SalivaPRINT Toolkit – Protein profile evaluation and phenotype stratification

Igor Cruz[1], Eduardo Esteves[1], Mónica Fernandes[1], Nuno Rosa[1], Maria José Correia[1], Joel P. Arrais[2], Marlene Barros[1*]

[1] Universidade Católica Portuguesa, Center for Interdisciplinary Research in Health (CIIS), Institute of Health Sciences (ICS), Viseu, Portugal

[2] Department of Informatics Engineering (DEI), Centre for Informatics and Systems of the University of Coimbra (CISUC), University of Coimbra, Coimbra, Portugal

* Corresponding author

Address correspondence to: Marlene Barros, PhD, Director of Center for Interdisciplinary Research in Health (CIIS), Senior Scientist at SalivaTec Universidade Católica Portuguesa, Estrada da Circunvalação 3504-505 Viseu – Portugal; Tel. +351232430200 - Fax +351232428344

email: mbarros@viseu.ucp.pt

Abstract

The value of the molecular information obtained from saliva is dependent on the use of *in vitro* and *in silico* techniques. The main proteins of saliva when separated by capillary electrophoresis enable the establishment of individual profiles with characteristic patterns reflecting each individual phenotype. Different physiological or pathological conditions may be identified by specific protein profiles. The association of each profile to the particular protein composition provides clues as to which biological processes are compromised in each situation. Patient stratification according to different phenotypes often within a particular disease spectrum is especially important for the management of individuals carrying multiple diseases and requiring personalized interventions. In this work we present the SalivaPRINT Toolkit, which enables the analysis of protein profile patterns and patient phenotyping. Additionally, the SalivaPRINT Toolkit allows the identification of molecular weight ranges altered in a particular condition and therefore potentially involved in the underlying dysregulated mechanisms. This tutorial introduces the use of the SalivaPRINT Toolkit command line interface (https://github.com/salivatec/SalivaPRINT) as an independent tool for electrophoretic protein profile evaluation. It provides a detailed overview of its functionalities, illustrated by the application to the analysis of profiles obtained from a healthy population *versus* a population affected with inflammatory conditions.

Keywords

Protein profiling; Protein pattern recognition; Protein phenotypes

1. Introduction

In the age of precision medicine, diagnostics are based on the use of big data from genomic, proteomic and transcriptomic research. These techniques enable the establishment of molecular phenomes associated with different functional profiles which characterize the phenotypes of individuals sharing similar conditions and may direct a personalized intervention.

Omics results have revealed information on molecules which are dysregulated in certain health and disease situations. This information is stored in several public databases [1–4].

Saliva is a fluid increasingly used in diagnostics [5]. Several techniques have been used to acquire molecular information from this fluid. Such information is available in several public databases such as OralOme [3,6] or SalivaOmics [7]. One of the techniques used to characterize the main protein content of saliva is electrophoresis, and capillary electrophoresis is one of the most sensitive variants. Despite the wide availability of capillary electrophoresis-based techniques, the challenge remains in the exploration of the technique´s full power. In particular, the fact that the currently available tools for result analysis require manual and visual inspection of the profiles and are not amenable to high throughput result analysis, has created a bottleneck in the generation of powerful analysis of the results from large number of profiles such as those generated in large population studies.

Few studies have been developed to surpass these problems mainly in the analysis of nucleic acid results [8,9] but also for total protein profiles [10,11].

In spite of the existence of studies to recognize patterns of capillary electrophoresis profiles [12] there is, to our knowledge, no approach developed and applied to the use of total protein profiles of complex samples for patient stratification or sample quality control.

The possibility of establishing protein profile patterns corresponding to specific clinical situations is an opportunity for the development of new diagnostics strategies essential for the analysis of large samples characteristic of population wide and large epidemiologic studies.

The Experion™ automated electrophoresis system [12] (from Bio-Rad

Laboratories, USA) was used to provide the data in the example presented in this tutorial. This system integrates protein analysis into a single process in which protein separation, staining, band detection and quantitation are automatically executed and produces protein profiles in about 30 minutes (10 samples) through an automated process.

By performing capillary electrophoresis it is possible to obtain a protein profile of the sample within the molecular weights (MW) in the range of 10–260 kiloDaltons (kDa) while separating and detecting protein concentrations in the 2.5–2000 ng/mL range [12].

The system software is responsible for plotting the fluorescence index as a function of migration time to produce an electropherogram. A virtual gel image is generated from the electropherogram data. Proteins bands or peaks are identified by migration time relative to the known MW markers.

After running the samples, relevant peak heights and density of protein bands are calculated by the software and the output is exported in a file containing multiple information such as MW, peak height, protein concentration, and total sample concentration among others. This information can be used with data analysis techniques in order to characterize each individual and/or the population to which it belongs.

Capillary electrophoresis technology has been used efficiently to detect *Listeria monocytogene*s in foods [13] and to measure ovarian cancer or cancer-related proteins biomarkers in serum [14], however the methodology followed for result processing was to manually select individuals and check which molecular weights were different according to the individual’s conditions.

The development of solutions for automatic analysis of the results produced by capillary electrophoresis technology, to obtain typical profiles or molecular weight ranges, revealing altered protein quantities, are a first approach to evaluate the functional status of each individual. These solutions are also useful for the identification of the molecular weight ranges in which there are dysregulated proteins associated to specific pathologies or phenotypes and therefore may be used for diagnosis or stratification.

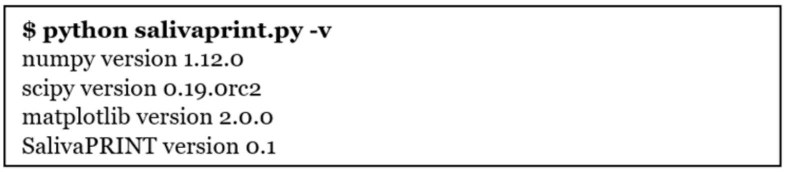
SalivaPRINT Toolkit provides a set of functionalities to analyze the output data provided by capillary electrophoresis techniques. This tool can be widely applied for the analysis of data from protein separation techniques resulting in an output of migration/molecular weight data and respective protein quantification in each sample.

2. The SalivaPRINT Toolkit command line tools

a. Installation

SalivaPRINT Toolkit command line tools are written in Python and work on Windows, macOS and Unix. Python 3.0 (https://www.python.org/downloads/) is required along with the modules numpy (http://www.numpy.org/), scipy (https://www.scipy.org/), configparser (https://docs.python.org/2/library/configparser.html) and matplotlib (https://matplotlib.org/).

After successful installation of Python and the required libraries for running the program, the user should decompress the file *salivaprint.zip* to a new directory and use the salivaprint.py as a normal program passing commands as arguments. In order to check if everything is working properly, the command *salivaprint.py –v* should print the version number as follows.



```
$ python salivaprint.py -v
numpy version 1.12.0
scipy version 0.19.0rc2
matplotlib version 2.0.0
SalivaPRINT version 0.1
```

b. Available commands

SalivaPRINT Toolkit is a command line tool, which allows data extraction and analysis from capillary electrophoresis systems output files.

The functionalities available allow the construction of a matrix of molecular weights from an output file provided by Experion™ systems, which can then be used with data analysis and machine learning tools in order to find similarities between individuals and/or populations. By implementing a naïve Bayes classification algorithm, a probabilistic classifier based on the application of the Bayes' theorem with strong independence assumptions between features, it becomes possible to achieve an overview of important features for the stratification of the individuals in study.

SalivaPRINT Toolkit available commands can be checked anytime by using **-h** as argument. The following commands are currently implemented (version 0.1):

- **-v**: Displays the program and required libraries version;
- **-h**: Displays the help menu. Lists the available commands;
- **-build** output_file: Builds a new molecular feature matrix from capillary electrophoresis output files using *config.cfg* as the configurations file;

-view input_file: Shows a visual representation of the dataset previously built using the **–build** flag;

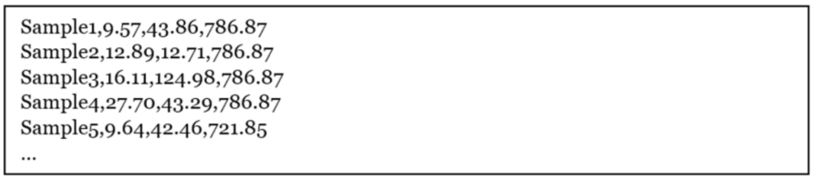
-learn input_file output_file: Builds a classifier from input_file dataset. Uses the name given as output_file for saving the created classifier;

-classify classifier_file dataset: Classifies the dataset using the previously trained classifier.

c. Dataset preparation

The main data file accepted by SalivaPRINT Toolkit is composed by a Comma Separated File (**CSV**) file with peak information collected with Experion™ (or other equivalent system) in the format: Sample, Molecular Weight, Protein Concentration, Sample Concentration without header information.

An example is shown below.



```
Sample1,9.57,43.86,786.87
Sample2,12.89,12.71,786.87
Sample3,16.11,124.98,786.87
Sample4,27.70,43.29,786.87
Sample5,9.64,42.46,721.85
...
```

Linux command line tools provides an easy way to prepare the Experion™ output files as datasets which can be used with SalivaPRINT Toolkit. Assure the use of a **CSV** format files containing the data encoded to **UTF-8** with Unix Line Feed **(LF)** as line break special characters. Note that it is important to use this file encoding since the **awk** language for processing text, available on the standard

Linux bash, may fail to correctly recognize columns if the file encoding is not correctly set.

Using **awk** is a fast option to select the correct columns for creating the dataset file. The following command selects rows 7,10,13 and 17 from all the data available. Note that these row positions (7,10,13,17) correspond to the columns which provide information as sample name, MW, concentration and sample concentration in the standard output file, and are the ones we need in order to use SalivaPRINT Toolkit .

```
awk -F',' '{print $7,$10,$13,$17}' output_experion.csv > dataset.csv
```

d. Configurations file

Config.cfg is the file that contains all the configurations necessary for the program to run. In order to extract data from the original MW from the capillary electrophoresis output file the following configurations are necessary.

MIN_MOL_WEIGHT – (Default 9) Minimum molecular weight, defined in kDa to consider while extracting data from the input dataset file.

MAX_MOL_WEIGHT – (Default 120) Maximum molecular weight, defined in kDa to consider while extracting data from the input dataset file.

N_SLICES – (Default 120) Number of slices to consider from the MIN_MOL_WEIGHT to MAX_MOL_WEIGHT.

DATASET - Input file containing all the capillary electrophoresis molecular weights at which protein concentration peaks occur.

CONTROL – A list of healthy individuals, or control individuals, present in the DATASET file. It should contain the sample IDs as found in the DATASET one by each line. Ideally, it should have the same length of STUDY list for generating a balanced classifier.

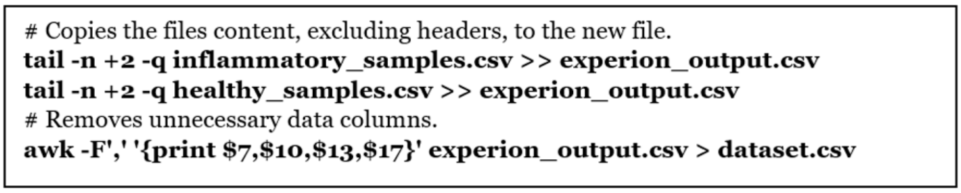
STUDY - A list of unhealthy, or disease carrier individuals, present in the DATASET file. It should contain the sample IDs as found in the DATASET one by each line. Ideally, it should have the same length of CONTROL list for generating a balanced classifier.

3. Case study: What can we learn from patients with inflammatory conditions?

In order to build this tutorial, 184 salivary electrophoretic profiles from Experion™ automated electrophoresis system were used. The data was split into two classes regarding the health status of individuals. The healthy population was composed of 92 individuals without acute or chronic inflammation, as far as could be discerned from the clinical history, ranging from 18 to 89 years of age (average: 23.7, standard deviation: 9.4). The unhealthy population was represented by 92 individuals, ranging from 7 to 84 years of age (average: 39.4, standard deviation: 25.3). These individuals presented a broad spectrum of diseases, from oral problems such as gingivitis, to whole systemic and chronic diseases as diabetes or celiac disease, all related to an underlying inflammatory condition.

a) Preparing the dataset

For this part of the tutorial, the saliva protein profiles from 164 individuals (82 healthy and 82 inflammatory), were used. Considering that we have two output files from SalivaPRINT Toolkit, one for patients with inflammation and one without, we can process them using the following commands:



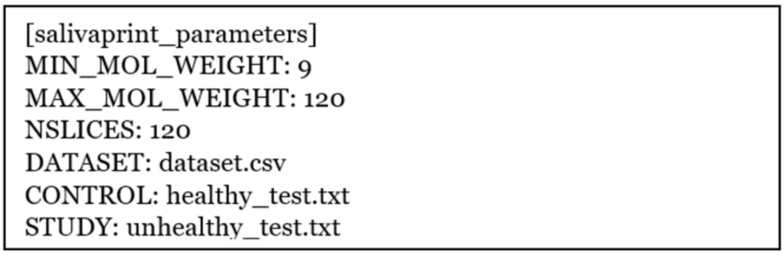
```
# Copies the files content, excluding headers, to the new file.
tail -n +2 -q inflammatory_samples.csv >> experion_output.csv
tail -n +2 -q healthy_samples.csv >> experion_output.csv
# Removes unnecessary data columns.
awk -F',' '{print $7,$10,$13,$17}' experion_output.csv > dataset.csv
```

After this procedure, dataset.csv should have the format shown in 2.c) and should be is ready to be used with SalivaPRINT Toolkit.

b) Building the Molecular Weight Matrix

First, it is necessary to properly set the configurations file. Using a minimum MW of 9 kDa and a maximum MW of 120 kDa with 120 slices we will get a description of each individual protein profile. Experion™ does not account for MW below 10kDa (~9.5) and identifications with MW above 120 kDa since these larger MW are often protein aggregates easily formed in saliva [16]. Note that if using a different sample it may be useful to include MWs above 120 kDa.

The configurations used are shown below.



```
[salivaprint_parameters]
MIN_MOL_WEIGHT: 9
MAX_MOL_WEIGHT: 120
NSLICES: 120
DATASET: dataset.csv
CONTROL: healthy_test.txt
STUDY: unhealthy_test.txt
```

The next step is to run **SalivaPRINT** Toolkit **–build matrix.csv** using the standard configurations available in the configurations file. Make sure you build two lists of individuals using the same IDs provided on the dataset file and edit the config.cfg file to point to these files. One should list the healthy individuals and the other the unhealthy. The program will then use the dataset in order to build a matrix of relative concentration of protein per MW. This matrix represents the presence of a ratio of protein.

By using the command **salivaprint.py –view matrix.csv** is possible to obtain a visual representation of the matrix created.

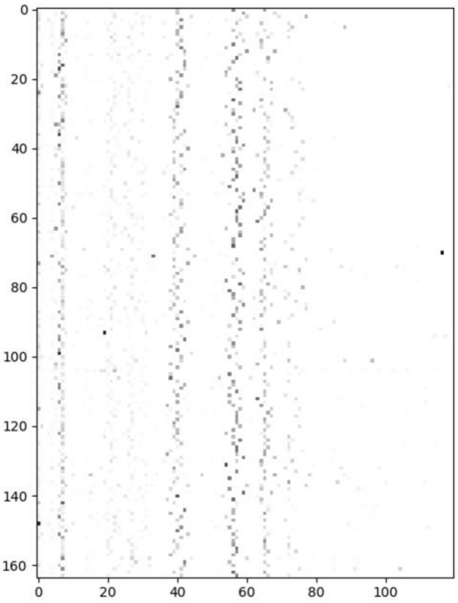


Figure 1 - Graphical representation of the population. Each line represents one individual and each column represents a small range of molecular weights (In this case approximately 1kDa).

c) Creating a Classifier

Using this matrix, which represents protein concentration per MW per individual, it is possible to use SalivaPRINT Toolkit and create a classifier. The command **salivaprint.py –learn matrix.csv classifier.pkl** will create a naïve Bayes classifier with the examples provided in the matrix.csv file and save it with the name classifier.pkl. When a Graphical User Interface (GUI) is available, it will also show a graphical representation of the influence of each MW towards the classification of samples according to condition state (healthy or inflammatory). In Figure 2, we show the influence of molecular weights over the classification of healthy individuals and individuals with inflammatory conditions obtained from the dataset used on this tutorial. Y-axis values correspond to the influence of each MW as learned from the naïve Bayes classifier. Negative values are associated to the influence of a given MW over the population in study, in this case a population with inflammatory conditions, while positive values are associated towards the control population, in this case the healthy population.

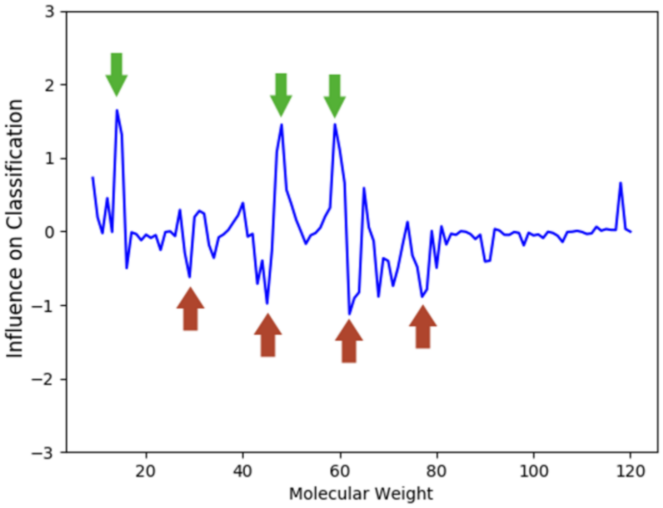


Figure 2 - Graphical representation of the MW influence towards classification of individuals. The green arrows point to MWs related to a tendency towards healthiness and the red arrows to MW related with inflammatory states.

From this graphic representation it is possible to analyze the influence of different MW towards the classification of individuals given their protein profiles. Profiles containing some of the same MW as the positive values on Figure 2 are expected to be related to healthy individuals and profiles containing some of the same MWs as the negative values are expected to be related with individuals suffering from inflammatory conditions.

d) Using a Classifier With a Different Dataset

In table I a set of molecular weights and proteins within the ranges identified in the previous section is shown. In this table, the corresponding proteins are absent or present in different quantities. These MW ranges with the greatest variability in the proteins present enable through the identification of which proteins are present (using Omics databases) and the potentially compromised molecular mechanisms. The potentially dysregulated proteins presented in each MW range were identified according to the data from Rosa *et. al,* 2016 [15]. Proteins with molecular weights with a ± 8.56% interval were considered since this is the largest variation in Experion™ efficiency as reported by the manufacturer [16].

Table I – Proteins present in the MW ranges with greater influence in distinguishing protein profiles of healthy or inflammation challenged individuals.

Molecular Weights	Proteins	
14 – 15 kDa	P09228	Cystatin-SA (Cystatin-2)
	P01037	Cystatin-SN (Cystain-SA-I)
	P01036	Cystatin-S (Cystatin-4)
	P01034	Cystatin-C (Cystatin-3)
	P07737	Profilin-1
	Q01469	Fatty acid-binding protein
	P06702	Protein S100-A9 (Calgranulin-B)
46 – 49 kDa	P52209	6-phosphogluconate dehydrogenase
	P80303	Nucleobindin-2
	P01871	Ig mu chain C region
	Q8N4F0	BPI fold-containing family B member 2
	P01009	Alpha-1-antitrypsin
	Q9UIV8	Serpin B13
	P30740	Leukocyte elastase inhibitor (LEI)
58 – 61 kDa	P14618	Pyruvate kinase PKM
	P04745	Alpha-amylase 1
	P07237	Protein disulfide-isomerase (PDI)
	Q9UBG3	Cornulin
	P52209	6-phosphogluconate dehydrogenase
28 – 29 kDa	P06870	Kallikrein-1
	P31947	14-3-3 protein sigma
	Q96DR5	BPI fold-containing family A member 2
42 – 46 kDa	P80303	Nucleobindin-2
	P01871	Ig mu chain C region
	Q8N4F0	BPI fold-containing family B member 2

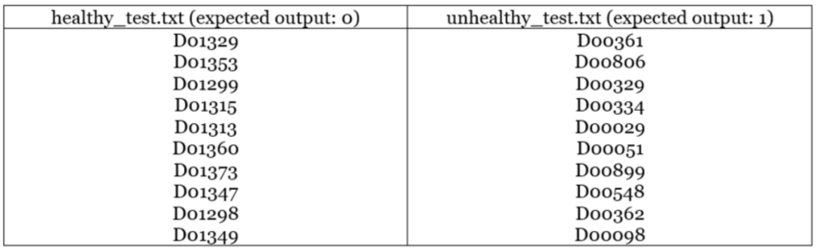
	P01009	Alpha-1-antitrypsin
	Q9UIV8	Serpin B13
	P30740	Leukocyte elastase inhibitor (LEI)
	P04083	Annexin A1
	P01876	Ig alpha-1 chain C region
	Q6P5S2	Protein LEG1 homolog
62 – 64 kDa	P02768	Serum albumin
	P15311	Ezrin (Cytovillin)
	P14618	Pyruvate kinase PKM
	P04745	Alpha-amylase 1
	P07237	Protein disulfide-isomerase (PDI)
76 – 78 kDa	P01833	Polymeric immunoglobulin receptor (PIgR)
	P22079	Lactoperoxidase (LPO)
	P02788	Lactotransferrin (Lactoferrin)
	Q08188	Protein-glutamine gamma-glutamyltransferase E
	P02768	Serum albumin
	P15311	Ezrin (Cytovillin)

e) Using SalivaPRINT Toolkit as a Classification Tool

Another functionality implemented in SalivaPRINT Toolkit is the possibility to run the previously created classifier on an independent set of individuals. This allows to verify if the classifier has correctly learned to differentiate the molecular weights related with the two populations (when the expected output is known), as well as providing a way to test if a particular individual is more similar to a population or another.

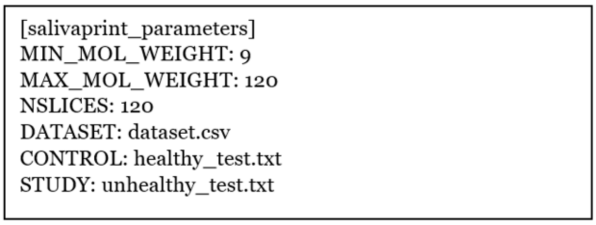
In this step, a new set of individuals from the original dataset, not used in the training of the algorithm, was used for testing the previously created classifier.

A list of 10 healthy individuals and 10 unhealthy individuals was created:



healthy_test.txt (expected output: 0)	unhealthy_test.txt (expected output: 1)
D01329	D00361
D01353	D00806
D01299	D00329
D01315	D00334
D01313	D00029
D01360	D00051
D01373	D00899
D01347	D00548
D01298	D00362
D01349	D00098

Next, the configurations file was adapted to create a testing dataset, note that it must provide the same configurations as the ones used to extract the data, which was used to create the classifier.

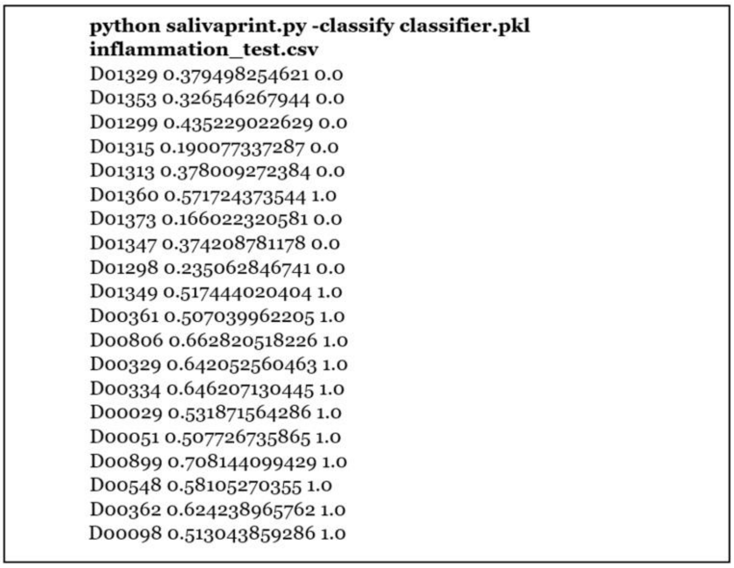


```
[salivaprint_parameters]
MIN_MOL_WEIGHT: 9
MAX_MOL_WEIGHT: 120
NSLICES: 120
DATASET: dataset.csv
CONTROL: healthy_test.txt
STUDY: unhealthy_test.txt
```

Then it is necessary to generate the test dataset, as follows:

python salivaprint.py –build inflammation_test.csv

And, finally, classify the test dataset:



```
python salivaprint.py -classify classifier.pkl
inflammation_test.csv
D01329 0.379498254621 0.0
D01353 0.326546267944 0.0
D01299 0.435229022629 0.0
D01315 0.190077337287 0.0
D01313 0.378009272384 0.0
D01360 0.571724373544 1.0
D01373 0.166022320581 0.0
D01347 0.374208781178 0.0
D01298 0.235062846741 0.0
D01349 0.517444020404 1.0
D00361 0.507039962205 1.0
D00806 0.662820518226 1.0
D00329 0.642052560463 1.0
D00334 0.646207130445 1.0
D00029 0.531871564286 1.0
D00051 0.507726735865 1.0
D00899 0.708144099429 1.0
D00548 0.58105270355 1.0
D00362 0.624238965762 1.0
D00098 0.513043859286 1.0
```

The values closer to zero are related with a tendency towards healthier states and values closer to 1 are related with inflammatory states. As shown, using this independent dataset, the classifier was able to correctly identify 18 out of 20 samples. Note that the misclassified examples occurred when the expected class was 0 and present values closer to 0.5 than most of the samples where the expected class was 1. This means that, despite misclassified, they are closer to the threshold line, which splits the two classes.

It is also important to keep in mind that the inflammatory process is not a binary classification problem in its origin; there are no absolute healthy or unhealthy individuals from which the classifier can learn from. Thus, it is expected that small changes in the classification threshold line (here considered to be 0.5) lead to adaptations on the sensibility and specificity of the algorithm.

4. Case study: Celiac patients a distinct phenotype within the inflammatory process

In this case study, a dataset of individuals diagnosed as celiac was used. These individuals share a chronic inflammation status and therefore it is expected that their salivary protein profile reflects the underlying functional dysregulation.

To test this hypothesis celiac patients were chosen according to time since diagnostic and grouped in 1-5 years or more than 5 years since diagnostics. Form each of these groups the individuals presenting the most dysregulated protein profiles were selected. This selection was based on complementary clinical data.

SalivaPRINT Toolkit commands were run following the steps above, considering, the two groups of celiac patients. The goal was to find which MW is important in the distinction of these groups.

The plot below represents the output of salivaprint.py –learn using this dataset. The molecular differences found between the two groups were minimal occurring on a small number of MW and with small values of influence (<0.5).

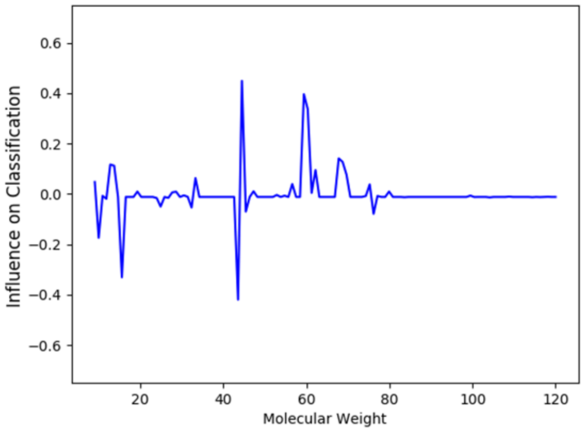


Figure 3 - Graphical representation of the molecular weight related with dysregulated proteins in the celiac groups.

The small differences found between the two groups with different diagnostic times are characterized by different profiles in the MW ranges presented in table II. The potential dysregulated proteins are also presented in each MW range according to the data from Rosa *et. al,* 2016 [15]. Proteins with molecular weights with a ± 8.56% interval were considered since this is the largest variation in Experion™ efficiency as reported by the manufacturer [16].

Table II – Proteins present in the MW ranges with greater influence in distinguishing protein profiles of healthy or inflammation challenged individuals.

Molecular Weights	Proteins	
44 – 45 kDa (1-5 years)	P01871	Ig mu chain C region
	Q8N4F0	BPI fold-containing family B member 2
	P01009	Alpha-1-antitrypsin
	Q9UIV8	Serpin B13
	P30740	Leukocyte elastase inhibitor (LEI)
	P04083	Annexin A1
59 – 60 kDa (1-5 years)	P14618	Pyruvate kinase PKM
	P04745	Alpha-amylase 1
	P07237	Protein disulfide-isomerase (PDI)
	Q9UBG3	Cornulin
15 – 16 kDa (+5 years)	P27482	Calmodulin-like protein 3
	P12273	Prolactin-inducible protein (PIP)
	P02810	Salivary acidic proline-rich phosphoprotein 1/2
	P09228	Cystatin-SA (Cystatin-2)
	P01037	Cystatin-SN (Cystain-SA-I)
	P01036	Cystatin-S (Cystatin-4)
	P01034	Cystatin-C (Cystatin-3)
	P07737	Profilin-1
	Q01469	Fatty acid-binding protein
43 – 44 kDa (+5 years)	P01009	Alpha-1-antitrypsin
	Q9UIV8	Serpin B13
	P30740	Leukocyte elastase inhibitor (LEI)
	P04083	Annexin A1

Conclusions

SalivaPRINT Toolkit is a command line tool that uses machine learning to analyze and learn from capillary electrophoresis data set experiments.

The analysis of individual protein profiles stratified by health condition has enabled the proposal of which MW ranges and respective proteins are altered in each group, leading to the inference of which molecular processes might be compromised.

In this tutorial, two scenarios were selected to demonstrate the use of the SalivaPRINT toolkit. First, a dataset composed of healthy individuals and individuals suffering from inflammatory conditions. Second, a group of individuals all with celiac disease, but stratified by date of diagnosis and treatment.

In both cases, the use of the proposed toolkit enabled the finding of protein MWs ranges, which characterizes the protein phenotype of these individuals.

The true power of using SalivaPRINT Toolkit as protein profile analysis tool, relies on the fact that the information of a large number of profiles is analyzed simultaneously and large amounts of data are accounted for, enabling the inference of which proteins may be involved with the underlying molecular process compromised in a particular condition. In this way, the identification of the protein profile patterns in saliva corresponding to different clinical situations, or the existence of different patterns within the same pathology may constitute a first approach to establish patient stratification according to the individual molecular profile (phenotype).

Acknowledgments

This work was made possible by the support of Centre for Interdisciplinary Research in Health through funds from the FCT (Portuguese national funding agency for science, research and technology) granted to Unit 4279 - UID/MULTI/4279/2016 and by the support of SalivaTec through + Centro - Programa Operacional Regional do Centro under the Quadro de Referência Estratégico Nacional (QREN) and through the Fundo Europeu de Desenvolvimento Regional (FEDER) CENTRO-07-CT62-FEDER-005004.

Igor Cruz were supported by Fundação para a Ciência e Tecnologia (Grant No. UID/MULTI/4279/2016.

References

[1] K.G. Becker, K.C. Barnes, T.J. Bright, S.A. Wang, The genetic association database., Nat. Genet. 36 (2004) 431–2. doi:10.1038/ng0504-431.

[2] A. Kozomara, S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data., Nucleic Acids Res. 42 (2014) D68-73. doi:10.1093/nar/gkt1181.

[3] J.P. Arrais, N. Rosa, J. Melo, E.D. Coelho, D. Amaral, M.J. Correia, M. Barros, J.L. Oliveira, OralCard: a bioinformatic tool for the study of oral proteome., Arch. Oral Biol. 58 (2013) 762–72. doi:10.1016/j.archoralbio.2012.12.012.

[4] M. Uhlén, L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C.A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J.M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Proteomics. Tissue-based map of the human proteome., Science. 347 (2015) 1260419. doi:10.1126/science.1260419.

[5] F.M.L. Amado, R.P. Ferreira, R. Vitorino, One decade of salivary proteomics: current approaches and outstanding challenges., Clin. Biochem. 46 (2013) 506–17. doi:10.1016/j.clinbiochem.2012.10.024.

[6] N. Rosa, M.J. Correia, J.P. Arrais, P. Lopes, J. Melo, J.L. Oliveira, M. Barros, From the salivary proteome to the OralOme: comprehensive molecular oral biology., Arch. Oral Biol. 57 (2012) 853–64. doi:10.1016/j.archoralbio.2011.12.010.

[7] D.T.W. Wong, Salivaomics., J. Am. Dent. Assoc. 143 (2012) 19S–24S. http://www.ncbi.nlm.nih.gov/pubmed/23034834.

[8] S. Yoon, J. Kim, J. Hum, H. Kim, S. Park, W. Kladwang, R. Das, HiTRACE: high-throughput robust analysis for capillary electrophoresis., Bioinformatics. 27 (2011) 1798–805. doi:10.1093/bioinformatics/btr277.

[9] S. Lee, H. Kim, S. Tian, T. Lee, S. Yoon, R. Das, Automated band annotation for RNA structure probing experiments with numerous capillary electrophoresis profiles., Bioinformatics. 31 (2015) 2808–15. doi:10.1093/bioinformatics/btv282.

[10] M. Jonsson, J. Carlson, Computer-supported interpretation of protein profiles after capillary electrophoresis., Clin. Chem. 48 (2002) 1084–93. http://www.ncbi.nlm.nih.gov/pubmed/12089178.

[11] G.A. Ceballos, J.L. Paredes, L.F. Hernández, Pattern recognition in capillary electrophoresis data using dynamic programming in the wavelet domain, Electrophoresis. 29 (2008) 2828–2840. doi:10.1002/elps.200700831.

[12] H. Laboratories, Electrophoresis System, (2000) 1–6. doi:10.1016/S0960-9822(00)80094-9.

[13] E. Delibato, A. Gattuso, A. Minucci, B. Auricchio, D. De Medici, L. Toti, M.

Castagnola, E. Capoluongo, M.V. Gianfranceschi, PCR experion automated electrophoresis system to detect Listeria monocytogenes in foods., J. Sep. Sci. 32 (2009) 3817–21. doi:10.1002/jssc.200900166.

[14] J.H. Kim, Y.-W. Kim, I.-W. Kim, D.C. Park, Y.W. Kim, K.-H. Lee, C.K. Jang, W.S. Ahn, Identification of candidate biomarkers using the Experion^{TM} automated electrophoresis system in serum samples from ovarian cancer patients., Int. J. Oncol. 42 (2013) 1257–62. doi:10.3892/ijo.2013.1803.

[15] N. Rosa, J. Marques, E. Esteves, M. Fernandes, V.M. Mendes, Â. Afonso, S. Dias, J.P. Pereira, B. Manadas, M.J. Correia, M. Barros, Protein Quality Assessment on Saliva Samples for Biobanking Purposes., Biopreserv. Biobank. 14 (2016) 289–97. doi:10.1089/bio.2015.0054.

[16] K. Zhu, M. Nguyen, W. Strong, C. Whitman-guliaev, B. Laboratories, P. Samples, Performance Comparison of the Experion TM Automated Electrophoresis System and SDS-PAGE for Protein Analysis, Methods. (n.d.) 0–5.