The salivary proteome reflects some traits of dietary habits in diabetic and non-diabetic older adults

Christophe Chambon, Eric Neyraud, Thierry Sayd, Pauline Bros, Romane Di Biagio, Frank Hyvrier, Catherine Féart, Perrine André, Fernando Rodriguez-Artalejo, Esther Lopez-Garcia, et al.

The salivary proteome reflects some traits of dietary habits in diabetic and non-diabetic older adults

Christophe Chambon[a], Eric Neyraud[b], Thierry Sayd[a], Pauline Bros[a], Romane Di Biagio[b], Frank Hyvrier[b], Catherine Féart[c], Perrine André[c], Fernando Rodriguez-Artalejo[d,e,f], Esther Lopez-Garcia[d,e,f], Esther Garcia Garcia-Esquinas[d,e], David Gomez Cabrero Lopez[g], Gordon Proctor[g], Martine Morzel [b]*

[a] : INRAE, Plateforme d'Exploration du Métabolisme Composante Protéome PFEMcp, St-Genès-Champanelle, France

[b] : Centre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRAE, Université de Bourgogne Franche-Comté, Dijon, France

[c] : Université de Bordeaux, Inserm, BPH, team LEHA, UMR 1219, Bordeaux, France

[d] : Department of Preventive Medicine and Public Health, Universidad Autónoma de Madrid and CIBERESP, Madrid, Spain

[e]: Cardiovascular and Nutritional Epidemiology Group. IdiPAZ (La Paz University Hospital – Universidad Autónoma de Madrid), Madrid, Spain

[f] : IMDEA-Food Institute, Madrid, Spain

[g] : Centre for Host Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College, London, UK

ORCID

Christophe Chambon : 0000-0003-0357-9807

Eric Neyraud : 0000-0001-7060-1336

Thierry Sayd : 0000-0003-1370-8959

Frank Hyvrier : 0000-0003-3633-9885

Catherine Féart : 0000-0002-7959-1610

Fernando Rodriguez-Artalejo : 0000-0001-9317-5755

Esther Lopez-Garcia : 0000-0001-6202-4970

Esther Garcia Garcia-Esquinas : 0000-0002-8688-5174

David Gomez Cabrero Lopez : 0000-0003-4186-3788

Gordon Proctor : 0000-0002-5684-841X

Martine Morzel : 0000-0002-3589-3641

* **corresponding author**. Phone number : +33 223 48 56 17. E-mail address : martine.morzel@inrae.fr

Abstract

Purpose : Objective markers of usual diet are of interest as alternative or validating tools in nutritional epidemiology research. The main purpose of the work was to assess whether saliva protein composition can reflect dietary habits in older adults, and how type 2 diabetes impacted on the saliva-diet correlates.

Methods : 214 participants were selected from two European cohorts of community-dwelling older adults (3C-Bordeaux and Seniors-ENRICA-2), using a case-control design nested in each cohort. Cases were individuals with type 2 diabetes. Dietary information was obtained using the Mediterranean Diet Adherence Screener (MEDAS). Saliva was successfully obtained from 211 subjects, and its proteome analyzed by liquid chromatography-tandem mass spectrometry.

Results : The relative abundance of 246 saliva proteins was obtained across all participants. The salivary proteome differed depending on the intake level of some food groups (especially vegetables, fruits, sweet snacks and red meat), in a diabetic status- and cohort-specific manner. Gene Set Enrichment Analysis suggested that some biological processes were consistently affected by diet across cohorts, for example enhanced platelet degranulation in high consumers of sweet snacks. Minimal models were then fitted to predict dietary variables by sociodemographic, clinical and salivary proteome variables. For the food group « sweet snacks », selected salivary proteins contributed to the predictive model and improved its performance in the Seniors-ENRICA-2 cohort and when both cohorts were combined.

Conclusion : Saliva proteome composition of elderly individuals can reflect some aspects of dietary patterns.

Keywords : salivary biomarkers, proteomics, usual diet, ageing, diabetes, Gene Set Enrichment Analysis

Declarations

Funding : This work was initiated by the Joint Programming Initiative A Healthy Diet for a Healthy Life (HDHL). The SALAMANDER project is funded by Agence Nationale de la Recherche, France (grant ANR-16-HDHL-0005-01), Biotechnology & Biological Sciences Research Council, UK (grant BB/R000891/1) and Ministry of Economy and Competitiveness, Spain (APCIN grant 2016-145)

Conflicts of interests : nothing to disclose

Ethics approval : Study protocols of the 3C and Seniors-ENRICA-2 cohorts were approved respectively by the Commitee for Protection of Persons (CPP) participating in Biomedical Research of the Kremlin-Bicêtre University Hospital (Paris) and the Ethics Research Commitee of the « La Paz » University Hospital (Madrid).

Consent to participate / Consent to publish : Participants provided written informed consent to participate and for the results of the study to be published.

Availability of data and material : Data described in the manuscript will be made available upon request addressed to the corresponding author.

Code Availibility : The R scripts will be made available upon request addressed to the corresponding author.

Author's contributions : EN, CF, FRA, ELG, GP & MM conceived and designed the study. CC, TS & PB performed saliva analysis. RD, FH & DGCL performed statistical analyses. PA, EGGE, CF, FRA & ELG provided epidemiological data and sampled saliva. CC & MM drafted the manuscript. All authors revised the manuscript and approved its final version.

Introduction

There is vast evidence linking intakes of nutrients, specific foods, food groups or even dietary patterns with health and well-being outcomes. Some studies have focused on the older population, targeting health issues more prevalent in this age group. For example, a link has been established between usual intake of protein and frailty [1] or between coffee consumption and risk of falling [2]. Higher adherence to the so-called Mediterranean diet (MeDi) has also been associated with decreased cognitive decline [3] or decreased risk of frailty [4,5]. Concerning another pathology with higher prevalence among the elderly population, higher adherence to MeDi diet has been associated with lower risk of type 2 diabetes mellitus (T2DM) [6,7] and not only through its impact on body weight [8]. Recording one's diet can therefore be a useful tool for the clinician as a starting point to implement dietary guidelines and/or to follow adherence to dietary advice.

One challenge in nutritional epidemiology research is to capture real dietary intakes. Assessment usually relies on questionnaires (such as 24-hour recalls or Food Frequency Questionnaire) which may be prone to memory or social desirability bias. Therefore, objective markers of dietary intake are required as alternative or validating tools to increase the reliability and accuracy of diet information [9]. With this objective in mind, markers of dietary intake have been sought for mainly in serum and in urine. For example, metabolome markers of usual consumption of citrus fruit or fish were identified in serum of adult participants [10]. Urinary metabolome markers of usual nut intake were also identified and correlated to cognitive decline in older adults [11]. Recently, plasma metabolome markers of adherence to the Mediterranean diet have been associated with cardiovascular disease risk [12].

Saliva is a biological fluid which presents some advantages over blood or urine, particularly regarding its sampling which minimizes pain, privacy or safety issues. Saliva sampling also presents some limitations, for example in case of minimal saliva flow. Young children and the oldest or frail elderly subjects are thus two groups for which it may be difficult to obtain saliva samples [13]. However, saliva was sucessfully sampled and its protein or peptide composition analyzed on 3- and 6-month-

old infants [14], on premature babies [15] or on old persons with a mean age of 82 years [16]. Saliva composition was successfully associated with usual intake of carbohydrates in adults [17], with diet transition in infants [14,18] or with dietary patterns in children with or without eating difficulties [19]. In addition, saliva (together with blood) proved more resilient than urine to recent dietary intake when focusing on metabolome composition [20]. For all these reasons, the primary objective of this work was to evaluate whether saliva is a suitable source of objective protein markers of usual dietary intake in older adults. We evaluated this with participants from two different cohorts, in France and Spain, which provided information on the cross-cultural and/or geographical sensitivity of the results. Finally, participants were part of a wider project on salivary biomarkers of Mediterranean diet and type 2 diabetes mellitus [21]. The design of participants' selection therefore also enabled to address a secondary objective, namely to assess the impact of this pathology on the association between saliva composition and dietary habits.

Methods

Population-based cohorts

Participants originated from two population-based cohorts on ageing, the Bordeaux sample of the Three-City Study (3C) in France [22] and the Seniors-ENRICA-2 cohort in Spain [23]. Study protocols of the 3C and Seniors-ENRICA-2 cohorts were approved respectively by the Commitee for Protection of Persons (CPP) participating in Biomedical Research of the Kremlin-Bicêtre University Hospital (Paris) and the Ethics Research Commitee of the « La Paz » University Hospital (Madrid).

Participants

The 214 Individuals included in the study were selected using a case-control design nested in each cohort. Cases were individuals affected by T2DM based on self-reported physician's diagnosis and/or being on antidiabetic treatment (oral medication or insulin) at the time of data collection. Controls were selected concurrently and were free of diabetes at the time of data collection. For the purpose of the present study, we used the following information from each participant: cohort, sex, age

(years), body mass index (BMI; kg/m^2), type 2 diabetic status, smoker status (never, former, current), saliva flow (g/min) and food consumption.

Dietary surveys and assessment of adherence to the Mediterranean diet

Food consumption data were collected using a FFQ in the 3C cohort [24] and a validated electronic diet history in the Seniors-ENRICA-2 cohort [25]. Adherence to the Mediterranean dietary pattern was assessed by calculating a MEDAS score [26] modified by omitting the question on *sofrito*, since this cooking technique is specific of the Spanish population. The MEDAS score could therefore range from 0 to 13, with higher values indicating higher adherence to the Mediterranean diet. Differences between cohorts or diabetic status was tested by a Wilcoxon test. In addition, to study the link between saliva proteome and diet, we restricted the analysis to the 11 MEDAS items describing the consumption frequency of food groups, namely olive oil, vegetables, fruits, red meat, butter/margarine/cream, sweet drinks, wine, legumes, fish/shellfish, sweet snacks (confectionary, biscuits and commercial pastries) and nuts. For each food group, the participants were classified as high or low consumers based on the cut-off points defined for MEDAS score calculation (Online Resource 1).

Saliva sampling

Sampling of unstimulated saliva was conducted at the participants' home. Sampling was proposed early in the morning after overnight fasting. Drinking water was permitted up to 5 minutes before saliva collection. Participants were instructed to sit comfortably and to tilt their head slightly downwards. At their own rythmn, they spat the saliva pooling on the floor of the mouth into 40mL polypropylene tubes. Sampling was performed for 10 minutes. In case a participant wished to stop before the end of the 10 minutes, the time was recorded in order to be able to calculate the saliva flow (expressed in g/min). Saliva samples were immediately placed on ice, transported to the laboratory and placed at -80°C as soon as possible (never after 4 hours on ice). At the end of the collection wave, samples were shipped to the analytical facilities in dry-ice.

Saliva proteome analyses

Sample preparation

Saliva was thawed at 4°C and vortexed. One mL of saliva (or the total volume of saliva when it was lower than 1mL) was centrifuged at 14000 g for 20 minutes. The supernatant was used for proteome analyses. Protein concentration was measured by an infrared spectroscopy-based method using a Direct Detect® spectrometer (Merck). Samples were diluted in water in order to adjust all samples to the same protein concentration, then mixed with 1 volume of Laemmli denaturing buffer and heated at 90°C for 5 minutes. Sample volumes corresponding to 3.5 µg of protein were loaded onto SDS-PAGE gels containing 12% and 5% acrylamide in the resolving and stacking gels, respectively. Electrophoresis was performed using a Mini-Protean II unit (BioRad, Marnes-La-Coquette, France) at 100 V until the dye front entered the resolving gel. Gels were stained for one hour in R-250 Coomassie. Bands were manually excised, reduced in 10 mM dithiotreitol in 50 mM ammonium bicarbonate, and alkylated in 55 mM iodoacetamide in 50 mM ammonium bicarbonate. Destaining of the excised bands was obtained by sucessive rinses in 25 mM ammonium bicarbonate / acetonitrile (1 :1 v/v). Gel pieces were then dried by incubation in 100% acetonitrile for 10 min followed by vacuum-drying in a SpeedVac. Finally, gel pieces were incubated overnight at 37°C with 30 µL of a trypsin solution (V5111, Promega) at 10 ng/µL in 25 mM ammonium bicarbonate. Peptide extraction was performed by addition of 40µL of acetonitrile 100%, 0.5% formic acid and sonication for 15 min. The trypsin digests were vacuum-dried in a SpeedVac and stored at 20 °C in a solution of 0.05% trifluoroacetic acid before Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) analysis

Mass Spectrometry analyses

Five µL of the protein digests were injected into a nanoHPLC (Ultimate 3000, ThermoFisher Scientific). The peptide mixture was first concentrated and desalted on a microcolumn (Acclaim, 300 µm, 5mm) equilibrated with trifluoroacetic acid (TFA) 0.05% in water. After 6 min, the microlumn was switched on-line to an analytical C18 nanocolumn (Acclaim,75 µm, 25 cm, Pepmap) equilibrated with 94.9 % H_20, 5% dimethyl sulfoxide (DMSO), 0.1% formic acid (FA) and peptides were separated

at 35°C according to their hydrophobicity with a 4 to 25% linear gradient of acetonitrile (94.9 % ACN, 5 % DMSO, 0.1% FA) at a flow rate of 300 nL/min for 50 min. Peptides were electro-eluted with an ESI nanosource (1.6 kV) in the mass spectrometer (Orbitrap Velos, ThermoFisher scientific). The Orbitrap was used in top 15 data dependent mode, with gas phase fractionation (GPF1 400-480 m/z; GPF2 480-560 m/z; GPF3 560-660 m/z; GPF4 660-850 m/z, GPF5 850-1401 m/z).

Samples were analyzed in 6 different series (different days). To monitor and normalize LC-MS performance over time, a Quality Control (QC) was injected every 10 samples. This QC was prepared by diluting 20µL of a commercial mixture of 15 synthetic peptides (Thermo Scientific™ Pierce™ Peptide Retention Time Calibration Mixture) into 180µL of a trypsin digest of a pool of 20 salivas. Five µL of the QC was injected and analyzed similarly to samples, except that gas phase fractionation was not performed.

Raw files were imported into ProgenesisQI Proteomics (Nonlinear dynamics) and the label-free quantification workflow was followed (see www.nonlinear.com). This includes peak alignment, establishment of a single ion map, normalization of intensities using data from 2 to 5x charged ions and XIC type quantification of all ions detected. Peptide identification were performed using Mascot interrogating the database homosapiens UniProt 201804 (71,600 sequences). Peptide mass tolerance was set to 10 ppm and fragment mass tolerance was set to 0.5 Da. Two miscleavages were authorized, and methionine oxidation, carbamidomethylation of cysteine and deamidated asparagine or acid aspartic were set as variable modifications. Protein identification was then validated when at least two unique peptides from one protein showed significant identification Mascot scores with False Discovery Rate (FDR < 1%). For protein quantification, the sum of all unique normalized peptide ion abundances for a specific protein was calculated for each sample. The final dataset consisted in 246 proteins quantified in 211 samples.

Proteome data handling

Protein abundance as obtained above was corrected for any series effect using QC data. For each synthetic peptide, the average abundance was calculated per series and for all runs combined. This

enabled calculting a correcting coefficient applicable to each series for normalization. The mean correcting coefficient for the 15 synthetic peptides was calculated per series, and it was applied to the protein abundance within that series. Missing values (0.3% of overall data) were handled following two distinct imputation methods as described by Wei et al. 2018 [27]. Eighty-two percent of the missing values corresponded to proteins which always showed low abundance values. These missing values were regarded as Missing Not At Random and thus handled using the quantile regression imputation of left-censored method. Remaining missing values were regarded as Missing At Random thus handled using the Random Forest imputation method.

Statistics

First, for each food group, we tested differences in the salivary proteome between high and low consumers Student t-tests corrected by the Benjamini-Hocberg method to control the False Discovery rate (FDR set at 5%). This was performed for the entire sample or when participants were stratified by diabetic status or by cohort.

Second, the Gene Set Enrichment Analysis (GSEA) method [28] was used to highlight biological processes enriched depending on the level of consumption of a food group. Given the case-control design of the study, we performed the analyses separately for diabetic and non-diabetic subjects. For each of the 246 proteins of the dataset, ontology terms and related annotations were retrieved from quickGO and the GO database. GO terms corresponding to at least two proteins (1850 in total) were selected. For each condition (one condition = one diabetic status, one food group), proteins were ranked in descending order of their p-values after attributing a negative sign to p-values when proteins were under-expressed in “high consumers”. This way, a GO term is regarded as of interest if its protein members are mostly in the top (over-expressed) or bottom (under-expressed) region of the list. GO terms were tested by unweighted GSEA estimation algorithm with at least a million permutations. With multifactorial dependencies occurring in the GO terms classification, (correlated and/or co-occurring annotations, co-occurring proteins …), controlling false positive occurrence is challenging and usual methods may be over-restrictive. Post-hoc correction was therefore conducted

as follows: the community structure of the GO network (terms as nodes, relationships as edges) was determined using the Girvan-Newman algorithm [29]. This revealed 223 separated communities with a modularity of 0.85. GO terms within a community were considered as strongly dependent from each other whereas communities were considered as independent from each other. P-values of the 223 communities were computed as the mean of p-values of its GO terms members, and the Benjamini-Hochberg correction was performed (calculation of a q-value). This generated a correcting factor for each community, which was then applied to p-values of all GO terms within this community to calculate individual q-values. GO terms were considered significant for q-values <0.01 (FDR 1%).

Finally, we fitted minimal models predicting dietary variables by sociodemographic, clinical and salivary proteome variables. For that purpose, the workflow was as follows for each food group. First we reduced the number of protein variables by three successive selection steps : 1- a Kernel Partial Least Square Regression model was computed and proteins with high Variable Importance in Projection (VIP) values were selected (cutoff=1). 2- For these proteins, the difference in abundance between high and low consumers of the food group was tested by a Student t-test corrected by the Benjamini-Hochberg method to control the False Discovery rate (FDR). Variables with a low FDR were selected (cutoff=0.05). 3- Correlation among these proteins were calculated. Candidates sharing at least one high correlation (cutoff=0.7) were ordered by Ascendant Hierarchical Clustering, with adequate number of clusters determined by bootstrapping. The final set of proteins was made of all uncorrelated variables plus one variable per cluster (the one with the highest contribution). Second, we fitted the minimal model by logistic model learning using selected proteins and sociodemographic and clinical confounding factors (diabetic status, cohort, sex, age, BMI, smoker status and saliva flow). Subjects were split into a learning set (75%) and a validation set (25%). From the learning set, a logistic model was adjusted. Two stepwise procedures (Akaike Information Criterion, Bayesian Information Criterion) were performed separately : the AUC (Area under the Curve) of the ROC (receiver operating characteristic) curves were computed from the validation set and the model with

the highest AUC was selected. This process was repeated through 1000 simulations of learning/validation sets, and the final model retained the variables selected in at least half of the simulations. For the food groups where proteome variables were retained, we evaluated the performance of the models by calculating the AUC of the ROC curve. More specifically, three models were tested : Model 1 was the optimal model obtained as described above, Model 2 was Model 1 into which we forced some basic descriptive variables (cohort, age, sex, diabetic status), and Model 3 was Model 1 from which protein variables were removed. The comparison of Model 1 and Model 3 allows documenting how much value the saliva proteome variables add to the model.

Data management and statistical analyses were performed using the Rgui open-source software (https://cran.r-project.org).

Results

Participants' characteristics

Among 570 participants from wave 8 (2017-2018, 18y after baseline) of the 3C-Bordeaux cohort, 65 were diabetics. Among them, 37 participants agreed to participate and a random sample of 71 participants were selected as controls. Among 3273 participants from baseline in the Seniors-ENRICA-2 cohort, 669 were diabetics. A random sample of 53 diabetics and 53 matched controls were selected. Out of those 214 participants, 3 were excluded because of unsufficient saliva production (Online Resource 2). The 211 remaining participant's characteristics are presented in Table 1. The French sample was older on average (87.3 vs 71.0 y) and comprised more women (61.3 vs 36.2%) than the Spanish one. As expected, the diabetic participants exhibited higher BMI on average than controls in both cohorts.

Saliva amounts collected and salivary flows

The amounts of saliva collected varied from 0.09 to 8.68 g. As reported in Table 1, the average at-rest saliva flow varied from 0.18 to 0.26 g/min depending on the cohort and diabetic status. The difference between the two cohorts tested by a Wilcoxon test was significant ($p<0.001$), with higher

flow among Seniors-ENRICA-2 participants. The difference between diabetic and non-diabetic subjects was not significant either when combining the two cohorts or within each cohort.

MEDAS scores and dietary intakes

The mean MEDAS scores and the proportions of high vs low consumers of 11 food groups among all participants or stratified by diabetic status or cohort are presented in Table 2. Adherence to the Mediterranean dietary pattern was slightly (but not significantly) lower in diabetic participants compared to controls, while it was significantly ($p<0.05$) lower in 3C-Bordeaux participants compared to Seniors-ENRICA-2 participants. The dietary patterns were overall comparable between diabetic and non-diabetic individuals : the only two food groups which were clearly different according to the diabetic status were « nuts » and « sweets snacks » (confectionary, biscuits and commercial pastries), with a higher proportion of low consumers among diabetic subjects. In contrast, there were more differences between the two cohorts, with five food groups clearly different between Seniors-ENRICA-2 and 3C-Bordeaux. In the Spanish cohort there were more high consumers of olive oil, while in the French cohort there were more high consumers of vegetables, nuts, wine and butter.

Proteins differentially expressed between high and low consumers of 11 food groups

Table 3 provides the number of proteins significantly different (FDR 5%) between high and low consumer of the 11 food groups recorded. When considering all participants, there were differentially expressed proteins for three food groups : vegetables, butter and sweet snacks. The list of proteins significantly different, details of mass spectrometry identifications and abundance in high vs low consumers are available in Online resource 3. The intake level of vegetables, red meat and butter altered more the saliva of non-diabetic subjects (20, 58 and 5 differential proteins, respectively) than that of diabetic subjects (4, 0 and 0 differential proteins, respectively). In contrast, the intake level of sweet snacks had a major impact on saliva proteome (65 differential proteins) but only for diabetic subjects. More generally, the link between saliva proteome and diet was more pronounced among non-diabetic controls, with 85 proteins in total linked to 4 food groups (vegetables, legumes, red meat, butter), than among diabetic participants with 69 proteins linked to

2 food groups (vegetables, sweet snacks). The link between saliva proteome and diet was also more expressed among Seniors-ENRICA-2 participants with 77 proteins linked to two food groups (fruits, sweet snacks), than among 3C-Bordeaux participants with 2 proteins linked to the intake level of olive oil.

Gene Set Enrichment Analysis (GSEA)

Based on protein expression levels in the entire proteomic dataset, GSEA identified the biological processes that were significantly enriched or depleted depending on the participants' intake level of a given food group. This was performed for diabetic and control participants separately and for the food group « sweet snacks » because of its major and robust impact on the salivary proteome (see Table 3). Results for all participants are presented in Figure 1 and results of the analyses performed for each cohort separately are available in Online resource 4.

For the two cohorts combined, some biological processes were in common and similarly affected for all participants. Thus, *regulation of cellular macromolecule biosynthetic process* was repressed while *platelet degranulation*, *leukocyte migration involved in inflammatory response and regulation of immune system* were enhanced in high consumers of sweet snacks, regardless of their diabetic status.

In addition, the impact of sweet snacks intake level was particularly evident among diabetic participants (Fig. 1). In those participants, three GO categories related to apoptosis were repressed in high consumers of sweet snacks, as well as the category regulation of peptidase activity. In contrast, all the other GO terms were enriched in diabetic high consumers of sweet snacks. These terms covered a very large panel of biological functions such as hemostasis, protein and peptide expression, modification or secretion, immunity, cell-cell communication or metabolism. Of special interest is a group of GO terms related to carbohydrate metabolism, (black box in Fig. 1). In controls, high consumption of sweet snacks also modified specifically some biological processes but to a lesser extent.

Analyses performed for each cohort separately confirmed that the impact of sweet snacks intake level was higher for diabetic participants than for controls in both cohorts. A vast majority of the biological processes significant for the two cohorts combined were also identified and significant in at least one of the two cohorts. Results were mostly consistent between the two cohorts combined and one of the cohorts (lines highlighted in blue in Online resource 4, e.g. *platelet degranulation*). Five biological processes (highlighted in yellow in Online resource 4, e.g. *regulation of phosphorylation*) were significant for diabetics and controls in the Seniors-ENRICA-2 cohort, while they were significant only for diabetics in the 3C-Bordeaux cohort and in the joint analysis. Finally, 14 biological processes were consistent across cohorts for diabetics subjects: *regulation of apoptotic process, platelet degranulation, regulation of immune response, regulation of gene expression, regulation of transport, regulation of cellular component organization*, *regulation of protein/peptide secretion*, *regulation of cell morphogenesis involved in differentiation*, *interleukin-12-mediated signaling pathway*, *regulation of substrate adhesion-dependent cell spreading*, *Arp2/3 complex-mediated actin nucleation*, *actin polymerization-dependent cell motility* and *microtubule-based movement*. The latter four processes, related to cell spreading and motility, are likely linked to the oral epithelium repair potential.

Salivary markers of specific food groups' intake levels

Minimal models were sought for all food groups. Tables 4 and 5 present an overview of the variables retained in such models for all participants combined or cohort by cohort, respectively.

Considering both cohorts combined, apart from fruit for which no minimal model was retained, the descriptive characteristics of the subjects could predict the level of consumption (defined as meeting the MeDi requirements or not) for all food groups. The descriptor « cohort » was most frequently retained (4 occurrences for vegetables, olive oil, wine and butter) followed by sex (3 occurrences for wine, red meat and sweet snacks), diabetic status (nuts and sweet snacks) and smoker status (legumes and sweet drinks). Of special interest are the three food groups for which abundance of salivary proteins are predictors of intake : vegetables (5 proteins : P22079 lactoperoxidase, O00391

sulfhydryl oxidase 1, P10909 clusterin, P37802 transgelin-2 and Q8TAX7 mucin-7), butter (4 proteins : Q96DR5 BPI fold-containing family A member 2, P22079 lactoperoxidase, P07476 involucrin and P23280 carbonic anhydrase 6) and sweet snacks (2 proteins : E7EQB2 lactotransferrin, P02749 beta-2-glycoprotein 1). For the three food groups vegetables, butter and sweet snacks, the distribution of AUCs for 10000 simulations and the corresponding means and 95% confidence intervals are presented in Fig. 2. These show that compared to the minimal models (top panels), forcing into the model additional subjects' basic descriptors (middle panels) did not improve the model performance. Moreover, removing the protein descriptors from the minimal models (bottom panel) clearly reduced the model performance with an AUC mean shift of approximately -0.05 for vegetables and butter and -0.1 for sweet snacks. In other words, we confirmed that the salivary proteome added value to the prediction models.

Results obtained cohort by cohort (Table 5) did not directly confirm the protein markers identified when the two cohorts were combined. In agreement with the very low number of proteins significantly different between high and low consumers of the different food groups among 3C-Bordeaux participants (Table 3), no minimal models retained salivary proteins for French participants. Among Seniors-ENRICA-2 participants, salivary proteins were predictors of intake of fruits, red meat and sweet snacks.

Discussion

In this study, the salivary proteome differed depending on the intake level of some food groups (especially vegetables, fruits, sweet snacks and red meat), in a diabetic status- and cohort-specific manner. In addition, some saliva proteins were predictive of the intake level of sweet snacks after adjusting for several sociodemographic and clinical confounding factors.

The first challenge of the study was the collection of saliva. Only 3 participants out of 214 were excluded due to unsufficient saliva production, and only one of the 211 donors requested to stop before the end of the 10 min collection time. The average saliva flows were below the 0.3 g-0.4/min generally reported [30,31], in accordance with the well-known reducing impact of ageing on both at-

352 rest and stimulated saliva flow [32]. Age is likely the main factor explaining the statistical difference
353 between the cohorts, with lower average flow in French participants who are on average 16 years
354 older than Spanish participants. The sex-ratio and the cases-controls proportions also differ between
355 the two cohorts. However, several studies reported that gender had no significant impact on saliva
356 flows of either healthy [33] or diabetic elderly subjects [34]. Furthermore, the diabetic status had no
357 significant impact on saliva flow in our study, in each cohort or combining the two. To conclude on
358 saliva collection, our study highlighted that it was feasible, well-tolerated and although the volumes
359 were sometimes limited especially in the oldest elderlies, they remained compatible with the needs
360 of proteome analytical methods.
361 The assessment of dietary intakes is also a challenge for any population. Diet history and Food
362 Frequency Questionnaire are often used to capture usual nutrient intake or dietary habits with a high
363 level of details. Here, we used two different tools which are standard and validated instruments for
364 collecting food consumption data. In a previous study where these same cohorts and respective diet
365 recording tools were used, consistent results for the association between fruit and vegetable intake
366 and functional outcomes in older adults were obtained [35]. Based on the data obtained, we used
367 the MEDAS score to represent adherence to the MeDi diet [26]. We confirmed previous
368 observational studies reporting no difference in adherence to MeDi between persons with diagnosed
369 diabetes and controls [36]. The MEDAS questionnaire also enabled to classify participants into high
370 and low consumers for 11 food groups. There was a limited impact of the diabetic status on the
371 intake levels, at the exception of nuts and sweet snacks. This latter finding is in accordance with
372 previous results observed within the Seniors-ENRICA cohort, showing that the diet of diabetics differs
373 mainly by a greater avoidance of sweet products while other food groups are little affected [36].
374 Other studies are also in general agreement with this idea. For example, a cross-cultural study
375 investigated the diet of diabetic and non-diabetic elderly men in Finland, the Netherlands and Italy
376 and highlighted that the only food group that was significant different in all three countries was the
377 consumption of added sugar [37]. In another study [38], the food group *sweets* differed largely
378 between individuals with or without diabetes both in a pan-European cohort (the EPIC study) and a

379 USA-based cohort (the MEC study). The two latter articles also evidenced that differences between
380 countries exceeded those between diabetics and non-diabetics, in line with our results on the
381 MEDAS score or the individual food groups. Some dietary differences that we observed between
382 cohorts corresponded to expected cultural specificities (more high consumers of wine in the
383 Bordeaux area in France, more high consumers of olive oil in Spain), while others were more
384 surprising. In particular, the proportion of high consumers of vegetables was extremely high in the 3-
385 C Bordeaux cohort, well above other estimations for France [39]. In the 3-C Bordeaux cohort, fruits
386 and vegetables consumption has already been associated with a lower risk of death [40] which may
387 explain that high consumers of vegetables are over-represented in our sample of very old
388 participants. In contrast, the proportion of high consumers of vegetables in the Seniors-ENRICA-2
389 cohort was extremely low (3.8%) compared to the 12.3% or 17.6% in the general Spanish population
390 [41] or in diabetic Spanish patients [36], respectively.
391 Keeping these issues in mind, and with the objective of describing broadly the impact of dietary
392 intake on salivary proteome, we used the GSEA approach. Many salivary proteins are multifunctional
393 [42] and thus GSEA provided an overwhelming wealth of information, but a first obvious result was
394 that the biological processes enriched according to the intake level were mostly different between
395 diabetic subjects and controls. The salivary proteome is therefore shaped differently by diet
396 depending on the physiopathological status of the subject.
397 The second finding is that intake of sweet snacks had a major effect on biological processes in
398 diabetics, regardless of the cohort considered. This is not surprising since dietary sugars have a
399 profound physiological effect in diabetic individuals, who by definition have difficulty in controlling
400 glycemia. For example 8 enzymes of the *glucose catabolic process to pyruvate* (out of the 10 of the
401 core glycolysis pathway) were over-represented in high consumers of sweet snacks among diabetic
402 participants. This is consistent with the long-known increase in glycolytic enzymes activity or
403 expression induced by glucose in various tissues or cells such as pancreatic cells [43] or the jejunum
404 [44]. Inter-individual variability in salivary glycolytic enzymes has been hypothesized to be at least
405 partly related to dietary habits of the healthy adult saliva donors [45]. Finally, several glycolytic

406 enzymes were found in saliva of diabetic subjects and these were overexpressed in subjects with the
407 most severe forms of retinopathy [46].
408 More unexpectedly, when analyzing the two cohorts jointly, the *endothelial cell apoptotic process*
409 was repressed in diabetic high consumers of sweets, while high glucose is known to trigger apoptosis
410 in endothelial cells [47]. However, the three proteins attached to this GO term in our dataset are the
411 fibrinogen α, β and γ-chains. They are linked to negative regulation of the apoptotic process in GO,
412 but their primary function is rather blood clotting: their over-expression could thus reflect the more
413 frequent gum bleeding in diabetic individuals with inadequate intake of sweet snacks. This is then in
414 accordance with a consensus report indicating that poor glycaemic control in diabetes is associated
415 with poorer periodontal status and outcomes [48]. This example illustrates that GSEA should be
416 regarded as indicative of biological processes to be investigated further, and findings should be
417 examined considering representativity of the proteins attached to each GO category. In that respect,
418 the example of *platelet degranulation* deserves special attention since it is associated with 30
419 proteins in our dataset (out of 129 in the human genome). Increased platelet degranulation is
420 observed in T2DM patients compared to healthy subjects [49,50] and a mechanistic study concluded
421 that high glucose *per se* increased platelet reactivity in blood of both diabetics and controls [51]. In
422 our study, combining both cohorts, the saliva proteome composition also suggested that high intake
423 of sweet snacks was linked with enhanced platelet reactivity both in diabetics or controls. This was
424 confirmed in the Spanish cohort and among diabetic subjects in the French cohort.
425 We also looked for markers of diet and adjusted the models for confounding factors (diabetic status,
426 cohort, age, sex etc.). Combining both cohorts, salivary proteins were retained in the model for three
427 food groups. For « vegetables » and « butter », one should note that the proportions of high
428 consumers are much higher in the 3C-Bordeaux cohort. Results should therefore be taken cautiously
429 because there is a risk that results are partly confounded with a cohort effect. In addition, the
430 analyses performed for each cohort separately highlighted the differences between the two cohorts
431 and illustrated the difficulty of identifying universal markers. Nevertheless, whatever the subjects'
432 characteristics in terms of cohort, age, sex, diabetic status, smoker status, BMI or saliva flow, higher

abundance of lactotransferrin (E7EQB2) and beta-2-glycoprotein 1 (P02749) in saliva was associated with a higher chance of being a high consumer of sweet snacks. Interestingly, we had previously observed in healthy children that another protein of the transferrin family (serotransferrin) was positively associated with a number of food groups including biscuits & sweets [19]. There is also some biological coherence among some of the identified markers. For example, the two positive markers of vegetables intake (lactoperoxidase, sulhydryl oxidase 1) catalyze reactions involving H_2O_2 and contribute to cellular redox homeostasis. However, the main general lesson from those results is that the abundance of some salivary proteins was linked to the consumption level of some food groups in elderly subjects after adjusting for several sociodemographic and clinical confounding factors. In addition, since including salivary proteins data improved the prediction models, this is the proof-of-concept that saliva might be a source of objective markers of usual diet.

To conclude, it should be reminded that the studied population originated from two very contrasted cohorts, especially in terms of age and dietary habits. This represented a challenge, particularly for statistical methods which can not fully account for confounding factors. Dietary intake was also here simply evaluated as compliance or not to the requirements of MeDi diet. However, the salivary proteome data suggested biological functions affected by the dietary intake level of sweet snacks, across cohorts and diabetic status (e.g. *platelet degranulation*) or across cohorts and specific to one diabetic status (e.g. functions linked to cell spreading and cell motility). Proteome data also added value to minimal models predicting the intake level of some food groups. It is now necessary to confirm these results on a validation population, but also to assess more finely the link between the potential markers' expression and the actual quantitative intakes.

References

1. Coelho-Júnior H, Rodrigues B, Uchida M, Marzetti E (2018) Low Protein Intake Is Associated with Frailty in Older Adults: A Systematic Review and Meta-Analysis of Observational Studies. Nutrients. https://doi.org/10.3390/nu10091334

2. Machado-Fragua MD, Struijk EA, Ballesteros J-M, Ortolá R, Rodriguez-Artalejo F, Lopez-Garcia E (2019) Habitual coffee consumption and risk of falls in 2 European cohorts of older adults. Am J Clin Nutr 109:1431‐1438. https://doi.org/10.1093/ajcn/nqy369

3. Singh B, Parsaik AK, Mielke MM, Erwin PJ, Knopman DS, Petersen RC, Roberts RO (2014) Association of Mediterranean Diet with Mild Cognitive Impairment and Alzheimer's Disease: A Systematic Review and Meta-Analysis. J Alzheimers Dis 39:271‐282. https://doi.org/10.3233/JAD-130830

4. Kojima G, Avgerinou C, Iliffe S, Walters K (2018) Adherence to Mediterranean Diet Reduces Incident Frailty Risk: Systematic Review and Meta-Analysis. J Am Geriatr Soc 66:783‐788. https://doi.org/10.1111/jgs.15251

5. Struijk EA, Hagan KA, Fung TT, Hu FB, Rodríguez-Artalejo F, Lopez-Garcia E (2020) Diet quality and risk of frailty among older women in the Nurses' Health Study. Am J Clin Nutr 111:877‐883. https://doi.org/10.1093/ajcn/nqaa028

6. The InterAct Consortium (2011) Mediterranean Diet and Type 2 Diabetes Risk in the European Prospective Investigation Into Cancer and Nutrition (EPIC) Study: The InterAct project. Diabetes Care 34:1913‐1918. https://doi.org/10.2337/dc11-0891

7. Salas-Salvadó J, Guasch-Ferré M, Lee C-H, Estruch R, Clish CB, Ros E (2015) Protective Effects of the Mediterranean Diet on Type 2 Diabetes and Metabolic Syndrome. J Nutr 146:920S-927S. https://doi.org/ 10.3945/jn.115.218487

8. André P, Proctor G, Driollet B, Garcia-Esquinas E, Lopez-Garcia E, Gomez-Cabrero D, Neyraud E, Rodriguez-Artalejo F, Morzel M, Féart C (in press) The role of overweight in the association between the Mediterranean diet and the risk of type 2 diabetes mellitus: a mediation analysis among 21 585 UK biobank participants. Int J Epidemiol. https://doi.org/10.1093/ije/dyaa103

9. Hu FB, Willett WC (2018) Current and Future Landscape of Nutritional Epidemiologic Research. JAMA. https://doi.org/10.1001/jama.2018.16166

10. Guertin KA, Moore SC, Sampson JN, Huang W-Y, Xiao Q, Stolzenberg-Solomon RZ, Sinha R, Cross AJ (2014) Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am J Clin Nutr 100:208‐217. https://doi.org/10.3945/ajcn.113.078758

11. Rabassa M, Zamora-Ros R, Palau-Rodriguez M, Tulipani S, Miñarro A, Bandinelli S, Ferrucci L, Cherubini A, Andres-Lacueva C (2020) Habitual Nut Exposure, Assessed by Dietary and Multiple Urinary Metabolomic Markers, and Cognitive Decline in Older Adults: The InCHIANTI Study. Mol Nutr Food Res https://doi.org/10.1002/mnfr.201900532

12. Li J, Guasch-Ferré M, Chung W, Ruiz-Canela M, Toledo E, Corella D et al (2020) The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. Eur Heart J 41:2645‐2656. https://doi.org/10.1093/eurheartj/ehaa209

13. Granger DA, Kivlighan KT, Fortunato C, Harmon AG, Hibel LC, Schwartz EB, Whembolua G-L (2007) Integration of salivary biomarkers into developmental and behaviorally-oriented research: Problems and solutions for collecting specimens. Physiol Behav 92:583‐ 590. https://doi.org/10.1016/j.physbeh.2007.05.004

14. Morzel M, Jeannin A, Lucchi G, Truntzer C, Pecqueur D, Nicklaus S, Chambon C, Ducoroy P (2012) Human infant saliva peptidome is modified with age and diet transition. J Proteomics 75:3665‐ 3673. https://doi.org/10.1016/j.jprot.2012.04.028

15. Castagnola M, Inzitari R, Fanali C, Iavarone F, Vitali A, Desiderio C, et al (2011) The Surprising Composition of the Salivary Proteome of Preterm Human Newborn. Mol Cell Proteomics. https://doi.org/10.1074/mcp.M110.003467

16. Wang K, Wang X, Zheng S, Niu Y, Zheng W, Qin X, Li Z, Luo J, Jiang W, Zhou X, Li W, Zhang L (2018) iTRAQ-based quantitative analysis of age-specific variations in salivary proteome of caries-susceptible individuals. J Transl Med. https://doi.org/10.1186/s12967-018-1669-2

17. Méjean C, Morzel M, Neyraud E, Issanchou S, Martin C, Bozonnet S, Urbano C, Schlich P, Hercberg S, Péneau S, Feron G (2015) Salivary Composition Is Associated with Liking and Usual Nutrient Intake. PLOS ONE. https://doi.org/10.1371/journal.pone.0137473

18. Morzel M, Palicki O, Chabanet C, Lucchi G, Ducoroy P, Chambon C, Nicklaus S (2011) Saliva electrophoretic protein profiles in infants: Changes with age and impact of teeth eruption and diet transition. Arch Oral Biol 56:634‐ 642. https://doi.org/10.1016/j.archoralbio.2010.12.015

19. Morzel M, Truntzer C, Neyraud E, Brignot H, Ducoroy P, Lucchi G, Canlet C, Gaillard S, Nicod F, Nicklaus S, Peretti N, Feron G (2017) Associations between food consumption patterns and saliva composition: Specificities of eating difficulties children. Physiol Behav 173:116‐ 123. http://dx.doi.org/10.1016/j.physbeh.2017.02.005

20. Walsh MC, Brennan L, Malthouse JPG, Roche HM, Gibney MJ (2006) Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. Am J Clin Nutr 84:531‐ 539.

21. Proctor GB, André P, Lopez-Garcia E, Gomez Cabrero Lopez D, Neyraud E, Féart C, Rodriguez Artalejo F, Garcıa-Esquinas E, Morzel M (2017) The SALAMANDER project: SALivAry bioMarkers of mediterraneAN Diet associated with long-tERm protection against type 2 diabetes. Nutr Bull 42:369‐ 374. https://doi.org/10.1111/nbu.12298

22. The 3C Study Group (2003) Vascular Factors and Risk of Dementia: Design of the Three-City Study and Baseline Characteristics of the Study Population. Neuroepidemiology 22:316‐ 325. https://doi.org/10.1159/000072920

23. Cabanas-Sánchez V, Esteban-Cornejo I, Migueles JH, Banegas JR, Graciani A, Rodríguez-Artalejo F, Martínez-Gómez D (2020) Twenty four-hour activity cycle in older adults using wrist-worn accelerometers: The seniors-ENRICA-2 study. Scand J Med Sci Sports 30:700‐ 708. https://doi.org/10.1111/sms.13612

24. Rahi B, Ajana S, Tabue-Teguo M, Dartigues J-F, Peres K, Feart C (2018) High adherence to a Mediterranean diet and lower risk of frailty among French older adults community-dwellers: Results from the Three-City-Bordeaux Study. Clin Nutr 37:1293‐ 1298. https://doi.org/10.1016/j.clnu.2017.05.020

25. Guallar-Castillón P, Sagardui-Villamor J, Balboa-Castillo T, Sala-Vila A, Ariza Astolfi MJ, Sarrión Pelous MD, Leon-Muñoz LM, Graciani A, Laclaustra M, Benito C, Banegas JR, Rodriguez Artalejo F (2014) Validity and Reproducibility of a Spanish Dietary History. PLoS ONE. https://doi.org/10.1371/journal.pone.0086074

26. Schröder H, Fitó M, Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, et al (2011) A Short Screener Is Valid for Assessing Mediterranean Diet Adherence among Older Spanish Men and Women. J Nutr 141:1140‐1145. https://doi.org/10.3945/jn.110.135566

27. Wei R, Wang J, Su M, Jia E, Chen S, Chen T, Ni Y (2018) Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. Sci Rep. https://doi.org/10.1038/s41598-017 19120-0

28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci 102:15545‐15550. https://doi.org/10.1073/pnas.0506580102

29. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci. 99:7821‐7826. https://doi.org/10.1073/pnas.122653799

30. Humphrey SP, Williamson RT (2001) A review of saliva: Normal composition, flow, and function. J Prosthet Dent 85:162‐169. https://doi.org/10.1067/mpr.2001.113778

31. Dawes C, Pedersen AML, Villa A, Ekström J, Proctor GB, Vissink A, et al (2015) The functions of human saliva: A review sponsored by the World Workshop on Oral Medicine VI. Arch Oral Biol 60:863‐874. https://doi.org/10.1016/j.archoralbio.2015.03.004

32. Affoo RH, Foley N, Garrick R, Siqueira WL, Martin RE (2015) Meta-Analysis of Salivary Flow Rates in Young and Older Adults. J Am Geriatr Soc 63:2142‐2151. https://doi.org/10.1111/jgs.13652

33. Vandenberghe-Descamps M, Labouré H, Prot A, Septier C, Tournier C, Feron G, Sulmont-Rossé C (2016) Salivary Flow Decreases in Healthy Elderly People Independently of Dental Status and Drug Intake. J Texture Stud 47:353‐360. https://doi.org/10.1111/jtxs.12191

34. Lima DLF, Carneiro SDRM, Barbosa FT de S, Saintrain MV de L, Moizan JAH, Doucet J (2017) Salivary flow and xerostomia in older patients with type 2 diabetes mellitus. PLOS ONE. https://doi.org/10.1371/journal.pone.0180891

35. García-Esquinas E, Rahi B, Peres K, Colpo M, Dartigues J-F, Bandinelli S, Feart C, Rodríguez-Artalejo F (2016) Consumption of fruit and vegetables and risk of frailty: a dose-response analysis of 3 prospective cohorts of community-dwelling older adults. Am J Clin Nutr 104:132-142.

36. Muñoz-Pareja M, León-Muñoz LM, Guallar-Castillón P, Graciani A, López-García E, Banegas JR, Rodriguez-Artalejo F (2012) The Diet of Diabetic Patients in Spain in 2008–2010: Accordance with the Main Dietary Recommendations—A Cross-Sectional Study. PLoS ONE. https://doi.org/10.1371/journal.pone.0039454

37. Virtanen S, Feskens E, Räsänen L, Fidanza F, Tuomilehto J, Giampaoli S, Nissinen A, Kromhout D (2000) Comparison of diets of diabetic and non-diabetic elderly men in Finland, The Netherlands and Italy. Eur J Clin Nutr 54:181‐186. https://doi.org/10.1038/sj.ejcn.1600916

38. Nöthlings U, Boeing H, Maskarinec G, Sluik D, Teucher B, Kaaks R, et al (2011) Food intake of individuals with and without diabetes across different countries and ethnic groups. Eur J Clin Nutr 65:635‐ 641. https://doi.org/10.1038/ejcn.2011.11

39. Péneau S, Galan P, Jeandel C, Ferry M, Andreeva V, Hercberg S, Kesse-Guyot E, the SU.VI.MAX 2 Research Group (2011) Fruit and vegetable intake and cognitive function in the SU.VI.MAX 2 prospective study. Am J Clin Nutr 94:1295‐ 1303. https://doi.org/10.3945/ajcn.111.014712

40. Letois F, Mura T, Scali J, Gutierrez L-A, Féart C, Berr C (2016) Nutrition and mortality in the elderly over 10 years of follow-up: the Three-City study. Br J Nutr 116:882‐ 889. https://doi.org/10.1017/S000711451600266X

41. León-Muñoz LM, Guallar-Castillón P, Graciani A, López-García E, Mesas AE, Aguilera MT, Banegas JR, Rodriguez-Artalejo F (2012) Adherence to the Mediterranean Diet Pattern Has Declined in Spanish Adults. J Nutr 142:1843‐ 1850. https://doi.org/10.3945/jn.112.164616

42. Proctor GB, Carpenter GH (1998) The function of salivary proteins and the regulation of their secretion by salivary glands. Biomed Rev 9:3‐ 9. https://doi.org/10.14748/bmr.v9.132

43. Roche E, Yaney G, Corkey B, Asfari M, Prentki M (1997) Induction by Glucose of Genes Coding for Glycolytic Enzymes in a Pancreatic beta-Cell Line (INS-1). J Biol Chem 272:3091‐ 3098. https://doi.org/10.1074/jbc.272.5.3091

44. Stifel FB, Herman RH, Rosenweig NS (1969) Dietary regulation of glycolytic enzymes. III. Adaptive changes in rat jejunal pyruvate kinase, phosphofructokinase, fructosediphosphatase and glycerol-3-phosphate dehydrogenase. Biochim Biophys Acta 184:29‐ 34. https://doi.org/10.1016/0304-4165(69)90094-4

45. Quintana M, Palicki O, Lucchi G, Ducoroy P, Chambon C, Salles C, Morzel M (2009) Inter-individual variability of protein patterns in saliva of healthy adults. J Proteomics 72:822‐ 830. https://doi.org/10.1016/j.jprot.2009.05.004

46. Chee CS, Chang KM, Loke MF, Angela Loo VP, Subrayan V (2016) Association of potential salivary biomarkers with diabetic retinopathy and its severity in type-2 diabetes mellitus: a proteomic analysis by mass spectrometry. PeerJ. https://doi.org/10.7717/peerj.2022

47. Baumgartner-Parzer SM, Wagner L, Pettermann M, Grillari J, Gessl A, Waldhäusl W (1995) High-glucose-triggered apoptosis in cultured endothelial cells. Diabetes 44:1323‐ 1327. https://doi.org/10.2337/diab.44.11.1323

48. Sanz M, Ceriello A, Buysschaert M, Chapple I, Demmer RT, Graziani F, et al (2018) Scientific evidence on the links between periodontal diseases and diabetes: Consensus report and guidelines of the joint workshop on periodontal diseases and diabetes by the International diabetes Federation and the European Federation of Periodontology. Diabetes Res Clin Pract 137:231‐ 241. https://doi.org/10.1016/j.diabres.2017.12.001

49. Cerbone AM, Macarone-Palmieri N, Saldalamacchia G, Coppola A, Di Minno G, Rivellese AA. (2009) Diabetes, vascular complications and antiplatelet therapy: open problems. Acta Diabetol 46:253‐ 261. https://doi.org/10.1007/s00592-008-0079-y

50. Randriamboavonjy V, Isaak J, Elgheznawy A, Pistrosch F, Frömel T, Yin X, Badenhoop K, Heide H, Mayr M, Fleming I (2012) Calpain inhibition stabilizes the platelet proteome and reactivity in diabetes. Blood 120:415‐ 423. https://doi.org/10.1182/blood-2011-12-399980

51. Keating FK, Sobel BE, Schneider DJ (2003) Effects of increased concentrations of glucose on platelet reactivity in healthy subjects and in patients with and without diabetes mellitus. Am J Cardiol 92:1362- 1365. https://doi.org/10.1016/j.amjcard.2003.08.033

Table 1. Main characteristics of the 211 participants whose saliva was analyzed in the study

	Seniors-ENRICA-2		3C-Bordeaux	
	diabetics (n=52)	controls (n=53)	diabetics (n=37)	controls (n=69)
Age (years, mean ± SD)	70.6 ± 4.3	71.4 ± 3.7	87.4 ± 4.1	87.3 ± 2.9
Sex (% men)	69	58	48	33
BMI (mean ± SD)	30.0 ± 5.0	26.6 ± 3.9	29.0 ± 3.9	24.4 ± 4.0
Smoker (never/former/current) [a]	9/37/6	15/33/5	17/7/5	33/20/5
Saliva flow (g/min, mean ± SD)	0.26 ± 0.17	0.23 ± 0.16	0.18 ± 0.13	0.20 ± 0.18

[a] 19 missing values (3C-Bordeaux)

Table 2. Modified MEDAS score (0 to 13 points) and percentage of low and high consumers of 11 food groups among the 211 older adults, altogether or when individuals are stratified by diabetic status or cohort. The cut-off points to define low and high intake are provided in Online Resource 1. Cells shaded in gray highlight the food groups for which there is a large (>10%) difference between the two sub-samples.

		All subjects	**Diabetic status**		**Cohort**	
			Diabetics	Controls	Seniors-ENRICA-2	3C-Bordeaux
		(n=211)	(n= 89)	(n=122)	(n=105)	(n=106)
13-point MEDAS score		6.24 ± 1.57	6.13 ± 1.52	6.32 ± 1.61	6.47 ± 1.39	6.02 ± 1.70
Vegetables						
	Low	56.4	60.7	53.3	***96.2***	***17.0***
	High	43.6	39.3	46.7	***3.8***	***83.0***
Fruits						
	Low	54.0	58.4	50.8	53.3	54.7
	High	46.0	41.6	49.2	46.7	45.3
Olive oil						
	Low	88.2	85.4	90.2	***80.0***	***96.2***
	High	11.8	14.6	9.8	***20.0***	***3.8***
Legumes						
	Low	91.0	88.8	92.6	88.6	93.4
	High	9.0	11.2	7.4	11.4	6.6
Nuts						
	Low	82.9	***91.0***	***77.0***	***89.5***	***76.4***
	High	17.1	***9.0***	***23.0***	***10.5***	***23.6***
Wine						
	Low	65.4	69.7	62.3	***82.9***	***48.1***
	High	34.6	30.3	37.7	***17.1***	***51.9***
Fish						
	Low	68.2	70.8	66.4	64.8	71.7
	High	31.8	29.2	33.6	35.2	28.3
Red meat						
	Low	92.4	91.0	93.4	90.5	94.3
	High	7.6	9.0	6.6	9.5	5.7
Butter						
	Low	61.6	66.3	58.2	***89.5***	***34.0***
	High	38.4	33.7	41.8	***10.5***	***66.0***
Sweet snacks						
	Low	63.5	***70.8***	***58.2***	63.8	63.2
	High	36.5	***29.2***	***41.8***	36.2	36.8
Sweet drinks						
	Low	91.5	86.5	95.1	88.6	94.3
	High	8.5	13.5	4.9	11.4	5.7

Table 3. Number of proteins significantly differently expressed (FDR<5%) in saliva of low vs high consumers of 11 food groups, altogether or when individuals are stratified by diabetic status or by cohort. The cut-off points to define low and high intake are provided in Online Resource 1.

	Vegetables	Fruits	Olive oil	Legumes	Nuts	Wine	Fish	Red meat	Butter	Sweet snacks	Sweet drinks
All participants (n=211)	22	0	0	0	0	0	0	0	3	8	0
Diabetics (n=89)	4	0	0	0	0	0	0	0	0	65	0
Controls (n=122)	20	0	0	2	0	0	0	58	5	0	0
Seniors-ENRICA-2 (n=105)	0	63	0	0	0	0	0	0	0	14	0
3C Bordeaux (n=106)	0	0	2	0	0	0	0	0	0	0	0

Table 4. Variables retained in minimal models of prediction of intake of 11 food groups (two cohorts combined). Quantitative variables are separated into positive (chance of being a high consumer increases with value) or negative (chance of being a high consumer decreases with value) predictors. For categorial variables, the category associated with high intake is reported. Proteins are identified by their UniProt entry reference.

Food groups	**Quantitative variables**		**Categorial variables**
	positive	**negative**	
Vegetables	P22079, O00391	P10909, P37802, Q8TAX7	French cohort
Fruits			
Olive oil	BMI		Spanish cohort
Legumes			Former or current smokers
Nuts			Non-diabetics
Wine			French cohort, Men
Fish		Saliva flow	
Red meat			Men
Butter	Q96DR5, P22079	P23280, P07476	French cohort
Sweet snacks	E7EQB2, P02749		Non-diabetics, Men
Sweet drinks		Age	Former or current smokers

Table 5. Variables retained in minimal models of prediction of intake of 11 food groups (for each cohort separately). Quantitative variables are separated into positive (chance of being a high consumer increases with value) or negative (chance of being a high consumer decreases with value) predictors. For categorial variables, the category associated with high intake is reported. Proteins are identified by their UniProt entry reference.

Food groups	Quantitative variables		Categorial variables
	positive	negative	
Seniors-ENRICA-2			
Vegetables	BMI	Age	Women
Fruits	Q86YZ3, P0DMV8	P06753, P01024	
Olive oil	BMI		
Legumes			Women
Nuts			Non-diabetics
Wine			Men
Fish			
Red meat		Q96DR5, Q8TDL5, P01024, Age	Men
Butter			
Sweet snacks	P61158, P07195, P02763, P61769, saliva flow	P06744	Non-diabetics
Sweet drinks		Age	
3C-Bordeaux			
Vegetables	Age		
Fruits		Age, saliva flow	
Olive oil			Non-diabetics
Legumes			Men
Nuts		Saliva flow	
Wine			Men
Fish		Saliva flow	
Red meat			Diabetics
Butter			
Sweet snacks	BMI		Non-diabetics, Men
Sweet drinks			

Figure Captions

Fig. 1 Biological processes (Gene Ontology terms) associated with saliva proteins in high consumers of sweet snacks. Gene Set Enrichment Analysis was conducted separately for diabetic participants or controls. A « positive » or « negative » effect refers to the situation where proteins attached to the GO term are over-expressed and under-expressed, respectively, in high consumers of sweet snacks. The size of the dots indicates the number of proteins attached to each GO term in the experimental dataset, and the colour of the dots translates the gene ratio, i.e. the number of proteins attached to each GO term in the experimental dataset divided by the total number of proteins linked to that GO category in the human genome. In addition, four biological processes are common to all high consumers of sweet snacks, regardless of their diabetic status : *regulation of cellular macromolecule biosynthetic process* (negative effect), *platelet degranulation, leukocyte migration involved in inflammatory response* and *regulation of immune system* (positive effect).

Fig. 2 Estimation of the performance of minimal models predicting the consumption level of vegetables, butter and sweet snacks : distribution of AUCs (Area under the Curve) of the ROC (receiver operating characteristic) curves for 10000 simulations. For each food group, results are presented for three models: minimal model selected as described in the Material & Methods section (upper panel), minimal model with systematic inclusion of age, sex, cohort, diabetic status (middle panel), minimal model without protein variables (lower panel).