

## REGULAR ARTICLE

# A functional annotation of subproteomes in human plasma

Peipei Ping<sup>1</sup>, Thomas M. Vondriska<sup>1</sup>, Chad J. Creighton<sup>2</sup>, TKB Gandhi<sup>3</sup>, Ziping Yang<sup>4</sup>, Rajasree Menon<sup>2</sup>, Min-Seok Kwon<sup>5</sup>, Sang Yun Cho<sup>5</sup>, Garry Drwal<sup>6</sup>, Markus Kellmann<sup>7</sup>, Suraj Peri<sup>8</sup>, Shubha Suresh<sup>3</sup>, Mads Gronborg<sup>8</sup>, Henrik Molina<sup>8</sup>, Raghobhama Chaerkady<sup>3</sup>, B. Rekha<sup>3</sup>, Arun S. Shet<sup>9</sup>, Robert E. Gerszten<sup>10</sup>, Haifeng Wu<sup>6</sup>, Mark Raftery<sup>11</sup>, Valerie Wasinger<sup>11</sup>, Peter Schulz-Knappe<sup>7</sup>, Samir M. Hanash<sup>12</sup>, Young-ki Paik<sup>5</sup>, William S. Hancock<sup>4</sup>, David J. States<sup>2</sup>, Gilbert S. Omenn<sup>12</sup> and Akhilesh Pandey<sup>8</sup>

<sup>1</sup> Departments of Physiology and Medicine, Division of Cardiology, Cardiovascular Research Laboratories, University of California, Los Angeles, CA, USA

<sup>2</sup> Bioinformatics Program, University of Michigan, Ann Arbor, MI, USA, Departments of Pediatrics<sup>12</sup> and Internal Medicine<sup>13</sup>, University of Michigan, Ann Arbor, MI, USA

<sup>3</sup> Institute of Bioinformatics, International Tech Park, Bangalore, India

<sup>4</sup> Chemistry and Chemical Biology Department, Barnett Institute, Northeastern University, Boston, MA 02115, USA

<sup>5</sup> Yonsei Proteome Research Center and Biomedical Proteome Research Center, Dept. of Biochemistry, Yonsei University, Seoul, Korea

<sup>6</sup> Department of Pathology, College of Medicine and Public Health, Ohio State University, Columbus, OH, USA

<sup>7</sup> BioVisioN AG, Hannover, Germany

<sup>8</sup> McKusick-Nathans Institute of Genetic Medicine and Departments of Biological Chemistry and Oncology, Johns Hopkins University, Baltimore, MD, USA

<sup>9</sup> Laboratory of Blood and Vascular Biology, Rockefeller University, NY, USA

<sup>10</sup> Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Charlestown, MA, USA

<sup>11</sup> Bioanalytical Mass Spectrometry Facility, University of NSW, Sydney, Australia

<sup>12</sup> University of Michigan, Ann Arbor, MI, USA

The data collected by Human Proteome Organization's Plasma Proteome Pilot project phase was analyzed by members of our working group. Accordingly, a functional annotation of the human plasma proteome was carried out. Here, we report the findings of our analyses. First, bioinformatic analyses were undertaken to determine the likely sources of plasma proteins and to develop a protein interaction network of proteins identified in this project. Second, annotation of these proteins was performed in the context of functional subproteomes involved in the coagulation pathway, the mononuclear phagocytic system, the inflammation pathway, the cardiovascular system, and the liver; as well as the subset of proteins associated with DNA binding activities. Our analyses contributed to the Plasma Proteome Database (<http://www.plasmaproteomedatabase.org>), an annotated database of plasma proteins identified by HPPP as well as from other published studies. In addition, we address several methodological considerations including the selective enrichment of post-translationally modified proteins by the use of multi-lectin chromatography as well as the use of peptidomic techniques to characterize the low molecular weight proteins in plasma. Furthermore, we have performed additional analyses of peptide identification data to annotate cleavage of signal peptides, sites of intramembrane proteolysis and post-translational modifications. The HPPP-organized, multi-laboratory effort, as described herein, resulted in much synergy and was essential to the success of this project.

Received: March 7, 2005

Accepted: April 16, 2005



## Keywords:

Annotation / Bioinformatics / MS / Proteomics / Serum

**Correspondence:** Dr. Akhilesh Pandey, McKusick-Nathans Institute of Genetic Medicine, 733 N. Broadway, BRB 569, Johns Hopkins University, Baltimore, MD 21205, USA

**E-mail:** [pandey@jhmi.edu](mailto:pandey@jhmi.edu)

**Fax:** +1-410-502-7544

**Abbreviations:** GO, gene ontology; HPPP, human plasma proteome project; HPRD, human protein reference database; PPP, Plasma Protein Project

## 1 Introduction

Proteomic technologies and applications have the potential to facilitate the development of novel diagnostic tools for clinical medicine. The proteome, representing the functional translation of the genome, affords great opportunities for the discovery of novel biomarkers/biosignatures for human diseases. More closely linked than the genome to physiologic and pathologic human conditions, proteins offer an opportunity to serve as functional biomarkers for human diseases. Despite considerable progress in this regard [1], many challenges remain, including proteomic-based diagnostic tools for routine clinical applications. One impediment to this process has been the absence of informative proteomic map(s), similar to the blueprint of the human genome. Among all possible proteomes, the Human Plasma Proteome is of particular interest because of its broad interface with other organs in the body. Many of the plasma protein constituents are considered as attractive candidates for biomarkers as they can reflect the diseased or healthy states of various tissues and organ systems. Despite the obvious challenges—as plasma contains a wide dynamic range of component concentrations (*i.e.*, nine to ten orders of magnitude) [2]—several recent investigations have discussed the potential of plasma proteome in biomarker research [3–7].

The approaches used for protein identification by HUPO's Plasma Proteome Project (HPPP) laboratories were based on tandem mass spectrometry, although the exact instruments and pre-analytical separation methods varied across laboratories. A critical step subsequent to mass spectrometry-based protein identification is careful data mining and annotation, which place the identified proteins in the context of what is known about the human biology. To address this challenge, the initial list of proteins identified by HPPP was annotated for their relevance to physiological and pathological states. This included annotation by researchers studying the coagulation pathway, the mononuclear phagocytic system, the liver, the cardiovascular system, the glycosylated proteins and DNA binding proteins. In addition, a bioinformatic analysis was conducted to discover signal peptide cleavage sites by aligning semi-tryptic peptides to the amino termini of proteins. Finally, an analysis of the interactome of plasma proteins with reference to the sites of protein localization was also performed.

## 2 Materials and methods

Eighteen participating laboratories collected data on the plasma proteome using tandem MS/MS. These findings constituted the basis of the initial dataset. A core dataset of 3020 proteins were selected from the total based upon a pre-determined criterion of two or more identified peptides [9]. These were unambiguously mapped (as described ref. [10]) to 2446 distinct gene products and used in the following analyses.

### 2.1 Coagulation pathway and protein interaction network analysis

A list of proteins involved in the coagulation pathway was prepared from the catalog of plasma proteins identified by HPPP based on scientific evidence documented in the literature as well as review by experts in coagulation. The protein interaction network was analyzed by using the 2446 proteins as 'seed' proteins to extract their interacting proteins from the Human Protein Reference Database (HPRD) [11]. The subcellular component of both seed and interactor proteins was also obtained from HPRD and an interactome map was drawn using the Osprey tool [12]. Each protein was represented as a node and the nodes were arranged according to their subcellular localizations in the protein interaction map.

### 2.2 Gene ontology annotations

The proteins identified by the HPPP were assigned gene ontology (GO) terms, which rely on a controlled vocabulary for describing a protein in terms of its molecular function, biological process, or subcellular localization [13].

### 2.3 Analysis of MS-derived data for identification of proteolytic events and post-translational modifications

A bioinformatic analysis of the data generated using a quadrupole time-of-flight mass spectrometer by one of the participating laboratories (Waniger and Raftery) was carried out. We focused on those peptides that were semi-tryptic in nature (*i.e.*, cleaved by trypsin only at one end) and could, therefore, arise from *in vivo* cleavage events. All the identified semi-tryptic peptides were mapped to the corresponding proteins and overlaid with the protein domain architecture information including signal peptides and transmembrane domains. In addition, we carried out a database search using N-acetylation and hydroxylation as variable modifications to identify peptides containing these post-translational modifications. RefSeq database was searched and only peptides with a MASCOT score greater than 30 were considered for further analysis. All MS/MS spectra were also manually interpreted to validate the assignment.

## 3 Results and discussion

The primary criterion that we used for assembling the initial list for annotation in this study was that the proteins were identified on the basis of two or more peptides from tandem mass spectrometry experiments. All of these identifications were deemed 'high confidence' by the individual HUPO laboratories (for details, see ref [8]). Peptide mass fingerprinting data were not included in these analyses. In all, 3020 IPI protein entries were identified based on these criteria. These IPI proteins were unambiguously mapped as 2446

gene products. The major reasons we were unable to map the remaining identifications was because the IPI entries were deleted (203 proteins), redundant (81 proteins), corresponded to pseudogenes (32 proteins) or immunoglobulin fragments (9 proteins), did not have a complete open reading frame (33 proteins) or were not identical to any other protein entry in the non-redundant database (216 proteins). All of these proteins, in addition to proteins that have been described by previous studies to exist in plasma, have been annotated in the Plasma Proteome Database (<http://www.plasmaproteomedatabase.org>) (see ref [10]).

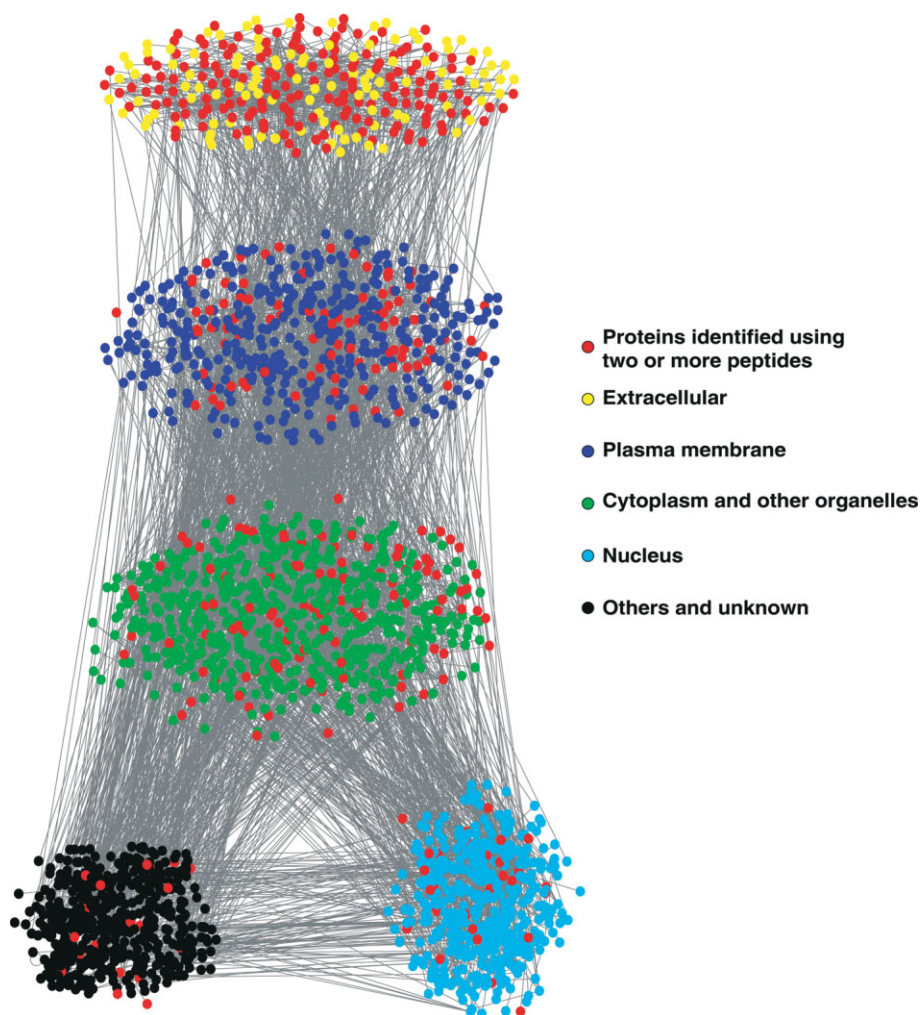
### 3.1 Bioinformatic analyses of the functional subproteomes

Projects of the magnitude such as the HPPP are likely to provide interesting and biologically significant information when protein identification is coupled to bioinformatics analyses. To this end, we carried out an analysis on protein-protein interaction “network” of the HPPP identified proteins. We assigned probable functions to the proteins

according to the Gene Ontology convention. Our goal for these analyses was to use the static protein identification data to gain functional insights pertaining to the role(s) of these proteins in biological processes.

#### 3.1.1 An interaction map of human plasma proteins

The access to this large plasma protein data base provided an opportunity to explore the global and systemic properties of the underlying molecular networks/pathways of the plasma proteome. The afore-mentioned 2446 gene products were further analyzed with respect to their subcellular localization and potential interacting partners. These ‘seed’ proteins were used to extract all documented interactors from the Human Protein Reference Database [11]. The subcellular components of both seed and interacting proteins were determined and subsequently, an interactome map was created utilizing the Osprey tool [12]. Most of the identified proteins possess multiple interacting partners. The average number of interactions per protein were 5.8. In the plasma interaction map (Figure 1), each protein is



**Figure 1.** Protein-interaction map of the plasma proteome. The figure shows the organization of protein interactions in human plasma. Each node represents a protein; the colors depict their localization in the cell. Red nodes represent 652 out of 2446 unique gene products that were identified by HUPPO PPP that had interacting proteins listed in the HPRD. Yellow, blue, green, magenta and grey colors represent proteins that are localized to extracellular, plasma membrane, cytoplasm and other organelles, nucleus and others/unknown subcellular compartments, respectively.

represented as a node, and the nodes are arranged according to their subcellular localization. The lines connecting the nodes represent protein interactions.

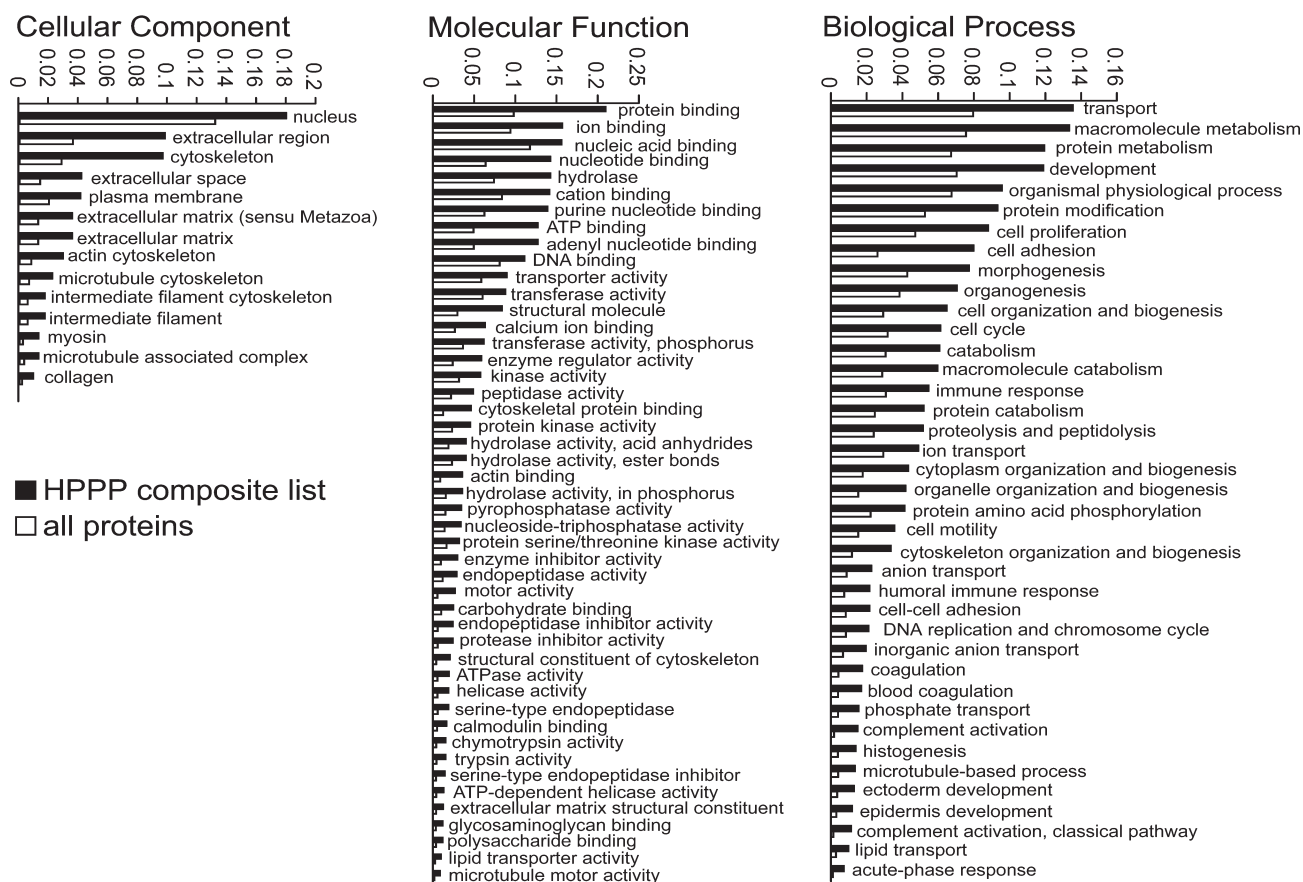
Among the 2446 proteins studied, 652 proteins were determined to be involved in potentially 3811 interactions based on published literature. The proteins were found to be distributed evenly in the extracellular compartment, plasma membrane, cytoplasm and nucleus. Extracellular proteins, in addition to interactions among themselves and to members on the plasma membrane, were found to have an equally high number of interactors with cytoplasmic and nuclear proteins. There are several possible biological explanations for this result. First, it suggests that many extracellular proteins may exhibit heretofore unrecognized intracellular localization. Second, it is conceivable that these interactions may occur when a large number of intracellular molecules interact with plasma membrane-bound proteins. Finally, our protein network analysis may illustrate interactions that are dependent on physiological/pathological or experimentally-induced release of intracellular proteins into the plasma.

It is also noteworthy that this investigation represents one of the first studies to consider the plasma as containing a network of interacting proteins. Many recent investigations in other tissue types have emphasized the role of multiprotein complexes as building blocks of hierarchical cellular machinery, and a goal of these bioinformatic analyses was to provide an initial framework to conceptualize proteins in the plasma in a similar manner.

### 3.1.2 Gene Ontology annotation of protein function

We linked the 2,446 proteins to their Gene Ontology (GO) annotations, a procedure relying upon a controlled vocabulary for describing proteins with respect to their molecular functions, biological processes, or subcellular localization [13]. About 100 GO annotation terms showed significant enrichment in the HPPP list ( $p < 0.001$ , Bonferroni-adjusted), as compared to the entire list of proteins in the LocusLink database (Figure 2). Many of these enriched GO terms

## Gene Ontology term enrichment (fraction of proteins)



**Figure 2.** Significantly enriched GO terms for the set of 2446 HUPO PPP proteins. GO term annotations were obtained from LocusLink (<http://www.ncbi.nlm.nih.gov/LocusLink/>), 1943 of which were in the HUPO PPP composite list. GO terms that applied either to less than 20 genes or to more than 2800 genes in LocusLink were not considered. All GO terms shown are significant with  $p < 0.001$  (after correcting for multiple term testing by Bonferroni, using 1000 random gene lists of 1943).



represent protein classes commonly associated with plasma, such as *extracellular region* (191 proteins found), *cell adhesion* (155), *protease inhibitor activity* (47), *blood coagulation* (33), and *complement activation* (29). Other enriched terms are somewhat unexpected, such as *DNA binding* (216 proteins found) and *nucleus* (350), which observations may in fact be the result of the secretion of cellular breakdown products into the circulation. For each tissue represented in the Novartis atlas [14], we obtained an ordered list of genes that were ranked on the basis of higher mRNA expression in the given tissue as compared to that in other tissues (by rank sum). When the HUPO PPP proteins are ranked by the number of peptides used to identify them across all laboratories, we observe a very high correspondence (approaching  $p < 10E-50$ ) between genes expressed in the liver over other tissues and HUPO PPP proteins that had the most peptide hits. This correspondence is measured by taking the top genes from each ranked list, identifying the genes in common between these lists, and determining whether this overlap significantly exceeds that expected due to random occurrence (Supplementary Fig. 1). This observation probably reflects the fact that the liver is a major contributor to proteins in plasma. Such a correspondence was not found for any other tissues represented in the atlas. It is worthy to emphasize that this does not imply that tissues other than liver do not contribute to the plasma proteome, rather, that proteins derived from these other tissues may not be the most abundant in plasma (or the most readily detectable by proteomics). This trend also does not exclude the contribution of proteins to the plasma from other tissues, as has been described elsewhere in this article with regard to organ-specific protein components of plasma (such as heart and smooth muscle-derived proteins). Another important implication from these bioinformatics analyses is that the number of times a protein is identified across laboratories might correlate with its relative abundance in the plasma. Likewise, the results of this analysis suggest that many of the proteins identified by only a single laboratory may indeed be present in plasma, perhaps in low abundance, and making their detection a rare event. This important information could not have been obtained without a combined effort of multiple laboratories to analyze a standardized sample. Future studies will provide absolute quantitation of these proteins and can be directed to unequivocally reveal the tissue(s) of origin on a protein-by-protein basis.

### 3.2 Proteins involved in the blood coagulation pathway

Coagulation of blood is a critical physiological process that prevents excessive bleeding at the site of injury. It is a proteolysis-driven pathway in which pro-coagulant factors and cofactors interact to constitute a clotting cascade, which ultimately leads to fibrin clot formation [15, 16]. Concomitant with this, regulatory mechanisms, such as the thrombomodulin/protein C pathway, are activated to limit the extent of

clot formation and to prevent thrombosis [17]. Coagulation-related proteins identified by the HPPP (Table 1) can be classified into three categories from a functional standpoint:

The first category includes the major coagulation factors and their cofactors. The plasma concentration of the major coagulation proteins has been described previously [16] and thus it was not surprising that all of the most abundant procoagulant proteins were identified in the HPPP studies. Some coagulation proteins with concentration below 1  $\mu\text{g}/\text{mL}$ , such as factor VIII, were also successfully identified in the present study. The second category of proteins includes components that limit the extent of blood coagulation. This category includes protein C, a serine protease, and its cofactor, protein S, both of which possess anticoagulant activity. Serine protease inhibitors (SERPINS) also represent another major regulatory mechanism controlling enzyme activity of activated coagulation factors. Numerous serpins such as antithrombin, heparin cofactor II, and protein C inhibitor were identified. Proteins involved in regulating fibrinolysis – of which most known members were identified – are also included in this category. One example of a protein identified in this category is carboxypeptidase B, also known as thrombin activable fibrinolysis inhibitor (TAFI), which is a 58-kDa enzyme that circulates in plasma as a zymogen [18]. The third category contains other coagulation-related proteins: ADAMTS-13 is a newly identified metalloprotease enzyme, which cleaves von Willebrand factor (vWF) multimers [19]. Impairment of ADAMTS-13 activity is caused either by a hereditary deficiency or by acquired autoantibodies that specifically inhibit ADAMTS-13 function. Both lead to an excess of ultra large vWF and thereby to thrombotic thrombocytopenic purpura [20, 21]. Interestingly, there are several well-known coagulation factors that escaped identification by all HPPP participating laboratories; these include platelet glycoprotein IX, factor VII, tissue factor and thrombomodulin.

### 3.3 Proteins potentially derived from mononuclear phagocytes

The mononuclear phagocytic system consists of bone marrow monoblasts, promonocytes, blood monocytes, and tissue macrophages. Under normal circumstances, monocytes derived from bone marrow circulate in the blood for less than 48 h and enter into various tissues to differentiate into macrophages [22, 23]. Along with polymorphonuclear neutrophils, mononuclear cells (MCs) provide a defense against microbial invasion via chemotaxis, phagocytosis and release of inflammatory cytokines. Additionally, monocyte-derived tissue macrophages such as dendritic cells are primarily responsible for processing antigen and interact with CD4 positive T cells to initiate host adaptive immunity [22]. HPPP proteins in Supplementary Table 1 are potentially of MC origin and the possible biological implications of these findings are discussed below.

**Table 1.** Proteins known to be involved in blood coagulation

	IPI ID	PPD ID	Gene Symbol	Name of the protein
1	IPI00019568	HPRD_01488	F2	Prothrombin
2	IPI00022937	HPRD_01964	F5	Coagulation factor V
3	IPI00017603	HPRD_02384	F8	Coagulation factor VIII
4	IPI00296176	HPRD_02385	F9	Coagulation factor IX
5	IPI00019576	HPRD_01966	F10	Coagulation factor X
6	IPI00008556	HPRD_07524	F11	Coagulation factor XI
7	IPI00019581	HPRD_01992	F12	Coagulation factor XII
8	IPI00297550	HPRD_00604	F13A1	Coagulation factor XIII, A1
9	IPI00025862	HPRD_00404	C4BPB	Complement component 4 binding protein, beta
10	IPI00009920	HPRD_01956	C6	Complement component 6
11	IPI00007240	HPRD_00605	F13B	Factor XIII, B subunit
12	IPI00021885	HPRD_00619	FGA	Fibrinogen, alpha chain
13	IPI00298497	HPRD_00620	FGB	Fibrinogen, beta chain
14	IPI00021891	HPRD_00621	FGG	Fibrinogen, gamma chain
15	IPI00022418	HPRD_00626	FN1	Fibronectin 1
16	IPI00011255	HPRD_01976	GP1BA	Glycoprotein Ib, platelet, alpha polypeptide
17	IPI00027410	HPRD_01431	GP5	Glycoprotein V platelet
18	IPI00292950	HPRD_00795	SERPIND1	Heparin cofactor II
19	IPI00008558	HPRD_01971	KLKB1	Kallikrein B
20	IPI00032328	HPRD_01970	KNG1	Kininogen
21	IPI00032256	HPRD_00072	A2M	Macroglobulin, alpha 2
22	IPI00019580	HPRD_01417	PLG	Plasminogen
23	IPI00296180	HPRD_01883	PLAU	Plasminogen activator urokinase
24	IPI00021817	HPRD_01466	PROC	Protein C
25	IPI00007221	HPRD_03503	SERPINA5	Protein C inhibitor
26	IPI00294004	HPRD_01473	PROS1	Protein S, alpha
27	IPI00027843	HPRD_07182	PROZ	Protein Z
28	IPI00296099	HPRD_01765	THBS1	Thrombospondin I
29	IPI00021834	HPRD_01064	TFPI	Tissue factor pathway inhibitor
30	IPI00023014	HPRD_01906	VWF	von Willebrand factor
31	IPI00032215	HPRD_00120	SERPINA3	Alpha 1 antichymotrypsin
32	IPI00305457	HPRD_02463	SERPINA1	Alpha 1 antitrypsin
33	IPI00032179	HPRD_00122	SERPINC1	Antithrombin III
34	IPI00293057	HPRD_04374	CPB2	Carboxypeptidase B2, plasma

Macrophages produce an array of cytokines at sites of inflammation and contribute to the development of overall host inflammatory responses [23]. Plasma Interleukin 1 $\beta$ , 6, 10, 12 $\alpha$ , and 15 are all potentially contributed by activated MCs. Two chemokines CCL3 and CCL5 were also observed in plasma. CCL3, or macrophage inflammatory protein 1 $\alpha$  (MIP-1 $\alpha$ ), regulates the migration of various effector cells such as monocytes, T cells, neutrophils, eosinophils, basophils, and natural killer cells [24]. CCL5 is a chemoattractant for CD4+ memory T lymphocytes, monocytes, and eosinophils. CCL5 is expressed in many cell types including MCs [25, 26]. MCs constitutively express Colony Stimulating Factor 1 receptor (CSF1R, CD115), and Colony stimulating Factor 2 receptor alpha chain (CSF2RA, CD116) and thus these plasma proteins may have arisen from MCs. Upon activation, both of these receptors mediate monocyte differentiation and prolong cell survival. Inflammatory stimulation also triggers release of a soluble isoform of CSF2RA potentially by metalloprotease action [27, 28].

CD14 is the lipopolysaccharide (LPS)-receptor protein in MCs; interaction between which triggers an inflammatory response in monocytes via the TLR4/MD-2 complex [29, 30]. CD163 is a hemoglobin-haptoglobin scavenger, a monocyte/macrophage-restricted member of the scavenger receptor family. In addition to being expressed on the MC surface, a soluble form of CD163, resulting from proteolytically cleavage of membrane bound CD163, has been reported [31]. Plasma level of CD163 may be used as a marker for inflammatory disorders [32], or as immunophenotypic marker of monocytic lineage during the diagnosis of acute myeloid leukemia [33]. The mannose receptor (MR) is primarily expressed on macrophages. It recognizes a range of carbohydrates present on, or shed from the surface of microorganisms and can mediate endocytosis and phagocytosis of microbial pathogens [34, 35]. Mannose receptor C type 2 (CD69) is a newly described transmembrane glycoprotein with a C-type lectin binding domain that is constitutively expressed on human monocytes [37] and potentially contributes to innate immunity [36].

MCs contain numerous lysosomal thiol proteases and aspartic proteases, supporting their activities in proteolysis of pathogens and antigen processing [38–42]. Several thiol proteases (Cathepsins L, S, Z) and one aspartic protease (Cathepsin D), identified in plasma, were potentially derived from MCs. Urokinase inhibitor, also named plasminogen activator inhibitor 2 (PAI2), is primarily synthesized by monocytes [43–45]. The physiological function of PAI2 in blood coagulation is incompletely understood; however, it is known to inhibit intracellular proteases and has been implicated in coronary heart disease [46].

### 3.4 Proteins involved in inflammation

Systemic or local inflammation is a pathology associated with a number of human disease. Therefore, the inflammatory subproteomes within the plasma has the potential to be relevant for the management of a host of inflammatory human disorders. As seen in Supplementary Table 2, a number of inflammatory mediators were identified by the HUPO PPP, including chemokines, adhesion molecules, as well as other proinflammatory cytokines. It must also be pointed out that many classic protein markers of leukocytes were not identified including CD19, 20, 21, 23, 24 and 25 (markers of B cell lineage), CD 1, 2, 3, 5, 8, and 28 (markers of T cell lineage), CD 15, 16, 33, 35 (markers of granulocytic lineages), CD 61 (marker of platelets) and CD 68 (marker of MCs). This probably reflects the low abundance of many of these molecules coupled with the lack of shedding from cell surface. Nevertheless, some of the proteins that were identified are believed to be present in relative low abundance in normal serum or plasma. For example, vascular cell adhesion molecule (VCAM)-1 has been reported in healthy individuals at a concentration of 200–300 ng/mL [47]. Surprisingly, multiple laboratories identified a far less abundant cytokine, Interleukin-6, which is present at levels as low as 1.0 pg/mL [48]. It is interesting to note that several of the proteins identified are more closely associated with ongoing inflammation, rather than classic homeostatic processes *per se*. Intriguingly, the HPPP also identified multiple viral and bacterial proteins in purportedly normal plasma [8]. Inflammatory mediators may thus reflect ongoing subclinical challenges by antigens or infectious agents. These groundwork analyses of plasma notwithstanding, the “normal” spectrum of inflammatory mediators defined in broader cohorts of human subjects remains the subject of future investigation. In summary, currently available technologies can already achieve sufficient depth so as to identify immune mediators, which has clear implications for biomarker discovery for the diagnosis and prognosis of human inflammatory diseases.

### 3.5 Analyzing the peptide subproteome of human plasma

Endogenous peptides play an important role in many physiological processes in their capacity as messengers, hormones or cytokines. Furthermore, alterations in peptide levels

under pathophysiological conditions implicate this class of molecules as potential drug targets or biomarkers. Alterations in peptide levels under pathophysiological conditions indicate their usefulness as potential drug targets. In addition, peptides can be derived from turnover processes of larger blood- or tissue-borne proteins. A limited number of studies have been published that concentrate on native, circulating peptides, also referred to as the peptidome (low molecular weight proteome) in blood [3, 4, 49], and therefore, we investigated these peptide constituents of plasma in the HPPP.

In one of our labs (Kellman and Schulz-Knappe), Peptidomics-Technologies<sup>®</sup> were developed specifically for peptidome analysis. This method combines a multiplexed peptide display, differential analysis (Differential Peptide Display<sup>®</sup>, DPD) and sequence analysis on the basis of liquid chromatography and mass spectrometry, providing single markers, and marker panels or patterns. Peptidomics Technologies provides a multiplexed display of approximately 2000 endogenous peptides derived from 1 mL human plasma down into the picomolar concentration range. Further details are available in the accompanying paper by Tammen *et al.* [50]. In the near future, population-based approaches that display thousands of circulating peptides in a (semi-)quantitative manner will allow for the search of biomarkers from the blood-peptidome.

### 3.6 Liver related plasma proteins

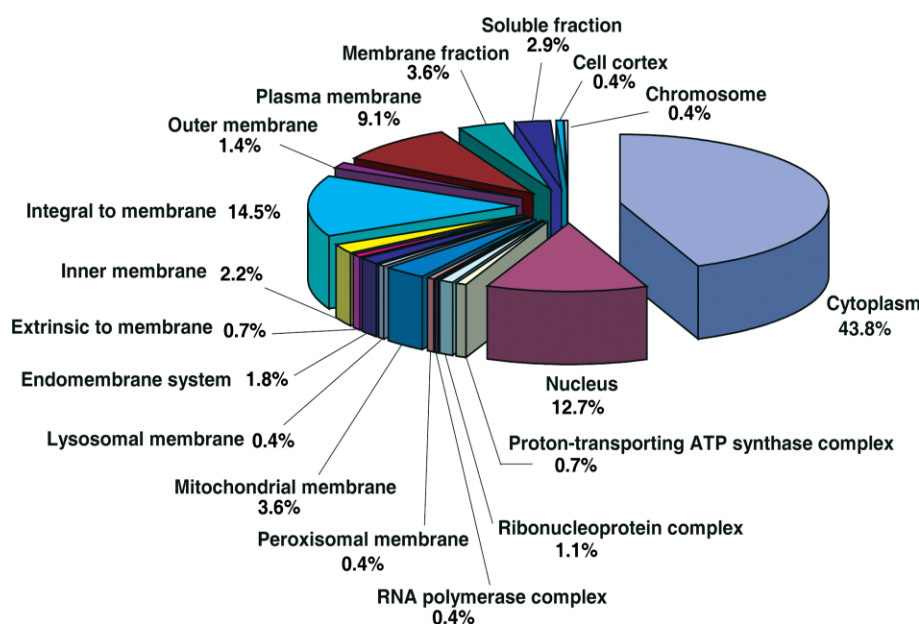
In addition to modification and function-specific annotation of proteins described in the preceding sections, we also annotated the proteins based on whether they could be derived from the liver or the cardiovascular system. The liver is one of the largest, most functionally complex, organs in the body and plays a major role in the metabolism of carbohydrate, amino acids and lipids. Currently, the functional status of the liver is surveyed by quantifying blood components to reveal the extent and type of liver damage. These measurements include chemicals (*e.g.*, bilirubin) and enzyme activities (*e.g.*, aminotransferases) in a serum specimen [51]. Furthermore, the presence and level of certain proteins in the plasma is commonly used for diagnosing hepatic disorders. Levels of albumin, the most abundant protein in plasma, can aid in determining generalized liver disease, since hypoalbuminaemia is a feature of advanced chronic liver disease [52]. Serum alkaline phosphatase activity indicates cholestasis or blockage of bile flow [53]. Elevated activity of aminotransferases in serum indicates acute damage to hepatocytes irrespective of its etiology [54]. Other proteins such as  $\alpha$ 1-antitrypsin and ceruloplasmin are measured in the diagnosis of specific diseases affecting the liver. The existing plasma protein markers for liver function provide clear impetus to examine the plasma as a source of liver and liver disease-associated proteins.

We prepared a partial list of 362 proteins of liver-origin, based on HPPP master list. Annotation was performed on these confirmed lists using the information present in the published literature (using PubMed) and the online Mendelian inheritance in man (OMIM) database. Proteins of hepatic origin were classified according to their subcellular localization (Fig. 3) (for details, see [55]). The proteins are highly localized in the cytoplasmic (121), followed by integral membrane (40), nuclear (35), extracellular space (26) and plasma membrane (25). When these proteins of liver-origin were examined in the context of various liver diseases using NCBI Mesh DB, only a few represent known proteins used to assess liver function (*e.g.*, alpha-1 anti-trypsin and ceruloplasmin), which was not surprising because plasma samples were obtained from 'normal' individuals, *i.e.*, persons without overt liver disease. Alanine or aspartate aminotransferases or alkaline phosphatase enzymes, three key proteins that are elevated in liver disorders, were not identified. However, two proteins, termed insulin-like growth factor-I (IGF-I), a marker of hepatic function [56] and transforming growth factor beta (TGF  $\beta$ ) [57, 58], a marker with known involvement in hepatocellular carcinoma, were identified. Hepatocellular carcinoma is the fifth most common malignant disorder that causes nearly 1 million deaths worldwide. The circulating IGF-I is mainly liver-derived and the development of hepatocellular carcinoma is accompanied by a significant reduction of serum IGF-I levels independent of the grade of impairment of liver function. Importantly, alteration in IGF-I levels precedes the morphologic appearance of the disease, permitting a diagnosis of the tumor [59, 60]. TGF beta levels are elevated in the sera of patients with hepatocellular carcinoma [58]. In conclusion, many proteins potentially of liver-origin, including some disease markers, were found in normal human plasma.

### 3.7 Cardiovascular system related plasma proteins

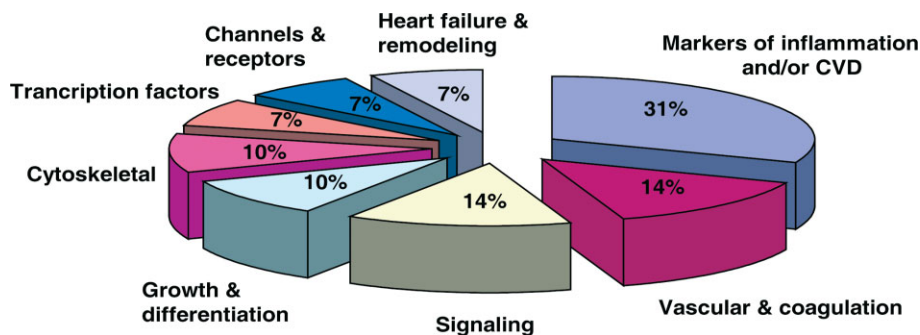
To obtain information regarding proteins of cardiac or vascular origin that were identified in the HPPP analyses, we used PubMed to individually search the proteins for relevance to cardiovascular function using the key words "heart" or "vascular." When searches of protein names returned no results, we searched abbreviations of the protein names (*e.g.*, protein kinase C would also be searched as PKC) and/or non-isoform-specific names (*e.g.*, PKC $\beta$  would also be searched as PKC). In-house expertise with cardiovascular medicine and physiology was used to examine the database search results for possible importance to CV function and/or disease. Our analyses identified cardiovascular-related functions for at least 345 of these 2446 proteins and suggested that these proteins can be divided into at least eight different categories based on function as shown in Figure 4 (a comprehensive analysis of these cardiovascular-associated proteins can be found in ref [61]). These categories include: markers of inflammation, myocardial ischemia and/or cardiovascular disease, signaling proteins, growth- and differentiation-associated proteins, transcription factors, cytoskeletal proteins, vascular proteins, heart failure- and remodeling-related proteins and channel and receptor proteins.

There are at least two reasons why the organ-based annotative approach performed herein can yield further important information regarding plasma proteins. First, if these proteins in the plasma indeed originate from the heart or other organs, they may herald pathologic (such as cell death-associated) or physiologic (normal interplay between functioning tissues and the plasma) relationships between plasma and organ systems that were previously not appreciated. Second, if the individual proteins are of plasma or hematopoietic cell origin, insights into their function in



**Figure 3.** Distribution of 362 proteins of liver-origin according to the sites of subcellular localization.





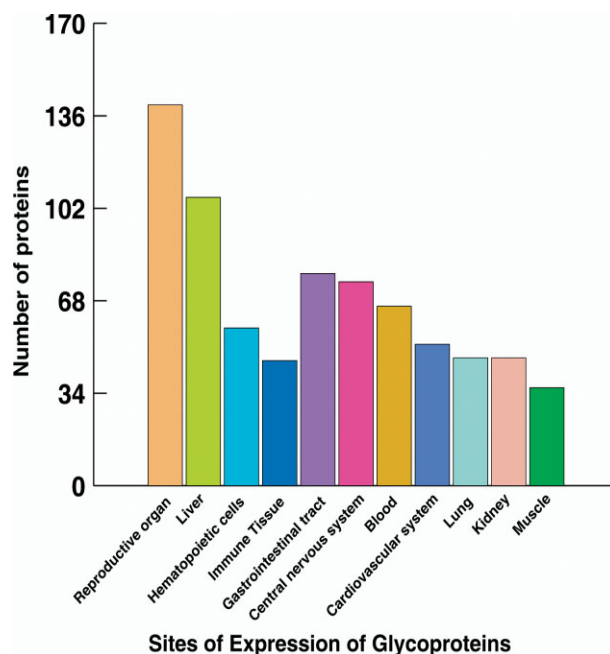
**Figure 4.** Annotative breakdown based on known function of cardiovascular-related proteins identified in plasma. The initial subset of plasma proteins utilized for this study (3,020) was searched individually with the keywords “heart” and “vascular” on the PubMed search engine. Of this initial group, 345 proteins were ascribed cardiovascular-related function on the basis of previous studies, and have been divided into eight groups: markers of inflammation and/or cardiovascular disease; signaling; growth- and differentiation-associated; transcription factors; cytoskeletal; vascular; heart failure and remodeling; and channels and receptors. Percentages of total cardiovascular-related proteins represented by each group are labeled.

other tissues will advance the understanding of their potential actions in the plasma. The potential roles of cardiac- or vascular-derived signaling molecules in the plasma to influence cardiovascular function, and the understanding of whether their release into the plasma is a controlled or catastrophic process, will be the pursuit of future basic science investigations.

In addition, the list of cardiovascular-related proteins itself constitutes a blueprint from which physiologically relevant information regarding cardiovascular diseases can be further mined. It is reasonable to hypothesize that changes in the levels and/or presence of these cardiovascular-related proteins may serve as a diagnostic tool, or readout, of the underlying function of the heart.

### 3.8 Glycoproteins

Various post-translational modifications, including glycosylation, can lead to altered protein function and activity. Selective detection of such modified proteins is a useful tool for modification-based subproteomes mining. A total of 170 proteins were identified using multi-lectin affinity chromatography out of which 84 (50%) were previously shown to be glycosylated. The exact glycosylation site was known in a majority of cases (Supplementary Table 3). Further, almost 66 of 170 (40%) of the proteins identified had previously been shown to be present in serum, plasma or whole blood. The details of the experimental protocol are available in reference [62]. We also carried out a functional annotation of the glycoproteins. Figure 5 and Supplementary Table 3 show the variety of tissues where the proteins identified by multi-lectin affinity chromatography are expressed. The existence of glycosylation-based subproteomes in plasma demonstrate the utility of this approach and suggest that it could be used to identify modifications of potential physiological relevance.



**Figure 5.** Sites of expression of proteins enriched by multi-lectin affinity chromatography. The histogram shows the sites of expression of 170 proteins that were identified using multi-lectin affinity chromatography. The sites of expression are from the published literature based on mRNA and/or protein expression as annotated in the Human Protein Reference Database.

### 3.9 DNA-binding proteins

The next component of our annotative effort included the characterization of DNA-binding proteins, a functional subproteome within the plasma. DNA-binding proteins were classified as such based on analysis of the published literature and a bioinformatics analysis of the protein domains.

DNA-binding proteins bind to double-stranded as well as ssDNA. Because DNA is normally localized to the nucleus, DNA-binding proteins are expected to localize to the nucleus or to shuttle between the cytoplasm and nucleus. Thus, it was of particular interest that normal human plasma proteins were found to include a number of proteins, which have previously been attributed DNA-binding functions. Several independent classes of DNA-binding proteins were seen on the HPPP list, including histones, helicases, zinc finger proteins and proteins involved in cell division (Supplementary Table 4). In addition to the HPPP identifications discussed herein, previous studies [63] have also reported the presence of members of these protein families in human serum. Another possible explanation for the presence of nuclear proteins in plasma or serum could be due to the contamination by broken/dead cells caused by the isolation procedure.

### 3.9.1 Histones

Histones are small basic proteins that form the structural core of chromatin nucleosomes, the macromolecular conformation of DNA. Each nucleosome is composed of two H2A and H2B subunits and two H3 and H4 subunits, however, many variants and histone related proteins have been identified in the genome. Histone H1 and related proteins bind to DNA in the linker region between nucleosomes. Histone proteins, which were identified in HPPP samples, include H2AFJ, HIST1H4A, and HIST1H2AL. H2AFJ and HIST1H2AL belong to the H2A histone family. The H2A subunit performs essential roles in maintaining structural integrity of the nucleosome, chromatin condensation, and binding of specific chromatin-associated proteins. There are previous reports which support the presence of histones in the serum. Waga *et al.* [64] identified and isolated soluble histones from bovine milk and serum. Using Western blot analysis, they detected H4 and partially degraded H2A in bovine serum. Holdenrieder *et al.* [65] reported circulating nucleosomes in serum. During cell death, particularly apoptosis, endonucleases are activated and cleave chromatin into multiple oligo and mononucleosomes. Subsequently, these nucleosomes are packed into apoptotic bodies and engulfed by macrophages or neighboring cells. Under conditions of high cellular turnover and cell death, nucleosomes can also be released into the circulation and can be detected in serum or plasma. As enhanced cell death occurs under a number of pathologic conditions, the presence of elevated amounts of circulating nucleosomes is not specific for any specific benign or malignant disorder.

Serum amyloid P component (SAP) is a highly conserved plasma protein named for its universal presence in amyloid deposits. SAP is the single normal circulating protein that shows specific calcium-dependent binding to DNA and chromatin under physiological conditions. Binding of SAP displaces H1-type histones and thereby solubilizes native chro-

matin, which is otherwise insoluble at the physiological ionic strength of extracellular fluids. Also, SAP binds *in vivo* both to apoptotic cells, the surface blebs of which bear chromatin fragments, and to nuclear debris released by necrosis. SAP may therefore participate in handling of chromatin exposed by cell death [66]. The role played by nucleosomes in the pathogenesis of diseases is not well studied. In autoimmune diseases such as SLE (Systemic Lupus Erythematosus), circulating nucleosomes might be one of the elements that initiate and maintain the activated state of the immune system [67]. In malignant diseases, it has been proposed that the presence of large amounts of circulating nucleosomes might be a part of the tumor counterattack that overwhelms the immune system [68].

### 3.9.2 Helicases

Helicases are proteins that promote unwinding of duplex DNA during replication by binding cooperatively to single-stranded regions of DNA or to short regions of duplex DNA that are undergoing transient opening. Helicases, identified in HPPP samples include HELLS (helicase, lymphoid specific or Lsh), RECQL5, CHD1, 5 and 7 (chromodomain helicase DNA binding proteins). Lsh is thought to be involved with cellular proliferation and may play a role in leukemogenesis. A study by Yan *et al.* demonstrated that Lsh is an exclusively nuclear protein, and they defined a nuclear localization domain within the N-terminal portion of Lsh [69]. Lsh strongly associates with chromatin and requires the internal and C-terminal regions of the protein for this interaction. Interestingly, many of the HPPP participating labs have confirmed the presence of Lsh in plasma. Thus, tissue leakage or release of nuclear debris into the extracellular fluid could account for its presence in the plasma. The CHD family of proteins is characterized by the presence of chromo (chromatin organization modifier) domains and SNF2-related helicase/ATPase domains.

### 3.9.3 Zinc finger proteins

Zinc fingers include multiple distinct structures with the commonality that they require zinc for correct formation of their tertiary and/or quaternary structure. These structural motifs are involved in a broad range of biological activities including double and single stranded DNA binding, RNA recognition, and coordination of protein-protein interactions [70]. The zinc finger proteins that were identified in HPPP samples include RAG 1, MGC26914, FLJ30791, ZNF21, ZNF291, ZNF22 and KAISO-L1. A reasonable explanation for the presence of high number of zinc finger proteins in the plasma could be that some of these domains are involved in the formation of multiprotein complexes and thereby in regions of the cells exposed to exocytosis. Likewise, these proteins may have been released into the plasma by protein transport or cell death.

A number of proteins either directly shown to bind DNA or containing DNA domains thought to confer DNA binding capacity are present in normal human blood [7, 71]. A non-redundant list of 1175 gene products was developed by combination of four separate sources [7]. The methodologies used in this study were 1) literature search for proteins reported to occur in plasma or serum; 2) multidimensional chromatography of proteins followed by two-dimensional electrophoresis and mass spectrometric (MS) identification of resolved proteins; 3) tryptic digestion and multidimensional chromatography of peptides followed by MS identification; and 4) tryptic digestion and multidimensional chromatography of peptides from low-molecular-mass plasma components followed by MS identification. At least 10 transcription factors were observed in the experimental sets, and none of these were found in the literature accession set. In some cases, such as SAP, these proteins may have a physiologic role in binding and solubilizing DNA and chromatin that leak into the plasma. In other cases, these proteins most likely represent breakdown products released in the process of cell death and disintegration. Concrete determination of the roles of these individual DNA-binding-like molecules in the plasma to bind to nucleotides or other functions will require additional studies.

### 3.10 Annotation through reanalysis of mass spectrometry data

Although the primary endpoint in the HPPP was to identify as many proteins as possible in normal human plasma, an interesting offshoot is that the raw data from such mass spectrometric experiments are also a potential goldmine of information regarding post-translational modifications. To cultivate this resource, we carried out a bioinformatic analysis of the data generated using a quadrupole time-of-flight mass spectrometer by the one of the laboratories participating in HPPP (Rafferty and Waniger) to specifically investigate signal peptides, post-translationally modified (e.g. acetylated) proteins or proteins that potentially underwent regulated intramembranous proteolysis.

#### 3.10.1 Cleavage of signal peptides and transmembrane domains

The plasma proteome is rich in proteins that have undergone at least one major proteolytic cleavage event—the removal of the signal peptide. These signal peptides are found at the N termini of proteins that are either extracellular or generally bound to the plasma membrane. The exact location of these signal peptides is usually assigned on the basis of prediction by algorithms that recognize general features of signal peptides including the presence of a hydrophobic stretch flanked on one end by basic amino acids. Although this is currently the most popular method of annotating the location of a signal peptide, such assignments can be inaccurate. Precise identification of signal peptides thus requires experimental determination of the location of the cleavage event that results in the formation of mature proteins. After this cleavage has occurred, the mature proteins lack a signal peptide. In many instances, the mature proteins undergo further cleavage and the resultant cleavage products have different biological activities. If these proteins have to be produced recombinantly (as is the case with cytokines and growth factors administered in humans such as erythropoietin, IL-2, insulin etc.), addition or deletion of a few amino acids because of incorrect assignment of signal peptides can have disastrous consequences ranging from immunological reactions to lack of biological activity.

We focused our attention on those peptides that were semitryptic in nature (*i.e.*, cleaved by trypsin only at one end) and could, therefore, arise from *in vivo* cleavage events. All the semi-tryptic peptides identified through this analysis were mapped to the corresponding proteins and overlaid with the protein domain architecture information including signal peptides and transmembrane domains. As shown in Table 2, we were able to validate the correct signal peptide assignments in the case of macroglobulin alpha-2 and fibrinogen, alpha chain as the semi-tryptic peptide that we identified matched the assigned site of signal peptide cleavage. However, our analysis showed the presence of one instance where the identified peptide indicated the site of cleavage of signal peptide but did not coincide with the

**Table 2.** A list of semi-tryptic peptides indicating signal peptide cleavage sites

Accession #	Protein Name	Gene Symbol	Peptide Sequence
1. NP_000005	Macroglobulin, Alpha-2	A2M	<u>MGKNKLLHPSLVLLLLVLLPTDA</u> <b>SVSG</b> <u>KPQYMVLVPSLLHTETTEKGCVL</u>
2. NP_068657	Fibrinogen, alpha chain	FGA	<u>MFSMRIVCLVLSVVGTAWTADS</u> <b>EGEGD</b> <u>FLAEGGGVRGPRVVERHQSACKDS</u>
3. NP_001076	Alpha-1 antichymotrypsin	SERPINA3	<u>MERMLPLLALGLLAAGFCPAVL</u> <b>CHPN</b> <u>SPLDEENLTQENQDRGTHVDLGLA</u>

Gray color signifies predicted signal peptide sequence and underline indicated the semitryptic peptide that was identified.

predicted site. The signal peptide in the case of alpha-1-antichymotrypsin (SERPINA3) was predicted to be cleaved after residue 25 whereas the semi-tryptic peptide that we identified indicated that the cleavage occurred after residue 23. Thus a shorter mature protein was predicted in the databases for alpha-1-antitrypsin because of this incorrect prediction of signal peptide. Similar validation and correction of assignment of signal peptide cleavage sites has recently been shown for proteins identified from hemodialysis fluid [72].

Regulated intramembrane proteolysis (RIP) is a recently discovered phenomenon that refers to the cleavage of the hydrophobic stretches that comprise the transmembrane domain(s) of membrane-bound proteins [73]. The cleaved product is rendered soluble and might possess a biological activity not contained within the protein itself. An example of RIP is SREBP, an ER membrane protein whose cytoplasmic domain is cleaved in response to low sterol concentrations and this soluble protein fragment then translocates to the nucleus and facilitates the transcription of genes involved in lipid synthesis or uptake [73]. We found 2 previously undescribed candidates for RIP – the first protein is a syntaxin in which the cleavage event would lead to release of SynN and t-SNARE domains into the cytoplasm, which are known to mediate binding to v-SNARES to facilitate membrane fusion. Thus, the loss of these domains by a cleavage event could represent a possible mechanism to regulate membrane targeting. The second protein is the programmed cell death 1 ligand 2. A cleavage event in this case would lead to the release of the extracellular immunoglobulin domain, which is involved in interactions with other proteins and its release would probably be a point of regulation of the immune function [74]. Table 3 shows the peptides sequences that were identified from these proteins.

**Table 3.** A list of proteins with TM domain and a semi-tryptic peptide corresponding to intramembrane cleavage

Accession #	Protein Name	Gene Symbol	Peptide Sequence
1. NP_003754	Syntaxin 16	STX16	IVLIVLVGVKSR
2. NP_079515	Programmed cell death 1 ligand 2	PDCD1LG2	IATVIALR

**Table 4.** A list of peptides with post-translational modifications (PTMs)

Accession #	Protein Name	Gene Symbol	Type of PTM	Peptide Sequence
1. NP_068657	Fibrinogen, alpha chain	FGA	Hydroxylation	TFPGFFSP*MLGEFVSETSR
2. NP_005521	Isocitrate dehydrogenase 3 alpha Subunit	IDH3A	N-acetylation	Ac-AGPAWISK

\* Denotes hydroxylation and Ac denotes an acetyl group

### 3.10.2 Identification of PTMs

MS/MS spectra obtained during the project by the above-mentioned laboratory were also carefully reanalyzed with the specific intent of identifying post-translational modification(s). We specifically searched the RefSeq database for the presence of modified peptides and found one N-terminally acetylated peptide as well as a peptide containing a hydroxyproline residue. Neither of these post-translational modifications had previously been reported regarding the proteins in question (Table 4). A detailed analysis of the entire dataset is likely to reveal many additional examples as illustrated by a recent proteomic analysis of hemodialysis fluid [72]. Supplementary Fig. 2 shows the MS/MS spectra for all semitryptic and post-translationally modified peptides that were identified in this analysis.

## 4 Concluding remarks

Large scale proteomics investigations across numerous laboratories, notwithstanding the advantages of using standardized samples, face challenges for data analyses. The HPPP approach presented in this study capitalized on an international array of expertise from across a spectrum of human physiology and pathophysiology toward the common goal of analysis and annotation of human plasma. It is essential to reiterate that the investigators involved in the annotation process, as presented in this manuscript, were also centrally involved in the plasma analysis/protein identification stage. This integration of analysis and annotation provides for an accurate assessment of physiological context of proteins and peptides by investigators with direct knowledge of the experimental parameters under which the data were obtained.

Through this annotative effort, we have characterized functional subproteomes in the human plasma with potential relevance to cardiovascular and liver disease, DNA binding, coagulation and mononuclear phagocytosis. We have also performed a detailed bioinformatic analysis of the peptides used to identify these proteins, both with regard to the potential independent roles of these peptides in plasma, and as their unique properties can provide insights into the proteins from which they originate. These results suggest



that the proteins that were rarely identified may either be lower abundance plasma components, and/or may arise from non-hematopoietic cell types. Furthermore, these findings suggest that the *groups* of proteins enriched in the lower confidence dataset are classes that are either under-represented in plasma or that have been released into the plasma through cellular breakdown. Resolution of this dichotomy will clearly be one of the most exciting challenges of the future of plasma proteomics. Specifically, the field now faces the challenge to decipher the physiological/pathological roles of these proteins in the plasma while simultaneously working from these annotated datasets to discover and develop novel biomarkers/biosignatures of disease. The model of subproteome analysis employed in this study can be applied to numerous other disease states and in other bodily fluids and tissues. Ultimately, it is hoped that subproteome analysis will couple changes in subsets of proteins to alterations in human phenotype.

*We would like to thank all the laboratories that participated in the HPPP. We thank Dr. David Ginsburg for helpful discussions regarding the coagulation pathway. This work is supported in part by the UCLA Theodore C. Laubisch Endowment to Dr. Ping. Dr. Pandey serves as Chief Scientific Advisor to the Institute of Bioinformatics. The terms of this arrangement are being managed by the Johns Hopkins University in accordance with its conflict of interest policies.*

## 5 References

- [1] Granger, C. B., Van Eyk, J. E., Mockrin, S. C., Anderson, N. L., *Circulation* 2004, **109**, 1697–1703.
- [2] Anderson, N. L., Anderson, N. G., *Mol. Cell. Proteomics* 2002, **1**, 845–867.
- [3] Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J., *et al.*, *Mol. Cell. Proteomics* 2003, **2**, 1096–1103.
- [4] Schulz-Knappe, P., Raida, M., Meyer, M., Quellhorst, E. A., Forssmann, W. G., *Eur. J. Med. Res.* 1996, **1**, 223–236.
- [5] Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S., *et al.*, *Proteomics* 2003, **3**, 1345–1364.
- [6] Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J., *et al.*, *Mol. Cell. Proteomics* 2002, **1**, 947–955.
- [7] Anderson, N. L., Polanski, M., Pieper, R., Gatlin, T., *et al.*, *Mol. Cell. Proteomics* 2004, **3**, 311–326.
- [8] Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., *et al.*, *Proteomics* 2005, **5**. DOI 10.1002/pmic.200500358
- [9] Adamski, M., Blackwell, T., Menon, R., Martens, L., *et al.*, *Proteomics* 2005, **5**. DOI 10.1002/pmic.200500186
- [10] Muthusamy, B., Hanumanthu, G., Suresh, S., Rekha, B., *et al.*, *Proteomics* 2005, **5**. DOI 10.1002/pmic.200400588
- [11] Peri, S., Navarro, J. D., Amanchy, R., Kristiansen, T. Z., *et al.*, *Genome Res.* 2003, **13**, 2363–2371.
- [12] Breitkreutz, B. J., Stark, C., Tyers, M., *Genome Biol.* 2003, **4**, R22.
- [13] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., *et al.*, *Nat. Genet.* 2000, **25**, 25–29.
- [14] Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., *et al.*, *Proc. Natl. Acad. Sci. USA* 2004, **101**, 6062–6067.
- [15] Butenas, S., Mann, K. G., *Biochemistry (Mosc)* 2002, **67**, 3–12.
- [16] Mann, K. G., *Thromb. Haemost.* 1999, **82**, 165–174.
- [17] Bourin, M. C., Lindahl, U., *Biochem. J.* 1993, **289** (Pt. 2), 313–330.
- [18] Sakharov, D. V., Plow, E. F., Rijken, D. C., *J. Biol. Chem.* 1997, **272**, 14477–14482.
- [19] Furlan, M., Robles, R., Galbusera, M., Remuzzi, G., *et al.*, *N. Engl. J. Med.* 1998, **339**, 1578–1584.
- [20] Tsai, H. M., Lian, E. C., *N. Engl. J. Med.* 1998, **339**, 1585–1594.
- [21] Triplett, D. A., *Arch. Pathol. Lab. Med.* 2002, **126**, 1424–1429.
- [22] Furie, B., Bouchard, B. A., Furie, B. C., *Blood* 1999, **93**, 1798–1808.
- [23] Cerundolo, V., Hermans, I. F., Salio, M., *Nat. Immunol.* 2004, **5**, 7–10.
- [24] van Furth, R., *Res. Immunol.* 1998, **149**, 719–720.
- [25] Rollins, B. J., *Blood* 1997, **90**, 909–928.
- [26] Melchjorsen, J., Pedersen, F. S., Mogensen, S. C., Paludan, S. R., *J. Virol.* 2002, **76**, 2780–2788.
- [27] Fessele, S., Boehlk, S., Mojaat, A., Miyamoto, N. G., *et al.*, *Faseb. J.* 2001, **15**, 577–579.
- [28] Flamant, S., Lebastard, M., Pescher, P., Besmond, C., *et al.*, *Microbes. Infect.* 2003, **5**, 1064–1069.
- [29] Prevost, J. M., Pelley, J. L., Zhu, W., D'Egidio, G. E., *et al.*, *J. Immunol.* 2002, **169**, 5679–5688.
- [30] Guha, M., Mackman, N., *Cell. Signal* 2001, **13**, 85–94.
- [31] O'Neill, L. A., *Trends Immunol.* 2002, **23**, 296–300.
- [32] Sulahian, T. H., Pioli, P. A., Wardwell, K., Guyre, P. M., *J. Leukoc. Biol.* 2004, **76**, 271–277.
- [33] Moller, H. J., Aerts, H., Gronbaek, H., Peterslund, N. A., *et al.*, *Scand. J. Clin. Lab Invest. Suppl.* 2002, **237**, 29–33.
- [34] Walter, R. B., Bachli, E. B., Schaer, D. J., Ruegg, R., Schoedon, G., *Blood* 2003, **101**, 3755–3756.
- [35] Apostolopoulos, V., McKenzie, I. F., *Curr. Mol. Med.* 2001, **1**, 469–474.
- [36] Stahl, P. D., Ezekowitz, R. A., *Curr. Opin. Immunol.* 1998, **10**, 50–55.
- [37] Natarajan, K., Sawicki, M. W., Margulies, D. H., Mariuzza, R. A., *Biochemistry* 2000, **39**, 14779–14786.
- [38] Marzio, R., Mauer, J., Betz-Corradin, S., *Immunopharmacol. Immunotoxicol.* 1999, **21**, 565–582.
- [39] Menard, R., Nagler, D. K., Zhang, R., Tam, W., *et al.*, *Adv. Exp. Med. Biol.* 2000, **477**, 317–322.
- [40] Sakai, K., Nii, Y., Ueyama, A., Kishino, Y., *Cell. Mol. Biol.* 1991, **37**, 353–358.
- [41] Chapman, H. A., Jr., Munger, J. S., Shi, G. P., *Am. J. Respir. Crit. Care Med.* 1994, **150**, S155–S159.
- [42] Rossman, M. D., Maida, B. T., Douglas, S. D., *Cell. Immunol.* 1990, **126**, 268–277.
- [43] Riese, R. J., Chapman, H. A., *Curr. Opin. Immunol.* 2000, **12**, 107–113.
- [44] Ritchie, H., Jamieson, A., Booth, N. A., *Thromb. Haemost.* 1997, **77**, 1168–1173.

- [44] Ritchie, H., Booth, N. A., *Thromb. Haemost.* 1998, 79, 813–817.
- [45] Ritchie, H., Booth, N. A., *Exp. Cell Res.* 1998, 242, 439–450.
- [46] Risse, B. C., Chung, N. M., Baker, M. S., Jensen, P. J., *J. Cell. Physiol.* 2000, 182, 281–289.
- [47] Sampson, M., Davies, I., Gavrilovic, J., Sussams, B., *et al.*, *Cardiovasc. Diabetol.* 2004, 3, 7.
- [48] Yun, K. A., Lee, W., Min, W. K., Chun, S., *et al.*, *Scand. J. Clin. Lab Invest.* 2004, 64, 223–228.
- [49] Harper, R. G., Workman, S. R., Schuetzner, S., Timperman, A. T., Sutton, J. N., *Electrophoresis* 2004, 25, 1299–1306.
- [50] Tammen, H., Schulte, I., Hess, R., Menzel, C., *et al.*, *Proteomics* 2005, 5. DOI 10.1002/pmic.200400419
- [51] Rajkumar, S. V., Richardson, R. L., Goellner, J. R., *Mayo Clin. Proc.* 1998, 73, 533–536.
- [52] Hanif, M., Raza, J., Qureshi, H., Issani, Z., *J. Pak. Med. Assoc.* 2004, 54, 119–122.
- [53] Sapey, T., Mendler, M. H., Guyader, D., Morio, O., *et al.*, *J. Clin. Gastroenterol.* 2000, 30, 259–263.
- [54] Goris, R. J., Boekhorst, T. P., Nuytinck, J. K., Gimbere, J. S., *Arch. Surg.* 1985, 120, 1109–1115.
- [55] Cho, S. Y., Lee, E.-Y., Chun, Y.-W., Lee, J. S., *et al.*, *Proteomics* 2005, 5. DOI 10.1002/pmic.200400497
- [56] Vyzantiadis, T., Theodoridou, S., Giouleme, O., Harsoulis, P., *et al.*, *Hepatogastroenterology* 2003, 50, 814–816.
- [57] Murawaki, Y., Ikuta, Y., Nishimura, Y., Koda, M., Kawasaki, H., *J. Gastroenterol. Hepatol.* 1996, 11, 443–450.
- [58] Song, B. C., Chung, Y. H., Kim, J. A., Choi, W. B., *et al.*, *Cancer* 2002, 94, 175–180.
- [59] Mazziotti, G., Sorvillo, F., Morisco, F., Carbone, A., *et al.*, *Cancer* 2002, 95, 2539–2545.
- [60] Stuver, S. O., Kuper, H., Tzonou, A., Lagiou, P., *et al.*, *Int. J. Cancer* 2000, 87, 118–121.
- [61] Berhane, B., Zong, C., Liem, D. A., Huang, A., *et al.*, *Proteomics* 2005, 5. DOI 10.1002/pmic.200401084
- [62] Yang, Z., Hancock, W. S., Richmond-Chew, T., Bonilla, L., *Proteomics* 2005, 5. DOI 10.1002/pmic.200400411
- [63] King, C., Chan, D. A. L., Hise, D. M., Schaefer, C. F., Xiao, Z., Janini, G. M., Buetow, K. H., *et al.*, *Clin. Proteomics* 2004, 1, 101–226.
- [64] Waga, S., Tan, E. M., Rubin, R. L., *Biochem. J.* 1987, 244, 675–682.
- [65] Holdenrieder, S., Stieber, P., Bodenmuller, H., Busch, M., *et al.*, *Ann. NY Acad. Sci.* 2001, 945, 93–102.
- [66] Bickerstaff, M. C., Botto, M., Hutchinson, W. L., Herbert, J., *et al.*, *Nat. Med.* 1999, 5, 694–697.
- [67] Amoura, Z., Chabre, H., Koutouzov, S., Lotton, C., *et al.*, *Arthritis Rheum.* 1994, 37, 1684–1688.
- [68] Igney, F. H., Behrens, C. K., Krammer, P. H., *Eur. J. Immunol.* 2000, 30, 725–731.
- [69] Yan, Q., Cho, E., Lockett, S., Muegge, K., *Mol. Cell. Biol.* 2003, 23, 8416–8428.
- [70] Leon, O., Roth, M., *Biol. Res.* 2000, 33, 21–30.
- [71] Girschick, H. J., Grammer, A. C., Nanki, T., Mayo, M., Lipsky, P. E., *J. Immunol.* 2001, 166, 377–386.
- [72] Molina, H., Bunkenborg, J., Reddy, G. H., Muthusamy, B., *et al.*, *Mol. Cell. Proteomics* 2005, 4, 637–650.
- [73] Brown, M. S., Ye, J., Rawson, R. B., Goldstein, J. L., *Cell* 2000, 100, 391–398.
- [74] Liu, X., Gao, J. X., Wen, J., Yin, L., *et al.*, *J. Exp. Med.* 2003, 197, 1721–1730.