

# Supporting Information

## Optimized Sample Preparation Workflow for Improved Identification of Ghost Proteins

Tristan Cardon<sup>1\*</sup>, Flore Hervé<sup>1\*</sup>, Vivian Delcourt<sup>1,2</sup>, Xavier Roucou<sup>1,2</sup>, Michel Salzet<sup>1,3</sup>, Julien Franck<sup>1\*\*</sup> and Isabelle Fournier<sup>1,3\*\*</sup>

<sup>1</sup>Université de Lille, Inserm, U1192 - Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM), F-59000 Lille, France

<sup>2</sup>Department of Biochemistry, Université de Sherbrooke, Quebec, Canada

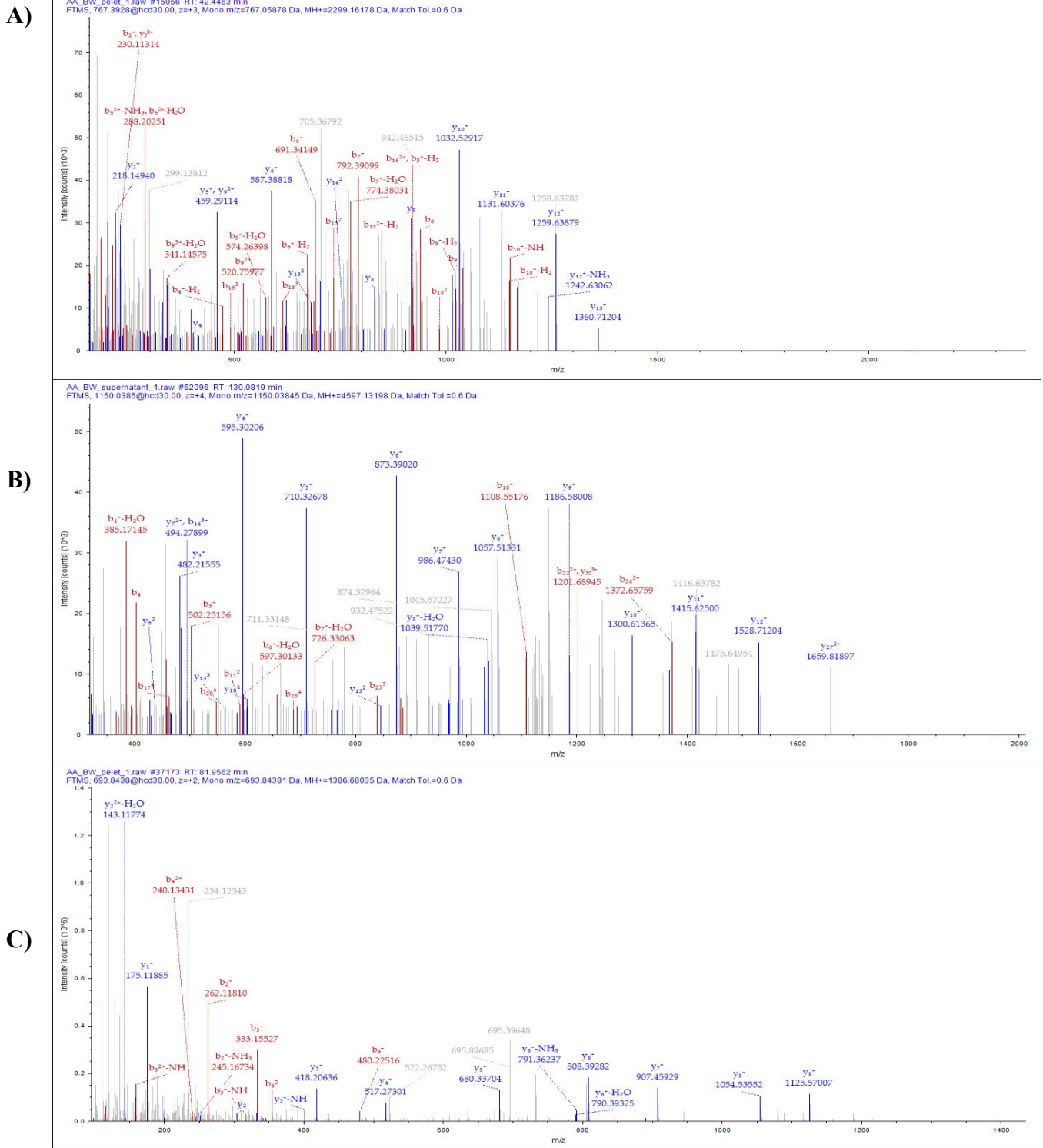
<sup>3</sup>Institut Universitaire de France (IUF), Paris, France

### **\*\*Corresponding authors**

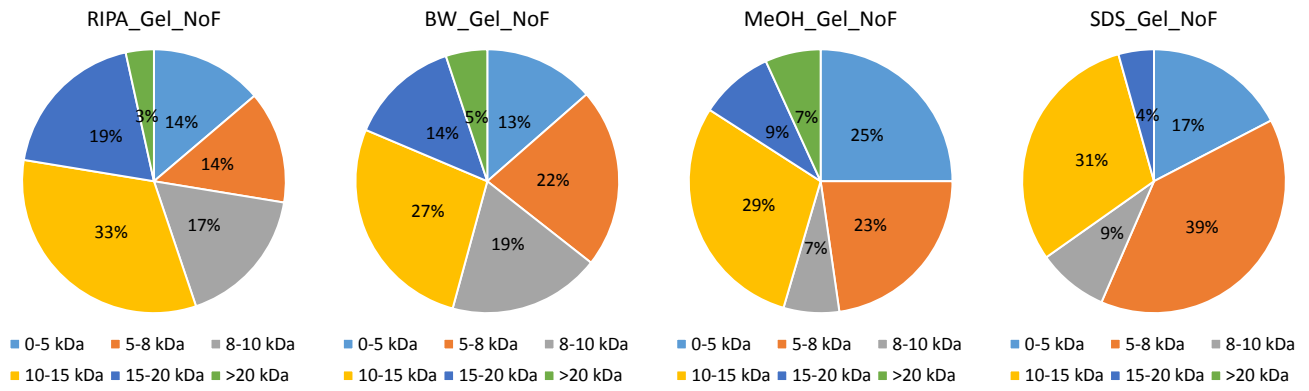
[isabelle.fournier@univ-lille.fr](mailto:isabelle.fournier@univ-lille.fr) and [julien.franck@univ-lille.fr](mailto:julien.franck@univ-lille.fr)

ORCID : Isabelle Fournier: 0000-0003-1096-5044 & Julien Franck: 0000-0002-7443-1706

# Supplementary Figures



**Supplementary Figure 1.** MS<sup>2</sup> spectra of 3 AltProt identified using boiling water (BW) for extraction and acetic acid (AA) precipitation for enrichment. These AltProt were identified based on a unique peptide. **A)** IP\_592880 (KTCTTVTFTQVNSEDKGALAK) **B)** IP\_572421 (VSDTVIESYNATLSVHQLVENTDETYCIDNEALYDICFR) and **C)** IP\_774870 (NFAFVQYVNER).



**Supplementary Figure 2.** Distribution in molecular weight (MW) of the AltProts according to the 4 extractions methods. No fractionation was performed post-extraction.

Supplementary Tables:

A		Mean ratio RefProts $\leq$ 15kDa/RefProts (%)	Associated Mean Standard deviation
extraction	enrichment		
RIPA	NoF	2.94	0.19
	Gel	8.23	0.43
	AA	2.74	0.25
	TCA	6.79	0.71
BW	NoF	9.51	0.35
	Gel	7.37	0.16
	AA	2.86	0.20
	TCA	4.68	1.29
MeOH	NoF	4.35	0.86
	Gel	4.07	1.83
SDS	NoF	2.83	1.50
	Gel	7.42	0.67
B		Mean ratio AltProts $\leq$ 15 kDa /RefProts (%)	Associated Mean Standard deviation
extraction	enrichment		
RIPA	NoF	7.81	0.85
	Gel	9.03	0.84
	AA	75.39	16.41
	TCA	7.20	0.51
BW	NoF	6.91	4.92
	Gel	21.72	0.63
	AA	71.54	3.92
	TCA	6.51	4.36
MeOH	NoF	3.89	2.59
	Gel	6.86	1.68
SDS	NoF	38.10	41.27
	Gel	10.72	1.12

**Supplementary Table 1.** Ratio of low MW ( $\leq$  15 kDa) to the total proteins. **(A)** For the RefProts ( $\leq$ 15 kDa RefProts /Total RefProts) and **(B)** for the AltProts ( $\leq$ 15 kDa AltProts/total RefProts).

Extraction	Enrichment	AltProts/RefProts mean (%)	Mean standard deviation (%)
RIPA	NoF	0,22	0,02
	Gel	0,80	0,03
	AA	1,08	0,14
	TCA	0,67	0,03
BW	NoF	0,81	0,59
	Gel	1,85	0,02
	AA	1,23	0,10
	TCA	0,22	0,08
	C18	0,30	0,01
MeOH	NoF	0,22	0,15
	Gel	0,45	0,05
SDS	NoF	0,29	0,25
	Gel	1,03	0,21

**Supplementary Table 2.** Ratio of the number of identified AltProts to the number of identified RefProts for the 4 different extractions (RIPA, BW, MeOH, SDS) and different enrichments. As indicated the experiments were conducted in triplicates (n = 3, SD).

Extraction	Enrichment	mean (%)	Mean standard deviation (%)
RIPA	NoF	9,23	1,42
	Gel	8,37	2,85
	AA	11,24	4,47
	TCA	9,19*	1,87
BW	NoF	11,20	0,64
	Gel	7,91	1,16
	AA	10,57	1,71
	TCA	15,88	2,82
	C18	7,79	1,91
MeOH	NoF	18,48	6,02
	Gel	28,91*	9,77
SDS	NoF	8,30*	1,07
	Gel	11,80	0,58

**Supplementary Table 3.** Variation in the identified RefProts for the different extractions and enrichment methods. Variation are calculated as the ratio of the number of specific proteins identified to the number of shared proteins in percentage for the triplicates (\* n=2).

protein accession	protein length (a.a.)	molecular weight (kDa)	isoelectric point	gene symbol	type	AA_BW	AA_RIPA	Gel_BW	Gel_MeOH	Gel_RIPA	Gel_SDS	TCA_BW	TCA_RIPA
IP_587085	148	16,62	11,73	RP11-475C16.1	ncRNA	X	X	X	X	X	X	X	X
IP_595471	204	24,17	12,13	RPL15P3	ncRNA	X	X	X	X	X		X	X
IP_759887	254	28,85	7,2	RP11-490H24.5	ncRNA	X	X	X		X	X	X	X
IP_591792	153	17,24	4,62	ACTBP8	ncRNA	X	X	X		X	X	X	X
IP_746392	266	30,11	11,44	RPL7AP6	ncRNA	X		X		X	X	X	X
IP_781237	145	17,27	11,11	RPL26P30	ncRNA	X	X	X			X	X	X
IP_584395	160	18,59	11,17	RPL21P75	ncRNA	X	X	X		X	X		X
IP_734708	71	7,77	6,48	RP11-24M17.3	ncRNA	X	X	X		X			X
IP_592880	90	10,1	10,08	RP3-486D24.1	ncRNA	X	X	X		X	X		
IP_736003	264	29,97	10,47	RPS3AP6	ncRNA	X	X					X	X
IP_688853	58	6,26	7,51	RP11-15H20.6	ncRNA			X	X	X	X		
IP_637160	255	28,4	9,5	RP11-395L14.17	ncRNA	X						X	X
IP_602798	92	10,25	6,79	RP11-509M23.1	ncRNA		X	X					X
IP_572421	229	25,27	4,43	TUBBP1	ncRNA	X	X						X
IP_559683	115	13,04	11,63	GS1-184P14.2	ncRNA	X	X						X
IP_613138	44	5,06	4,72	TMSB4XP8	ncRNA					X	X	X	
IP_592855	165	17,99	7,41	PPIAP9	ncRNA			X		X	X		
IP_662403	141	15,58	4,28	NPM1P19	ncRNA			X		X	X		
IP_556680	58	6,62	10,39	USMG5P1	ncRNA			X		X	X		
IP_603809	189	20,68	10,41	RP11-79P5.10	ncRNA			X		X	X		

IP_691726	74	8,32	10,53	CTC-398G3.1	ncRNA			X		X	X		
IP_624545	108	12,55	10,57	AC092798.2	ncRNA			X		X	X		
IP_723386	88	10,2	10,98	RPL18P13	ncRNA			X		X	X		
IP_565092	115	13,02	11,35	RPS26P3	ncRNA			X		X	X		
IP_557390	109	12,26	6,77	RP5-878I13.1	ncRNA	X	X			X			
IP_565117	92	10,68	8,46	SNRPEP2	ncRNA	X	X			X			
IP_641478	101	11,42	7,67	HNRNPA1P66	ncRNA						X		X
IP_592933	64	7,11	10,33	BTF3P7	ncRNA			X					X
IP_670480	76	8,43	9,49	HNRNPA1P68	ncRNA	X							X
IP_612504	136	15,23	11,61	H3F3AP6	ncRNA	X							X
IP_648412	86	9,11	7,81	ERCC3	ncRNA					X	X		
IP_737334	106	12,03	9,82	ACTBP7	ncRNA					X	X		
IP_556923	142	16,18	11,25	H3F3AP5	ncRNA					X	X		
IP_593099	225	24,94	4,31	TUBB2BP1	ncRNA			X			X		
IP_595290	130	14,48	8,68	ASS1P1	ncRNA			X			X		
IP_709935	222	23,77	9,41	UBE2SP1	ncRNA			X			X		
IP_634654	129	15,44	10,88	AC009302.2	ncRNA		X			X			
IP_689114	394	44,7	6,33	AC078899.1	ncRNA			X		X			
IP_557834	87	9,73	4,09	ACTBP1	ncRNA	X	X						
IP_655967	83	9,42	10,6	CTA-243E7.4	ncRNA	X	X						
IP_612062	215	23,4	4,13	NACA3P	ncRNA								X
IP_613753	74	8,75	4,53	TUBB4BP5	ncRNA								X
IP_572422	212	24,18	4,64	TUBBP1	ncRNA								X



IP_673532	144	16,46	4,73	EIF1AXP1	ncRNA													X
IP_667681	160	18	6,23	RP11-193H5.1	ncRNA													X
IP_2264871	84	9,92	4,62	LOC105372741	ncRNA													X
IP_776097	132	14,6	6,9	RP11-864N7.2	ncRNA													X
IP_649269	74	8,41	8,22	NCOA1	mRNA													X
IP_660647	73	7,16	4,93	UBASH3A	mRNA													X
IP_671420	212	23,3	5,57	PSAT1P3	ncRNA													X
IP_737074	85	9,48	7,6	EEF1A1P22	ncRNA													X
IP_755869	335	36,71	8,07	GAPDHP69	ncRNA													X
IP_761518	123	13,79	8,1	RPS4XP1	ncRNA													X
IP_671464	249	26,94	8,21	TPI1P1	ncRNA													X
IP_765292	56	6,51	10,91	RPL21P136	ncRNA													X
IP_592773	135	14,92	9,85	RP11-378G13.2	ncRNA				X									
IP_592972	95	10,67	5,18	PKMP5	ncRNA				X									
IP_619702	154	17,24	3,74	RPSAP69	ncRNA													X
IP_721455	187	21,62	4,32	TUBB8P7	ncRNA													X
IP_602542	239	27,24	4,78	RP11-274E7.2	ncRNA													X
IP_2390011	57	6,48	8,23	LOC105372901	ncRNA													X
IP_774832	168	17,95	8,66	GAPDHP70	ncRNA													X
IP_2364146	35	3,83	9,61	CFI	mRNA													X
IP_623043	80	8,83	9,61	EEF1A1P8	ncRNA													X
IP_669889	100	11,58	10,47	RP11-389O22.4	ncRNA													X

IP_639671	108	12,68	10,87	RP11-416L21.1	ncRNA					X			
IP_639834	149	17,59	10,94	RPL7P13	ncRNA					X			
IP_559678	225	24,93	4,34	EEF1B2P3	ncRNA			X					
IP_563312	44	5,11	4,72	TMSB4XP4	ncRNA			X					
IP_755940	36	4,11	4,87	HNRNPA1P30	ncRNA			X					
IP_671453	306	34,92	6,64	RP11-181C21.4	ncRNA			X					
IP_789374	103	11,09	7,15	GAPDHP21	ncRNA			X					
IP_580245	131	14,49	7,44	AC091654.7	ncRNA			X					
IP_592506	112	12,48	8,25	TUBBP9	ncRNA			X					
IP_587041	271	29,39	8,29	LDHAL6FP	ncRNA			X					
IP_624042	58	7	10,96	NDUFB1P1	ncRNA			X					
IP_579441	41	4,63	11,55	EEF1GP1	ncRNA			X					
IP_624921	50	5,66	4,19	ACTG1P12	ncRNA		X						
IP_593685	128	14,51	5,7	EEF1A1P42	ncRNA		X						
IP_591881	151	16,44	7,57	GAPDHP63	ncRNA		X						
IP_590800	228	24,4	9,7	GAPDHP72	ncRNA		X						
IP_641652	92	10,26	10,32	RP11-3304.2	ncRNA		X						
IP_623047	111	12,14	10,69	EEF1A1P8	ncRNA		X						
IP_622873	152	17,65	11,51	RP11-234A1.1	ncRNA		X						
IP_756756	135	15,15	5	FABP5P2	ncRNA	X							
IP_2286622	60	6,67	8,11	CDYL2	mRNA	X							
IP_612311	236	25,52	8,76	RTN3P1	ncRNA	X							
IP_736298	70	8,09	9,01	LINC01579	ncRNA	X							

IP_173179	82	9,52	11,85	NAA35	mRNA	x							
-----------	----	------	-------	-------	------	---	--	--	--	--	--	--	--

**Supplementary Table 4.** List of AltProts identified in the experiments. Red font accession numbers correspond to AltProts identified in each of the three replicates and black font those which were identified in two of the three replicates according the different extraction and enrichment methods (as indicated in the headers of the columns). Colored rows indicate AltProts that were detected for a single extraction/enrichment method.

	Protein Accession	Type	Transcript Accession	Gene	Gene Description	chromosome
Gel/Precipitation	IP_591792	ncRNA	ENST00000403258	ACTBP8	Actin, Beta Pseudogene 8	6
	IP_592880	ncRNA	ENST00000401715	AL136226.1	ribosomal protein L7A (RPL7A) pseudogene	6
	IP_734708	ncRNA	ENST00000567565	PPIAP47	peptidylprolyl isomerase A pseudogene 47	15
Precipitation	IP_655967	ncRNA	ENST00000623888	AL022323.3	novel transcript	22
	IP_602798	ncRNA	ENST00000425840	PPIAP79	peptidylprolyl isomerase A pseudogene 79	5
	IP_557834	ncRNA	ENST00000417985	ACTBP1	ACTB pseudogene 1	X
	IP_572421	ncRNA	ENST00000248151, ENST00000518096	TUBBP1	tubulin beta pseudogene 1	8
GEL	IP_688853	ncRNA	ENST00000598599	RP11-15H20.6		19
	IP_624545	ncRNA	ENST00000458542	AC092798.1	ribosomal protein L32 (RPL32) pseudogene	3
	IP_689114	ncRNA	ENST00000521432	AC078899.1	ARP2 actin-related protein 2 homolog (yeast) (ACTR2) pseudogene	19
	IP_592855	ncRNA	ENST00000403866	PPIAP9	peptidylprolyl isomerase A pseudogene 9	6
	IP_662403	ncRNA	ENST00000450070	NPM1P19	nucleophosmin 1 pseudogene 19	20
	IP_593099	ncRNA	ENST00000404155	TUBB2BP1	tubulin beta 2B class IIb pseudogene 1	6
	IP_691726	ncRNA	ENST00000483614	AC008481.2	ribosomal protein L18a (RPL18A) pseudogene	19
	IP_723386	ncRNA	ENST00000478088	RPL18P13	ribosomal protein L18 pseudogene 13	16
	IP_737334	ncRNA	ENST00000418351	ACTBP7	Actin, Beta Pseudogene 7	15

**Supplementary Table 5. Common and specific AltProts identified for gel and precipitation.** Each identified protein is present in all of the triplicates for each of the extraction method and at least three times out of four for the precipitations (AA\_BW, AA\_RIPA, TCA\_BW, TCA\_RIPA) and gel (Gel\_SDS, Gel\_MeOH, Gel\_RIPA, Gel\_BW).