Investigation of Potential Biomarkers in Whole Blood, Urine, and Saliva for Precision Medicine Applications

Hugo Save
hugo.save@outlook.com

under the direction of
Ph.D. Fredrik Edfors
Division of Systems Biology, Department of Protein Science,
School of Engineering Sciences in Chemistry
KTH, Royal Institute of Technology

Research Academy for Young Scientists
July 10, 2019

Abstract

Biomarkers provide biological information that can help clinic personnel give more correct diagnoses. Today, most protein biomarkers are individual proteins, that when crossing a static concentration threshold, is said to indicate a certain disease or biological state. It would be beneficial for future biomarker search to analyze if protein patterns and changes of protein levels over time also could be used as biomarkers. Mass spectrometry (MS) is capable of analyzing a broad range of proteins in a single run and this study's aim is to investigate and compare the potential of using whole blood, urine, and saliva together with MS to find potential biomarkers.

Three samples each of whole blood, urine, and saliva were collected from three, healthy non fasting individuals. The samples were collected by the individuals themselves in a non clinic environment and the proteins were cleaved with trypsin before being analyzed with shotgun-tandem MS. The raw data was then analyzed in MaxQuant and a list of found proteins were compiled.

The results show that saliva had the biggest protein coverage, followed by blood and then urine (877, 585, 194 respectively). The protein composition, between patients, varied the most for the saliva samples, while urine and blood were similar. A total of 1146 proteins were found and 80 of those were known biomarkers. Of the 80 known biomarkers, 66 were found in blood, 53 in saliva and 32 in urine.

Blood is a known source of proteins and biomarkers but this study also proposes saliva as a candidate for protein profiling, especially over a longer time period. That is because of it being very low invasive, easy to sample and having a rich proteome. To see how well the saliva proteome can be correlated to diseases in practice, is an important issue left for future studies.

Acknowledgements

I would like to warmly thank my mentor Ph.D. Fredrik Edfors for introducing me to his team and giving me an incredible two weeks at SciLifeLab. He has been an inspiring and knowledgeable mentor and I am truly grateful for him taking his time to make this experience possible. I would also like to thank those who read my report and given valuable feedback and guidance. Lastly, I would like to thank Rays - for excellence and their partners Kungliga Patriotiska Sällskapet and Kjell och Märta Beijers Stiftelse for making this summer possible.

Contents

List of Abbreviations

ACN Acetonitrile.

BSA Bovine Serum Albumin.

BV Biological Variation.

CAA 2-Chloroacetamide.

CV Coefficient of Variation.

DBS Dried Blood Spot.

DTT Dithiothreitol.

ELISA Enzyme-Linked Immunosorbent Assay.

FA Formic Acid.

FDA U.S. Food and Drug Administration.

LC Liquid Chromatography.

MS Mass Spectrometry.

MS/MS Tandem Mass Spectrometry.

PBS Phosphate-Buffered Saline.

RCF Relative Centrifugal Field.

SDC Sodium Deoxycholate.

SRM Selected Reaction Monitoring.

TFA Trifluoroacetic acid.

TV Technical Variation.

VAMS Volumetric Absorbtive Microsampler.

w/v Weight per Volume.

1 Introduction

For each diagnosis that is placed, there is also a decision taken by clinic personnel. To help clinic personnel gather biological information about the patients, there are biomarkers. Biomarkers are indicators of a biological state and are capable of giving crucial information about the state of the patient [1]. An example of biomarkers is proteins. Proteins are part of most functions in the body and their concentrations reflects thus also the state of the body. Proteins can be found in all tissues but it is common to search for proteins in blood plasma. Two benefits with blood plasma are that it is considered less invasive than taking samples from tissues while also including a wider range of proteins [2, 3]. Even proteins, that only functions inside cells, are present in blood plasma. That is because parts of the cell interior gets released when the cell dies. It's likely that all diseases would show some sign on the protein levels. That also means it would be possible to detect the diseases if we have a good understanding of our protein levels [3]. Because of that potential, it's crucial to deepen our understanding of the human proteome, and improve our protein analysis methods.

1.1 Sampling Methods

Before any analyzing can be done, samples have to be collected. One common method to collect blood samples is dried blood spot (DBS). In DBS, the patient pierces the skin, often a finger, and applies a few drops of blood onto a absorbent paper which is then sent for analyzing. Advantages of DBS over venepuncture is that it is easy to perform, cheap and does not involve risks associated with needles and syringes [2]. This also makes it possible for patients to take blood samples themselves. A disadvantage, however, is that depending on the individual's blood viscosity, you get a different amount of blood volume. Even for a blood spot with fixed size. One way to address this issue has been by developing volumetric absorbtive microsampler (VAMS), or Mitra™ as it is called commercially. VAMS are capable of absorbing a specific volume of a body fluid, no

matter the viscosity [4].

Even though DBS and VAMS has made it easier to extract blood compared to venepuncture, they are still partly invasive. Urine or preferably saliva samples are even less invasive and would be more convenient, especially over longer time periods, for the persons sampling themselves.

1.2 Identifying Biomarkers Today

Today, protein biomarkers are often quantified using enzyme-linked immunosorbent assay (ELISA). ELISA is a well established assay and used worldwide to give more correct diagnoses [5]. One downside however is that each ELISA test can only measure one single protein from each patient and when wanting to measure a new a protein a new antibody for the ELISA test also has to be produced. Already developed ELISA tests are cost effective but the development of new antibodies are expensive [6]. ELISA tests are also biased in the way that they can only analyze proteins that we are already are aware of and it can not see protein patterns [7, 8]. Patterns of proteins could give more detailed information about the state of the patient but ELISA is limited in this regard since it only can measure individual proteins [8].

1.3 Mass Spectrometry

To get a better view of protein patterns and that way get a more thorough understanding of the state of the patient, mass spectrometry (MS) can be used. MS functions by ionizing a compound that then is put into an electric field. By analyzing how the charged compound acts in an electric field, it is possible to determine its mass to charge ratio (m/z) and, by comparing it to its naturally appearing heavy-isotopes, also its mass with very high accuracy and precision [9]. The theoretical protein masses are known from our knowledge of the human genome and how it codes proteins. If the MS detects a unique mass peak, it is thus possible to assign the peak to a certain protein [10].

There are, however, technical difficulties of measuring whole proteins because of their high mass. What can be done instead, is cleaving the proteins into smaller peptides [11]. It is possible to predict what peptides will be formed since enzymes, like trypsin, can cleave the protein in a regular manner [12]. By knowing what peptides will be cleaved from a protein it is possible to deduce what proteins existed by only looking at the peptides [11]. Two other challenges in analyzing proteins in plasma is the dynamic range of the proteins and the vast amount of different proteins. The range of protein concentration stretches over 10 magnitudes, with the 22 most abundant proteins representing approximately 99% of the plasma's protein mass, and the amount of different proteins is approximated up to a million [13, 14].

To address the dynamic range and the complexity of protein mixtures, different methods have been developed. Two of those methods are liquid chromatography (LC), and tandem MS (MS/MS). LC reduces the amount of peptides the MS has to analyze at a given time by gradually inputting the peptides, and MS/MS helps peptide analysis by sending peptides into a second MS where more information about the peptide is provided [15]. The method of determining proteins by cleaving them and analyzing the peptides is called "bottom up proteomics". If you only select a narrow window of masses to analyze for the second MS in MS/MS it is called selected reaction monitoring (SRM), and when you try to see as many different proteins as possible over a wide mass range, shotgun proteomics is used. Shotgun proteomics has the advantage of being able to detect more unique proteins at cost of reproducibility. Running identical samples in shotgun proteomics can thus give slightly different results each time. SRM is more reproducible but you have to know in advance what masses you want to monitor and the range of proteins that can be seen is more limited. SRM is suited for quantification of proteins that you know is in the sample while shotgun is good for discovering and getting a broad view of the protein content. The two methods can of course also be used in succession to complement each other [11].

1.4 Precision Medicine

MS can detect and quantify hundreds of protein biomarkers from a single sample in a single run. In that way, doctors do not have to prioritize what biomarkers to analyze and more information is provided about the biological state of the patient. Having more information about the individual also makes precision medicine easier to apply. For example, the drug amount given to patients who are diagnosed for a particular disease is determined experimentally. With no way of differentiating the individuals, the drug amount becomes an average of what works best for the most people. But that does not mean it is the best dose for everyone. Since everyone is biologically unique, we also have slightly different protein levels. Different protein levels can, in turn, vary the optimal drug dose. Using MS it is possible to detect each individual's protein levels and thus create a protein profile for each person [16]. If the drug dose experiments also includes protein profiling of each participant, it might be possible to connect different protein profiles to different doses. Then, new patients could get their protein levels analyzed and a more optimal dose of drugs can be given. Similarly, it is not certain that the most commonly used drug is optimal for everyone. One drug might be optimal to a percentage of the patients but another drug might be optimal towards the rest. Precision medicine is about taking individual variability into account, and mass spectroscopy can help us measure that variability [17].

1.5 Previous Research

There are a vast amount of proteins in the body and many, but far from all, have been identified. The human proteome project estimates that it exist a million different proteoforms (proteoforms includes all posttranslational modifications of a protein) and Uniprots proteome database for Homo Sapiens lists 70 000 found proteoforms [14, 18]. Of those 70 000, 20 000 have been reviewed by humans and the other 50 000 proteoforms are waiting to be manually annotated [18]. A little more than a hundred proteins are approved

biomarkers by the U.S. Food and Drug Administration (FDA) [19]. Blood is often used when searching for proteins and biomarkers because of its high protein concentration but other body fluids can also be used [20]. Studies have shown that urine biomarkers can be used to detect kidney infection and the saliva proteome has been analyzed and some potential biomarkers have been found [21, 22]. The body fluid samples from the above studies were, for those who specified, collected by clinical personnel and taken from fasting patients. To be able to do large scale studies of individual's proteome, it would be cost beneficial and reduce the strain on primary care, to let the patients send samples from home directly to the lab. Fasting is also not an option for regular protein profiling over a longer time and it would be more convenient for the individual and less invasive if the patient could take saliva or urine samples instead of blood.

1.6 Aim of Study

The aim of this study is to compare the use of urine, saliva and blood to find known and potential biomarkers using mass spectrometry, when the samples are taken by non fasting individuals themselves in a non clinic environment.

2 Method

Before the samples are sent into the MS, the proteins needs to be extracted and cleaved. That can be done by solid phase extraction and trypsin digestion. To be able to use the same setup of phase extraction for all samples, the amount of peptides also had to be similar. The protein concentrations were thus measured and different amounts of each sample had to extracted before cleavage.

2.1 Protein Extraction

Three samples each of blood, urine, and salivia were collected from three healthy individuals (27 samples total). Each individual sent their samples from home using Mitra™

tips.

The samples were put in 1.5 ml low-binding tubes and each sample was extracted with 496 µl of Phosphate-buffered saline (PBS), 5 µl of 10% weight per volume (w/v) sodium deoxycholate and 0.25 µl of 1 M dithiothreitol (DTT). All samples were then incubated at room temperature on a Rotomixer with setting uu at 60 rpm for 1 hour.

2.2 Protein Concentration Determination

The sample's protein concentration was measured using absorption values and a standard curve. 5 standard tubes and a reference (100% solvent) were made using a bovine serum albumin (BSA) standard (16 mg/ml) diluted with a solvent mix of BIO-RAD protein assay regent A and B (50:1, Reagent A:B). The final concentrations were between 0.1-0.6 mg/l and 0 mg/l for the reference tube. Each standard tube had a volume of 100 µl and was was vortexed and incubated for 30 min at 37°C. A standard curve was then made using the absorption values from the standard tubes and linear regression.

100 µl of each sample was extracted and analyzed at 595 nm in the absorption spectrometer. Only 20 µl of the blood samples was extracted and then diluted to 100 µl with the A B BIO-Rad protein assay mix to get their concentration in range of the standard samples. 30 µg of proteins was then extracted from each sample and put in new tubes. Each tube was diluted to 300 µl with PBS.

2.3 Peptide Sample Preparation

10% (w/v) of sodium deoxycholate (SDC) was added to each tube to a final concentration of 1%. 1 M DDT was added to a final concentration of 10 mM. The samples were then incubated at 56°C for 30 minutes. 1 M 2-chloroacetamide (CAA) was added to a final concentration of 50 mM and the samples were then placed in the dark for 30 minutes. 1.5 µg trypsin (modified, MS-Grade, from Thermo Scientific) was then added to make a 1:20 protein ratio and the samples were incubated at 37°C over the night. The digestion

was quenched with 10% trifluoroacetic acid (TFA) to a final concentration of 0.5% volume concentration. The SDC was then let to percipitate for 30 minutes at room temperature. All samples was then centrifuged at 370 relative centrifugal field (RCF) for 5 minutes.

2.4 Solid Phase Extraction

All centrifuge steps in the following paragraph was done with 2000 RCF for 5 minutes.

For each sample, stage tips with 6 layers of Empore octadecyl C18 47 mm extraction disks were prepared. All stage tips were centrifuged after each of the following steps: the C18 layers were activated with 50 µl of 100% acetonitrile (ACN), and the tips were then treated with 50 µl 0.1% TFA. The pH was checked to be below 3, and the samples with the digested proteins were added into each stage tip and all stage tips were centrifuged. All samples were then washed 2 times with 80 µl of 0.1% TFA and then eluted 2 times with 30 µl of 80% ACN and 30 µl of 0.1% of formic acid (FA). All stage tips were centrifuged after each wash and elution. The eluted peptides were then speed evaporated for 20 min at 45°C. The samples were then resuspended in 30 µl of 3% ACN, 0.1% FA and 96.9% water for MS analysis.

6 µg each of the 27 peptide samples were analyzed in LC-MS/MS and all mass spectras were processed via the MaxQuant software to give a list of potential proteins and peptides in the respective samples. The protein lists were then analyzed and compared against themselves and against a list of already known biomarkers. This is shown in section 3.

3 Results

A total of 1146 proteins were found in the 27 samples. The 27 samples consisted of nine blood, nine saliva and nine urine samples and those samples came from three different patients who all gave three of each. Figure 1 shows the distribution of the number of proteins in the categories: blood, saliva and urine. The graph shows that the saliva samples included the most amount of different proteins (877) followed by blood (585), and urine

(194). Saliva was also the sample that had the most amount of different proteins that were found in that sample only (463). There was an overlap of 292 proteins between saliva and blood and a common overlap of 85 proteins between all samples. For a protein to be counted, it had to be present in 1 or more of the 9 samples in a category.

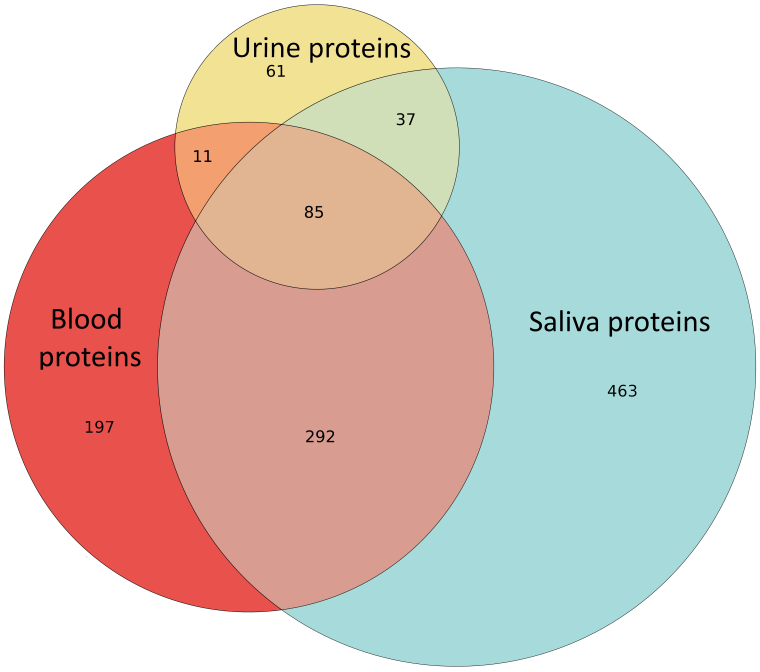

Figure 1: Distribution of proteins between blood, saliva and urine

The proteins of each category is shown in figures 2-4 and are sorted after intensity from highest to lowest. The intensity can be used to see relative concentration between proteins and it stretches between 6 magnitudes of order. Figure 2,3 and 4 represents 585 proteins found in blood, 877 proteins found in saliva and 194 proteins found in urine respectively. Of those proteins, 66 were approved or known biomarkers found in blood, 53 were found in saliva and 32 in urine. The distribution and overlap of approved and known biomarkers are shown in figure 5. 42 of the 66 biomarkers found in the blood could also be detected in the saliva and 26 of those 66 in urine.

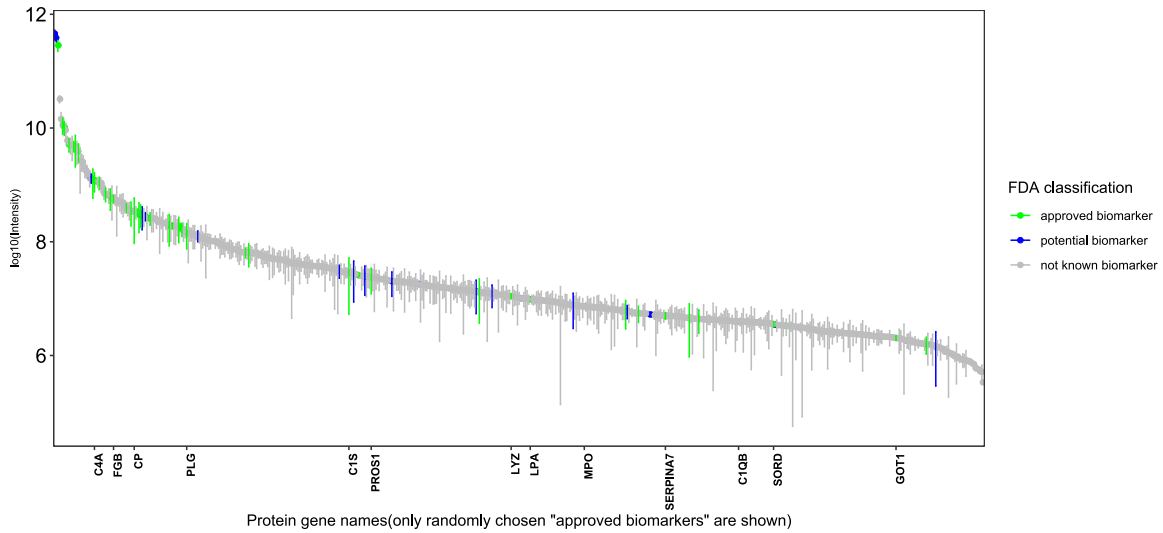

Figure 2: Proteins found in blood and their intensity (Green represent FDA approved biomarkers and blue represent potential biomarkers that have been found but not FDA approved. Each dot represents the intensity mean of all measurements of a protein and the vertical lines are the standard deviation for that proteins intensity)

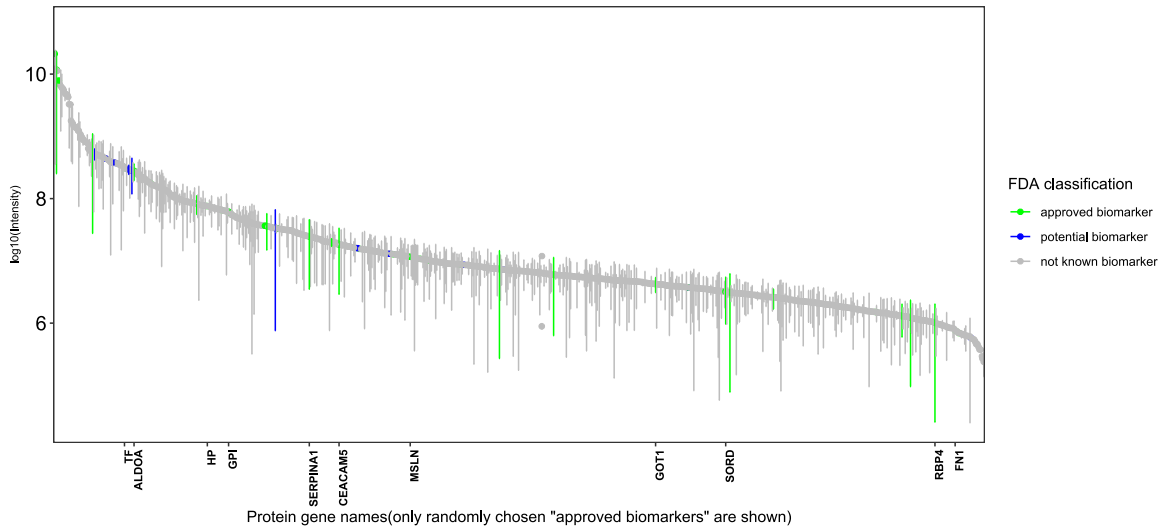

Figure 3: Proteins found in saliva and their intensity (Green represent FDA approved biomarkers and blue represent potential biomarkers that have been found but not FDA approved. Each dot represents the intensity mean of all measurements of a protein and the vertical lines are the standard deviation for that proteins intensity)

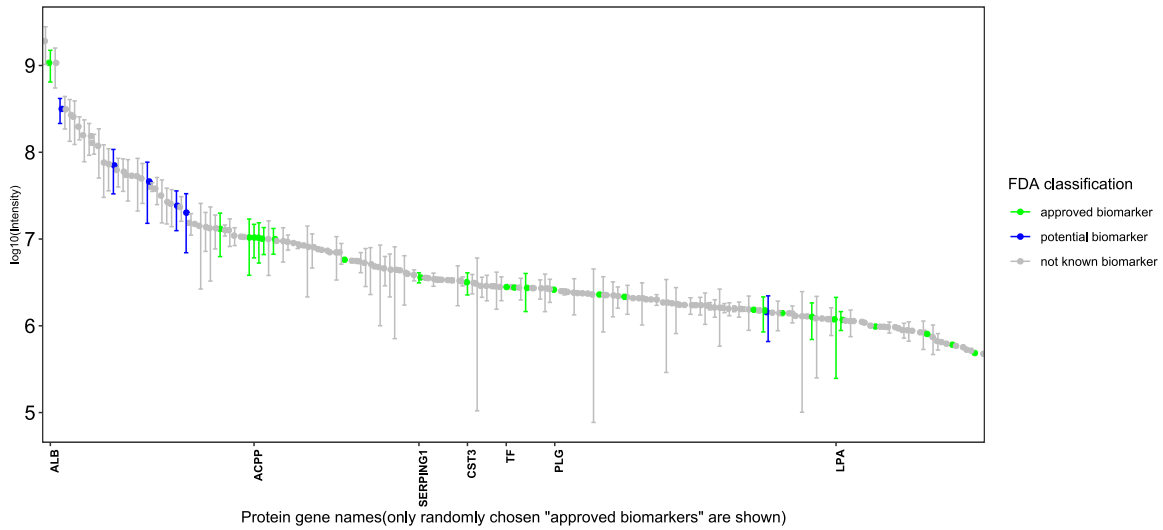

Figure 4: Proteins found in urine and their intensity (Green represent FDA approved biomarkers and blue represent potential biomarkers that have been found but not FDA approved. Each dot represents the intensity mean of all measurements of a protein and the vertical lines are the standard deviation for that proteins intensity)

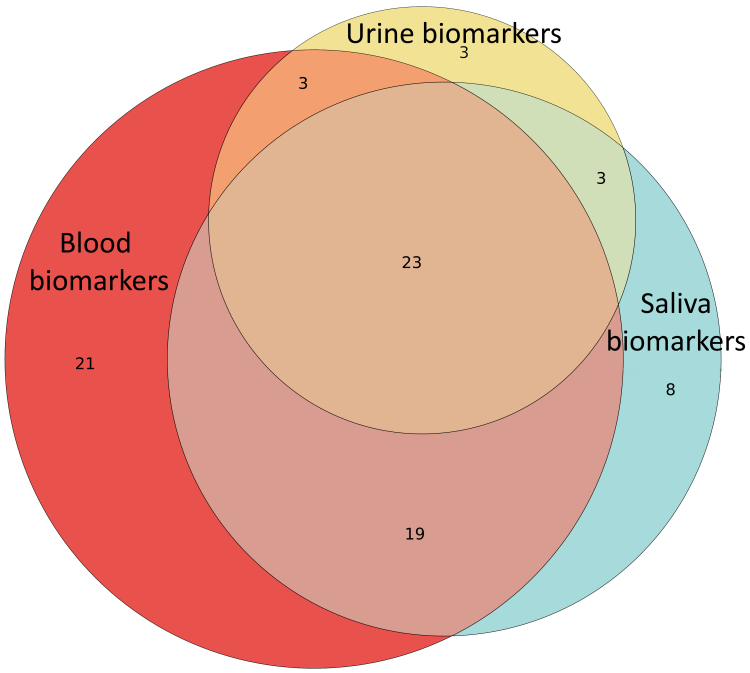

Figure 5: Distribution of found and known biomarkers between blood, saliva and urine

Table 1 shows an excerpt of a table that is linked in the appendix. The table shows the abundance of approved or known biomarkers in the different patients and samples. Table 2 is an extract of the same table but shows some of the biomarkers that were only found in the saliva or urine samples. The complete table also includes the biological variation between each patient for each protein.

Table 1: Sample of found biomarkers and where they were found. (First digit represents in the number of patients the protein was found and the second digit after the / represents the number of samples where the protein was found)

Blood	Saliva	Urine	Uniprot name	FDA protein name
3/7	3/9	0/0	P00338	Lactate Dehydrogenase Liver Fraction (LLDH)
3/9	3/8	2/4	P00450	Ceruloplasmin
1/3	0/0	0/0	P00488	Factor XIII
3/9	0/0	3/9	P00734	Factor II Activity Assay, Plasma
3/9	0/0	0/0	P00736	Complement C1
3/9	3/9	1/3	P00738	Haptoglobin
3/9	2/3	1/1	P00747	Plasminogen
3/9	0/0	0/0	P00748	Factor XII Activity Assay, Plasma
3/9	3/5	0/0	P00751	PROPERDIN FACTOR B
3/9	2/2	0/0	P01008	ANTITHROMBIN III (ATIII)
...	...	...	...	...

Table 2: Sample of biomarkers only found in saliva or urine. (First digit represents in the number of patients the protein was found and the second digit after the / represents the number of samples where the protein was found)

Blood	Saliva	Urine	Uniprot name	FDA protein name
0/0	3/9	3/9	P01034	Cystatin C
0/0	0/0	3/9	P01133	Epidermal Growth Factor
0/0	3/9	0/0	P06731	Carcinoembryonic Antigen (CEA)
0/0	3/8	0/0	P08727	CYFRA 21-1 (Soluble Cytokeratin 19 Fragment)
0/0	3/5	2/6	P15309	Prostatic Acid Phosphatase (PAP)
0/0	3/9	0/0	P18510	Interleukin-1 Receptor Antagonist
0/0	3/9	0/0	P29508	Squamous Cell Carcinoma Antigen, Serum
...	...	...	...	...

Table 3 and 4 shows the technical and biological variation respectively. The technical variation is equal to the coefficient of variation (CV) over the samples that were taken from the same patient and category. The biological variation is also equal to the CV but taken over the different patients.

Table 3: Mean technical variation (TV) for the three samples that each patient sent in each category

Sample name		TV
Patient 1	Blood	62.3%
	Saliva	71.9%
	Urine	63.9%
Patient 2	Blood	43.3%
	Saliva	53.6%
	Urine	49.8%
Patient 3	Blood	19.7%
	Saliva	27.2%
	Urine	23.1%

Table 4: Mean biological variation (BV) for the three patients in each category

Sample category	BV
Blood	42.0%
Saliva	78.8%
Urine	49.1%

4 Discussion

The highest biological variation, between patients, was found in the saliva tests. The biological variation numbers are shown in table 4, but they should not be analyzed too strictly. That is because the biological variation is affected by the technical variation, which is noticeably high in shotgun proteomics (the technical variation is necessary when wanting to profile samples with unknown proteins). That combined with only including three patients makes the biological variation values just approximates. There is, however, a big difference of 30% units between saliva and the other two categories. That means it is still quite certain that the variation is higher in the saliva. Experiments using other MS methods, such as SRM, would have to be used to determine the biological variation more exactly by delivering more accurate and precise data.

4.1 Technical Variation

Shotgun proteomics has a noticeable technical variation compared to other MS methods. What is interesting, is that the variation is not the same for each triplet of samples.

Each patient sent 9 samples, 3 each of blood, saliva, and urine. 3 samples from the same patient and category is called a triplet. Each triplet should be close to identical, since the patients took the three samples of the triplet at the same time. The technical variation over the samples in a triplet should thus also be the same for all triplets. What is seen in table 3 is that, as predicted, each patient indeed has similar technical variations in each triplet of blood, saliva, and urine. There is, however, a clear difference between the patients. Patient 1 has technical variations around 60-70% while patient 3 has technical variations around 20-30%. This is probably the result of bad sampling technique, storage differences and other biasing factors that has interfered for some of the patients' samples.

4.2 Identifying Known Biomarkers

Saliva was the test with the most amount of proteins. It had almost 300 hundred more than blood (877 against 585), but blood was still the sample that covered most known biomarkers (66 against 53). Two possible reasons for that might either be that saliva is not as a reliable source of biomarkers as blood, or that most biomarker studies have been done with blood samples and thus also mostly found biomarkers that are present in blood.

With saliva samples covering 64% of the proteins found in the blood samples, it seems likely that saliva could be used to replace blood in search of certain biomarkers. The documented concentration thresholds for known biomarkers needs, however, to be reevaluated from blood to saliva. It is also not certain that known correlations between a protein and a disease is applicable in both blood and saliva. Future studies has to examine the saliva proteome in connection to diseases to determine it's effectiveness, finding new threshold values and finding protein patterns that have correlations to the patients' health.

4.3 Identifying Alternative Biomarkers

Figures 2-4 show greyed out proteins. These are proteins that are not known biomarkers. But that none of those 1066 greyed out proteins, seen in figures 2-4, would be able to give valuable information about the state of a patient, is unlikely. The known biomarkers covers only 80 of the 1146 proteins that was found in this study. Many proteins might not individually have a strong enough correlation to a disease to be used as separate biomarkers. It is, however, possible that relationships between different proteins could create protein patterns that could be correlated to a disease. It would also be of interest to profile proteins over time. That way, it would be possible to see if certain protein changes, or changes of multiple protein relations, can be correlated with biological states.

It is especially important to measure proteins over time for proteins with high variation between individuals. Those proteins are in the current system not very valuable biomarkers because it is difficult to draw conclusions from proteins with high individual variation when only measuring one patient's protein levels once. That does not mean, however, that proteins with high biological variation can not be used as biomarkers. It only means that static thresholds can not be used and instead, the thresholds has to be relative to the individual. The change in protein concentration, relative to the patient's normal values, might then be correlated to a disease.

5 Conclusion

It is important to monitor protein levels over time to find new correlations between protein levels and biological states. This study proposes that saliva samples might be favorable over blood when conducting protein profiling over a longer time. Saliva samples covered a broad range of proteins, while also being less invasive than blood, and easier to sample for the research participants themselves. By making the participants sample themselves, stress is lifted from the clinics and costs are reduced. MS is a suitable method for analyzing the samples since it can analyze and quantify a broad range of proteins in

a single run. By measuring a broad range of proteins it would be possible to not only correlate individual proteins to diseases but also patterns of proteins, and by measuring over time it might also be possible to correlate changes of protein patterns to diseases.

With improved knowledge about the correlation between protein patterns and the individual's biological state, it would be possible to provide a more individualized, and optimal health care.

References

[1] Strimbu K, Tavel JA. What are biomarkers? *Current Opinion in HIV and AIDS*. 2010;5(6):463.

[2] Parker S, Cubitt W. The use of the dried blood spot sample in epidemiological studies. *Journal of clinical pathology*. 1999;52(9):633.

[3] Anderson NL, Polanski M, Pieper R, Gatlin T, Tirumalai RS, Conrads TP, et al. The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Molecular & Cellular Proteomics*. 2004;3(4):311–326.

[4] Spooner N, Denniff P, Michielsen L, De Vries R, Ji QC, Arnold ME, et al. A device for dried blood microsampling in quantitative bioanalysis: overcoming the issues associated blood hematocrit. *Bioanalysis*. 2015;7(6):653–659.

[5] Lequin RM. Enzyme Immunoassay (EIA)/Enzyme-Linked Immunosorbent Assay (ELISA). *Clinical Chemistry*. 2005;51(12):2415–2418. Available from: "http://clinchem.aaccjnls.org/content/51/12/2415".

[6] Sakamoto S, Putalun W, Vimolmangkang S, Phoolcharoen W, Shoyama Y, Tanaka H, et al. Enzyme-linked immunosorbent assay for the quantitative/qualitative analysis of plant secondary metabolites. *Journal of natural medicines*. 2018;72(1):32–42.

[7] Bazenet C, Lovestone S. Plasma biomarkers for Alzheimer's disease: much needed but tough to find. *Biomarkers in medicine*. 2012;6(4):441–454.

[8] Seibert V, Ebert MP, Buschmann T. Advances in clinical cancer proteomics: SELDI-ToF-mass spectrometry and biomarker discovery. *Briefings in Functional Genomics*. 2005;4(1):16–26.

[9] Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422(6928):198–207. Available from: https://doi.org/10.1038/nature01511.

[10] Yates III JR. Mass spectrometry: from genomics to proteomics. *Trends in Genetics*. 2000;16(1):5–8.

[11] Kumar S, Gaur V, Khurana S, Bose S, Kiran M, Kiran M, et al. Proteomics Tools–An Update. *Clin Oncol*. 2017;2:1358.

[12] Uniprot. UniProtKB P07477 (TRY1_HUMAN); 2019. [Online; accessed 5 July 2019]. https://www.uniprot.org/uniprot/P07477.

[13] Domon B, Aebersold R. Mass spectrometry and protein analysis. *science*. 2006;312(5771):212–217.

[14] Legrain P, Aebersold R, Archakov A, Bairoch A, Bala K, Beretta L, et al. The human proteome project: current state and future direction. *Molecular & cellular proteomics*. 2011;10(7):M111–009993.

[15] Qian WJ, Jacobs JM, Liu T, Camp DG, Smith RD. Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. *Molecular & Cellular Proteomics*. 2006;5(10):1727–1744.

[16] Beger RD, Dunn W, Schmidt MA, Gross SS, Kirwan JA, Cascante M, et al. Metabolomics enables precision medicine:"a white paper, community perspective". *Metabolomics*. 2016;12(9):149.

[17] Collins FS, Varmus H. A New Initiative on Precision Medicine. *New England Journal of Medicine*. 2015;372(9):793–795. PMID: 25635347. Available from: `https://doi.org/10.1056/NEJMp1500523`.

[18] Uniprot. Proteomes Homo sapiens (Human); 2019. [Online; accessed 29 June 2019]. `https://www.uniprot.org/proteomes/UP000005640`.

[19] Anderson NL. The Clinical Plasma Proteome: A Survey of Clinical Assays for Proteins in Plasma and Serum. *Clinical Chemistry*. 2009;Available from: `http://clinchem.aaccjnls.org/content/early/2009/11/02/clinchem.2009.126706`.

[20] Crosley LK, Duthie SJ, Polley AC, Bouwman FG, Heim C, Mulholland F, et al. Variation in protein levels obtained from human blood cells and biofluids for platelet, peripheral blood mononuclear cell, plasma, urine and saliva proteomics. *Genes & Nutrition*. 2009 Apr;4(2):95. Available from: `https://doi.org/10.1007/s12263-009-0121-x`.

[21] TESCH GH. Review: Serum and urine biomarkers of kidney disease: A pathophysiological perspective. *Nephrology*. 2010;15(6):609–616. Available from: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1440-1797.2010.01361.x`.

[22] Wang Q, Yu Q, Lin Q, Duan Y. Emerging salivary biomarkers by mass spectrometry. *Clinica Chimica Acta*. 2015;438:214 – 221. Available from: `http://www.sciencedirect.com/science/article/pii/S0009898114003891`.

Appendix

The following files can be found in the Github repository: The data provided from the mass spectrum directly after MaxQuant analysis. The complete summary table of all found biomarkers with their presence in the different patients and samples. Biomarker tables that were used for biomarker comparison. R code that provided the data for the graphics and tables found in the paper. `https://github.com/HugoSave/Investigation-of-potential-biomarkers-supplementary-data`