

Table of Contents

1	Overview of the HUPO Plasma Proteome Project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database	1
	<i>Gilbert S. Omenn, David J. States, Marcin Adamski, Thomas W. Blackwell, Rajasree Menon, Henning Hermjakob, Rolf Apweiler, Brian B. Haab, Richard J. Simpson, James S. Eddes, Eugene A. Kapp, Robert L. Moritz, Daniel W. Chan, Alex J. Rai, Arie Admon, Ruedi Aebersold, Jimmy Eng, William S. Hancock, Stanley A. Hefta, Helmut Meyer, Young-Ki Paik, Jong-Shin Yoo, Peipei Ping, Joel Pounds, Joshua Adkins, Xiaohong Qian, Rong Wang, Valerie Wasinger, Chi Yue Wu, Xiaohang Zhao, Rong Zeng, Alexander Archakov, Akira Tsugita, Ilan Beer, Akhilesh Pandey, Michael Pisano, Philip Andrews, Harald Tammen, David W. Speicher and Samir M. Hanash</i>	
1.1	Introduction	2
1.2	PPP reference specimens	4
1.3	Bioinformatics and technology platforms	5
1.3.1	Constructing a PPP database for human plasma and serum proteins	5
1.3.2	Analysis of confidence of protein identifications	14
1.3.3	Quantitation of protein concentrations	15
1.4	Comparing the specimens	17
1.4.1	Choice of specimen and collection and handling variables	17
1.4.2	Depletion of abundant proteins followed by fractionation of intact proteins	19
1.4.3	Comparing technology platforms	22
1.4.4	Alternative search algorithms for peptide and protein identification	23
1.4.5	Independent analyses of raw spectra or peaklists	24
1.4.6	Comparisons with published reports	25
1.4.7	Direct MS (SELDI) analyses	27
1.4.8	Annotation of the HUPO PPP core dataset(s)	27
1.4.9	Identification of novel peptides using whole genome ORF search	30
1.4.10	Identification of microbial proteins in the circulation	30

1.5	Discussion	31
1.6	References	33

2 Data management and preliminary data analysis in the pilot phase of the HUPO Plasma Proteome Project 37

Marcin Adamski, Thomas Blackwell, Rajasree Menon, Lennart Martens, Henning Hermjakob, Chris Taylor, Gilbert S. Omenn and David J. States

2.1	Introduction	37
2.2	Materials and methods	39
2.2.1	Development of the data model	39
2.2.1.1	Laboratory	39
2.2.1.2	Experimental protocol	39
2.2.1.3	Protein identification data set	39
2.2.1.4	Peak list	41
2.2.1.5	Summary of technologies and resources	41
2.2.1.6	MS/MS spectra	41
2.2.1.7	SELDI peak list	42
2.2.2	Data submission process	42
2.2.3	Design of the data repository	42
2.2.4	Receipt of the data	43
2.3	Inference from peptide level to protein level	44
2.4	Summary of contributed data	46
2.4.1	Cross-laboratory comparison, confidence of the identifications	49
2.5	False-positive identifications	51
2.6	Data dissemination	56
2.7	Discussion	57
2.8	Concluding remarks	58
2.9	Computer technologies applied	60
2.10	References	61

3 HUPO Plasma Proteome Project specimen collection and handling: Towards the standardization of parameters for plasma proteome samples 63

Alex J. Rai, Craig A. Gelfand, Bruce C. Haywood, David J. Warunek, Jizu Yi, Mark D. Schuchard, Richard J. Mehig, Steven L. Cockrill, Graham B. I. Scott, Harald Tammen, Peter Schulz-Knappe, David W. Speicher, Frank Vitzthum, Brian B. Haab, Gerard Siest and Daniel W. Chan

3.1	Introduction	63
3.2	Materials and methods	65
3.2.1	HUPO reference sample collection protocol	65
3.2.2	Differential peptide display	66
3.2.3	Stability studies and SELDI analysis	66
3.2.4	SDS-PAGE analysis for stability studies	67

3.2.5	2-DE for stability studies	67
3.2.6	SELDI-TOF analysis for protease inhibitor studies	67
3.2.7	2-DE for plasma protease inhibition studies	68
3.2.8	Tryptic digestion and protein identification for protease inhibition studies	69
3.2.9	Antibody microarray analysis using two-color rolling circle amplification	69
3.3	Results	69
3.3.1	Comparisons of specimen types	71
3.3.1.1	Analysis of serum	71
3.3.1.2	Analysis of plasma	71
3.3.2	Evaluation of storage and handling conditions	71
3.3.3	Evaluations of the use of protease inhibitors	73
3.3.3.1	Analysis with SELDI-TOF MS of "time zero" effects of protease inhibitors in plasma	73
3.3.3.2	Analysis by 2-DE	73
3.3.3.3	Analysis with antibody arrays	76
3.4	Discussion	77
3.4.1	Other pre-analytical variables and control considerations	83
3.4.2	Reference materials	84
3.5	Concluding remarks	87
3.6	References	88

4 Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: Systematic variation between sample types and calibration of mass spectrometry data

Brian B. Haab, Bernhard H. Geierstanger, George Michailidis, Frank Vitzthum, Sara Forrester, Ryan Okon, Petri Saviranta, Achim Brinker, Martin Sorette, Lorah Perlee, Shubha Suresh, Garry Drwal, Joshua N. Adkins and Gilbert S. Omenn

4.1	Introduction	92
4.2	Materials and methods	93
4.2.1	Reference specimens	93
4.2.2	DB immunoassays	93
4.2.3	Antibody arrays at GNF	94
4.2.3.1	Antibodies, reagents, microarray printing, and platform	94
4.2.3.2	Microarray layout and processing	94
4.2.3.3	Array imaging and data analysis	95
4.2.4	Antibody microarrays at MSI	95
4.2.4.1	Chip manufacture	95
4.2.4.2	Rolling circle amplification (RCA) immunoassay	96
4.2.4.3	Conversion of mean fluorescent intensity to concentration	96

4.2.5	Antibody microarrays at VARI	96
4.2.5.1	Fabrication of antibody microarrays	96
4.2.5.2	Serum labeling	97
4.2.5.3	Processing of antibody microarrays	97
4.2.5.4	Analysis	97
4.2.6	Retrieval and matching of IPI numbers for the analytes	97
4.3	Results	98
4.3.1	Antibody-based measurements of the HUPO reference specimens	98
4.3.2	Systematic variation between the preparation methods of the PPP reference specimens	100
4.3.3	Consistent alterations in specific protein abundances	107
4.3.4	Linkage of MS data and antibody-based measurements	108
4.4	Discussion	110
4.5	References	113
5	Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma	115
	<i>Lynn A. Echan, Hsin-Yao Tang, Nadeem Ali-Khan, KiBeom Lee and David W. Speicher</i>	
5.1	Introduction	116
5.2	Materials and methods	117
5.2.1	Serum/plasma collection	117
5.2.2	MARS	118
5.2.3	Multiple affinity removal spin cartridge	118
5.2.4	Microscale solution IEF (MicroSol IEF) (ZOOM™-IEF) fractionation	118
5.2.5	2-DE	119
5.2.6	LC-MS/MS	119
5.3	Results	120
5.3.1	Depletion of major proteins to enhance detection of lower abundance proteins	120
5.3.2	Evaluation of high-abundance protein removal using 2-DE	121
5.3.3	Specificity of major protein depletion	123
5.3.4	Impact of Top-6 protein depletion on detection of lower abundance proteins using 2-D gels	125
5.3.5	Combining Top-6 protein depletion with microSol IEF prefractionation and narrow pH range gels	125
5.3.6	Analysis of Top-6 depleted serum and plasma using protein array pixelation	128
5.4	Discussion	130
5.5	References	134

6	A novel four-dimensional strategy combining protein and peptide separation methods enables detection of low-abundance proteins in human plasma and serum proteomes	135
	<i>Hsin-Yao Tang, Nadeem Ali-Khan, Lynn A. Echan, Natasha Levenkova, John J. Rux and David W. Speicher</i>	
6.1	Introduction	135
6.2	Materials and methods	138
6.2.1	Materials	138
6.2.2	Top six protein depletion	138
6.2.3	MicroSol-IEF fractionation	139
6.2.4	Protein array pixelation	139
6.2.5	LC-ESI-MS/MS methods	140
6.2.6	Data analysis	140
6.3	Results and discussion	141
6.3.1	Protein array pixelation strategy	141
6.3.2	Optimization of protein array pixelation	143
6.3.3	Total analysis time for protein array pixelation of human plasma proteome	146
6.3.4	Systematic protein array pixelation of the human plasma proteome	147
6.3.5	Systematic protein array pixelation of the human serum proteome	150
6.3.6	Analyses of human plasma and serum proteomes using HUPO filter criteria	153
6.4	Concluding remarks	157
6.5	References	157
7	A study of glycoproteins in human serum and plasma reference standards (HUPO) using multilectin affinity chromatography coupled with RPLC-MS/MS	159
	<i>Ziping Yang, William S. Hancock, Tori Richmond Chew and Leo Bonilla</i>	
7.1	Introduction	159
7.2	Materials and methods	160
7.2.1	Materials	160
7.2.2	Isolating glycoproteins using multilectin affinity columns	161
7.2.3	Analysis of glycoproteins on LC-LCQ MS	161
7.2.4	Analysis of glycoproteins on LC-LTQ MS	162
7.2.5	Protein database search	162
7.3	Results and discussion	162
7.3.1	Protein IDs from the plasma and serum samples	162
7.3.2	Comparison between serum and plasma glycoproteomes	179
7.3.3	Comparison of the glycoproteins present in the samples collected from three ethnic groups	179

7.4	Concluding remarks	182
7.5	References	183

8 Evaluation of prefractionation methods as a preparatory step for multidimensional based chromatography of serum proteins 185

Eilon Barnea, Raya Sorkin, Tamar Ziv, Ilan Beer and Arie Admon

8.1	Introduction	185
8.1.1	The HUPO Plasma Proteome Project (PPP) goals and the serum as a complex sample	185
8.1.2	The scope of this manuscript	187
8.2	Materials and methods	187
8.2.1	Depletion from serum albumin and antibodies	187
8.2.2	MudPIT and mass segmentation	187
8.2.3	Protein separation by SDS-PAGE	188
8.2.4	SCX separation of intact proteins followed by MudPIT	188
8.2.5	Liquid-phase IEF followed by MudPIT	188
8.2.6	Capillary RP-LC-MS/MS	189
8.2.7	MS data processing and peptide/protein identifications	189
8.3	Results	189
8.3.1	Comparisons between the prefractionation methods	190
8.3.2	Identification of different protein subsets	191
8.3.3	Proteins identified by only one prefractionation method	193
8.3.4	Different methods resulted in diverse peptide coverage	193
8.4	Discussion	196
8.4.1	Giving every peptide a chance	196
8.4.2	How to identify more of the marginal proteins	197
8.4.3	Clustering and comparing raw data	197
8.4.4	High throughput and ruggedness versus high sensitivity	197
8.4.5	The cost effectiveness of the different methods	198
8.5	Concluding remarks	198
8.6	References	199

9 Efficient prefractionation of low-abundance proteins in human plasma and construction of a two-dimensional map 201

Sang Yun Cho, Eun-Young Lee, Joon Seok Lee, Hye-Young Kim, Jae Myun Park, Min-Seok Kwon, Young-Kew Park, Hyoung-Joo Lee, Min-Jung Kang, Jin Young Kim, Jong Shin Yoo, Sung Jin Park, Jin Won Cho, Hyon-Suk Kim and Young-Ki Paik

9.1	Introduction	202
9.2	Materials and methods	203
9.2.1	Plasma sample preparation	203
9.2.2	Depletion of major abundance proteins with an immunoaffinity column	203

9.2.3	2-DE	204
9.2.4	Identification of proteins by MS	204
9.2.5	Fractionation of the plasma samples by FFE	204
9.2.6	LC-MS/MS	205
9.2.7	Bioinformatics	206
9.3	Results and discussion	206
9.3.1	2-DE map of human plasma devoid of high-abundance proteins	206
9.3.2	Expression of different anticoagulant-treated plasma	214
9.3.3	FFE/1-DE/nanoLC-MS/MS and 2-DE/MALDI-TOF	215
9.4	Concluding remarks	219
9.5	References	219
10	Comparison of alternative analytical techniques for the characterisation of the human serum proteome in HUPO Plasma Proteome Project	221
	<i>Xiaohai Li, Yan Gong, Ying Wang, Songfeng Wu, Yun Cai, Ping He, Zhuang Lu, Wantao Ying, Yangjun Zhang, Liyan Jiao, Hongzhi He, Zisen Zhang, Fuchu He, Xiaohang Zhao and Xiaohong Qian</i>	
10.1	Introduction	222
10.2	Materials and methods	223
10.2.1	Materials	223
10.2.2	Human serum samples	223
10.2.3	Integrated strategy for characterising analytical approaches	223
10.2.4	Depletion of the highly abundant serum proteins by MARS	224
10.2.5	Desalting and concentrating the flow-through fractions by centrifugal ultrafiltration	224
10.2.6	Fractionation of depleted serum samples by anion-exchange HPLC	225
10.2.7	Protein fractionation by 2-D HPLC with nonporous RP-HPLC	225
10.2.8	The 2-DE strategy for the analysis of serum proteins	226
10.2.8.1	2-DE	226
10.2.8.2	In-gel digestion <i>via</i> automated workstation	227
10.2.8.3	Protein spot identification by MALDI-TOF-MS/MS	227
10.2.9	Shotgun strategy for the analysis of serum proteins	228
10.2.9.1	Trypsin digestion of serum proteins	228
10.2.9.2	Protein identification by micro2-D LC-ESI-MS/MS	228
10.2.9.3	Data processing	229
10.2.10	Protein fractionation strategy for the analysis of serum proteins	229
10.2.10.1	2-D LC fractionation of serum proteins	229
10.2.10.2	Digestion of the 2-D LC separated fractions	229
10.2.10.3	1-D microRP-HPLC-ESI-MS/MS identification of digested serum proteins	230
10.2.11	Offline shotgun strategy for the analysis of serum proteins	230

10.2.11.1	Offline SCX for first-dimension chromatographic separation of peptides	230
10.2.11.2	1-D capillary RP-HPLC/microESI-IT-MS/MS analysis for the SCX-separated peptide fractions	231
10.2.12	Optimised nanoRP-HPLC-nanoESI IT-MS/MS for the reanalysis of offline SCX-separated peptides (offline-nanospray strategy)	231
10.3	Integrated analysis of the whole data sets	231
10.3.1	Protein grouping analysis	231
10.3.2	Sequence clustering	232
10.4	Results and discussion	233
10.4.1	Depletion of the highly abundant serum proteins	233
10.4.2	The 2-DE strategy for the analysis of serum proteins	233
10.4.3	2-D HPLC fractionation for the analysis of serum proteins	234
10.4.4	Shotgun strategy for the analysis of serum proteins with online SCX	237
10.4.5	Shotgun strategy for the analysis of serum proteins with offline SCX	237
10.4.6	Offline SCX shotgun-nanospray strategy for the analysis of serum proteins	239
10.4.7	Comparison of the five strategies for the analysis of the human serum proteome	241
10.5	Concluding remarks	246
10.6	References	246
11	A proteomic study of the HUPO Plasma Proteome Project's pilot samples using an accurate mass and time tag strategy	249
	<i>Joshua N. Adkins, Matthew E. Monroe, Kenneth J. Auberry, Yufeng Shen, Jon M. Jacobs, David G. Camp II, Frank Vitzthum, Karin D. Rodland, Richard C. Zangar, Richard D. Smith and Joel G. Pounds</i>	
11.1	Introduction	250
11.2	Materials and methods	251
11.2.1	Human blood serum and plasma	251
11.2.2	Depletion of Igs and trypsin digestion	252
11.2.3	Peptide cleanup	252
11.2.4	Capillary RP-LC	253
11.2.5	IT-MS	254
11.2.6	SEQUEST identification of peptides	254
11.2.7	Putative mass and time tag database from SEQUEST results	254
11.2.8	FT-ICR-MS	255
11.2.9	cLC-FT-ICR MS data analysis	255
11.2.10	OmniViz cluster and visual analysis	257
11.3	Results	257
11.3.1	PuMT tag database	257

11.3.2	Summary of peptide/protein identifications by AMT tags	258
11.3.3	Protein concentration estimates from ion current	260
11.3.4	Global protein analysis	261
11.4	Discussion	264
11.4.1	Application of FT-ICR MS as a proteomic technology bridge	264
11.4.2	Confidence in any MS-based proteomic approach	266
11.4.3	Peptide/protein redundancy	267
11.4.4	Identification sensitivity versus specificity	267
11.4.5	Throughput and differential analysis	269
11.5	References	270
12	Analysis of Human Proteome Organization Plasma Proteome Project (HUPO PPP) reference specimens using surface enhanced laser desorption/ionization-time of flight (SELDI-TOF) mass spectrometry: Multi-institution correlation of spectra and identification of biomarkers	273
	<i>Alex J. Rai, Paul M. Stemmer, Zhen Zhang, Bao-ling Adam, William T. Morgan, Rebecca E. Caffrey, Vladimir N. Podust, Manisha Patel, Lih-yin Lim, Natalia V. Shipulina, Daniel W. Chan, O. John Semmes and Hon-chiu Eastwood Leung</i>	
12.1	Introduction	273
12.2	Materials and methods	275
12.2.1	Sample preparation	275
12.2.2	Sample preprocessing	275
12.2.3	Target (CM10) chip preparation and sample incubation	275
12.2.4	Scanning protocol	276
12.2.5	Data processing	276
12.2.6	Bioinformatics analysis of data and correlation coefficient matrix	276
12.2.7	Protein purification, SDS-PAGE analysis, and extraction of proteins	276
12.2.8	Peptide mass fingerprinting (PMF)	277
12.2.9	MS/MS analysis	277
12.2.10	Western blot analysis	277
12.3	Results	278
12.4	Discussion	283
12.5	References	286
13	An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis	289
	<i>Eugene A. Kapp, Frédéric Schütz, Lisa M. Connolly, John A. Chakel, Jose E. Meza, Christine A. Miller, David Fenyo, Jimmy K. Eng, Joshua N. Adkins, Gilbert S. Omenn and Richard J. Simpson</i>	
13.1	Introduction	289
13.1.1	Heuristic algorithms	291

13.1.2	Probabilistic algorithms	292
13.2	Materials and methods	292
13.2.1	HUPO-PPP reference specimens	292
13.2.2	Sample preparation and MS analysis	293
13.2.3	Protein sequence databases	293
13.2.4	MS/MS database search strategy	293
13.2.4.1	SEQUEST and MASCOT workflow performed by the JPSL research group	294
13.2.4.2	SEQUEST and PeptideProphet workflow performed by the ISB research group	294
13.2.4.3	Spectrum Mill workflow performed by the Agilent group	295
13.2.4.4	Sonar and X!Tandem workflow performed by David Fenyo	295
13.2.5	Web interface for data validation, integration, and cross annotation	295
13.2.6	ROC curve generation	297
13.3	Results and discussion	298
13.3.1	Comparison of MS/MS search algorithms	299
13.3.1.1	Sensitivity and concordance between MS/MS search algorithms	299
13.3.1.2	Specificity and discriminatory power of the primary score statistic for the different MS/MS search algorithms: Distribution of scores and ROC plots	301
13.3.1.3	Calculation of score thresholds based on specified FP identification error rates	304
13.3.1.4	Benchmarking of the different MS/MS search algorithms at 1% FP error rate	310
13.3.1.5	Effect of database size and search strategy	311
13.3.1.6	Utility of reversed sequence searches	311
13.3.1.7	Consensus scoring between MS/MS search algorithms	312
13.4	Concluding remarks	313
13.5	References	314
14	Human Plasma PeptideAtlas	317
	<i>Eric W. Deutsch, Jimmy K. Eng, Hui Zhang, Nichole L. King, Alexey I. Nesvizhskii, Biaoyang Lin, Hookeun Lee, Eugene C. Yi, Reto Ossola and Ruedi Aebersold</i>	
14.1	References	322
15	Do we want our data raw? Including binary mass spectrometry data in public proteomics data repositories	323
	<i>Lennart Martens, Alexey I. Nesvizhskii, Henning Hermjakob, Marcin Adamski, Gilbert S. Omenn, Joël Vandekerckhove and Kris Gevaert</i>	
15.1	References	328

16	A functional annotation of subproteomes in human plasma	329
	<i>Peipei Ping, Thomas M. Vondriska, Chad J. Creighton, TKB Gandhi, Ziping Yang, Rajasree Menon, Min-Seok Kwon, Sang Yun Cho, Garry Drwal, Markus Kellmann, Suraj Peri, Shubha Suresh, Mads Gronborg, Henrik Molina, Raghothama Chaerkady, B. Rekha, Arun S. Shet, Robert E. Gerszten, Hoifeng Wu, Mark Raftery, Valerie Wasinger, Peter Schulz-Knappe, Samir M. Hanash, Young-ki Paik, William S. Hancock, David J. States, Gilbert S. Omenn and Akhilesh Pandey</i>	
16.1	Introduction	330
16.2	Materials and methods	330
16.2.1	Coagulation pathway and protein interaction network analysis	331
16.2.2	Gene ontology annotations	331
16.2.3	Analysis of MS-derived data for identification of proteolytic events and post-translational modifications	331
16.3	Results and discussion	331
16.3.1	Bioinformatic analyses of the functional subproteomes	332
16.3.1.1	An interaction map of human plasma proteins	332
16.3.1.2	Gene Ontology annotation of protein function	334
16.3.2	Proteins involved in the blood coagulation pathway	335
16.3.3	Proteins potentially derived from mononuclear phagocytes	337
16.3.4	Proteins involved in inflammation	338
16.3.5	Analyzing the peptide subproteome of human plasma	339
16.3.6	Liver related plasma proteins	339
16.3.7	Cardiovascular system related plasma proteins	341
16.3.8	Glycoproteins	342
16.3.9	DNA-binding proteins	342
16.3.9.1	Histones	343
16.3.9.2	Helicases	344
16.3.9.3	Zinc finger proteins	345
16.3.10	Annotation through reanalysis of mass spectrometry data	345
16.3.10.1	Cleavage of signal peptides and transmembrane domains	346
16.3.10.2	Identification of PTMs	347
16.4	Concluding remarks	348
16.5	References	349
17	Cardiovascular-related proteins identified in human plasma by the HUPO Plasma Proteome Project Pilot Phase	353
	<i>Beniam T. Berhane, Chenggong Zong, David A. Liem, Aaron Huang, Steven Le, Ricky D. Edmondson, Richard C. Jones, Xin Qiao, Julian P. Whitelegge, Peipei Ping and Thomas M. Vondriska</i>	
17.1	Introduction	353
17.1.1	HUPO Plasma Proteome Project pilot phase	354

17.1.2	Need for novel insights into cardiovascular disease	354
17.2	Materials and methods	355
17.3	Groups of cardiovascular-related proteins	356
17.3.1	Markers of inflammation and CVD	356
17.3.2	Vascular and coagulation proteins	357
17.3.3	Signaling proteins	359
17.3.4	Growth- and differentiation-associated proteins	360
17.3.5	Cytoskeletal proteins	360
17.3.6	Transcription factors	361
17.3.7	Channel and receptor proteins	363
17.3.8	Heart failure- and remodeling-related proteins	364
17.4	Functional analyses and implications	365
17.4.1	Organ specific cardiovascular-related proteins in plasma	365
17.4.2	Novel cardiovascular-related proteins identified in plasma	366
17.5	Methodology considerations	368
17.6	Conclusions and future directions	368
17.7	References	370