Research Article

Improved Label-Free LC-MS Analysis by Wavelet-Based Noise Rejection

Salvatore Cappadona,[1] Paolo Nanni,[2] Marco Benevento,[2] Fredrik Levander,[3] Piera Versura,[4] Aldo Roda,[2] Sergio Cerutti,[1] and Linda Pattini[1]

[1] *Department of Bioengineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy*
[2] *Department of Pharmaceutical Science, University of Bologna, Via Belmeloro 6, 40126 Bologna, Italy*
[3] *Department of Immunotechnology, Lund University, BMC D13, 221 84 Lund, Sweden*
[4] *Ophthalmology Unit, University of Bologna S.Orsola-Malpighi Hospital, Via Massarenti 9, 40138 Bologna, Italy*

Correspondence should be addressed to Salvatore Cappadona, salvatore.cappadona@polimi.it

Label-free LC-MS analysis allows determining the differential expression level of proteins in multiple samples, without the use of stable isotopes. This technique is based on the direct comparison of multiple runs, obtained by continuous detection in MS mode. Only differentially expressed peptides are selected for further fragmentation, thus avoiding the bias toward abundant peptides typical of data-dependent tandem MS. The computational framework includes detection, alignment, normalization and matching of peaks across multiple sets, and several software packages are available to address these processing steps. Yet, more care should be taken to improve the quality of the LC-MS maps entering the pipeline, as this parameter severely affects the results of all downstream analyses. In this paper we show how the inclusion of a preprocessing step of background subtraction in a common laboratory pipeline can lead to an enhanced inclusion list of peptides selected for fragmentation and consequently to better protein identification.

1. Introduction

In the last years, clinical proteomics has witnessed an increased interest towards mass spectrometry-based methods for quantitative differential analysis of protein content in biological samples (i.e., biological fluids from drug-treated versus untreated subjects, or from healthy versus ill patients).

MS-based proteomics approaches for comparative analysis include both methods based on the use of stable isotopes [1], such as iTRAQ (Isobaric tags for relative and absolute quantitative) and SILAC (Stable isotope labeling with amino acids in cell culture), and so-called label-free approaches [2].

Label-free liquid chromatography-mass spectrometry (LC-MS) differential analysis allows determining the differential expression level of proteins in multiple samples without presenting any limit to the number of samples being compared and without increasing the complexity of mass spectra. It is based on the direct comparison of peak intensi-

ties between multiple runs obtained by continuous detection in MS mode, followed by MS/MS fragmentation of only differentially expressed peptides. This procedure avoids the bias toward abundant peptides, typical of data-dependent tandem MS, and allows an increased identification of low-abundant peptides.

In a typical label-free LC-MS experiment each analysis is performed independently and it is followed by comparison of the multiple LC-MS images. The computational framework includes the steps of peaks detection, maps alignment and normalization, peaks matching across multiple sets, and a statistical analysis of the detected features for the evaluation of the differentially expressed peptides.

Several open-source, commercial, and custom software packages that address one or more of these processing steps have been described in the literature [3]. Nevertheless, most of the available tools show little or no care in assessing a minimum quality standard for the LC-MS maps

entering the pipeline. Baseline subtraction and denoising, for instance, are still often neglected, despite their strong impact on all downstream analyses [4, 5].

In order to show the importance of noise rejection, in this study we report a label-free LC-MS differential analysis of protein abundance in tears samples performed with and without inclusion of a preprocessing step [6] into an established analytical and computational strategy [7, 8].

The preprocessing is performed on a whole LC-MS map. The algorithms work iteratively by first extracting all Single Ion Chromatograms (SICs) and by then processing independently each SIC by means of a wavelet decomposition to identify and remove the components of the chemical and the random noise.

Several other papers have introduced algorithms that exploit the two-dimensional nature of the data to minimize the noise in the mass domain by signal processing in the chromatographic time domain. The advantage of our denoising strategy over other algorithms, though, mainly comes from characterising and subtracting the noise features from all SICs independently. The limit of other common approaches like CODA [9] or MEND [10], in fact, is that they often process only a selection of SICs. CODA, for instance, automatically retains only chromatograms with high S/N ratio and combines them to form a reduced total ion hromatogram (TIC) trace. Similarly, MEND divides the whole mass range in consecutive regions and for each region it determines a global model of the noise by combining a fixed number of "vacant" SICs, that contain no chromatographic peaks.

In the first step of the present work, tears proteins from healthy (H) subjects and patients affected by hyperevaporative dry eye (HDE) were subjected to tryptic digestion and analyzed by reverse-phase chromatography nano-LC ESI-QTOF MS in order to evaluate the protein changes related to the disease. The msInspect software [11] was used for alignment and normalization of the LC-MS maps while the open-source Proteios Software Environment (ProSE) [12] was used for statistical analysis and for the creation of the list of peptides to be identified by RP nano-LC ESI/QTOF MS/MS analysis followed by database search.

In a second phase of the work the same LC-MS raw data files were first filtered to remove chemical and random noise and then reprocessed by the same computational pipeline. The results are compared with those obtained by means of the standard procedure and the influence of noise rejection on the selection of peptides for MS/MS fragmentation is commented according to previously obtained outcomes [13].

2. Experimental

2.1. Materials and Reagents.
All the analytical grade reagents, the Myoglobin, and the solvents were purchased from Sigma Aldrich (St. Louis, MO, USA).

2.2. Subjects Studied.
A total of 4 subjects, including 2 healthy volunteers (H; 1 M and 1 F) and 2 hyperevaporative dry eye patients (HDE; 2 M), were admitted to this study.

Inclusive criteria for patients were Schirmer test I value $\geq$ 10 mm/5 min, Tear Film Break Up Time (T.F.B.U.T.) < 10 seconds, and symptoms of ocular discomfort from at least two months. Inclusive criteria for healthy control subjects were Schirmer test I value $\geq$ 10 mm/5 min, T.F.B.U.T. $\geq$ 10 seconds, and no ocular discomfort symptoms. In both groups exclusion criteria were considered the presence of punctuate cheratopaty and/or autoimmune diseases, the use of contact lenses and any ocular surgery in the last 6 months.

All the tear samples were provided by the Ophthalmology Unit at the University of Bologna (Italy) after obtaining informed consent from the subjects studied and according to DEWS guide lines [14] A minimum of 5 μL tears was collected using a micropipette with sterile tip, centrifuged and stored as previously described [15].

2.3. Tear Samples Preparation.
Total protein quantification of each tear sample was performed by Bradford protein assay using bovine serum albumin (BSA) as a standard according to the manufacturers' instructions (Bio-Rad, Laboratories Inc., CA, USA).

For each sample, 10 μg of proteins were diluted to 10 μL with 6 M Urea in 100 mM ammonium bicarbonate pH 8.2 and 2 picomoles of a Myoglobin (Myo, P68082) were added as internal standard. The protein mixtures were reduced by adding 1 μL of 100 mM dithiothreitol (DTT, Sigma) in 100 mM ammonium bicarbonate for 1 hour at 37°C and alkylated by addition of 3 μL of 100 mM iodoacetamide (IAA, Sigma) in 100 mM ammonium bicarbonate for 1 hour at room temperature in the dark. The resulting samples were incubated overnight at 37°C with trypsin 12 ng/μL (Promega) in a 50 : 1 (w : w) ratio. The tryptic digestions were blocked after 15 hours incubation with 1 μL of formic acid, and afterwards the samples were lyophilized to dryness and resolubilized with 20 μL of 0.1% formic acid (FA).

2.4. Liquid Chromatography and Mass Spectrometry.
For each sample 4 μL were analyzed by LC-MS analysis using a CapLC (Waters, Manchester, U.K) with flow splitting from 5 μL/min to 200 nL/min, connected with a nanoelectrospray interface to a QTOF Ultima (Waters) using MassLynx v4.0 software as operating software. The peptide separation was performed on an Atlantis dC18 NanoEase column (150 $\times$ 0.075 mm, 3 μm) (Waters) with an Atlantis dC18 NanoEase precolumn (0.3 $\times$ 5 mm, 5 μm particle size) (Waters), using as mobile phase A H_2O/acetonitrile (95 : 5) 0.1% FA while the mobile phase B was acetonitrile/H_2O (95 : 5) 0.1% FA. A 90-minute chromatographic gradient was used to give a linear increase after 3 minutes from 2% B to 35% B in 70 minutes and from 35% B to 80% B in 2 minutes, and after 3 minutes at 80% B the column is conditioned again at 2% B for 15 minutes. One blank injection with a 30-minute gradient was run between samples to reduce sample carry over, and every six samples 2 pmol of Myo tryptic digest were analyzed as quality control using the same 90 minutes gradient to evaluate the experimental variation. During MS analysis the QTOF was set to scan in profile mode m/z 400–1800 with
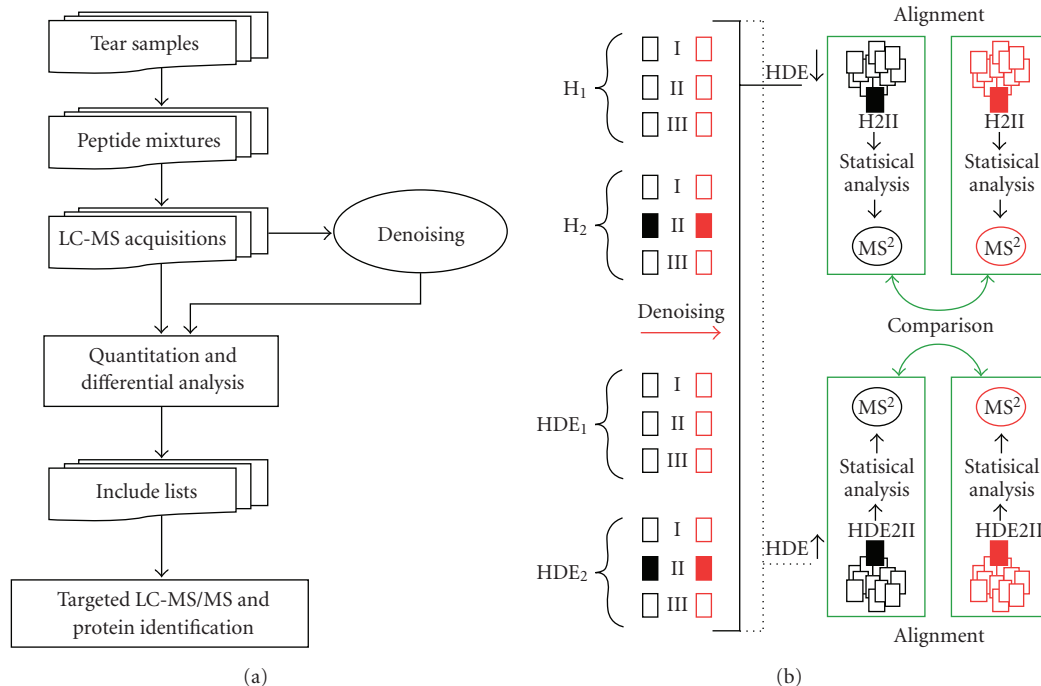
(a)

(b)

FIGURE 1: Typical pipeline for comparative LC-MS analysis. (a) The preprocessing step of noise rejection is shown as optional to the common workflow. (b) Details of the bioinformatic steps: in black the standard workflow, in red the alternative workflow subsequent to noise rejection, and in green the comparison of the results obtained by the two approaches.

1.9 seconds per scan and 0.1 seconds of scan delay. The samples were analyzed in triplicates. Four microliters of sample were injected for targeted MS/MS and the same LC gradient was used. The survey scan time was set to 1 second and a peak limit of 15 counts to switch to MS/MS mode. For inclusion lists the time tolerance was set to 300 seconds.

2.5. MS Data Analysis. The pipeline of MS data analysis has been already described elsewhere by our research group [7, 8] and is shown in Figure 1(a). Briefly, massWolf (v2.0, http://sashimi.sourceforge.net) was used for the conversion of Micromass LC-MS raw data file to mzXML, while the peptide feature detection was performed using msInspect v2.1. Two alignments were performed for the analysis of all the LC-MS features: one with an H sample and one with an HDE sample as master. The normalization of the LC-MS maps and their alignments were performed by means of the peptide Array tool in msInspect [11], using a mass window of 0.2 m/z and a time window of 250 scans. After alignment, significantly upregulated features were scheduled for targeted MS/MS in inclusion lists generated using the ProSE 2.1 platform [12]. The inclusion limit was a fold change of at least 1.5 and a P-value of 0.05 in a Student's t-test. For the t-test the total intensities, which represent the integrated peak volumes, were used. For features where peaks could not be found in the healthy control samples, a value of 50 ion counts was used, which was an estimate for the detection level in the present setup. Selected features were sorted according to intensity and put into include lists with a maximum of 300 peaks per include list. The retention

times of the second technical replicate acquired sample were used in the include list. Targeted MS/MS analysis was finally performed to identify the peptides contained in the include lists.

2.6. Tandem MS Data Analysis. To generate peak lists for peptide identification, ProteinLynx Global Server 2.2 (Waters) was used. The XML format peak lists were converted to mzData using ProSE. Mascot version 2.2 (www.matrixscience.com) was used for peptide identification.

The Sprot human database, version 57.3, was used, 468851 sequences in total. The search settings were 0.2 Da precursor and 0.6 Da fragment tolerances, carbamidomethylation of cysteine as fixed modification, methionine oxidation as variable modification, and semiTrypsin with one missed cleavage as enzymatic digestion. The search results were exported as XML and matched with MS features using a ProSE plug-in, with a retention time tolerance of 100 seconds and a mass tolerance of 0.12 Da. Proteins were considered correctly identified when at least two different peptides (with significant individual score, i.e., $P < 0.05$) were present.

2.7. Noise Rejection. In the second phase of this work, the original mzXML files entering the pipeline described in Figure 1(a) were cleaned from extraneous noise by a wavelet-based algorithm already described by our research group [6]. Briefly, the algorithm works on a whole LC-MS map by first extracting all Single Ion Chromatograms (SICs) from the spectrographic data and by then decomposing each SIC to

TABLE 1: Alignment of technical replicates. Mass Window = 0.2 Th, Scan Window = 250 scans, Min Charge = 2^+.

|		Total	Only I	Only II	Only III	I-II	I-III	II-III	I-II-III
H1 I-II-III	*UNPROCESSED*	4167	767	568	564	315	288	208	1457
	PROCESSED	5736	1061	915	1018	379	361	338	1664
	% increase	**37.65**	**38.33**	**61.09**	**80.50**	**20.32**	**25.35**	**62.50**	**14.21**
H2 I-II-III	*UNPROCESSED*	4476	890	933	625	208	420	347	1053
	PROCESSED	5783	1178	1307	925	228	519	472	1154
	% increase	**29.20**	**32.36**	**40.09**	**48.00**	**9.62**	**23.57**	**36.02**	**9.59**
HDE1 I-II-III	*UNPROCESSED*	4456	861	889	607	314	475	245	1065
	PROCESSED	6324	1301	1386	1051	384	650	347	1205
	% increase	**41.92**	**51.10**	**55.91**	**73.15**	**22.29**	**36.84**	**41.63**	**13.15**
HDE2 I-II-III	*UNPROCESSED*	3921	842	543	563	345	253	288	1087
	PROCESSED	5087	1068	754	962	356	299	394	1254
	% increase	**29.74**	**26.84**	**38.86**	**70.87**	**3.19**	**18.18**	**36.81**	**15.36**
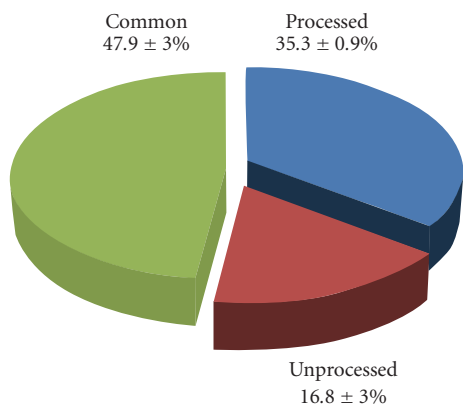
FIGURE 2: Results of the alignment of the original and the filtered files of all 12 LC-MS run. Common: features found in both maps; processed: features found only after noise rejection; unprocessed: features found only in the original maps.
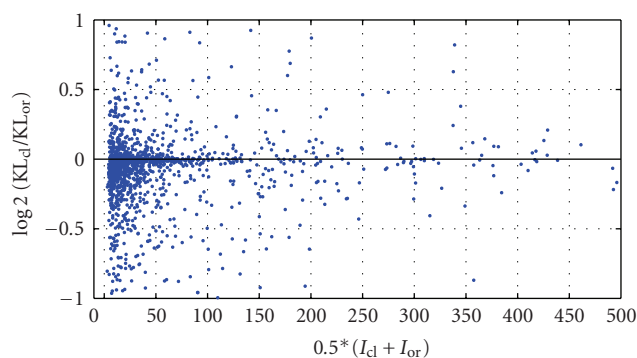

FIGURE 3: KL ratio of corresponding peptides after the alignment of the processed and the unprocessed maps of HDE2III. Most of the peptides have $KL_{Clean} < KL_{Original}$.

identify and remove the noise components. This cleaning step is simply added to the standard pipeline (Figures 1(a) and 1(b)) to selectively remove random and chemical noise while leaving the peptide peaks unaffected.

All data were processed by a stand-alone Java application on a 2.66 GHz iMac running Mac OS X with 1 GB of RAM allocated for the JVM heap. The decomposition was performed on 6 scales and by means of the Coifmann wavelet of degree 1.

3. Results and Discussion

3.1. Effects of Denoising on Peptide Feature Detection. The peptide Array tool of msInspect was run on the technical replicates of all subjects and the results are shown in Table 1. The first column of the table shows that the number of peptide features detected by msInspect increases on average by about 35%. In order to access sensitivity of the cleaning, only peptides with a charge ≥ 2 were included in the table. Uncharged features were excluded to avoid spurious peaks, while singly charged peptides were excluded to avoid false

positive identifications caused by chemical noise, whose regular pattern of peaks occurring at every Th can easily be mistaken for the isotopic distribution of a 1^+ peptide. Furthermore, by only considering peptides that are aligned through at least two of the replicates or through all of the replicates, the average increase in detection can be estimated, respectively, around 20% or 13%. A second alignment was performed between the original and the filtered files of each LC-MS run. In this case, since each file was practically aligned to itself, stricter parameters of Mass Window = 0.05 Da and Scan Window = 5 scans were imposed. Figure 2 shows that on average half of the peptides are found both before and after the cleaning. The figure shows also that about one sixth of the peptides have disappeared from the original file because of the cleaning, while one third have emerged in the processed files after background subtraction.

The biggest region of the pie chart relates to the peaks that are left unaffected by the cleaning, typically high-intensities peptides. A look into the quality of these peptide features shows a 30% average increase in the number of consecutive scans in which a peptide is detected (data not shown). This improvement is usually obtained by unveiling the lowest peaks of its isotopic distribution.
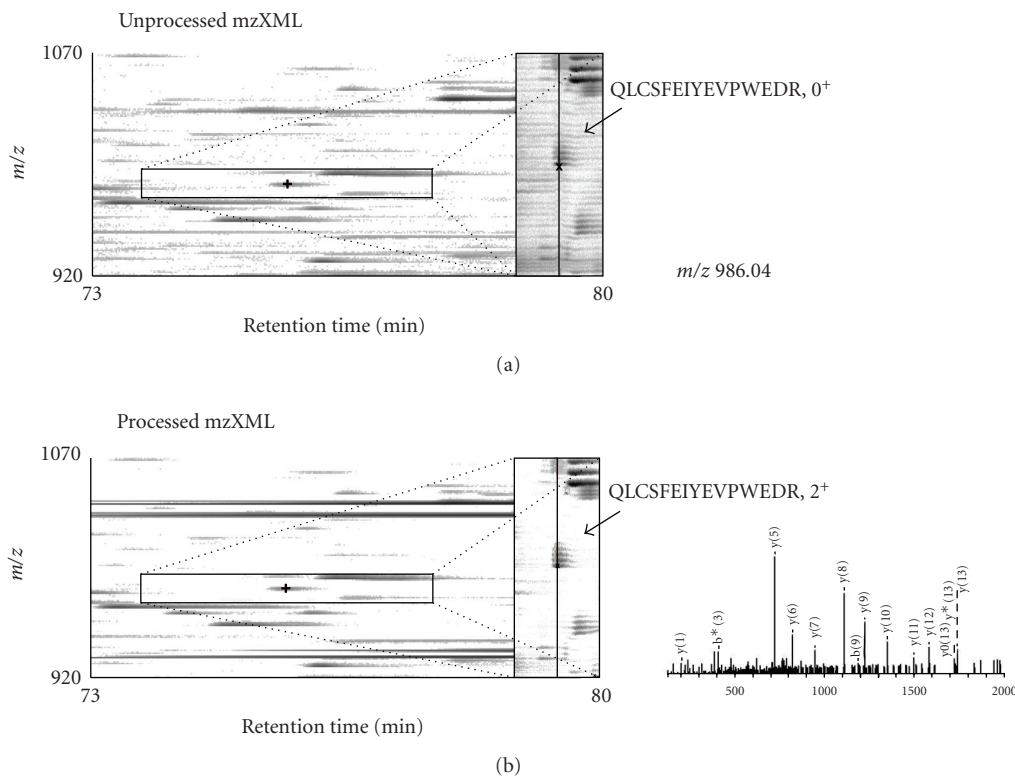
(a)

(b)

FIGURE 4: Details of unprocessed and processed LC-MS maps. Despite its high intensity, the highlighted doubly charged peptide QLCSFEIYEVPWEDR (m/z 986.04) was not assigned a charge in the original map, because some of its isotopes were altered by the chemical noise.

The general improvement of the quality of the common peptides is also evident in Figure 3. The x-axis gives the average intensity of two aligned peaks, while the y-axis shows the logarithm of their KL ratio. KL is the Kullback-Leibler score and it is used in msInspect as a measure of how closely the observed isotopic distribution of a peptide feature matches its own theoretical distribution at a given mass. This score is always nonnegative and approaches zero for better matches. Therefore all peptides which benefit from a correct denoising will gain a lower KL, will have a ratio $KL_{cl}/KL_{or} < 1$, and will be located below zero on the logarithmic axis. On the opposite, the peptides located above zero are those negatively affected by noise rejection, while points close to the horizontal axis indicate unaffected scores.

The figure thus shows that most of the peptides gained a better (lower) score and that the best improvements were obtained at very low intensities, therefore by those peptides originally masked or hidden by the chemical noise.

The second and the third sectors of the chart show peptides which have appeared or disappeared in the processed maps as a consequence of the preprocessing step. This result is more difficult to interpret because of the high complexity of the samples, which makes it almost impossible to infer which of the found peptides were true positives and which of the lost peptides were true negatives. Nevertheless, visual comparison of original versus clean maps confirms the detection of new low-intensity peaks [6], belonging to sector 2. It also shows that most of the peptides lost in sector 3

are uncharged features which are assigned a charge only after cleaning, like the peptide highlighted in Figure 4.

3.2. Effects of Denoising on Protein Identification. The ProSE toolkit was used to identify over-expressed and under-expressed peptides in the maps of the HDE subjects compared to the H controls. The same statistical analysis was then repeated for the processed data, thus producing a total of 4 include lists (IL), subsequently used for peptide identification by tandem MS (Figure 1(b)).

The results of the MS/MS analysis are summarized in Table 2. Since no processing is performed on tandem MS spectra, the quality of protein identification can be directly ascribed to the quality of the include lists. Despite an expected variability of the identified peptides, which is mostly related to experimental conditions of the targeted acquisitions, the general trend shows a clear improvement of protein identification obtained from the processed data.

Considering the protein under-expressed in HDE patients, the denoising strategy allowed the identification of 4 new proteins and achieved a higher sequence coverage and a better protein score for three proteins already found with the unprocessed data. In particular, Secretoglobin (SG1D1) had 0 peptides in the raw IL and 5 in the clean one, Mammaglobin-B (SG2A1) had 0 versus 3, Ig kappa chain C region (IGKC) had 0 versus 2, Cystatin (CYTS) had 0 versus 4, while Lipocalin-1 (LCN1) had 6 versus 12, Proline-rich

TABLE 2: List of the peptides identified using the include lists from unprocessed and processed data. Bold italic proteins indicate abundance variations in agreement with previous outcomes [13].

(a) Over-expressed proteins in hyperevaporative dry eye patients

Name	Entry name	m/z	Charge	Sequence	Score	
					Unprocessed	Processed
Hemoglobin subunit bete	HBB_HUMAN	590.89	2	GVANALAHKYH	16	18
		862.91	3	GTFATLSELHCDKLHVDPENFR	40	59
		637.89	2	LLVVYPWTQR	31	30
		737.508	3	TLSELHCDKLHVDPENFR	51	—
		725.44	2	VVAGVANALAHKYH	33	27
		835.59	2	VLGAFSDGLAHLDNLK	—	74
		1030.15	2	FFESFGDLSTPDAVMGNPK	—	17
		888.95	2	LLGNVLVCVLAHHFGK	—	49
		899.64	2	KVLGAFSDGLAHLDNLK	—	53
		575.39	2	VVAGVANALAHK	—	28
		524.805	2	VVYPWTQR	—	14
		810.203	3	FATLSELHCDKLHVDPENFR	—	15
		626.39	2	AGVANALAHKYH	—	17
		786.03	2	LGAFSDGLAHLDNLK	—	22
				Total number of peptides	**5**	**13**
				Protein score	**86**	**169**
Serum albumin	ALBU_HUMAN	756.49	2	VPQVSTPTLVEVSR	47	—
		613.88	2	FKDLGEENFK	5	17
		717.46	2	KECCEKPLLEK	51	23
		820.49	2	KVPQVSTPTLVEVSR	—	85
		722.35	2	YICENQDSISSK	—	30
		871.95	2	HPYFYAPELLFFAK	—	35
		926.8	3	LVRPEVDVMCTAFHDNEETFLKK	—	29
				Total number of peptides	**3**	**_6_**
				Protein score	**71**	**96**
Hemoglobin subunit alpha	HBA_HUMAN	523.96	3	FPHFDLSHGSAQVK	—	2
		663.37	2	HFDLSHGSAQVK	—	30
		942.98	3	DALTNAVAHVDDMPNALSALSDLHAHK	—	46
		785.92	4	KVADALTNAVAHVDDMPNALSALSDLHAHK	—	6
				Total number of peptides	**—**	**4**
				Protein score	**—**	**55**

(b) Under-expressed proteins in hyperevaporative dry eye patients

Name	Entry name	m/z	Charge	Sequence	Score	
					Unprocessed	Processed
Lipocalin-1	LCN1_HUMAN	669.86	2	TDEPGKYTADGGK	34	48
		722.94	2	YCEGELHGKPVR	16	—
		707.44	3	DHYIFYCEGELHGKPVR	97	36
		643.64	4	SHVKDHYIFYCEGELHGKPVR	30	15
		874.54	2	NNLEALEDFEKAAGAR	67	84
		934.6	2	YIFYCEGELHGKPVR	14	—
		496.84	2	GELHGKPVR	—	10
		561.36	2	EGELHGKPVR	—	20
		593.88	2	AVLEKTDEPGK	—	28
		641.38	2	CEGELHGKPVR	—	2
		796.48	2	FYCEGELHGKPVR	—	36
		885.5	2	SDEEIQDVSGTWYLK	—	56
		1171.21	2	HHLLASDEEIQDVSGTWYLK	—	102
		857.88	3	SHVKDHYIFYCEGELHGKPVR	—	47
				Total number of peptides	**6**	*12*
				Protein score	**136**	*257*
Polymeric immunoglobulin receptor	PIGR_HUMAN	703.95	2	VLDSGFREIENK	38	44
		876.04	2	TVTINCPFKTENAQK	11	—
		854.01	2	QSSGENCDVVVNTLGK	78	47
		573.35	2	QGARGGCITLI	—	3
		775.97	2	TINCPFKTENAQK	—	6
				Total number of peptides	**3**	**4**
				Protein score	**81**	**63**
Lactotransferrin	TRFL_HUMAN	694.94	2	GPQYVAGITNLKK	16	—
		624.37	2	WCAVGEQELR	15	—
		790.48	2	NLLFNDNTECLAR	37	—
		874.03	2	RSVQWCAVSQPEATK	74	—
		630.91	2	GPQYVAGITNLK	—	21
		702.89	2	QWCAVSQPEATK	—	9
		706.93	2	PIQCIQAIAENR	—	62
		841.97	2	CSTSPLLEACEFLR	—	70
		886.06	2	RDSPIQCIQAIAENR	—	6
		953.55	3	SQQSSDPDPNCVDRPVEGYLAVAVVR	—	30
		982.55	2	SASCVPGADKGQFPNLCR	—	56
				Total number of peptides	**4**	*8*
				Protein score	**77**	*141*
Serum albumin	ALBU_HUMAN	1020.23	2	VFDEFKPLVEEPQNLIK	38	—
		773.99	2	LKECCEKPLLEK	35	—
				Total number of peptides	**2**	—
				Protein score	**46**	—
Ig kappa chain C region	IGKC_HUMAN	714.45	3	HKVYACEVTHQGLSSPVTK	—	79
		1071.18	2	HKVYACEVTHQGLSSPVTK	—	34
				Total number of peptides	—	**2**
				Protein score	—	**79**

Name	Entry name	m/z	Charge	Sequence	Score	
					Unprocessed	Processed
Proline-rich protein 4	PROL4_HUMAN	815.2	2	DRPARHPQEQPLW	43	15
		801.48	2	FPSVSLQEASSFFR	—	84
		537.66	3	PSVSLQEASSFFRR	—	13
		598.38	2	HPPPPPFQNQQRPPR	—	15
		552.68	3	PPPPPFQNQQRPPR	—	27
				Total number of peptides	**1**	**5**
				Protein score	**43**	**84**
Secretoglobin family 1D member 1	SG1D1_HUMAN	869.57	1	APLEAVAAK	—	24
		572.9	2	FKAPLEAVAAK	—	39
		988.67	2	QALGSEITGFLLAGKPVFK	—	96
		712.8	3	CQALGSEITGFLLAGKPVFK	—	71
		622.86	2	KCVDTMAYEK	—	10
				Total number of peptides	—	*5*
				Protein score	—	*188*
Mammaglobin-B	SG2A1_HUMAN	732.95	2	FKQCFLNQSHR	—	45
		969	2	ELLQEFIDSDAAAEAMGK	—	81
		977.54	2	ELLQEFIDSDAAAEAMGK + oxM	—	7
				Total number of peptides	—	*3*
				Protein score	—	*89*
Cystatin-S	CYTS_HUMAN	964.05	2	PNLDTCAFHEQPELQK	—	21
		685.74	3	PNLDTCAFHEQPELQKK	—	29
		1135.65	2	SQPNLDTCAFHEQPELQKK	—	6
		986.04	2	QLCSFEIYEVPWEDR	—	67
				Total number of peptides	—	**4**
				Protein score	—	**70**

protein 4 (PROL4) had 1 versus 5, and Lactotransferrin (TRFL) had 4 versus 8. One protein was found with a better sequence coverage but a lower protein score (Polymeric Immunoglobulin Receptor (PIGR): 3 versus 4) and a last one was found only before filtering (Serum Albumin (ALBU): 2 versus 0).

Among the proteins over-expressed in HDE patients, 1 protein was found only after noise rejection (Hemoglobin subunit alpha (HBA): 0 versus 4) and 2 proteins achieved a higher sequence coverage and a better score (Hemoglobin subunit beta (HBB): 5 versus 13 and ALBU: 3 versus 6). The obtained results have been compared with the outcomes of a previous study performed by our research group [13], in which the differential expression of proteins in HDE patients over H controls was monitored by mono-dimensional gel electrophoresis and western blot analysis. Under-expressions of LCN1, SG1D1, SG2A1, TRFL, and over-expression of ALBU are in perfect agreement with the previous study and these proteins are shown in bold italic in Table 2. In particular, noise filtering allowed to identify ALBU as over-expressed in HDE patients, whereas the standard pipeline wrongly identified the protein as both under- and over-expressed.

As regards the variations in the abundances of HBA, HBB, IGKC, PROL4, PIGR, and CYTS associated to HDE, this could not be validated by comparison with previously published data. Nevertheless, a clear proof of their proper identification can be observed in Figure 4, where the correct charge assignment of the CYTS peptide QLCSFEIYEVP-WEDR allowed its inclusion in the list of peptides under-expressed in tear samples from HDE patients.

4. Conclusions

We have previously shown that wavelet denoising in the RT domain achieves selective rejection of chemical and random noise while preserving peptides features and morphology. The approach has proved to unveil low-intensity peptides originally masked by the chemical noise and to reduce false positive identification, by filtering noise peaks originally mimicking the peptide morphology.

In this work we have applied our noise filtering strategy to a label-free LC-MS differential analysis of protein abundance in tears samples. The mzXML files have been simply intercepted, processed by our algorithm and reinserted in the standard workflow just before the analyses by msInspect.

The results show that noise rejection allows to increase the sensitivity of msInspect to real peptides and to obtain more accurate include lists for further targeted MS/MS analysis. These results are validated by comparison to previous outcomes which confirm an improvement in terms of number of identified proteins, higher sequence coverage, and better protein scores.

Acknowledgments

References

[1] A. Iliuk, J. Galan, and W. A. Tao, "Playing tag with quantitative proteomics," *Analytical and Bioanalytical Chemistry*, vol. 393, no. 2, pp. 503–513, 2009.

[2] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, and B. Kuster, "Quantitative mass spectrometry in proteomics: a critical review," *Analytical and Bioanalytical Chemistry*, vol. 389, no. 4, pp. 1017–1031, 2007.

[3] A. H. P. America and J. H. G. Cordewener, "Comparative LC-MS: a landscape of peaks and valleys," *Proteomics*, vol. 8, no. 4, pp. 731–749, 2008.

[4] M. Hilario, A. Kalousis, C. Pellegrini, and M. Müller, "Processing and classification of protein mass spectra," *Mass Spectrometry Reviews*, vol. 25, no. 3, pp. 409–449, 2006.

[5] J. Listgarten and A. Emili, "Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry," *Molecular and Cellular Proteomics*, vol. 4, no. 4, pp. 419–434, 2005.

[6] S. Cappadona, F. Levander, M. Jansson, P. James, S. Cerutti, and L. Pattini, "Wavelet-based method for noise characterization and rejection in high-performance liquid chromatography coupled to mass spectrometry," *Analytical Chemistry*, vol. 80, no. 13, pp. 4960–4968, 2008.

[7] P. Nanni, L. Mezzanotte, G. Roda, et al., "Differential proteomic analysis of HT29 Cl.16E and intestinal epithelial cells by LC ESI/QTOF mass spectrometry," *Journal of Proteomics*, vol. 72, no. 5, pp. 865–873, 2009.

[8] P. Nanni, F. Levander, G. Roda, A. Caponi, P. James, and A. Roda, "A label-free nano-liquid chromatography-mass spectrometry approach for quantitative serum peptidomics in Crohn's disease patients," *Journal of Chromatography B*, vol. 877, no. 27, pp. 3127–3136, 2009.

[9] W. Windig, J. M. Phalp, and A. W. Payne, "A noise and background reduction method for component detection in liquid chromatography/mass spectrometry," *Analytical Chemistry*, vol. 68, no. 20, pp. 3602–3606, 1996.

[10] V. P. Andreev, T. Rejtar, H.-S. Chen, E. V. Moskovets, A. R. Ivanov, and B. L. Karger, "A universal denoising and peak picking algorithm for LC-MS based on matched filtration in the chromatographic time domain," *Analytical Chemistry*, vol. 75, no. 22, pp. 6314–6326, 2003.

[11] M. Bellew, M. Coram, M. Fitzgibbon, et al., "A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS," *Bioinformatics*, vol. 22, no. 15, pp. 1902–1909, 2006.

[12] J. Häkkinen, G. Vincic, O. Månsson, K. Wårell, and F. Levander, "The proteios software environment: an extensible multiuser platform for management and analysis of proteomics data," *Journal of Proteome Research*, vol. 8, no. 6, pp. 3037–3043, 2009.

[13] P. Versura, P. Nanni, A. Bavelloni, et al., "Tear protein changes in mild evaporative dry eye," submitted.

[14] "Report of the Dry Eye Workshop (DEWS)," *The Ocular Surface*, vol. 5, no. 2, 2007.

[15] P. Versura, M. Frigato, R. Mulé, N. Malavolta, and E. C. Campos, "A proposal of new ocular items in Sjögren's syndrome classification criteria," *Clinical and Experimental Rheumatology*, vol. 24, no. 5, pp. 567–572, 2006.