Title

Comparison of Salivary Proteome Signatures With and Without Starch Treatment

Author

Smith, Hannah

Publication Date

2022

Peer reviewed|Thesis/dissertation

Comparison of Salivary Proteome Signatures With and Without Starch Treatment

By

Hannah Smith

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Forensic Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Cecilia Giulivi, Chair

Kent Pinkerton

James Angelastro

Committee in Charge

2022

ABSTRACT

Identification of subjects, including perpetrators, is one of the most crucial goals of forensic science. Saliva is among the most common biological fluids found at crime scenes containing identifiable components. DNA has been the most prominent identifier to date, but its analysis can be complex due to low DNA yields and issues preserving its integrity at the crime scene. Proteins are emerging as viable candidates for subject identification. Previous work has shown that the salivary proteome of the least abundant proteins may be helpful for subject identification, but more optimized techniques are needed. Among them, is the removal of the most abundant proteins, such as salivary α-amylase. We will test the hypothesis that the saliva proteome profile depleted of α-amylase and enriched with the least abundant proteins allows a more nuanced subject identification.

INTRODUCTION

In forensic science, identifying unidentified bodies and perpetrators is crucial to solving crimes. Identifying people using biological evidence as either non-perpetrators or perpetrators is the most obvious, but biological evidence can also help build event timelines. Saliva can shed light on both questions. Saliva can be found on everyday objects, such as cigarette butts and envelopes, or recovered from bite marks (1). Markers in saliva can also help reconstruct events, such as identifying the type of food or drink the suspect had before or during a crime (2).

Saliva is already used in forensic analysis for various purposes since it is ubiquitous at crime scenes and easily obtainable through non-invasive procedures (3). For instance, salivary drug screening can detect illegal substances (4), salivary DNA can be used for subject identification (5), and saliva can be used to perform microbiome profiling (6). Thus, the next step in the research is optimizing the use of salivary proteomes for subject identification.

A biological fluid to be identified as “saliva” relies on presumptive and confirmatory tests. Presumptive tests establish that a biological fluid may be present, while confirmatory tests conclusively identify the fluid as saliva. The most common presumptive test for saliva is the Phadebas test. This analysis uses a starch-dye complex that reacts with salivary α-amylase based on saliva’s high abundance of amylases (7). Amylases are secreted proteins that hydrolyze 1,4-

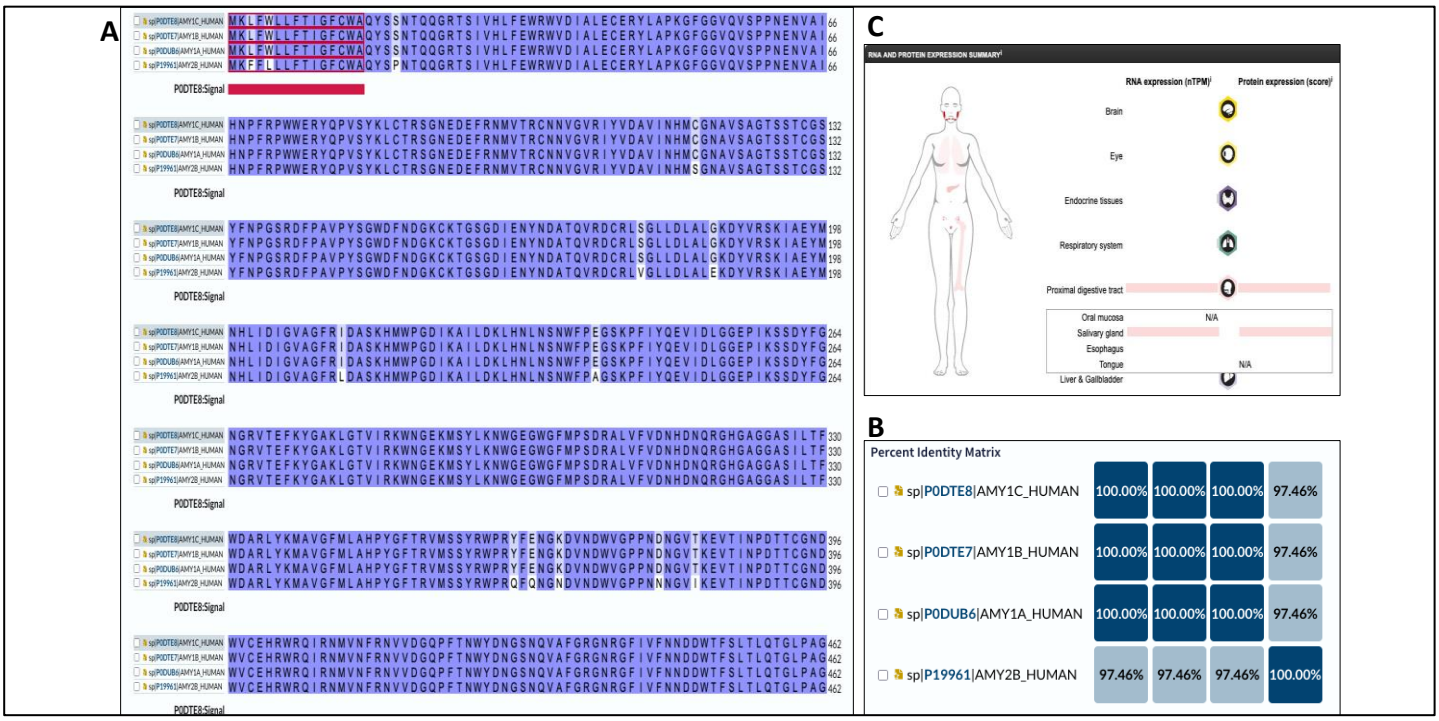


Figure 1. Alignment and expression of α-amylase isoforms in the body.
A. Alignment of primary sequences of isoforms AMY1A, AMY1B, AMY1C and AMY2B. Alignment performed with CLUSTAWL under Uniprot. Similar amino acids are highlighted in blue whereas the signal peptide is indicated in red. B. Percentage of identity across isoforms. C. Protein expression of AMY1A in the female body. Information retrieved from Human Protein Atlas.

alpha-glucoside bonds in oligosaccharides and polysaccharides and thus catalyze the first step in the digestion of dietary starch and glycogen (8). The human genome has a cluster of several amylase genes encoding for proteins with high similarity (**Figure 1A, B**) and are expressed at high levels in either the salivary gland (AMY1A, AMY1B, AMY1C; **Figure 1C**) or pancreas (AMY2A, AMY2B). Alternative splicing results in multiple transcript variants encoding the

same protein. However, this test can result in false positives as α-amylase is present in other bodily fluids such as semen, vaginal fluid, serum, and sweat (9).

The confirmatory test for saliva is the Rapid Stain Identification Test. This analysis uses a lateral flow immunochromatographic strip to test for the specific isoform of α-amylase in saliva. Mass spectrometry has also been suggested to identify saliva based on biological matrices (2).

Regarding subject identification using proteomes, proteins are usually used when DNA cannot be utilized because DNA molecules have less stability and degrade faster than proteins based on chemical, biological, and environmental processes (10). In cases where the DNA present is too degraded or low yield, proteins can be used to detect identifying characteristics (11-13). The salivary proteome in humans has been found to have differences based on many factors. The composition can be affected by how the saliva was collected, age, medication, sex, circadian rhythm, disease, physical activity, and oral hygiene (14-18).

The salivary proteome's main proteins are not only amylases, but also other critical proteins such as histatins, statherins, cystatins, proline-rich proteins, mucins, and immunoglobulins (19,20). Histatins show anti-fungal activity and the most common histatins found in saliva are histatins 1, 3, and 5 (21,22). Statherins are pivotal in maintaining the remineralization capacity of saliva by being the only salivary protein to inhibit primary and secondary calcium phosphate precipitation (23-26). The primary function of cystatins is protease inhibition (27,28). Like statherins, proline-rich proteins protect oral health by maintaining calcium concentration (29). Mucins have many functions including protecting the epithelium, aiding mastication, and guarding teeth (30,31). Immunoglobulin A is the dominant immunoglobulin isotype in saliva and protects against pathogens (32,33).

The study aims to determine if the removal of α-amylase, the most abundant protein found in saliva (7, 19, 34), will lead to an enrichment in the least abundant proteins, which are the most useful for subject identification. These proteins are seen less frequently and when looked at could add more points of comparison for subject identification.

MATERIALS AND METHODS

Sample Collection

Fifteen 1.5 mL saliva samples were taken from female volunteers aged 21-61 years (32.5 ± 10.8). All samples were from subjects who worked in the VetMed 3B School of Veterinary Medicine building at the University of California Davis. The only identifying information received from each subject was age and sex. The collection of samples occurred on a single day (February 4th, 2022) from 11:00 am to 2:00 pm. Subjects had not eaten, drunk, or performed oral hygiene routines for 30 minutes before providing samples. Using a sterile Corning tube given by the researcher, subjects provided saliva after tilting their heads back and letting it pool for 60 seconds. Samples were collected according to informed consent policies by the institutional review board (approved by the IRB (IRBNet ID: 1544585-1, 4/17/2020).

Saliva Preparation

Two aliquots (1 mL) from each of the ten saliva samples were prepared and treated with (n=10) or without (n=10) starch to remove amylase. The starch used in this study was from potato because, compared to other starches, has a low lipid level (Sigma-Aldrich, S2004, batch #0000129501). The starch was cleaned of any spurious material by washing it 3 times in distilled ~~deionized (~~deionized (ddi) water and centrifuged at 5000 rpm for 5 minutes. The starch solution was 20 g/l in ddi Milli-Q water. One mL of the clean and hydrated starch solutions was transferred to ten microcentrifuge tubes. Ten saliva samples were treated with the starch solutions. All saliva samples (1 mL) were added to each microcentrifuge tube, stirred for 5 minutes, and then centrifuged at 30,000 rpm for 5 minutes. The proteins in the supernatants of amylase-depleted saliva and the non-treated saliva samples were concentrated by acetone precipitation. Four volumes of -20°C acetone (analytical grade; Sigma-Aldrich, St Louis) were

added to each sample and left overnight at 4 °C. The samples were centrifuged at 16,000 x g for 10 minutes at 4 °C. The supernatant was discarded, and the pellet was washed twice with -20°C acetone at 16,000 x g for 10 minutes at 4 °C. The protein-containing pellets were dried under vacuum for 15 minutes (34).

Proteomics

Samples were prepared for protein analysis following a method previously published (34). Essentially, proteins were reduced and alkylated before digestion with LysC and the digestion into smaller peptides was performed by incubating with porcine trypsin. Depending on sample amount, 10–100 μg of the digest prepared from each sample was analyzed by mass spectrometry (34). Protein identification was performed using Mascot.

RESULTS

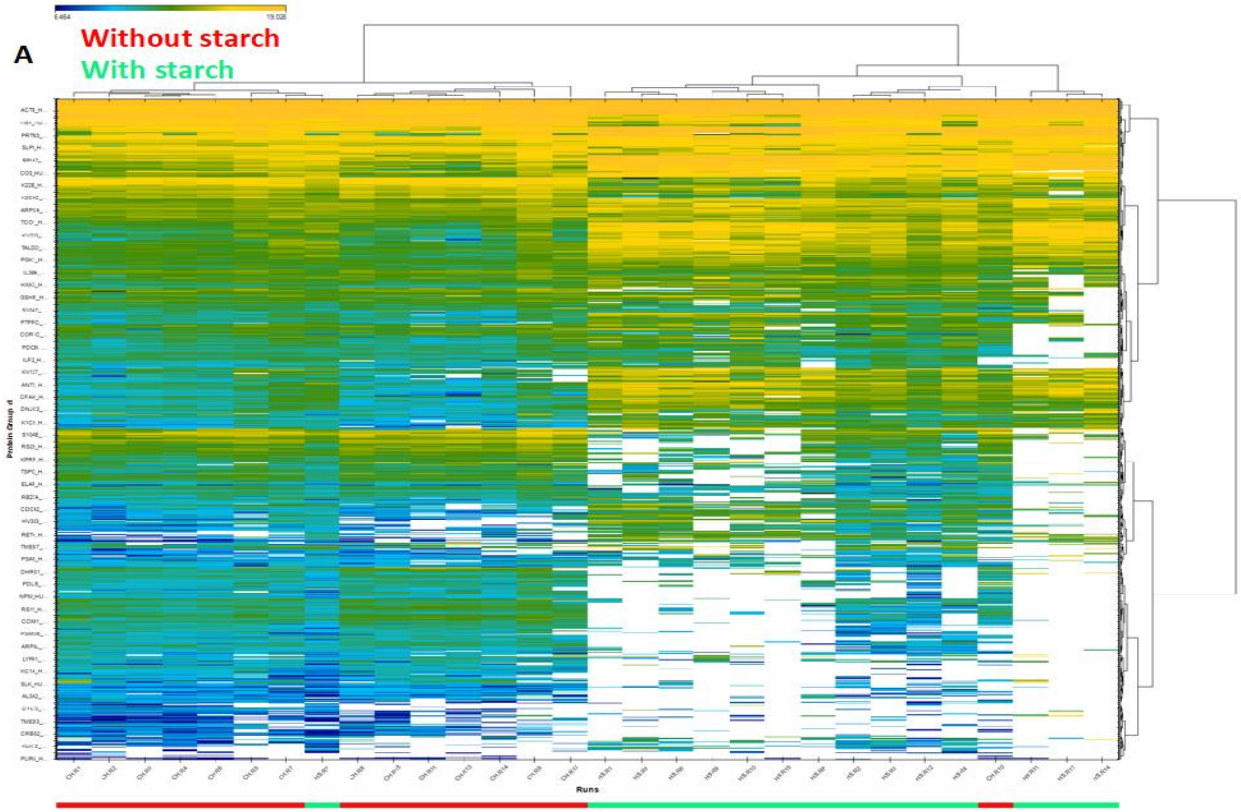
Analysis of starch treatment on the salivary proteome

The salivary proteome identified 954 proteins. Treatment with starch resulted in the enrichment of some proteins and depletion of others, which can be visualized in a heat map (**Figure 2A**). Statistical analyses between the two paired groups (with and without starch treatment) indicated that samples after starch treatment were depleted by 15 proteins and enriched by 35 (**Figure 2B**). Notably those depleted by starch treatment were associated with immunity as 60% of them were immunoglobulins (**Figure 2C**). In contrast those proteins enriched by starch treatment were associated with the keratin and ribosomal families (25.7% each, **Figure 2D**). Therefore, enriched proteins were associated with cytoskeleton organization and cytoplasmic translation (**Figure 2D**). To understand why some proteins were enriched while others were depleted, compositional bias (basic, acidic, polar, or apolar repeats), isoelectric points, glycosylation, or amino acid length were investigated to explain these results. From the

variables investigated, two were significant: amino acid length and glycosylation. Those depleted by starch had an average of 239 amino acids whereas those enriched averaged 600 ($p = 0.018$; Kruskall-Wallis). Proteins depleted by starch were mainly *N*-linked glycosylated (27% vs. 5.7%; $p = 0.036$; Chi-squared test). Thus, starch treatment depleted the samples of relatively small proteins and/or with high glycosylation.

As visualized in the heat map and due to starch treatment, the total abundance of proteins was 40% of those without starch (**Figure 2E**; $p < 0.0001$). The abundance of amylase in samples after starch treatment was decreased by about 60% (**Figure 2F**; $p = 0.069$). It could be wrongly concluded that as the abundance of all proteins/sample and that of amylase after starch treatment were reduced by a similar extent, then unspecific starch adsorption of proteins other than amylase had taken place. However, as discussed, the adsorption of proteins to starch seemed to be favored by not only its natural enzyme amylase but also by those proteins with relatively low molecular weight and high glycosylation pattern.

A



B

Genes	AVG Log2 Ratio	neg log Pvalue	neg log Qvalue	ProteinDescriptions
IGHV3-13	5.16	4.25	2.46	Immunoglobulin heavy variable 3-13
IGHV3OR16-12	4.24	3.32	1.68	Immunoglobulin heavy variable 3
IGHV3-74	3.83	-0.08	1.53	Immunoglobulin heavy variable 3-74
IGHV3-30	3.81	3.42	1.74	Ig-like domain-containing protein
AHSG	3.47	2.48	1.79	Alpha-2-HS-glycoprotein
IGKV2D-29	3.24	0.95	1.44	Immunoglobulin kappa variable 2D-29
IGKV4-1	3.23	1.87	1.38	Immunoglobulin kappa variable 4-1
PGM1	2.87	0.35	1.91	Phosphoglucomutase-1
PRSS8	2.63	3.06	1.45	Prostasin
IGHV3-35	0.45	3.37	1.72	Probable non-functional immunoglobulin heavy variable 3-35
IGHV5-51	0.45	1.46	2.01	Immunoglobulin heavy variable 5-51
CTSZ	0.37	2.99	1.40	Cathepsin Z
IGHV3-7	0.35	4.77	2.88	Immunoglobulin heavy variable 3-7
CORO1A	0.28	3.65	1.94	Coronin-1A
ACTB	0.12	3.83	3.09	Actin, cytoplasmic 1
RPL10	-0.02	8.08	6.68	60S ribosomal protein L10
ATP5F1B	-0.17	3.34	2.53	ATP synthase subunit beta, mitochondrial
LGALS7B	-0.18	4.94	4.02	Galectin-7
ANXA1	-0.20	8.33	6.74	Annexin A1
KRT14	-0.20	3.95	4.02	Keratin, type I cytoskeletal 14
KRT76	-0.22	1.91	1.33	Keratin, type II cytoskeletal 2 oral
KRT6A	-0.23	8.15	4.97	Keratin, type II cytoskeletal 6A
EPS8L2	-0.26	6.02	3.97	Epidermal growth factor receptor kinase substrate 8-like protein 2
RPL34	-0.28	2.34	1.79	60S ribosomal protein L34
HSP90AB1	-0.90	3.34	2.53	Heat shock protein HSP 90-beta
TUBB4B	-1.33	0.94	1.44	Tubulin beta-4B chain
RHCG	-1.45	4.17	3.39	Ammonium transporter Rh type C
EEF2	-1.56	10.03	6.68	Elongation factor 2
ARF3	-1.58	3.51	2.66	ADP-ribosylation factor 3
RPS11	-1.59	3.55	2.68	40S ribosomal protein S11
AHNAK	-1.59	8.15	6.68	Neuroblast differentiation-associated protein AHNAK
RPS3	-1.69	2.81	6.07	40S ribosomal protein S3
DSP	-1.70	4.97	3.95	Desmoplakin
MACROH2A1	-1.79	3.20	1.58	Core histone macro-H2A.1
RPS27	-1.86	2.32	1.78	40S ribosomal protein S27
S100A10	-1.99	1.82	3.25	Protein S100-A10
ANXA2	-2.01	2.39	1.73	Annexin A2
KRT17	-2.06	4.58	5.59	Keratin, type I cytoskeletal 17
KRT6B	-2.15	7.12	4.97	Keratin, type II cytoskeletal 6B
MAL2	-2.19	7.50	6.26	Protein MAL2
S100A14	-2.19	6.15	5.07	Protein S100-A14
RPS9	-2.22	10.60	7.40	40S ribosomal protein S9
RPS15A	-2.25	8.41	6.74	40S ribosomal protein S15a
JUP	-2.28	3.01	1.41	Junction plakoglobin
KRT5	-2.33	2.59	1.89	Keratin, type II cytoskeletal 5
HSPB1	-2.57	10.07	7.17	Heat shock protein beta-1
KRT16	-2.61	6.96	4.85	Keratin, type I cytoskeletal 16
RPL27A	-2.67	6.92	5.63	60S ribosomal protein L27a
RPS8	-2.71	5.19	4.22	40S ribosomal protein S8
KRT6C	-3.00	4.12	3.36	Keratin, type II cytoskeletal 6C

C

Term name	Term ID	p_{adj}
phagocytosis, recognition	GO:0006910	1.168×10^{-10}
complement activation, classical pathway	GO:0006958	1.647×10^{-10}
humoral immune response mediated by circulating imm...	GO:0002455	3.745×10^{-10}
phagocytosis, engulfment	GO:0006911	5.307×10^{-10}
B cell receptor signaling pathway	GO:0050853	5.938×10^{-10}
complement activation	GO:0006956	6.633×10^{-10}
plasma membrane invagination	GO:0099024	8.681×10^{-10}
membrane invagination	GO:0010324	1.244×10^{-9}
positive regulation of lymphocyte activation	GO:0051251	1.593×10^{-9}

D

Term name	Term ID	p_{adj}
intermediate filament cytoskeleton organization	GO:0045104	3.011×10^{-11}
intermediate filament-based process	GO:0045103	3.346×10^{-11}
keratinocyte differentiation	GO:0030216	2.385×10^{-10}
skin development	GO:0043588	2.741×10^{-9}
cytoplasmic translation	GO:0002181	4.886×10^{-9}
epidermal cell differentiation	GO:0009913	7.666×10^{-9}
epidermis development	GO:0008544	2.088×10^{-8}
keratinization	GO:0031424	1.286×10^{-7}
supramolecular fiber organization	GO:0097435	4.289×10^{-7}

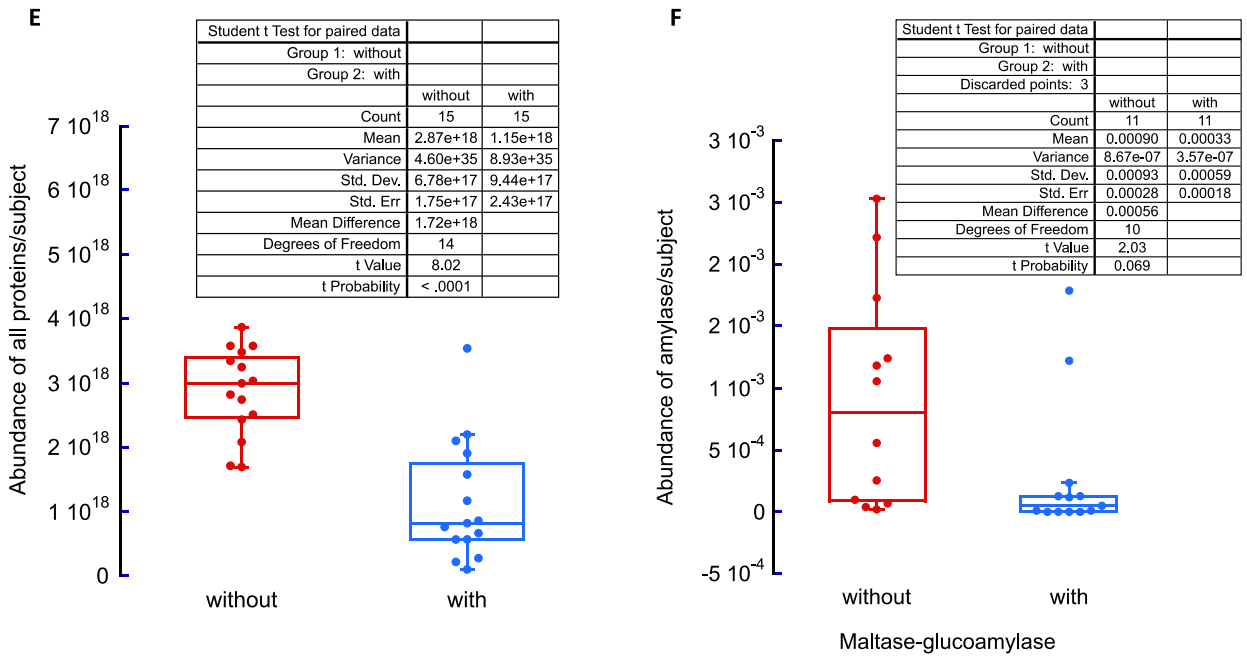


Figure 2. Effect of starch treatment on the salivary proteome of 15 females
A. Heat map visualizing the abundance of protein distribution across all samples; B. Proteins enriched or depleted by starch treatment. The first column is the gene name; the second column is the average log2 ratio of abundance before over after starch treatment; the third column is the negative log of p-value; and the fourth column is the negative log of q value; protein description according to Uniprot database. Shown in pink are immunoglobulins; in light blue are ribosomal proteins, and in dark blue are keratins. C. Gene ontology analysis of proteins depleted by starch. The analysis was performed with gProfiler. Only the top 10 genes are shown. D. Gene ontology analysis of proteins enriched by starch. This analysis was performed with gProfiler with only the top 10 shown. E. Abundance of proteins/subject with and without starch treatment. F. Abundance of amylase/subject with and without starch treatment.

Use of salivary proteome for subject identification

From the total 954 proteins detected in any sample, only those proteins (n=176) that had a value above the limit of detection in all samples, untreated and treated, were used for subject identification. Two methods of normalization were used to calculate the variances of each protein/group. The first method used Z-scores to normalize the data. The mean of the protein abundances in each sample was subtracted by the value of the protein in question and divided by the standard deviation of all proteins/sample. The second method used the sum of the protein

abundance in each sample. Each protein value was then divided by the sum of all protein abundancies in that specific sample and expressed in log scale.

For each normalization protocol, we calculated the cut-off values to filter out those proteins with low variance across subjects to maximize the differences among them (**Table 1**). For each normalization protocol, using the average of the variances of the 176 proteins/group, the upper 95% confidence interval could be calculated. The second cut-off was calculated by multiplying the first cut-off by two.

Samples	Normalization method	Average Variance	Standard Deviation	95% Confidence Interval	Cut-off	Number of proteins remaining after cut-off
Untreated no cut-off	Z-score	34.67	82.09	12.21	None	176
Untreated 1st cut-off	Z-score	34.67	82.09	12.21	46.88	36
Untreated 2nd cut-off	Z-score	34.67	82.09	12.21	93.77	14
Treated no cut-off	Z-score	15.54	70.93	10.55	None	176
Treated 1st cut-off	Z-score	15.54	70.93	10.55	26.09	15
Treated 2nd cut-off	Z-score	15.54	70.93	10.55	52.18	13

Untreated no cut-off	Sum	203.2	118.3	17.60	None	176
Untreated 1st cut-off	Sum	203.2	118.3	17.60	220.8	69
Untreated 2nd cut-off	Sum	203.2	118.3	17.60	441.6	5
Treated no cut-off	Sum	220.5	122.5	18.22	None	176
Treated 1st cut-off	Sum	220.5	122.5	18.22	238.8	64
Treated 2nd cut-off	Sum	220.5	122.5	18.22	477.5	5

Table 1. Summary of data used to calculate cut-off variances for all treatments

Principal component analysis (PCA) was created for each set of normalized samples (**Figures 3-8**). There were twelve PCAs in total, six for each normalization method. For each normalization method, three charts were prepared for the different cut-offs (none, 1st cut-off, and 2nd cut-off).

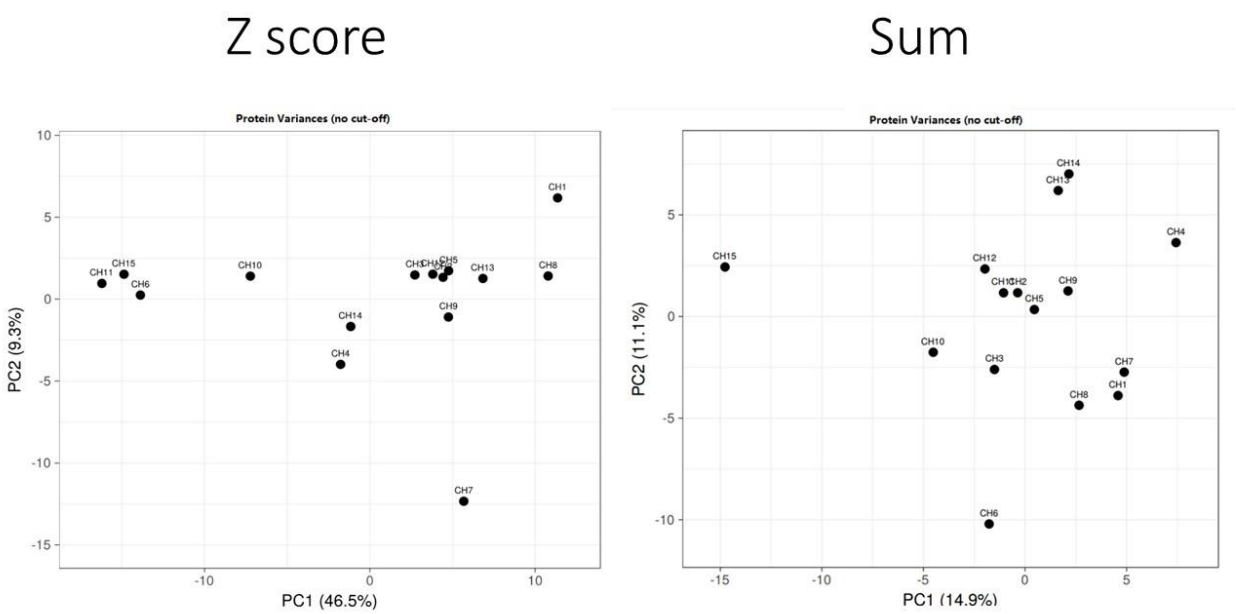


Figure 3. Principal component analysis (PCA) of samples without starch for each normalization protocol with no cut-offs

For the first method of normalization, Z scores, ~~aare~~are on the left while for the second method, sum, is on the right. Untreated samples are labeled as CH, while treated (with starch) samples are labeled as HS. The number associated with CH or HS corresponds to the sample number, and each number represents the same sample (ex. CH15 and HS15 both come from the same subject, but HS15 was treated with starch). The separation of samples is slightly better using the sum normalization here. With the Z score normalization, there are 2 clusters (CH6,10,11,15 and CH1,2,3,4,5,8,9,12,13,14) and 1 outlier (CH7), while CH2 and CH5 are overlapping. With the sum normalization, there is 1 cluster (CH1,2,3,4,5,7,8,9,10,11,12,13,14) and 2 outliers (CH6 and CH15) with no samples overlapping completely. Although there is separation, it is not sufficient for our purpose of subject identification.

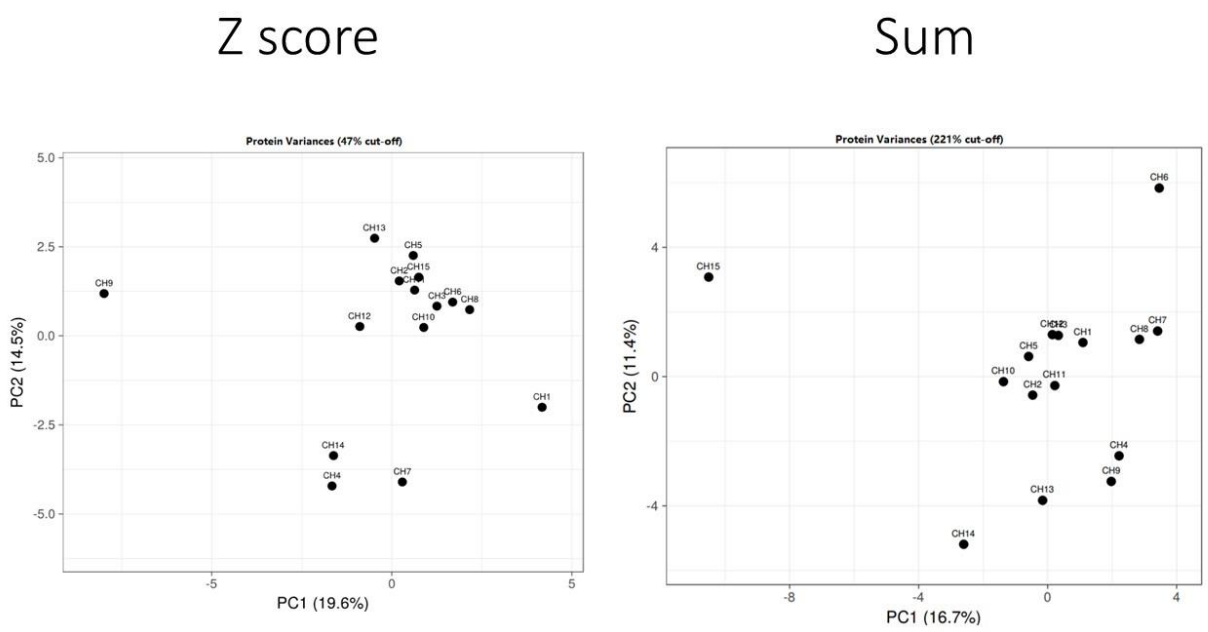


Figure 4. PCA of normalized samples without starch when using the first cut-off for each normalization protocol

For the 176 proteins, 36 proteins had a higher variance than 47 (left), while 69 proteins had a higher variance than 221 (right).

In this case, the separation using the Z score method is slightly better. There are 2 clusters (CH4,7,14, and CH2,3,5,6,8,10,11,12,13,15) and 2 outliers (CH1 and CH9), with no overlapping samples using the Z score normalization. There are 2 clusters (CH4,9,13,14 and CH1,2,3,5,7,6,10,11,12) and 2 outliers (CH6 and CH15) with CH3 and CH12 overlapping using the sum normalization. The separation is still not useful for our goal of subject identification.

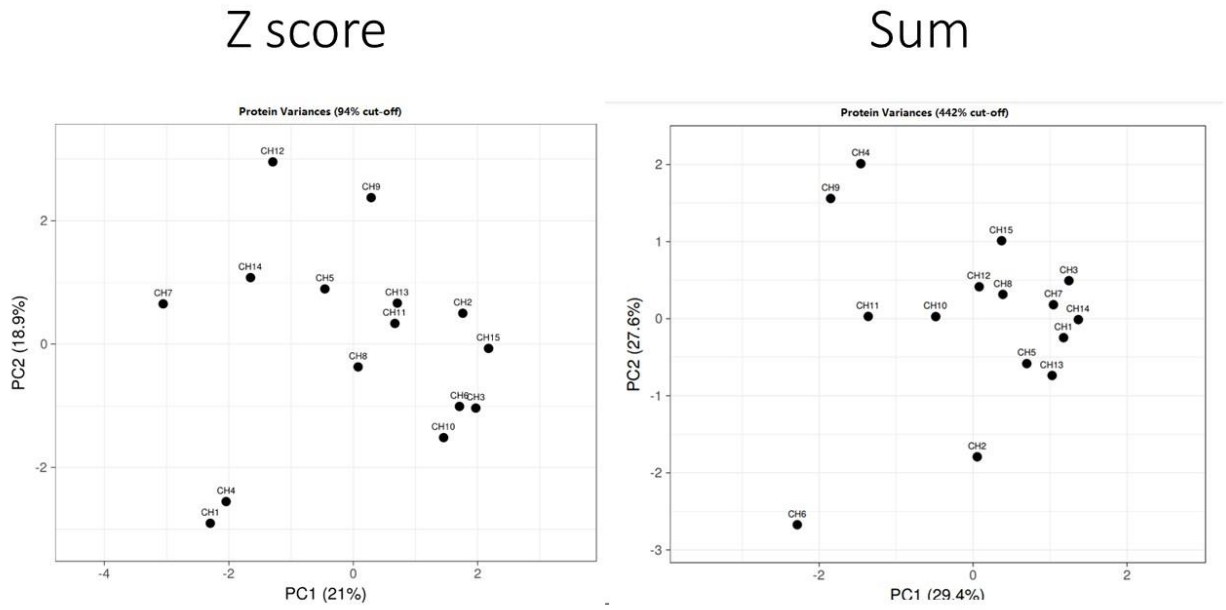


Figure 5. PCA of normalized samples without starch when using the second cut-off for each normalization protocol

Out of 176 proteins, 14 proteins had a higher variance than 94 (left), while 5 proteins had a higher variance than 442 (right).

With these latest PCAs, the distance between samples in clusters is more spaced out and makes it difficult to define clusters. With the Z score normalization, there are 3 clusters (CH1,4, CH9,12, and CH2,3,5,6,7,8,10,11,13,14,15) with no outliers or overlapping samples. With the sum normalization, there are 3 clusters (CH2,6, CH4,9, and CH1,3,5,7,8,10,11,12,13,14,15) with no outliers or overlapping samples. The sum normalization has a very slight edge on separation because the Z score normalization has more close pairs of samples (CH1,4, CH11,13, and CH3,8). This separation is suitable for subject identification.

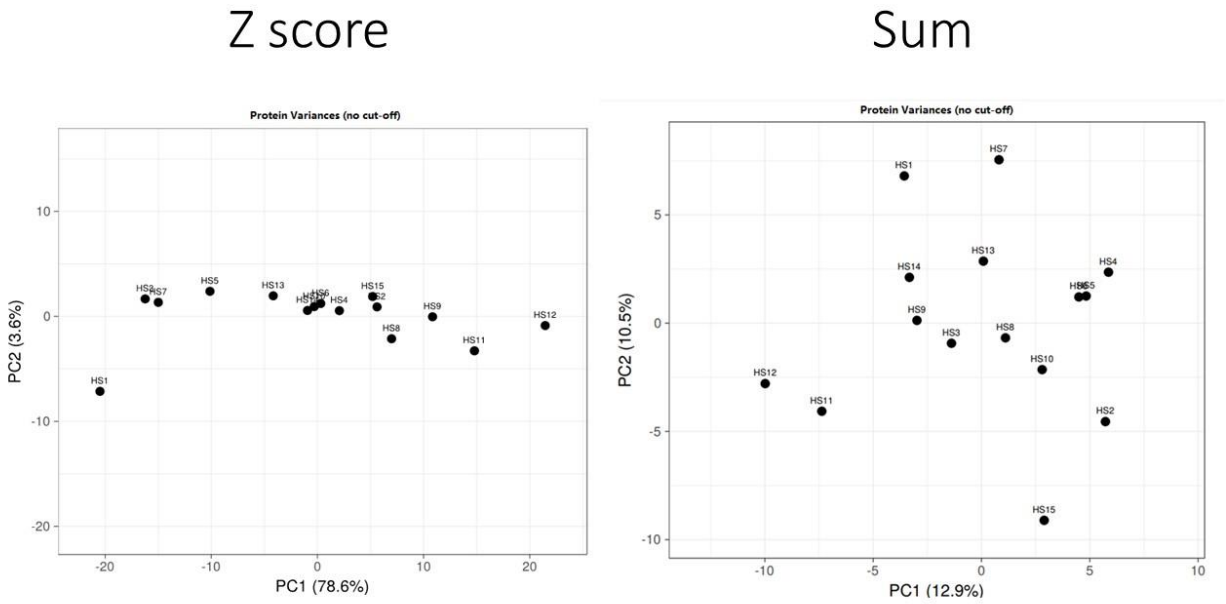


Figure 6. Principal component analysis (PCA) of samples after starch treatment for each normalization protocol with no cut-offs

The sum normalization clearly has a better separation of samples. The Z score normalization has 1 cluster with HS3,6,10,14 all overlapping. The sum normalization has 3 clusters (HS1,7, H11,12, and HS2,3,4,5,6,8,9,10,13,14) with 1 outlier (HS15) and 2 overlapping samples (HS5,6). This separation is not good enough for our purpose of subject identification.

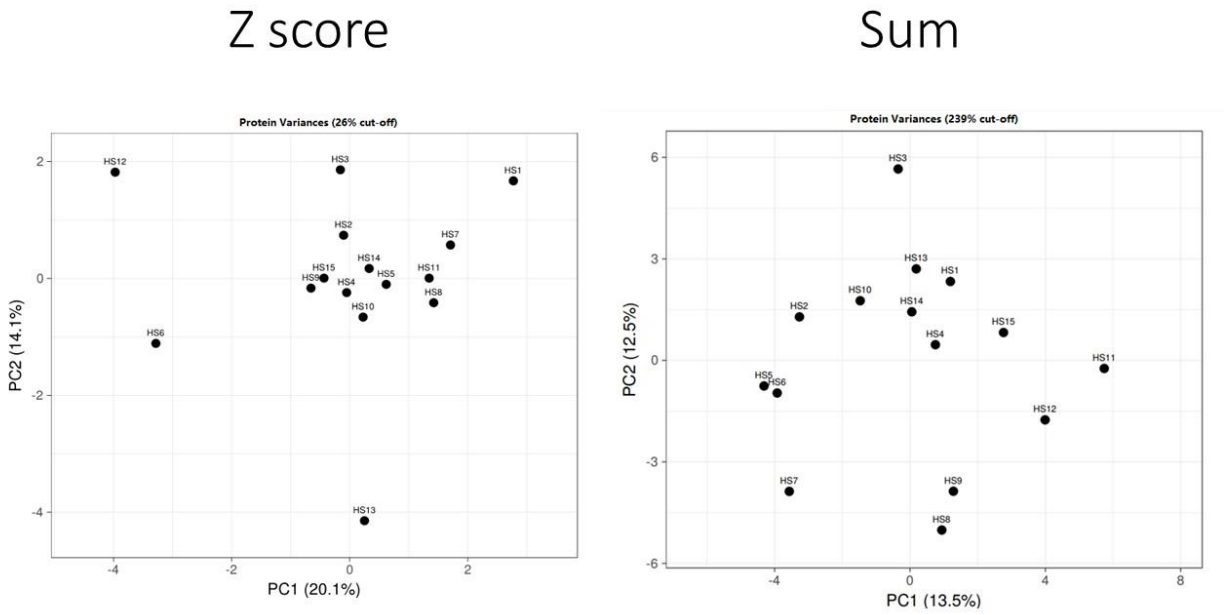


Figure 7. PCA of normalized samples with starch when using the first cut-off for each normalization protocol

Out of 176 proteins, 15 proteins had a higher variance than 26 (left), while 64 proteins had a higher variance than 239 (right).

The sum normalization is better for separation in this case. There is 1 cluster (HS1,2,3,4,5,7,8,9,10,11,14,15) and 3 outliers (HS6, HS12, and HS13) with no overlapping samples using Z score normalization. There is 1 cluster (HS1,2,4,5,6,7,8,9,10,11,12,13,14,16) and 1 outlier with no overlapping samples using sum normalization. The distance between samples in the cluster using Z score normalization is generally less than the distance between samples in the cluster using sum normalization, so the separation is better. The separation at this cut-off still is not great for subject identification.

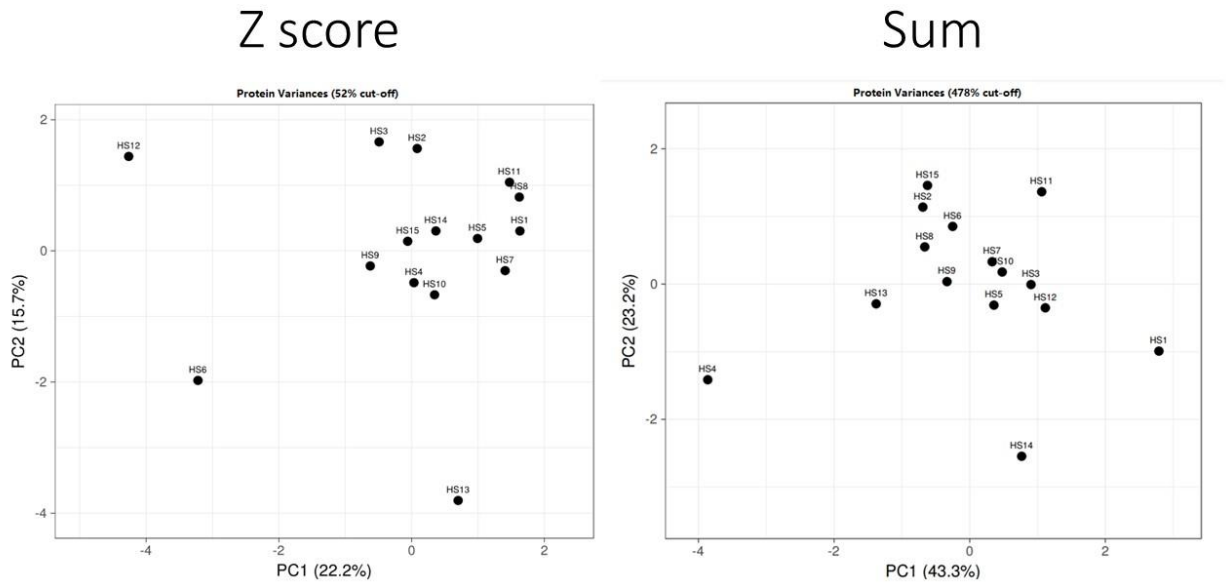


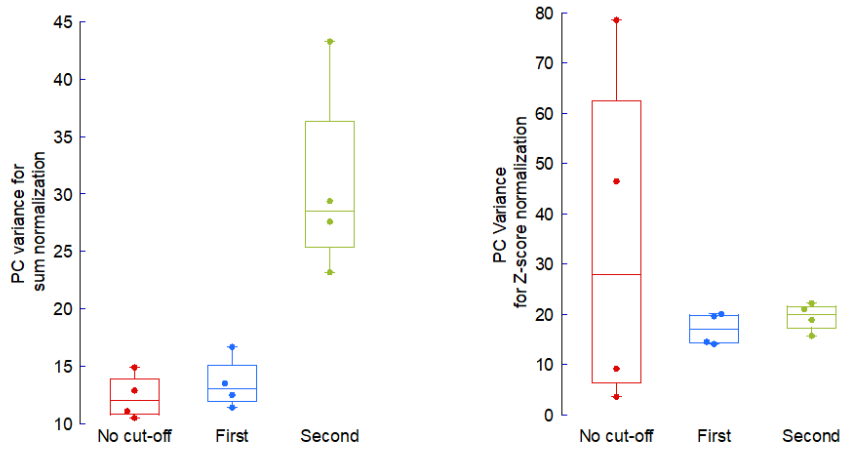
Figure 8. PCA of normalized samples with starch when using the second cut-off for each normalization protocol

Out of 176 proteins, 13 proteins had a higher variance than 52 (left), whereas 5 proteins had a higher variance than 478 (right).

Using Z score normalization, there is 1 cluster (HS1,2,3,4,5,7,8,9,10,11,14,15) and 3 outliers (HS6, HS12, HS13) with no overlapping samples. Using sum normalization, there is 1 cluster (HS2,3,5,6,7,8,9,10,11,12,13,15) and 3 outliers (HS1, HS4, and HS14) with no overlapping samples. The sum normalization is better for separation because of the significant difference between PC1 percentages (43.3% to 22.2%). The separation is sufficient for subject identification.

To compare the different cut-offs and normalization protocols, PC variances of the components 1 and 2 were analyzed (**Figure 9**). While the Z-score protocol did not result in any

differences among the mean of PC variances with different cut-offs (**Figure 9**, top right panel), the protocol utilizing the sum for normalization was significantly better when the second cut-off was used vs. no or first cut-off (**Figure 9**, left, bottom panel). Visual inspection of the PCA plots with and without starch both normalized by the sum at the second cut-off indicated that the starch treatment increased the variance of PC1 from 29.4% to 43.3% without significant changes PC2 (27.6% vs. 23.2%). As the objective of this work was to obtain the maximum dispersion of data points, the combination of the sum normalization with the second cut-off seemed the most significant (**Figure 9**) and suitable one for this purpose.



One Way ANOVA					
Data Table: Data 4					
Factor A: 3 Groups					
No cut-off, First, Second					
Analysis of Variance Results					
Source	DF	SS	MS	F	P
Total	11	1114.39	101.30		
A	2	860.77	430.38	15.27	0.0012
Error	9	253.62	28.18		
Tukey's All Pairs Comparison					
Comparison	Mean Difference	|q|	P	95% CL	
Second vs No cut-off	18.25	6.97	0.0021	8.04 to 29.00	
Second vs First	17.35	6.53	0.0032	6.86 to 27.83	
First vs No cut-off	1.175	0.44	0.947	-9.30 to 11.65	

Figure 9. PC1 and PC2 variances for each normalization protocol and cut-off utilized

Notably, from the original 176 proteins present in all samples, only 5 normalized by the sum after using the second cut-off, seemed to be sufficient to identify each sample, with (**Figure 8**) or without (**Figure 5**) starch treatment. We wanted to ascertain whether these 10 discriminating proteins obtained with and without starch treatment were related in terms of cell type and biological pathway. Ten proteins were used to mine two databases, Tabula Sapiens (for cell of origin) and Gene Ontology Biological Pathways (for function and pathway) using EnrichR.

	Gene Names	Protein names
without starch	ENO1*	Alpha-enolase
	ANXA2*	Annexin A2
	KRT16*	Keratin, type I cytoskeletal 16
	B2M	Beta-2-microglobulin
	DNAJC3	DnaJ homolog subfamily C member 3
with starch		
	H2BC12	Histone H2B type 1-K
	PSAP	Prosaposin (Proactivator polypeptide)
	CD63	CD63 antigen (Granulophysin)
	RHOA	Transforming protein RhoA
	ERO1A	ERO1-like protein alpha

Cell type database Tabula Sapiens

term	p-value	q-value	Overlap genes
Tongue-schwann Cell	4.86	3.22	CD63,ANXA2, B2M
Lung-macrophage	4.85	3.22	CD63,ANXA2, B2M
Lung-type I Pneumocyte	4.85	3.22	CD63,ANXA2, B2M
Lymph Node-endothelial Cell	4.85	3.22	CD63,ANXA2, B2M
Lymph Node-macrophage	4.85	3.22	CD63,ANXA2, B2M
Eye-limbal Stem Cell	4.85	3.22	ANXA2, ENO1, B2M
Eye-corneal Epithelial Cell	4.85	3.22	ANXA2, ENO1, B2M
Trachea-mucus Secreting Cell	4.85	3.22	CD63,ANXA2, B2M
Trachea-tracheal Goblet Cell	4.85	3.22	CD63,ANXA2, B2M
Skin-melanocyte	4.85	3.22	CD63,ANXA2, B2M

Gene Ontology Biological Pathways

term	p-value	q-value	Overlap genes
neutrophil degranulation (GO:0043312)	7.44	5.40	DNAJC3, CD63, ANXA2, PSAP, B2M, RHOA
neutrophil activation involved in immune response (GO:0002283)	7.42	5.40	DNAJC3, CD63, ANXA2, PSAP, B2M, RHOA
neutrophil mediated immunity (GO:0002446)	7.40	5.40	DNAJC3, CD63, ANXA2, PSAP, B2M, RHOA
positive regulation of receptor-mediated endocytosis (GO:0048260)	5.93	4.05	CD63, ANXA2, B2M
positive regulation of receptor binding (GO:1900122)	5.47	3.69	ANXA2, B2M
modulation by symbiont of host process (GO:0044003)	5.00	3.30	ENO1, B2M
regulation of vacuole organization (GO:0044088)	4.29	2.65	ANAXA2, ENO1
negative regulation of neurogenesis (GO:0050768)	3.88	2.30	B2M, RHOA
epithelial cell migration (GO:0010631)	3.74	2.22	KRT16, RHOA
innate immune response (GO:0045087)	3.42	1.95	H2BC12, KRT16, B2M

Figure 10. Ten discriminating proteins and their cell origin and biological pathway
The 10 discriminating proteins obtained after sum normalization and filtered by the second cut-off (top left) were used to identify the cell of origin (top right) and the biological pathway (bottom right). The p-values and q-values of significant terms are shown for the selected libraries. The q-value was an adjusted p-value calculated using the Benjamini-Hochberg method for correction for multiple hypotheses testing. Only the top 10 significant results are displayed in this Table. Marked with * are those proteins among the most discriminating ones from our previous study (34).

From the 10 proteins identified in this study, 3 were identified before by us in a previous study as having discriminating power across subjects (34) (**Figure 10**, marked with asterisk). Statistical analyses indicated that some of these proteins were associated with Schwann cells in the tongue having a role in immunity.

DISCUSSION

The treatment of human saliva samples with starch was intended to remove amylase to enhance the detection of less abundant proteins as a means to facilitate subject identification through implementation of the salivary proteome.

. While the treatment with starch decreased 63% of all protein abundance, similar to the decrease in amylase content, the proteins depleted from the starch treatment had relatively low molecular weight and high glycosylation.

Two normalization methods were used to analyze the suitability of the data for subject identification. The purpose of the normalization protocol was to create PCAs with clear separation across different subjects. The two methods used, Z score and sum, only produced useful results after utilizing a second variance cut-off. With no cut-off, PC1 was over 30% (high variance) with and without starch using Z score normalization. However, the separation in these PCAs was not optimal. With the first cut-off, PC1 was below 30% using Z score normalization. Using sum normalization, PC1 was below 30% with no and the first cut-off with and without starch. The second cut-off using Z score normalization had better separation but PC1 was below 30%. Sum normalization created better separation compared to Z score normalization five out of six times, making it a more suitable protocol for the purposes of this study. The best separation and highest variation was PCA using sum normalization and after the second variance cut-off (**Figure 9**).

From the ten proteins with the most discriminating power (**Figure 10**), seven were not identified as discriminating in our previous study (34). This could be because of starch treatment depleting the smaller and glycosylated proteins. Most of these proteins had roles in the immune response (CD63, DNAJC3, ANXA2, PSAP, B2M, RHOA, ENO1, H2BC12, KRT16). Six of the

proteins (DNAJC3, CD63, ANXA2, PSAP, B2M, RHOA) share the neutrophil degranulation pathway. Notably, two proteins (ENO1, CD63) are also found to be related to diseases (endometriosis, cancer-associated retinopathy, Hashimoto encephalopathy, lung carcinoma, Hermansky-Pudlak syndrome) suggesting that this could be another piece of evidence to be added when trying to identify a subject.

While the use of saliva fingermarks (the salivary proteome of each individual) has the potential to identify subjects, further research is required to determine if they are more useful than DNA profiles. DNA degradation is the most important reason for finding alternative solutions, and this study did not evaluate the change of the proteome with exposure to blue light (450 nm) or to the reagent on Phadebas paper, direct methods used to locate saliva stains, on the stability of the protein markers. Also, this study did not evaluate the sensitivity and stability of the protein markers under various storage conditions, however, several studies showed no differences in the proteomes of dry or wet saliva samples (35,36). Further research should also test other populations of different gender and larger sizes.

Amylase removal is also the next step in research from this study (37), indicating it "may increase the stability of other salivary proteins and eases the characterization of low abundant proteins." Our study did show that starch treatment could have increased stability because the treatment also depleted two proteases (**Figure 2B**): prostasin and cathepsin Z. It is possible that by lowering the content proteases, the stability of the left proteins is enhanced. Another study also found that amylase removal enhanced the stability of proteins (38). Our study shows that amylase removal does not allow for smaller proteins to appear in abundance, as the proteins that were depleted had lower molecular weight (**Figure 2E-F**). The findings of our study somewhat agree with the expectations put forth by Hu.

Treating samples with starch intended to find more varied proteins to use for saliva fingermarks. This study showed that some protein abundance was lost, but some proteins were enriched because of the starch treatment. This study comes close to offering a viable alternative to DNA fingerprinting.

References

1. Anzai-Kanto E, Hirata MH, Hirata RD, Nunes FD, Melani RF, Oliveira RN. DNA extraction from human saliva deposited on skin and its use in forensic identification procedures. Braz Oral Res. 2005 Jul-Sep;19(3):216-22. doi: 10.1590/s1806-83242005000300011.
2. Van Steendam K, De Ceuleneer M, Dhaenens M, Van Hoofstat D, Deforce D. Mass spectrometry-based proteomics as a tool to identify biological matrices in forensic science. Int J Legal Med. 2013 Mar;127(2):287-98. doi: 10.1007/s00414-012-0747-x.
3. Chatterjee S. Saliva as a forensic tool. J Forensic Dent Sci. 2019;11(1):1-4. doi:10.4103/jfo.jfds_69_18
4. Toennes SW, Steinmeyer S, Maurer HJ, Moeller MR, Kauert GF. Screening for drugs of abuse in oral fluid – Correlation of analysis results with serum in forensic cases. J Anal Toxicol. 2005;29:22–7.
5. Lee YH, Zhou H, Reiss JK, Yan X, Zhang L, Chia D, et al. Direct saliva transcriptome analysis. Clin Chem. 2011;57:1295–302.
6. Spradbury P. Restriction fragment length polymorphisms of mutans streptococci in forensic odontological analysis. Biosci Horiz. 2010;3:166–78.
7. Scannapieco FA, Torres G, Levine MJ. Salivary α-amylase: role in dental plaque and caries formation. Crit Rev Oral Biol Med 1993;4(3):301–7.
8. Peyrot des Gachons C, Breslin PA. Salivary Amylase: Digestion and Metabolic Syndrome. Curr Diab Rep. 2016;16(10):102. doi:10.1007/s11892-016-0794-7
9. Wornes DJ, Speers SJ, Murakami JA. The evaluation and validation of Phadebas((R)) paper as a presumptive screening tool for saliva on forensic exhibits. Forensic Sci Int. 2018 Jul;288:81-8. doi: 10.1016/j.forsciint.2018.03.049.
10. Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993;362(6422):709–15. Epub 1993/04/22. 10.1038/362709a0
11. Parker GJ, Leppert T, Anex DS, Hilmer JK, Matsunami N, Baird L, Stevens J, Parsawar K, Durbin-Johnson BP, Rocke DM, Nelson C, Fairbanks DJ, Wilson AS, Rice RH, Woodward SR, Bothner B, Hart BR, Leppert M. Demonstration of Protein-Based Human Identification Using the Hair Shaft Proteome. PloS one. 2016;11(9), e0160653. https://doi.org/10.1371/journal.pone.0160653
12. Mason KE, Anex D, Grey T, Hart B, Parker G. Protein-based forensic identification using genetically variant peptides in human bone. Forensic Sci Int. 2018 Jul;288:89-96. doi: 10.1016/j.forsciint.2018.04.016. Epub 2018 Apr 22. PMID: 29738994.

13. Borja T, Karim N, Goecker Z, Salemi M, Phinney B, Naeem M, Rice R, Parker G. Proteomic genotyping of fingermark donors with genetically variant peptides, Forensic Sci Int: Genetics. 2019;42: 21-30,ISSN 1872-4973, https://doi.org/10.1016/j.fsigen.2019.05.005.
14. Messana I, Inzitari R, Fanali C, Cabras T, Castagnola M. Facts and artifacts in proteomics of body fluids. What proteomics of saliva is telling us? J Sep Sci. 2008; 31(11):1948–63. doi: 10.1002/jssc. 200800100 PMID: 18491358
15. Hoek GH, Brand HS, Veerman ECI, Nieuw Amerongen AV. Toothbrushing affects the protein composition of whole saliva. Eur J Oral Sci. 2002; 110(6):480–1. PMID: 12507223
16. Rudney JD. Does variability in salivary protein concentrations influence oral microbial ecology and oral health? Crit Rev Oral Biol M. 1995; 6(4):343–67. doi: 10.1177/10454411950060040501
17. Dodds MWJ, Johnson DA, Yeh C-K. Health benefits of saliva: a review. J Dent. 2005; 33(3):223–33. http://dx.doi.org/10.1016/j.jdent.2004.10.009. PMID: 15725522
18. Caseiro A, Ferreira R, Padrão A, Quintaneiro C, Pereira A, Marinheiro R, Vitorino R, Amado F. Salivary proteome and peptidome profiling in type 1 diabetes mellitus using a quantitative approach. J Proteome Res. 2013 Apr 5;12(4):1700-9. doi: 10.1021/pr3010343. Epub 2013 Feb 25. PMID: 23406527.
19. Oppenheim, F. G., Salih, E., Siqueira, W. L., Zhang, W., & Helmerhorst, E. J. Salivary Proteome and Its Genetic Polymorphisms. Ann. N. Y. Acad. Sci. 2007; 1098(1): 22–50.
20. Woof JM, Kerr MA. The function of immunoglobulin A in immunity. J Pathol. 2006; 208(2): 270-82.
21. Oppenheim FG, Xu T, McMillian FM, Levitz SM, Diamond RD, Offner GD, Troxler RF. Histatins, a novel family of histidine-rich proteins in human parotid secretion. Isolation, characterization, primary structure, and fungistatic effects on Candida albicans. J Biol Chem. 1988 Jun 5;263(16):7472-7. PMID: 3286634.
22. Tsai H, Bobek LA. Human salivary histatins: promising anti-fungal therapeutic agents. Crit Rev Oral Biol Med. 1998;9(4):480-97. doi: 10.1177/10454411980090040601. PMID: 9825223.
23. Hay DI, Smith DJ, Schluckebier SK, Moreno EC. Relationship between concentration of human salivary statherin and inhibition of calcium phosphate precipitation in stimulated human parotid saliva. J Dent Res. 1984 Jun;63(6):857-63. doi: 10.1177/00220345840630060901. PMID: 6429216.
24. Schlesinger DH, Hay DI. Complete covalent structure of statherin, a tyrosine-rich acidic peptide which inhibits calcium phosphate precipitation from human parotid saliva. J. Biol. Chem. 1997; 252: 1689–1695.
25. Hay, DI, Schluckebier SK, Moreno EC. Equilibrium dialysis and ultrafiltration studies of calcium and phosphate binding by human salivary proteins. Implications for salivary supersaturation with respect to calcium phosphate salts. Calcif. Tissue Int. 1982;34: 531–538. 57.

26. Moreno, EC, Varughese K, Hay DI. Effect of human salivary proteins on the precipitation kinetics of calcium phosphate. Calcif. Tissue Int. 1979; 28: 7–16.
27. Dickinson, DP. Cysteine peptidases of mammals: their biological roles and potential effects in the oral cavity and other tissues in health and disease. Crit. Rev. Oral Biol. Med. 2002; 13: 238–275.
28. Lamkin, MS, Oppenheim FG. Structural features of salivary function. Crit. Rev. Oral Biol. Med. 1993; 4: 251–259.
29. Wong RS, Bennick A. The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A. J Biol Chem. 1980 Jun 25;255(12):5943-8. PMID: 7380845.
30. Reddy MS, Bobek LA, Haraszthy GG, Biesbrock AR, Levine MJ. Structural features of the low-molecular-mass human salivary mucin. Biochem J. 1992;287 (Pt 2). 639-43. 10.1042/bj2870639.
31. Offner GD, Nunes DP, Keates AC, Afdhal NH, Troxler RF. The amino-terminal sequence of MUC5B contains conserved multifunctional D domains: implications for tissue-specific mucin functions. Biochem Biophys Res Commun. 1998 Oct 9;251(1):350-5. doi: 10.1006/bbrc.1998.9469. PMID: 9790959.
32. Trochimiak T, Hübner-Woźniak E. Effect of exercise on the level of immunoglobulin a in saliva. Biol Sport. 2012;29(4):255-261. doi:10.5604/20831862.1019662
33. Yel L. Selective IgA deficiency. J Clin Immunol. 2010;30(1):10-16. doi:10.1007/s10875-009-9357-x
34. Thomas C, Giulivi G. Saliva protein profiling for subject identification and potential medical applications. 2021 Medicine in Omics (3). https://doi.org/10.1016/j.meomic.2021.100012.
35. Schulte F, Hasturk H, Hardt M. Mapping Relative Differences in Human Salivary Gland Secretions by Dried Saliva Spot Sampling and nanoLC-MS/MS. Proteomics. 2019 Oct;19(20):e1900023. doi: 10.1002/pmic.201900023.
36. Hedman J, Jansson L, Akel Y, Wallmark N, Gutierrez Liljestrand R, Forsberg C, et al. The double-swab technique versus single swabs for human DNA recovery from various surfaces. Forensic Sci Int Genet. 2020 May;46:102253. doi: 10.1016/j.fsigen.2020.102253
37. Hu S, Loo J, Wong D. Human body fluid proteome analysis. Proteomics. 2006;6(23):6326-6353. doi: 10.1002/pmic.200600284.
38. Xiao H, Wong D. Method development for proteome stabilization in human saliva. Analytica Chimica Acta. 2012;722:63-69. https://doi.org/10.1016/j.aca.2012.02.017