

# SCIENTIFIC REPORTS

**OPEN**

## Comprehensive and quantitative proteomic analyses of zebrafish plasma reveals conserved protein profiles between genders and between zebrafish and human

Received: 01 December 2015

Accepted: 24 March 2016

Published: 13 April 2016

Caixia Li, Xing Fei Tan, Teck Kwang Lim, Qingsong Lin &amp; Zhiyuan Gong

Omic approaches have been increasingly used in the zebrafish model for holistic understanding of molecular events and mechanisms of tissue functions. However, plasma is rarely used for omic profiling because of the technical challenges in collecting sufficient blood. In this study, we employed two mass spectrometric (MS) approaches for a comprehensive characterization of zebrafish plasma proteome, i.e. conventional shotgun liquid chromatography-tandem mass spectrometry (LC-MS/MS) for an overview study and quantitative SWATH (Sequential Window Acquisition of all THEoretical fragmentation spectra) for comparison between genders. 959 proteins were identified in the shotgun profiling with estimated concentrations spanning almost five orders of magnitudes. Other than the presence of a few highly abundant female egg yolk precursor proteins (vitellogenins), the proteomic profiles of male and female plasmas were very similar in both number and abundance and there were basically no other highly gender-biased proteins. The types of plasma proteins based on IPA (Ingenuity Pathway Analysis) classification and tissue sources of production were also very similar. Furthermore, the zebrafish plasma proteome shares significant similarities with human plasma proteome, in particular in top abundant proteins including apolipoproteins and complements. Thus, the current study provided a valuable dataset for future evaluation of plasma proteins in zebrafish.

The Zebrafish (*Danio rerio*) has been an increasingly popular vertebrate model not only in developmental biology but also in toxicological and biomedical research. To investigate fundamental molecular mechanisms, omic tools, especially those for transcriptomic analyses, have been increasingly used for profiling important molecules from embryos to various organs such as the liver, kidney, brain, heart, gut, swimbladder, gonads, etc<sup>1–8</sup>. However, relatively few proteomic studies have been conducted for zebrafish. Early proteomic studies have mainly relied on gel electrophoresis coupled with tandem mass spectrometry analysis (MS/MS) and zebrafish proteomes was reported for embryonic stages<sup>9–11</sup>, liver<sup>12</sup> and lens<sup>13</sup>. The number of proteins identified is quite limited, ranging from less than a hundred to about two hundreds. In the past five years, liquid chromatography (LC) for protein separation before MS/MS has been increasingly used. With this approach, zebrafish proteomes have been reported for multiple organs, including gills<sup>14</sup>, brain<sup>15</sup>, olfactory bulb<sup>16</sup>, fin<sup>17</sup>, skeletal muscle<sup>18</sup>, etc. The number of proteins identified with the new technology varies from a few hundreds to over five thousand. In recent years, there is also an increasing trend to apply proteomics in zebrafish for developmental analyses, disease studies and assessments of chemical toxicity.

Unlike other tissues and organs, zebrafish blood and plasma are rarely used for omic studies. Most studies on zebrafish blood aimed at hematopoiesis, immunity, hemostasis and thrombosis, and hematologic disorders through genetic manipulation, biochemical and imaging analyses<sup>19</sup>. Other blood studies included measurement of hematological parameters and assays for specific components as a monitoring tool for abnormal conditions, e.g. cholesterol and triglycerides in plasma for lipid accumulation in arteries<sup>20</sup> and blood glucose for pancreas

Department of Biological Sciences, National University of Singapore, Singapore 117543. Correspondence and requests for materials should be addressed to Z.G. (email: [dbsgzy@nus.edu.sg](mailto:dbsgzy@nus.edu.sg)) or Q.L. (email: [dbslinqs@nus.edu.sg](mailto:dbslinqs@nus.edu.sg))

function<sup>21</sup>. However, taking into account the growing interest in the zebrafish models and the great potential of plasma biomarker discovery, there is an urgent need to understand the basic plasma protein profile.

Plasma is the most commonly used clinical sample and there have been increasing numbers of studies in pursuing plasma/serum biomarkers for various diseases, for it contains not only blood proteins but also proteins secreted from almost all tissues. Plasma Proteome Database (PPD), an online resource for human plasma proteins, was developed in 2005 as an effort to characterize human plasma proteome. By 2014, it has been updated with over 10,000 reported plasma proteins based on more than 500 studies<sup>22</sup>. The characterization of mouse plasma proteome has also been conducted and Zhang *et al.* have published a repository composed of 568 LC-MS/MS runs, which contains over 3000 mouse plasma proteins with abundance spanning seven orders of magnitude<sup>23</sup>. After immunodepletion of highly abundant proteins prior to LC-MS/MS, Lai *et al.* have detected 4,727 mouse plasma proteins that span the range of concentrations from 10 pg/mL to 2.65 mg/mL<sup>24</sup>. In contrast, studies of zebrafish plasma remain rather limited, partly due to the technical difficulties in blood sampling and the small volume of blood available from each fish. The amount of blood collected from each adult fish varies between 1–10  $\mu$ L<sup>25–28</sup>. Volume of plasma collectable is typically 40–50% of whole blood. With the minute amount of plasma, two research groups have applied LC-MS/MS to profile the proteome: one group has detected about 150 proteins in a pooled sample from two fish<sup>17</sup> and the other has detected about 130 proteins from six individual fish<sup>27</sup>. Neither study has attempted to estimate the concentration of individual proteins identified. These numbers are far below those commonly identified from mammalian plasma, which generally range from several hundreds to over a thousand proteins<sup>29</sup>, with the highest number of 9,087 proteins identified in a single study<sup>30</sup>.

Thus, with the current advances of proteomic technology, it is necessary to have a more comprehensive examination of proteomic profiles of fish plasma and the information should be valuable for future plasma related research in fish, in particular in discovery of non-invasive biomarkers for diseases and chemical toxicity. In this study, we employed two MS approaches to plasma samples collected from both male and female zebrafish: first profiled by conventional shotgun LC-MS/MS for an overview of the proteome and then followed by a quantitatively analytical method called Sequential Window Acquisition of all Theoretical fragment-ion spectra (SWATH)<sup>31</sup> to examine proteins with differential abundance between two genders. We have identified a total of 959 proteins with estimated concentrations ranging from 0.9  $\mu$ g/mL to 8.6 mg/mL. 35 of these proteins are significantly gender-biased in the SWATH profile.

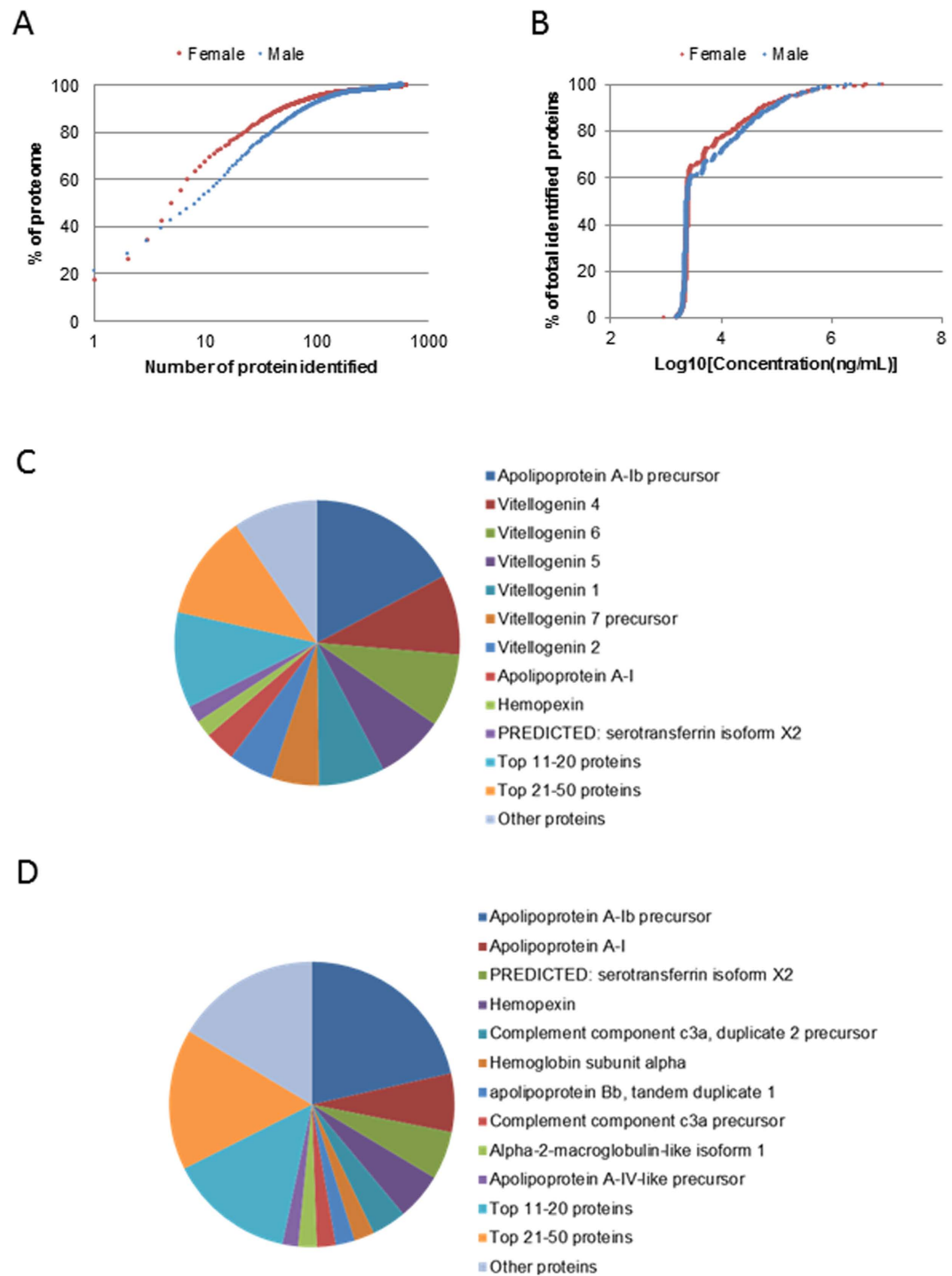
## Results and Discussion

**Abundance and distribution of zebrafish plasma proteins.** Methods of blood collection from zebrafish mainly include tail ablation, heart puncture and dorsal aorta cut. After testing these methods, we chose a modified tail ablation method for its short processing time and reproducibly high yield. Using this method, 5–10  $\mu$ L of blood was routinely collected from male fish and 8–15  $\mu$ L from female fish. Occasionally up to 12  $\mu$ L was collected from a male and 20  $\mu$ L from a female. Plasma was analyzed using 10% SDS-PAGE for quick assessment of protein abundance patterns. Silver staining of individual fish plasma revealed high similarity within the same gender as well as divergence between genders (Supplementary Fig. S1). The protein patterns of plasma collected from 6-month-old fish were similar to those from 1.5-year-old fish. Plasma protein concentration was generally lower in males ( $37.7 \pm 8.4$  mg/mL) than in female ( $53.1 \pm 9.3$  mg/mL), with  $p < 0.011$  in two-tailed heteroscedastic *t*-test. Unlike human plasma where the albumin constitutes at least half of total proteins and forms predominantly large band in gels, the zebrafish does not have albumin<sup>32</sup> and had no overwhelmingly strong protein bands in gel electrophoresis of plasma proteins, except for two bands of about 150 kDa (putatively vitellogenins) in females.

For comprehensive proteomic profiling of male and female zebrafish plasma, fresh plasma of nine individuals of each gender were pooled. The total protein concentrations of plasma were 34.6 mg/mL for the male sample and 49.8 mg/mL for the female sample. An aliquot of 2  $\mu$ g of total plasma protein was used for MS analysis without depletion or enrichment. Six technical replicates of shotgun MS were performed to increase the coverage of proteins. Unique proteins were identified based on various numbers of significant peptides, among which about 40% were identified with two or more significant peptide hits (Supplementary Fig. S2).

A total of 666 protein forms were identified in female plasma and 624 in male plasma, resulting in 959 unique proteins in total, 289 of which were common in both genders. 949 unique proteins (642 in female and 594 in male) had semi-quantitative information (Supplementary Tables S1 and S2). Concentrations of individual proteins were estimated using the total protein concentration, molecular weight of the proteins, and their corresponding emPAI scores in the shotgun profiles. Distribution patterns of molecular weights for female and male plasma proteins were similar (Supplementary Fig. S3A). The number of top abundant proteins contributing to 90% of plasma proteins was smaller in female (49) than in male (80) (Fig. 1A). Proteins detected in female plasma ranged from 0.9  $\mu$ g/mL to 8.6 mg/mL. Similar range of protein concentration distribution (1.6  $\mu$ g/mL to 7.4 mg/mL) was also found in male plasma. For both genders, most proteins were 2–10  $\mu$ g/mL (Fig. 1B). Highly abundant proteins (>100  $\mu$ g/mL) were essentially the top 50 proteins (54 in female and 49 in male). Proteins of intermediate concentrations (10–100  $\mu$ g/mL) included various serine protease inhibitors (Serpins), complement factors and components, coagulation factors, and some metabolic enzymes (e.g. Muscle creatine kinase a, Carboxylesterase 3, Glyceraldehyde-3-phosphate dehydrogenase, Aldolase a, Glutathione peroxidase 4 and beta-enolase). In this study, we also detected a few low-abundant proteins (<2  $\mu$ g/mL), including Eukaryotic translation elongation factor 2b, Retinol-binding protein 3, Semaphorin-5B and Catenin delta-2.

Female plasma contained abundant vitellogenins (Vtgs) which were not detected in the male shotgun profile. Vtgs are egg yolk precursor proteins and all seven known Vtgs in zebrafish<sup>33</sup> were detected in this study. Six of them (Vtg1, 2, 4, 5, 6 and 7) were among the top 10 abundant proteins, while Vtg3, whose gene is located on a different chromosome from other Vtg genes<sup>33</sup>, was ranked 18 (Table 1). Among these Vtgs, Vtg3, 4, 5, 6, and 7 had molecular weight of ~150 kDa and probably together (~31% of total plasma protein mass) gave rise to the most abundant band in the SDS PAGE of female plasma (Supplementary Fig. S1), while Vtg1 (~7, 0.5% of total



**Figure 1. Overview of plasma protein profiles based on shotgun MS.** (A) Cumulative percentage of plasma proteome ranked according to protein abundance. (B) Cumulative plasma protein numbers (shown as % of total identified proteins, 642 in female and 595 in male) ranked based on estimated concentrations of proteins detected. Most proteins (~70%) were in the range of 2–10  $\mu\text{g/mL}$ . (C,D) Abundance distribution of plasma proteins in females (C) and in males (D). Top 10 proteins account for about 65% of plasma proteins in female and about 50% in male.

plasma protein mass), with a molecular size of 211 kDa, probably accounted for the prominent band of ~200 kDa. However, if Vtgs were excluded, both female and male plasma had similarly abundance in apolipoproteins and the top 30 abundant proteins largely overlapped (22/23) (Table 1). In both genders, the most abundant protein was Apolipoprotein A-Ib, accounting for about 20% of total plasma proteins detected. Other common abundant proteins included various apolipoproteins, hemopexin (Hpx), complement component C3, hemoglobin subunits,

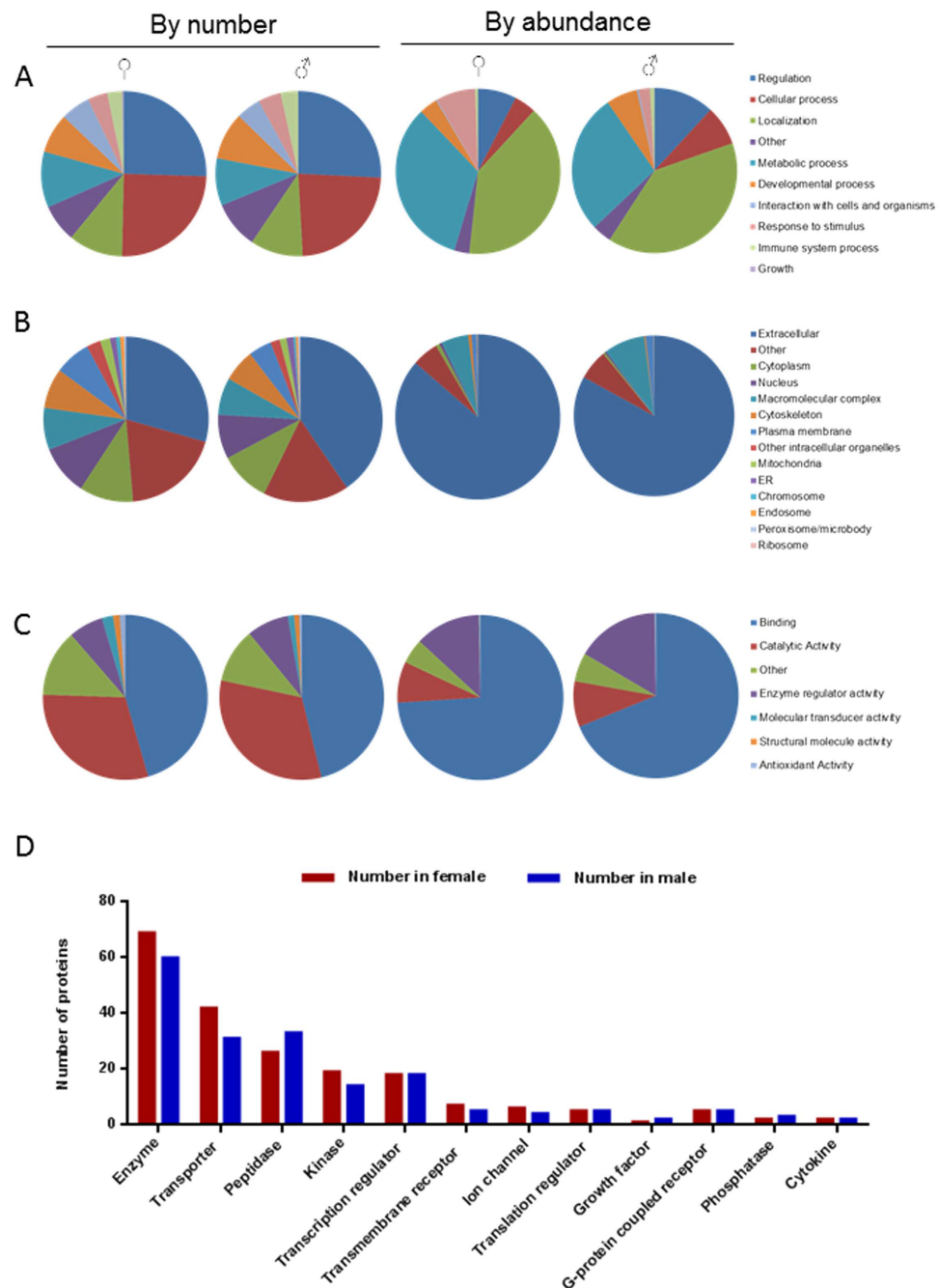
Female			Male		
Protein Name	Gene Symbol	Estimated Conc. (mg/mL)	Protein Name	Gene Symbol	Estimated Conc. (mg/mL)
*Apolipoprotein A-Ib precursor	<i>apoa1b</i>	8.60	*Apolipoprotein A-Ib precursor	<i>apoa1b</i>	7.43
Vitellogenin 4	<i>vtg4</i>	4.51	*Apolipoprotein A-I	<i>apoa1</i>	2.30
Vitellogenin 6	<i>vtg6</i>	4.14	*PREDICTED: serotransferrin isoform X2	<i>tfa</i>	1.89
Vitellogenin 5	<i>vtg5</i>	3.82	*Hemopexin	<i>hpx</i>	1.85
Vitellogenin 1	<i>vtg1</i>	3.73	*Complement component c3a, duplicate 2 precursor	<i>c3a.2</i>	1.36
Vitellogenin 7 precursor	<i>vtg7</i>	2.69	*Hemoglobin subunit alpha	<i>hbaa1</i>	0.80
Vitellogenin 2	<i>vtg2</i>	2.52	*Uncharacterized protein (apolipoprotein Bb, tandem duplicate 1)	<i>apobb.1</i>	0.76
*Apolipoprotein A-I	<i>apoa1</i>	1.74	*Complement component c3a precursor	<i>c3a.1</i>	0.73
*Hemopexin	<i>hpx</i>	0.96	Alpha-2-macroglobulin-like isoform 1	<i>LOC100006895</i>	0.73
*PREDICTED: serotransferrin isoform X2	<i>tfa</i>	0.95	*Apolipoprotein A-IV-like precursor	<i>apoa4b.2</i>	0.61
*Uncharacterized protein (PREDICTED: apolipoprotein Bb, tandem duplicate 1 isoform X1)	<i>apobb.1</i>	0.88	*PREDICTED: apolipoprotein B-100	<i>LOC100330435</i>	0.59
*Apolipoprotein A-IV-like precursor	<i>apoa4b.2</i>	0.79	*Hemoglobin subunit beta-2	<i>ba2</i>	0.57
*Complement component c3a, duplicate 2 precursor	<i>c3a.2</i>	0.58	*Hemoglobin subunit beta-1	<i>ba1</i>	0.57
*PREDICTED: apolipoprotein A-I-like	<i>apoa4b.3</i>	0.57	Ceruloplasmin	<i>cp</i>	0.50
*Apolipoprotein A-IV	<i>apoa4b.1</i>	0.51	*Complement component C3-2	<i>c3a.3</i>	0.50
*Complement component c3a precursor	<i>c3a.1</i>	0.45	*Apolipoprotein A-IV	<i>apoa4b.1</i>	0.49
*Complement component C3-2	<i>c3a.3</i>	0.43	*Fibrinogen, B beta polypeptide	<i>fgb</i>	0.47
Vitellogenin 3	<i>vtg3</i>	0.42	Uncharacterized protein LOC322327 precursor (Apolipoprotein A-II)	<i>apoa2</i>	0.45
*Hemoglobin subunit alpha	<i>hbaa1</i>	0.39	*PREDICTED: apolipoprotein A-I-like	<i>apoa4b.3</i>	0.44
*Fibrinogen, B beta polypeptide	<i>fgb</i>	0.37	*Fibrinogen, alpha polypeptide	<i>fga</i>	0.36
*PREDICTED: apolipoprotein B-100	<i>LOC100330435</i>	0.37	Uncharacterized protein (Fibronectin)	<i>fn1b</i>	0.31
Apolipoprotein Eb	<i>apoeb</i>	0.35	*Uncharacterized protein (PREDICTED: apolipoprotein B-100)	<i>apoba</i>	0.31
*Hemoglobin subunit beta-2	<i>ba2</i>	0.33	*Serpina1 protein like	<i>serpina1l</i>	0.31
*Hemoglobin subunit beta-1	<i>ba1</i>	0.33	*Alpha-2-macroglobulin	<i>sb:cb37</i>	0.29
*Uncharacterized protein (PREDICTED: apolipoprotein B-100)	<i>apoba</i>	0.29	*Serpina1 protein (Fragment)	<i>serpina1</i>	0.28
*Fibrinogen, alpha polypeptide	<i>fga</i>	0.28	Immunoglobulin heavy constant mu	<i>ighm</i>	0.28
*Serpina1 protein like	<i>serpina1l</i>	0.26	*Alpha-2-macroglobulin-like precursor	<i>a2ml</i>	0.28
*Alpha-2-macroglobulin	<i>sb:cb37</i>	0.26	Fibrinogen, gamma polypeptide	<i>fgg</i>	0.24
*Serpina1 protein (Fragment)	<i>serpina1</i>	0.25	PREDICTED: alpha-2-macroglobulin-like	<i>si:dkey-105h12.2</i>	0.23
*Alpha-2-macroglobulin-like precursor	<i>a2ml</i>	0.24	Complement component 5	<i>c5</i>	0.22

**Table 1. Top 30 abundant proteins in female and male zebrafish plasma.** Some entries were labeled as “uncharacterized” in protein database but unambiguously mapped to a known zebrafish gene. The gene name was given in brackets in such cases. \*Proteins that are among top 30 abundant ones in both genders.

and fibrinogen alpha and beta chains. Top 10 abundant proteins constituted 67.6% of plasma proteins in female and 53.3% in male (Fig. 1C,D).

We compared the 959 plasma proteins detected with those reported in the two published studies<sup>17,27</sup> and found that most proteins (820/959) in this study were newly identified in zebrafish plasma (Supplementary Fig. S3B). About 80% of these newly identified proteins were below 7.5 µg/mL and 95% below 50 µg/mL. Based on the concentrations estimated, we detected proteins across almost five orders of magnitude without depletion of abundant proteins or enrichment of low-abundant proteins.

**Characteristics of plasma protein composition.** To obtain an overview of the biological functions of these plasma proteins, Gene ontology (GO) annotations were retrieved using software STRAP. Combining proteins detected in both genders, a total of 795 Uniprot IDs were obtained. GO distributions for Biological Process, Cellular Component and Molecular Function are presented in Fig. 2A–C. Female and male plasma shared very similar distribution patterns. Top terms in Biological Process were Regulation (~25%), Cellular process (~25%), Localization (~10%), Metabolic process (~10%) and Developmental process (~10%). However, in term of protein abundance, Localization (~30% in female and ~40% in male) and Metabolic process (~25% in female and ~35% in male) were the two largest categories. For Cellular Component, the top term was Extracellular (~30% in female and 40% in male). Interestingly, a significant portion of the proteins was found to be in the cytoplasm (~10%), nucleus (~10%), and plasma membrane (~7–8%), while a small portion was from mitochondria, chromosome,

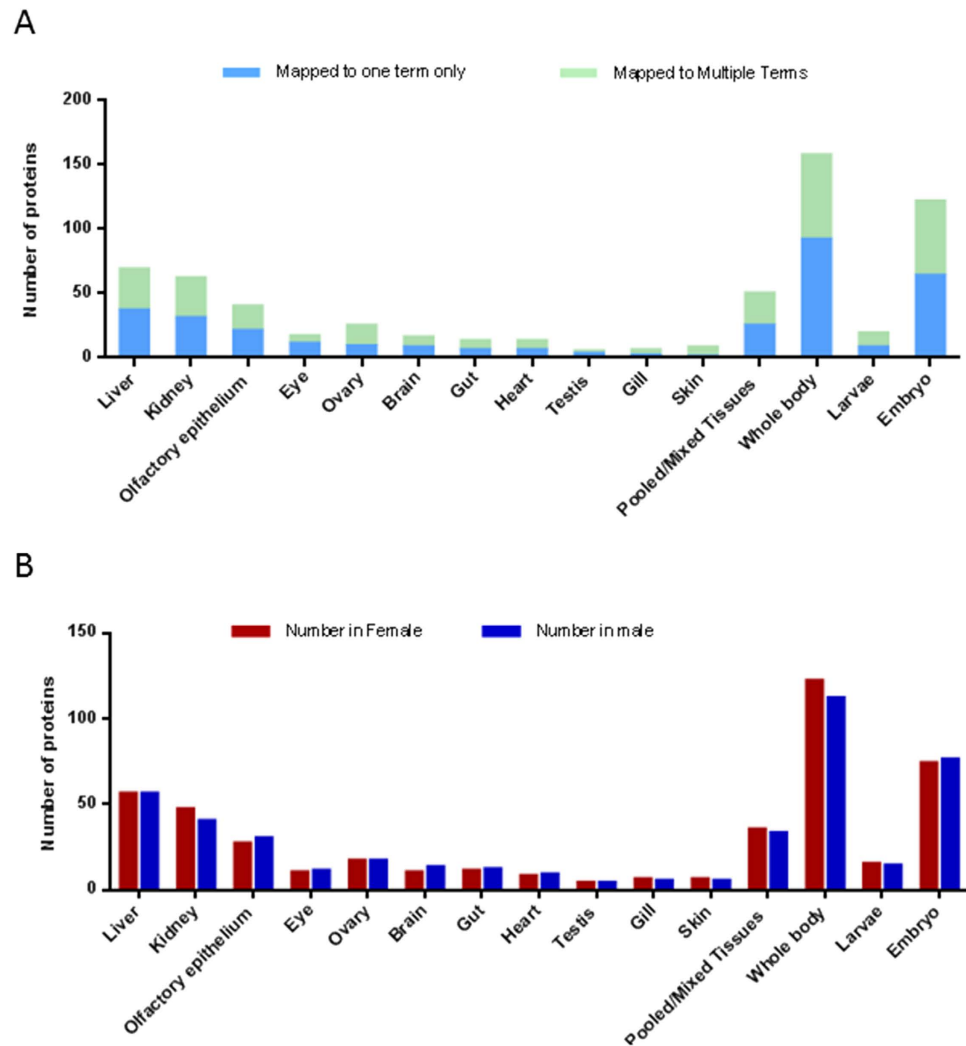


**Figure 2. Characteristics of plasma protein composition.** (A–C) GO distribution of plasma proteins based on number of protein entries (left) and protein abundance (right) in both genders as indicated. GO annotation was retrieved for a total of 791 unique UniProt entries of zebrafish plasma proteins. Categories of GO are: Biological Process (A), Cellular Component (B) and Molecular Function (C). (D) Comparison of types of plasma proteins between the two genders as classified by IPA.

endoplasmic reticulum and endosome (~0.8–2%). These indicate that zebrafish plasma contained many cellular proteins that probably leaked from different tissues. As expected, in term of protein abundance, Extracellular proteins are predominant in both females (~40%) and males (~60%). For Molecular Function, most plasma proteins fell in the categories of Binding (~45%) and Catalytic activity (~30%), with a small percentage of proteins in the category of Enzyme regulator activity (~7% in female and 8% in male), Molecular transducer activity (~2% in female and ~1% in male) and Antioxidant activity (~1% in female and 0.5% in male). If protein abundance is taken into consideration, higher protein contents were attributed to Binding for Molecular function (Fig. 2C).

To further understand the composition of plasma proteins, we mapped the 959 total proteins with corresponding zebrafish Gene IDs in NCBI database and input 910 mapped IDs to IPA to retrieve annotation on types of proteins. Annotation for 518 plasma proteins (57% was retrieved and the composition was similar in both





**Figure 3. Tissue expression of plasma proteins.** (A) Distribution of plasma proteins based on tissues of expression. (B) Comparison of tissue expression of plasma proteins between females and males. Only those proteins expressed in a single tissue are used for comparison.

genders (Fig. 2D and Supplementary Table S2). Based on number of proteins belonging to each defined type, the largest category was Enzyme, followed by Transporters, Peptidase, Kinase and Transcription regulator. However, in term of total weight fraction in plasma, enzymes account for only 1.7% in female and 3.9% in male, while transporters account for 33% in female and 54% in male. These transporters include macroglobins, apolipoproteins, transferrin and hemoglobins. Notably, two growth factors (angiotensinogen and Hepatocyte growth factor-like protein) and two cytokines (Complement component 5 and Interleukin-27 subunit beta) were also detected, which ranged from 2.3  $\mu\text{g}/\text{mL}$  to 222.0  $\mu\text{g}/\text{mL}$ . There were 242 IDs mapped to the term “Others”, yet accounting for only about 5.1% protein content in female plasma and 9.5% in male.

**Tissues expressing plasma proteins.** To determine which organs are major contributors to plasma proteins, the 910 Gene IDs were input to DAVID database for information on tissue expression. Information on tissue expression was retrieved for 452 genes (Supplementary Table S3). Most of the proteins were found to be expressed in whole body or in multiple organs (Fig. 3A). 135 genes had reported expression in only one individual organ, including liver (37), kidney (31), olfactory epithelium (21), eye (11), ovary (9) and brain (8) (blue bars in Fig. 3A). If the abundance of proteins is taken into account, the liver is the single predominant site of production for plasma proteins, with 37 proteins accounting for about 30% of total plasma protein content in both genders (Supplementary Table S3). These 37 proteins included high abundance proteins such as Vtgs and some complement components. There were also other abundant proteins that were expressed in both the liver and other organs, such as apolipoproteins. In fact, according to our in-house zebrafish liver transcriptome generated by RNA sequencing, mRNAs for at least 323 plasma proteins detected in the present study are also expressed in the liver, among which Vtg mRNAs take up as much as 78% of female liver transcriptome<sup>6</sup>. The extremely high expression of Vtgs in the liver leads to the high abundance of total Vtgs (>40%) in the plasma of female. Since the expression level of liver secreted proteins is generally in agreement with corresponding plasma protein concentrations<sup>34</sup>, alteration of liver functions could result in prominent changes in plasma proteome. Unlike those from

liver, proteins from other organs are present at low concentrations (Supplementary Table S3). For example, 31 proteins from kidney only make up about 1–2% of total mass, while 21 proteins from olfactory epithelium only make up 0.1%.

**Comparison of zebrafish and human plasma proteome.** The range of human total plasma protein concentration is 60–80 mg/mL, with concentrations of detected proteins spanning at least 12 orders of magnitudes<sup>35</sup>. As compared to human, the total plasma protein concentration seems to be lower in zebrafish (30–55 mg/mL), with females often having higher total protein content than males.

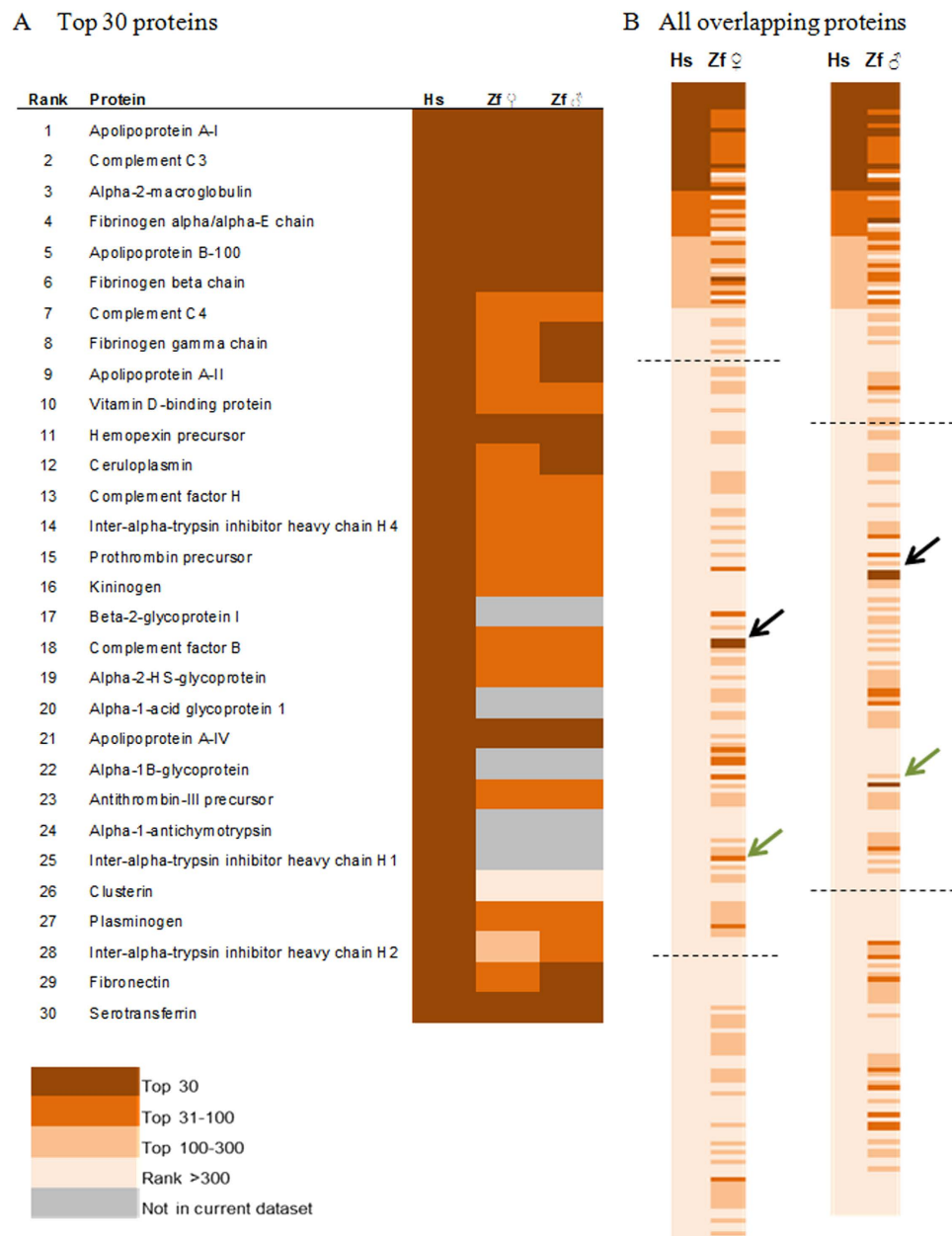
To compare the zebrafish plasma proteins with those reported in human, we searched against NCBI database and mapped 882 out of the 959 zebrafish plasma proteins to 598 unique human homologs. These human homologs were examined against PPD (<http://www.plasmaproteomedatabase.org/>). Most of these proteins (435/598, or 73%) have been detected in human plasma and 140 of them have concentration values available. In comparison with human plasma, zebrafish plasma has at least three prominent differences: the absence of albumin, the presence of vitellogenins, and the presence of multiple homologs of the same human protein. For example, in zebrafish plasma, there are at least three homologs of human Apolipoprotein A and seven homologs of Alpha-2-macroglobulin. In terms of protein composition, human albumin comprises approximately 50% of the total plasma content and the top 22 proteins make up 99% of the total proteins<sup>30</sup>. In zebrafish, the highly abundant proteins take up a smaller percentage. The top most abundant protein, Apolipoprotein A-Ib precursor, comprises approximately 20% of total proteins detected. Furthermore, it takes at least top 50 proteins to account for 90% of the total content and top 400 proteins for 99% content (Supplementary Tables S1 and S2).

To compare the abundance ranking of plasma proteins between zebrafish and human homologs, we mapped the human homologs of detected zebrafish plasma proteins to those reported by Liu *et al.*<sup>36</sup>. Among the top 30 most abundant proteins identified in human after depletion of several highly abundant proteins, we found that at least 13 of them, such as Apolipoprotein A-I, Complement C3, Alpha-2-macroglobulin, alpha and beta Fibronogens, Apolipoprotein B-100, etc., were also among the top 30 abundant proteins in zebrafish (Fig. 4A). Essentially all the top 30 serum proteins in human (except for four proteins that were not identified in zebrafish) were among the top 100 zebrafish serum proteins and these proteins were enriched with apolipoproteins and complement proteins. Thus, there are similar profiles of abundant plasma protein with conserved main functions in fatty acid transport and immunity between human and zebrafish. When the ranks of all overlapping plasma proteins were compared between human and zebrafish (Fig. 4B), we found that there were higher similarities in the high (top 100) and medium ranks (top 500) for both male and female zebrafish but much lower similarities in the low ranks. We speculate that, due to the much smaller coverage in zebrafish with non-depleted plasma (~1,000 proteins) as compared to the large profile (~9,000 proteins) of immune-depleted human plasma, concentration for zebrafish proteins were generally over-estimated and the effect was much more significant for those of lower ranks. It is interesting to note that some top abundant zebrafish proteins correspond to low abundant protein in human, such as Alpha-1-antitrypsin precursor and Ig mu chain C region (Fig. 4B); this is apparently due to immunodepletion of abundant proteins (including antitrypsin and immunoglobins) in human samples for enriching low abundant proteins.

To compare the distribution of GO annotation terms between zebrafish and human plasma, through PPD, we downloaded proteins that have been reported in at least two human plasma studies and obtained a list of 3,638 unique UniProtKB accessions. GO annotation of these proteins was retrieved using STRAP and plotted (Fig. 5A–C). As compared to zebrafish plasma (Fig. 2A–C), for both Biological process and Molecular function, top two terms were the same and the corresponding percentages are also similar. Terms at lower ranks overlapped between the two species with both similarities and differences in percentages. As for Cellular component, more human plasma proteins fell under the terms such as “cytoplasm”, “nucleus” and “plasma membrane”, while higher percentages of zebrafish proteins belong to “extracellular” and “others” (Figs 5B and 2B). This is likely due to two reasons: (1) better GO annotation is available for human than zebrafish; (2) the coverage of human plasma proteins from a composite database is much higher than our single study, such that the human dataset contains more low-abundant proteins secreted or leaked from cellular compartments, which were not detected in this study.

To test whether there is a direct correlation in corresponding plasma proteins, we compared the estimated concentrations of zebrafish proteins with published concentration values of their human homologs and performed Pearson correlation test (Fig. 4D–F). While high correlation was observed between zebrafish genders ( $r = 0.95$ ), correlation across two species were much lower. Yet, decent positive correlation was still observed between human and both zebrafish genders ( $p < 0.0001$ ), with a correlation score of 0.53 for female and 0.52 for male. The detection limit of this study was apparent in the correlation plots, which might have compromised the accuracy of concentration estimation for low-abundant proteins and thus gave rise to an under-estimated correlation score between the two species.

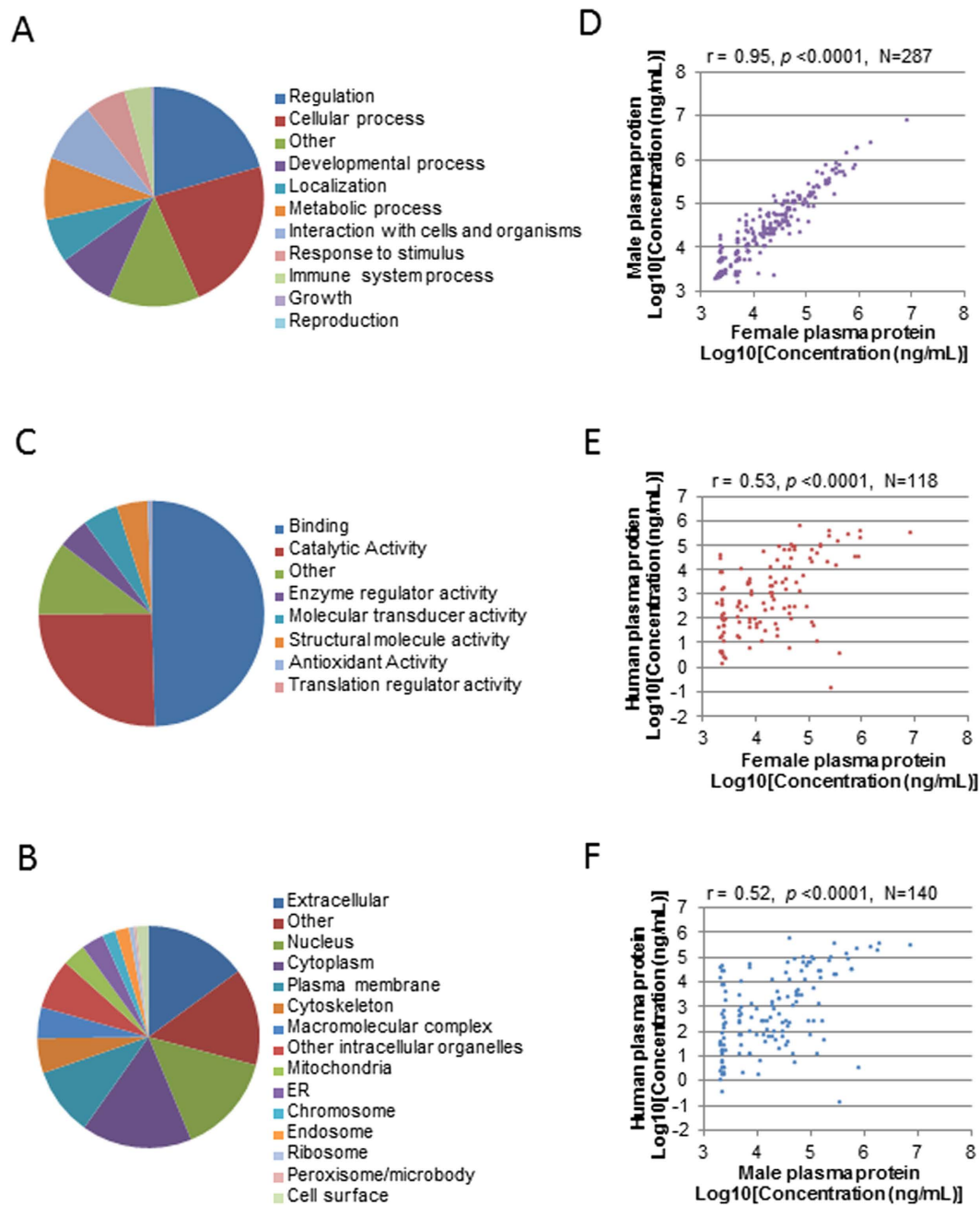
The major organ for synthesis of human plasma proteins is the liver and at least 362 plasma proteins of liver origin have been reported<sup>37</sup>. Liver diseases are often accompanied with manifestations in hematological disorders<sup>38</sup>; thus, routine clinical screening of functional status of liver are generally based on amount and activity of selected enzymes of liver origin in serum samples<sup>37</sup>. Similarly, in the present zebrafish plasma proteome, at least 323 plasma proteins whose transcripts were also found in zebrafish liver<sup>39</sup>. These include proteins whose human homologs are liver disease markers; for example, apolipoprotein E and M, AHSG, and complement C3 are biomarkers for hepatocellular carcinoma (HCC)<sup>40</sup>. In human, plasma biomarkers for disease status of other organs have also been studied and homologs of a few such biomarkers (putative or in use) are also present in zebrafish plasma. For example, biomarkers for ovarian cancer (apolipoprotein A-I and transferrin)<sup>41</sup>, Lung cancer (LDH and haptoglobin)<sup>42,43</sup>, and pancreatic cancer (Plasminogen, Hemopexin, Complement factor H, and alpha-2 macroglobulin)<sup>44</sup>.



**Figure 4. Overlaps in ranks of plasma proteins between human and zebrafish based on the study of Liu *et al.*<sup>36</sup>** (A) Rank of top 30 abundant human plasma proteins in zebrafish. (B) Comparison in ranks of all mapped overlapping plasma proteins between human and zebrafish female (left, 255 proteins) and male (right, 250 proteins). Proteins are presented based on ranks in human plasma. Two dashed lines in (B) mark approximately the rank of 500 and 5000 in human plasma proteins. Hs: human; Zf: zebrafish. Black arrows indicate Alpha-1-antitrypsin precursor (human) and Serpina1 protein (Zebrafish), and green arrows indicate Ig mu chain C region.

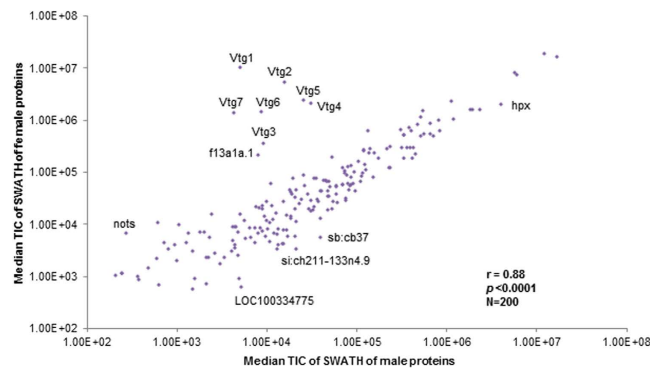
**Gender-biased plasma proteins.** Sexual dimorphism in gene expression pattern has long been observed in human and there is increasing evidence on the differential risks and responses to diseases due to gender disparity<sup>45</sup>. In terms of human plasma proteins, the gender-associated differences between healthy subjects in general are not well studied. Based on a small-scale proteomic study of five men and five women, five female-biased proteins (pregnancy zone Protein, coagulation factor V,  $\alpha_1$ -antitrypsin,  $\beta_2$ -microglobulin, and complement factors H) and five male-biased proteins (Fc binding protein, protein Z-dependent protease inhibitor, phosphatidylinositol-glycan specific phospholipase, protein S-100 and transgelin-2) have been identified<sup>46</sup>. A few more studies focusing on the concentration difference of a specific protein or few related proteins as well as the consequence of such difference under a disease condition have been reported. For example, the sex-differential abundances of candidate proteomic markers in cardiovascular diseases<sup>47</sup> and in non-small cell lung cancer and asthma<sup>48</sup> have been reported.





**Figure 5. Comparison of zebrafish and human plasma proteomes based on Plasma Proteome Database (PPD).** (A–C) GO distribution of human plasma proteins in the categories of Biological Process (A), Cellular Component (B), and Molecular Function (C). Both the ranks of terms and the corresponding percentages are similar to those in zebrafish plasma (Fig. 2A–C). (D) Pearson correlation between zebrafish female and male plasma proteins based on concentration of overlapping proteins. (E,F) Pearson correlation between human plasma proteins and zebrafish female (E) and male (F) plasma proteins, based on corresponding concentrations of homologous proteins.

In this study, SWATH was employed to investigate the difference between male and female zebrafish plasma. SWATH MS is a data-independent acquisition method which combines the advantages of high throughput shotgun proteomics and quantitative MRM (multiple reaction monitoring) approach. Through systematical fragmentation of all precursors in a moving mass window, a complete record of all detectable precursors and their MS/MS ion fragments is generated. This allow for the identification of proteins via matching with a pre-generated spectra ion library and MRM-like quantitation of identified proteins<sup>31</sup>. An ion spectra library was built from the shotgun profiling and triplicate runs were then performed in SWATH mode for quantitative comparison of individual plasma proteins between two genders. There was a good correlation between technical triplicates (Supplementary Fig. S5A,B). A total of 200 proteins were quantified for both male and female plasma samples



**Figure 6. Overview of SWATH profile of female and male plasma proteins depicting abundance of detected proteins.** Median total ion current (TIC) of proteins identified in male (x-axis) and female (y-axis) plasma were compared. Majority of plasma protein have similar abundance in both sexes (Pearson correlation = 0.88), with representative sex-biased plasma proteins as outliers.

with high confidence (Supplementary Table S5), which include most abundant proteins as well as some relatively low-abundant proteins such as Nothepsin (7.8  $\mu\text{g}/\text{mL}$ ) and Vitelline membrane outer layer 1 homolog b (2.4  $\mu\text{g}/\text{mL}$ ). Not surprisingly, variation between replicates was smaller for proteins of high abundance than those of lower abundance.

Most proteins were of similar abundance across genders and only a few proteins showed prominent differences between genders, among which Vtgs were the most obvious ones (Fig. 6). To identify gender-biased proteins, the average total ion current (TIC) values for each protein were compared and 35 differential proteins were obtained based on two criteria:  $p < 0.05$  and  $\text{FC} > 2$  (Table 2). Quantitation of these proteins in shotgun profiles was mostly consistent (Pearson correlation  $r = 0.76$ ), except that some proteins of lower concentrations (mostly  $< 10 \mu\text{g}/\text{mL}$ ) were either not detected or with inconsistent quantitation (Supplementary Table S6). Based on cross-gender difference smaller than 5% and small coefficient of variation (CV) in SWATH profile, we also identified apolipoprotein A1 (Apoa1) as a stable protein across genders and technical replications. It has only 2% difference between genders ( $\text{FC} = 0.98$ ) and a small CV (12.8% in female and 6.6% in male) in SWATH profile. It was among the top 10 proteins in both genders according to the shotgun profile ( $\sim 2 \text{ mg}/\text{L}$ ). In comparison, beta-actin, the popularly used as a loading control in western blot of zebrafish plasma<sup>49,50</sup>, had a moderate concentration of  $\sim 20 \mu\text{g}/\text{L}$  in both genders in shotgun profile. In the SWATH profile, beta-actin showed 8% difference ( $\text{FC} = 1.12$ ) between the two genders and much greater CVs (29.5% in female and 32.4% in male). Thus, in addition to beta-actin, Apoa1 could be a potential reference protein in zebrafish plasma studies for normalization of quantification of other proteins.

Among the 26 female-biased proteins, all seven vitellogenins (Vtg1 to Vtg7) were among the top with high fold changes ranging from 20 (Vtg2) to 1552 (Vtg1) fold. Another ovary-associated protein, Vitelline membrane outer layer 1 homolog a (Vmo1a), was also 6.4 fold more abundant than that in male. Two other proteins, Nothepsin and an uncharacterized homolog of human Coagulation factor XIII (FXIII), showed about 23 fold changes. In addition, apolipoprotein E related proteins (Apolipoprotein Eb and Apolipoprotein E precursor) were also 4.2 and 3.5 fold more abundant in female. Other female-biased proteins included Phosphoglycerate kinase 1 (Pgk1), a predicted keratin 8 homolog, and Muscle creatine kinase a and b (Ckma and Ckmb), and a few other apolipoproteins. The fold changes of the 9 male-biased plasma proteins were much smaller. A Carboxylesterase 2-like protein showed the greatest fold change of 3.6, followed by Ependymin (3.4 fold) and Serpina 1 protein like (3.1 fold). Other male-biased proteins include Sex hormone binding globulin (Shbg), a predicted complement C4-B, Myoglobin and uncharacterized proteins.

As compared to a previous study<sup>27</sup>, only the Vtgs have been earlier reported as gender-biased proteins, while the rest were newly found in this study. As there have not been many studies on the sexual dimorphism of plasma proteins, only few of these 35 proteins have been shown to have gender-difference in plasma or other tissues. In zebrafish, *vtg* genes are highly expressed in the liver of female fish and estrogen-treated male fish<sup>6,33</sup>. In females, Vtgs were transported in blood and taken up by oocytes, whereas in males, they remain in blood until degradation<sup>51</sup>. The induction of Vtgs in the liver and plasma of male fish have been widely used as molecular markers for environmental estrogens<sup>6,51</sup>. Other than the Vtgs<sup>27</sup>, Nothepsin, a paralog of Cathepsin, has been reported to express only in the liver of female zebrafish<sup>52</sup>, though the protein has also been detected in the ovary<sup>53</sup>. The gene for Nothepsin has been reported to be among the top 20 most regulated genes during vitellogenesis in female liver and after E2 treatment of male liver<sup>54</sup>. In human, Cathepsin D is a secreted protein from various types of cancer cells and its serum level has been proposed as a biomarker for breast cancer and a few other cancers<sup>55</sup>. Coagulation factor XIII (FXIII) circulates in human plasma and is involved in the formation of fibrin structure as well as regulation of fibrinolysis; in zebrafish, this female-biased homologous protein, named FXIIIa-42, has not been well studied but it has been reported to be upregulated prior to vertebral calcification of larvae<sup>56</sup>. Gender-biased expression of FXIII in human or zebrafish has not been reported before; however, in a transcriptomic analyses of gene targets of estrogen, *f13a1a* (coding gene for FXIIIa-42) was highly induced by E2 in both embryos and male adults<sup>57</sup>. In addition, in human, FXIII also plays a role in maintaining pregnancy<sup>58</sup>. All this suggests that FXIIIa-42 may also play important roles in females. Another interesting female-biased protein is

Protein	Gene Symbol	Entrez GI	HUMAN Homolog	Fold change
<b>Female-biased</b>				(F/M)
Vitellogenin 1	vtg1	559475		1552.54
Vitellogenin 2 (Q1MTC4)	vtg2	559931		414.70
Vitellogenin 7 precursor	vtg7	559856		218.59
Vitellogenin 6	vtg6	559229		186.10
Vitellogenin 5	vtg5	64260		96.13
Vitellogenin 4	vtg4	678536		73.91
Vitellogenin 3	vtg3	30518		30.95
Uncharacterized protein (Coagulation factor XIII)	f13a1a.1	767742		23.16
Uncharacterized protein (Nothepsin)	nots	114367		22.64
Vitellogenin 2 (A8WGJ1)	vtg2	559931		19.73
Vitelline membrane outer layer 1 homolog a	vmo1a	793369	VMO1	6.38
Apolipoprotein Eb	apoeb	30314	APOE	4.19
Phosphoglycerate kinase 1	pgk1	406696	PGK1	4.16
PREDICTED: saxitoxin and tetrodotoxin-binding protein 2-like	LOC795897	795897		4.02
Branched chain keto acid dehydrogenase E1, beta polypeptide, like	bckdhbl	555648	BCKDHB	3.95
PREDICTED: keratin, type II cytoskeletal 8 isoform X1	krt8	797433	KRT8	3.75
Apolipoprotein E precursor	apoea	553587	APOE	3.46
Ckmb protein (Muscle creatine kinase b)	ckmb	794752	CKM	3.33
Triosephosphate isomerase B	tpi1b	560753		3.02
PREDICTED: apolipoprotein Bb, tandem duplicate 1 isoform X1	apobb.1	321166	APOB	2.65
Nucleoside diphosphate kinase	nme2b.2	30084	NME2	2.48
Muscle creatine kinase a	ckma	30095	CKM	2.37
Retinol binding protein 4, plasma	rbp4	30077	RBP4	2.18
Apolipoprotein C-II	apoc2	568972	APOC2	2.15
Apolipoprotein A-IV	apoa4b.1	322543	APOA4	2.12
Uncharacterized protein (Parvalbumin 1)	pvalb1	402805		2.08
<b>Male-biased</b>				(M/F)
Uncharacterized (Carboxylesterase 2-like)	ces2	566132	CES1	3.59
Ependymin	epd	30199		3.39
Serpina1 protein like	serpina1l	321195	SERPINA1	3.13
Uncharacterized protein ( si:ch1073-126c3.2)	si:ch1073-126c3.2	555816		3.12
Uncharacterized protein LOC791587	zgc:174259	791587	SERPINA9	2.53
Hypothetical protein LOC334459	wu:fi74c02	334459		2.25
Sex hormone binding globulin	shbg	322604	SHBG	2.21
PREDICTED: complement C4-B	c4	566261	C4A/C4B	2.09
Myoglobin	mb	393558	MB	2.05

**Table 2. Gender-biased proteins ( $p < 0.05$ , fold change  $> 2$  or  $< -2$ ) in zebrafish plasma.**

Pgk1, which is an erythrocyte glycolytic enzyme gene located on X chromosome in human; mutation of this gene causes non-spherocytic haemolytic anemia and consequently various neurological disorders<sup>59</sup>.

The top male-biased protein Carboxylesterase 2-like (Ces2), is a homolog of human CES1. The human plasma CES1 has been proposed to be a serological marker candidate for HCC and it has performed better in discriminating HCC from other liver diseases in a recent clinical study<sup>60</sup>. It is interesting to note that Ces2-like protein is male-biased in zebrafish plasma, a fact consistent with the high HCC incidence in men<sup>61</sup>. Another noteworthy male-biased protein is Shbg. In human, SHBG is a major plasma transporter of sex steroids and is associated with insulin resistance; however, the concentration of SHBG tends to be higher in women than men<sup>62</sup>. This difference may indicate functional variation of SHBG between human and fish.

As the liver is the major organ for plasma protein production, it is also of interest to compare the gender-biased plasma proteins with differentially expressed genes between male and female livers. We have previously reported sexual dimorphism of gene expression in zebrafish liver and relevant effects of sex hormones<sup>6</sup>. Consistent with the highly female-biased abundance of Vtgs and Nothepsin in female plasma, gene expression of *vtgs* and *nots* was highly enriched in liver of naïve female and E2-induced male but down-regulated in KT11-treated female. An uncharacterized male-biased protein, si:ch1073-126c3.2, also showed male-biased gene expression in liver and it was down-regulated in E2-treated male. Interestingly, this protein was also down-regulated in KT11-treated

male but not affected by either hormone treatment in female. Another gene *shbg*, coding for male-biased plasma protein Shbg, also showed 10-fold enrichment in the control male liver as compared to female. Different from what was observed in plasma proteins, gene expression for four female-biased apolipoproteins, Apobb.1, Apoc2, Apoea and Apoeb, were 3–9 fold higher in control male and not deregulated by either hormone in either gender. However, it should be noted that these proteins are also expressed in other tissues and secretion of apolipoproteins from liver to blood is also post-translationally regulated. For example, Apolipoprotein B is expressed in the liver, intestine and heart and its secretion can be affected by factors such as oleate and TGF- $\beta$  with or without change at mRNA level<sup>63–65</sup>. Other gender-biased proteins identified are not known to be liver-enriched and not reported as differentially expressed genes in the transcriptomic study either<sup>6</sup>.

Combining a non-targeted, semi-quantitative shotgun MS profile and a quantitative SWATH MS profile, we have provided a comprehensive overview of zebrafish plasma proteome. 959 proteins were identified with estimated concentrations spanning almost five orders of magnitude. Based on the SWATH profile, we identified 35 gender-biased proteins, many of which were newly reported. This study has provided a valuable dataset for further studies on identification or evaluation of plasma biomarkers using zebrafish as a model.

## Methods

**Fish maintenance.** This study involving zebrafish was carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and the protocol was approved by the Institutional Animal Care and Use Committee (IACUC) of the National University of Singapore (Protocol Number: 096/12). Adult zebrafish of 6-month old and 1.5-year old were used for testing methods of blood collection, estimation of total plasma protein content, and initial SDS-PAGE examination. For shotgun MS and SWATH, adult zebrafish (~9 months old) were purchased from a local fish farm (Mainland Fish Farm, Singapore) and allowed to acclimatize in aquaria for two weeks before sacrifice for plasma sampling. Fish maintenance and experiments were carried out following the protocols approved by Institutional Animal Care and Use Committee (IACUC) of National University of Singapore.

**Plasma sample preparation and assessment.** Blood was collected using a modified tail ablation method. Briefly, fish was euthanized in ice-water and its surface was wiped dry. Tail was removed by a cut at the end of anal fin and the fish was held with the wound facing down. Whole blood was collected at the dorsal aorta using P20 micropipette fitted to an elongated tip (Prot/Elec Tips, Bio-Rad) and aspirated into pre-chilled Eppendorf tubes. Both pipette tips and tubes were pre-coated with EDTA by submerging in 18 mg/mL EDTA solution for 24 hrs and then dried prior to use. Plasma was obtained as clear supernatant after centrifugation at 1000 g for 10 min at 4 °C. Typical yield of whole blood is 5–10  $\mu$ l from male fish and 8–15  $\mu$ L from female fish. Whole blood was collected to pre-chilled tubes on ice and then centrifuged within 10 min to obtain plasma. For MS analyses, each plasma sample was pooled from 9 fish. Protein concentration was determined using RCDC protein assay kit (Bio-Rad) according to the manufacturer's protocol. For quick evaluation of protein abundance patterns, six individual plasma samples (3 males and 3 females, 0.5  $\mu$ L each) were analyzed using 10% SDS gel followed by silver staining.

**Database compilation.** A non-redundant *Danio rerio* protein database was compiled using databases from International protein index (40,470), UniProtKB (13,383) and NCBI GenBank (22,288) for a combined total of 76,141 non-redundant sequence entries.

**Shotgun LC-MS/MS with MASCOT search.** LC-MS/MS analyses were performed in six technical replicates for each pooled sample. 2  $\mu$ g of each sample from male and female was trapped on a pre-column (200  $\mu$ m  $\times$  0.5 mm) and then eluted on an analytical column (75  $\mu$ m  $\times$  150 mm). Both columns were packed with ChromXP C18-CL 3  $\mu$ m 120 Å (Eksigent, Dublin, CA). Samples were run in a gradient formed by a mobile phase A (0.1% formic acid in 2% acetonitrile) and a mobile phase B (0.1% formic acid in 98% acetonitrile) at a flow rate of 300 nL/min. LC gradient was programmed as follows: 1 min of 95% A, 29 min of 95–88% A, 90 min of 88–70% A, 2 min of 70–10% A, 7 min of 10% A, 3 min of 10–95% and 14 min of 95% A.

Tandem MS analysis was performed using a 5600 TripleTOF analyzer (QqTOF; SCIEX) in Information Dependent Mode. Precursor ions were selected across the mass range of 400–1800 m/z using 250 ms accumulation time per spectrum. A maximum of 20 precursors per cycle from each MS spectra were selected for MS/MS analyses, with 100 ms accumulation time for each precursor and exclusion time set at 15 s. Tandem mass spectrometry was recorded in high sensitivity mode with rolling collision energy on.

The group file generated from ProteinPilot<sup>TM</sup> (SCIEX) was converted to MGF peak list by its built-in converter. Protein identification was performed with MASCOT Server 2.4.0 (Matrix Science). User-defined search parameters were as follows: (1) Type of search: MS/MS Ion Search; (2) Enzyme: Trypsin; (3) Variable modifications: Acetyl (N-term), Methylthio (C) and Oxidation (M); (4) Mass Values: Monoisotopic; (5) Protein mass: Unrestricted; (6) Peptide mass tolerance:  $\pm$ 200 ppm; (7) Fragment mass tolerance:  $\pm$ 0.4 Da; (8) Max missed cleavages: 1; (9) Instrument type: ESI\_QUAD-TOF; (10) Number of queries: 126,513; and (11) Taxonomy: *Danio rerio* (zebrafish). The data was searched against the compiled *Danio rerio* database (76,141 sequences) as described above with the integrated false discovery rate (FDR) analysis function. The protein identification threshold was set at 1% FDR based on reversed protein sequences. Concentrations of detected proteins were estimated using two steps: (1) the weight fraction (weight %) of each protein in plasma was obtained based on its emPAI (exponentially modified protein abundance index) score in MS profile and its molecular weight<sup>66</sup>; (2) the weight% was multiplied by the total plasma concentration of respective gender.

**SWATH.** Ion library for SWATH was built using the same LC-MS/MS set up as in shotgun MS using 2  $\mu$ g of each sample. Duplicates of male and female plasma samples were separated on the same LC gradient as



described above and analyzed using a 5600 TripleTOF system (QqTOF; SCIEX,) in Information Dependent Mode. Precursor ions were selected across the mass range of 350–1250 m/z using 250 ms accumulation time per spectrum. A maximum of 20 precursors per cycle from each MS spectra were selected for MS/MS analyses with 100 ms accumulation time for each precursor and dynamic exclusion for 15 s. Tandem mass spectrometry was recorded across the mass range of 100–1800 m/z in high sensitivity mode with rolling collision energy on. Protein identification was performed with ProteinPilot™ Software 4.1 (SCIEX) as described before<sup>67</sup>. User-defined search parameters were as follows: (1) Sample Type: Identification; (2) Cysteine Alkylation: MMTS; (3) Digestion: Trypsin; (4) Instrument: TripleTOF5600; (5) Special Factors: None; (6) Species: *Danio rerio*; (7) ID Focus: Biological modifications; (8) Database: compiled *Danio rerio* database as described above; (9) Search Effort: Thorough; (10) the False Discovery Rate Analysis: Yes; and (11) User Modified Parameter Files: No. The proteins identified were used to generate a SWATH™ Ion library.

Triplicate samples from male and female fish were analyzed on a 5600 TripleTOF system (QqTOF; SCIEX,) in SWATH mode. Each acquisition method had a cycle time of 3 s to include 36 acquisition windows of 25 Da. Each acquisition window had 80 ms accumulation time and 50 ms of TOF/MS survey scan from 350 to 1250 Da. Each MS/MS acquisition window had a scan of 100 to 1,800 Da and was performed using collision energy of 35 V with a spread of 15 V. The spectra from SWATH acquisition and the ion library generated were imported into PeakView software V2.1 (SCIEX) supplemented with MS/MS (ALL) with SWATH Acquisition Microapp, version 1.0.0.653 add-on where peak alignment and data interrogation was performed. The following filters were used: (1) Peptide confidence: 99%; (2) Exclude modifications: Yes; and (3) Exclude Shared: Yes. Extracted areas were exported into MarkerView for statistical analysis. With MarkerView, a global normalization was performed on the dataset. Mean area intensity and fold change of each protein between female and male were generated and exported. Intensity values for female plasma were then multiplied by a concentration scaling factor, which was determined as the ratio of total protein concentration of female plasma over that of male plasma.

**Data analysis and biological interpretation.** Gene ontology (GO) annotation for proteins detected was retrieved using Software Tool for Rapid Annotation of Proteins (STRAP, version 1.5)<sup>68</sup>, which extracts information from UniProt databases. Protein accessions were mapped to respective Entrez Gene IDs (GIs) and human homologous gene symbols, if available, by searching against NCBI database. Human orthologs were taken as the homologous proteins whenever available (~50%, including ~30% one-to-one ortholog pairs). When orthologs were not established, the respective homologous human genes in NCBI HomoloGene were used. Database for Annotation, Visualization and Integrated Discovery (DAVID) was used for functional annotation analysis including GO enrichment and tissue expression<sup>69,70</sup>. Ingenuity Pathway Analysis (IPA) was used for annotation and biological interpretation. For SWATH analysis, two-tailed heteroscedastic *t*-test was performed and gender-biased proteins were defined as the ones that met two criteria:  $p < 0.05$  and fold change (FC)  $> 2$ .

## References

- Aanes, H. *et al.* Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res* **21**, 1328–1338 (2011).
- Kobayashi, I. *et al.* Comparative gene expression analysis of zebrafish and mammals identifies common regulators in hematopoietic stem cells. *Blood* **115**, E1–E9 (2010).
- Mathavan, S. *et al.* Transcriptome analysis of zebrafish embryogenesis using microarrays. *Plos Genet* **1**, 260–276 (2005).
- Renn, S. C. P., Aubin-Horth, N. & Hofmann, H. A. Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BMC Genomics* **5**, 42–54 (2004).
- Shih, Y. H. *et al.* The Cardiac Transcriptome and Dilated Cardiomyopathy Genes in Zebrafish. *Circ. Cardiovasc. Genet.* **8**, 261–269 (2015).
- Zheng, W. *et al.* Transcriptomic analyses of sexual dimorphism of the zebrafish liver and the effect of sex hormones. *PLoS one* **8**, e53562 (2013).
- Zheng, W. L. *et al.* Comparative Transcriptome Analyses Indicate Molecular Homology of Zebrafish Swimbladder and Mammalian Lung. *PLoS One* **6**, e24019 (2011).
- Wang, Z. *et al.* Morphological and molecular evidence for functional organization along the rostrocaudal axis of the adult zebrafish intestine. *Bmc Genomics* **11**, 392 (2010).
- Tay, T. L. *et al.* Proteomic analysis of protein profiles during early development of the zebrafish, *Danio rerio*. *Proteomics* **6**, 3176–3188 (2006).
- Link, V., Shevchenko, A. & Heisenberg, C. P. Proteomics of early zebrafish embryos. *BMC Dev. Biol.* **6**, 1–9 (2006).
- Lucitt, M. B. *et al.* Analysis of the zebrafish proteome during embryonic development. *Mol Cell Proteomics* **7**, 981–994 (2008).
- Wang, N. *et al.* Proteome profile of cytosolic component of zebrafish liver generated by LC-ESI MS/MS combined with trypsin digestion and microwave-assisted acid hydrolysis. *J Proteome Res* **6**, 263–272 (2007).
- Posner, M. *et al.* A proteome map of the zebrafish (*Danio rerio*) lens reveals similarities between zebrafish and mammalian crystallin expression. *Mol Vis* **14**, 806–814 (2008).
- De Souza, A. G., MacCormack, T. J., Wang, N., Li, L. & Goss, G. G. Large-Scale Proteome Profile of the Zebrafish (*Danio rerio*) Gill for Physiological and Biomarker Discovery Studies. *Zebrafish* **6**, 229–238 (2009).
- Singh, S. K., Sundaram, C. S., Shanbhag, S. & Idris, M. M. Proteomic Profile of Zebrafish Brain Based on Two-Dimensional Gel Electrophoresis Matrix-Assisted Laser Desorption/Ionization MS/MS Analysis. *Zebrafish* **7**, 169–177 (2010).
- Singh, S. K., Saxena, S., Lakshmi, M. G. M., Saxena, P. & Idris, M. M. Proteome Profile of Zebrafish *Danio rerio* Olfactory Bulb Based on Two-Dimensional Gel Electrophoresis Matrix-Assisted Laser Desorption/Ionization MS/MS Analysis. *Zebrafish* **8**, 183–189 (2011).
- Abramsson, A. *et al.* Proteomics Profiling of Single Organs from Individual Adult Zebrafish. *Zebrafish* **7**, 161–168 (2010).
- Kessels, M. Y. *et al.* Proteomics Analysis of the Zebrafish Skeletal Extracellular Matrix. *PLoS One* **9**, e90568 (2014).
- Boatman, S. *et al.* Assaying hematopoiesis using zebrafish. *Blood Cell Mol Dis* **51**, 271–276 (2013).
- Stoletov, K. *et al.* Vascular Lipid Accumulation, Lipoprotein Oxidation, and Macrophage Lipid Uptake in Hypercholesterolemic Zebrafish. *Circ Res* **104**, 952–960 (2009).
- Moss, J. B. *et al.* Regeneration of the Pancreas in Adult Zebrafish. *Diabetes* **58**, 1844–1851 (2009).
- Nanjappa, V. *et al.* Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic acids research* **42**, D959–D965 (2014).
- Zhang, Q. *et al.* A mouse plasma peptide atlas as a resource for disease proteomics. *Genome biology* **9**, R93 (2008).



24. Lai, K. K., Kolippakkam, D. & Beretta, L. Comprehensive and quantitative proteome profiling of the mouse liver and plasma. *Hepatology* **47**, 1043–1051 (2008).
25. Jagadeeswaran, P. & Sheehan, J. P. Analysis of blood coagulation in the zebrafish. *Blood Cell Mol Dis* **25**, 239–249 (1999).
26. Eames, S. C., Philipson, L. H., Prince, V. E. & Kinkel, M. D. Blood Sugar Measurement in Zebrafish Reveals Dynamics of Glucose Homeostasis. *Zebrafish* **7**, 205–213 (2010).
27. Babaei, F. *et al.* Novel Blood Collection Method Allows Plasma Proteome Analysis from Single Zebrafish. *J Proteome Res* **12**, 1580–1590 (2013).
28. Pedroso, G. L. *et al.* Blood Collection for Biochemical Analysis in Adult Zebrafish. *Jove-Journal Of Visualized Experiments* **26**, e3865 (2012).
29. Pernemalm, M. & Lehtio, J. Mass spectrometry-based plasma proteomics: state of the art and future outlook. *Expert review of proteomics* **11**, 431–448 (2014).
30. Nanjappa, V. *et al.* Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic acids research* **42**, D959–965 (2014).
31. Gillet, L. C. *et al.* Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics : MCP* **11**, O111.016717 (2012).
32. Noel, E. S., dos Reis, M., Arain, Z. & Ober, E. A. Analysis of the Albumin/alpha-Fetoprotein/Afamin/Group specific component gene family in the context of zebrafish liver differentiation. *Gene Expr Patterns* **10**, 237–243 (2010).
33. Wang, H., Tan, J. T., Emelyanov, A., Korzh, V. & Gong, Z. Hepatic and extrahepatic expression of vitellogenin genes in the zebrafish, *Danio rerio*. *Gene* **356**, 91–100 (2005).
34. Jiang, Y. *et al.* First Insight into the Human Liver Proteome from PROTEOMESKY-LIVERHu 1.0, a Publicly Available Database. *J Proteome Res* **9**, 79–94 (2010).
35. Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**, 845–867 (2002).
36. Liu, X. *et al.* Mapping the human plasma proteome by SCX-LC-IMS-MS. *Journal of the American Society for Mass Spectrometry* **18**, 1249–1264 (2007).
37. Ping, P. *et al.* A functional annotation of subproteomes in human plasma. *Proteomics* **5**, 3506–3519 (2005).
38. Marks, P. W. Hematologic manifestations of liver disease. *Seminars in hematology* **50**, 216–221 (2013).
39. Xu, H., Lam, S. H., Shen, Y. & Gong, Z. Genome-wide identification of molecular pathways and biomarkers in response to arsenic exposure in zebrafish liver. *PLoS one* **8**, e68737 (2013).
40. Niu, D. D., Feng, H. X. & Chen, W. N. Proteomic analysis of HBV-associated HCC: Insights on mechanisms of disease onset and biomarker discovery. *J Proteomics* **73**, 1283–1290 (2010).
41. Leung, F., Diamandis, E. P. & Kulasingam, V. Ovarian Cancer Biomarkers: Current State and Future Implications from High-Throughput Technologies. *Adv Clin Chem* **66**, 25–77 (2014).
42. Harmsma, M., Schutte, B. & Ramaekers, F. C. S. Serum markers in small cell lung cancer: Opportunities for improvement. *Bba-Rev Cancer* **1836**, 255–272 (2013).
43. Yotsukura, S. & Mamitsuka, H. Evaluation of serum-based cancer biomarkers: A brief review from a clinical and computational viewpoint. *Crit Rev Oncol Hemat* **93**, 103–115 (2015).
44. Hanash, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nature* **452**, 571–579 (2008).
45. Ober, C., Loisel, D. A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nature reviews. Genetics* **9**, 911–922 (2008).
46. Silliman, C. C. *et al.* Proteomic analyses of human plasma: Venus versus Mars. *Transfusion* **52**, 417–424 (2012).
47. Kim, C. X. *et al.* Sex and ethnic differences in 47 candidate proteomic markers of cardiovascular disease: the Mayo Clinic proteomic markers of arteriosclerosis study. *PLoS one* **5**, e9065 (2010).
48. Izbicka, E. *et al.* Plasma biomarkers distinguish non-small cell lung cancer from asthma and differ in men and women. *Cancer genomics & proteomics* **9**, 27–35 (2012).
49. Holford, N. C., Sandhu, H. S., Thakkar, H., Butt, A. N. & Swaminathan, R. Stability of beta-actin mRNA in plasma. *Annals of the New York Academy of Sciences* **1137**, 108–111 (2008).
50. Zhang, R. *et al.* beta-actin as a loading control for plasma-based Western blot analysis of major depressive disorder patients. *Analytical biochemistry* **427**, 116–120 (2012).
51. Deans, N. D., Chow, M. C., Kroll, K. J. & Green, L. Vitellogenin as a biomarker of exposure for estrogen or estrogen mimics. *Ecotoxicology* **8**, 385–398 (1999).
52. Riggio, M., Scudiero, R., Filosa, S. & Parisi, E. Sex- and tissue-specific expression of aspartic proteinases in *Danio rerio* (zebrafish). *Gene* **260**, 67–75 (2000).
53. Knoll-Gellida, A. *et al.* Molecular phenotype of zebrafish ovarian follicle by serial analysis of gene expression and proteomic profiling, and comparison with the transcriptomes of other animals. *Bmc Genomics* **7**, 46 (2006).
54. Levi, L. *et al.* Revealing genes associated with vitellogenesis in the liver of the zebrafish (*Danio rerio*) by transcriptome profiling. *Bmc Genomics* **10**, 141 (2009).
55. Benes, P., Vetvicka, V. & Fusek, M. Cathepsin D—many functions of one aspartic protease. *Critical reviews in oncology/hematology* **68**, 12–28 (2008).
56. Deasey, S., Grichenko, O., Du, S. & Nurminskaya, M. Characterization of the transglutaminase gene family in zebrafish and *in vivo* analysis of transglutaminase-dependent bone mineralization. *Amino acids* **42**, 1065–1075 (2012).
57. Hao, R. *et al.* Identification of estrogen target genes during zebrafish embryonic development through transcriptomic analysis. *PLoS one* **8**, e79020 (2013).
58. Fadoo, Z., Merchant, Q. & Rehman, K. A. New developments in the management of congenital Factor XIII deficiency. *Journal of blood medicine* **4**, 65–73 (2013).
59. Beutler, E. PGK deficiency. *Brit J Haematol* **136**, 3–11 (2007).
60. Na, K. *et al.* Human liver carboxylesterase 1 outperforms alpha-fetoprotein as biomarker to discriminate hepatocellular carcinoma from other liver diseases in Korean patients. *Int J Cancer* **133**, 408–415 (2013).
61. Yeh, S. H. & Chen, P. J. Gender disparity of hepatocellular carcinoma: the roles of sex hormones. *Oncology* **78** Suppl 1, 172–179 (2010).
62. Geer, E. B. & Shen, W. Gender Differences in Insulin Resistance, Body Composition, and Energy Balance. *Gender Med* **6**, 60–75 (2009).
63. Dixon, J. L., Furukawa, S. & Ginsberg, H. N. Oleate stimulates secretion of apolipoprotein B-containing lipoproteins from Hep G2 cells by inhibiting early intracellular degradation of apolipoprotein B. *The Journal of biological chemistry* **266**, 5080–5086 (1991).
64. Pullinger, C. R. *et al.* The Apolipoprotein-B Gene Is Constitutively Expressed In Hepg2 Cells - Regulation Of Secretion by Oleic-Acid, Albumin, And Insulin, And Measurement Of the Messenger-Rna Half-Life. *J Lipid Res* **30**, 1065–1077 (1989).
65. Singh, K. *et al.* Differential, tissue-specific, transcriptional regulation of apolipoprotein B secretion by transforming growth factor beta. *The Journal of biological chemistry* **277**, 39515–39524 (2002).
66. Ishihama, Y. *et al.* Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**, 1265–1272 (2005).
67. Ghosh, D. *et al.* iTRAQ based quantitative proteomics approach validated the role of calyculin binding protein (CacyBP) in promoting colorectal cancer metastasis. *Mol Cell Proteomics* **12**, 1865–1880 (2013).

68. Bhatia, V. N., Perlman, D. H., Costello, C. E. & McComb, M. E. Software tool for researching annotations of proteins: open-source protein annotation software with data visualization. *Analytical chemistry* **81**, 9819–9823 (2009).
69. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).
70. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1–13 (2009).

### Acknowledgements

This work was supported by the Singapore National Research Foundation under its Environmental and Water Technologies Strategic Research Programme and administered by the Environment and Water Industry Programme Office (EWI) of the PUB, grant number R-154-000-328-272, and National Medical Research Council of Singapore, grant number R-154-000-547-511. Caixia Li is supported by the NRF Singapore under its NRF (EWT) PhD Scholarship Programme and administered by the Environment & Water Industry Programme Office (EWI) of the Public Utilities Board, Singapore. We thank Ms Li Zhou for her help in making the non-redundant zebrafish protein database.

### Author Contributions

C.L., Q.L. and Z.G. conceived experiments and wrote manuscript. C.L., X.F.T. and T.K.L. performed all experiments and analyzed data. All authors reviewed and approved the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Li, C. *et al.* Comprehensive and quantitative proteomic analyses of zebrafish plasma reveals conserved protein profiles between genders and between zebrafish and human. *Sci. Rep.* **6**, 24329; doi: 10.1038/srep24329 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>