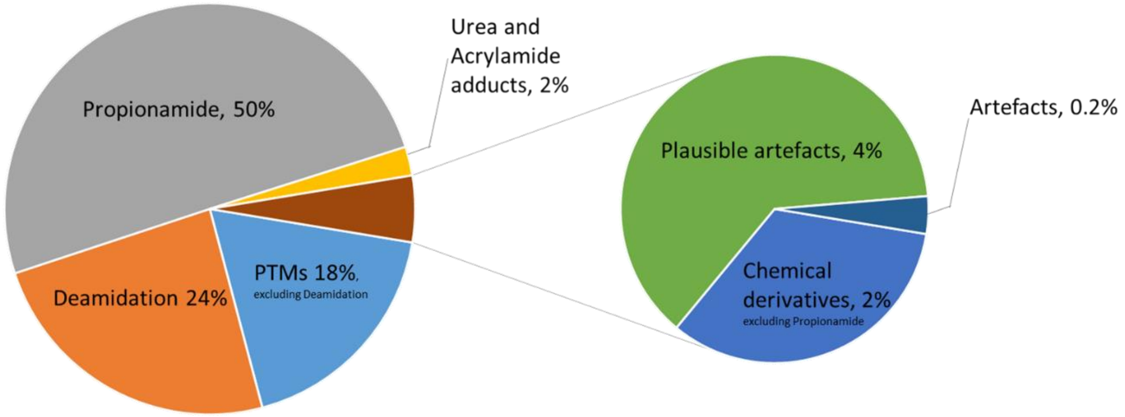
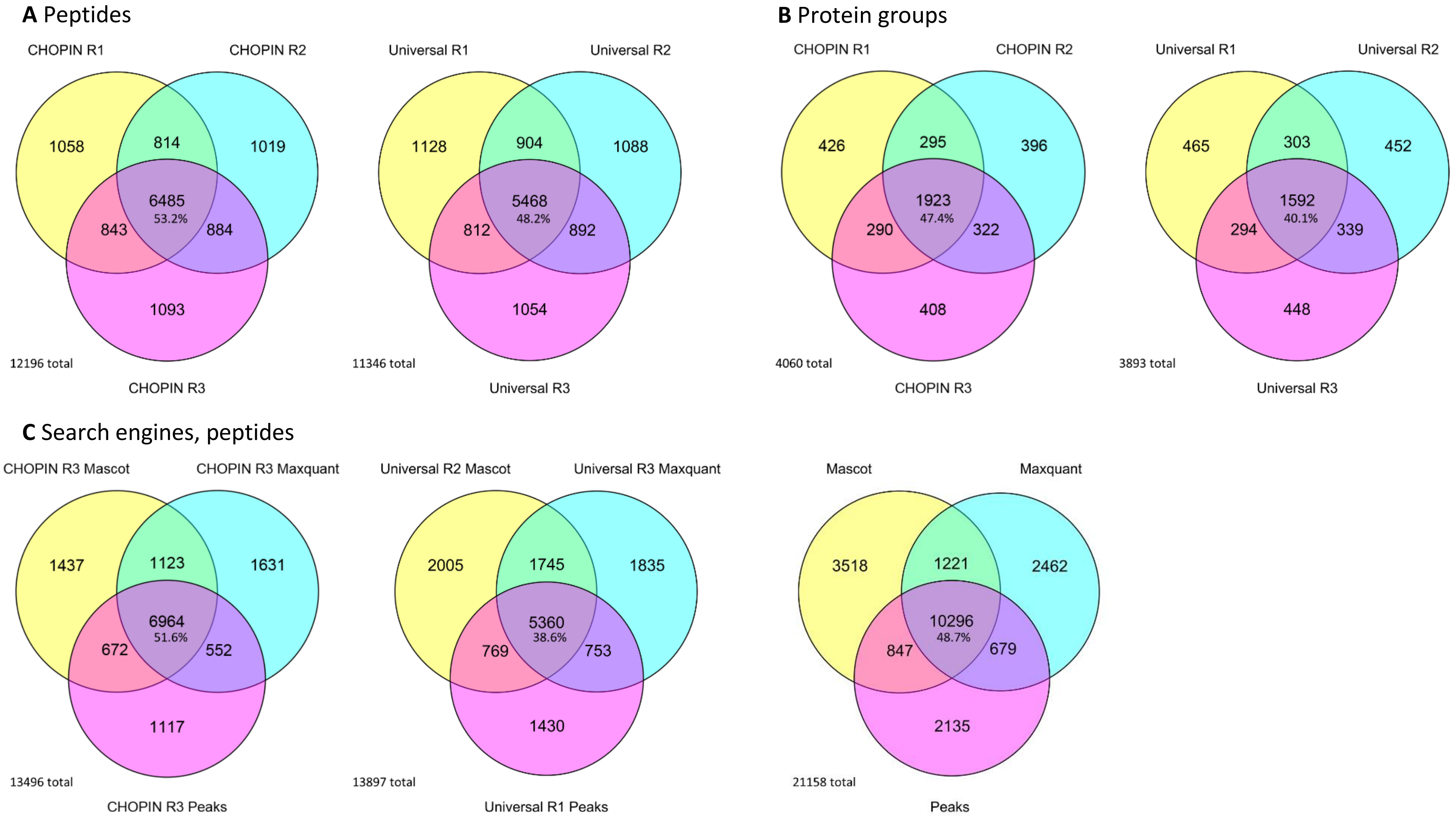


Figure S6 Detected modification landscape in MCF-7 cells. *We detected a total of 206 different types of peptide modifications in the combined data (see also Table S3). 91 modification types classify as PTMs (Unimod), 28 as artefacts and 87 as chemical derivatives (for details see Tab. S4). The vast majority in this classification are introduced during the sample preparation. The GASP derived propionamide modification of Cysteines, Lysines and N-termini sum up to 50 % of all detected modifications, while incorrect precursor assignment of the precursor and deamidation account for 24 % of detected modifications (compare Fig. S5). Artefacts and chemical derivatives (excluding propionamide) represent 2.2 % of all detected modifications. This group cannot be explained by sample handling, wrong precursor assignment or plausible artefacts, and may be in fact false discoveries, even though the chemical derivative classification includes metabolic modifications with biological origin as well. In absence of a false positive estimation at the PTM level, the percentage of implausible modifications (2.2%) can be used to evaluate the validity of modified peptide identifications.*


A Peptides
CHOPIN R1
CHOPIN R2
1058
814
1019
6485
53.2%
843
884
1093
12196 total
CHOPIN R3
Universal R1
Universal R2
1128
904
1088
5468
48.2%
812
892
1054
11346 total
Universal R3
B Protein groups
CHOPIN R1
CHOPIN R2
426
295
396
1923
47.4%
290
322
408
4060 total
CHOPIN R3
Universal R1
Universal R2
465
303
452
1592
40.1%
294
339
448
3893 total
Universal R3
C Search engines, peptides
CHOPIN R3 Mascot
CHOPIN R3 Maxquant
1437
1123
1631
6964
51.6%
672
552
1117
13496 total
CHOPIN R3 Peaks
Universal R2 Mascot
Universal R3 Maxquant
2005
1745
1835
5360
38.6%
769
753
1430
13897 total
Universal R1 Peaks
Mascot
Maxquant
3518
1221
2462
10296
48.7%
847
679
2135
21158 total
Peaks

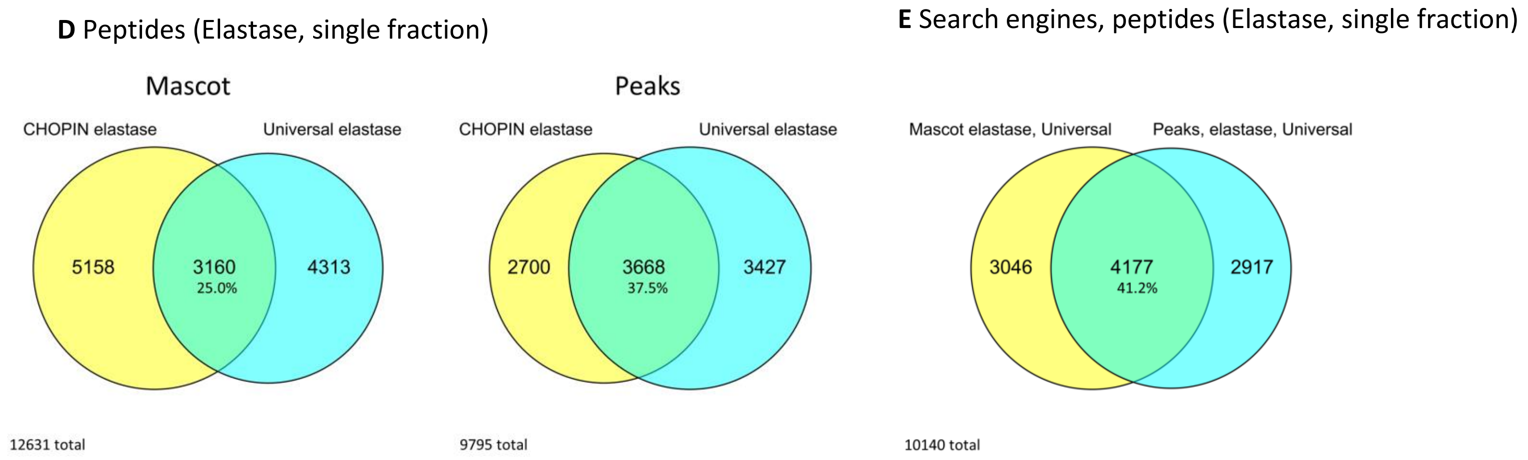
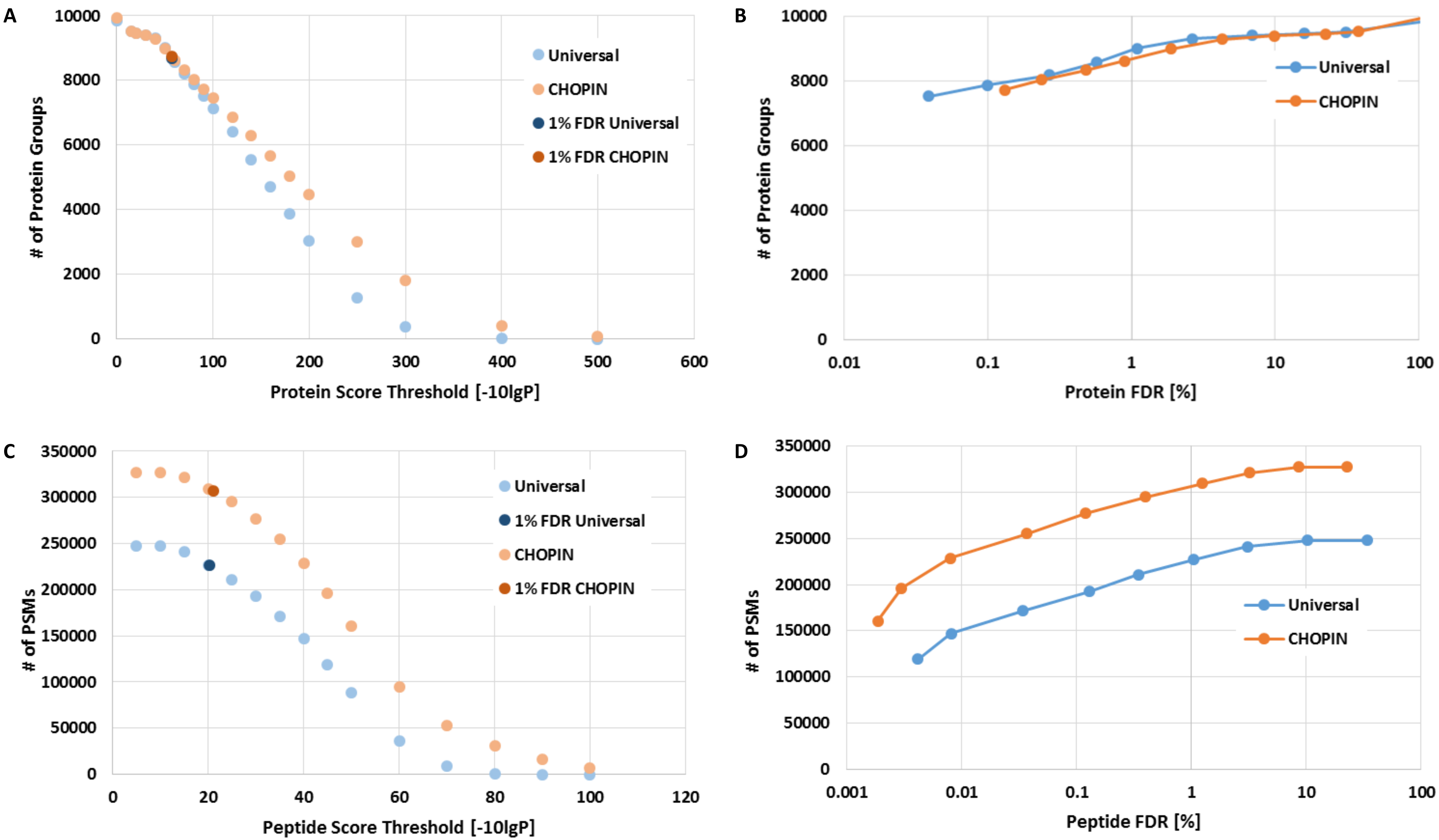


Figure S7 Reproducibility of CHOPIN and Universal methods assessed using multiple search engines. *In order to assess reproducibility and robustness of the acquisition methods used, we analyzed fraction F16/31 of the tryptic digest in technical triplicates (R1-R3) with both methods in PEAKS.* ***(A)*** *Using CHOPIN we identified a total of 12196 peptide sequences with 53% being identified in all three replicates. The Universal method produced less identified sequences (11346) and a lower percentage (48.2%) of peptide identified in all replicates. The same trends carry through to the protein group level.* ***(B)****. If searches are conducted in MaxQuant*[2] *(v. 1.5.6.5) and identifications are allowed to be transferred between the technical replicates (the "Match-between-runs" option), the percentage of proteins detected in all three replicates for both the CHOPIN and Universal Method reaches 90.5% (not shown). We also assessed potential biases of search engines by comparing the best peptide level result within the replicates from different search engines* ***(C)****. While the number of peptides identified with Mascot, MaxQuant and PEAKS in the best replicate is comparable, we see a significantly higher percentage of peptides identified in all three search engines when CHOPIN is used (51.6 versus 38.6%), indicating a generally higher robustness for peptide identification overall. When all identified peptides (from CHOPIN and Universal methods) are pooled, we observe 48.7% agreement between the used search engines. Interestingly PEAKS has consistently the lowest number of peptides exclusively identified by a single search engine. We also compared the single analyses of fraction F16/31 after elastase digest and PEAKS and Mascot database search* ***(D)****. While MaxQuant was unable to finish the search, Mascot and PEAKS produced remarkably different results. Comparing the data generated by the CHOPIN and Universal/FT methods on the peptide level, we identified 9795 peptides with*

PEAKS and a limited overlap of 37.5% between CHOPIN and the Universal/FT Method. This is expected as the elastase digested sample is more than an order of magnitude more complex than a tryptic digest and precursor selection is a more random event as a consequence. From the same data, Mascot identified a total of 12631 peptide sequences with an even lower overlap of 25.0%. The higher number of peptides exclusively identified with CHOPIN suggests that the high resolution MS/MS spectra are also beneficial to a precursor mass based search algorithm, when no cleavage specificity is applied. When Mascot and PEAKS peptide identifications (Universal Method results only) are compared, we do not observe an advantage of one search engine over the other, even when different peptide filtering mechanisms are applied (parsimony vs. ranked, see suppl. Notes) ***(E)****. Venn diagrams were created with VennDis 1.0.1*[3]*.*

Trypsin digest



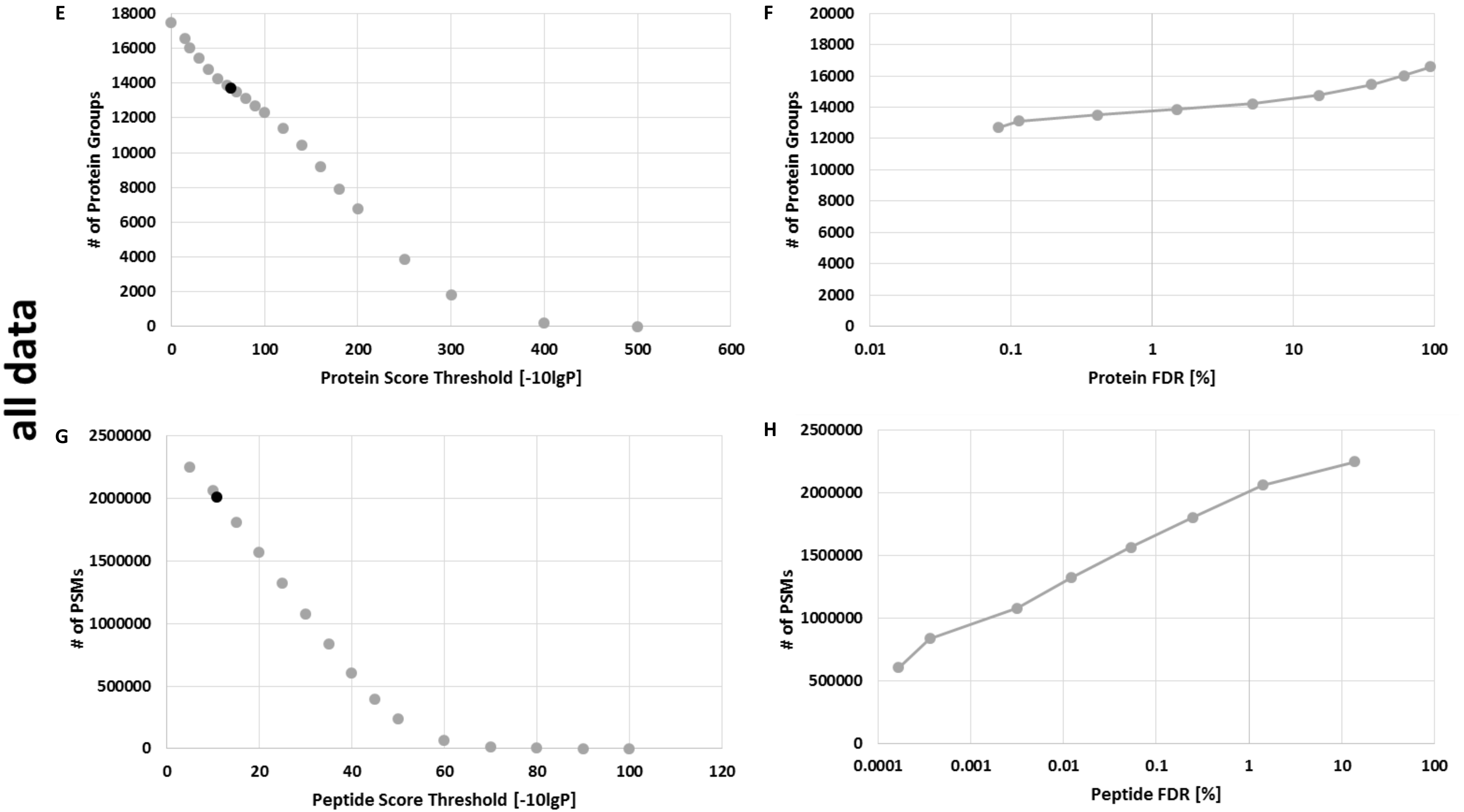
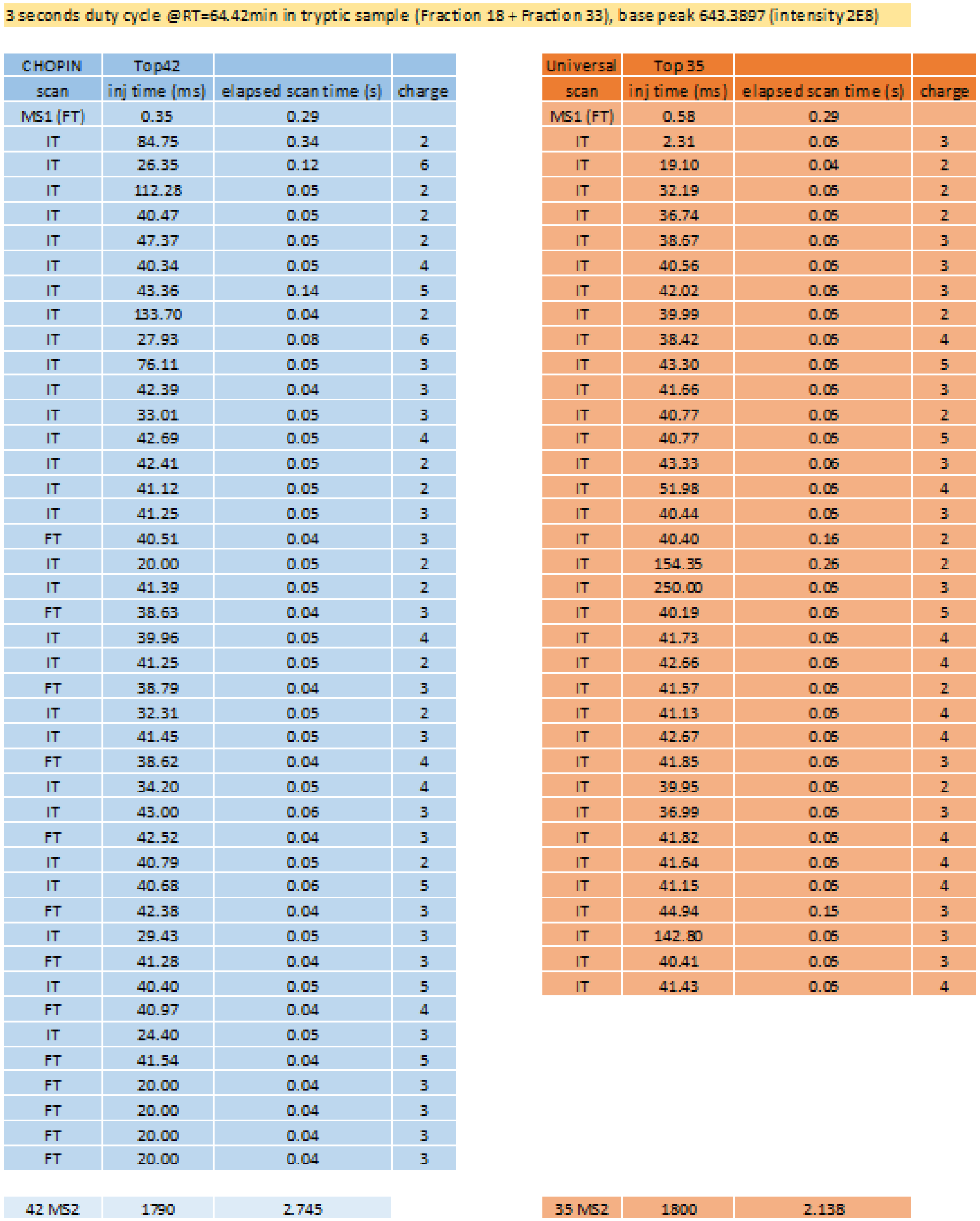

all data
E
# of Protein Groups
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
0
100
200
300
400
500
600
Protein Score Threshold [-10lgP]
F
# of Protein Groups
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
20000
0.01
0.1
1
10
100
Protein FDR [%]
G
# of PSMs
0
500000
1000000
1500000
2000000
2500000
0
20
40
60
80
100
120
Peptide Score Threshold [-10lgP]
H
# of PSMs
0
500000
1000000
1500000
2000000
2500000
0.0001
0.001
0.01
0.1
1
10
100
Peptide FDR [%]

Figure S8 FDR modelling demonstrates enhanced peptide, but not protein group identifications by CHOPIN. (A) *Using trypsin digested samples**,** we analyzed the number of protein groups that were identified dependent on different protein score thresholds settings. We evaluated whether CHOPIN analysis not only leads to more identified peptides, but also to enhanced identification rates of protein groups as compared to the Universal method. The protein FDR observed is consequential to the score threshold selected. At high protein score thresholds, CHOPIN consistently outperforms the Universal method in terms of protein groups. However, as the score threshold is lowered towards an effective FDR of 1% (in both datasets @ protein score 57, compare table 1), the trends of both methods converge towards a protein group upper limit, with the additional peptide identifications from CHOPIN contributing to a greater extent towards increased coverage of existing protein groups, rather than adding to the total group count (see main text and Figure S4 above). Similarly, the FDR of both methods (with respect to total protein group count) across this 'plateau' region is very similar* ***(B)****. The protein score threshold region where the difference in protein group count between CHOPIN and Universal becomes more pronounced corresponds to estimated protein FDRs below 1%. However, estimation of FDR beyond this point becomes inaccurate (the numbers of inferred decoy protein groups rapidly approaches zero).* ***(C)*** *At the peptide level, the benefits of CHOPIN are easier to observe. The improvement to peptide identification metrics using CHOPIN is pronounced, as generally more MS/MS spectra are acquired and are identified. This result is carried through to the peptide FDR* ***(D)****, in which a similar number of spectra can be matched with 0.01% FDR using CHOPIN compared to 1% FDR using the Universal method.* ***(E-H)*** *The combined data aver all acquisition methods and both digests shows similar distributions to the trypsin derived data, indicating that the FDR model applied is consistent, even when no-enzyme results are included (1% FDR data is shown in black). In particular the protein group number does not collapse even at 0.1% protein FDR* ***(F)****.*

Table S1 Comparative examples of a typical duty cycle using the CHOPIN and Universal methods. *A 3 second duty cycle using CHOPIN (blue) and Universal method (orange) in the same sample at the same retention time (same base peak and intensity) are compared to illustrate the additional parallelization gained by using CHOPIN. While the accumulated total injection time is similar, the time spent on MS/MS scans is increased from 2.14 seconds to 2.75 seconds. The additional scan time is used by HCD/FT scans of seven additional precursors without prolonging the duty cycle.*

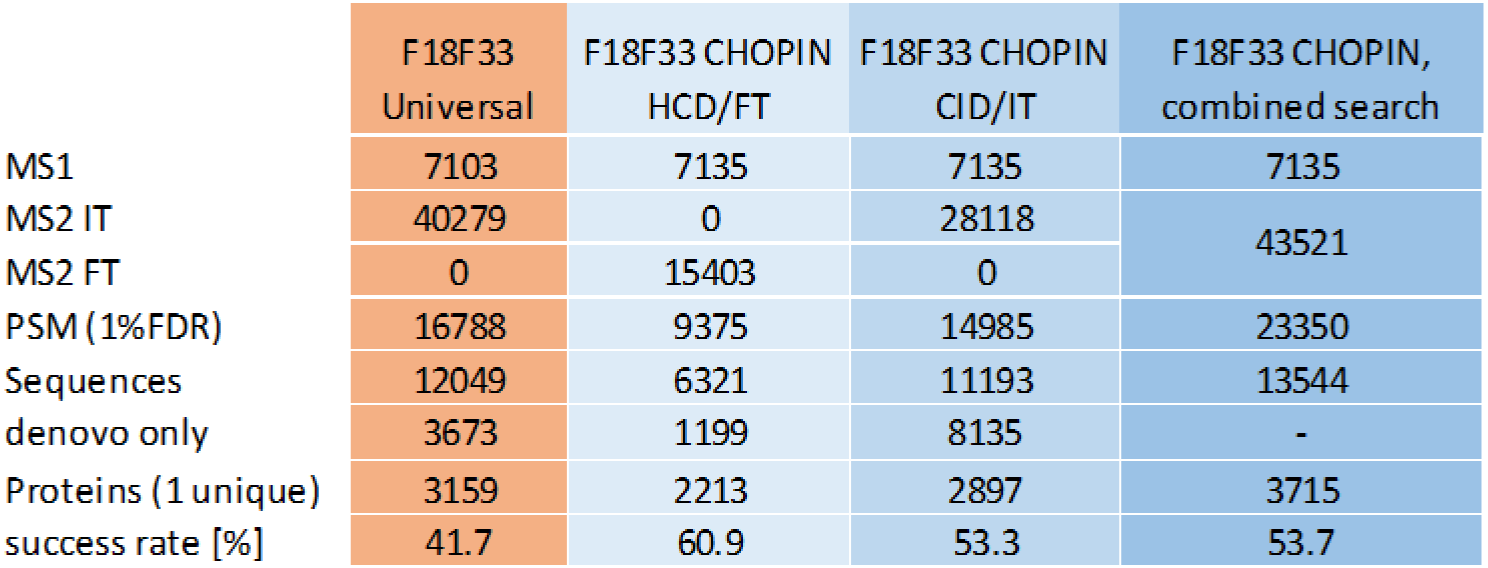


3 seconds duty cycle @RT=64.42min in tryptic sample (Fraction 18 + Fraction 33), base peak 643.3897 (intensity 2E8)

CHOPIN	Top42		
scan	inj time (ms)	elapsed scan time (s)	charge
MS1 (FT)	0.35	0.29	
IT	84.75	0.34	2
IT	26.35	0.12	6
IT	112.28	0.05	2
IT	40.47	0.05	2
IT	47.37	0.05	2
IT	40.34	0.05	4
IT	43.36	0.14	5
IT	133.70	0.04	2
IT	27.93	0.08	6
IT	76.11	0.05	3
IT	42.39	0.04	3
IT	33.01	0.05	3
IT	42.69	0.05	4
IT	42.41	0.05	2
IT	41.12	0.05	2
IT	41.25	0.05	3
FT	40.51	0.04	3
IT	20.00	0.05	2
IT	41.39	0.05	2
FT	38.63	0.04	3
IT	39.96	0.05	4
IT	41.25	0.05	2
FT	38.79	0.04	3
IT	32.31	0.05	2
IT	41.45	0.05	3
FT	38.62	0.04	4
IT	34.20	0.05	4
IT	43.00	0.06	3
FT	42.52	0.04	3
IT	40.79	0.05	2
IT	40.68	0.06	5
FT	42.38	0.04	3
IT	29.43	0.05	3
FT	41.28	0.04	3
IT	40.40	0.05	5
FT	40.97	0.04	4
IT	24.40	0.05	3
FT	41.54	0.04	5
FT	20.00	0.04	3
FT	20.00	0.04	3
FT	20.00	0.04	3
FT	20.00	0.04	3
42 MS2	1790	2.745	

Universal	Top 35		
scan	inj time (ms)	elapsed scan time (s)	charge
MS1 (FT)	0.58	0.29	
IT	2.31	0.05	3
IT	19.10	0.04	2
IT	32.19	0.05	2
IT	36.74	0.05	2
IT	38.67	0.05	3
IT	40.56	0.05	3
IT	42.02	0.05	3
IT	39.99	0.05	2
IT	38.42	0.05	4
IT	43.30	0.05	5
IT	41.66	0.05	3
IT	40.77	0.05	2
IT	40.77	0.05	5
IT	43.33	0.06	3
IT	51.98	0.05	4
IT	40.44	0.05	3
IT	40.40	0.16	2
IT	154.35	0.26	2
IT	250.00	0.05	3
IT	40.19	0.05	5
IT	41.73	0.05	4
IT	42.66	0.05	4
IT	41.57	0.05	2
IT	41.13	0.05	4
IT	42.67	0.05	4
IT	41.85	0.05	3
IT	39.95	0.05	2
IT	36.99	0.05	3
IT	41.82	0.05	4
IT	41.64	0.05	4
IT	41.15	0.05	4
IT	44.94	0.15	3
IT	142.80	0.05	3
IT	40.41	0.05	3
IT	41.43	0.05	4
35 MS2	1800	2.138	

Table S2 Example of data acquisition statistics for one of the tryptic fraction using Universal or CHOPIN method. *Results obtained from fraction 18/33 with CHOPIN and Universal method are listed to illustrate the differences in acquired data. While the number of total MS/MS spectra for both methods is similar, the success rate especially for the HCD/FT spectra using CHOPIN is increased, leading to overall better peptide and protein identification results.*



	F18F33 Universal	F18F33 CHOPIN HCD/FT	F18F33 CHOPIN CID/IT	F18F33 CHOPIN, combined search
MS1	7103	7135	7135	7135
MS2 IT	40279	0	28118	43521
MS2 FT	0	15403	0	
PSM (1%FDR)	16788	9375	14985	23350
Sequences	12049	6321	11193	13544
denovo only	3673	1199	8135	-
Proteins (1 unique)	3159	2213	2897	3715
success rate [%]	41.7	60.9	53.3	53.7

Table S6: Comparison of different search engines. *The analysis of the combined tryptic data acquired with CHOPIN and Universal Methods shows comparable results in SEQUEST, MaxQuant, PEAKS and Mascot. MaxQuant (via Andromeda) and PEAKS show highly similar metrics, as parsimony rules are used for protein grouping. The number of protein groups identified for each search algorithm is similar between CHOPIN and the Universal Method. However, the number of PSMs and the success rate for identification is significantly higher for the CHOPIN results across all compared search engines. The Mascot search was repeated using the Human Swiss-Prot protein database, excluding isoforms and splice variants. As the result obtained is very similar to the Mascot search with the UniProt Reference database, this indicates that only a low number of isoforms and splice variants may be distinguished from tryptic digests alone.*

Trypsin	Sequest HT (Proteome Discoverer 2.1)		Maxquant 1.5.6.5		PEAKS 7.5		Mascot 2.5		Mascot 2.5	
Database	UPR human (92910 protein entries, 05.10.2016)								Swissprot, human (20268 protein entries)	
	CHOPIN	Universal	CHOPIN	Universal	CHOPIN	Universal	CHOPIN	Universal	CHOPIN	Universal
Protein Groups	7883	7982	8559	8687	8745	8692	7424	7574	7222	7361
PSMs	179373	128471	301987	218576	307318	226143	241553	185055	241477	184816
Success rate (%)	30.8	23.8	51.2	39.9	52.8	41.9	41.0	33.8	41.0	33.7

Supplementary Notes and Results

Score distribution

Instead of the original Universal Method, we used a pure HCD/FT method ("Universal FT") for the data acquisition following elastase digestion in order to improve the identification success rate. The non-protease specificity-restricted search dramatically improves when high mass accuracy MS/MS spectra are available [4]. However, it also resulted in a lower number of acquired MS/MS spectra as compared to the original Universal method. As expected, the success rate was greatly improved. CHOPIN performed equally well with regards to total PSMs, but had to acquire almost twice the number of spectra for this result.

The Post Digest Mix produced a very similar profile to the tryptic digest. However, we identified less peptides than in the tryptic sample. Given the further increased complexity as compared to the trypsin and elastase digests alone, this sample benefits from a higher resolved pre-fractionation approach as shown in Tab. 1.

Success rate and Protein ID

Fig. S4A demonstrates that a tryptic digest generates peptides, better suited for MS detection than the loosely restricted elastase digest. If precursors are streamlined for their optimized detection with CHOPIN, peptide identification rates can be further improved. Interestingly, CHOPIN copes very well with the largely increased sample complexity of the Post Digest Mix, while the Universal Method is severely challenged. Not only generates the "free" HCD/FT mode (integrated in CHOPIN) data with high identification success rate, but we also see improved IDs in the CID/IT spectra generated with CHOPIN over the Universal Method as a result of being applied with higher priority on doubly charged ions.

The data acquisition method has a direct effect on the depth of the detected proteome. In CHOPIN, we prioritize low abundant, doubly charged precursor ions for CID/IT detection. At the same time, highly abundant large peptides will be scanned with HCD/FT and put on the dynamic exclusion list. In this way, the more sensitive CID/IT scan is not spent on ions which have a high chance to generate good quality spectra through their high abundance/better ionisation. This effect can be observed when the CHOPIN elastase result is compared to the Universal/FT Method. The instrument only gets a chance to look at the more abundant ions in the Universal/FT Method. The success rate of MS/MS spectra is remarkable, given that we need to conduct the database search without proteolytic restriction and the application of high mass accuracy seems beneficial. Using CHOPIN, we generated a similar number of PSMs and observe an inferior success rate. However, we detected more than 3000 additional protein groups using CHOPIN, which are mostly proteins, identified with a single short peptide. These peptides are mostly razor peptides and lead to an overly optimistic results when proteins are grouped under standard rules and no protein FDR is applied. To account for this affect we used a protein score threshold (Table 1) to achieve a protein FDR of 1%.

In Fig. S8 we evaluated the FDR modelling comparing Universal with the CHOPIN methods based on the tryptic data (A-D). In summary we observed very similar protein group metrics at a 'standard' 1% FDR for both methods, indicating that the increased sequence coverage from CHOPIN (as seen in the peptide FDR comparison) will not always translate to a higher number of protein groups following protein inference. This observation agrees with our assumption that a further increase in detected protein group numbers analysing only tryptic peptides may require either an increase in sensitivity of the detector, or targeting of certain protein entities (complexes, subcellular fractions, etc.). We also modelled the FDR metrics for the complete dataset (E-H) including trypsin and elastase digests, as well as CHOPIN and Universal method data acquisition. Reassuringly across all metrics the used 1% FDR cut-off on peptide and protein level is within the linear range of the plotted curves, which indicates a valid FDR estimation for our data.

Modified peptides

In low complexity samples, elastase has been used to increase protein sequence coverage with great success [5, 6]. However, the nearly unrestricted proteolysis stresses peptide/protein identification through classic database comparison approaches as are used in Mascot [7], MaxQuant (Andromeda) [2], SEQUEST [8], etc. Second generation search tools integrate a de novo based search into the workflow, which also allows an unbiased detection of modifications and amino acid exchanges. In our study we employed the PEAKS PTM search algorithm [9], which is able to detect 485 modifications based on the Unimod database [10]. Using this search algorithm we detected 206 different modifications on a total of 193548 sites. Next to common modifications such as phosphorylation, acetylation, methylation and ubiquitylation, we detected hydroxylation, sulfenylation, succinylation and formylation at notable frequencies. We also identified modifications caused by sample processing/handling including deamidation (46547 sites), oxidation (11217 sites), propionamide (97137 sites) and others (38647 sites in total). Only 2.2 % of the detected modifications cannot be explained immediately by sample processing, instrument error at precursor selection or biological processes, indicating potential false positive identifications (Fig. S6). We also did not observe a higher precursor mass error distribution for modified peptides (Fig. S5) with the exception of incorrectly assigned precursor masses, when the 13C precursor ion of the peptide was selected. Many PTM types, for which enrichment strategies exist (for example phosphorylation, ubiquitination, acetylation), appear underrepresented in our data (Tab. S3). However, the unbiased PTM detection in combination with the wealth of MS/MS spectra generated with CHOPIN in this study greatly enriches the scope for PTM analysis by modifications which are not routinely identified or even considered to be relevant.

Comparison of search engines for the interrogation of very deep proteome data

We primarily used the PEAKS search engine, as it allows an unbiased search for modified peptides and also works exceptionally well with no enzyme specificity due to its de novo approach. However, PEAKS is less commonly used than other search engines such as MaxQuant/Andromeda, Mascot or SEQUEST. The performance of these search engines have been compared extensively elsewhere [9, 11-13]. However, to illustrate how different search engines handle deep proteome data, we analysed the fractionated tryptic digest with SEQUEST, Mascot, MaxQuant and PEAKS.

In table S6 we show data obtained with CHOPIN and Universal Method following tryptic digest, searching with Mascot against two Homo sapiens databases; the UniProt Reference database (which includes protein isoforms and splice variants) and the more conserved manually annotated Swiss-Prot database. Interestingly, protein and peptide metrics are very similar (see also Figure S7), indicating that tryptic digest results in only very few peptides, that can be used to distinguish protein isoforms and splice variants. Furthermore, we observed that the results between different search engines are broadly comparable, although PEAKS appears to have a small advantage due to its capability to do an unbiased PTM search. Comparing CHOPIN with the Universal

Method, we did not observe large differences in the identification of protein groups across the panel. However, it should be noted that the number of PSMs and also the ID success rate is significantly higher in all search engines when CHOPIN is applied for MS/MS acquisition, leading to significantly increased protein scores across the board (~1.5-2 fold, not visualized).

Robustness and reproducibility

In principle, CHOPIN should generate more reproducible results as it specifically addresses under-sampling by the mass spectrometer. The higher number of MS spectra acquired and also the generation of high quality spectra from abundant precursors should increase the percentage of peptides identified in all three analyses of a technical replicate. To address this question, we acquired three datasets from a tryptic fraction using CHOPIN and Universal Method.

Protein and peptide numbers identified with PEAKS are visualized in Fig. S7A and B. The results show a higher number of proteins and peptides identified in all three replicates when CHOPIN is used, as well as higher total number of identified peptides with CHOPIN overall. The increased reproducibility is replicated when Mascot and MaxQuant are employed for database searches (Fig. 7C). When identifications are transferred between technical replicates (using MaxQuant, "match between runs", not visualized), the percentage of proteins detected in all three replicates increases to 90.5% for both CHOPIN (4463 protein groups in total) and the Universal Method (4339 protein groups in total).

We repeated the analysis for single runs of one fraction of the elastase sample set and observed a striking difference between how Mascot and PEAKS handle "no-enzyme" searches. While all search results are reported at 1% peptide FDR for Mascot and 1% peptide/protein FDR for PEAKS, we did observe a limited overlap in peptides identified between CHOPIN and the Universal Method results (Fig. S7D). However this can be explained by the extreme complexity of such a sample and the increased arbitrariness of precursor selection of the under-sampling MS instrument, which is less of an issue with tryptic samples. Comparing all identified peptides of the Mascot search and the PEAKS search in the Universal Method generated peptide lists, we found reasonable overlap between the reported peptide identifications (Fig. S7E). However, discrimination between equal-scoring matches of an MS/MS spectrum to multiple sequences can only be achieved in both PEAKS and Mascot via reference to inferred protein groups (here "razor peptides" vs. "require bold red"). Without re-analysing the data with an independent inference algorithm (which introduces its own complications and biases), we can only make an imperfect comparison between the two search engines. We were unable to complete MaxQuant searches of the elastase data, as the software became unresponsive at the search preparation stage.

Practicalities of using CHOPIN/broad specificity digest workflows

CHOPIN has been very useful to improve confidence in peptide identifications and parallelisation of scan events in the Orbitrap Fusion mass spectrometer. While the method setup is the one of a data dependent decision tree, we provided all necessary information to reproduce this method. We have applied CHOPIN here in pre-fractionated samples. However, CHOPIN can also be applied to unfractionated samples to increase peptide and protein ID confidence and MS/MS success rate, even though it may not necessarily lead to higher protein identification numbers. Here, we used an ion count threshold of 500,000 to trigger HCD/FT MS2 acquisition, which works well with high sample loads (>500ng of lysate on column). If the threshold is set too high, or the *sample is too low abundant, HCD/FT scans will be triggered less frequently, and the benefit of additional parallelisation is diminished. Consequently the HCD/FT triggering threshold should be evaluated for each MS workflow and set with sample abundance in mind.*

Precursor intensity based quantitation methods such as label-free quantitation or SILAC are not affected by using CHOPIN, as the number of MS1 scans can be defined by the Top Speed setting. Product ion based quantitation such as TMT™ or iTRAQ™ are not compatible with CHOPIN at its current method design. Identification reproducibility using CHOPIN is significantly improved (Fig. S7), as CHOPIN essentially addresses under-sampling.

The dual nature of the MS/MS spectra obtained with CHOPIN (CID/IT and HCD/FT) complicates the data analysis, as slightly different fragment ion distributions can be expected and different mass error tolerances need to be applied. While MaxQuant can handle the presence of two different detectors for MS/MS spectra in the same data file, search engines such as Mascot or PEAKS require the separation of CID/IT and HCD/FT spectra for an optimal analysis. This separation can be achieved in Proteome Discoverer, which can be set to generate two separate peaklist files for their individual search. As a consequence the search results of the two peaklist files need to be re-combined, following protein grouping to address protein inference. This procedure is cumbersome and time consuming and we hope that multi-detector/fragmentation data will be searchable in future versions of Mascot and other search engines. It should be noted that the data analysis of the elastase digest in particular was very time consuming. MaxQuant was unable to complete a search without protease specificity using the UniProt Reference database, and PEAKS search time was three weeks on a workstation fitted with an 8 core CPU, 16GB memory and SSD to complete the searches for the elastase samples.

*In summary, we believe that CHOPIN is a generally easy to use method to improve parallelisation in an Orbitrap Fusion (Lumos) instrument. To allow an easy integration of CHOPIN we provided the method files (*Xcalibur Tune v. 2.0.1258.14) *and also method transcripts with additional explanations in the supplementary material.*

References

1. Wilhelm, M.; Schlegl, J.; Hahne, H.; Moghaddas Gholami, A.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; Mathieson, T.; Lemeer, S.; Schnatbaum, K.; Reimer, U.; Wenschuh, H.; Mollenhauer, M.; Slotta-Huspenina, J.; Boese, J. H.; Bantscheff, M.; Gerstmair, A.; Faerber, F.; Kuster, B., Mass-spectrometry-based draft of the human proteome. *Nature* **2014,** 509, (7502), 582-7.
2. Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M., Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **2011,** 10, (4), 1794-805.
3. Ignatchenko, V.; Ignatchenko, A.; Sinha, A.; Boutros, P. C.; Kislinger, T., VennDIS: a JavaFX-based Venn and Euler diagram software to generate publication quality figures. *Proteomics* **2015,** 15, (7), 1239-44.
4. Zubarev, R.; Mann, M., On the proper use of mass accuracy in proteomics. *Mol Cell Proteomics* **2007,** 6, (3), 377-81.

5. Schlosser, A.; Pipkorn, R.; Bossemeyer, D.; Lehmann, W. D., Analysis of protein phosphorylation by a combination of elastase digestion and neutral loss tandem mass spectrometry. *Anal Chem* **2001,** 73, (2), 170-6.

6. Getie, M.; Schmelzer, C. E.; Neubert, R. H., Characterization of peptides resulting from digestion of human skin elastin with elastase. *Proteins* **2005,** 61, (3), 649-57.

7. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** 20, (18), 3551-67.

8. Eng, J. K.; McCormack, A. L.; Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **1994,** 5, (11), 976-89.

9. Han, X.; He, L.; Xin, L.; Shan, B.; Ma, B., PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications. *J Proteome Res* **2011,** 10, (7), 2930-6.

10. Creasy, D. M.; Cottrell, J. S., Unimod: Protein modifications for mass spectrometry. *Proteomics* **2004,** 4, (6), 1534-6.

11. Bruce, C.; Stone, K.; Gulcicek, E.; Williams, K., Proteomics and the analysis of proteomic data: 2013 overview of current protein-profiling technologies. *Curr Protoc Bioinformatics* **2013,** Chapter 13, Unit 13 21.

12. Paulo, J. A., Practical and Efficient Searching in Proteomics: A Cross Engine Comparison. *Webmedcentral* **2013,** 4, (10).

13. Tu, C.; Sheng, Q.; Li, J.; Ma, D.; Shen, X.; Wang, X.; Shyr, Y.; Yi, Z.; Qu, J., Optimization of Search Engines and Postprocessing Approaches to Maximize Peptide and Protein Identification for High-Resolution Mass Data. *J Proteome Res* **2015,** 14, (11), 4662-73.