SalivaPrint as a tool for individual's health status monitorization

Eduardo Esteves
Ana Cristina Esteves
Liliana Bernardino
Marlene Barros
Nuno Rosa
Questions
The SalivaPrint concept

Is it possible to access the panorama of all saliva proteins, using a single approach?

Does the salivary protein profile reflect the individual’s health changes?
Motivation
The SalivaPrint concept

Changes in protein concentrations in the body may indicate the presence of pathologies.

Identifying those protein alterations, assists in early diagnosis and disease monitoring.
Type 2 Diabetes Mellitus as disease model for SalivaPrint

- Type 2 Diabetes Mellitus (T2DM) is a chronic disease, affecting about 150 million people world-wide.

- It is characterized by a complex metabolic dysregulation with a very varied clinical picture.

- It has been previously shown that salivary proteomes present alterations in diabetic patients.

- From literature, it is known that glycosylation can affect the electrophoretic mobility of proteins.

- The application of SalivaPrint approach using T2DM as disease model has the clinical support to ensure the experimental results.
1. Establish a saliva protein profile by capillary electrophoresis - the **SalivaPrint**

2. Develop a **Machine Learning algorithm** to classify salivary total protein profiles in healthy or unhealthy individuals.

   i. Use Type 2 Diabetes Mellitus as model disease.
Methodology

• **Specific Objectives:**
  
  • Saliva Collection and Sample Characterization;
  
  • Individual protein profile determination using Capillary Electrophoresis;
  
  • Supervision Machine Learning (ML) algorithm development on T2DM.
Results

1. SalivaPrint Determination
2. SalivaPrint Algorithm Development


## SalivaPrint Determination

### Sample Dataset

*Table 1. Examples of data set entries. The first column contains the identifiers, the last column contains the individual’s health status and the columns between them represent protein molecular weights*

<table>
<thead>
<tr>
<th>sample</th>
<th>10</th>
<th>10.1</th>
<th>10.2</th>
<th>10.3</th>
<th>...</th>
<th>121</th>
<th>health status</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1127</td>
<td>30.00929794</td>
<td>25.47068097</td>
<td>21.09144563</td>
<td>16.56765483</td>
<td>...</td>
<td>2.8623154385</td>
<td>healthy</td>
</tr>
<tr>
<td>d1132</td>
<td>1.189318499</td>
<td>0.311820769</td>
<td>-0.650945071</td>
<td>-1.482789145</td>
<td>...</td>
<td>-6.6426125758</td>
<td>unhealthy</td>
</tr>
<tr>
<td>d1143</td>
<td>-68.75426899</td>
<td>-66.33690751</td>
<td>-62.77299866</td>
<td>-58.76343131</td>
<td>...</td>
<td>0.7623547533</td>
<td>unhealthy</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
SalivaPrint Determination

Total Protein Profiles of healthy and T2DM individuals

Figure 1. **Representation of the distribution of fluorescence values for each molecular weight (kDa).** The average values of the molecular weight’s fluorescence for the healthy individuals (n=118) are represented by the blue line and for the unhealthy individuals (n=57) represented by the dashed orange line. The shaded areas around the lines represent the standard deviation of the fluorescence values.
Figure 2. Graphical representation of the pairwise relationship. Diagonal axis represent the distribution of individuals for each peak.

- Peaks “26-29”, “42-45”, “46-49,” and “58-61” are present mostly in unhealthy individuals, making them good candidates for use in the process of classification by the algorithm.

Healthy individuals in peaks “34-37”, “62-65”, “70-73” and “74-77” are concentrated in lower height values, while the unhealthy individuals are better distributed.

- Peak “42-45” has a very similar distribution for both profiles, making it not an exciting feature to be used to classify the health of the base entries.
Figure 3 - Graphical representation of the hierarchical classification using the average distance between groups. It’s visible the grouping of unhealthy individuals on the left side of the graph, represented by the black rectangle,
Conclusions

• Protein profile observations comparing the diabetic and healthy individuals indicates that the saliva protein signature is related with individual's health condition.

• Capillary electrophoresis data can be used to perform a cluster analysis (hierarchical), and define groups based on systemic health status.

• We identified ranges where protein molecular weights are different among diabetic individuals and healthy individuals.

• We implemented an artificial intelligence algorithm able to distinguish healthy and diabetic individuals, using their saliva protein profiles.
Future work

• Develop stronger Supervised Machine Learning algorithms able to classify an individual – using its protein saliva profile – as healthy or non-healthy;

• Apply the same strategy to COVID-19.
Thank You!

SalivaPrint as a tool for individual's health status monitorization

3rd November 2020
Ciência 2020