Spectroscopic molecular-fingerprint profiling of saliva

Buchan, Emma; Kelleher, Liam; Clancy, Michael; Rickard, Jonathan James Stanley; Oppenheimer, Pola Goldberg

DOI:
10.1016/j.aca.2021.339074

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Buchan, E, Kelleher, L, Clancy, M, Rickard, JJS & Oppenheimer, PG 2021, 'Spectroscopic molecular-fingerprint profiling of saliva', *Analytica Chimica Acta*, vol. 1185, 339074. https://doi.org/10.1016/j.aca.2021.339074

Analytica Chimica Acta 1185 (2021) 339074

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Spectroscopic molecular-fingerprint profiling of saliva

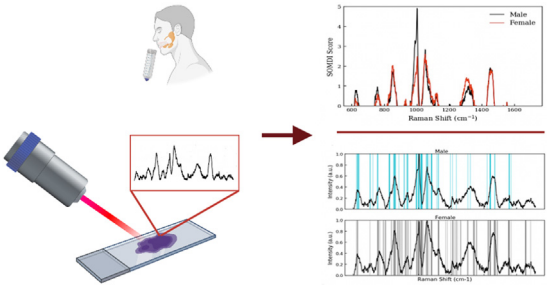
Emma Buchan [a], Liam Kelleher [a], Michael Clancy [a], Jonathan James Stanley Rickard [b], Pola Goldberg Oppenheimer [a, c, *]

[a] School of Chemical Engineering, Advanced Nanomaterials Structures and Applications Laboratories, College of Engineering and Physical Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
[b] Department of Physics, Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge, CB3 0HE, UK
[c] Healthcare Technologies Institute, Institute of Translational Medicine, Mindelsohn Way, Birmingham, B15 2TH, UK

HIGHLIGHTS

- Rapid profiling of saliva as a function of age, gender and temporal evaluation.
- Spectral data classified *via* artificial neural networks algorithm - 'SKiNET'.
- Molecular fingerprints distinguishing *finite* changes.
- "Multichemical barcode" as a characteristic tool for spectroscopic analytical studies.
- Reliable Raman spectroscopy protocols for sensing as analytical tool for diagnostics.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 May 2021
Received in revised form
4 September 2021
Accepted 15 September 2021
Available online 18 September 2021

Keywords:
Raman spectroscopy
Saliva profiling
Multivariate analysis
Diagnostic forensic biofluid

ABSTRACT

Saliva analysis has been gaining interest as a potential non-invasive source of disease indicative biomarkers due to being a complex biofluid correlating with blood-based constituents on a molecular level. For saliva to cement its usage for analytical applications, it is paramount to gain underpinning molecular knowledge and establish a 'baseline' of the salivary composition in healthy individuals as well as characterize how these factors are impacting its performance as potential analytical biofluid. Here, we have systematically studied the molecular spectral fingerprint of saliva, including the changes associated with gender, age, and time. *Via* hybrid artificial neural network algorithms and Raman spectroscopy, we have developed a non-destructive molecular profiling approach enabling the assessment of salivary spectral changes yielding the determination of gender and age of the biofluid source. Our classification algorithm successfully identified the gender and age from saliva with high classification accuracy. Discernible spectral molecular 'barcodes' were subsequently constructed for each class and found to primarily stem from amino acid, protein, and lipid changes in saliva. This unique combination of Raman spectroscopy and advanced machine learning techniques lays the platform for a variety of applications in forensics and biosensing.

1. Introduction

In the past decade saliva has been emerging as a potential candidate bodily fluid for medical diagnostics and forensic

* Corresponding author. Healthcare Technologies Institute, Institute of Translational Medicine, Mindelsohn Way, Birmingham, B15 2TH, United Kingdom.
E-mail address: GoldberP@bham.ac.uk (P.G. Oppenheimer).

https://doi.org/10.1016/j.aca.2021.339074

applications [1–3]. Owing to its non-invasive collection, simple, or no, pre-processing required prior to analysis and yet, a similar rich molecular composition as of blood, this complex biofluid has been increasingly gaining scientific interest [4,5]. Like blood, saliva is rich in biochemical information and is comprised of a complex mixture of proteins, lipids and many biological and inorganic components including hormones, antibodies, and growth factors [1,6,7] as well as a breadth of disease biomarkers [4,8–10]. Thus, saliva is an attractive candidate biofluid to complement the routine use of blood for patient screening in diagnostics [11], dentistry [12,13] or for evidence collection in forensics [14–16]. In contrast to blood however, saliva exhibits added advantages including the straightforward and less-demanding collection procedures, alleviating the discomfort to the subject in question and, the less physically intrusive but more cost-effective analyses options. Saliva procurement is non-invasive and painless, does not require highly trained personnel, and the collected samples are easily stored and transported. Importantly, in contrast to blood, saliva does not clot with time. Additionally, salivary secretions from individuals diagnosed with blood borne pathogens such as HIV have negligible rates of oral transmission due to factors that inhibit the infectivity of pathogens and therefore, are considerably safer to handle.

Saliva therefore, has been increasingly exploited for biomedical applications in diagnostic assays including, the detection of acute myocardial infarction and Amyotrophic Lateral Sclerosis, diagnosis of Tamiflu-resistant Influenza virus and tracking urease activity [17–20]. Salivary research is also ideal for investigations relating to cancer research and diagnostics. It has been applied in oral, gastric, breast, lung, prostate and ovarian cancers to name a few [4,5,8,10,21]. Li et al. established an 80% accuracy and 83% specificity in identifying lung cancer using surface enhanced Raman spectroscopy (SERS) [22]. Likewise, Connolly et al. revealed a sensitivity of 89% and diagnostic accuracy of 73% in the determination of oral squamous cell cancer and Chen et al. determined a specificity of >87.7% and sensitivity <80% in distinguishing early and advanced gastric cancer [4,5].

In the field of forensics, several recent studies have examined suitability of saliva in the identification of victims and suspects [23–25]. This application has a particular importance at a crime scene, where bite marks are found, indicating involvement of an oral cavity [26]. Currently, DNA is the primary source used for evidence gathering in forensics however, it is only conclusive if the genetic profile of the individual is fully known. Therefore, an additional source for these purposes, which is also inherently more cost-effective could be used for targeted diagnostics, with saliva posing as a complementary and sometimes, preferred biofluid for screening and spot-checks due to availability and ease of collection. Despite these favourable attributes, the use of saliva as a diagnostic and forensic biofluid is yet to become mainstream.

Concurrently, Raman spectroscopy (RS), a form of a sensitive vibrational spectroscopy, which by exciting molecular bonds within a sample, provides a unique biomolecular spectral fingerprint of target analytes, has been increasingly studied for diagnostic applications [4–6,27]. RS enables non-destructive quantitative analysis of chemical composition and structure, requires no complex sample preparation and poses an inherently straightforward detection in aqueous conditions [8,27]. Spectroscopic 'fingerprints' of Raman spectra have been shown to identify the health/disease state of the subject from which the biofluid has been collected [4,5,8,9]. In contrast to the *in-vitro* bioassays, the availability of inexpensive, portable RS instruments makes this technique particularly attractive for point-of-care sampling, analysis and screening of biofluids such as, saliva.

There are a number of Raman spectroscopy studies where saliva has been deployed towards exploring applications in forensics [27,28], therapeutics in chemotherapy [29–31], detection of oral cancer [21,32,33] and dental disease diagnostics [12,34,35]. SERS and microfluidics were further used to enhance the spectral response of saliva and to create small test platforms to be used [8,36,37]. However, none of these studies paid attention to the protein content of saliva samples and understanding of the role of proteins in the differentiation of abnormal cells from the control, protein backbone vibrations and the amide III regions. Therefore, understanding the protein content of saliva is necessary to take full advantage of the synergistic functions of saliva. The inherent inhomogeneity of saliva is most often overlooked by researchers, hence missing the fundamental characteristics of the sample. The utilisation of Raman spectroscopic profiling of healthy saliva also remains rather unexplored. Currently, there are no confirmatory tests specific to saliva in the field of forensic sciences. Although the presumptive testing is based on the activity of amylase in the sample, it is found in two different forms within the human body and is therefore, not exclusive to saliva.

Whilst research continues unravelling saliva as a potential alternative to blood- and tissue-based diagnostics, it is imperative to establish a 'baseline' of inherent characteristics of this biofluid in healthy subjects, where the detected changes *via* Raman spectroscopy could be attributed to the underpinning variations in gender, age and function of temporal evolution. The only study, which has examined the gender related spectroscopic variations, is by Muro et al. who used internal validation of saliva samples *via* Support Vector Machine Discriminant Analysis (SVMDA) for forensics discrimination [38]. While the SVMDA model proved successful, no comparison of spectral changes due to gender was established to indicate spectral regions of change, and the authors concluded that there is a need for advanced multivariate analysis due to the high standard deviation of measurements [38].

Here, we present a proof-of-concept Raman spectroscopy profiling and classification of saliva as a function of age and gender as well as temporal evaluation, towards establishing an important baseline of healthy saliva in a form of a "multi-biochemical barcode", as a characteristic tool for ongoing and future spectroscopic studies for diagnostic and forensic applications. We examine and systematically evaluate what is the spectral variability between healthy individuals, divided to female and male groups, as well as the main three representative age groups of young, middle-age and older adults and how these affect Raman spectra while gaining insights on the biochemical interpretation of spectral data. The effect of time from collection of saliva on spectral inter-variability is further examined to establish whether saliva as a biofluid exhibits major temporal changes in comparison to blood, which has been shown to considerably vary as a function of time from collection. The acquired spectral data is classified using our new artificial neural networks algorithm, SKiNET as a decision support tool [39]. SKiNET is based on the separation of data classes in a self-organising map (SOM) with characterisation using a self-organising map discriminant index (SOMDI) enabling the subsequent classification of the tested data (Fig. S1, Tables S1–S2). The SOM is a model inspired by nature and the way that neurons in the visual cortex are spatially organised according to the type of visual stimuli. The SOM defines a 2D map of neurons, typically arranged as a grid of hexagons. Each neuron is assigned a weight vector, which is initialised randomly and has a length equal to the number of variables in a spectrum. The weight vector effects which neuron will be activated for a given sample and neighbouring neurons will have similar weights. Spatial clustering is therefore observed in the trained map, with spectra that exhibit distinct properties activating different neurons. To understand which features in the data cause certain neurons to activate over others, SOMDI is used. The SOMDI introduces class vectors as labels for each spectrum and

corresponding weight vectors for each neuron, without influencing the training process. These allow for the identification of what type of data a given neuron activates, which can be used to inspect the weights across all neurons and extract prominent features belonging to each class. Neurons (hexagons) are coloured according to the modal class they activate, from the training set of Raman spectra. Neurons that have no majority class or activate none of the training data are shown in white. Coloured circles within each neuron represent spectra from the training data that have been activated for that neuron. For each class, there is a clearly defined block of neurons, with many of these activating only a single tissue type. An approximately even distribution in the number of neurons required to identify each class observed. The SOMDI provides a representation of weights associated with neurons that identify a particular class. A higher SOMDI intensity indicates a greater importance of particular inverse centimetres along the axis of a spectrum. Well defined peaks can be easily resolved, which are either more prominent or unique to each class. Automated classification of Raman spectra and assignment to a particular tissue type or disease state is perhaps the most important step for the translation of Raman based diagnostic techniques to real world, clinical applications.

Through inspection of key differences between neuron weights and class weight vectors, the algorithm enables identification of the key spectral changes. Training parameters used for the SOM included the grid size of 7, the learning rate of 0.2 and 10 epochs. From the separation of classes, it is evident that there are characteristic differences due to the obvious classification of certain neurons. As such, there is a clear basis for differentiation enabling characteristic weight vectors to be derived in the SOMDI. The identified barcodes from this study act as a reference, constituting a solid basis towards developing standard protocols as an essential prerequisite for reliable studies aimed at establishing the feasibility of Raman spectroscopy as analytical tool for diagnostics and forensics. Molecular barcodes can further be constructed for distinguishing between disease and healthy states associated with spectral changes *via* an easy subtraction of the variations from the reference sample spectra. In conjunction with the emergence of state-of-the-art machine learning techniques, the development of reliable, rapid spectroscopic analytical tools, ultimately promises improved forensic science technologies, helping to identify a possible victim, suspect or criminal as well as a better quality and timeliness of disease diagnostics and tailored treatments.

2. Results and discussion

We have established a process aimed to identify how age and gender of the participants vary and impact the characteristic Raman fingerprints of the extracted saliva samples as well as to uncover the effects of time on saliva. Namely it includes saliva collection, centrifugation, deposition onto an aluminium slide and measuring the corresponding Raman spectra (Fig. 1).

The measurement procedure was optimized to achieve high signal-to-noise spectra, with a 785 nm laser used for probing the samples composition due to its low levels fluorescence induction, minimal sample damage and good response in the fingerprint region [42]. Fresh saliva samples, from an overall of 38 female and 32 male participants with age groups ranging between 22 and 70 y (<39>±15 y) and 24–80 y (<40>±16 y) respectively, were analysed using our recently developed neural network SKiNET algorithm [41]. The derived weight vectors characteristic to each class are shown in Fig. 2c–e. The greater the normalised intensity of each wavenumber, the more important that scatter band is to a class spectral fingerprint.

Participants have been further sub-divided into age groups consisting of young adults (20–25 and 26–30 years), mid-aged adults (31–55 years) and older adults (56+ years) and the representative mean spectra of male ($n = 32$) and female ($n = 38$) saliva for each group are shown in Fig. 2a and b. Age selection criteria was based on healthy adults with no known medical conditions that fit into each of the age ranges (*see* **'Materials and Methods'**) [15,38,68,70]. The most prominent spectral features of the saliva for these groups are found at 630, 760, 855, 930, 1003, 1051, 1125, 1205, 1304, 1335, 1455, and 1655 cm^{-1}, with the associated assignments summarised in Table 1.

Gender based spectral differences of male saliva (black) and female saliva (red) are found to vary significantly (Fig. 2g-h). SOMDI scores show that it is intensity changes of the saliva rather than peak shifts that are responsible for the main differences. The male spectra were found to be distinct from female spectra due to the increased response at 630, 760, and 1003 cm^{-1} bands, with the female spectra showing an increased 855 cm^{-1} and 1300-1400 cm^{-1} response. Subtle differences are further noticeable at the 1051 and 1455 cm^{-1} shifts. The overall gender of saliva samples (independent of age) has been classified with an accuracy of 93.0 ± 0.5% with a 10-fold cross validation. Protein, lipid, and amino acid assigned peaks are identified to represent the main changes in the spectra, accordingly.

The peak at 630 cm^{-1}, attributed to phenylalanine and linked to metabolites concentration, is decreased in the female spectra where salivary metabolites such as, the taurine and lactate are commonly identified [43]. The 1003 cm^{-1} peak is typically assigned to phenylalanine as well, but may also be attributed to lysine, which is known to be one of the most abundant amino acids found in saliva [44]. Phenylalanine is found in both the sequences for amylase and lipase, two of the major components of saliva. A subtle peak shift is observed for the amide III region (1205-1300 cm^{-1}) showing that the composition of saliva in both males and females is distinct from each other. A difference in salivary composition can be linked to flow rate, where Inoue et al. and Li-Hui et al. showed that the flow rate of saliva in females is considerably lower than in males as their submandibular glands, where unstimulated saliva is mainly derived from, are much smaller than their male counterparts [45,46]. Lastly, the Raman spectra of saliva show a lack of response in the amide I region. Amide I along with aromatic breathing and C–H stretching peaks typically indicate a significant presence of protein in a sample [8,13].

Of a further consideration in a gender related spectroscopic study of female saliva is the effect of oestrogen on the salivary metabolome [47]. Salivary flow rate is significantly reduced in post-menopausal women in comparison with those who are menstruating (typically, ages of 45–50 are considered as the average menopause range) [48–50]. It has also been observed that salivary pH is significantly lower in post-menopausal women and the pH of the mouth plays a key role in the salivary Raman response [48]. The normal pH of saliva is 6–7 however, this can vary from 5.3 (low flow) to 7.8 (peak flow). The buffer capacity of saliva includes protein (amylase, mucin and immunoglobulin A), bicarbonate and phosphate buffer systems and their main role is to maintain a constant pH [51,52]. The concentration of bicarbonate detected in saliva is dependent on salivary flow rate. Thus, when the flow rate is increased, both bicarbonate levels and the pH also increase, leading to variance in the saliva spectra. Bicarbonate is observed at 1070 cm^{-1} and Fig. 2h shows a higher peak intensity in males than females indicating a higher pH and thus, a faster flow rate. In addition, women are affected more often than men with autoimmune diseases, which are known to affect salivary gland function [53]. Therefore, this could also have an effect on the salivary flow rate, as for example in the Sjögren's syndrome, where one of the most common symptoms is a dry mouth [54].

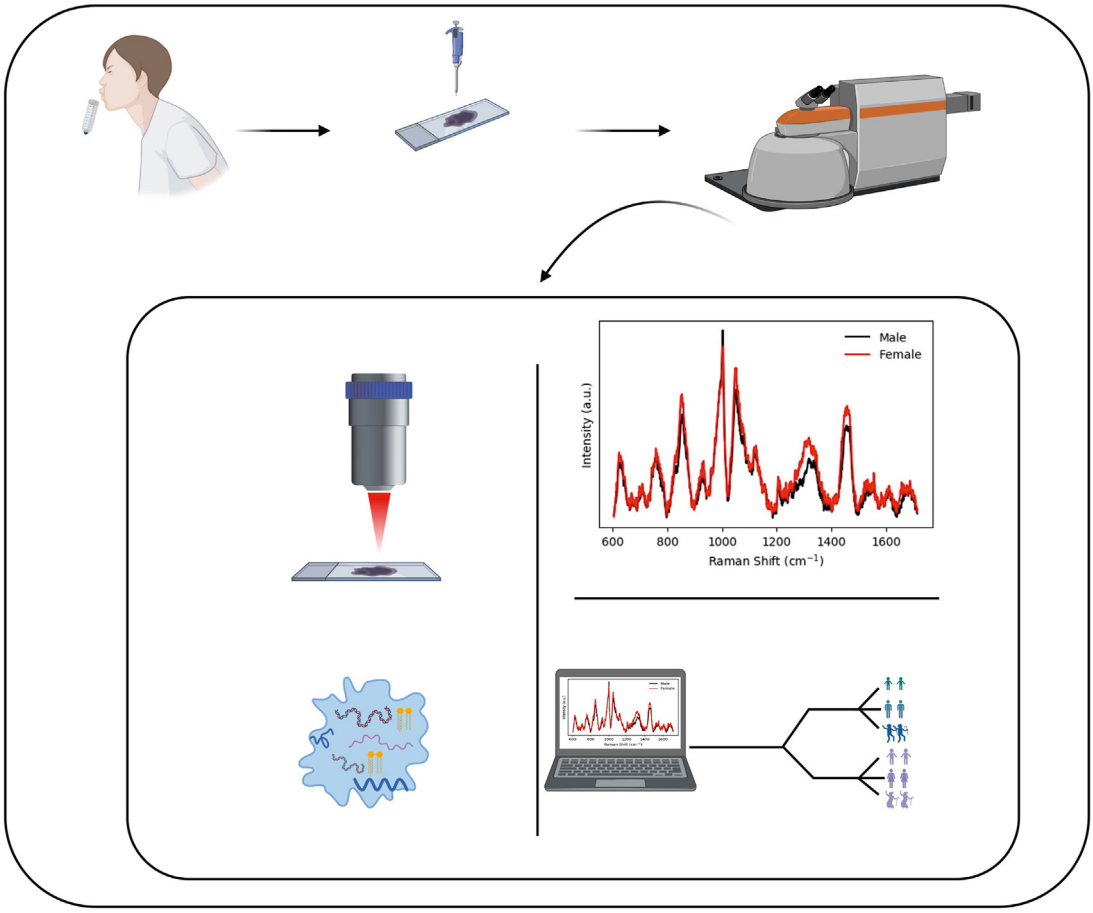


Fig. 1. Schematic representation of the experimental procedure for collecting and measuring saliva *via* Raman spectroscopy with the subsequent analysis of age, gender and time parameters.

The peak at 1340 cm^{-1}, known to be associated with collagen, has a higher intensity in females than males. Collagenase 3 (MMP-13) expression was first observed in breast cancer and is readily associated with diseases such as the osteoarthritis, which females are more likely to develop than men. A recent study by Virtanen et al. indicated an increased MMP-13 levels in female saliva [55]. The link between MMP-13 and its initial detection in breast cancer may potentially explain the increase observed in women and it is plausible that its metabolism could be an indication of hormonal associations. MMP-13 is also known to initiate bone resorption and activate proMMP-9. Both matrix metalloproteinases have been associated with periodontitis. Hernandez *et* al. identified elevated levels of MMP-13 in patients with chronic periodontitis indicating MMP-13 may play a role in the destruction of periodontal soft tissue [56].

Interestingly, age related spectral changes of saliva split the male and female datasets further into chronological age categories of young (20–30 years), mid (31–55 years) and older (56+ years) adults. These results, for the first time, indicate that it is possible to detect and monitor the age of a person using Raman spectroscopy of fresh saliva for both male and females.

Age based results empirically showed that the 20–25 and 26–30 age groups are closely clustered together and therefore, these have been grouped together for the further analysis. The variation in signal intensity for the four most prominent peaks as a function of the various ages is shown in Fig. 3a and b for female and male saliva. The largest changes noted are for the 56+ age category, where the female spectral fingerprints are found to change the most. The largest differences in peak intensity are observed in the 56+ age group at 855, 1076 and 1456 cm^{-1}, accounting for the ring breathing mode and C–C stretching, respectively. For female saliva samples, there is a clear separation according to the age groups of 20–30 y (black), 31–55 y (red), and >56 y (green), classified with an accuracy of 70.0 ± 1.2% (Fig. 2c, left). The SOMDI, akin to a PCA loading plot, shows in order of decreasing influence that the 1051, 1003, 1456, 855 cm^{-1} peaks as well as the region of 1300–1400 cm^{-1} account for the most variance in the classes. It is further noted that the older adult category, *i.e.,* >56 y, is the most distinctive age group, where an increase in peak intensity in the 1076, 1455, and 855 cm^{-1} is identified. The younger adult category of female spectra is distinctive by the decreased intensity in the 1300-1400 cm^{-1} region and the increased intensity at the 1003, 630, and 760 cm^{-1} peaks. Similarly, male age groups of saliva also exhibit the biggest variance in the ring breathing mode, C–N stretches for the >56 y group with an increased peak intensity at 1455, 1051, and 855 cm^{-1}. However, unique to the >56 y male group, is the increase in the 760 cm^{-1} and in the 1300-1400 cm^{-1} peaks, attributed to the tryptophan ring breathing mode and amide III region (proteins). For the 20–30 y and 31–55 y age groups, a Raman shift occurs in the 1200-1300 cm^{-1} region to 1300-1375 cm^{-1}.

It is well-established that as people age, the rate, volume and composition of saliva changes [57]. Histomorphometric examinations of healthy salivary gland tissues have revealed a decrease in the number of acinar cells as in individual ages. Xu *et* al.

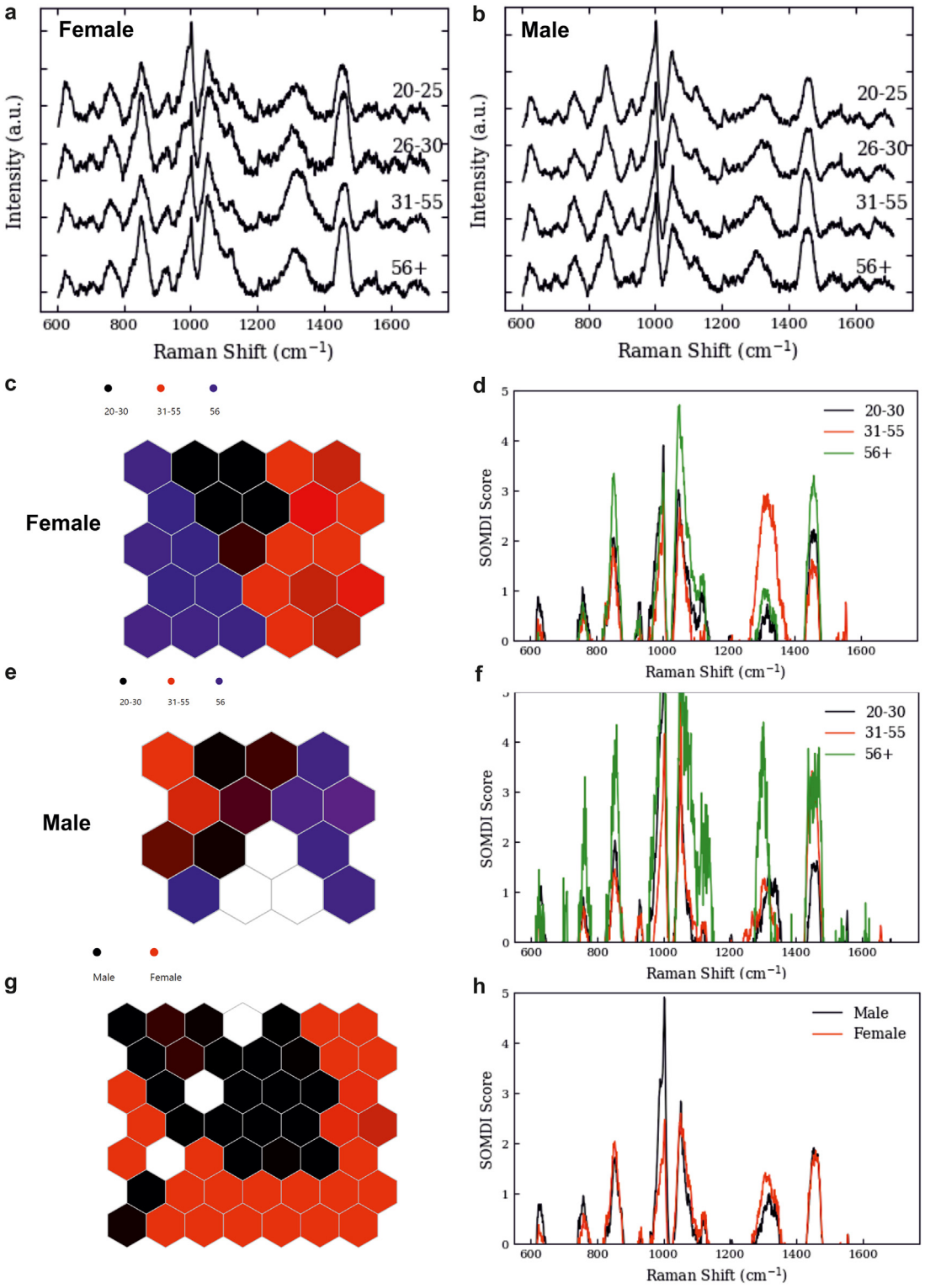


Fig. 2. Representative mean spectra categorised by age for **(a)** female and **(b)** male saliva. SOM (left) and SOMDI (right) of **(c)** female and **(e)** male saliva, successfully classifying the saliva samples according to age with an accuracy of 70.0 ± 1.2% and 78.0 ± 0.1%, respectively. The sub-classification according to age for female and male samples are represented by black hexagons for 20–30 years, red hexagons for 31–55 years and purple hexagons for the age group above 56 years. **(g-h)** Classification of female (red) *versus* male (black) saliva with an accuracy of 93.0 ± 0.5%. The biggest changes are observed at 630, 760, 855 and 1300 cm^{-1} bands. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

hypothesised that despite a loss of acinar cells, a secretory reserve exists to maintain the function of the salivary gland, thus potentially accounting for salivary changes with age. We indeed observe this effect by the shift in the amide III peak of saliva for both genders although, it is more distinct in the male saliva. Both lysine (1003 cm^{-1}) and glycine (1327 cm^{-1}) are known to be the most

Table 1
Characteristic Raman peak assignments for the identified peak wavenumbers [9,10,33,49].

Peak Wavenumber (cm^{-1})	Assignment
628	C–S stretch and C–C twist Protein; tyrosine
760	Ring breathing mode tryptophan; proteins
855	C–C; ring breathing mode tyrosine
930	C–C stretch amino acids (proline, hydroxyproline and valine); proteins
960	Calcium-phosphate stretching band (cholesterol), α-helix Proline, Valine (ν (C–C))
1003	Symmetric ring breathing mode (phenylalanine, tryptophan)
1051	C–O and C–N stretch
1070	HCO_3
1076	C–C (Lipids); symmetric stretch of phosphates in hydroxyapatite
1087	Vibration of DNA backbone; C–N stretch (proteins); PO_3^{-2} anti-symmetric stretch
1125	C–C skeletal stretch (lipids); C–N stretch (proteins)
1205	Amide III; CH_2 wagging and vibrations (glycine, proline, tyrosine and phenylalanine)
1300	Amide III; CH_2 twisting and wagging (lipids and/or proteins)
1337/1339	CH_2/CH_3 wagging and twisting (proteins, nucleic acid, lipids), nucleic acid bases (ν (C–H))
1456	CH_2 and CH_3 deformation vibrations (proteins and lipids)
1655	Amide I region; C=C stretch (lipids); C=O stretch (proteins)

abundant amino acids in saliva and increase in concentration in males and females as they age [57]. In addition, Nagler *et* al. discuss differences in salivary gland size and the implication this would have on salivary components in the ageing process of both males and females [58]. The authors determined that older patients tend to commonly experience xerostomia, causing dry mouth and affecting salivary production and speech, causing oral infections and dental caries. At least 25% of the elderly population have xerostomia and many display symptoms without realising it. Although all patients are screened for health conditions, elderly patients may not have been officially diagnosed with xerostomia and thus, it may explain the significant variation in the salivary spectra of the elderly.

Furthermore, the intensity changes observed in the 630 and 1003 cm^{-1} peaks between the 20–30 y and the >56 y groups indicate that the concentration of the associated amino acids is decreasing with increasing age. This is attributed to the reduction of the excretion rate of saliva with age and subsequently resulting in the decrease in the amino acids concentration [59]. In addition, a strong intensity change is observed between the age classes, with a higher response in the 56+ group at 1076 cm^{-1}. This band is attributed to lipids, with the spectra indicative of increased biological matter in older saliva (56+). Underpinning this, might be the lower saliva production rates, which can act as a discernible factor when compared to saliva of younger people. A meta-study by Affoo *et* al. found that saliva flow rate was significantly lower in older adults with such a decrease specifically attributed to the saliva from submandibular and sublingual salivary glands [60]. This decrease in flow rate can be also attributed to the loss of acinar cells, increase in adipose tissue and neurophysiological deterioration. In histopathological samples of salivary glands, glands of young adults show a more even and compact lobar structure with a uniform appearance of parenchymal elements when compared to those of an older individual [57]. With age, histological studies indicate that although the number of salivary glands remain constant, the volume of fat and fibromuscular tissue increases in both the parotid and submandibular glands. Conversely, the authors also state that in elderly patients a reduced volume of acinar cell secretion is observed and is considered as one of the main causes of dry mouth. Combined, these changes can also lead to salivary gland hypofunction [57]. Also, with age the number of salivary glycoproteins increases as indicated in a mass spectroscopy study by Sun *et* al., where the N-glycoproteins were found to be related to innate immunity an individual has built over their lifetime against microorganisms and oral cavity protection [61]. In our case, the detected increase in proteins and the higher response in the over 56 group may indicate that a higher level of innate immunity has been established, thus a decrease in salivary flow rate appears to have an indirect influence on the overall quality of an individual's saliva.

One of the major components of saliva is amylase. However, two different forms are found in the human body, the AMYI and the AMY2. AMYI (sAA) is detected more in saliva than in any of the other biofluids and therefore, has the potential to act as a unique biomarker. Amylase contains phenylalanine in its sequence, resulting in a strong Raman peak at 1003 cm^{-1}. This peak is detected with a higher intensity in younger adults, 20–30 years of age, as opposed to the older 56+ group. Different studies have conflicting opinions regarding levels of AMYI and its age-related differences. Ben-Aryeh et al. indicated there was no statistical difference in AMY1 between age groups [62], however Yim *et* al., determined the opposite [63]. Rutherfurd-Markwick *et* al. have shown that the amylase content of saliva increases steadily up to the middle age prior to stabilising and remaining constant in older adults [64]. The authors observed that amylase activity at rest was higher in a group of women than men. AMYI is secreted by the salivary glands but at birth AMYI is not detected. Levels increase as an individual ages to adulthood. Similarly, salivary glands continually grow throughout childhood and mature during adulthood and thus, levels of AMYI indicate developmental differences and potentially influence Raman spectra of saliva. Other studies have found that diet and stress affect AMYI levels. Perry et al. have shown that populations or individuals with a high starch diet had a higher AMYI level than those with low-starch diets [65]. Furthermore, AMYI is secreted by the salivary glands in response to stressors. Chatterton et al. have shown that psychosocial and physical stressors can rapidly increase levels of AMYI.

In addition to amylase, lysozyme makes up another major component of saliva. The lysozyme peak was observed at 1343 $cm^{-1}$with distinct intensity changes between the age groups. In both males and females, the peak intensity was higher in older adults (30+). Lysozyme, a salivary protein plays a role in the non-immunological bacterial defence system. A study by Lira-Junior et al. found that elderly individuals (>64 years of age) had elevated counts of 24 out of the 41 investigated bacteria when compared to younger adults (<64 years of age) [66]. An increase in bacteria in older adults may account for the increase observed in lysozyme due to lysozyme providing antimicrobial activity. The lysine peak found at 1000 cm^{-1} also increased in intensity as individuals aged. The peak was stronger for 56+ age group in both males and females. A study by Tanaka et al. observed similar results with significantly increased lysine levels ($p < 0.05$) with ageing regardless of gender [67].

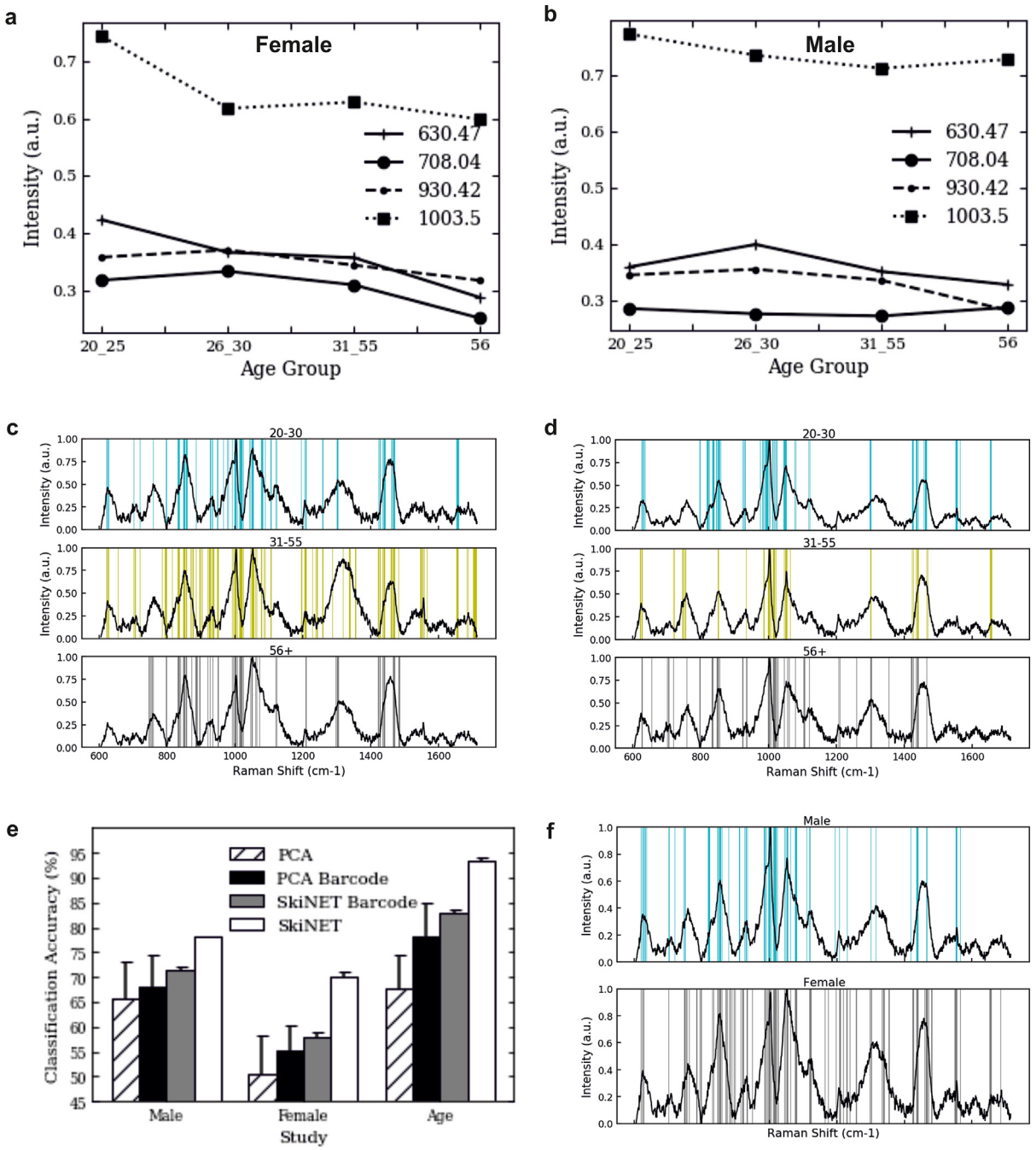


Fig. 3. Variation in Raman intensity with age for **(a)** female and **(b)** male saliva. Barcodes derived from Raman spectra of saliva for age groups of 20–30 y, 31-55 y and >56 y for **(c)** females and (**d**) males. **(e)** Comparison of PCA and SKiNET based classification of Raman profiling of saliva according to gender and age. **(f)** Barcoding analysis of male and female saliva for overall age groups.

Total protein concentration in saliva also varies with age [66]. Average protein concentration has been shown to decrease from 3.25 mg/ml for 20–30 years young adult to 2.71 mg/ml for >56 years adult groups [67]. Furthermore, females exhibited a considerably higher total protein concentrations than males, of 3.06 and 2.60 mg/ml, respectively [68]. Mucins (observed in the amide I region) are also more readily detected in young adults (20–30 years) than older adults (>56 years) [69]. Lower molecular weight proteins such as, the Statherin, Histatin (997 cm^{-1}) and Cystin (678 cm^{-1}) are also detected in young adults but not in older adults' saliva. Peroxidases, proline rich proteins (856 and 920 cm^{-1}) and IgA (1480 cm^{-1}) Raman bands are present in saliva across all age groups however, the concentrations differ with age as evidenced in Fig. 2d and f. This indicates that there is an *age-dependant* influence on the specific secretion of one component but at other bands (930 and 1200 cm^{-1}) where a lower intensity is observed, the concentration effect is due to the drop in salivary output. Moreover, the higher peak intensities observed at 855, 1003 and 1449 cm^{-1} and an additional peak at 1550 cm^{-1} in older females (56+ age group), indicate the presence of different proteins, and age-related differences in the female parotid gland. This is in agreement with a study by Ambatipudi et al. who quantitatively analysed the variation and

abundance of salivary proteins [70]. The study identified 338 proteins from the 20–30 years age group and 460 from the 55–65 years age group. The authors further classified each protein with the largest percentage, 40, belonging to the immune function group.

Subsequently, spectral barcodes were generated to represent the Raman bands of the highest intensity and most significance, with spectral differences compared for age and gender. The second derivative of each spectrum was taken, with absolute values above 40% of the maximum peak assigned a value of 1. The cut-off threshold of 40% was selected from empirical testing, where bars overlaid below the threshold resulted in an overly crowded spectrum. The overcrowding is due to the noise in the derivative spectrum and the need for a higher cut-off to allow only the most significant peaks to be identified. The balance of selecting the threshold value needs to be tested for each study, accordingly [71–73]. Barcodes derived from the mean Raman spectra for male and female saliva across all age groups are shown in Fig. 3f and the barcodes derived for the specific age groups are shown in Fig. 3c and d. These are found to be in an agreement with the SOMDI spectral features of the most significant classifications shown in Fig. 2.

While considering the barcoded spectra for gender classification, an overlap of the male and female spectral features can be seen with the most notable differences within the female barcodes is the appearance of more bands in the amide I and III regions as well as in the 800-860 cm^{-1} region. The high frequency of these bands correlates with the higher classification significance shown by SOMDI in Fig. 2h, right. In the characteristic male barcode, the increased bands frequency found in the 900-1000 cm^{-1} region is also found in the corresponding SOMDI. Similarly, for the classification according to the age groups, the frequency of spectral bands in Fig. 3, c-d is in correspondence to the Raman shift values in SOMDI (Fig. 2 d-f, right). Therefore, these barcodes can be used to identify both the gender and age of the subject according to the unique salivary spectral profile.

We have further compared the classification accuracy of the Raman spectra based on standard and barcode-based approaches using SKiNET and PCA (Fig. 3e). Standard spectral classification *via* the artificial neural network SKiNET shows the highest accuracy for both gender and age classifications. In contrast, standard Principal Component Analysis (PCA) was found to exhibit the lowest accuracy, with a classification decrease of 27% for gender, 13% for male age groups and 26% for female age groups relative to SKiNET. A more enhanced classification is found when PCA is utilised in conjunction with the barcoded spectral approach. It is possible that the performance increase is due to the PCA classifying fewer complex spectra since the barcoded spectra pre-identifies the significant spectral features. SKiNET's inherent ability to classify complex data with a higher accuracy whilst embracing and making use of the noisy data, enables it to achieve an overall classification accuracy of 93% for gender discrimination, specifically identifying males with 78% and females with 70% accuracy according to their ages.

We have further studied the temporal profiles of saliva over a period of 7-days (Fig. 4a). Saliva (males, 20–30 y) was stored at room temperature in a non-airtight microslide storage box during the study. The stability (or variability) of saliva as a function of time is an important factor in its assessment for suitability to act as a potential diagnostic biofluid and could find further uses in forensic applications, where hours or days may pass before the sample is acquired and analysed. The spectroscopic variation of the identified peaks at 630, 1003, 1051, 1458 cm^{-1} responsible for gender and age classification is shown in Fig. 4c–f. The *C–C* stretching around the 1125 cm^{-1} and in the Amide III, region exhibit the largest variance across the whole-time range (Fig. 4b). The largest variation as a function of time is found at 855 cm^{-1}, followed by the 1458 cm^{-1} peaks. These peaks are attributed to ring breathing mode tyrosine and CH_2 and CH_3 deformation vibrations (proteins and lipids). Tyrosine -rich peptide statherin is a potent inhibitor of calcium phosphate precipitation. Human saliva is usually supersaturated with calcium phosphate salts that form the dental enamel. Therefore, tyrosine alongside proline rich proteins work to inhibit precipitation from calcium phosphate solutions. These inhibitors adsorb from the saliva onto surfaces such as, the tooth enamel and thus, work to recalcify the enamel and prevent the formation of mineral deposits on the tooth.

Numerous components of saliva have been studied over time under different conditions. At room temperature, analytes such as alpha amylase, cholinesterase, aspartate aminotransferase and catalase are all stable. However, lipase, creatine kinase and lactate dehydrogenase are considerably less stable after 2–3 days [74]. Salivary enzymes show the greatest degree of variation at room temperature however, the reason for these enzymatic activity changes is largely unknown and requires further study. Protein breakdown is expected to occur over the first 4 h after sampling and therefore, this is where we anticipate seeing the majority of protein variation. This was observed in Fig. 4 c, d and f, where the lowest saliva intensity was detected prior to stabilising over the proceeding days. Each of the three peaks are attributed to proteins. Esser *et* al. also shown that immediately after sampling, a large variation between the salivary contents of each patient's salivary profile was observed [74]. It is possible, as opposed to a direct change in the analyte under investigation over a period of time, the identified changes are due to variation within the volunteer groups saliva. Variation could also be attributed to biological variations or differing observable degrees of protein degradation between individuals. The concentration of protein/enzymes within a patient's saliva can also vary the length of time required to cause any observable breakdown within the saliva sample thus, leading to variation between the samples. After 4 h, most peptides decrease in abundance since during this time period they are further degraded into single amino acids. This could be observed on the spectroscopic time series as an initial dip in intensity before the increase when salivary stabilisation occurs. Overall, these results indicate that saliva appears to be stable over a time of 1 week with no significant changes in its composition ($p > 0.05$), making it a highly promising biofluid, which can maintain diagnostic significance over prolonged time. These results are in agreement with electrophoresis and mass spectroscopy observations by Chevalier et al., who studied saliva over 30 days [75], and by Kim et al., who used PCR analysis on saliva DNA found that unstimulated saliva remained unchanged for four weeks after initial room temperature storage. However, after 4 weeks there was a gradual reduction in its content [76].

3. Conclusions

In this study, Raman spectroscopy has been established as an efficient technique for profiling the molecular composition of saliva. Gender and age differentiation have been enabled from the spectral fingerprints, identifiable from the representative chronologically defined age groups of young (20–30 years), mid-aged (31–55 years) and older (>56 years) adults, establishing unique *molecular barcodes* for females and males as well as each age group. Saliva also showed no significant change in its composition with time over a period of seven days. This is of a further importance if saliva is to be employed for either diagnostics or forensic purposes and might be beneficial over other human biofluids such as blood or the cerebrospinal fluid, which are known to rapidly and

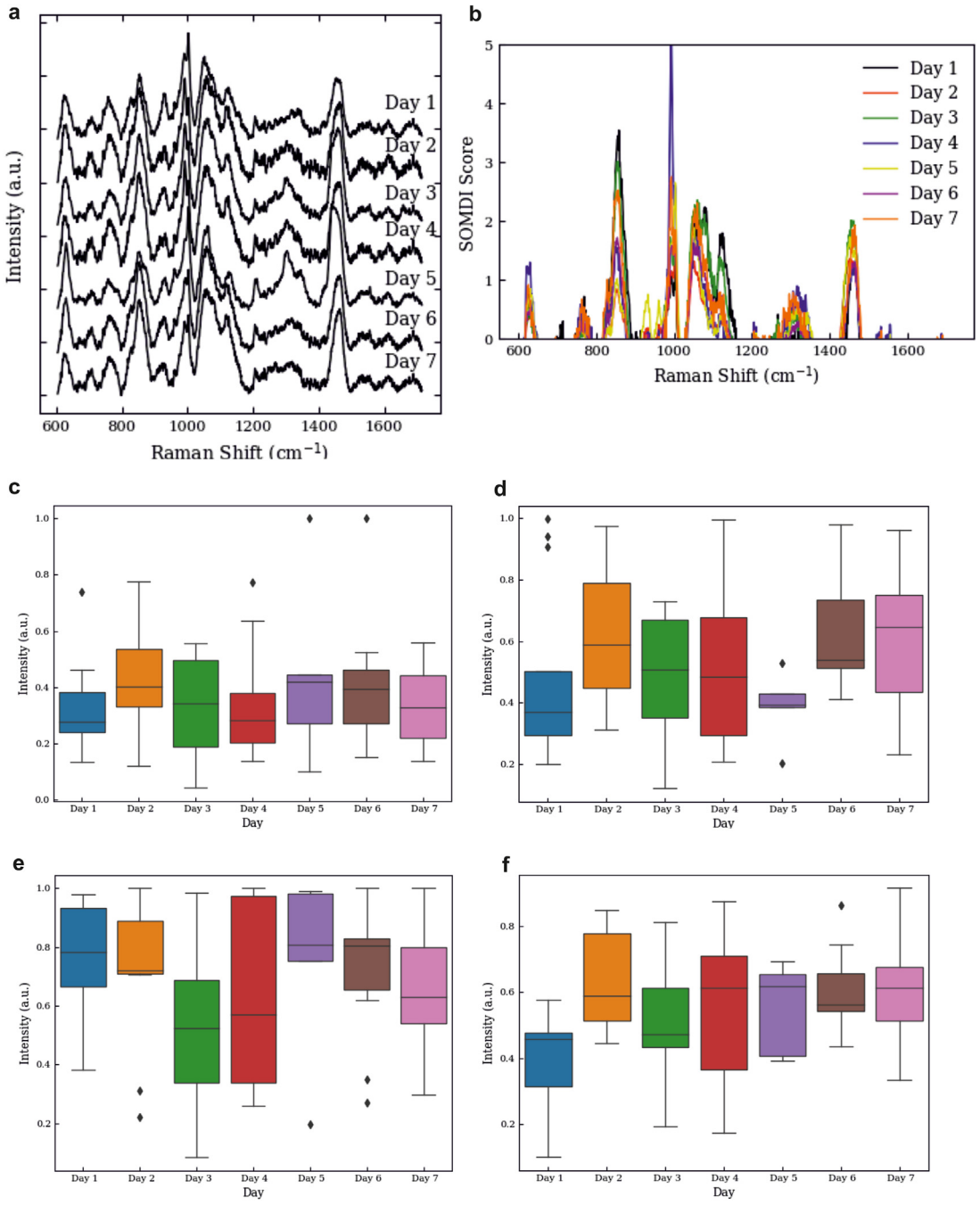


Fig. 4. Spectroscopic time series profiles of saliva between the ages of 23–27 ($n = 10$), with **(a)** mean spectra measured over 7 days and **(b)** representative Raman peaks identified through SOMDI SKiNET analysis. **(c)** Box and whisker plots represent the minima, maxima, interquartile ranges, whiskers and the median in saliva intensity at **(c)** 630 cm^{-1}, **(d)** 855 cm^{-1}, **(e)** 1003 cm^{-1} and **(f)** 1458 cm^{-1} peaks.

significantly change their properties from the time of excretion. The classification of the molecular spectroscopic profiles has been enhanced using our SKiNET algorithm with gender identification accuracy of 93% and average age sub-classification of 75%.

These results, addressing the underpinning molecular composition of healthy saliva, by considering changes due to gender, age, and time, allow the determination of sex and age of individuals. The identified unique molecular properties of the saliva combined with the remarkable progress in developments of Raman instrumentation, which has made great strides forward in the last few decades in terms of miniaturisation of sensing devices, lays the basis for the design of novel assays for on-site forensic and diagnostic point-of-need applications. Furthermore, the non-destructive nature of Raman spectroscopy, enables preservation of evidence and the subsequent diagnostic analyses of the salivary biofluid by additional techniques. However, with the obvious advantages for forensic applications in generating a primary victim or criminal profile, for diagnostic purposes, our results highlight the importance of the care that must be taken in designing and developing diagnostic assays. For efficient and accurate diagnostic significance of disease detection, the interpretation of Raman spectral differences, either in the form of peak shifts, broadening of peaks, change

in intensities or appearance/disappearance of bands, should consider the inherent spectral differences arising from the variation in gender and age of the patients. The further detailed interplay of the effects on saliva composition arising from smoking, BMI variation or different conditions such as morning/evening, moods, before and after meals are subject to further ongoing studies (Fig. S2) assessing the suitability of Raman spectroscopy as a potential physiological monitoring tool. Overall, Raman based diagnosis of the various disease pathologies *via* saliva detection is emerging as a promising tool for rapid point-of-care diagnostics. With the addition of the identified molecular profiles, a simple process of a '*background-like subtraction*' of the healthy salivary barcodes will ensure the variation due to the factors such as sex, age, and further, therapeutic treatments and various underlying conditions, will be taken into the account and reduce the error in the data interpretation of disease associated changes, leading towards improved diagnostic accuracy. Ultimate integration of Raman molecular fingerprinting as a powerful tool combined with emerging artificial intelligence techniques, such as our recently developed SkiNET method, will further provide important *real-time*, rapid spectroscopic saliva measurements for *on-site* forensic analysis and could also enable establishing insights into biological pathways underlying the associated diseases pathophysiology and conceivably allow tracking the passage and dosage of pharmacological therapeutics. There are ongoing efforts in the field for translation into *in-vivo* measurements, and diagnostics.

4. Materials and methods

Saliva Collection and Sample Preparation. Saliva has been collected from 70 healthy consenting volunteers (32 males and 38 females) with no previous health issues or chronic conditions. Samples were collected from healthy individuals only. All individuals recruited were subject to two-tier screening by the research team at the University Hospital Birmingham. The ethical framework for the laboratory studies for analysis of saliva samples from healthy volunteers using Raman Spectroscopy to explore age related variations comprises the Red Diamond ethics (ERN_18−2050) part of the NIHR portfolio study. Tobacco and alcohol usage were noted as factors preventing participants from donating saliva. The presence of pre-established pathology or chronic pathology (such as periodontitis, diabetes or cancer) was specific exclusion, and if this was identified or suspected in the recruited individual, they were excluded, and samples were dissuaded. Participants did not consume any food or beverages 45 min prior to each sample collection. The passive drool method was used to collect approximately 2 ml of unstimulated saliva (sampling time: 5 min) in a 50 ml conical tube (Thermo Fisher, UK). The method allowed for direct collection of saliva without need of parafilm or a citrus liquid droplet to promote stimulated salivary emission. An unstimulated saliva was used herein since the previous studies using the stimulated saliva and chewing of parafilm are known to change salivary composition [77,78]. Stimulated saliva has shown to contain lower quantities of protein and has a lower viscosity than unstimulated saliva [79,80]. 1 ml of the collected saliva was transferred into a 1.5 ml micro-centrifuge (Thermo Fisher, UK) tube and centrifuged for 30 min at 10,000×*g* to remove the larger debris. After centrifugation, 5 μl of the collected supernatant was pipetted onto a microscope glass slide coated with aluminium foil with subsequent Raman measurements carried out immediately after [81].

Raman Spectroscopy. Raman spectra were acquired *via* a Renishaw InVia Qontor confocal Raman microscope (Renishaw PLC, UK) equipped with a microscope Leica DMLM with a 50× objective, 785 nm excitation laser with an output power at the sample of 30−50 mW. Raman maps over a 100 μm × 100 μm grid were acquired for each sample with (100 spectra) using a step size of 1.5 μm between points and an acquisition time of 10 s per spectra.

Data Acquisition and Analysis. WiRE 5.1 (Renishaw PLC, UK) was used to acquire the data and for the polynomial (5th order) background subtraction. Next the spectra were normalised using the standard normal variate (SNV) and subsequently, screened to remove cosmic ray peaks using a custom Python script (Python 3.7) [40]. The calcium spectra were identified by the presence of two dominant peaks at 620 and 1002 cm^{-1}. Following this, sets of 10 spectra from different map locations were averaged to improve the SNR. Boxplots and *p*-value calculations were generated and calculated using Python. The Student's *t*-test was used to determine p-values. $p < 0.05$ was considered significant. Multivariate analysis was performed using SKiNET, an open source analysis tool [41], with the accompanying Raman Toolkit web interface to build SOM models using training data and perform predictions against test data. SKiNET models were optimized by performing 10-fold cross validation on the training data, and tuning the number of neurons, initial learning rate and number of training steps. The final model used a 20 × 20 grid of neurons, 57,600 training steps (5 epochs of the data), with an initial learning rate of 0.2. The initial neighbourhood size was maintained at 2/3 the edge length of the grid and cosine similarity used as the distance metric to determine the best matching unit. The optimized model was subsequently used to classify the previously unused test data. Classification using the test data were repeated 10 times from separate SOM initializations. The number for male/female participants were 32/38, respectively. Female age groups 20−30/31-55/56+ participants were 16/11/11 and male age groups 20−30/31-55/56+ participants were 22/6/4, respectively. Given the imbalances in the number of samples per class to be analysed, resampling was used to prevent bias and improve model training. For both the SOM and PCA models Synthetic Minority Oversampling Technique (SMOTE) was utilised to oversample the smaller minority sample classes but not the largest class (Imbalearn version 0.80). Using this approach additional new data is created to form a balanced model for analysis and classification. To achieve higher accuracy the SOM size, learning rate and number of epochs was empirically tested, with classification accuracy determined using a 10-fold cross validation. Stability of the model was further verified by running a repeat initialisation of the classification five times. Barcoding was completed in the custom Python script, with barcodes generated using the Savitzky-Golay filter to calculate second derivative of each spectra, with a smoothing window of 21 and polynomial order of 2. The absolute values over 40% of the maximum peak height were assigned a value of 1, with values below 40% assigned a value of 0. These were overlaid on the mean spectra referencing the underpinning peak positions. Principal component analysis (PCA) was carried out in Python using the Sklearn Decomposition package, with pre-processed spectra split into a testing and training group (split of 75% training and 25% test). The model was initially trained and then tested to assess the classification accuracy with the classification repeated five times to obtain the error values.

CRediT authorship contribution statement

Emma Buchan: Conceptualization, Investigation, Methodology, Formal analysis, Writing − original draft. **Liam Kelleher:** Writing − original draft, Investigation, SkiNET algorithm data classification. **Michael Clancy:** Methodology, Formal analysis, Software. **Jonathan James Stanley Rickard:** Validation, Writing − original draft, Writing − review & editing, Data curation. **Pola Goldberg Oppenheimer:** Conceptualization, Methodology, Writing − review & editing, Resources, Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge funding from the Wellcome Trust (174ISSFPP and 213458/Z/18/Z), the Royal Academy of Engineering (RF1415\14\28) and the EPSRC. P.G.O. is a Royal Academy of Engineering Research Fellowship holder.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2021.339074.

References

[1] G. Bahar, R. Feinmesser, T. Shpitzer, A. Popovtzer, R.M. Nagler, Salivary analysis in oral cancer patients: DNA and protein oxidation, reactive nitrogen species, and antioxidant profile, Cancer 109 (1) (2007) 54–59.
[2] S. Sindhu, N. Jagannathan, Saliva: a cutting edge in diagnostic procedures, J. Oral Dis. (2014) 1–8.
[3] I. Pence, A. Mahadevan-Jansen, Clinical instrumentation and applications of Raman spectroscopy, Chem. Soc. Rev. 45 (7) (2016) 1958–1979.
[4] J.M. Connolly, K. Davies, A. Kazakeviciute, A.M. Wheatley, P. Dockery, I. Keogh, M. Olivo, Non-invasive and label-free detection of oral squamous cell carcinoma using saliva surface-enhanced Raman spectroscopy and multivariate analysis, Nanomed. Nanotechnol. Biol. Med. 12 (6) (2016) 1593–1601.
[5] Y. Chen, S. Cheng, A. Zhang, J. Song, J. Chang, K. Wang, Y. Zhang, S. Li, H. Liu, G. Alfranca, M.A. Aslam, B. Chu, C. Wang, F. Pan, L. Ma, J.M. de la Fuente, J. Ni, D. Cui, Salivary analysis based on surface enhanced Raman scattering sensors distinguishes early and advanced gastric cancer patients from healthy persons, J. *Biomed. Nanotechnol.* 14 (10) (2018) 1773–1784.
[6] F. Zapata, Á. Fernández De La Ossa, C. García-Ruiz, Emerging spectrometric techniques for the forensic analysis of body fluids, Trends Anal. Chem. 64 (2015) 53–63.
[7] C.K. Muro, K.C. Doty, L. De, S. Fernandes, I.K. Lednev, Forensic body fluid identification and differentiation by Raman spectroscopy, Forensic Chem. 1 (2016) 31–38.
[8] X. Lin, D. Lin, X. Ge, S. Qiu, S. Feng, R. Chen, Noninvasive detection of nasopharyngeal carcinoma based on saliva proteins using surface-enhanced Raman spectroscopy, J. Biomed. Opt. 22 (10) (2017) 1–6.
[9] S. Qiu, Y. Xu, L. Huang, W. Zheng, C. Huang, S. Huang, J. Lin, D. Lin, S. Feng, R. Chen, J. Pan, Non-invasive detection of nasopharyngeal carcinoma using saliva surface-enhanced Raman spectroscopy, Oncol. Lett. 11 (1) (2016) 884–890.
[10] S. Feng, S. Huang, D. Lin, G. Chen, Y. Xu, Y. Li, Z. Huang, J. Pan, R. Chen, H. Zeng, Surface-enhanced Raman spectroscopy of saliva proteins for the noninvasive differentiation of benign and malignant breast tumors, Int. J. Nanomed. 10 (2015) 537–547.
[11] J.J. Zermeño-Nava, M.U. Martinez-Martinez, A.L. Ramirez-de-Avila, A.C. Hernandez-Arteaga, M.G. Garcia-Valdivieso, A. Hernandez-Cedillo, M. Jose-Yacaman, H.R. Navarro-Contreras, Determination of sialic acid in saliva by means of surface-enhanced Raman spectroscopy as a marker in adnexal mass patients: ovarian cancer vs benign cases, J. Ovarian Res. 11 (61) (2018) 1–9.
[12] P. Seredin, D. Goloshchapov, Y. Ippolitov, P. Vongsvivut, Pathology-specific molecular profiles of saliva in patients with multiple dental caries—potential application for predictive, preventive and personalised medical services, EPMA J. 9 (2) (2018) 195–203.
[13] A. Hernández-Cedillo, G. Garcia-Valdivieso, A.C. Hernandez-Arteaga, N. Patino-Marin, A.A. Vertiz-Hernandez, M. Jose-Yacaman, H.R. Navarro-Contreras, Determination of sialic acid levels by using surface-enhanced Raman spectroscopy in periodontitis and gingivitis, Oral Dis. 25 (6) (2019) 1627–1633.
[14] V. Sikirzhytski, A. Sikirzhytskaya, I.K. Lednev, Multidimensional Raman spectroscopic signature of sweat and its potential application to forensic body fluid identification, Anal. Chim. Acta 718 (2012) 78–83.
[15] K. Virkler, I.K. Lednev, Forensic body fluid identification: the Raman spectroscopic signature of saliva, Analyst 135 (2010) 512–517.
[16] K. Virkler, I.K. Lednev, Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene, Forensic Sci. Int. 188 (2009) 1–17.
[17] G. Cao, M. Chen, Y. Chen, Z. Huang, J. Lin, J. Lin, Z. Xu, S. Wu, W. Huang, G. Weng, G. Chen, A potential method for non-invasive acute myocardial infarction detection based on saliva Raman spectroscopy and multivariate analysis, Laser Phys. Lett. 12 (12) (2015).
[18] C. Carlomagno, P.I. Banfi, A. Gualerzi, S. Picciolini, E. Volpato, M. Meloni, A. Lax, E. Colomvo, N. Ticozzi, F. Verde, V. Silani, M. Bedoni, Human salivary Raman fingerprint as biomarker for the diagnosis of Amyotrophic Lateral Sclerosis, Sci. Rep. 10 (1) (2020) 1–13.
[19] G. Eom, A. Hwang, H. Kim, S. Yang, D.K. Lee, S. Song, K. Ha, J. Jeong, J. Jung, E. Lim, T. Kang, Diagnosis of tamiflu-resistant influenza virus in human nasal fluid and saliva using surface-enhanced Raman scattering, ACS Sens. 4 (9) (2019) 2282–2287.
[20] S. Hu, Y. Gao, Y. Wu, X. Guo, Y. Ying, Y. Wen, H. Yang, Raman tracking the activity of urease in saliva for healthcare, Biosens. Bioelectron. J. 129 (2019) 24–28.
[21] A. Falamas, C.I. Faur, M. Baciut, H. Rotaru, M. Chirila, S. Cinta Pinzaru, M. Hedesiu, Raman spectroscopic characterization of saliva for the discrimination of oral squamous cell carcinoma, Anal. Lett. (1–13) (2020).
[22] X. Li, T. Yang, J. Lin, Spectral analysis of human saliva for detection of lung cancer using surface-enhanced Raman spectroscopy, J. Biomed. Opt. 17 (3) (2012) 1–5.
[23] V. Sikirzhytski, A. Sikirzhytskaya, I.K. Lednev, Multidimensional Raman spectroscopic signatures as a tool for forensic identification of body fluid traces: a review, Appl. Spectrosc. 65 (11) (2011) 1223–1232.
[24] K. Virkler, I.K. Lednev, Raman spectroscopy offers great potential for the nondestructive confirmatory identification of body fluids, Forensic Sci. Int. J. 181 (2008), 1–5.
[25] D. Nilendu, A. Kundu, A. Chand, A. Johnson, Forensic implications of saliva: an overview, Indian J. Forensic Med. Toxicol. 14 (1) (2020) 189–194.
[26] S. Pandit, D. Desai, P. Jeergal, S. Venkatesh, Awareness of forensic odontology among police personnel: a new ray of hope in forensic odontology, J. Forensic Dent. Sci. 8 (1) (2016) 1–6.
[27] S. Farquharson, C. Shende, A. Sengupta, H. Huang, F. Inscore, Rapid detection and identification of overdose drugs in saliva by surface-enhanced Raman scattering using fused gold colloids, Pharmaceutics 3 (3) (2011) 425–439.
[28] V. Sikirzhytski, K. Virkler, I.K. Lednev, Discriminant analysis of Raman spectra for body fluid identification for forensic purposes, Sensors 10 (4) (2010) 2869–2884.
[29] C. Shende, F. Inscore, P. Maksymiuk, S. Farquharson, Five minute analysis of chemotherapy drugs in saliva, Opt. Methods Life Sci. 6386 (2006) 1–6.
[30] A.D. Gift, C.S. Shende, F.E. Inscore, S. Farquharson, Analysis of chemotherapy drug 5- fluorouracil and its metabolites by surface-enhanced Raman spectroscopy, Smart Med. Biomed. Sens. Technol. II 5588 (December 2004) 70–77, 2004.
[31] S. Farquharson, A.D. Gift, C. Shende, P. Maksymiuk, F.E. Inscore, J. Murran, Detection of 5-fluorouracil in saliva using surface-enhanced Raman spectroscopy, Vib. Spectrosc. 38 (2005) 79–84.
[32] S. Han, L.A. Oaks, A.K. Locke, Y.-S.L. Cheng, G.L. Coté, Development of a free-solution SERS-based assay for point-of-care oral cancer biomarker detection using DNA-conjugated gold nanoparticles, in: SPIE 10501, 2018, pp. 1–6.
[33] P. Rekha, P. Aruna, E. Brindha, D. Koteeswaran, M. Baludavid, S. Ganesan, Near-infrared Raman spectroscopic characterization of salivary metabolites in the discrimination of normal from oral premalignant and malignant conditions, J. Raman Spectrosc. 47 (7) (2016) 763–772.
[34] R. Condò, L. Cerroni, G. Pasquantonio, M. Mancini, A. Pecora, A. Convertino, V. Mussi, A. Rinaldi, L. Maiolo, A deep morphological characterization and comparison of different dental restorative materials, BioMed Res. Int. (–16) (2017) 1.
[35] D.E. Romonti, A.V. Gomez Sanchez, I. Milošev, I. Demetrescu, S. Ceré, Effect of anodization on the surface characteristics and electrochemical behaviour of zirconium in artificial saliva, Mater. Sci. Eng. C 62 (2016) 458–466.
[36] I.J. Jahn, O. Zukovskaja, X.S. Zheng, K. Weber, T.W. Bocklitz, D. Cialla-May, J. Popp, Surface-enhanced Raman spectroscopy and microfluidic platforms: challenges, solutions and potential applications, Analyst 142 (7) (2017) 1022–1047.
[37] (a) C. Andreou, M.R. Hoonejani, M.R. Barmi, M. Moskovits, C.D. Meinhart, Rapid detection of drugs of abuse in saliva using surface enhanced Raman spectroscopy and microfluidics, ACS Nano 7 (8) (2013) 7157–7164;
(b) P.D. Gomes, J.J. Rickard, P. Goldberg Oppenheimer, Electrofluidynamic patterning of tailorable nanostructured substrates for surface-enhanced Raman scattering, ACS Appl. Nano Mater. 3 (7) (2020) 6774–6784.
[38] C.K. Muro, L. De, S. Fernandes, I.K. Lednev, Sex determination based on Raman spectroscopy of saliva traces for forensic purposes, Anal. Chem. 88 (2016) 12489–12593.
[39] C. Banbury, R. Mason, I. Styles, N. Eisenstein, M. Clancy, A. Belli, A. Logan, P.G. Oppenheimer, Development of the self optimising kohonen index network (SKiNET) for Raman spectroscopy based detection of anatomical eye tissue, Sci. Rep. 9 (2019) 1–9.
[40] L. Kelleher, Raman spectrscopy barcode and PCA analysis scripts [Online]. Available: https://github.com/Liam-Kelleher123/Raman_code_snips, 2020.
[41] C. Banbury, Raman Toolkit - analysis and data management tool for Raman spectra [Online]. Available: https://github.com/cbanbury/raman-tools, 2018.
[42] H.J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N.J. Fullwood, B. Gardner, P.L. Martin-Hirsch, M.J. Walsh, M.R. McAinsh, N. Stone, F.L. Martin, Using Raman spectroscopy to characterize biological materials, Nat. Protoc. 11 (4) (2016) 664–687.
[43] I. Takeda, C. Stretch, P. Barnaby, K. Bhatnager, K. Rankin, H. Fu, A. Weljie, N. Jha, C. Slupsky, Understanding the human salivary metabolome, NMR

Biomed. 22 (6) (2009) 577–584.
[44] I. Reddy, H. Sherlin, P. Ramani, P. Premkummar, A. Natesan, T. Chandrasekar, Amino acid profile of saliva from patients with oral squamous cell carcinoma using high performance liquid chromatography, J. Oral Sci. 54 (3) (2012) 279–283.
[45] H. Inoue, K. Ono, W. Masuda, Y. Morimoto, T. Tanaka, M. Yokota, K. Inenaga, Gender difference in unstimulated whole saliva flow rate and salivary gland sizes, Arch. Oral Biol. 51 (12) (2006) 1055–1060.
[46] A. Cydejko, A. Kusiak, M.E. Grzybowska, B. Kochanska, J. Ochocinska, A. Maj, D. Sweietlik, Selected physicochemical properties of saliva in menopausal women-a pilot study, Int. J. Environ. Res. Publ. Health 17 (7) (2020) 1–8.
[47] W. Li-Hui, L. Chuan-Quan, Y. Long, L. Ru-Liu, C. Long-Hui, C. Wei-Wen, Gender differences in the saliva of young healthy subjects before and after citric acid stimulation, Clin. Chim. Acta 460 (2016) 142–145.
[48] J.N. Rukmini, R. Sachan, N. Sibi, A. Meghana, C. Malar, Effect of menopause on saliva and dental health, J. Int. Soc. Prev. Community Dent. 8 (6) (2018) 529–533.
[49] D.R. Mahesh, Evaluation of salivary flow rate, pH and buffer in pre, post & post-menopausal women on HRT, J. Clin. Diagn. Res. 8 (2014) 233–236.
[50] National Health Service - Menopause, February 2021. Last Updated August 2018.
[51] C.F. Streckfus, Advances in Salivary Diagnostics, Springer-Verlag, Berlin Heidelberg, 2015, p. 1204.
[52] K.E. Kaczor-Urbanowicz, C.M. Carreras-Presas, K. Aro, M. Tu, F. Garcia-Godoy, D.T.W. Wong, Saliva diagnostics, Curr. Rev. Direct. 242 (5) (2017) 459–472.
[53] A. Chandra Sekhar Talari, Z. Movasaghi, S. Rehman, I. ur Rehman, Raman spectroscopy of biological tissues, Raman Spectrosc. Biol. Tissues 50 (1) (2015) 46–111.
[54] S. Derruau, J. Robinet, V. Untereiner, O. Piot, G.D. Sockalingum, S. Lorimier, Vibrational Spectroscopy saliva profiling as a biometric tool for disease diagnostics: a systematic literature review, Molecules 25 (18) (2020) 4142.
[55] R. Leimola-Virtanen, H. Helenius, M. Laine, Hormone replacement therapy and some salivary antimicrobial factors in post- and perimenopausal women, Maturitas 27 (2) (1997) 145–151.
[56] M. Hernández Ríos, T. Sorsa, F. Obregón, T. Tervahartiala, M. Antonieta Valenzuela, P. Pozo, N. Dutzan, E. Lesaffre, M. Molas, J. Gamonal, "Proteolytic roles of matrix metalloproteinase (MMP)-13 during progression of chronic periodontitis: initial evidence for MMP-13/MMP-9 activation cascade," J. Clin. Periodontol., vol. 36, no. 12, pp. 1011-1017.
[57] F. Xu, L. Laguna, A. Sarkar, Aging-related changes in quantity and quality of saliva: where do we stand in our understanding? J. Texture Stud. 50 (1) (2019) 27–35.
[58] R.M. Nagler, Salivary glands and the aging process: mechanistic aspects, health-status and medicinal-efficacy monitoring, Biogerontology 5 (2004) 223–233.
[59] C.K. Yeh, D.A. Johnson, M.W.J. Dodds, Impact of aging on human salivary gland function: a community-based study, Aging Clin. Exp. Res. 10 (1998) 421–428.
[60] R.H. Affoo, N. Foley, R. Garrick, W.L. Siqueira, R.E. Martin, Meta-analysis of salivary flow rates in young and older adults, J. Am. Geriatr. Soc. 63 (10) (2015) 2142–2151.
[61] S. Sun, F. Zhao, Q. Wang, Y. Zhong, T. Cai, P. We, Analysis of age and gender associated n-glycoproteome in human whole saliva, Clin. Proteonomics 11 (1) (2014) 1–10.
[62] H. Ben-Aryeh, M. Fisher, R. Szargel, D. Laufer, Composition of whole unstimulated saliva of healthy children: changes with age, Arch. Oral Biol. 35 (11) (1990) 929–931.
[63] I.S. Yim, D.A. Granger, J.A. Quas, "Children's and adults' salivary alpha-amylase responses to a laboratory stressor and to verbal recall of the stressor, Dev. Psychobiol. 52 (6) (2010) 598–602.
[64] K. Rutherfurd-Markwick, C. Starck, D.K. Dulson, A. Ali, Salivary diagnostic markers in males and females during rest and exercise, J. Int. Soc. Sports Nutr. 14 (1) (2017) 1–8.
[65] G.H. Perry, N.J. Dominy, K.G. Claw, A.S. Lee, H. Fiegler, R. Redon, J. Werner, F.A. Vilanae, J.L. Mountain, R. Misra, N.P. Carter, C. Lee, A.C. Stone, Diet and the evolution of human amylase gene copy number variation, Nat. Genet. 39 (10) (2007) 1256–1260.
[66] R. Lira-Junior, S. Åkerman, B. Klinge, E.A. Boström, A. Gustafsson, Salivary microbial profiles in relation to age, periodontal, and systemic diseases, PloS One 13 (3) (2018) 1–14.
[67] S. Tanaka, M. Machino, S. Akita, Y. Yokote, H. Sakagami, Changes in salivary amino acid composition during aging, In Vivo (Brooklyn). 24 (6) (2010) 853–856.
[68] D. Bhuptani, S. Kumar, M. Vats, R. Sagav, Age and gender related changes of salivary total protein levels for forensic application, J. Forensic Odontostomatol. 36 (1) (2018) 26–33.
[69] W.I. Chang, J.Y. Chang, Y.Y. Kim, G. Lee, H.S. Kho, MUC1 expression in the oral mucosal epithelial cells of the elderly, Arch. Oral Biol. 56 (9) (2011) 885–890.
[70] K.S. Ambatipudi, B. Lu, F.K. Hagen, J.E. Melvin, J.R. Yates, Quantitative analysis of age specific variation in the abundance of human female parotid salivary proteins, J. Protein Chem. 8 (11) (2009) 5093–5102.
[71] J.J.S. Rickard, V. Di-Pietro, D.J. Smith, D.J. Davies, A. Belli, P.G. Oppenheimer, Rapid optofluidic detection of biomarkers for traumatic brain injury via surface-enhanced Raman spectroscopy, Nat. Biomed. Eng. 4 (2020) 610–623.
[72] V. Mlinar, A. Zunger, Spectral barcoding of quantum dots: deciphering structural motifs from the excitonic spectra, Phys. Rev. B Condens. Matter 80 (3) (2009) 1–7.
[73] I.S. Patel, W.R. Premasiri, D.T. Moir, L.D. Ziegler, Barcoding bacterial cells: a SERS-based methodology for pathogen identification, J. Raman Spectrosc. 39 (11) (2008) 1660–1672.
[74] D. Esser, G. Alvarez-Llamas, M.P. de Vries, D. Weening, R.J. Vonk, H. Roelofsen, Sample stability and protein composition of saliva: implications for its use as a diagnostic fluid, Biomark. Insights 3 (2008) 25–27.
[75] F. Chevalier, C. Hirtz, S. Chay, F. Cuisinier, N. Sommerer, M. Rossignol, D. Deville de Périère, Proteomic studies of saliva: a proposal for a standardized handling of clinical samples, Clin. Proteonomics 3 (1–4) (2007) 13–21.
[76] Y. Kim, Y. Kim, D. Ph, The Effects of Storage of Human Saliva on DNA Isolation and Stability, vol. 31, 2006, pp. 1–16, 1.
[77] E. Neyraud, T. Sayd, M. Morzel, E. Dransfield, Proteomic analysis of human whole and parotid salivas following stimulation by different tastes, J. Proteome Res. 5 (2006) 2474–2480.
[78] R. Schipper, A. Loof, J. de Groot, L. Harthoom, E. Dransfield, W. van Heerde, SELDI TOF-MS of saliva: methodology and pre-treatment effects, J. Chromatogr., B 15 (2007) 45–53.
[79] M. Edgar, C. Dawes, D. O'Mullane, Saliva and Oral Health, third ed., British Dental Association, London, 2004, pp. 6–7.
[80] S. Gomar-Vercher, A. Simón-Soro, J.M. Montiel-Company, J.M. Almerich-Silla, A. Mira, Stimulated and unstimulated saliva samples have significantly different bacterial profiles, PloS One 13 (2018) 1–12.
[81] I. Maitra, C.L.M. Morais, K.M.G. Lima, K.M. Ashton, R.S. Date, F.L. Martin, Raman spectral discrimination in human liquid biopsies of oesophageal transformation to adenocarcinoma, J. Biophot. 13 (3) (2019).