12-2018

The First In Vivo Human Methionine Sulfide Proteome and the Impact of Smoking

Abdullah Qassab
University of Arkansas, Fayetteville

The First In Vivo Human Methionine Sulfoxide Proteome and the Impact of Smoking

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Cell and Molecular Biology

by

Abdullah M. Qassab
Salahaddin University
Bachelor of Science in Biological Sciences, 2002
Salahaddin University
Master of Science in Animal Genetics, 2008

December 2018
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Wesley E. Stites, Ph.D.
Dissertation Director

_____
Suresh Kumar Thallapuranam, Ph.D.
Committee Member

_____
Ralph Henry, Ph.D.
Committee Member

_____
Joshua Sakon, Ph.D.
Committee Member

ABSTRACT

Reactive oxygen species are naturally generated within the human body and they are known to modulate signaling pathway and mediate other physiological activities. However, excessive generation of ROS and the inability of body defense system in detoxifying them results in the so called “oxidative stress”. Methionine has powerful antioxidant properties due to the presence of electronegative sulfur in its structure. Therefore, Met is readily oxidized, and methionine sulfoxide has been linked to several pathological conditions.

The urinary proteome is an attractive candidate for the discovery of biomarkers to diagnose and classify health conditions because of the non-invasive collection procedure. However, protocols developed for the analysis of the urinary proteome have limitations both in identification of proteins and eliminating false positives.

Urine samples were collected from smoker and non-smokers participants. Proteins were concentrated and concentrates then were subjected to separation via SDS-PAGE gel electrophoresis technique. Gel was cut out into five sections including the band representing intact HSA. Gel sections then were digested with trypsin. Total proteins and levels of methionine sulfoxide in the resulting peptides were assessed by mass spectrometry. Data analysis was performed using Scaffold and Skyline software packages.

Correspondence with predicted retention times, MS/MS sequence information, and a minimum of two tryptic peptides were required to reduce false positives. Proteins were matched to 1023 UniProtKB identification numbers, 467 of which have not been previously reported in the human urinary proteome. Analysis of methionine sulfoxide proteome revealed 392 peptides mapped to 198 proteins that were detected with a significant fold change in smokers. Regions of gels with proteins of lower mass showed significant higher levels of oxidation in smokers compared to

control, suggesting an association between oxidative caused by smoking and fragmentation of proteins.

ACKNOWLEDGEMENTS

This work would not have been possible without the guidance, encouragement, and support of my advisor, Dr. Wesley Stites. I am deeply thankful to all his efforts, recommendations and more importantly for finding me a graduate assistantship position in the department of Chemistry and Biochemistry for the years I am involved in the PhD program. His brilliant ideas with no doubt have paved and shaped my path toward a successful career in my near future.

I am also grateful Dr. Rohana Liyanage. He was always available whenever I needed his help. I learned a lot about mass spectrometry and analysis of data from him.

I would like thank my graduate research committee; Dr. Suresh Kumar Thallapuranam, Dr. Josh Sakon, and Dr. Ralph Henry for all their support and valuable recommendations throughout the years I have been pursuing my degree at the University of Arkansas, especially during my candidacy exam back in Fall 2014. I would like thank the Cell and Molecular Biology program, particularly the director of program, Professor Douglas Rhoads for his endless support, without his support it would be very hard for me to pursue this degree.

Nobody has been more important to me in the pursuit of this degree than the members of my family. I would like to thank my mother, I know she never quit praying for me, her love and guidance are with me in whatever I pursue. I thank them all for their endless inspiration.

My thanks to all of those with whom I have had the pleasure to work during this project especially (Samrat, Matthew, Collin, Alexis, and Samuel) and the two former graduate students worked in Dr. Stites lab (Dr. Jeffery Froude and Dr. Chris Saunders) whom I have not met in person but they are collaborating in accomplishing some related project.

Thanks to the Human Capacity Development Program (HCDP), Kurdistan Regional Government for the partial funding and support.

TABLE OF CONTENTS

CHAPTER ONE

1.1 INTRODUCTION

The complete set of proteins and peptides within a particular body compartment is known as proteome. Proteomes include a wide spectrum ranging from proteins with high molecular masses to short polypeptide chains, proteins that are very abundant to extremely rare, and a wide variety of pre- and post-translational modifications. In contrast to genomes, which are relatively stable, proteomes are highly affected by factors including gene expression and protein secretion and turnover. Advances in protein detection and identification enables detecting even low abundance proteins or protein variants generated in response to certain health conditions or environmental factors.

Due to the inherent complexity of proteomes, samples are usually separated in some fashion prior to protein identification. The SDS-PAGE developed by Laemmli and the two-dimensional gel electrophoresis conducted by O'Farrell in mid 1970s were the earliest studies of what is known as proteomics today (1,2). The technique separates proteins based on their molecular masses. However, the challenge of protein identification in 2DE remained until Burnette described Western blot assay using antibodies (3). Such techniques have been used successfully for decades now. However, advancements in protein identification using mass spectrometry was a revolution in studying a wide array of proteins with greater accuracy. Today, experimental procedures rely on gel electrophoresis and/or liquid chromatography followed by mass spectrometry for identification. Moreover, LC can be easily coupled to almost all MS instruments (4-6). Identifying proteins using these two separation techniques often involves enzymatic digestion (7,8) of polypeptides at some stage to produce fragments that can be easily detected and identified by MS. Usually, the techniques uses at least two of the produced fragments to search in a protein database

for identification purposes. A key advantage of using MS techniques is that the peptide identification can be confirmed by MS/MS sequencing. However, 1DE technique has limitations such as low reproducibility and the significant long time required for the analysis. Liquid chromatography (LC) on the other hand, is a robust tool used in fractionation and separating complex biological samples.

As these techniques have improved, it is not surprising that proteomics has gained a considerable amount of attention over last few decades. Body fluids are attractive targets for clinical proteomics. Urinary proteomics has become a popular topic in clinical biomarker discovery. The simple and noninvasive nature of urine sampling provides large quantities of material Urine contains around 150 mg of protein in each 24 hour period (9).. It offers the potential for significant insights into normal and pathological conditions due to the wide range of its proteome profile. Urine directly reflects the condition of the urinary system, but also contains considerable amounts of plasma proteins and proteins from a wide range of cell types throughout the body (10). Thus, pathological and physiological changes can be detected in urine samples (11-13). By means of comparative proteomics, many urinary proteins such as albumin, cystatin C, α-1and β-2 microglobulin (14), have been identified as potential biomarkers reflecting different illnesses including glomerular dysfunction, bladder cancer, renal cell carcinoma, renal failure, diabetic nephropathy, acute kidney injury, and non-renal diseases including cardiovascular diseases (15,16) and atherosclerosis (17). One of the first studies that sought to more fully describe the urinary proteome using LC-MS was published by Spahr et al. (6,18). They pooled urine samples and the analysis resulted in identifying 124 proteins. Although the study did not attempt to discover any biomarkers, it paved the road toward further urine proteome examinations. Pang et al. (19) attempted to discover biomarkers specific for inflammation in urine samples of healthy

individuals. Later, scientists tried to study urine proteome more in depth. Thonboonkerd et al (20) defined the first human normal urinary proteome in 2002. They identified 47 unique proteins using 2DE electrophoresis followed by matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectrometry. Further expansion of human urinary proteome was reported by studies conducted later by Pieper *et al.* in 2004 (21), Sun *et al* (13), and Castagna *et al.* (22) in 2005. These studies collectively identified less than 1,000 proteins. In the following years, studies continued mining urinary proteome and, as will be discussed in a subsequent chapter, the current urine proteome map encompass over 11,000 proteins.

Beyond identifying and quantifying proteins, another area of biomarker discovery using MS approaches is identification of posttranslational modifications (PTMs). The number of proteins in an organism exceeds the DNA coding capacity in response to the RNA metabolism and alternative splicing, also due to the various covalent posttranslational modifications (PTMs) occur at different sites (23). Many PTMs of proteins are intentional, specific, and vital for cellular survival. Examples would include proteolytic maturation, glycosylation, and phosphorylation events. Others may be more random in nature and possibly deleterious to the organism. Examples include proteolysis by pathogens, damage from ultraviolet light, Amadori modification of proteins by glucose in diabetes, and oxidation. PTMs can be observed in many pre-MS separations, since modification often changes mass or charge of the proteins or fragments being analyzed.

The use of or MS/MS approaches is vital if one to determine the precise location and nature of PTMs in the proteome. The first MS instrument is employed as a mass filter of the so called "parent ions" and the following MS step serves in analyzing products of fragmentation "daughter ions" induced via collision with other molecules in a process called collision-induced fragmentation (24). Even with MS/MS techniques to pinpoint the location of PTMs, their analysis can be

potentially difficult due to several factors including low abundance of the modified peptide, low stability of the modification, and the degree to which the modification itself impacts ionization efficiency. Samples handled for proteomic studies are typically treated with SDS, thus, extraction of proteins with PTMs could be different from unmodified proteins (25). While the location of some modifications can be determined via MS/MS, this requires the PTM to withstand different energetics used in MS and MS/MS analysis. Additionally, the shift in the mass due to the modification should not shift the mass of the peptide beyond the mass range being detected by MS/MS (~800-2500Da) (26).

The PTM of particular interest here is protein oxidation, which is of interest in different disciplines including for example, in studies of aging and age-related diseases, such as Alzheimer's disease; in the field of therapeutic proteins; food-storage and in industrial biotechnology (27). The concept of oxidative stress was first introduced to redox biology by Helmut Sies in 1985 (28). Reactive oxygen species (ROS) are inevitable consequences of aerobic metabolism. Various cellular processes and enzymes are capable of generating oxidative PTMs. Later, oxidation of macromolecules in response to ROS was linked to a variety of pathologies. ROS leading to oxidative stress include free radicals such as superoxide ($O^{\cdot -}_2$), hydroxyl (OH^-), peroxy (RO_2), alcoxyl (RO^-), hydroperoxyl ($HO^{\cdot}_2$), and non-radical species such as hydrogen peroxide (H_2O_2), hydrochlorous acid (HOCl), ozone (O_3), singlet oxygen ($^1O_2$), and peroxinitrite ($ONOO^-$) (29,30). Reuter *et al.* reviewed the link between oxidative stress, inflammation and cancer (31). They pointed that oxidative stress can activate transcriptional factor responsible for expression of over 500 genes including those for cytokines and chemokines. They concluded that the three elements (oxidative stress, inflammation and cancer) are closely related (31). The main source of endogenous ROS can be summarized as the addition of an electron to molecular oxygen to form

superoxide anion, which is mediated by NADPH that occurs in mitochondria (32). In addition, upon phagocytosis, macrophages evolve a bactericidal mechanism through the formation of oxidative burst. Superoxide is converted into hydrogen peroxide through the action of superoxide dismutase enzymes and the product can easily diffuse across the plasma membrane. Hydrogen peroxide then can breakdown to OH^- in the presence of transition metals in Haber-Weiss and Fenton reactions (33,34). Hydroxyl radical is the most reactive among ROS and it can impair the function and damage proteins, lipids and nucleic acids (35).

Of course, the body is naturally able in some degree to detoxify generated oxygen species by producing antioxidants capable of capturing ROS. Imbalance between the formation and detoxification of ROS can result in oxidative stress and non-enzymatic modification of biomolecules (36). Oxidative damage to proteins can virtually affect all amino acids. However, methionine and cysteine residues react more readily due to the presence of sulfur atoms in their structures (37-39). It has been proposed that these amino acids are among the main ROS scavengers and thus they protect proteins and other macromolecules from deleterious effects of oxidative damage (40).

The major product of oxidation of methionine is methionine sulfoxide and this PTM is the main focus of this dissertation. Methionine sulfoxide adducts are of great interest in that they can be reversed back to methionine via the action of methionine sulfoxide reductases. Further oxidation of methionine can lead to the formation of irreversible methionine sulfone (41).

Not all methionines are equally susceptible to oxidation. Most methionine residues are localized in the hydrophobic core of proteins due their hydrophobicity and they mainly interact with lipid bilayer in the membrane proteins, therefore, many are fairly protected from oxidation

by polar oxidants. However, surface exposed methionine amino acids are susceptible to oxidation (42-46). But even on the surface, the rate of formation of methionine sulfoxide is dependent on the amino acids found in vicinity of Met. Ghesquiere *et al.* found that Met are more susceptible to oxidation when surrounded by acidic amino acids than basic ones (41).

The consequences of oxidative modification of proteins in general and methionine sulfoxide in particular can undergo biological degradation or tend to form larger aggregates due to increased surface hydrophobicity (47). In a review article, Stadtman *et al.*, relying on a study conducted using *E coli* glutamine synthetase, stated that oxidation of methionine residues leads to conformational changes. The enzyme has ten surface-exposed Met residues. The study observed no detectable changes in surface hydrophobicity when seven of the residues were oxidized, but with further oxidation of two additional Met residues, a sharp increase in surface hydrophobicity was observed suggesting a conformational change led to expose buried residues to the outer surface, as result of the change, proteolytic degradation via 20S proteasome was detected (48). Fujino *et al.* investigated cleavage sites of bovine serum albumin treated with H_2O_2. Computer model of BSA revealed that the products of cleavage induced by oxidized protein hydrolase are located in the internal core of the protein, and the oxidation influenced the structure in a way hydrophobic Met residues became accessible by the proteolytic enzyme. They replaced Met and tryptophan residues found in the hydrolysis products with their oxidized forms. Later, they found the fragments became more exposed to the solvent than un-oxidized BSA (49). Furtunatelly, oxidation of methionine is reversible via the action of methionine sulfoxide reductases (MsrA and MsrB). Several studies have examined the physiological importance of Msr enzymes in prokaryotes and eukaryotes. Gebendorfer *et al.* noticed increased sensitivity and reduced viability of mutant *E. coli* lacking the gene producing MsrA (50). Pennington *et al.* described an alternative

pathway in which Nrf2 transcription factor (binds to the antioxidant response element region of the DNA) initiates the transcription of genes responsible in the production of antioxidant enzymes under non-stressed conditions in mice with MsrA knockout (51).

While many research teams have sought to reveal the prevalence and location in the proteome of PTMs such as glycosylation, phosphorylation, sulfation, and proteolysis, fewer groups have tried to decipher oxidation post-translational modification and more specifically the formation of methionine sulfoxide. Madian *et al.* isolated sixty-five carbonylated proteins from plasma of male subjects. They also identified that in addition to the carbonylation, the proteins bore an oxidative modification and that methionine oxidation was the predominant followed by tryptophan oxidation (52). Ghesquière *et al.* studied methionine bound oxidation PTM *in vitro* using human Jurkat cells stressed with hydrogen peroxide. They identified over 2000 oxidation sensitive Met residues mapped to more than 1600 different proteins (41). Hsieh *et al.* studied the global Met oxidation changes of proteins in photofrin-mediated photodynamic therapy (PDT) in human carcinoma A431(53). They observed 431 Met residues linked to 302 proteins that suffered from severe oxidation. In a time-course study monitoring protein-bound Met oxidation in *Bacillus cereus*, Madeira *et al.* detected total of 8720 and 3417 Met(O) peptides in the cellular proteomes and exoproteomes, respectively (54).

Albumin represents the most abundant serum protein with a normal concentration ranging from 35-50 g/l (55-57). Human Serum Albumin (HSA) participates in a multitude of functions such as transporting of molecules, maintaining osmotic pressure and serum redox state in the extracellular fluids (58-60). Furthermore, HSA is the most important antioxidant in plasma (61-63). Albumin is synthesized in liver with a half-life of ~27 days. HSA is single non-glycosylated polypeptide,

66438 Da, organized in three domains I, II and III each of which is comprised of A and B subunits. About 67% of albumin structure is α-helix It contains a single polypeptide chain of 585 amino acids. The structure owns 35 Cys residues forming 17 disulfide bridges with a single free Cys34 that constitutes about 80% of all free thiols in plasma and it is therefore considered as quantitatively important oxygen scavenger in plasma (57,60,64,65). A study conducted by Finch *et al.* concluded that Cys34, Met123, Met298, Met446 and Met548 are the main oxidation sites within the structure of HSA (66).

The exogenous factor of cigarette smoke (CS) is rich in free radicals and causes serious oxidative stress. Tobacco smoke contains over 4,000 components, of which over 1,000 are very toxic chemicals such as (3,4-benzopyrene, heavy metals, free radicals, hydrogen cyanide, nitrogen oxides and N-nitrosamines) which are proven carcinogens (67-71). The response of the human body to stressing factors can be studied in its protein expression profile. Therefore, identifying changes in protein expression upon smoking will lead to understanding physiological responses due to smoking. Reznick *et al.* studied carbonyl (the appearance of carbonyl groups in the structure of proteins such as aldehydes or ketones due to oxidation) formation in plasma proteins of smokers. They found that isolated albumin and creatinine kinase exposed to CS showed much faster carbonyl formation than did the whole plasma (72). Yeh *et al.* found that protein carbonyl levels was higher among current smokers compared to former smokers, but the difference did not attend statistical significance (68). D'Anna *et al.* examined changes in proteome profile of lung fibroblast exposed to CS extract. Eleven proteins demonstrated changes, of which, seven were shown to be significantly higher and the remaining proteins were lower in cells exposed to CS (73). In a different approach, Kulikowska (74) *et al.* measured the concentration of the genotoxic reagent (8-hydroxy-2'-deoxyguanosine) in the urine of smokers. They estimated a 3-fold increase in the

concentration of the reagent urine samples of smokers. Witherell *et al.* also noticed an elevated levels of oxidative DNA adduct 8-hydroxy-2'-deoxyguanosine excreted in urine after infection *with Helicobacter pylori* (75). A study conducted Panda et al. confirmed that the tar phase of CS is responsible for oxidation of microsomal proteins in vitro and the oxidized proteins were rapidly degraded by the action of proteases found in the microsomes (76). Clerici et al. observed an impaired ligand-binding capacity of albumin oxidized with CS (77).

In brief, proteome analysis still requires further expansion in order to display its full potential to become the predominant tool used in clinical diagnosis. However, increasing number of proteomics studies especially those recruited larger number of patients clearly demonstrates diagnostic accuracy. It might take long years for the full exploration of urinary proteome. A crucial question would be whether the focus should be turned into exploring the whole proteome of urine or we should take advantage of studying a subset of proteins with important clinical information.

1.2 REFERENCES

1. O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007-4021

2. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685

3. Burnette, W. N. (1981) "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal Biochem* **112**, 195-203

4. Zürbig, P., Renfrow, M. B., Schiffer, E., Novak, J., Walden, M., Wittke, S., Just, I., Pelzing, M., Neusüss, C., Theodorescu, D., Root, K. E., Ross, M. M., and Mischak, H. (2006) Biomarker discovery by CE-MS enables sequence analysis via MS/MS with platform-independent separation. *Electrophoresis* **27**, 2111-2125

5. Issaq, H. J., Conrads, T. P., Janini, G. M., and Veenstra, T. D. (2002) Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis* **23**, 3048-3061

6. Spahr CS 1 , D. M., McGinley MD , Robinson JH , Bures EJ , Beierle J , Mort J , Courchesne PL , Chen K , Wahl RC , Yu W , Luethy R , Patterson SD . (2001) Towards defining the urinary proteome using liquid chromatography - tandem mass spectrometry . I. Profiling an unfractionated tryptic digest., Proteomics

7. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A* **90**, 5011-5015

8. Henzel, W. J., Watanabe, C., and Stults, J. T. (2003) Protein identification: the origins of peptide mass fingerprinting. *J Am Soc Mass Spectrom* **14**, 931-942

9. González-Buitrago, J. M., Ferreira, L., and Lorenzo, I. (2007) Urinary proteomics. *Clin Chim Acta* **375**, 49-56

10. Pisitkun, T., Shen, R. F., and Knepper, M. A. (2004) Identification and proteomic profiling of exosomes in human urine. *Proc Natl Acad Sci U S A* **101**, 13368-13373

11. Siwy, J., Mullen, W., Golovko, I., Franke, J., and Zürbig, P. (2011) Human urinary peptide database for multiple disease biomarker discovery. *Proteomics Clin Appl* **5**, 367-374

12. Moon, P. G., You, S., Lee, J. E., Hwang, D., and Baek, M. C. (2011) Urinary exosomes and proteomics. *Mass Spectrom Rev* **30**, 1185-1202

13. Sun, W., Li, F., Wu, S., Wang, X., Zheng, D., Wang, J., and Gao, Y. (2005) Human urine proteome analysis by three separation approaches. *Proteomics* **5**, 4994-5001

14. Kentsis, A., Monigatti, F., Dorff, K., Campagne, F., Bachur, R., and Steen, H. (2009) Urine proteomics for profiling of human disease using high accuracy mass spectrometry. *Proteomics Clin Appl* **3**, 1052-1061

15. Napoli, C., Zullo, A., Picascia, A., Infante, T., and Mancini, F. P. (2013) Recent advances in proteomic technologies applied to cardiovascular disease. *J Cell Biochem* **114**, 7-20

16. Sharma, P., Cosme, J., and Gramolini, A. O. (2013) Recent advances in cardiovascular proteomics. *J Proteomics* **81**, 3-14

17. von Zur Muhlen, C., Schiffer, E., Zuerbig, P., Kellmann, M., Brasse, M., Meert, N., Vanholder, R. C., Dominiczak, A. F., Chen, Y. C., Mischak, H., Bode, C., and Peter, K. (2009) Evaluation of urine proteome pattern analysis for its potential to reflect coronary artery atherosclerosis in symptomatic patients. *J Proteome Res* **8**, 335-345

18. Davis, M. T., Spahr, C. S., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Yu, W., Luethy, R., and Patterson, S. D. (2001) Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. II. Limitations of complex mixture analyses. *Proteomics* **1**, 108-117

19. Pang, J. X., Ginanni, N., Dongre, A. R., Hefta, S. A., and Opitek, G. J. (2002) Biomarker discovery in urine by proteomics. *J Proteome Res* **1**, 161-169

20. V ISITH T HONGBOONKERD , K. E. R. M. C. L. E., J OHN M. A RTHUR , and J ON B. K LEIN. (2002) Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation. *Kidney International* **62**, 1461–1469

21. Pieper, R., Gatlin, C. L., McGrath, A. M., Makusky, A. J., Mondal, M., Seonarain, M., Field, E., Schatz, C. R., Estock, M. A., Ahmed, N., Anderson, N. G., and Steiner, S. (2004) Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. *Proteomics* **4**, 1159-1174

22. Castagna, A., Cecconi, D., Sennels, L., Rappsilber, J., Guerrier, L., Fortis, F., Boschetti, E., Lomas, L., and Righetti, P. G. (2005) Exploring the hidden human urinary proteome via ligand library beads. *J Proteome Res* **4**, 1917-1930

23. Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J. (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**, 7342-7372

24. Fliser, D., Novak, J., Thongboonkerd, V., Argilés, A., Jankowski, V., Girolami, M. A., Jankowski, J., and Mischak, H. (2007) Advances in urinary proteome analysis and biomarker discovery. *J Am Soc Nephrol* **18**, 1057-1071

25. Larsen, M. R., Sørensen, G. L., Fey, S. J., Larsen, P. M., and Roepstorff, P. (2001) Phospho-proteomics: evaluation of the use of enzymatic de-phosphorylation and differential mass spectrometric peptide mass mapping for site specific phosphorylation assignment in proteins separated by gel electrophoresis. *Proteomics* **1**, 223-238

26. Parker, C. E., Mocanu, V., Mocanu, M., Dicheva, N., & Warren, M. R. (2010) *Mass spectrometry for post-translational modifications*, CRC Press/Taylor & Francis, Boca Raton (FL)

27. Törnvall, U. (2010) Pinpointing oxidative modifications in proteins—recent advances in analytical methods. *Analytical Methods*

28. Sies, H., and Cadenas, E. (1985) Oxidative stress: damage to intact cells and organs. *Philos Trans R Soc Lond B Biol Sci* **311**, 617-631

29. Dalle-Donne, I., Rossi, R., Giustarini, D., Milzani, A., and Colombo, R. (2003) Protein carbonyl groups as biomarkers of oxidative stress. *Clin Chim Acta* **329**, 23-38

30. Sies, H. (1997) Oxidative stress: oxidants and antioxidants. *Exp Physiol* **82**, 291-295

31. Reuter, S., Gupta, S. C., Chaturvedi, M. M., and Aggarwal, B. B. (2010) Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic Biol Med* **49**, 1603-1616

32. Miller, D. M., Buettner, G. R., and Aust, S. D. (1990) Transition metals as catalysts of "autoxidation" reactions. *Free Radic Biol Med* **8**, 95-108

33. Carocci, A., Catalano, A., Sinicropi, M. S., and Genchi, G. (2018) Oxidative stress and neurodegeneration: the involvement of iron. *Biometals* **31**, 715-735

34. Gutteridge, J. M. (1994) Hydroxyl radicals, iron, oxidative stress, and neurodegeneration. *Ann N Y Acad Sci* **738**, 201-213

35. Yan, L. J., and Sohal, R. S. (1998) Gel electrophoretic quantitation of protein carbonyls derivatized with tritiated sodium borohydride. *Anal Biochem* **265**, 176-182

36. Dunlop, R. A., Brunk, U. T., and Rodgers, K. J. (2009) Oxidized proteins: mechanisms of removal and consequences of accumulation. *IUBMB Life* **61**, 522-527

37. Sen, K. I., Hepler, R., and Nanda, H. (2017) Detection and Measurement of Methionine Oxidation in Proteins. *Curr Protoc Protein Sci* **87**, 14.16.11-14.16.11

38. Drazic, A., and Winter, J. (2014) The physiological role of reversible methionine oxidation. *Biochim Biophys Acta* **1844**, 1367-1382

39. Cabreiro, F., Picot, C. R., Friguet, B., and Petropoulos, I. (2006) Methionine sulfoxide reductases: relevance to aging and protection against oxidative stress. *Ann N Y Acad Sci* **1067**, 37-44

40. Luo, S., and Levine, R. L. (2009) Methionine in proteins defends against oxidative stress. *FASEB J* **23**, 464-472

41. Ghesquière, B., Jonckheere, V., Colaert, N., Van Durme, J., Timmerman, E., Goethals, M., Schymkowitz, J., Rousseau, F., Vandekerckhove, J., and Gevaert, K. (2011) Redox proteomics of protein-bound methionine oxidation. *Mol Cell Proteomics* **10**, M110.006866

42. Schechter, Y., Burstein, Y., and Patchornik, A. (1975) Proceedings: Selective oxidation of methionine residues in proteins. *Isr J Med Sci* **11**, 1171

43. Shechter, Y., Burstein, Y., and Patchornik, A. (1975) Selective oxidation of methionine residues in proteins. *Biochemistry* **14**, 4497-4503

44. Levine, R. L., Berlett, B. S., Moskovitz, J., Mosoni, L., and Stadtman, E. R. (1999) Methionine residues may protect proteins from critical oxidative damage. *Mech Ageing Dev* **107**, 323-332

45. Levine, R. L., Moskovitz, J., and Stadtman, E. R. (2000) Oxidation of methionine in proteins: roles in antioxidant defense and cellular regulation. *IUBMB Life* **50**, 301-307

46. Friguet, B. (2006) Oxidized protein degradation and repair in ageing and oxidative stress. *FEBS Lett* **580**, 2910-2916

47. Grune, T., Reinheckel, T., and Davies, K. J. (1997) Degradation of oxidized proteins in mammalian cells. *FASEB J* **11**, 526-534

48. Stadtman, E. R., Moskovitz, J., and Levine, R. L. (2003) Oxidation of methionine residues of proteins: biological consequences. *Antioxid Redox Signal* **5**, 577-582

49. Fujino, T., Kojima, M., Beppu, M., Kikugawa, K., Yasuda, H., and Takahashi, K. (2000) Identification of the cleavage sites of oxidized protein that are susceptible to oxidized protein hydrolase (OPH) in the primary and tertiary structures of the protein. *J Biochem* **127**, 1087-1093

50. Gebendorfer, K. M., Drazic, A., Le, Y., Gundlach, J., Bepperling, A., Kastenmüller, A., Ganzinger, K. A., Braun, N., Franzmann, T. M., and Winter, J. (2012) Identification of a hypochlorite-specific transcription factor from Escherichia coli. *J Biol Chem* **287**, 6892-6903

51. Pennington, S. M., Klutho, P. R., Xie, L., Broadhurst, K., Koval, O. M., McCormick, M. L., Spitz, D. R., and Grumbach, I. M. (2018) Defective protein repair under methionine sulfoxide A deletion drives autophagy and ARE-dependent gene transcription. *Redox Biol* **16**, 401-413

52. Madian, A. G., and Regnier, F. E. (2010) Profiling carbonylated proteins in human plasma. *J Proteome Res* **9**, 1330-1343

53. Hsieh, Y. J., Chien, K. Y., Yang, I. F., Lee, I. N., Wu, C. C., Huang, T. Y., and Yu, J. S. (2017) Oxidation of protein-bound methionine in Photofrin-photodynamic therapy-treated human tumor cells explored by methionine-containing peptide enrichment and quantitative proteomics approach. *Sci Rep* **7**, 1370

54. Madeira, J. P., Alpha-Bazin, B., Armengaud, J., and Duport, C. (2018) Time-course proteomics dataset to monitor protein-bound methionine oxidation. *Data Brief* **18**, 394-398

55. Bourdon, E., Loreau, N., Lagrost, L., and Blache, D. (2005) Differential effects of cysteine and methionine residues in the antioxidant activity of human serum albumin. *Free Radic Res* **39**, 15-20

56. Grigoryan, H., Li, H., Iavarone, A. T., Williams, E. R., and Rappaport, S. M. (2012) Cys34 adducts of reactive oxygen species in human serum albumin. *Chem Res Toxicol* **25**, 1633-1642

57. Turell, L., Radi, R., and Alvarez, B. (2013) The thiol pool in human plasma: the central contribution of albumin to redox processes. *Free Radic Biol Med* **65**, 244-253

58. Matsuyama, Y., Terawaki, H., Terada, T., and Era, S. (2009) Albumin thiol oxidation and serum protein carbonyl formation are progressively enhanced with advancing stages of chronic kidney disease. *Clin Exp Nephrol* **13**, 308-315

59. Kovàcs, A., and Guttman, A. (2013) Medicinal chemistry meets proteomics: fractionation of the human plasma proteome. *Curr Med Chem* **20**, 483-490

60. Anraku, M., Yamasaki, K., Maruyama, T., Kragh-Hansen, U., and Otagiri, M. (2001) Effect of oxidative stress on the structure and function of human serum albumin. *Pharm Res* **18**, 632-639

61. Donadio, E., Piccolomini, F., Dimuccio, V., Felicioli, A., Balestreri, E., Cianti, R., Armini, A., Bini, L., Felicioli, R., and Donadio, C. (2009) Serum albumin fragmentation in end-stage renal disease patients--a pilot study. *Clin Chem Lab Med* **47**, 1373-1379

62. Donadio, C., Tognotti, D., and Donadio, E. (2012) Albumin modification and fragmentation in renal disease. *Clin Chim Acta* **413**, 391-395

63. Roche, M., Rondeau, P., Singh, N. R., Tarnus, E., and Bourdon, E. (2008) The antioxidant properties of serum albumin. *FEBS Lett* **582**, 1783-1787

64. Hackett, T. L., Scarci, M., Zheng, L., Tan, W., Treasure, T., and Warner, J. A. (2010) Oxidative modification of albumin in the parenchymal lung tissue of current smokers with chronic obstructive pulmonary disease. *Respir Res* **11**, 180

65. Maciążek-Jurczyk, M., Szkudlarek, A., Chudzik, M., Pożycka, J., and Sułkowska, A. (2018) Alteration of human serum albumin binding properties induced by modifications: A review. *Spectrochim Acta A Mol Biomol Spectrosc* **188**, 675-683

66. Finch, J. W., Crouch, R. K., Knapp, D. R., and Schey, K. L. (1993) Mass spectrometric identification of modifications to human serum albumin treated with hydrogen peroxide. *Arch Biochem Biophys* **305**, 595-599

67. Kulikowska-Karpińska, E., Zdanowicz, M., and Gałażyn-Sidorczuk, M. (2017) [Estimation of mercury in the urine of cigarette smokers]. *Wiad Lek* **70**, 697-702

68. Yeh, C. C., Barr, R. G., Powell, C. A., Mesia-Vela, S., Wang, Y., Hamade, N. K., Austin, J. H., and Santella, R. M. (2008) No effect of cigarette smoking dose on oxidized plasma proteins. *Environ Res* **106**, 219-225

69. Pryor, W. A., Prier, D. G., and Church, D. F. (1983) Electron-spin resonance study of mainstream and sidestream cigarette smoke: nature of the free radicals in gas-phase smoke and in cigarette tar. *Environ Health Perspect* **47**, 345-355

70. Valavanidis, A., and Haralambous, E. (2001) A comparative study by electron paramagnetic resonance of free radical species in the mainstream and sidestream smoke of cigarettes with conventional acetate filters and 'bio-filters'. *Redox Rep* **6**, 161-171

71. Haniff, A. N., and Gam, L. H. (2016) Identification of urinary protein biomarkers for tobacco smoking. *Biotechnol Appl Biochem* **63**, 266-272

72. Reznick, A. Z., Cross, C. E., Hu, M. L., Suzuki, Y. J., Khwaja, S., Safadi, A., Motchnik, P. A., Packer, L., and Halliwell, B. (1992) Modification of plasma proteins by cigarette smoke as measured by protein carbonyl formation. *Biochem J* **286 (Pt 2)**, 607-611

73. D'Anna, C., Cigna, D., Costanzo, G., Bruno, A., Ferraro, M., Di Vincenzo, S., Bianchi, L., Bini, L., Gjomarkaj, M., and Pace, E. (2015) Cigarette smoke alters the proteomic profile of lung fibroblasts. *Mol Biosyst* **11**, 1644-1652

74. Kulikowska-Karpińska, E., and Czerw, K. (2015) [Estimation of 8-hydroxy-2'-deoxyguanosine (8-OHdG) concentration in the urine of cigarette smokers]. *Wiad Lek* **68**, 32-38

75. Witherell, H. L., Hiatt, R. A., Replogle, M., and Parsonnet, J. (1998) Helicobacter pylori infection and urinary excretion of 8-hydroxy-2-deoxyguanosine, an oxidative DNA adduct. *Cancer Epidemiol Biomarkers Prev* **7**, 91-96

76. Panda, K., Chattopadhyay, R., Chattopadhyay, D., and Chatterjee, I. B. (2001) Cigarette smoke-induced protein oxidation and proteolysis is exclusively caused by its tar phase: prevention by vitamin C. *Toxicol Lett* **123**, 21-32

77. Clerici, M., Colombo, G., Secundo, F., Gagliano, N., Colombo, R., Portinaro, N., Giustarini, D., Milzani, A., Rossi, R., and Dalle-Donne, I. (2014) Cigarette smoke induces alterations in the drug-binding properties of human serum albumin. *Blood Cells Mol Dis* **52**, 166-174

CHAPTER TWO

Separation and analysis methodologies to expand the urinary proteome and comparison of smokers and non-smokers

2.1 ABSTRACT

The urinary proteome is an attractive candidate for the discovery of biomarkers to diagnose and classify health conditions because of the non-invasive collection procedure. However, protocols developed for the analysis of the urinary proteome have limitations both in identification of proteins and eliminating false positives. Urine was collected from 40 adults (24 smokers and 16 non-smokers). SDS-PAGE gel electrophoresis was followed by tryptic digestion, and then separation and identification by HPLC-MS/MS. Correspondence with predicted retention times, MS/MS sequence information, and a minimum of two tryptic peptides were required to reduce false positives. Proteins were matched to 1023 UniProtKB identification numbers, 467 of which have not been previously reported in the human urinary proteome. Gene ontology showed many were originally localized in the extracellular compartment but large numbers also originated intracellularly. Two proteins, annexin A1 and fatty acid binding protein 5 (FABP5), had statistically significant, after correction for multiple hypothesis testing, higher levels in urine of females. While proteins were found more commonly or at higher levels in the smoking or non-smoking population, after multiple hypothesis testing correction, these differences are not statistically significant. This lack of significant smoking related differences is in contrast to several earlier, smaller proteomic studies.

2.2 INTRODUCTION

Identifying protein biomarkers specific for different diseases using biological fluids is of great interest. Due to the ease and non-invasive nature of urine collection, this body fluid provides an alternate to blood plasma as a potential source for disease diagnosis and monitoring (1-3). In addition to potentially monitoring general health conditions, urine may provide more specific information about kidney function and disease (4-7).

Protein excretion in the urine of healthy individuals is ~150 mg/day (8,9) but the high abundance proteins serum albumin and IgG can complicate analysis of lower abundance proteins. Several separation approaches have been employed in the field of urinary proteomics. Identification by mass spectrometry has advanced continuously and emerged as a prominent tool for proteomic bioanalysis (10), including application to the study of the urinary proteome (8,11-19). However, each new study finds previously unidentified proteins in urine, implying that more remain to be discovered.

The emergence and development of spectral library search and bioinformatics tools for complements MS-based proteomic studies is assisting proteomic efforts in identifying and quantifying proteins and their post-translational modifications in biological samples (20-22). Advanced software packages increase the confidence in protein identification through the use of several statistical methods.

Smokers suffer from many illnesses linked to smoking and several studies have examined the proteomes of smokers in comparison to non-smokers (23-37), usually in sputum or blood samples. To the best of our knowledge only two studies (38,39) have examined differences in abundance of proteins in the urine of smoker and non-smokers and that on limited numbers of proteins and

individuals. As part of a larger study to determine differences in post-translational modification of proteins in smokers and non-smokers, we felt it appropriate to apply the latest instrumentation and analysis techniques to examine the occurrence of proteins in the urine of these two populations.

2.3 MATERIALS AND METHODS

2.3.1 Sample collection and processing

The University of Arkansas Institutional Review Board approved all experimental protocols using human subjects. First or second morning urine samples of at least 100 mL were obtained from forty individuals (Table 1 and Supporting Information Table S1). Immediately after sample collection, one protease inhibitor cocktail tablet (Millipore Sigma, SIGMA*FAST* containing 4-(2-aminoethyl) benzenesulfonyl fluoride, *trans*-epoxysuccinyl-L-leucylamido (4-guanidino), butane, bestatin, leupeptin, aprotinin, and sodium EDTA) was added. Samples were stored at 4°C for no longer than 2 hours prior to centrifugation at 4°C for 20 minutes in a Sorvall GS3 rotor at 4000 rpm (1700 RCF) to remove particulates. One hundred ml of the supernatant was transferred to six 15 ml 2000 Da MWCO centrifugal concentrators (Sartorius Vivaspin) and sample size was reduced to approximately 1 ml in each concentrator by spinning at 4700 rpm (3700 RCF) in a Sorvall SH-3000 rotor for 6-8 hours at 4°C. The concentrates were combined into a single concentrator. Concentrators not used in the next step were rinsed with small volumes of 100mM ammonium bicarbonate (pH 7.8). After this rinse solution was added to the combined concentrate, additional 100mM ammonium bicarbonate was added to bring the total volume up to 15 ml and this was centrifuged as before for approximately 4 hours to reduce the final volume to 1.5 ml; 1.5% of the original sample volume. Early experiments showed a steady decrease in the amount of

protein visible on gels with storage at -80° for more than a few days, so this concentrate was stored only overnight at -80° C.

On the next day, 200 microliters of the concentrate were desalted using C18 spin columns (Pierce). C18 columns were first prepared according to manufacturer's directions by adding 200µL of activation solution (50% methanol). Columns were placed into a microcentrifuge (receiver) tube and centrifuged at 1500 × g for 1 minute. Activation was repeated then 200µL of equilibration solution (0.5% TFA in 5% acetonitrile (ACN)) was added and spun at 1500 × g for 1 minute. The equilibration step was repeated and then the concentrated urinary protein sample was loaded on top of the resin bed. Columns were placed into a clean microcentrifuge tubes and spun at 1500 × g for 1 minute. The flow-through was recovered and the binding step was repeated to ensure complete binding. 200µL of washing buffer (0.5% TFA in 5% ACN) was added to the columns and they were centrifuged at 1500 × g for 1 minute. The flow-through was discarded and washing was repeated. Finally, proteins were eluted with 20µL of 70% ACN and centrifuged at 1500 × g for 1 minute. The elution was repeated with another 20µL of 70% ACN and the combined eluent was flash frozen and dried under vacuum using a SpeedVac concentrator for 5-7 hours. The resulting protein pellets were dissolved in 300 µl of 100 mM ammonium bicarbonate (pH 7.8). In a few cases these solutions were run immediately on a gel, but usually were frozen at -80° until the next day

2.3.2 One-dimensional SDS-PAGE and in-gel digestion of urinary proteins

Dissolved pellets were mixed in a 2:1 ratio with loading buffer (100 mM TrisHCl, pH 6.8, 4% SDS, 20% glycerol, 200 mM β-mercaptoethanol, 0.2% bromophenol blue). The mixture then was heated to 100°C for 5 minutes to denature and reduce the proteins. Equal amounts of the mixture

were then loaded into each well of a pre-cast tris-glycine, 4-20% acrylamide gradient gel (NuSep) and separated by SDS-PAGE gel electrophoresis in a BioRad MiniProtean Tetra at 150 volts for 1.5 to 2.0 hours until the tracking dye reached the bottom of the gel. The gel was stained overnight with Coomassie dye prepped by dissolving 0.25 g of Coomassie brilliant blue R-250 in 90 ml of methanol: 10 ml of H_2O and glacial acetic acid (1:1 v/v). The next day, after excess stain in the gel was removed by destaining briefly in methanol/water/acetic acid, the gel was cut into five bands using a clean razor blade as illustrated in Figure 1. Each band was further cut into 1 × 1mm pieces and transferred to clean microcentrifuge tubes. Most bands would fit in a single tube, but the larger D band and, occasionally, the C band were split into two or three microcentrifuge tubes because of larger gel volumes. Sufficient 50% 100mM ammonium bicarbonate/50% acetonitrile (ACN) was added to barely submerge the gel slices then incubated at 37°C for 30 minutes to remove the Coomassie stain bound to protein from the gel slices. The liquid was removed and destaining was repeated, a total of 3-7 cycles, until stain was completely removed.

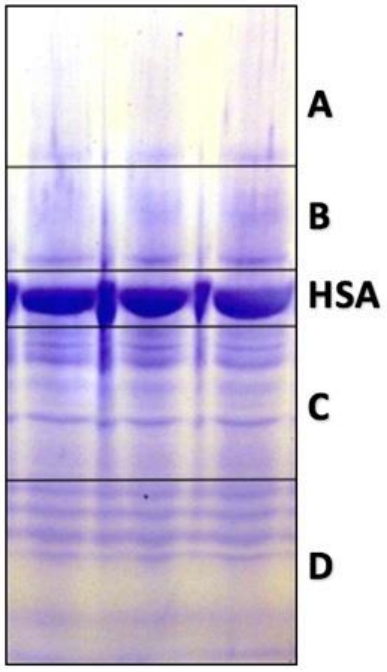


Figure 1. Urinary protein separation by SDS-PAGE. 200 uL of concentrate was divided across all wells of a BIS-TRIS 4-20% gel. After electrophoresis, the gel was stained with Coomassie blue, cut into segments as shown in the figure, destained, reduced and alkylated, and then each segment was separately digested with trypsin.

2.3.3 Reduction and Alkylation

Enough fresh *tris*-2-carboxyethylphosphine (TCEP) solution (5mM in 100mM ammonium bicarbonate, pH 7.8) was added to cover the destained gel slices which were then incubated at 95°C for 10 minutes. The TCEP solution then was pipetted out and discarded. Freshly prepared 100 mM iodoacetamide (IAA) in 100 mM ammonium bicarbonate solution was added so that the gel slices were completely covered and then incubated at 37°C for 15 minutes with shaking. The IAA solution was pipetted from the gel slices and discarded. Gel slices were rinsed with 50% 100mM ammonium bicarbonate/50% ACN and incubated at 37°C for 15 minutes with shaking and the rinse solution was pipetted away. The rinse and removal were then repeated.

2.3.4 Digestion

Gel pieces were immediately reduced in size after reduction and alkylation by adding 50µL 99.9% ACN and incubating for 15 minutes at room temperature. ACN then was removed by pipetting and gel pieces were left to air dry for 5-10 minutes in the open microfuge tube.

A stock solution was made from freeze dried porcine mass spectrometry grade trypsin (G Biosciences) in 100 mM ammonium bicarbonate (pH 7.8) at a concentration of 1 mg/ml. This stock was immediately diluted 100-fold in 100mM ammonium bicarbonate to a final trypsin concentration of 0.01 mg/ml. Aliquots of the final 0.01 mg/ml trypsin solution were either used immediately or stored at -80° for no more than one month before use.

The 0.01 mg/ml trypsin solution was added until the dried gel slices were covered. Tubes were incubated at 37°C for 24 hours. The digest solution was pipetted to a new microcentrifuge tube. Gel pieces were then extracted three times by adding 50µL of 50% ACN/50% water, 0.1% trifluoroacetic acid (TFA) and incubation at 37°C for 5-15 minutes. This extraction liquid was

combined with the digest solution. The combined solution was frozen with liquid nitrogen and the volatiles removed under vacuum over 2-3 hours using a Speed-Vac concentrator to produce a peptide pellet.

2.3.5 Peptide desalting

Each pellet was dissolved in 50 µl of 2% TFA in 20% ACN. If the volume of the pieces of a particular gel band required splitting it into to multiple tubes, the samples were combined at this step by using the same 50 µl to sequentially dissolve the pellets. Sample was desalted using C18 spin columns as described above. After the eluent was vacuum dried, peptide pellets were dissolved in 0.1% formic acid (FA) and stored at -80°C until LC-MS/MS.

2.3.6 Analysis of the digest via ESI-LC-MS/MS

Digests were analyzed by ESI-LC-MS/MS. Data dependent analysis (DDA) for the in-gel trypsin digested samples were performed by using an Agilent 1200 series micro flow HPLC in line with Bruker Amazon-SL quadrupole ion trap ESI mass spectrometer (QIT-ESI-MS). Tryptic peptides were separated by reverse-phase high-performance liquid chromatography (RP-HPLC) using a Zorbax SB C18 column, (150 × 0.3 mm, 3.5 µm particle size, 300 Å pore size, Agilent Technologies), with a solvent flow rate of 4 µL/minute, and a gradient of 5%–38% consisting of 0.1% FA (solvent A) and ACN (solvent B) over a time period of 320 minutes. As indicated in Table S1 (Supporting Information), some fractions were run multiple times. All the ESI-MS analyses were performed in a positive ion mode using Bruker captive electrospray source with a dry nitrogen gas temperature of 200°C, with nitrogen flow rate of 3 L/minute. LC-MS/MS data were acquired in the Auto MS(n) mode with optimized trapping condition for the ions at m/z 1000.

MS scans were performed in the enhanced scanning mode (8100 m/z/second), while the collision-induced dissociation or the MS/MS fragmentation scans were performed automatically for top ten precursor ions with a set threshold for one minute in the UltraScan mode (32,500 m/z/second).

2.3.7 Data analysis

Protein identifications were performed via MASCOT (Matrix Science, London, UK) (40) searching all mass spectra data against human protein entries obtained from UniProt (41). Scaffold (version Scaffold_4.8.4, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 80.0% probability to achieve an FDR less than 1.0% by the Peptide Prophet algorithm (42). Protein identifications were accepted if they could be established at greater than 5.0% probability to achieve an FDR less than 1.0% and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (43). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. Proteins were annotated with GO terms from goa.gaf (downloaded Sept. 7, 2017). (44). Protein levels in populations (smoker vs. non-smoker, heavy smoker vs. non-smoker, and female vs. male) were compared using the spectral counting method in Scaffold, normalizing protein spectral counts within different spectra for each individual.

2.4 RESULTS AND DISCUSSION

Table 1. Distribution of urine donors by sex, smoking status, and age group[a]

Sex	Group	Age range	Number
Female	Light smoker	30-40	2
		40-50	1
	Heavy smoker	20-30	1
		30-40	3
		60-70	1
	Non-smoker	18-20	2
		30-40	4
		40-50	2
Male	Light smoker	18-20	1
		20-30	3
		30-40	5
	Heavy smoker	20-30	1
		30-40	4
		50-60	1
		60-70	1
	Non-smoker	18-20	2
		20-30	3
		30-40	3

[a]Heavy smokers are those currently smoking 5 or more cigarettes a day, light smokers are those who are currently smoking at least once a week, but less than 5 cigarettes a day, non-smokers are those who have never smoked or not smoked in over 18 months. Sample donors are grouped by their age in years at time of sample collection. Number is the number of individuals in the study who meet the criteria to the left in the table.

Slightly over one thousand UniProtAC IDs were identified in urine samples from 24 smokers and 16 non-smokers. The distribution of sample donors by gender, smoking status, and age range is shown in Table 1 with more detailed information found in Table S1. These 1023 proteins were identified by Scaffold in the UniProt database with high confidence as a minimum of two tryptic peptides from each protein, with one or more being unique to that protein, was required, falsely identified proteins were reduced by performing a decoy database search, as described elsewhere (45), and the strictest threshold, a 1% False Discovery Rate, available in Scaffold was adopted. The Scaffold-assigned probability of identification was 95% or higher for 788 UniProtACs

(77.03%). Only two peptides were detected for just 61 proteins, and just three peptides were found for 20 more. The remaining 942 identifiers each had four or more peptides recognized. The complete list of 1023 UniProtACs is found in Table S2 (Supporting Information). Of these 1023 accession (UniProtAC) numbers, 1002 were successfully mapped to UniProt Knowledge Base (UniProtKB) identifiers. Six of the unmapped UniProtAC entries were fragments of immunoglobin variable regions and the remaining three unmapped were merged with other UniProtKB entries. Eleven were mapped to obsolete entries. Of the 991 active UniProtKB entries, 754 were mapped to 623 genes. This latter number, in many respects, may be the most reliable estimate of unique, non-redundant gene products. Of the 991 active UniProtKB IDs, 431 were reviewed and 560 are currently unreviewed. Once our data was mapped to UniProtKB there were, for example, eight IDs for unreviewed entries for serum albumin in addition to the one reviewed entry. Two of the unreviewed entries map to the ALB gene while the others are not mapped to any gene. Presumably all nine entries in this case represent a single protein product. However, this case is an extreme and not all unreviewed entries are duplicates of a reviewed entry. Scaffold clusters related proteins, in the case of the data here resulting in 467 groups. However, in some cases, the program is combining closely related, but distinct proteins, such as the lactate dehydrogenase A and B chains, into a single cluster. The lack of automated tools to reliably remove redundant entries without dropping non-redundant entries led us to use the complete list for most of the following analysis. The presence of redundant entries has no effect on our major conclusions.

The identified proteins were subjected to Gene Ontology (GO) analysis (46,47) in Scaffold to determine typical subcellular localizations. Proteins normally localized in the extracellular region were 37.7% of those found. However intracellular proteins were well represented with cytoplasmic proteins being 28.1% of those detected and proteins originating in intracellular organelles were 26.6%. Proteins localized to membranes comprised 20.6% of those identified. Gene ontology molecular function terms showed that 31.9% of the proteins detected had some sort of a binding function. Proteins with catalytic activity were also abundant (14.7%). Thirty-six uncharacterized (putative) proteins were observed in the current study.

The concentration and purification techniques were chosen to result in intact or large fragments of proteins being identified. The smallest protein chains found were the light chain of inter-alpha-trypsin inhibitor at 5.7 kDa and beta defensin 1, which after processing is just 3.9 kDa. The largest proteins found were titin, with a molecule weight as high as 3,816 kDa and mucins 5AC and 5B at 583 kDa and 595 kDa respectively, without considering any glycosylation. For larger proteins, the percent coverage was, not surprisingly, often low so these may represent the detection of large fragments of very large parent proteins. The median size of the parent proteins detected was approximately 45 kDa. These sizes are predicted for the intact protein and the methods employed may not allow separation and identification of isoforms such as transcriptional splicing variants or post-translational proteolytic cleavages.

Eighteen proteins were detected in the urine of all individuals and 74 proteins were found in at least 90% (≥36) of the individuals. Approximately one half of the proteins were found in ten or more individuals. Others were found in relatively few individuals. Approximately one fifth of the proteins, while scattered throughout the various study participants, were detected in just a single person. Most of these are proteins that are presumably present in more individuals and could be

detected if targeted, but low abundance prevented more frequent detection in the broad survey undertaken here.

Other studies (8,11-19) have previously explored the human urinary proteome and it might be wondered whether anything new remains to be learned. These studies are summarized in Table S3 (Supporting Information) and, when converted to UniProt and duplicates are eliminated, 11,551 unique UniProtKB ACs have been previously reported in human urine. Those 11,458 mapped to 11,473 UniProtKB IDs, of which 8,561 were reviewed and 2,419 were unreviewed. The majority, 10,481, of the 11,473 identifiers were successfully mapped to 8,867 genes. While many of the previous studies have detected more proteins overall than the results presented here, the present study is among the largest yet performed in terms of the number of individuals examined. The fact that one fifth of the UniProtKB IDs identified in the work presented here were each found in only a single individual implies that a complete description of the low abundance or rare proteins that can be found in human urine may still be incomplete. Indeed, this proved to be the case. Out of the total of 1023 different UniProtKB IDs reported here, 466 (45.6%) have not been previously reported in other proteomic studies of human urine. Of the 431 proteins that were identified in the current study and that are reviewed entries in UniProt, just eight had not been previously reported. However, of the 623 genes that entries found in the present study were linked to, 548 genes were among the 8,867 genes linked to entries that have been previously reported in urinary proteomic studies. Therefore, 75, or about one sixth of all the protein gene products we found, are newly reported here. These proteins are listed in Table S4 (Supporting Information). Roughly half, 33, of these gene products are identified by Scaffold with a probability of greater than 95% and most are found in multiple individuals. The fact that the vast majority of these novel gene are not curated

entries in UniProt lends credence to the general idea that these are low abundance proteins for which comparatively little information about physiological roles is available.

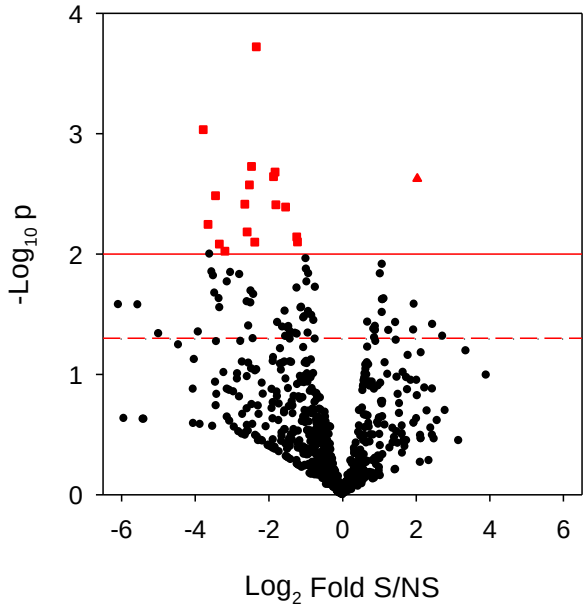
2.4.1 Comparison of protein levels and occurrence in smokers and non-smokers

Several studies have linked changes in levels of specific protein expression in cigarette smokers (48-51). However, little is known about the impact cigarette smoke may have on the level of plasma proteins, the immediate source of most proteins found in urine, more generally. It is no surprise that the total concentration of protein in urine varied widely. The total amount of protein was estimated using eight pictures of dyed gel smokers and a similar number for non-smokers. The total intensity was measured using the ImageJ application developed by NIH. The average ratio of dyed protein in smokers compared to non-smokers revealed an average doubled intensity of urine proteome in smokers (S/NS = 2.1) but with a standard deviation of nearly the same magnitude (1.8) these differences are clearly not statistically significant. We also compared the total mass spectrometry peak intensities of the entire proteome for all three replicates of each of the forty individuals used in the study. The ratio of the average intensities for smokers to non-smokers was 2.6 but the standard deviation was 6.4, so again the difference in average total protein between the two populations is not distinguishable from that expected by chance.

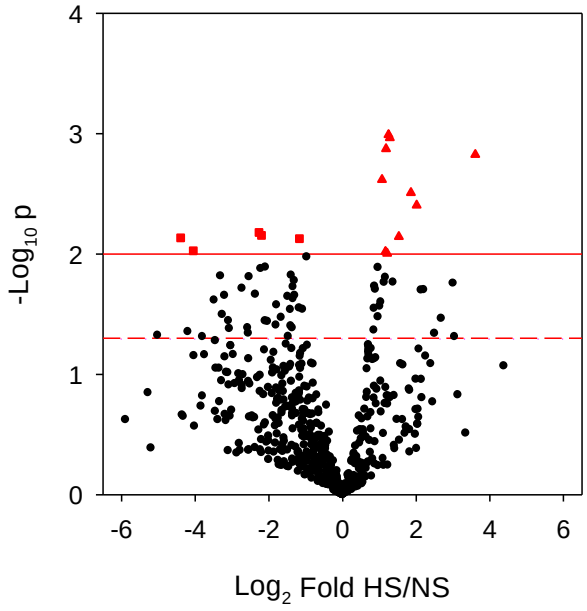
These results give confidence that any variations in the level of any given protein that might occur between two populations are not likely due to variation in overall urine protein levels between those two populations. The Scaffold software suite was used to determine differences between populations in the observed levels of each particular protein. Because of the large variation from sample to sample in total protein levels noted above, the level of each protein in each individual was normalized in Scaffold by the total spectrum count for that individual. Figure 2A is a volcano plot, showing the differential abundance (fold change) of each protein in the urine

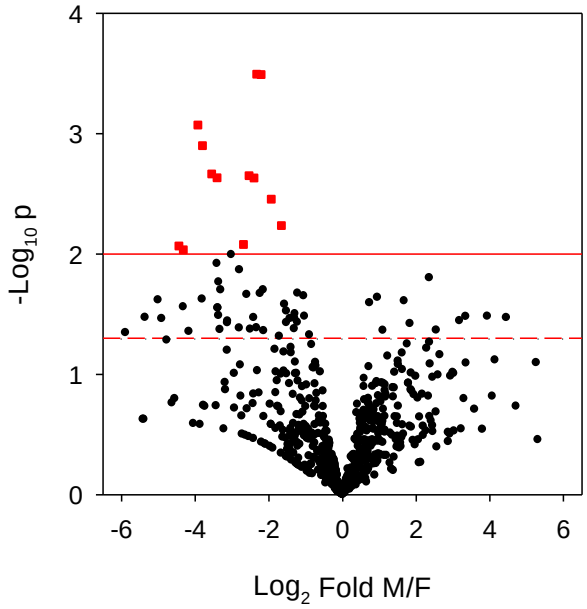
of smoking and non-smoking populations versus the t-test probability for that protein. Figure 2B is similar but compares non-smokers to the twelve individuals self-reporting as smoking five or more cigarettes a day (heavy smokers). As can be seen in both plots a number of proteins have a p of less than 0.05, often taken as a cutoff for statistical significance. In Tables 2 and 3, a summary of the proteins meeting the more stringent cutoff of $p \leq 0.01$ in Figure 2A and 2B, respectively, is given. Of course, when examining over one thousand UniProtKD IDs as we do here, one would expect approximately 10 proteins in a randomly distributed sample to have $p \leq 0.01$ and 50 or so with $p \leq 0.05$. So, not surprisingly, after applying the Benjamini-Hochberg multiple hypothesis testing correction (52), none of the proteins that are found at higher levels in either smokers versus non-smokers or heavy smokers versus non-smokers rises to the level of statistical significance, after correction, at $p \leq 0.05$ or even if the criteria for p is relaxed to ≤ 0.10.

A.


4
3
2
1
0
-Log10 p
-6
-4
-2
0
2
4
6
Log2 Fold S/NS

B.


4
3
2
1
0
-Log10 p
-6
-4
-2
0
2
4
6
Log2 Fold HS/NS



C.

Figure 2. Volcano plots of $-Log_{10}$ p versus the fold change of protein level between the two populations. In 2A the populations are smokers (S) and non-smokers (NS), in 2B heavy smokers (HS, more than five cigarettes a day) are compared to non-smokers (NS), and in 2C males (M) and females (F) are compared. The dashed red line corresponds to a p value of 0.05 and the solid red line to a p value of 0.01. UniProtKB IDs with $p \leq 0.01$ are shown in red and IDs found elevated in NS or F are squares, while those that are elevated in S or HS are triangles.

Table 2. Ratio of Protein Levels in Smoking and Non-Smoking Populations in Order of Statistical Significance[a]

UniProt Description	UniProt Acc. No.	p	Fold S/NS
High in NS, low in S			
SUN domain-containing protein 1 GN=SUN1	SUN1	0.00019	0.20
Keratin, type I cytoskeletal 13 GN=KRT13	K1C13	0.00092	0.07
cDNA FLJ58780, highly similar to lysosomal-associated membrane protein 2 (LAMP2B)	B4E2S7	0.00186	0.18
Beta-actin-like protein 2 GN=ACTBL2	ACTBL	0.00208	0.28
Putative beta-actin-like protein 3 GN=POTEKP	ACTBM	0.00227	0.27
Betaine-homocysteine S-methyltransferase 1 GN=BHMT	BHMT1	0.00266	0.17
cDNA FLJ78503, highly similar to keratin 13 (KRT13)	A8K2H9	0.00327	0.09
Keratin, type II cytoskeletal 3 GN=KRT3	K2C3	0.00384	0.16
L-lactate dehydrogenase B chain GN=LDHB	LDHB	0.00389	0.28
Keratin, type II cytoskeletal 6B GN=KRT6B	K2C6B	0.00406	0.34
Keratin, type II cytoskeletal 4 GN=KRT4	K2C4	0.00566	0.08
cDNA FLJ54604, highly similar to Betaine-homocysteine S-methyltransferase	B4DPF0	0.00653	0.17
Keratin, type II cytoskeletal 2 epidermal GN=KRT2	K22E	0.00719	0.42
Plasma serine protease inhibitor GN=SERPINA5	IPSP	0.00791	0.43
Prosaposin (Variant Gaucher disease and variant metachromatic leukodystrophy)	Q53FJ5	0.00794	0.19
Keratin, type II cytoskeletal 73 GN=KRT73	K2C73	0.00824	0.10
Phosphoglycerate kinase 1 GN=PGK1	PGK1	0.00944	ND in S
cDNA FLJ54596, highly similar to Proactivator polypeptide	B4DEK5	0.00945	0.11
High in S, low in NS			
Putative uncharacterized protein DKFZp686C02220 (Fragment) GN=DKFZp686C02220	Q6N091	0.00236	3.4

[a]Smokers (S) are those who are currently smoking at least once a week, non-smokers (NS) are those who have never smoked (n = 13) or not smoked in over eighteen months (n = 3). If the UniProtKB entry mapped to a gene, the gene ID is given as well. ND means not detected in the indicated population.

Table 3. Ratio of Protein Levels in Heavy-Smoking and Non-Smoking Populations in Order of Statistical Significance[a]

UniProt Description	UniProt Acc. No.	p	Fold HS/NS
High in NS, low in HS			
SUN domain-containing protein 1 GN=SUN1	SUN1	0.00662	0.21
Plasma serine protease inhibitor GN=SERPINA5	IPSP	0.00697	0.22
Keratin, type II cytoskeletal 4 GN=KRT4	K2C4	0.00731	0.05
Deoxyribonuclease-1 GN=DNASE1	DNAS1	0.00745	0.44
Betaine-homocysteine S-methyltransferase 1 GN=BHMT	BHMT1	0.00936	0.06
High in HS, low in NS			
Putative uncharacterized protein DKFZp686G21220 (Fragment) GN=DKFZp686G21220	Q6N090	0.00022	ND in NS
Putative uncharacterized protein DKFZp686L19235 GN=DKFZp686L19235	Q6MZV6	0.00102	2.4
IGH@ protein GN=IGH@	Q6P089	0.00108	2.4
cDNA FLJ41981 fis, clone SMINT2011888, highly similar to Protein Tro alpha1 H,myeloma	Q6ZVX0	0.00133	2.3
Putative uncharacterized protein DKFZp686J11235 (Fragment) GN=DKFZp686J11235	Q6MZW0	0.00149	12.2
Putative uncharacterized protein DKFZp686M08189 GN=DKFZp686M08189	Q6MZX9	0.00241	2.1
cDNA FLJ55716, highly similar to Desmocollin-2	B4DLJ5	0.00309	3.6
Complement decay-accelerating factor GN=CD55	H7BY55	0.00393	4.0
Desmocollin-2 GN=DSC2	DSC2	0.00717	2.9
Myosin-reactive immunoglobulin light chain variable region (Fragment)	Q9UL83	0.00955	2.2
Tyrosine-protein phosphatase non-receptor type substrate 1 GN=SIRPA	SHPS1	0.00988	2.3

[a]Heavy smokers (HS) are those who were currently smoking at least five cigarettes a day, non-smokers (NS) are those who have never smoked or not smoked in over eighteen months. If the UniProtKB entry mapped to a gene, the gene ID is given as well. ND means not detected in the indicated population.

Two studies have previously examined the impact of cigarette smoking upon the urinary proteome. In the first published, Airoldi *et al.* (38) examined the urine pooled from ten non-smokers and pooled from ten smokers, all male, using three replicate 2D electrophoresis gels for each pool to identify 165 spots present in all gels, followed by LC-MS/MS of the tryptic peptides of 6 spots that were significantly different in image analysis of the smoking and non-smoking gels. Pancreatic alpha-amylase, found in two distinct spots, S100A8, CD59, and inter-alpha-trypsin inhibitor heavy chain H4 were more abundant in smokers. One protein, zinc-alpha-2-glycoprotein, was the only protein less abundant in smokers. Haniff and Gam (39) more recently looked for differences in band intensities in 1D protein gels between eight smokers, eight passive smokers, and 8 non-smokers. They identified two bands that differed consistently. Using tryptic digests and MS/MS, they identified four proteins unique to smokers: pancreatic alpha-amylase, proepidermal growth factor, protein 4.1, and prostatic acid phosphatase. Further one protein, plasma serine protease inhibitor, was found in both populations in a 4.879 S/NS ratio.

While we did not observe either pancreatic alpha-amylase or protein 4.1, the others were present in the samples. The p values and ratios found in the present study for these proteins flagged in previous studies are summarized for non-smokers versus all smokers or heavy smokers in Table 4. In every case, and in contrast to the earlier work for all but one protein, we observed that the proteins in question were more abundant in non-smokers. The exception is zinc-alpha-2-glycoprotein which was the sole protein previously found less abundant in smokers. While our data agree with this, we note that the ratio found in smoker and non-smokers is near one and the statistical significance very low indeed. While plasma serine protease inhibitor is the lowest p value found for these proteins in our study, it still does not rise to the level of statistical significance once correction for multiple hypothesis testing is made.

Interestingly, a study of epithelial lining fluid showed decreased levels of S100A8 in acute response to smoking (24), a result in agreement with ours. A study of the human placental proteome response to maternal smoking (29) also had a number of differences in protein levels between smokers and non-smokers that were in qualitatively similar to those we observed in urine, although most were not of particular statistical significance in our data. On the other hand, in studies of proteomic changes in response to smoke exposure in platelets (26) or in plasma (25,30,53,54), relatively few of the proteins identified as showing significant changes in level between populations were detected in the present study, but none of those that were detected showed as statistically significant.

This is not to imply that previous work is unreliable. Indeed, one of these studies of plasma is particularly noteworthy for examining the plasma of 1,686 study participants, and having great statistical power (53). Rather, we believe that while different levels of various proteins may well be found in the urine of smokers and non-smokers, the high variability of levels of each urinary protein from sample to sample and the relatively small differences in protein levels between smokers and non-smokers (55,56) mean that a very large number of samples will be needed to find and confirm with any confidence that there are meaningful differences.

Table 4. Ratios of Urinary Protein Levels Between Populations in Present Study for Proteins Previously Discussed in Literature[a]

UniProt Description	UniProt Acc. No.	p	Fold S/NS	p	Fold HS/NS
Airoldi et al.					
CD59 glycoprotein GN=CD59	CD59	0.04016	0.36	0.06591	0.28
Inter-alpha-trypsin inhibitor heavy chain H4 GN=ITIH4	ITIH4	0.15690	0.66	0.08926	0.45
Protein S100-A8 GN=S100A8	S10A8	0.86680	0.91	0.91955	0.94
Zinc-alpha-2-glycoprotein GN=AZGP1	ZA2G	0.90871	0.98	0.63374	0.90
Haniff & Gam					
Plasma serine protease inhibitor GN=SERPINA5	IPSP	0.00791	0.43	0.00697	0.22
Pro-epidermal growth factor GN=EGF	EGF	0.24304	0.71	0.15999	0.57
Prostatic acid phosphatase ACPP	PPAP	0.28859	0.61	0.09216	0.29

[a]Proteins in this table are those previously identified by either Airoldi *et al.* (38) or Haniff & Gam (39) as being significantly elevated or depressed in smokers relative to non-smokers and also found in our samples. The ratio of protein levels found in smokers or heavy smokers relative to non-smokers and the p value are for the present study

2.4.2 Comparison of protein levels and occurrence in females and males

In contrast, in the comparison of protein levels found in females and males, Figure 2C and Table 5, after correcting for multiple hypothesis testing, three identifiers in the UniProt database are flagged as being statistically significant at $p \leq 0.05$. Although we list the three identifiers returned from Scaffold's search of UniProt, for reasons explained in more detail in the footnote below, there is strong reason to believe that there are really just two proteins due to the nature of the UniProt database. Those two proteins are annexin A1 and fatty acid binding protein 5 (FABP5).

Guo et al. (55) previously found that male and female urinary proteomes were distinguishable. They examined 7 males and 7 females and listed the 20 most significantly different proteins. We detected six of those proteins and three showed the same qualitative reduction or elevation in one gender, while the other three did not. Two of the proteins that qualitatively agreed were prostatic acid phosphatase and prostate-specific antigen, unsurprisingly elevated in males in both the earlier and present study. However, the M/F ratio of 2.23 for prostatic acid phosphatase and 1.94 prostate-specific antigen we observed did not have particularly high p values, 0.136 and 0.360 respectively. The fact that proteins that there is very good reason to believe are elevated in males do not show as statistically significant shows, again, that it is likely that there are real differences between levels of other proteins in the female/male and non-smoker/smoker populations that are masked statistically by the inherent noise in the data. On the other hand, the protein of the six that we found the most statistically significant was myeloperoxidase with a M/F ratio of 0.056 and $p = 0.0044$. In the previous study, the M/F ratio was 5.81. Again, we interpret

this to mean that highlighting as significant any difference in urinary protein levels between populations should only be done with great caution, especially when testing large numbers of

proteins. This interpretation is buttressed by the fact that few of the proteins identified as specific to the pooled urine of 20 men or 20 women by Oh *et al.* (57) were detected in the present study, and none of those that were detected here had statistically significant different levels in males and females.

That caution may even extend to those proteins whose difference in level are flagged as significant even after correction for multiple hypothesis testing. The present study finds FABP5 as having the most significant difference between males and females. There is, to the best of our knowledge, no data about differences in the levels in human urine for this protein. However, a study of human serum, the penultimate stop for proteins before urinary secretion, found no significant difference in FABP5 levels between men and women (58).

The other protein flagged as significantly different, after multiple hypothesis correction, was annexin A1. Other work supports the idea that this may be a real difference. Sun *et al.* (59) examined the urine of 3 men and 3 women and found annexin A1 in none of the men but only in the urine of a single female. Although in the blood rather than urine, premenopausal women expressed higher levels of annexin A1 on the surface of their circulating polymorphonuclear cells (60) than did men. Similarly, annexin A1 was identified as being significantly higher in human female aminocytes compared to human male aminocytes (61).

Table 5. Ratio of Protein Levels in Female and Male Populations in Order of Statistical Significance[a]

UniProt Description	UniProt Acc. No.	p	Fold F/M
Fatty acid binding protein 5 (Psoriasis-associated) GN=FABP5	E7DVW5	0.00001	69.7
Annexin GN=ANXA1	B5BU38	0.00002	12.6
Annexin (Fragment) GN=ANXA1	Q5T3N1	0.00003	21.7
Beta actin variant (Fragment)	Q53G99	0.00032	5.0
Actin-like protein (Fragment) GN=ACT	Q562Z4	0.00032	4.6
cDNA FLJ52761, highly similar to Actin, aortic smooth muscle	B4DUI8	0.00085	15.2
cDNA FLJ42347, highly similar to Actin, gamma-enteric smooth muscle	B3KW67	0.00125	14.0
cDNA FLJ57283, highly similar to Actin, cytoplasmic 2	B4E3A4	0.00215	11.7
Glyceraldehyde-3-phosphate dehydrogenase GN=GAPDH	G3P	0.00223	5.8
Annexin GN=ANXA2	A0A024R5Z7	0.00232	10.6
Protein S100-A9 GN=S100A9	S10A9	0.00233	5.3
Alpha-enolase GN=ENO1	ENOA	0.00350	3.8
Epididymis luminal protein 176 GN=HEL-176	V9HVZ7	0.00578	3.2
Pyruvate kinase PKM GN=PKM	KPYM	0.00829	6.4
Cystatin-B GN=CSTB	CYTB	0.00856	21.8
cDNA FLJ53963, highly similar to Leukocyte elastase inhibitor	B4E3A8	0.00922	20.1
Histone H2A.V GN=H2AFV	H2AV	0.00979	ND in M

[a]The three identifiers that remain significant at the $p \leq 0.05$ level after correction for multiple hypothesis testing are italicized. If the UniProtKB entry mapped to a gene, the gene ID is given as well. ND means not detected in the indicated population.

2.5 CONCLUSIONS

Our strategy in processing and analysis of urine samples resulted in a set of proteins among which about 20% were identified for the first time. More low abundance proteins likely remain to be identified in human urine. In contrast to earlier work, no statistically significant differences were found in the levels of proteins in the urine of non-smokers compared to either smokers or just heavy smokers. Indeed, in many cases the ratios found here were qualitatively inverted relative to earlier work. However, annexin A1 and fatty acid binding protein 5 (FABP5) both are significantly elevated in females compared to males. While other differences in protein abundance between these various populations may exist, the degree of difference is low enough that studies of larger numbers of individuals and greater statistical power will be required to tease them out. There are undoubted advantages to the collection of urine compared to other sources of human protein, but it seems doubtful that variations of specific protein levels in urine are likely to be tremendously useful as a clinical tool because of the statistical noise. Even for research purposes, using protein sources with less inherent individual variability will often make up for the collection difficulties if one is interested in physiologically or disease linked changes in protein levels.

2.6 REFERENCES

1. Pisitkun, T., Shen, R. F., and Knepper, M. A. (2004) Identification and proteomic profiling of exosomes in human urine. *Proc Natl Acad Sci U S A* **101**, 13368-13373

2. Decramer, S., Gonzalez de Peredo, A., Breuil, B., Mischak, H., Monsarrat, B., Bascands, J. L., and Schanstra, J. P. (2008) Urine in clinical proteomics. *Mol Cell Proteomics* **7**, 1850-1862

3. Wu, J., and Gao, Y. (2015) Physiological conditions can be reflected in human urine proteome and metabolome. *Expert Rev Proteomics* **12**, 623-636

4. Siwy, J., Zürbig, P., Argiles, A., Beige, J., Haubitz, M., Jankowski, J., Julian, B. A., Linde, P. G., Marx, D., Mischak, H., Mullen, W., Novak, J., Ortiz, A., Persson, F., Pontillo, C., Rossing, P., Rupprecht, H., Schanstra, J. P., Vlahou, A., and Vanholder, R. (2017) Noninvasive diagnosis of chronic kidney diseases using urinary proteome analysis. *Nephrol Dial Transplant* **32**, 2079-2089

5. Samavat, S., Kalantari, S., Nafar, M., Rutishauser, D., Rezaei-Tavirani, M., Parvin, M., and Zubarev, R. A. (2015) Diagnostic urinary proteome profile for immunoglobulin a nephropathy. *Iran J Kidney Dis* **9**, 239-248

6. Lv, Y., Cai, G., and Chen, X. (2015) Applications of urinary proteomics in renal disease research using animal models. *Adv Exp Med Biol* **845**, 145-150

7. Mischak, H. (2015) Pro: urine proteomics as a liquid kidney biopsy: no more kidney punctures! *Nephrol Dial Transplant* **30**, 532-537

8. Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V., and Mann, M. (2006) The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol* **7**, R80

9. Siwy, J., Mullen, W., Golovko, I., Franke, J., and Zürbig, P. (2011) Human urinary peptide database for multiple disease biomarker discovery. *Proteomics Clin Appl* **5**, 367-374

10. Norden, A. G., Lapsley, M., Lee, P. J., Pusey, C. D., Scheinman, S. J., Tam, F. W., Thakker, R. V., Unwin, R. J., and Wrong, O. (2001) Glomerular protein sieving and implications for renal failure in Fanconi syndrome. *Kidney Int* **60**, 1885-1892

11. Kentsis, A., Monigatti, F., Dorff, K., Campagne, F., Bachur, R., and Steen, H. (2009) Urine proteomics for profiling of human disease using high accuracy mass spectrometry. *Proteomics Clin Appl* **3**, 1052-1061

12. Zhao, M., Li, M., Yang, Y., Guo, Z., Sun, Y., Shao, C., Sun, W., and Gao, Y. (2017) A comprehensive analysis and annotation of human normal urinary proteome. *Sci Rep* **7**, 3024

13. Santucci, L., Candiano, G., Petretto, A., Bruschi, M., Lavarello, C., Inglese, E., Righetti, P. G., and Ghiggeri, G. M. (2015) From hundreds to thousands: Widening the normal human Urinome (1). *J Proteomics* **112**, 53-62

14. Zheng, J., Liu, L., Wang, J., and Jin, Q. (2013) Urinary proteomic and non-prefractionation quantitative phosphoproteomic analysis during pregnancy and non-pregnancy. *BMC Genomics* **14**, 777

15. Marimuthu, A., O'Meally, R. N., Chaerkady, R., Subbannayya, Y., Nanjappa, V., Kumar, P., Kelkar, D. S., Pinto, S. M., Sharma, R., Renuse, S., Goel, R., Christopher, R., Delanghe, B., Cole, R. N., Harsha, H. C., and Pandey, A. (2011) A comprehensive map of the human urinary proteome. *J Proteome Res* **10**, 2734-2743

16. Li, Q. R., Fan, K. X., Li, R. X., Dai, J., Wu, C. C., Zhao, S. L., Wu, J. R., Shieh, C. H., and Zeng, R. (2010) A comprehensive and non-prefractionation on the protein level approach for the human urinary proteome: touching phosphorylation in urine. *Rapid Commun Mass Spectrom* **24**, 823-832

17. Wang, Z., Hill, S., Luther, J. M., Hachey, D. L., and Schey, K. L. (2012) Proteomic analysis of urine exosomes by multidimensional protein identification technology (MudPIT). *Proteomics* **12**, 329-338

18. Gonzales, P. A., Pisitkun, T., Hoffert, J. D., Tchapyjnikov, D., Star, R. A., Kleta, R., Wang, N. S., and Knepper, M. A. (2009) Large-scale proteomics and phosphoproteomics of urinary exosomes. *J Am Soc Nephrol* **20**, 363-379

19. Hogan, M. C., Johnson, K. L., Zenka, R. M., Charlesworth, M. C., Madden, B. J., Mahoney, D. W., Oberg, A. L., Huang, B. Q., Leontovich, A. A., Nesbitt, L. L., Bakeberg, J. L., McCormick, D. J., Bergen, H. R., and Ward, C. J. (2014) Subfractionation, characterization, and in-depth proteomic analysis of glomerular membrane vesicles in human urine. *Kidney Int* **85**, 1225-1237

20. Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., King, N., Stein, S. E., and Aebersold, R. (2007) Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **7**, 655-667

21. Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., Stein, S. E., and Aebersold, R. (2008) Building consensus spectral libraries for peptide identification in proteomics. *Nat Methods* **5**, 873-875

22. Searle, B. C. (2010) Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **10**, 1265-1269

23. Xiao, D., Zhao, L. J., Chu, S. L., Jing, H., and Wang, C. (2012) Proteomic analysis for detecting serum biomarkers related to smoking in humans. *Chin Med J (Engl)* **125**, 3027-3032

24. Franciosi, L., Postma, D. S., van den Berge, M., Govorukhina, N., Horvatovich, P. L., Fusetti, F., Poolman, B., Lodewijk, M. E., Timens, W., Bischoff, R., and ten Hacken, N. H. (2014) Susceptibility to COPD: differential proteomic profiling after acute smoking. *PLoS One* **9**, e102037

25. Rice, S. J., Liu, X., Miller, B., Joshi, M., Zhu, J., Caruso, C., Gilbert, C., Toth, J., Reed, M., Rassaei, N., Das, A., Barochia, A., El-Bayoumy, K., and Belani, C. P. (2015) Proteomic profiling of human plasma identifies apolipoprotein E as being associated with smoking and a marker for squamous metaplasia of the lung. *Proteomics* **15**, 3267-3277

26. Della Corte, A., Tamburrelli, C., Crescente, M., Giordano, L., D'Imperio, M., Di Michele, M., Donati, M. B., De Gaetano, G., Rotilio, D., and Cerletti, C. (2012) Platelet proteome in healthy volunteers who smoke. *Platelets* **23**, 91-105

27. Lin, J. L., Bonnichsen, M. H., Nogeh, E. U., Raftery, M. J., and Thomas, P. S. (2010) Proteomics in detection and monitoring of asthma and smoking-related lung diseases. *Expert Rev Proteomics* **7**, 361-372

28. Uh, S. T., Koo, S. M., Jang, A. S., Park, S. W., Choi, J. S., Kim, Y. H., and Park, C. S. (2015) Proteomic differences with and without ozone-exposure in a smoking-induced emphysema lung model. *Korean J Intern Med* **30**, 62-72

29. Huuskonen, P., Amezaga, M. R., Bellingham, M., Jones, L. H., Storvik, M., Häkkinen, M., Keski-Nisula, L., Heinonen, S., O'Shaughnessy, P. J., Fowler, P. A., and Pasanen, M. (2016) The human placental proteome is affected by maternal smoking. *Reprod Toxicol* **63**, 22-31

30. Bortner, J. D., Richie, J. P., Das, A., Liao, J., Umstead, T. M., Stanley, A., Stanley, B. A., Belani, C. P., and El-Bayoumy, K. (2011) Proteomic profiling of human plasma by iTRAQ reveals down-regulation of ITI-HC3 and VDBP by cigarette smoking. *J Proteome Res* **10**, 1151-1159

31. Titz, B., Sewer, A., Schneider, T., Elamin, A., Martin, F., Dijon, S., Luettich, K., Guedj, E., Vuillaume, G., Ivanov, N. V., Peck, M. J., Chaudhary, N. I., Hoeng, J., and Peitsch, M. C. (2015) Alterations in the sputum proteome and transcriptome in smokers and early-stage COPD subjects. *J Proteomics* **128**, 306-320

32. Truedsson, M., Malm, J., Barbara Sahlin, K., Bugge, M., Wieslander, E., Dahlbäck, M., Appelqvist, R., Fehniger, T. E., and Marko-Varga, G. (2016) Biomarkers of early chronic obstructive pulmonary disease (COPD) in smokers and former smokers. Protocol of a longitudinal study. *Clin Transl Med* **5**, 9

33. Musharraf, S. G., Hashmi, N., Choudhary, M. I., Rizvi, N., Usman, A., and Atta-ur-Rahman. (2012) Comparison of plasma from healthy nonsmokers, smokers, and lung cancer patients: pattern-based differentiation profiling of low molecular weight proteins and peptides by magnetic bead technology with MALDI-TOF MS. *Biomarkers* **17**, 223-230

34. Lee, E. J., In, K. H., Kim, J. H., Lee, S. Y., Shin, C., Shim, J. J., Kang, K. H., Yoo, S. H., Kim, C. H., Kim, H. K., Lee, S. H., and Uhm, C. S. (2009) Proteomic analysis in lung tissue of smokers and COPD patients. *Chest* **135**, 344-352

35. Steiling, K., Kadar, A. Y., Bergerat, A., Flanigon, J., Sridhar, S., Shah, V., Ahmad, Q. R., Brody, J. S., Lenburg, M. E., Steffen, M., and Spira, A. (2009) Comparison of proteomic and transcriptomic profiles in the bronchial airway epithelium of current and never smokers. *PLoS One* **4**, e5043

36. Casado, B., Iadarola, P., Pannell, L. K., Luisetti, M., Corsico, A., Ansaldo, E., Ferrarotti, I., Boschetto, P., and Baraniuk, J. N. (2007) Protein expression in sputum of smokers and chronic obstructive pulmonary disease patients: a pilot study by CapLC-ESI-Q-TOF. *J Proteome Res* **6**, 4615-4623

37. Merkel, D., Rist, W., Seither, P., Weith, A., and Lenter, M. C. (2005) Proteomic study of human bronchoalveolar lavage fluids from smokers with chronic obstructive pulmonary disease by combining surface-enhanced laser desorption/ionization-mass spectrometry profiling with mass spectrometric protein identification. *Proteomics* **5**, 2972-2980

38. Airoldi, L., Magagnotti, C., Iannuzzi, A. R., Marelli, C., Bagnati, R., Pastorelli, R., Colombi, A., Santaguida, S., Chiabrando, C., Schiarea, S., and Fanelli, R. (2009) Effects of cigarette smoking on the human urinary proteome. *Biochem Biophys Res Commun* **381**, 397-402

39. Haniff, A. N., and Gam, L. H. (2016) Identification of urinary protein biomarkers for tobacco smoking. *Biotechnol Appl Biochem* **63**, 266-272

40. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567

41. UniProt Consortium, T. (2018) UniProt: The universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169

42. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**, 5383-5392

43. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646-4658

44. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29

45. Elias, J. E., Haas, W., Faherty, B. K., and Gygi, S. P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* **2**, 667-675

46. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P. D. (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* **45**, D183-D189

47. Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G. M., Blake, J. A., Bult, C., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J. M., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S., Fisk, D. G., Hirschman, J. E., Hong, E. L., Nash, R. S., Sethuraman, A., Theesfeld, C. L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S. Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E. M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., White, R., and Consortium, G. O. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**, D258-261

48. Pinto-Plata, V. M., Müllerova, H., Toso, J. F., Feudjo-Tepie, M., Soriano, J. B., Vessey, R. S., and Celli, B. R. (2006) C-reactive protein in patients with COPD, control smokers and non-smokers. *Thorax* **61**, 23-28

49. Lenártová, P., Kopčeková, J., Gažarová, M., Mrázová, J., and Wyka, J. (2017) Biochemical parameters as monitoring markers of the inflammatory reaction by patients with chronic obstructive pulmonary disease (COPD). *Rocz Panstw Zakl Hig* **68**, 185-190

50. Ozguner, F., Koyu, A., and Cesur, G. (2005) Active smoking causes oxidative stress and decreases blood melatonin levels. *Toxicol Ind Health* **21**, 21-26

51. Petersen, A. M., Magkos, F., Atherton, P., Selby, A., Smith, K., Rennie, M. J., Pedersen, B. K., and Mittendorfer, B. (2007) Smoking impairs muscle protein synthesis and increases the expression of myostatin and MAFbx in muscle. *Am J Physiol Endocrinol Metab* **293**, E843-848

52. Benjamini, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. (Hochberg, Y. ed., *Journal of the Royal Statistical Society*

53. Huang, B., Svensson, P., Ärnlöv, J., Sundström, J., Lind, L., and Ingelsson, E. (2016) Effects of cigarette smoking on cardiovascular-related protein profiles in two community-based cohort studies. *Atherosclerosis* **254**, 52-58

54. Pacheco, S. A., Torres, V. M., Louro, H., Gomes, F., Lopes, C., Marçal, N., Fragoso, E., Martins, C., Oliveira, C. L., Hagenfeldt, M., Bugalho-Almeida, A., Penque, D., and Simões, T. (2013) Effects of occupational exposure to tobacco smoke: is there a link between environmental exposure and disease? *J Toxicol Environ Health A* **76**, 311-327

55. Guo, Z., Zhang, Y., Zou, L., Wang, D., Shao, C., Wang, Y., Sun, W., and Zhang, L. (2015) A Proteomic Analysis of Individual and Gender Variations in Normal Human Urine and Cerebrospinal Fluid Using iTRAQ Quantification. *PLoS One* **10**, e0133270

56. Liu, X., Shao, C., Wei, L., Duan, J., Wu, S., Li, X., Li, M., and Sun, W. (2012) An individual urinary proteome analysis in normal human beings to define the minimal sample number to represent the normal urinary proteome. *Proteome Sci* **10**, 70

57. Oh, J., Pyo, J. H., Jo, E. H., Hwang, S. I., Kang, S. C., Jung, J. H., Park, E. K., Kim, S. Y., Choi, J. Y., and Lim, J. (2004) Establishment of a near-standard two-dimensional human urine proteomic map. *Proteomics* **4**, 3485-3497

58. Ishimura, S., Furuhashi, M., Watanabe, Y., Hoshina, K., Fuseya, T., Mita, T., Okazaki, Y., Koyama, M., Tanaka, M., Akasaka, H., Ohnishi, H., Yoshida, H., Saitoh, S., and Miura, T. (2013) Circulating levels of fatty acid-binding protein family and metabolic phenotype in the general population. *PLoS One* **8**, e81318

59. Sun, W., Chen, Y., Li, F., Zhang, L., Yang, R., Zhang, Z., Zheng, D., and Gao, Y. (2009) Dynamic urinary proteomic analysis reveals stable proteins to be potential biomarkers. *Proteomics Clin Appl* **3**, 370-382

60. Nadkarni, S., Cooper, D., Brancaleone, V., Bena, S., and Perretti, M. (2011) Activation of the annexin A1 pathway underlies the protective effects exerted by estrogen in polymorphonuclear leukocytes. *Arterioscler Thromb Vasc Biol* **31**, 2749-2759

61. Chen, C. P., Lai, T. C., Chern, S. R., Li, S. H., Chou, H. C., Chen, Y. W., Lin, S. T., Lu, Y. C., Wu, C. L., Li, J. M., and Chan, H. L. (2013) Proteome differences between male and female fetal cells in amniotic fluid. *OMICS* **17**, 16-26

62. Castagna, A., Cecconi, D., Sennels, L., Rappsilber, J., Guerrier, L., Fortis, F., Boschetti, E., Lomas, L., and Righetti, P. G. (2005) Exploring the hidden human urinary proteome via ligand library beads. *J Proteome Res* **4**, 1917-1930

63. Sun, W., Li, F., Wu, S., Wang, X., Zheng, D., Wang, J., and Gao, Y. (2005) Human urine proteome analysis by three separation approaches. *Proteomics* **5**, 4994-5001

64. Kim, K. H., and Moon, M. H. (2009) High speed two-dimensional protein separation without gel by isoelectric focusing-asymmetrical flow field flow fractionation: application to urinary proteome. *J Proteome Res* **8**, 4272-4278

65. Zhang, Y., Li, Y., Qiu, F., and Qiu, Z. (2010) Comprehensive analysis of low-abundance proteins in human urinary exosomes using peptide ligand library technology, peptide OFFGEL fractionation and nanoHPLC-chip-MS/MS. *Electrophoresis* **31**, 3797-3807

66. Nagaraj, N., and Mann, M. (2011) Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. *J Proteome Res* **10**, 637-645

67. Tyan, Y. C., Yang, M. H., Chen, S. C., Jong, S. B., Chen, W. C., Yang, Y. H., Chung, T. W., and Liao, P. C. (2011) Urinary protein profiling by liquid chromatography/tandem mass spectrometry: ADAM28 is overexpressed in bladder transitional cell carcinoma. *Rapid Commun Mass Spectrom* **25**, 2851-2862

68. Raj, D. A., Fiume, I., Capasso, G., and Pocsfalvi, G. (2012) A multiplex quantitative proteomics strategy for protein biomarker studies in urinary exosomes. *Kidney Int* **81**, 1263-1272

69. He, W., Huang, C., Luo, G., Dal Prà, I., Feng, J., Chen, W., Ma, L., Wang, Y., Chen, X., Tan, J., Zhang, X., Armato, U., and Wu, J. (2012) A stable panel comprising 18 urinary proteins in the human healthy population. *Proteomics* **12**, 1059-1072

70. Zubiri, I., Posada-Ayala, M., Sanz-Maroto, A., Calvo, E., Martin-Lorenzo, M., Gonzalez-Calero, L., de la Cuesta, F., Lopez, J. A., Fernandez-Fernandez, B., Ortiz, A., Vivanco, F., and Alvarez-Llamas, G. (2014) Diabetic nephropathy induces changes in the proteome of human urinary exosomes as revealed by label-free comparative analysis. *J Proteomics* **96**, 92-102

71. Di Meo, A., Batruch, I., Yousef, A. G., Pasic, M. D., Diamandis, E. P., and Yousef, G. M. (2017) An integrated proteomic and peptidomic assessment of the normal human urinome. *Clin Chem Lab Med* **55**, 237-247

72. Hirao, Y., Saito, S., Fujinaka, H., Miyazaki, S., Xu, B., Quadery, A. F., Elguoshy, A., Yamamoto, K., and Yamamoto, T. (2018) Proteome Profiling of Diabetic Mellitus Patient Urine for Discovery of Biomarkers by Comprehensive MS-Based Proteomics. *Proteomes* **6**, 9

2.7 SUPPORTING INFORMATION

Table S3: Large-scale urinary proteome studies

Reference	Year	#urine donors	#proteins identified	Identification system	Instrumentation
Oh *et al.* (57)	2004	40	113	UniProt	Voyager DE-STR MALDI-TOF mass spectrometer
Castagna *et al.* (62)	2005	8	383	IPI	Fourier Transform-Ion Cyclotron Resonance (FT-ICR) Mass Spectrometry
Sun *et al.* (63)	2005	10	226	GI	I-DE, 1-D LC/MS/MS, direct 1-D Lc/MS/MS and 2-D LC/MS/MS
Adachi *et al.* (8)	2006	10	1543	IPI	LTQ-Orbitrap, LTQ-FT
Kim & Moon (64)	2009	29	245	IPI	nLC-ESI-MS/MS, IEF-AF4
Kentsis *et al.* (11)	2009	12	2362	IPI, Swiss-Prot, Ref Seq	Linear iontrap-Orbitrap
Gonzales *et al.* (18)	2009	8	1132	GI	LTQ
Sun *et al.* (59)	2009	6	422	IPI, Swiss-Prot	LCQ Deca XP1 IT mass spectrometer
Li *et al.* (16)	2010	3	1310	IPI	LTQ-Orbitrap, LTQ-Orbitrap Velos
Zhang *et al.* (65)	2010	8	512	IPI	nanoHPLC-chip-MS/MS

Reference	Year	#urine donors	#proteins identified	Identification system	Instrumentation
Nagaraj & Mann (66)	**2011**	**7**	**808**	**UniProt**	**LTQ-Orbitrap**
Tyan *et al.* (67)	2011	16	2782	NCBI	nano-HPLC/ESIMS/ MS
Marimuthu *et al.* (15)	2011	24	1823	GI	LTQ-Orbitrap Velos
Raj *et al.* (68)	2012	4	378	UniProt	LC/ESIMS/ MS for iTRAQ
Wang *et al.* (17)	2012	9	3280	UniProt	LTQ-Orbitrap Velos
He *et al.* (69)	2012	200	1641	IPI	Four different fractionation techniques including in-gel, 2D-LC, OFFGEL, and mRP) coupled with HPLC-Chip-MS/MS
Zheng *et al.* (14)	2013	10	2579	IPI	LTQ-Orbitrap Velos
Zubiri *et al.* (70)	2014	16	352	IPI, Swiss-Prot	nLC–MS/MS, triple quad LC/MS/MS on-line connected to nano-chromatography in a Chip-format configuration (ChipCube interface, Agilent Technologies) and 1200 Series LC Modules
Hogan *et al.* (19)	2014	3 healthy 7 with glomerular disease	1830 5657	Swiss-Prot	LTQ-Orbitrap Velos
Santucci *et al.* (13)	2015	12	3429	UniProt	LTQ-Orbitrap Velos

Reference	Year	#urine donors	#proteins identified	Identification system	Instrumentation
Di Meo *et al.* (71)	2016	6	2091	Swiss-Prot	Q-Exactive Mass spectrometer
Zhao *et al.* (12)	2017	24	6085	Swiss-Prot	1D & 2D LC-MS/MS, gel-eluted liquid fraction entrapment electrophoresis/liquid-phase isoelectric focusing
Hirao *et al.* (72)	2018	10	640	UniProt/Swiss-Prot	QExactive plus
Current study	2018	40	1023	Uniprot	1D gel, HPLC in line with QIT-ESI-MS

CHAPTER THREE

Methionine Sulfoxide Proteome and the Impact of Smoking

3. 1 Abstract

There is accumulating evidence that high levels of methionine sulfoxide (MSO) has been found in plasma proteins in smokers compared to non-smokers. Oxidation plays a key role in the degradation of proteins. While oxidation may directly interfere with activity, the extent to which oxidation affects protein turnover is less clear. Here we quantify methionine sulfoxide containing tryptic peptides in plasma proteins of cigarette smokers versus non-smokers. Using Mass spectrometry-based platforms, we present the first in vivo human methionine sulfoxide proteome in addition to illustrating the impact of smoking.

To determine level of oxidized serum proteins cleared in the urine of smokers and non-smokers with a focus on MSO formation in human serum albumin (HSA) protein.

Urine samples were collected from smoker and non-smokers participants. Proteins were concentrated and concentrates then were subjected to separation via SDS-Gel electrophoresis technique. Gel was cut out into five sections including the band representing intact HSA. Gel sections then were digested with trypsin. Levels of MSO in the resulting peptides were assessed by mass spectrometry. Data analysis was performed using Skyline software package.

A group comparison between non-smokers (control) and smokers showed a slight increase in the levels of MSO found in intact HSA of smokers relative to non-smokers. Regions of gels with proteins of lower mass than intact HSA showed that degraded fragments of HSA in urine of both smokers and non-smokers have higher levels of MSO than oxidation level found in the intact HSA band. HSA in smokers has statistically significant higher levels of MSO than HSA in non-smokers. However, the higher levels of oxidation in smokers are concentrated in partially degraded

HSA. At the moment, it is not possible to say unequivocally whether oxidized HSA is more likely to be cleaved and cleared, if cleaved protein is more likely to be oxidized before clearance, or both.

3.2 INTRODUCTION

Reactive oxygen species (ROS) are naturally produced as intermediates during cellular respiration (1). Although low quantities of antioxidants are required to protect the body from deleterious effects of ROS, , an imbalance between the formation of ROS and detoxification results in so called “Oxidative Stress” (2,3). ROS are involved in several cellular processes including signal transduction (4,5), and antimicrobial action of immune system (6,7). On the other hand, the destructive effects of ROS on different macromolecules, including nucleic acids and proteins are known to be important contributors to degradation of proteins (3,8). Oxidative stress in proteins results in the modification of amino acids or the carbon skeleton of the peptide which consequently leads to fragmentation of the polypeptide chain predisposing the protein to structural and functional alterations (9-13). Plasma contains a small group of proteins exhibiting very high abundance, *e.g.* albumin, and a much larger group of proteins represented in lower abundances. While proteins are among the main targets of ROS, sulfur containing amino acids (Cys and Met) out of the twenty amino acids are potential reductants due to the electronegativity of the thiol group in the structure (14-16). Protein thiol concentrations are in the 400-600 uM range (17). Methionine is considered among the most hydrophobic residues in proteins and it resides mainly in the interior hydrophobic core or embedded in the lipid bilayer membrane and therefore fairly protected from oxidative stress (18-20). Meanwhile, surface exposed Met residues can be easily oxidized, subsequently, they are well known for their ROS scavenging actions (21-23).

The major product that arises from oxidation of Met is methionine sulfoxide (MSO). Further oxidation of Met results in the formation of methionine sulfone, which occurs to a much lesser extent (1) (Fig. 1). Mechanisms of MSO generation depend on the type of oxidant. Hydrogen peroxide, for instance, results in the formation of MSO via transfer of oxygen molecule in a two-

electron oxidation. Metal ions such as Fe^{3+} and $HO\cdot$ oxidative radicals utilize one electron oxidation mechanisms in generating MSO from methionine (24). Reaction of $HO\cdot$ with Met occurs at a rate constant of $7\times10^{9}\ M^{-1}\cdot S^{-1}$, while the reaction of Met with H_2O_2 proceeds at a much slower rate, $2\times10^{-2}\ M^{-1}\cdot S^{-1}$ (25).

Conversely, MSO can be recycled back enzymatically to its reduced state via methionine sulfoxide reductases (MSRs). Thus, both components are important in the damage and repair mechanisms of proteins (26,27) Methionine sulfoxide is known to modulate several pathways *e.g.* activation of calcium/calmodulin dependent protein kinase II (CaMKII) in the absence of calcium and in activating transcription factors (28-30). Also, methionine sulfoxide has been linked to the etiology of several diseases including cardiovascular, renal, neurodegenerative and age-related diseases (24,26,31), and malignancies including liver and breast cancers (32-35). Oxidation of certain methionine residues has been linked to impairment of protein functions such as Met 388 in thrombomodulin resulting in decreased coagulation activity (36,37), and oxidation of both Met 33 and 209 in F_c region of IgG1 impose negative consequences on the structure and stability of the protein (31).

Tobacco use is the leading cause of preventable death and diseases. It results in the death of approximately 500,000 Americans annually (38). Cigarette smoke is a complex mixture containing nearly 4000 chemicals (8), it also includes several reactive oxygen species (ROS) such as hydroxyl radical ($HO\cdot$), superoxide anion ($O_2^{-\cdot}$) and hydrogen peroxide (H_2O_2). Some of these species are highly reactive such as hydroxyl radical. It oxidizes almost all macromolecules. Most importantly, there is a well-established linkage between the disease of emphysema in smokers and methionine oxidation of a specific protein, α 1-antitrypsin (39). Oxidation of either methionine 351 or 358 in

the binding site of α 1-antitrypsin destroys its ability to bind elastase, consequently elastase activity may increase as a consequence of oxidative stress (40).

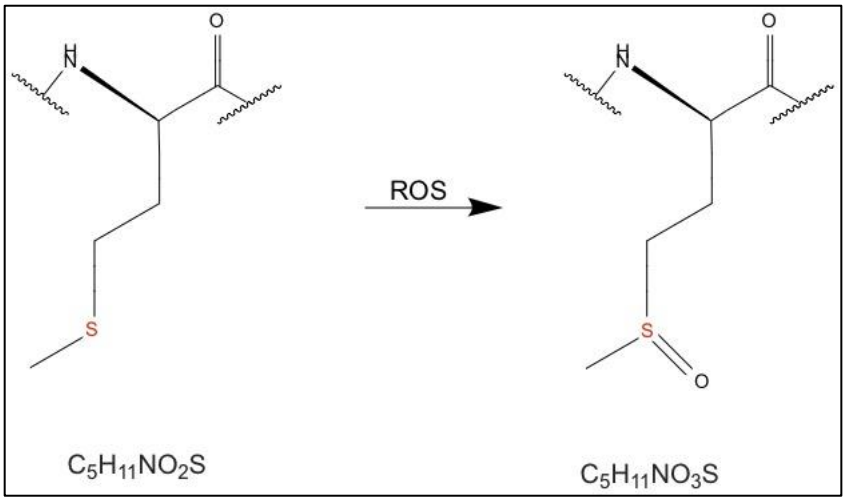


Figure 1: Met is oxidized to Met sulfoxide by ROS and to Met sulfone at higher concentrations of these oxidants

Human serum albumin is a monomer with a molecular mass of ~ 66.5 kDa and it accounts for 50% - 60% of total plasma protein. HSA has six Met and 35 Cys residues forming 17 disulfide bonds (Fig. 2). HSA is regarded as an important antioxidant in plasma due to the single free cysteine (Cys 34) (35,41). In addition to its antioxidant properties, HSA plays a significant role in binding and transporting multiple ligands such as cholesterol, fatty acids and bile pigments, in addition to playing a key role in regulating osmotic pressure in the body (42).

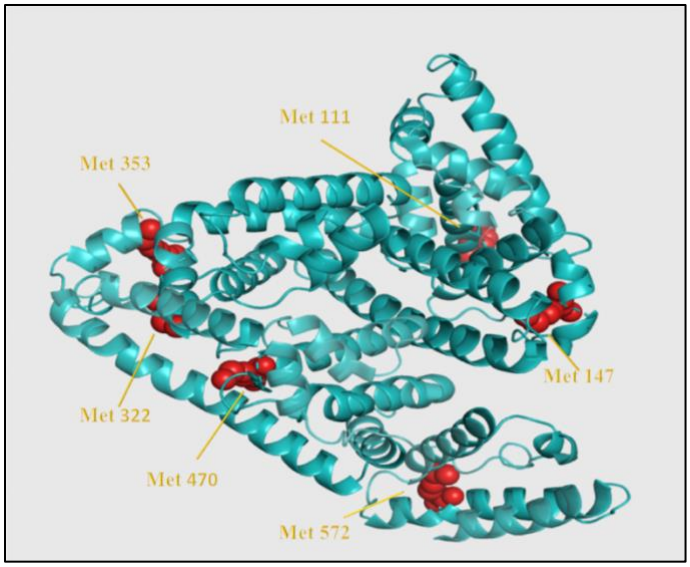


Figure 2: Methionine residues in the structure of HSA. Residues are shown in red spheres and are labeled according to their sequential order. PDB ID: 1BMO

Proteomic approaches such as 2D-electrophresis and mass spectrometry (MS) represent robust tools for identification and quantitation of post translational modifications (PTMs). Modern instruments possess the resolution to measure the shift in the mass caused by a single oxygen molecule (43). Mass spectrometry has evolved as the prevalent technique for the identification and characterization of proteins structure as well as quantification of PTMs. Oxidative PTM of proteins are of special interest as they appear to play an important role in many diseases and metabolic dysfunctions. Tandem MS (MS/MS) based techniques are powerful in localizing oxidation modification in proteins (44). It enables accurate measurement of mass to charge ratio (m/z) of ionized analytes. Since oxidation PTM alters the mass of a protein it changes the m/z ratio of the protein or peptide containing the oxidized amino acids (45).

Biofluids (e.g. urine and blood) have been used as valuable sources for proteomics analysis over decades due to their relative stable composition and the convenience of obtaining. It has been estimated that more than 1500 different proteins are cleared in the urine (44,46,47). Proteomic

studies dependent on urine samples demand a sensitive detection of total proteins. A study reported that different methods may respond inadequately for each protein present in a urine sample, consequently, standardization would be unachievable. This is mainly because of the complexity of urine and partially due to the fact that conformational changes in plasma proteins alter their filtration rate of the kidneys. In addition, proteins in urine (e.g. HSA) are exposed to a wider range pH than found in plasma (48).

3.3 MATERIALS AND METHODS

Procedural details including urine sample collection and processing, in-gel protein digestion, peptide desalting, and analysis of the digest via ESI-LC-MS/MS are extensively discussed in chapter 2. A schematic overview of overall procedure used in the current study is given in Fig. 3.

3.3.1 Data analysis

Protein identifications were performed via MASCOT (Matrix Science, London, UK) (49). searching all mass spectra data against human protein entries obtained from UniProt (50). Statistical analysis and extracting MS1 intensities of the identified tryptic peptides were achieved with Skyline-daily software package, an open source Windows client application for both Data Independent Acquisition (DIA/SWATH) and targeted Data Dependent Acquisition (DDA) with MS1 quantitative methods (MacCoss lab, University of Washington/Seattle), in accordance with

the number of replicates and the condition of samples (healthy and disease representing non-smoker and smoker samples, respectively) (51).

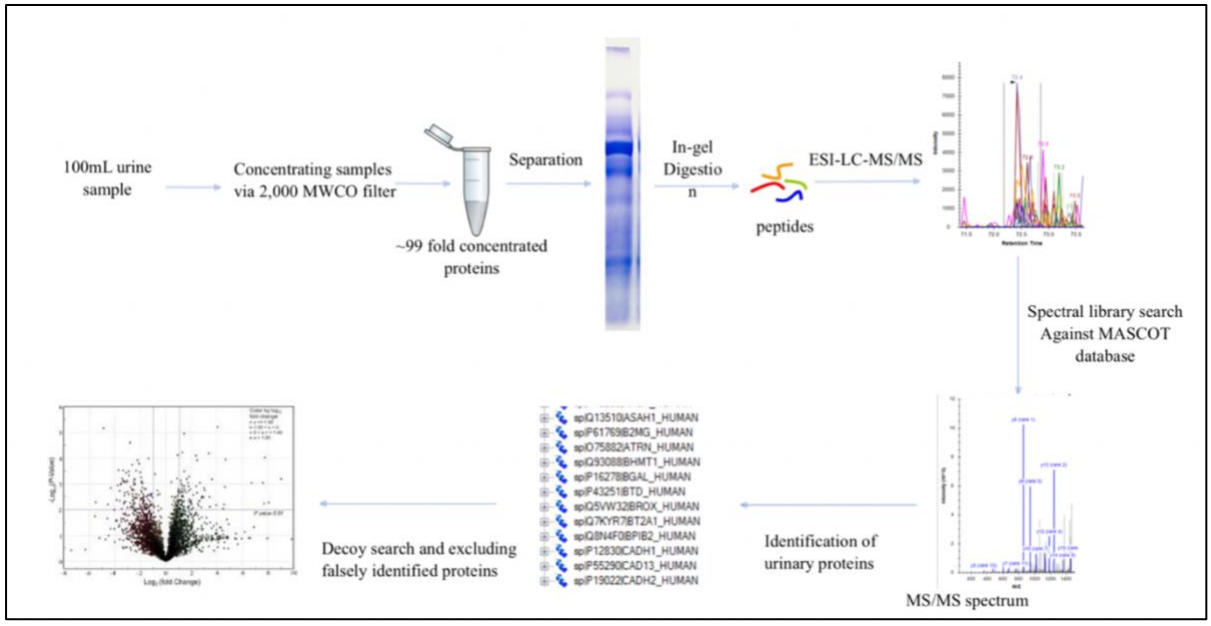


Figure 3. Schematic overview of the overall procedure used in the current study

3.3.2 Targeted proteomics and high accuracy protein quantitation

Targeted protein work was performed using a Shimadzu UPLC-20A coupled to a Shimadzu 8050 triple quadrupole ESI-MS with heated probe. Tryptic peptides were separated by reverse-phase ultra-high-performance liquid chromatography (RP-UPLC) compatible Shimadzu C18, 1.9-micron particle size, 50x2.1 mm column (SN # 16041880T), with a solvent flow rate of 0.3 mL/minute, and a gradient of 5%–90% consisting of 0.1% FA (solvent A) and 0.1% FA in ACN (solvent B) over a time period of 10 minutes. Sequence specific fragment ion intensities from all six Met containing (oxidized and reduced) tryptic peptides from HSA were used in the protein quantitation. Multiple reaction monitoring (MRM) events corresponding to sequence specific

fragment ions derived from the precursor tryptic peptides were targeted to operate at a certain specific retention time interval predicted by in-house retention time library. This library was generated using oxidized peptides. (Biomatik, Wilmington, DE). The library for the reduced peptides was generated using correlation of relative hydrophobicity of the tryptic peptides with their retention times (RT). While the RT correlated well within 99% confidence, a sufficient number of sequence specific fragment ions were present to use as basis for identification of the tryptic peptide by MS/MS alone. Further specificity and confidence were achieved by incorporating RT prediction.

3.4 RESULTS AND DISCUSSION

Urine samples were grouped according to sex, smoking status and age groups (chapter 2). Peptide settings used in Skyline allowed a maximum of one missed tryptic cleavage in identifying peptides ranging from 8-35 amino acids with carbamidomethylation of Cys as a fixed modification and oxidation of methionine as variable modification. The search resulted in identification of 198 proteins (table S1) harboring 392 methionine sulfoxide (MSO) containing peptides (table S2). About 70% of the proteins were found containing a single peptide with MSO. Skyline enables prediction of retention times using a function of sequence specific retention calculator (SSRCalc.) proposed by Krokhin and Spicer (52) (Fig.4). A combination of database search of spectral library and retention time prediction increases the accuracy by which proteins are identified.

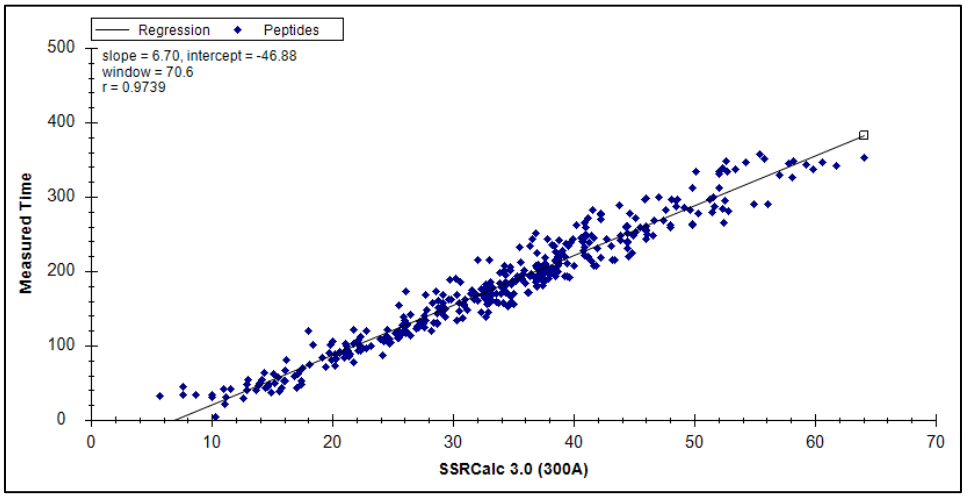


Figure 4. Skyline uses the retention times of the peptides to normalize the times of the replicates to get the maximum achievable linear regression.

The aim of using DDA maximize protein coverage and minimize redundant peptide precursors ions (53,54). However, studies have pointed technical issues arises using DDA. It has been shown that reproducibility of peptide identification among technical replicates results in overlaps (55,56). Bateman et al. addressed the issue of maximizing peptide identification using data-dependent acquisition (DDA). They proposed an analytical work flow to combine DDA with retention time (RT) aligned extracted ion chromatogram. They observed an increase in peptide identification using mouse blood plasma up to 30.5% due to the comparison of peptide MS1 extracted ion chromatogram followed by retention time alignment of co-identified peptides. Further, they found that the method is useful for quantitative measurements using dilution series of known standards of bovine peptides spiked in mouse plasma (57).

3.4.1 Comparison of methionine sulfoxide levels in smokers and non-smokers

To graphically represent oxidation in tryptic peptides, we generated the so-called volcano plot representing intensities acquired for all forty bio-samples (16 non-smokers and 24 smokers), in which $\log_{10}$ of the adjusted p-value for the difference in intensities between smokers and non-smokers is plotted against $\log_2$ of the actual fold changes between the two categories using non-smokers as reference. As it can be readily seen in Fig. 5 (a, b), a significant increase in the amounts of oxidized peptides from a wide variety of proteins. The large number of peptides with significant fold changes compared those that were not found significant is a quick way to illustrate increased oxidation of Met in smokers compared to non-smokers. In no case is a methionine found to be significantly more oxidized (down-regulated) in non-smokers than smokers. Testing results (table S2) show nearly 37% of the total identified proteins demonstrate elevated levels of MSO in smokers.

While previous studies examined biomarkers of oxidative stress and others investigated effects of specific methionine oxidations, there is a lack of sufficient evidence illustrating the proteome of oxidized methionine residues. Zurbig *et al.* (58) analyzed the low molecular weight urinary proteome of 324 healthy individuals aging between 2-73 looking for biomarkers indicating kidney aging and chronic kidney disease. The study focused on 49 peptides reflecting aging among which, fragments of collagen alpha-1 were predominant. They also identified three peptides for uromodulin that was spotted in our study with significant fold change in smoker with a fold change of 2.93 (p-value = 0). Surprisingly, the majority of peptides they identified were not methionine containing fragments. Similarly, several proteins were found to be significantly higher in an attempt by Siwy *et al.* to diagnose chronic kidney disease through analyzing urine proteome

including serum albumin, uromodulin and apolipoprotein, each with several fragments (59). Oxidation of methionine 35 in Alzheimer's amyloid beta-peptide 1-42 is shown to be associated with oxidative stress and neurotoxicity (60,61). Ghesquière *et al.* However, analyzed oxidation of methionine using human jurkat cellular inflammation as a model. So far, the study is regarded as the largest pool of protein-bound methionine oxidation (18). Other non-human protein-bound methionine oxidation were carried out using rats and *Bacillus cereus* (62,63).

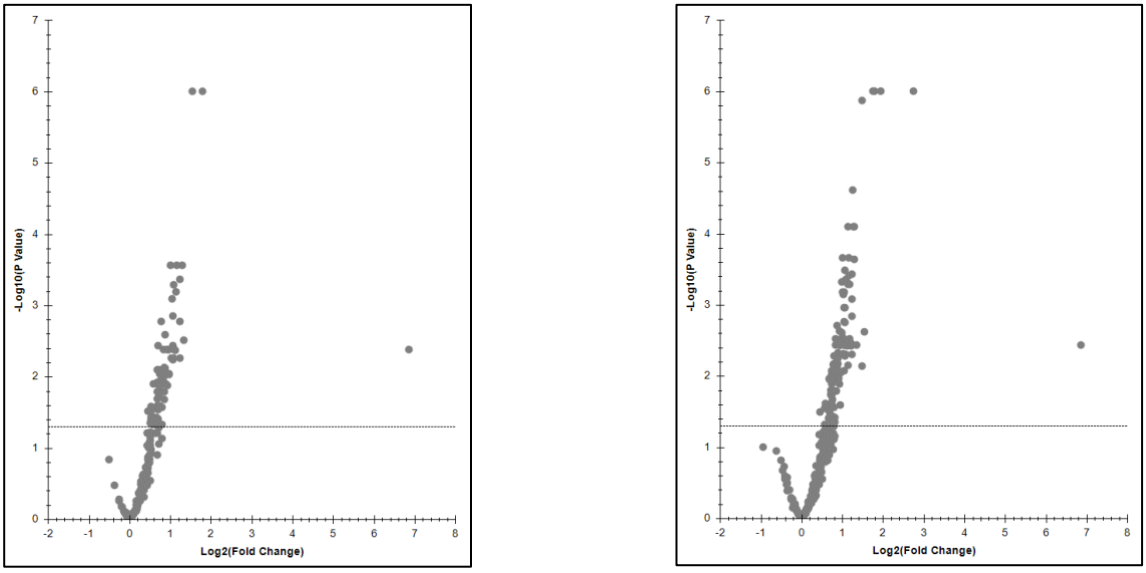


Figure 5: Volcano plot illustrates significant fold-changes of MSO. (Left) Significantly up-regulated methionine sulfoxide containing proteins. (right) Significantly up-regulated MSO containing peptides.

Higher levels of methionine sulfoxide were further confirmed via targeted proteomic approach using tryptic peptides of HSA. Serum albumin was identified in all bio-samples collected in the current study. The tryptic peptides containing all six methionine residues in HSA, length and masses are illustrated in table 1. A chromatogram of the peptides (oxidized and non-oxidized are shown in figure (6). MS intensities and spectral MS/MS spectral count acquired for HSA were used to draw a general picture of percent oxidation and redox ratios in both experimental and control populations, in addition to studying the effect smoking could have on protein turnover.

Targeted proteomics analysis of peak areas obtained for Met containing fragments (oxidized and reduced) of serum albumin agrees with the up-regulation of MSO in smoker compared to control, discussed in DDA. All six peptides were chosen for further analysis. First, peptides were confirmed via database search (Fig. 7). Three sequence specific fragment ions *b9,* and *y7, y9* were considered for this approach. Intensity comparison of these fragments elucidated higher levels of MSO in smokers (Fig. 8). Group comparison of MSO in intact HSA smokers compared to non-smokers show highly significant fold change demonstrated by each fragment (p-value < 0.05) (table 2).

Table 1: Methionine containing tryptic peptides of HSA, molecular masses, and their relative solvent accessibilities

Mass (Da)	Position	Peptide sequence	% Relative solvent accessibility
2917.3229	311-337	SHCIAEVENDEMPADLPSLA ADFVESK	37
2593.2425	139-160	LVRPEVDVMCTAFHDNEETF LK	20
2404.1709	470-490	MPCAEDYLSVVLNQLCVLHE K	24
1623.7875	348-360	DVFLGMFLYEYAR	15
1342.6347	570-581	AVMDDFAAFVEK	15
1320.4905	106-117	ETYGEMADCCAK	21

Table 2: Targeted proteomic data analysis reveals a significant fold change of methionine oxidation in smokers

Peptide	Fold Change	p-Value (peptide)	p-Value (protein)
ETYGEMADCCAK	7.43	0.0049	0.039
LVRPEVDVMCTAFHDNEETFLK	3.51	0.0099	
SHCIAEVENDEMPADLPSLAADFVESK	4.13	0.0049	
DVFLGMFLYEYAR	2.9	0.0049	
MPCAEDYLSVVLNQLCVLHEK	4.89	0.0049	
AVMDDFAAFVEK	2.49	0.0099	

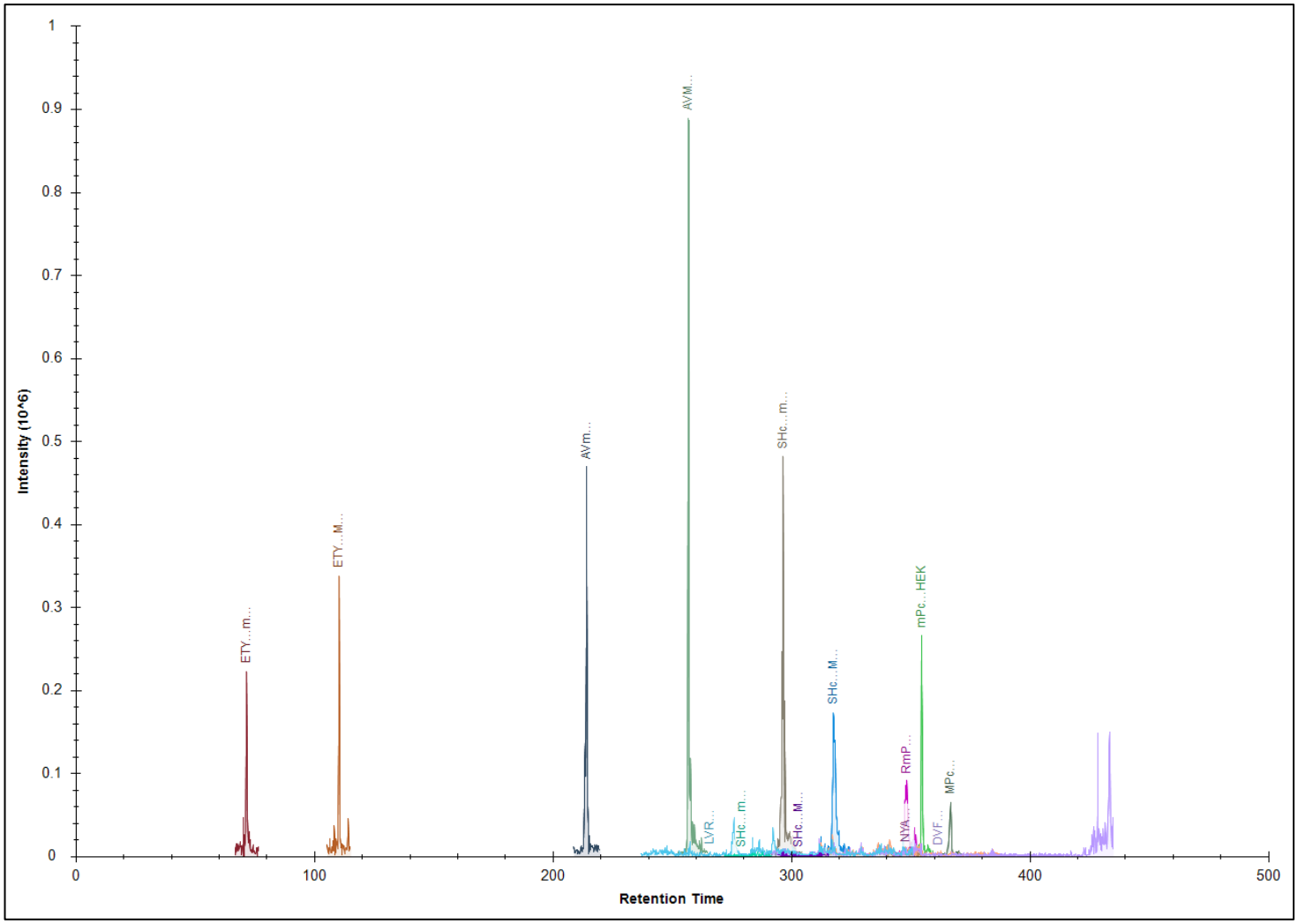


Figure 6: Reverse phCE HPLC-MS/MS separation and identification of tryptic peptides (oxidized and non-oxidized) of HAS. Skyline generates a color for each peptide based on the peptide sequence and modifications.

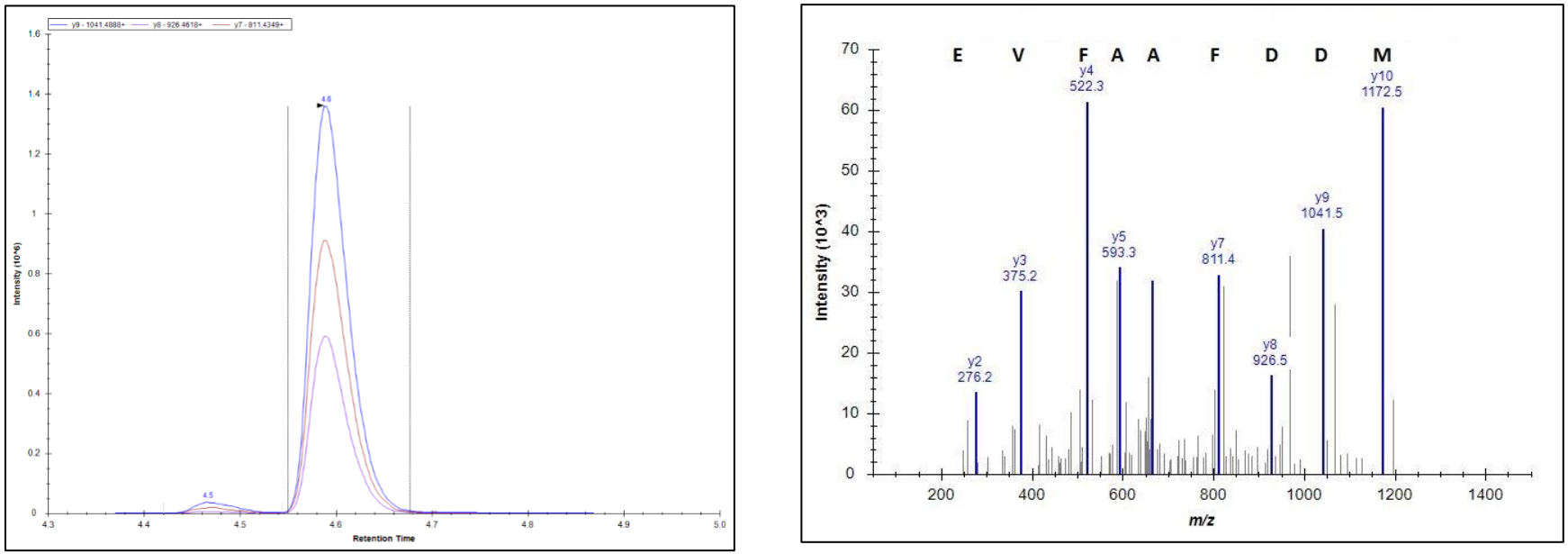


Figure 7: (left) Reverse phase HPLC-MS/MS separation and identification of tryptic peptides (oxidized and non-oxidized) of HSA. Skyline generates a color for each peptide based on the peptide sequence and modifications. Peptide (AVM…) was selected for targeted proteomics due to its high intensity. (right) Spectral library of peptide (AVM…) identified by MASCOT algorithm.

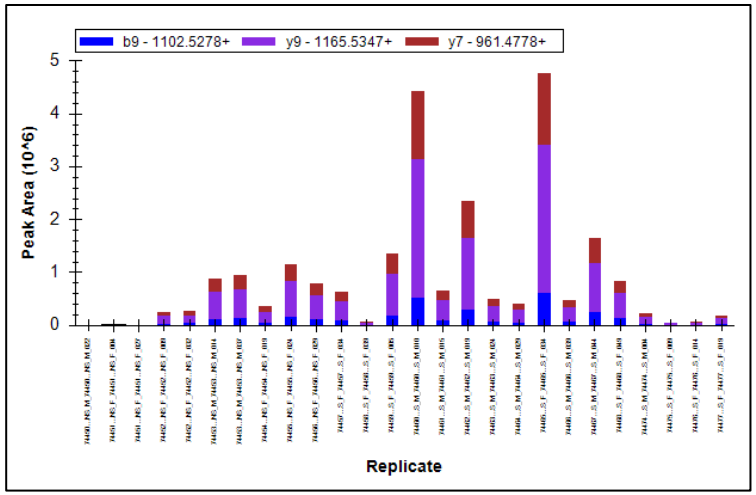


Figure 8: Intensity comparison of optimized MRM for the tryptic peptides between smokers and non-smokers. Three sequence specific fragment ions *b9* (N- terminus protected fragment ions), *y*7, *y*9 (C-terminus protected fragment ions) were considered for methionine sulfoxide containing fragments of HSA.

3.4.2 Larger amount of proteins is cleared in the urine of smokers compared to non-smokers

Intensities observed for the top 30 peptides representing proteins predominately exist in the collected samples indicate that more segments of an individual protein were detected in smokers (Fig. 9 a &b). The selected peptides in control group were mapped to 13 different proteins in which, seven peptides were derived from HSA. The intensity of the tryptic peptides ranged from 4.6×10^{9} to 2.9×10^{10} for IGKC and AMBP proteins, respectively. On the other hand, 30 tryptic peptides selected for smokers represented seven proteins only, in which, 14 peptides were mapped to HSA. Intensities of the selected peptides extended from 1.05×10^{10} to 6.44×10^{10} for HSA and AMBP proteins, respectively. The increased number of covered peptides per protein in smokers suggests secretion of higher amount of proteins in smokers. The increased level of proteins in the renal clearance of the experimental group is reflected in the intensities gained for oxidized and reduced tryptic peptides of HSA (Fig 9c). In which, almost all Met containing domains show higher intensities of the oxidized peptides in smokers compared to control group except for the peptide MPCAEDYLSVVLNQLCVLHEK which is non-significantly regulated in the opposite.

The increased level of proteins cleared in the urine samples of smokers could be mainly attributed the increased level of MSO in smokers. Studies have established a well connection between oxidative stress and protein degradation. Lai *et al.* tested the mechanism of oxidation and degradation of proteins caused by particulate matter (PM) (tobacco smoke is known to cause fine PM that exceeds the limits of outdoor pollutants) (64). Other studies have shown increased level of certain proteins expression in smokers, *e.g.* C-reactive protein, fibrinogen, interleukin-6 and PD-L1 (65,66) .

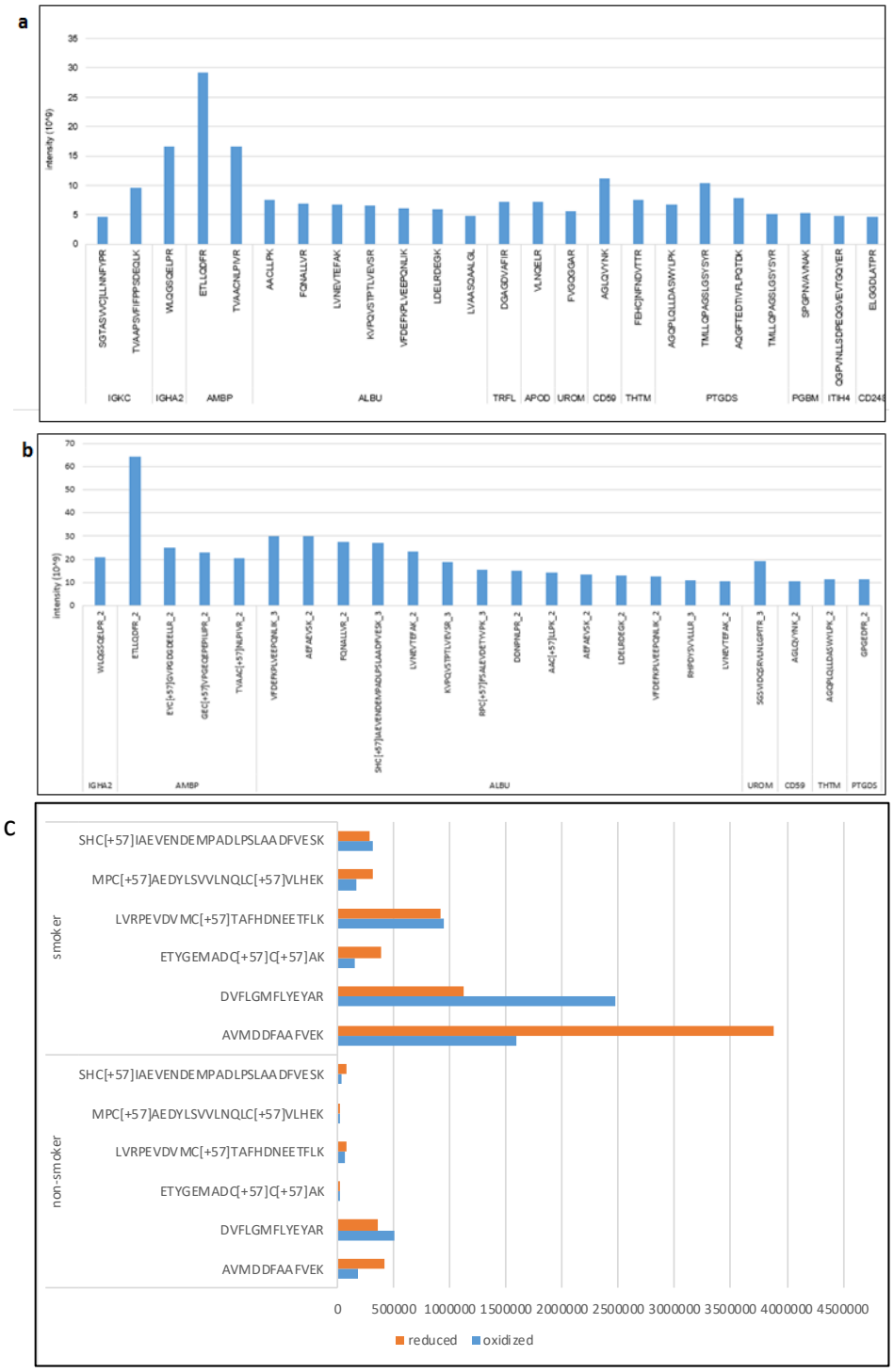


Figure 9: Higher level of plasma proteins is detected in the urine of smokers. (a) Top 30 abundant peptides match 13 proteins, of which 7 peptides were shown to represent HSA. (b) Top 30 abundant peptides are shown to match 7 proteins in which almost 50% of the selected peptides match Met containing tryptic fragments of HSA. (c) Intensity of reduced vs. oxidized Met containing peptides of HSA (smokers and non-smokers).

3.4.3 Methionine sulfoxide formation by cigarette smoke is associated with the degradation of plasma proteins

Increased number of peptides show large, statistically significant changes in the intensity of peptides with oxidized Met residues in smokers compared to non-smokers. Gel band labeled HSA contains predominately intact serum albumin and the other regions of the gel contains a mixture of proteins including peptides derived from HSA. Comparison of MSO in peptides digested from intact HSA (table 3) shows an increase in the intensity of oxidized peptides in smokers relative to non-smokers, but these changes have low statistical significance. However, comparison of MSO in the peptides derived from the HSA fragments found in HSA in band labeled D revealed an increased level of oxidative modification (p-value 0.053).

There is a significant association between oxidation of serum proteins mainly caused by tobacco smoke and their fragmentation. Degradation of proteins has been extensively studied. One possible causation is perhaps the deleterious effects of ROS. Proteolysis of cigarette-smoke induced oxidized proteins can occur in several ways. Agarwal *et al.* assessed the fragmentation and oxidation in proteins excreted in urine of 20 males with chronic kidney disease using immunoblotting. Infusion of iron was used as a source of carbonylation. They found that degradation of HSA was directly proportional with the increased level of total protein excretion (67). Panda *et al.* described a two-step process for the degradation of cigarette smoke induced degradation including oxidation followed by a rapid proteolytic via the action of proteases (68). Smoking is also known to activate transcription factors, mainly of those with proteolytic activity. Pedersen *et al.* noticed elevated levels of muscle-specific (MAFBx) factor that regulates ubiquitin-mediated proteolysis, in smokers (69). Li *et al.* described a pathway involved in the degradation

of proteins caused by cigarette smoke that relies on releasing of microvesicles with proteolytic activities when they exposed human macrophage to tobacco smoke extract (70).

3.5 CONCLUSIONS

Analysis of data revealed a comparable difference in protein intensities across all forty bio-samples used in the current study. Group comparisons of MSO content between both smoker and control populations demonstrated statistically significant fold change in smokers. Furthermore, we noticed significant fold change of oxidized methionine residues in the lower mass region suggesting an association between oxidation of Met and protein turnover (p-value = 0.053). Analysis of total proteins also showed an increased level of proteins cleared in the urine samples of smokers irrespective of methionine oxidation. The current study is the first to address methionine sulfoxide proteome in human serum proteins. However. It is early to confirm that the majority of oxidation occurred in vivo. Further evaluation of sample processing is necessary to observe the extent to which oxidation was introduced during sample preparations.

Peptide	Fold Change	P-Value (peptide)	P-value (protein)	Fold Change	P-Value (peptide)	P-value (protein)
	whole gel			HSA band		
ETYGEMADCCAK	2.79	0.0018		4.02	0.109	
SHCIAEVENDEMPADLPSLAADFVESK	1.75	0.0328		0.84	0.7431	
SHCIAEVENDEMPADLPSLAADFVESKDVCK	1.8	0.0018		1.41	0.5827	
RMPCAEDYLSVVLNQLCVLHEK	2.32	0.0017	0.0016	3.34	0.0729	0.085
MPCAEDYLSVVLNQLCVLHEK	2.28	0.0017		2.37	0.1427	
MPCAEDYLSVVLNQLCVLHEKTPVSDR	1.73	0.0018		1.46	0.3941	
AVMDDFAAFVEK	1.25	0.3237		1.32	0.6324	
	C band			D band		
ETYGEMADCCAK	2	0.5174		2.82	0.1849	
SHCIAEVENDEMPADLPSLAADFVESK	1.08	0.8703		2.85	0.1849	
SHCIAEVENDEMPADLPSLAADFVESKDVCK	1.68	0.5174		1.7	0.1849	
RMPCAEDYLSVVLNQLCVLHEK	1.22	0.8258	0.4163	2.02	0.1849	0.053
MPCAEDYLSVVLNQLCVLHEK	1.66	0.5174		2.91	0.1249	
MPCAEDYLSVVLNQLCVLHEKTPVSDR	1.75	0.5174		1.27	0.5009	
AVMDDFAAFVEK	0.79	0.8258		1.21	0.6796	
	C & D bands					
ETYGEMADCCAK	2.27	0.1659				
SHCIAEVENDEMPADLPSLAADFVESK	1.66	0.2899				
SHCIAEVENDEMPADLPSLAADFVESKDVCK	1.65	0.1659				
RMPCAEDYLSVVLNQLCVLHEK	1.51	0.2899	0.0987			
MPCAEDYLSVVLNQLCVLHEK	2.11	0.1659				
MPCAEDYLSVVLNQLCVLHEKTPVSDR	1.47	0.1659				
AVMDDFAAFVEK	0.94	0.8671				

Table 3: Group comparison of MSO in intact HSA found in HSA band between 24 smokers and 16 non-smokers showing significant fold changes and p-values for individual peptides as well as whole HSA. In the whole gel and nearly significant in D band, however, fold changes did not demonstrate a significant fold change including intact HSA band, C band and C&D bands together.

3.6 REFERENCES

1. Drazic, A., and Winter, J. (2014) The physiological role of reversible methionine oxidation. *Biochim Biophys Acta* **1844**, 1367-1382

2. Sies, H. (1997) Oxidative stress: oxidants and antioxidants. *Exp Physiol* **82**, 291-295

3. Dunlop, R. A., Brunk, U. T., and Rodgers, K. J. (2009) Oxidized proteins: mechanisms of removal and consequences of accumulation. *IUBMB Life* **61**, 522-527

4. Mieyal, J. J., and Chock, P. B. (2012) Posttranslational modification of cysteine in redox signaling and oxidative stress: Focus on s-glutathionylation. *Antioxid Redox Signal* **16**, 471-475

5. Ray, P. D., Huang, B. W., and Tsuji, Y. (2012) Reactive oxygen species (ROS) homeostasis and redox regulation in cellular signaling. *Cell Signal* **24**, 981-990

6. Lavrovsky, Y., Chatterjee, B., Clark, R. A., and Roy, A. K. (2000) Role of redox-regulated transcription factors in inflammation, aging and age-related diseases. *Exp Gerontol* **35**, 521-532

7. Wang, H., Han, X., Wittchen, E. S., and Hartnett, M. E. (2016) TNF-α mediates choroidal neovascularization by upregulating VEGF expression in RPE through ROS-dependent β-catenin activation. *Mol Vis* **22**, 116-128

8. Yeh, C. C., Barr, R. G., Powell, C. A., Mesia-Vela, S., Wang, Y., Hamade, N. K., Austin, J. H., and Santella, R. M. (2008) No effect of cigarette smoking dose on oxidized plasma proteins. *Environ Res* **106**, 219-225

9. Sung, C. C., Hsu, Y. C., Chen, C. C., Lin, Y. F., and Wu, C. C. (2013) Oxidative stress and nucleic acid oxidation in patients with chronic kidney disease. *Oxid Med Cell Longev* **2013**, 301982

10. Sanders, L. H., and Timothy Greenamyre, J. (2013) Oxidative damage to macromolecules in human Parkinson disease and the rotenone model. *Free Radic Biol Med* **62**, 111-120

11. Kryston, T. B., Georgiev, A. B., Pissis, P., and Georgakilas, A. G. (2011) Role of oxidative stress and DNA damage in human carcinogenesis. *Mutat Res* **711**, 193-201

12. AVERY, S. V. (2011) Molecular targets of oxidative stress. **434**, 201–210

13. Wang, L., and Chance, M. R. (2011) Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting. *Anal Chem* **83**, 7234-7241

14. Brosnan, J. T., and Brosnan, M. E. (2006) The sulfur-containing amino acids: an overview. *J Nutr* **136**, 1636S-1640S

15. Stipanuk, M. H. (2004) Sulfur amino acid metabolism: pathways for production and removal of homocysteine and cysteine. *Annu Rev Nutr* **24**, 539-577

16. Berlett, B. S., and Stadtman, E. R. (1997) Protein oxidation in aging, disease, and oxidative stress. *J Biol Chem* **272**, 20313-20316

17. Oliveira, P. V. S., and Laurindo, F. R. M. (2018) Implications of plasma thiol redox in disease. *Clin Sci (Lond)* **132**, 1257-1280

18. Ghesquière, B., Jonckheere, V., Colaert, N., Van Durme, J., Timmerman, E., Goethals, M., Schymkowitz, J., Rousseau, F., Vandekerckhove, J., and Gevaert, K. (2011) Redox proteomics of protein-bound methionine oxidation. *Mol Cell Proteomics* **10**, M110.006866

19. Nomura, T., Kamada, R., Ito, I., Chuman, Y., Shimohigashi, Y., and Sakaguchi, K. (2009) Oxidation of methionine residue at hydrophobic core destabilizes p53 tetrameric structure. *Biopolymers* **91**, 78-84

20. Levine, R. L., Berlett, B. S., Moskovitz, J., Mosoni, L., and Stadtman, E. R. (1999) Methionine residues may protect proteins from critical oxidative damage. *Mech Ageing Dev* **107**, 323-332

21. Parker, H., and Winterbourn, C. C. (2012) Reactive oxidants and myeloperoxidase and their involvement in neutrophil extracellular traps. *Front Immunol* **3**, 424

22. Luo, S., and Levine, R. L. (2009) Methionine in proteins defends against oxidative stress. *FASEB J* **23**, 464-472

23. Levine, R. L., Moskovitz, J., and Stadtman, E. R. (2000) Oxidation of methionine in proteins: roles in antioxidant defense and cellular regulation. *IUBMB Life* **50**, 301-307

24. Schöneich, C. (2005) Methionine oxidation by reactive oxygen species: reaction mechanisms and relevance to Alzheimer's disease. *Biochim Biophys Acta* **1703**, 111-119

25. Davies, M. J. (2005) The oxidative environment and protein damage. Biochimica et Biophysica Acta

26. Stadtman, E. R., Van Remmen, H., Richardson, A., Wehr, N. B., and Levine, R. L. (2005) Methionine oxidation and aging. *Biochim Biophys Acta* **1703**, 135-140

27. Vanhooren, V., Navarrete Santos, A., Voutetakis, K., Petropoulos, I., Libert, C., Simm, A., Gonos, E. S., and Friguet, B. (2015) Protein modification and maintenance systems as biomarkers of ageing. *Mech Ageing Dev* **151**, 71-84

28. Drazic, A., Miura, H., Peschek, J., Le, Y., Bach, N. C., Kriehuber, T., and Winter, J. (2013) Methionine oxidation activates a transcription factor in response to oxidative stress. *Proc Natl Acad Sci U S A* **110**, 9493-9498

29. Bertolotti-Ciarlet, A., Wang, W., Lownes, R., Pristatsky, P., Fang, Y., McKelvey, T., Li, Y., Drummond, J., Prueksaritanont, T., and Vlasak, J. (2009) Impact of methionine oxidation on the binding of human IgG1 to Fc Rn and Fc gamma receptors. *Mol Immunol* **46**, 1878-1882

30. Erickson, J. R., Joiner, M. L., Guan, X., Kutschke, W., Yang, J., Oddis, C. V., Bartlett, R. K., Lowe, J. S., O'Donnell, S. E., Aykin-Burns, N., Zimmerman, M. C., Zimmerman, K., Ham, A. J., Weiss, R. M., Spitz, D. R., Shea, M. A., Colbran, R. J., Mohler, P. J., and Anderson, M. E. (2008) A dynamic pathway for calcium-independent activation of CaMKII by methionine oxidation. *Cell* **133**, 462-474

31. Feng, B., Wang, Z., Liu, T., Jin, R., Wang, S., Wang, W., Xiao, G., and Zhou, Z. (2014) Methionine oxidation accelerates the aggregation and enhances the neurotoxicity of the D178N variant of the human prion protein. *Biochim Biophys Acta* **1842**, 2345-2356

32. Hecht, F., Pessoa, C. F., Gentile, L. B., Rosenthal, D., Carvalho, D. P., and Fortunato, R. S. (2016) The role of oxidative stress on breast cancer development and therapy. *Tumour Biol* **37**, 4281-4291

33. Fernández-Irigoyen, J., Santamaría, E., Sesma, L., Muñoz, J., Riezu, J. I., Caballería, J., Lu, S. C., Prieto, J., Mato, J. M., Avila, M. A., and Corrales, F. J. (2005) Oxidation of specific methionine and tryptophan residues of apolipoprotein A-I in hepatocarcinogenesis. *Proteomics* **5**, 4964-4972

34. Bakhtiari, S., Azimi, S., Mehdipour, M., Amini, S., Elmi, Z., and Namazi, Z. (2015) Effect of Cigarette Smoke on Salivary Total Antioxidant Capacity. *J Dent Res Dent Clin Dent Prospects* **9**, 281-284

35. Hackett, T. L., Scarci, M., Zheng, L., Tan, W., Treasure, T., and Warner, J. A. (2010) Oxidative modification of albumin in the parenchymal lung tissue of current smokers with chronic obstructive pulmonary disease. *Respir Res* **11**, 180

36. Wood, M. J., Helena Prieto, J., and Komives, E. A. (2005) Structural and functional consequences of methionine oxidation in thrombomodulin. *Biochim Biophys Acta* **1703**, 141-147

37. Stites, W. E., and Froude, J. W. (2007) Does the oxidation of methionine in thrombomodulin contribute to the hypercoaguable state of smokers and diabetics? *Med Hypotheses* **68**, 811-821

38. Xu, X., Bishop, E. E., Kennedy, S. M., Simpson, S. A., and Pechacek, T. F. (2015) Annual healthcare spending attributable to cigarette smoking: an update. *Am J Prev Med* **48**, 326-333

39. Gadek, J. E., Fells, G. A., and Crystal, R. G. (1979) Cigarette smoking induces functional antiprotease deficiency in the lower respiratory tract of humans. *Science* **206**, 1315-1316

40. Taggart, C. (2000) Oxidation of either methionine 351 or methionine 358 in α1-antitrypsin causes loss of anti-neutrophil elastase activity. (Cervantes-Laurean, D., Kim,, and G., M., N.G., Wehr, N., Moss, J. and Levine, R.L. eds.), *Journal of Biological Chemistry*

41. Turell, L., Radi, R., and Alvarez, B. (2013) The thiol pool in human plasma: the central contribution of albumin to redox processes. *Free Radic Biol Med* **65**, 244-253

42. Roche, M., Rondeau, P.,, and Singh, N. R., Tarnus, E., & Bourdon, E. (2008) The antioxidant properties of serum albumin. *FEBS letters*

43. Sen, K. I., Hepler, R., and Nanda, H. (2017) Detection and Measurement of Methionine Oxidation in Proteins. *Curr Protoc Protein Sci* **87**, 14.16.11-14.16.11

44. Verrastro, I., Pasha, S., Jensen, K. T., Pitt, A. R., and Spickett, C. M. (2015) Mass spectrometry-based methods for identifying oxidized proteins in disease: advances and challenges. *Biomolecules* **5**, 378-411

45. Spickett, C. M., and Pitt, A. R. (2012) Protein oxidation: role in signalling and detection by mass spectrometry. *Amino Acids* **42**, 5-21

46. Kalantari, S., Jafari, A., Moradpoor, R., Ghasemi, E., and Khalkhal, E. (2015) Human Urine Proteomics: Analytical Techniques and Clinical Applications in Renal Diseases. *Int J Proteomics* **2015**, 782798

47. Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V., and Mann, M. (2006) The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol* **7**, R80

48. Martin, H. (2011) Laboratory measurement of urine albumin and urine total protein in screening for proteinuria in chronic kidney disease. *Clin Biochem Rev* **32**, 97-102

49. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567

50. UniProt Consortium, T. (2018) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **46**, 2699

51. MacLean, B., Tomazela, D., Shulman, N., Chambers, M., Finney, G., Frewen, B., Kern, R., Tabb, D., Liebler, D., and MacCoss, M. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968

52. Krokhin, O. V., and Spicer, V. (2010) Predicting peptide retention times for proteomics. *Curr Protoc Bioinformatics* **Chapter 13**, Unit 13.14

53. Wu, C. C., and MacCoss, M. J. (2002) Shotgun proteomics: tools for the analysis of complex biological systems. *Curr Opin Mol Ther* **4**, 242-250

54. Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S. C., and Haynes, P. A. (2011) Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* **11**, 535-553

55. Nilsson, T., Matthias Mann, Ruedi Aebersold, John R Yates, Amos Bairoch & John J M Bergeron. (2010) Mass spectrometry in high-throughput proteomics: ready for the big time. *NATURE METHODS*

56. Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., Variyath, A. M., Ham, A. J., Bunk, D. M., Kilpatrick, L. E., Billheimer, D. D., Blackman, R. K., Cardasis, H. L., Carr, S. A., Clauser, K. R., Jaffe, J. D., Kowalski, K. A., Neubert, T. A., Regnier, F. E., Schilling, B., Tegeler, T. J., Wang, M., Wang, P., Whiteaker, J. R., Zimmerman, L. J., Fisher, S. J., Gibson, B. W., Kinsinger, C. R., Mesri, M., Rodriguez, H., Stein, S. E., Tempst, P., Paulovich, A. G., Liebler, D. C., and Spiegelman, C. (2010) Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res* **9**, 761-776

57. Bateman, N. W., Goulding, S. P., Shulman, N. J., Gadok, A. K., Szumlinski, K. K., MacCoss, M. J., and Wu, C. C. (2014) Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). *Mol Cell Proteomics* **13**, 329-338

58. Zürbig, P., Decramer, S., Dakna, M., Jantos, J., Good, D. M., Coon, J. J., Bandin, F., Mischak, H., Bascands, J. L., and Schanstra, J. P. (2009) The human urinary proteome reveals high similarity between kidney aging and chronic kidney disease. *Proteomics* **9**, 2108-2117

59. Siwy, J., Zürbig, P., Argiles, A., Beige, J., Haubitz, M., Jankowski, J., Julian, B. A., Linde, P. G., Marx, D., Mischak, H., Mullen, W., Novak, J., Ortiz, A., Persson, F., Pontillo, C., Rossing, P., Rupprecht, H., Schanstra, J. P., Vlahou, A., and Vanholder, R. (2017) Noninvasive diagnosis of chronic kidney diseases using urinary proteome analysis. *Nephrol Dial Transplant* **32**, 2079-2089

60. Qi, X. L., Xiu, J., Shan, K. R., Xiao, Y., Gu, R., Liu, R. Y., and Guan, Z. Z. (2005) Oxidative stress induced by beta-amyloid peptide(1-42) is involved in the altered composition of cellular membrane lipids and the decreased expression of nicotinic receptors in human SH-SY5Y neuroblastoma cells. *Neurochem Int* **46**, 613-621

61. Butterfield, D. A., and Boyd-Kimball, D. (2005) The critical role of methionine 35 in Alzheimer's amyloid beta-peptide (1-42)-induced oxidative stress and neurotoxicity. *Biochim Biophys Acta* **1703**, 149-156

62. Long, L. H., Wu, P. F., Guan, X. L., Zhang, J. Q., Jin, Y., Zhang, Z., Wang, Y., Li, Y. Y., Chen, J. G., and Wang, F. (2011) Determination of protein-bound methionine oxidation in the hippocampus of adult and old rats by LC-ESI-ITMS method after microwave-assisted proteolysis. *Anal Bioanal Chem* **399**, 2267-2274

63. Madeira, J. P., Alpha-Bazin, B., Armengaud, J., and Duport, C. (2018) Time-course proteomics dataset to monitor protein-bound methionine oxidation. *Data Brief* **18**, 394-398

64. Lai, C. H., Lee, C. N., Bai, K. J., Yang, Y. L., Chuang, K. J., Wu, S. M., and Chuang, H. C. (2016) Protein oxidation and degradation caused by particulate matter. *Sci Rep* **6**, 33727

65. Zhou, B., Liu, J., Wang, Z. M., and Xi, T. (2012) C-reactive protein, interleukin 6 and lung cancer risk: a meta-analysis. *PLoS One* **7**, e43075

66. Takada, K., Okamoto, T., Shoji, F., Shimokawa, M., Akamine, T., Takamori, S., Katsura, M., Suzuki, Y., Fujishita, T., Toyokawa, G., Morodomi, Y., Okano, S., Oda, Y., and Maehara, Y. (2016) Clinical Significance of PD-L1 Protein Expression in Surgically Resected Primary Lung Adenocarcinoma. *J Thorac Oncol* **11**, 1879-1890

67. Agarwal, R. (2005) On the nature of proteinuria with acute renal injury in patients with chronic kidney disease. *Am J Physiol Renal Physiol* **288**, F265-271

68. Panda, K., Chattopadhyay, R., Ghosh, M. K., Chattopadhyay, D. J., and Chatterjee, I. B. (1999) Vitamin C prevents cigarette smoke induced oxidative damage of proteins and increased proteolysis. *Free Radic Biol Med* **27**, 1064-1079

69. Petersen, A. M., Magkos, F., Atherton, P., Selby, A., Smith, K., Rennie, M. J., Pedersen, B. K., and Mittendorfer, B. (2007) Smoking impairs muscle protein synthesis and increases the expression of myostatin and MAFbx in muscle. *Am J Physiol Endocrinol Metab* **293**, E843-848

70. Li, C. J., Liu, Y., Chen, Y., Yu, D., Williams, K. J., and Liu, M. L. (2013) Novel proteolytic microvesicles released from human macrophages after exposure to tobacco smoke. *Am J Pathol* **182**, 1552-1562

3.7 SUPPORTING INFORMATION

Table S1: Proteins with MSO identified with significant fold changes in cigarette smokers

Protein Acc	Protein Description	Fold Change	P-Value
P07911	Uromodulin OS=Homo sapiens GN=UMOD PE=1 SV=1	2.93	0
F6KPG5	Albumin (Fragment) OS=Homo sapiens PE=2 SV=1	3.48	0
Q92820	Gamma-glutamyl hydrolase OS=Homo sapiens GN=GGH PE=1 SV=2	2.01	0.0003
Q7Z3Y9	Keratin, type I cytoskeletal 26 OS=Homo sapiens GN=KRT26 PE=1 SV=2	2.45	0.0003
B3KVV6	cDNA FLJ41607, highly similar to Homo sapiens alpha-2-macroglobulin-like 1 (A2ML1) PE=2 SV=1	2.24	0.0003
Q16610	Extracellular matrix protein 1 OS=Homo sapiens GN=ECM1 PE=1 SV=2	2.36	0.0004
Q6P1J6	Phospholipase B1, membrane-associated OS=Homo sapiens GN=PLB1 PE=1 SV=3	2.11	0.0005
P08571	Monocyte differentiation antigen CD14 OS=Homo sapiens GN=CD14 PE=1 SV=2	2.21	0.0006
P68871	Hemoglobin subunit beta OS=Homo sapiens GN=HBB PE=1 SV=2	2.07	0.0008
P02790	Hemopexin OS=Homo sapiens GN=HPX PE=1 SV=2	2.09	0.0014
P12830	Cadherin-1 OS=Homo sapiens GN=CDH1 PE=1 SV=3	2.36	0.0017
P05155	Plasma protease C1 inhibitor OS=Homo sapiens GN=SERPING1 PE=1 SV=2	1.71	0.0017
P15309	Prostatic acid phosphatase OS=Homo sapiens GN=ACPP PE=1 SV=3	1.82	0.0026
P11142	Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1	2.52	0.003
O43707	Alpha-actinin-4 OS=Homo sapiens GN=ACTN4 PE=1 SV=2	1.62	0.0037
P30530	Tyrosine-protein kinase receptor UFO OS=Homo sapiens GN=AXL PE=1 SV=3	2.09	0.0037
P19652	Alpha-1-acid glycoprotein 2 OS=Homo sapiens GN=ORM2 PE=1 SV=2	1.93	0.0042
P19835	Bile salt-activated lipase OS=Homo sapiens GN=CEL PE=1 SV=3	1.78	0.0042
Q7Z3Y8	Keratin, type I cytoskeletal 27 OS=Homo sapiens GN=KRT27 PE=1 SV=2	1.88	0.0042
P01617	Ig kappa chain V-II region TEW OS=Homo sapiens PE=1 SV=1	2.17	0.0042
P01625	Ig kappa chain V-IV region Len OS=Homo sapiens PE=1 SV=2	2.13	0.0042
O00560	Syntenin-1 OS=Homo sapiens GN=SDCBP PE=1 SV=1	1.96	0.0042
Q9NS89	Alpha1A-voltage-dependent calcium channel (Fragment) OS=Homo sapiens GN=CACNA1A PE=2 SV=1	115.58	0.0042
P05090	Apolipoprotein D OS=Homo sapiens GN=APOD PE=1 SV=1	2.36	0.0055
P24855	Deoxyribonuclease-1 OS=Homo sapiens GN=DNASE1 PE=1 SV=1	2.12	0.0055

Protein Acc	Protein Description	Fold Change	P-Value
Q03001	Dystonin OS=Homo sapiens GN=DST PE=1 SV=4	2.04	0.0055
P09668	Pro-cathepsin H OS=Homo sapiens GN=CTSH PE=1 SV=4	2.09	0.0058
Q9NQ84	G-protein coupled receptor family C group 5 member C OS=Homo sapiens GN=GPRC5C PE=1 SV=2	1.82	0.0075
B4DPR2	cDNA FLJ50830, highly similar to Serum albumin OS=Homo sapiens PE=2 SV=1	1.84	0.0075
P01011	Alpha-1-antichymotrypsin OS=Homo sapiens GN=SERPINA3 PE=1 SV=2	1.6	0.008
P04745	Alpha-amylase 1 OS=Homo sapiens GN=AMY1A PE=1 SV=2	1.64	0.008
P00450	Ceruloplasmin OS=Homo sapiens GN=CP PE=1 SV=1	1.63	0.008
P09211	Glutathione S-transferase P OS=Homo sapiens GN=GSTP1 PE=1 SV=2	1.77	0.008
P01023	Alpha-2-macroglobulin OS=Homo sapiens GN=A2M PE=1 SV=3	1.66	0.0091
P02768	Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2	1.96	0.0091
P60709	Actin, cytoplasmic 1 OS=Homo sapiens GN=ACTB PE=1 SV=1	1.78	0.0094
P19022	Cadherin-2 OS=Homo sapiens GN=CDH2 PE=1 SV=4	1.97	0.0094
Q96PD5	N-acetylmuramoyl-L-alanine amidase OS=Homo sapiens GN=PGLYRP2 PE=1 SV=1	1.71	0.0094
P0DMV8	Heat shock 70 kDa protein 1A OS=Homo sapiens GN=HSPA1A PE=1 SV=1	1.77	0.0098
Q99698	Lysosomal-trafficking regulator OS=Homo sapiens GN=LYST PE=1 SV=3	1.79	0.0102
C8C504	Beta-globin OS=Homo sapiens GN=HBB PE=3 SV=1	1.72	0.0108
P29375	Lysine-specific demethylase 5A OS=Homo sapiens GN=KDM5A PE=1 SV=3	1.61	0.0119
A8K2U0	Alpha-2-macroglobulin-like protein 1 OS=Homo sapiens GN=A2ML1 PE=1 SV=3	1.5	0.0126
R4H484	X-DING-CD4 (Fragment) OS=Homo sapiens PE=2 SV=1	1.85	0.0126
P01034	Cystatin-C OS=Homo sapiens GN=CST3 PE=1 SV=1	1.66	0.0132
Q71DI3	Histone H3.2 OS=Homo sapiens GN=HIST2H3A PE=1 SV=3	1.69	0.0132
Q5JSZ5	Protein PRRC2B OS=Homo sapiens GN=PRRC2B PE=1 SV=2	1.9	0.0132
P16444	Dipeptidase 1 OS=Homo sapiens GN=DPEP1 PE=1 SV=3	1.75	0.0164
O43451	Maltase-glucoamylase, intestinal OS=Homo sapiens GN=MGAM PE=1 SV=5	1.61	0.0164
P17900	Ganglioside GM2 activator OS=Homo sapiens GN=GM2A PE=1 SV=4	1.81	0.0164
L8E9E6	Alternative protein SOD1 OS=Homo sapiens GN=SOD1 PE=4 SV=1	1.64	0.0189
P07339	Cathepsin D OS=Homo sapiens GN=CTSD PE=1 SV=1	1.61	0.0205
P02760	Protein AMBP OS=Homo sapiens GN=AMBP PE=1 SV=1	1.81	0.0206
P15311	Ezrin OS=Homo sapiens GN=EZR PE=1 SV=4	1.6	0.0255

Protein Acc	Protein Description	Fold Change	P-Value
Q99456	Keratin, type I cytoskeletal 12 OS=Homo sapiens GN=KRT12 PE=1 SV=1	1.67	0.0255
P02751	Fibronectin OS=Homo sapiens GN=FN1 PE=1 SV=4	1.45	0.0261
P05067	Amyloid beta A4 protein OS=Homo sapiens GN=APP PE=1 SV=3	1.73	0.0271
P51654	Glypican-3 OS=Homo sapiens GN=GPC3 PE=1 SV=1	1.48	0.0279
P13645	Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6	1.5	0.0281
P02042	Hemoglobin subunit delta OS=Homo sapiens GN=HBD PE=1 SV=2	1.63	0.0287
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein OS=Homo sapiens GN=HSPG2 PE=1 SV=4	1.37	0.0307
P02787	Serotransferrin OS=Homo sapiens GN=TF PE=1 SV=3	1.46	0.0359
P04264	Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=6	1.44	0.037
P01859	Ig gamma-2 chain C region OS=Homo sapiens GN=IGHG2 PE=1 SV=2	1.59	0.0376
P54107	Cysteine-rich secretory protein 1 OS=Homo sapiens GN=CRISP1 PE=1 SV=1	1.63	0.0393
P98164	Low-density lipoprotein receptor-related protein 2 OS=Homo sapiens GN=LRP2 PE=1 SV=3	1.44	0.0405
P05154	Plasma serine protease inhibitor OS=Homo sapiens GN=SERPINA5 PE=1 SV=3	1.48	0.0439
P01009	Alpha-1-antitrypsin OS=Homo sapiens GN=SERPINA1 PE=1 SV=3	1.42	0.0447
P08473	Neprilysin OS=Homo sapiens GN=MME PE=1 SV=2	1.59	0.0447
P15144	Aminopeptidase N OS=Homo sapiens GN=ANPEP PE=1 SV=4	1.47	0.0456
Q9HC84	Mucin-5B OS=Homo sapiens GN=MUC5B PE=1 SV=3	1.48	0.047
E9PGN7	Plasma protease C1 inhibitor OS=Homo sapiens GN=SERPING1 PE=1 SV=1	1.54	0.047
Q6MZV7	Putative uncharacterized protein DKFZp686C11235 OS=Homo sapiens GN=DKFZp686C11235 PE=2 SV=1	1.73	0.047

Table S2: Tryptic peptides identified with MSO in the current study

Protein Acc#	Gene	Peptide Modified Sequence
Q8WZ42	TTN	DGGSPIKGYIVEM[+16]QEEGTTDWK
P05067	APP	AM[+16]ISR
P68871	HBB	FFESFGDLSTPDAVM[+16]GNPK
O95336	PGLS	FALGLSGGSLVSM[+16]LAR
P01023	A2M	VVSM[+16]DENFHPLNELIPLVYIQDPK
A8K2U0	A2ML1	YSM[+16]VELQDPNSNR
A8K2U0	A2ML1	YTYGKPM[+16]LGAVQVSVC[+57]QK
A8K2U0	A2ML1	FQM[+16]EDLVYNPEQVPR
A8K2U0	A2ML1	GSLVM[+16]EGQK
A8K2U0	A2ML1	WM[+16]AGNQLPSGC[+57]YANVGNLLHTAMK
A8K2U0	A2ML1	DVDDPM[+16]VSQGLR
A8K2U0	A2ML1	YATTAYM[+16]PSEEINLVVK
A8K2U0	A2ML1	M[+16]LSGFSPMEGTNQLLLQQPLVK
P68032	ACTC1	DLTDYLM[+16]K
P68032	ACTC1	LC[+57]YVALDFENEM[+16]ATAASSSSLEK
P01011	SERPINA3	DLDSQTMM[+16]VLVNYIFFK
P01011	SERPINA3	AKWEM[+16]PFDPQDTHQSR
P01011	SERPINA3	WEM[+16]PFDPQDTHQSR
P01011	SERPINA3	M[+16]EEVEAMLLPETLK
P01011	SERPINA3	MEEVEAM[+16]LLPETLK
P01011	SERPINA3	FNRPFLMIIVPTDTQNIFFM[+16]SK
P12821	ACE	SM[+16]YETPSLEQDLER
P19652	ORM2	TLM[+16]FGSYLDDEKNWGLSFYADKPETTK
P60709	ACTB	TTGIVM[+16]DSGDGVTHTVPIYEGYALPHAILR
P60709	ACTB	DLYANTVLSGGTTM[+16]YPGIADR
P01009	SERPINA1	LGM[+16]FNIQHC[+57]K
P01009	SERPINA1	GTEAAGAM[+16]FLEAIPMSIPPEVK
P01009	SERPINA1	GTEAAGAMFLEAIPM[+16]SIPPEVK
P01009	SERPINA1	FNKPFVFLM[+16]IEQNTK
O43707	ACTN4	TIQEM[+16]QQK
O43707	ACTN4	ISIEM[+16]NGTLEDQLSHLK
O43707	ACTN4	GISQEQM[+16]QEFR
O43707	ACTN4	M[+16]APYQGPDAVPGALDYK
P15144	ANPEP	DSQYEM[+16]DSEFEGELADDLAGFYR
P15144	ANPEP	SEYM[+16]EGNVR
P15144	ANPEP	KVVATTQM[+16]QAADAR
P15144	ANPEP	VVATTQM[+16]QAADAR
P15144	ANPEP	DLTALSNM[+16]LPK
P15144	ANPEP	DLM[+16]VLNDVYR
P15144	ANPEP	VM[+16]AVDALASSHPLSTPASEINTPAQISELFDAISYSK

Protein Acc#	Gene	Peptide Modified Sequence
P15144	ANPEP	M[+16]LSSFLSEDVFK
P15144	ANPEP	QYM[+16]PWEAALSSLSYFK
P04745	AMY1A	SGNEDEFRNM[+16]VTR
P04745	AMY1A	IYVDAVINHM[+16]C[+57]GNAVSAGTSSTC[+57]GSYFNPGSR
P04745	AMY1A	IAEYM[+16]NHLIDIGVAGFR
P04745	AMY1A	NWGEGWGFM[+16]PSDR
P04745	AMY1A	M[+16]AVGFMLAHPYGFTR
P04745	AMY1A	MAVGFM[+16]LAHPYGFTR
Q9H2P0	ADNP	SLPSQQM[+16]VNR
Q12802	AKAP13	IMSGVYSQGM[+16]M[+16]ADLLFEQQM[+16]VEK
P54802	NAGLU	SFGM[+16]TPVLPAFAGHVPEAVTR
P02768	ALB	ETYGEM[+16]ADC[+57]C[+57]AK
P02768	ALB	SHC[+57]IAEVENDEM[+16]PADLPSLAADFVESK
P02768	ALB	SHC[+57]IAEVENDEM[+16]PADLPSLAADFVESKDVC[+57]K
P02768	ALB	RM[+16]PC[+57]AEDYLSVVLNQLC[+57]VLHEK
P02768	ALB	M[+16]PC[+57]AEDYLSVVLNQLC[+57]VLHEK
P02768	ALB	M[+16]PC[+57]AEDYLSVVLNQLC[+57]VLHEKTPVSDR
P02768	ALB	AVM[+16]DDFAAFVEK
P05062	ALDOB	GILAADESVGTM[+16]GNR
P02760	AMBP	M[+16]TVSTLVLGEGATEAEISMTSTR
P02760	AMBP	MTVSTLVLGEGATEAEISM[+16]TSTR
P02760	AMBP	M[+16]TVSTLVLGEGATEAEISM[+16]TSTR
P02760	AMBP	ETLLQDFRVVAQGVGIPEDSIFTM[+16]ADR
P02760	AMBP	VVAQGVGIPEDSIFTM[+16]ADR
P02760	AMBP	VVAQGVGIPEDSIFTM[+16]ADRGEC[+57]VPGEQEPEPILIPR
P02760	AMBP	KEDSC[+57]QLGYSAGPC[+57]M[+16]GMTSR
P02760	AMBP	KEDSC[+57]QLGYSAGPC[+57]MGM[+16]TSR
P02760	AMBP	EDSC[+57]QLGYSAGPC[+57]M[+16]GMTSR
P02760	AMBP	EDSC[+57]QLGYSAGPC[+57]MGM[+16]TSR
P08519	LPA	TPEYYPNAGLIM[+16]NYC[+57]R
P63010	AP2B1	SQPDM[+16]AIM[+16]AVNSFVK
P05090	APOD	M[+16]TVTDQVNC[+57]PK
Q01814	ATP2B2	TGTLTTNRM[+16]TVVQAYVGDVHYK
P50895	BCAM	LEVPVEMNPEGYM[+16]TSR
O75882	ATRN	IMQSSQSM[+16]SK
P16278	GLB1	AGATLDLLVENM[+16]GR
Q15878	CACNA1E	AWIDKAEEVM[+16]LAEENK
Q8N4F0	BPIFB2	HVGTEGSM[+16]ATVGLSQQLFDSALLLLQK
P12830	CDH1	M[+16]ALEVGDYK
P55290	CDH13	M[+16]TAFDADDPATDNALLR
P19022	CDH2	M[+16]FVLTVAAENQVPLAK

Protein Acc#	Gene	Peptide Modified Sequence
Q9BY67	CADM1	EGDALELTC[+57]EAIGKPQPVM[+16]VTWVR
P07384	CAPN1	DM[+16]ETIGFAVYEVPPELVGQPAVHLK
P31944	CASP14	EGSEEDLDALEHM[+16]FR
P31944	CASP14	M[+16]AEAELVQEGK
P07339	CTSD	TM[+16]SEVGGSVEDLIAK
P07339	CTSD	FDGILGM[+16]AYPR
P07339	CTSD	ISVNNVLPVFDNLM[+16]QQK
P07339	CTSD	DPDAQPGGELM[+16]LGGTDSK
P07339	CTSD	AIGAVPLIQGEYM[+16]IPC[+57]EK
P08185	SERPINA6	NIFISPVSISMALAM[+16]LSLGTC[+57]GHTR
P08571	CD14	ITGTM[+16]PPLPLEATGLALSSLR
P09668	CTSH	GIM[+16]GEDTYPYQGK
Q49A88	CCDC14	GPQNSNTRGM[+16]EEASAPGIISALSK
P00450	CP	M[+16]YSVNGYTFGSLPGLSMC[+57]AEDR
P00450	CP	M[+16]YYSAVDPTK
P00450	CP	M[+16]YYSAVDPTKDIFTGLIGPMK
P00450	CP	MYYSAVDPTKDIFTGLIGPM[+16]K
P00450	CP	M[+16]YYSAVDPTKDIFTGLIGPM[+16]K
P00450	CP	DIFTGLIGPM[+16]K
P00450	CP	M[+16]FTTAPDQVDKEDEDFQESNK
P19835	CEL	AM[+16]IAYWTNFAK
P01024	C3	TVM[+16]VNIENPEGIPVK
P01024	C3	YFKPGMPFDLM[+16]VFVTNPDGSPAYR
P01024	C3	ILLQGTPVAQM[+16]TEDAVDAER
P0C0L4	C4A	LNM[+16]GITDLQGLR
P0C0L4	C4A	SM[+16]QGGLVGNDETVALTAFVTIALHHGLAVFQDEGAEPLK
Q12860	CNTN1	M[+16]NNGDVDLTSDR
Q12860	CNTN1	ILALAPTFEM[+16]NPM[+16]K
P10643	C7	M[+16]PYEC[+57]GPSLDVC[+57]AQDER
P22792	CPN2	SLM[+16]LSYNAITHLPAGIFR
P54107	CRISP1	M[+16]SWSEEAAQNAR
P54107	CRISP1	YC[+57]DM[+16]TESNPLER
P39059	COL15A1	SSQALAFESSAGIFM[+16]GNAGATGLER
Q7Z3J2	C16orf62	AAISLVPEVPKM[+16]INIDGK
P12109	COL6A1	FEPGQSYAGVVQYSHSQM[+16]QEHVSLR
P01034	CST3	LVGGPM[+16]DASVEEEGVR
P01034	CST3	ASNDM[+16]YHSR
P01034	CST3	AFC[+57]SFQIYAVPWQGTM[+16]TLSK
O60494	CUBN	SPENPM[+16]QVSSTGNELAIR
O60494	CUBN	FISDGSGSGTGFQATFM[+16]K
O60494	CUBN	ILEM[+16]DIEEIQNC[+57]YYDK

Protein Acc#	Gene	Peptide Modified Sequence
O60494	CUBN	STGEYM[+16]FIR
O60494	CUBN	FC[+57]GTSLPSSQWSSGEVM[+16]YLR
O60494	CUBN	FESSM[+16]EEC[+57]GGDLQGSIGTFTSPNYPNPNPHGR
O60494	CUBN	RITLM[+16]FNNLR
O60494	CUBN	ITLM[+16]FNNLR
O60494	CUBN	VGDADGPLM[+16]WR
O60494	CUBN	M[+16]SFTYQIADC[+57]NR
Q15828	CST6	AAQAAVASYNM[+16]GSNSIYYFR
P01040	CSTA	M[+16]IPGGLSEAKPATPEIQEIVDK
Q07507	DPT	AGM[+16]EWYQTC[+57]SNNGLVAGFQSR
P24855	DNASE1	GAVVPDSALPFNFQAAYGLSDQLAQAISDHYPVEVM[+16]LK
P16444	DPEP1	DSPVIDGHNDLPWQLLDM[+16]FNNR
Q03001	DST	VEPQLAEDQPVHGDIDLVM[+16]NLIDNHK
Q02413	DSG1	ALNSM[+16]GQDLERPLELR
Q02413	DSG1	QEPSDSPM[+16]FIINR
Q02413	DSG1	YVM[+16]GNNPADLLAVDSR
Q02413	DSG1	EM[+16]QDLGGGER
Q8WXX0	DNAH7	AM[+16]EKM[+16]ITEWDAVEFVIHSYR
Q16610	ECM1	M[+16]GTTARAALVLTYLAVASAASEGGFTATGQR
P01133	EGF	HPTQHNLFAM[+16]SLFGDR
P01133	EGF	LC[+57]SDIDEC[+57]EM[+16]GVPVC[+57]PPASSK
P08246	ELANE	LGNGVQC[+57]LAM[+16]GWGLLGR
P15311	EZR	IAQDLEM[+16]YGINYFEIK
Q12805	EFEMP1	TSSYLC[+57]QYQC[+57]VNEPGKFSC[+57]M[+16]C[+57]PQGYQVVR
Q12805	EFEMP1	FSC[+57]M[+16]C[+57]PQGYQVVR
Q12805	EFEMP1	C[+57]VC[+57]PVSNAM[+16]C[+57]R
Q12805	EFEMP1	QTSPVSAM[+16]LVLVK
P02751	FN1	VTIM[+16]WTPPESAVTGYR
P02751	FN1	FTNIGPDTM[+16]R
P02751	FN1	TEIDKPSQM[+16]QVTDVQDNSISVK
P02751	FN1	EINLAPDSSSVVVSGLM[+16]VATK
P15328	FOLR1	TELLNVC[+57]M[+16]NAK
Q9ULV1	FZD4	INIPIGPC[+57]GGM[+16]C[+57]LSVK
P06396	GSN	VPFDAATLHTSTAM[+16]AAQHGM[+16]DDDGTGQK
Q92820	GGH	M[+16]FQNFPTELLLSLAVEPLTANFHK
P15586	GNS	SM[+16]C[+57]GYQTFFAGK
P15586	GNS	SNFEPFFMM[+16]IATPAPHSPWTAAPQYQK
P15586	GNS	LM[+16]MLQSC[+57]SGPTC[+57]R
P51654	GPC3	IYDM[+16]ENVLLGLFSTIHDSIQYVQK
Q9NZH0	GPRC5B	M[+16]RETAFEEDVQLPR
Q9NZH0	GPRC5B	SNVYQPTEM[+16]AVVLNGGTIPTAPPSHTGR

Protein Acc#	Gene	Peptide Modified Sequence
P28799	GRN	C[+57]DM[+16]EVSC[+57]PDGYTC[+57]C[+57]R
Q9NQ84	GPRC5C	SSPEQSYQGDM[+16]YPTR
Q9NQ84	GPRC5C	AFSM[+16]DEPVAAK
P09211	GSTP1	DQQEAALVDM[+16]VNDGVEDLR
P68431	HIST1H3A	FQSSAVM[+16]ALQEAC[+57]EAYLVGLFEDTNLC[+57]AIHAK
Q71DI3	HIST2H3A	FQSSAVM[+16]ALQEASEAYLVGLFEDTNLC[+57]AIHAK
P02042	HBD	FFESFGDLSSPDAVM[+16]GNPK
P69905	HBA1	M[+16]FLSFPTTK
P69905	HBA1	VADALTNAVAHVDDM[+16]PNALSALSDLHAHK
P02790	HPX	DYFM[+16]PC[+57]PGR
P00738	HP	YVM[+16]LPVADQDQC[+57]IR
P00738	HP	SPVGVQPILNEHTFC[+57]AGM[+16]SK
P0DMV8	HSPA1A	SINPDEAVAYGAAVQAAILM[+16]GDK
P11142	HSPA8	NQVAM[+16]NPTNTVFDAK
P11142	HSPA8	RFDDAVVQSDM[+16]K
P01860	IGHG3	DTLM[+16]ISR
P30740	SERPINB1	IPELLASGMVDNM[+16]TK
P30740	SERPINB1	VLELPYQGEELSM[+16]VILLPDDIEDESTGLK
P30740	SERPINB1	ADLSGM[+16]SGAR
P05155	SERPING1	KVETNM[+16]AFSPFSIASLLTQVLLGAGENTK
P05155	SERPING1	VETNM[+16]AFSPFSIASLLTQVLLGAGENTK
P05155	SERPING1	VPM[+16]MNSK
P05155	SERPING1	LEDM[+16]EQALSPSVFK
P05155	SERPING1	VTTSQDM[+16]LSIMEK
P05155	SERPING1	FPVFM[+16]GR
Q12794	HYAL1	ALYPSIYM[+16]PAVLEGTGK
Q16270	IGFBP7	GYC[+57]APGM[+16]EC[+57]VK
P01859	IGHG2	TTPPM[+16]LDSDGSFFLYSK
P05154	SERPINA5	ALASAAPSQSIFFSPVSISM[+16]SLAMLSLGAGSSTK
P05154	SERPINA5	ALASAAPSQSIFFSPVSISMSLAM[+16]LSLGAGSSTK
P05154	SERPINA5	ALASAAPSQSIFFSPVSISM[+16]SLAM[+16]LSLGAGSSTK
P05154	SERPINA5	M[+16]QILEGLGLNLQK
P05154	SERPINA5	DGFQLSLGNALFTDLVVDLQDTFVSAM[+16]K
P05154	SERPINA5	TLYLADTFPTNFRDSAGAM[+16]K
P05154	SERPINA5	NLDSNAVVIM[+16]VNYIFFK
P05154	SERPINA5	M[+16]QQVENGLSEK
Q7Z3Y9	KRT26	VTM[+16]QNLNDR
P35908	KRT2	DVDNAYM[+16]IK
P35908	KRT2	VLYDAEISQIHQSVTDTNVILSM[+16]DNSR
P35908	KRT2	DYQELM[+16]NVK
Q14624	ITIH4	ETLFSVM[+16]PGLK

Protein Acc#	Gene	Peptide Modified Sequence
P13645	KRT10	NHEEEM[+16]KDLR
P13645	KRT10	NVSTGDVNVEM[+16]NAAPGVDLTQLLNNMR
P13645	KRT10	NVSTGDVNVEMNAAPGVDLTQLLNNM[+16]R
P13645	KRT10	NVSTGDVNVEM[+16]NAAPGVDLTQLLNNM[+16]R
P02533	KRT14	GQVGGDVNVEM[+16]DAAPGVDLSR
Q7Z3Y8	KRT27	VTM[+16]QNLNDRLASYLENVR
Q99456	KRT12	TDLEM[+16]QIESLNEELAYMK
P13646	KRT13	TLQGLEIELQSQLSM[+16]K
P08779	KRT16	GQTGGDVNVEM[+16]DAAPGVDLSR
P19013	KRT4	VLYDAELSQMQTHVSDTSVVLSM[+16]DNNR
P13647	KRT5	NM[+16]QDLVEDFK
P13647	KRT5	DVDAAYM[+16]NK
P13647	KRT5	EYQELM[+16]NTK
P02538	KRT6A	GM[+16]QDLVEDFK
P35527	KRT9	STM[+16]QELNSR
P35527	KRT9	M[+16]TLDDFR
P35527	KRT9	QVLDNLTM[+16]EK
P35527	KRT9	SDLEM[+16]QYETLQEELMALK
P35527	KRT9	SDLEMQYETLQEELM[+16]ALK
P35527	KRT9	SDLEM[+16]QYETLQEELM[+16]ALK
P35527	KRT9	SDLEM[+16]QYETLQEELMALKK
P35527	KRT9	SDLEMQYETLQEELM[+16]ALKK
P35527	KRT9	SDLEM[+16]QYETLQEELM[+16]ALKK
P35527	KRT9	TLNDM[+16]RQEYEQLIAK
P04264	KRT1	NM[+16]QDMVEDYR
P04264	KRT1	NMQDM[+16]VEDYR
P04264	KRT1	KDVDGAYM[+16]TK
P04264	KRT1	DVDGAYM[+16]TK
P04264	KRT1	LDNLQQEIDFLTALYQAELSQM[+16]QTQISETNVILSM[+16]DNNR
P04264	KRT1	DYQELM[+16]NTK
P04264	KRT1	M[+16]SGEC[+57]APNVSVSVSTSHTTISGGGSR
P29375	KDM5A	ILYPYELFQSGVSLM[+16]GVQM[+16]PNLDLKEK
P07288	KLK3	FLRPGDDSSHDLM[+16]LLR
P07288	KLK3	VM[+16]DLPTQEPALGTTC[+57]YASGWGSIEPEEFLTPK
P07288	KLK3	FM[+16]LC[+57]AGR
Q9UKR3	KLK13	M[+16]WPLALVIASLTLALSGGVSQESSK
P01596		DIQM[+16]TQSPSTLSASVGDR
P01600		DIQM[+16]TQSPSSLSASVGDR
O95835	LATS1	VGLSQDAQDQM[+16]RK
P01611		DIQM[+16]TQSPSSVSASVGDR
P01612		DVQM[+16]TQSPSSLSASVGDR

Protein Acc#	Gene	Peptide Modified Sequence
P01042	KNG1	IYPTVNC[+57]QPLGM[+16]ISLMK
P01042	KNG1	IYPTVNC[+57]QPLGMISLM[+16]K
P06870	KLK1	QADEDYSHDLM[+16]LLR
P06870	KLK1	VTDFM[+16]LC[+57]VGHLEGGK
P01617		DIVM[+16]TQSPLSLPVTPGEPASISC[+57]R
P01613		DIQM[+16]TQSPSSLSATVGDR
P01625		DIVM[+16]TQSPDSLAVSLGER
Q08380	LGALS3BP	AAFGQGSGPIM[+16]LDEVQC[+57]TGTEASLADC[+57]K
Q12907	LMAN2	LPTGYYFGASAGTGDLSDNHDIISM[+16]K
Q12907	LMAN2	LFQLM[+16]VEHTPDEESIDWTK
Q9NZU5	LMCD1	QLM[+16]HQLPIYDQDPSR
P61626	LYZ	LGM[+16]DGYR
P10253	GAA	LDVM[+16]METENR
P10253	GAA	LDVMM[+16]ETENR
P10253	GAA	AHFPLDVQWNDLDYM[+16]DSR
P10253	GAA	DGFRDFPAM[+16]VQELHQGGR
P10253	GAA	DFPAM[+16]VQELHQGGR
P10253	GAA	YM[+16]MIVDPAISSSGPAGSYRPYDEGLR
P10253	GAA	YM[+16]M[+16]IVDPAISSSGPAGSYRPYDEGLR
P10253	GAA	WTQLGAFYPFM[+16]R
P10253	GAA	NHNSLLSLPQEPYSFSEPAQQAM[+16]R
P98164	LRP2	IDM[+16]VNLDGSYR
P98164	LRP2	AFM[+16]DGSNRK
P98164	LRP2	LGWPAGVTLDM[+16]ISK
P98164	LRP2	VC[+57]GC[+57]PYGM[+16]R
P98164	LRP2	DNDC[+57]GDM[+16]SDEKDC[+57]PTQPFR
P98164	LRP2	TC[+57]EDIDEC[+57]DILGSC[+57]SQHC[+57]YNM[+16]R
P98164	LRP2	IASANM[+16]DGTSVK
P98164	LRP2	TVM[+16]SLDYDSVSDR
P98164	LRP2	ANKYDGSGQIAM[+16]TTNLLSQPR
P98164	LRP2	YDGSGQIAM[+16]TTNLLSQPR
P98164	LRP2	TC[+57]VDIDEC[+57]TEM[+16]PFVC[+57]SQK
Q99698	LYST	SLMSPGFMVISPSGFTASPYEGENSSNIIPQQM[+16]AAHM[+16]LR
Q7Z7M0	MEGF8	GAM[+16]YLLGGLTAGGVTR
Q2M2H8	MGAM2	WM[+16]QLGAFYPFSR
O43451	MGAM	YEYGTLDNM[+16]R
O43451	MGAM	VM[+16]AYVPDAVWYDYETGSQVR
Q9H8L6	MMRN2	TPVC[+57]TTGQGSGSTATVFAM[+16]AELQK
Q9H8L6	MMRN2	SLSGTAFGGFLM[+16]FK
P42345	MTOR	NLLPSNDPVVM[+16]EM[+16]ASKAIGR
Q8WXI7	MUC16	TTQSLGVM[+16]SSALPESTSR

Protein Acc#	Gene	Peptide Modified Sequence
Q9HC84	MUC5B	AAQLPDM[+16]PLEELGQQVDC[+57]DR
Q9HC84	MUC5B	M[+16]C[+57]FNYEIR
Q6W4X9	MUC6	FHGTC[+57]TYILLQSPQLPEDGALM[+16]AVYDK
Q6W4X9	MUC6	ETDPC[+57]SM[+16]SQLNK
Q6W4X9	MUC6	LSC[+57]PQRPQM[+16]FLASC[+57]QAPK
Q6W4X9	MUC6	FGAAC[+57]APTC[+57]QM[+16]LATGVAC[+57]VPTK
O00533	CHL1	VM[+16]TPAVYAPYDVK
P20929	NEB	SLQDDPKLVLSM[+16]NVAK
P08473	MME	LIQNM[+16]DATTEPC[+57]TDFFK
P08473	MME	FIM[+16]DLVSSLSR
P61916	NPC2	AVVHGILM[+16]GVPVPFPIPEPDGC[+57]K
Q6UX06	OLFM4	ITYGQGSGTAVYNNNM[+16]YVNM[+16]YNTGNIAR
Q9UN73	PCDHA6	GHPPM[+16]AGHC[+57]TVLVRILDK
Q92824	PCSK5	VSHLYGFGLMDAEAMVM[+16]EAEK
P05164	MPO	NQINALTSFVDASM[+16]VYGSEEPLAR
P05164	MPO	FWWENEGVFSM[+16]QQR
Q6P1J6	PLB1	QDWTERPQQVC[+57]MGVM[+16]TVLSDIIR
P14923	JUP	ALM[+16]GSPQLVAAVVR
P13797	PLS3	M[+16]INLSVPDTIDER
P98160	HSPG2	EHLLM[+16]ALAGIDTLLIR
Q96PD5	PGLYRP2	SPPTM[+16]VDSLLAVTLAGNLGLTFLR
P08F94	PKHD1	EDTVVGEDM[+16]R
P15309	ACPP	SPIDTFPTDPIKESSWPQGFGQLTQLGM[+16]EQHYELGEYIR
P15309	ACPP	ESSWPQGFGQLTQLGM[+16]EQHYELGEYIR
P15309	ACPP	SRLQGGVLVNEILNHM[+16]K
P15309	ACPP	LQGGVLVNEILNHM[+16]K
P15309	ACPP	LQGGVLVNEILNHM[+16]KR
P15309	ACPP	FAELVGPVIPQDWSTEC[+57]M[+16]TTNSHQGTEDSTD
Q5JSZ5	PRRC2B	AGEQGEAM[+16]K
P07737	PFN1	DSLLQDGEFSM[+16]DLR
P41222	PTGDS	AALSM[+16]C[+57]K
P41222	PTGDS	TM[+16]LLQPAGSLGSYSYR
P41222	PTGDS	M[+16]ATLYSR
P15151	PVR	VQLTGEPVPM[+16]AR
Q5JT25	RAB41	NLNVM[+16]FIETSAKTGYNVK
Q16769	QPCT	QIAEGTSISEM[+16]WQNDLQPLLIER
Q16769	QPCT	M[+16]ASTPHPPGAR
Q16769	QPCT	GVPVLHLIPSPFPEVWHTM[+16]DDNEENLDESTIDNLNK
P02753	RBP4	FSGTWYAM[+16]AK
P02753	RBP4	LLNNWDVC[+57]ADM[+16]VGTFTDTEDPAK
P10153	RNASE2	YAQTPANM[+16]FYIVAC[+57]DNR

Protein Acc#	Gene	Peptide Modified Sequence
P07998	RNASE1	QHM[+16]DSDSSPSSSSTYC[+57]NQMMR
P07998	RNASE1	QHMDSDSSPSSSSTYC[+57]NQMM[+16]R
Q8WZ75	ROBO4	M[+16]SC[+57]QASGQPPPTIR
P07602	PSAP	GEM[+16]SRPGEVC[+57]SALNLC[+57]ESLQK
P07602	PSAP	LGPGM[+16]ADIC[+57]K
P07602	PSAP	NYISQYSEIAIQM[+16]MMHMQPK
P07602	PSAP	NYISQYSEIAIQMMMHM[+16]QPK
P07602	PSAP	NYISQYSEIAIQMMM[+16]HM[+16]QPK
P17900	GM2A	IPC[+57]TDYIGSC[+57]TFEHFC[+57]DVLDM[+16]LIPTGEPC[+57]PEPLR
Q8WVN6	SECTM1	DSHAGLYM[+16]WHLVGHQR
O00560	SDCBP	LYPELSQYM[+16]GLSLNEEEIR
P29508	SERPINB3	ENNIFYSPISITSALGM[+16]VLLGAK
P29508	SERPINB3	GKDLSM[+16]IVLLPNEIDGLQK
P29508	SERPINB3	DLSM[+16]IVLLPNEIDGLQK
P29508	SERPINB3	LM[+16]EWTSLQNMR
P29508	SERPINB3	LMEWTSLQNM[+16]R
P29508	SERPINB3	TMGMVDIFNGDADLSGM[+16]TGSR
Q96P63	SERPINB12	NIFFSPLSLSAALGM[+16]VR
Q9UIV8	SERPINB13	ADYSGM[+16]SSGSGLYAQK
Q96PV0	SYNGAP1	DFLSDM[+16]AM[+16]SEVDR
Q8WXH0	SYNE2	KTEAELEM[+16]LK
P22105	TNXB	TITTM[+16]IDGPQDLR
P02787	TF	SM[+16]GGKEDLIWELLNQAQEHFGK
P02787	TF	M[+16]YLGYEYVTAIR
P02787	TF	IM[+16]NGEADAMSLDGGFVYIAGK
P02787	TF	IMNGEADAM[+16]SLDGGFVYIAGK
P02787	TF	IM[+16]NGEADAM[+16]SLDGGFVYIAGK
P02787	TF	TAGWNIPM[+16]GLLYNK
P02787	TF	LC[+57]M[+16]GSGLNLC[+57]EPNNK
P02787	TF	LC[+57]M[+16]GSGLNLC[+57]EPNNKEGYYGYTGAFR
P02788	LTF	GEADAM[+16]SLDGGYVYTAGK
Q9Y4A5	TRRAP	ELSEKDIGNQLHM[+16]LTNR
P30530	AXL	LAYQGQDTPEVLM[+16]DIGLR
P07911	UMOD	M[+16]AETC[+57]VPVLR
P07911	UMOD	LEC[+57]GANDM[+16]K
P07911	UMOD	LEC[+57]GANDM[+16]KVSLGK
P07911	UMOD	SLGFDKVFM[+16]YLSDSR
P07911	UMOD	VFM[+16]YLSDSR
P07911	UMOD	INFAC[+57]SYPLDM[+16]K
P07911	UMOD	INFAC[+57]SYPLDM[+16]KVSLK
P07911	UMOD	TALQPM[+16]VSALNIR

Protein Acc#	Gene	Peptide Modified Sequence
P07911	UMOD	VGGTGM[+16]FTVR
P02774	GC	VM[+16]DKYTFELSR
Q14508	WFDC2	DQC[+57]QVDSQC[+57]PGQM[+16]K
Q6ZQQ6	WDR87	EM[+16]AQAEGKFAQK
Q5THJ4	VPS13D	EKDDLSPQPLM[+16]TDFER
E9PFZ2	CP	M[+16]YYSAVEPTKDIFTGLIGPMK
Q53ET8		ELNVM[+16]FIETSAKAGYNVK
Q53FJ5		M[+16]VTDIQTAVR
E9PGN7	SERPING1	VTTSQDMLSIM[+16]EK
Q6MZV7	DKFZp686C11235	EPQVYTLPPSREEM[+16]TK
B3KVV6		NSATSTTNLYTQALLAYIFSLAGEM[+16]DIR
C8C504	HBB	FFKSFGDLSTPDAVM[+16]GNPK
Q9UL83		EIVM[+16]TQSPATLSVSPGER
Q9NS89	CACNA1A	EHM[+16]AHRQGSSSVSGSPAPSTSGTSTSR
Q4W4Y1	DRIP4	M[+16]VPVSVQQSLAAYNQR
E7ER45	MGAM	YTLLPYLYTLM[+16]QK
M0QY60	ZNF714	M[+16]NVM[+16]LENYK
F6KPG5		SHC[+57]IAEVENDEM[+16]RADLPSLAADFVESK
L8E9E6	SOD1	GMLETWAM[+16]
A2JA16		DIQM[+16]TQSPSSLPASVGDR
B4DPR2		SHC[+57]IAEVENDGM[+16]PADLPSLAADFVESK
R4H484		ITYM[+16]SPDYAAPTLAGLDDATK
A0SEH4	CCKBR	M[+16]LLVI
P02763	ORM1	TYM[+16]LAFDVNDEK

CHAPTER FOUR

4. CONCLUSIONS AND FUTURE STUDIES

Blood plasma is a powerful material used in discovering biomarkers of human physiological changes. However, the elevated concentration of plasma proteins often limits the study of blood biomarkers. Alternatively, the ease, noninvasive procedure of collecting urine samples, and as a result of being a filtrate of plasma enabled this biofluid to become a powerful tool substituting plasma in biomarker studies for a wide array of diseases. MS-based methods to profile urine with increasing depth in identifying thousands of fragments including unique peptides enables detecting of polypeptides with high accuracy and degree of confidence.

Oxidative posttranslational modification of proteins is known to impact the structure and consequently the function of these biomolecules. Increasing number of studies have examined the role of a variety of PTMs in signal transductions or the deleterious effects of these modifications might have on human physiology. The impact of oxidative stress on side chain residues especially those with sulfur group in the structure including methionine has been underpinned by several research groups.

As it was anticipated, analysis of data revealed a noticable variation in protein intensities across all forty bio-samples used in the current study, mainly because of the complexity of urine texture. Group comparisons of methionine sulfoxide content between both smoker and control populations demonstrated statistically significant fold change in smokers. Furthermore, we noticed significantly higher levels of oxidized methionine residues in the lower mass region, suggesting an association between oxidation of Met and protein turnover (p-value = 0.053). Analysis of total proteins also showed an increased level of proteins cleared in the urine samples of smokers

irrespective of methionine oxidation. The current study is the first to address and provide a quantitative measurement of methionine sulfoxide proteome in human serum proteins, and in evaluating the impact of cigarette smoking on plasma proteome. However. It is necessary to quantitatively evaluate the extent of oxidation that might be introduced during sample preparations. Moreover, integrating strategies to evaluate the level of antioxidants especially those directly linked to oxidation of methionine including methionine sulfoxide reductases (MsrA and MsrB) provide additional sights to the changes observed in the analysis of data since the two variables are interchangeably involved in the whole process of oxidative stress and in activating transcription of genes responsible in generating a wide spectrum of other antioxidants.