Methods for fungiform papillae assessment: a survey across European research units

[1] Maria Piochi, [2] Sara Spinelli, [3] Melania Melis, [3] Iole Tommassini Barbarossa, [1] Luisa Torri, [4] Monica Laureati, [4] Ella Pagliarini, [5] Emma Feeney, [6] Rebecca Ford, [6] Qian Yang, [7] Lisa Methven, [2] Erminio Monteleone, [2] Caterina Dinnella

[1] University of Gastronomic Sciences
[2] University of Florence
[3] University of Cagliari
[4] University of Milan
[5] University College Dublin
[6] University of Nottingham
[7] University of Reading

Abstract

Fungiform papillae (FP) are the anatomical structures dedicated to gustatory and somatosensory signal triggering. Whether the FP density (FPD=FP/cm^2) associates to the oral responsiveness is still controversial and could be partially due to diversity in methods used to quantify FP. The present study aims at mapping methods used to estimate FPD and at tentatively assessing the impact of main procedure variables in FPD estimation. An on-line survey was specifically developed and launched among European Sensory Science Society associates to collect information on methods and procedure variables and to share dataset on FP counting. Seven European research centres responded to the survey. Manual count resulted the most popular methodology and a merged dataset of 1006 observations was obtained. The type of device used for tongue picture acquisition (low-resolution, LR vs high-resolution, HR) resulted the main procedure variable. FPD mean values were lower when assessed by LR (35.8) than with HR devices (41.5). Characteristic distribution values of FPD for LR and HR datasets were similar. Procedure variables did not significantly affect FPD estimation obtained from HR devices, while both picture modification and data validation significantly affected FPD values collected with LR devices. Both HR and LR procedures resulted sensitive to differences in FPD due to demographic factors. Overall, measures from high resolution device appear the best option to depict the inter-individual variability in FPD.
Automated procedures are underutilized and an effort to widen the accessibility to the script/software, as well as the implementation of commercial versions is envisaged.

Keywords: oral responsiveness; manual count; device resolution; gender; age

Highlights:

- Methods for FPD estimation were mapped across 7 European research centres
- FPD were lower in estimation with low resolution- vs high resolution devices
- Overall, high resolution devices are suited to depict individual variability in FPD
- Automated procedures are still underutilized

1. Introduction

Fungiform papillae (FP) have been recently re-conceptualized as multimodal sensing organ responding to taste, tactile and thermal modalities (Mistretta & Bradley, 2021). The FP are a gustatory and somatosensory structure in which taste cells of the taste buds are in synaptic contact with sensory nerve endings (e.i. chorda timpani nerve). Additionally, taste buds in FP are surrounded by trigeminal endings, which contain mechanoreceptors in the somatosensory system and are responsible for the chemesthetic sensations (pungency and burning), temperature and food textural attributes (Engelen & Van Der Bilt, 2008).

Considerable individual variability has been reported in FP density (FPD) (from 0 to over 200 papillae/cm^2) (Piochi et al., 2018). Gender and age significantly affect FPD (Dinnella et al., 2018; Shahbake et al., 2005). Women generally show higher FPD mean value than men. FPD mean values significantly decreased with age, and the decline is more evident in men than in women (Dinnella et al., 2018).

Given the double innervation of FP, their accessibility on the anterior region of the tongue and their correlation with the density of taste buds, these anatomical structures are considered as one of the phenotypic markers of oral responsiveness.

Several studies showed a positive relationship between FPD and taste sensitivity, demonstrating that subjects with a higher FPD are more sensitive to taste stimuli (Miller & Reedy, 1990). However, studies on large population samples suggested a lack of straight association between the perception of prototypical taste solution intensity and FPD (Dinnella et al., 2018; Fischer et al., 2013). Conflicting associations of FPD with lingual tactile acuity have been also reported. Positive associations have been reported between FPD and the perception of food texture (Zhou et al., 2021), while other studies failed to find significant associations between FPD and food texture attributes, such as creaminess (Nachtsheim & Schlich, 2013).

Several studies have focused on the relationship between individual variation in oral responsiveness in association with PROP taster status (the ability to perceive bitterness of 6-n-propylthiouracil - PROP) (Tepper, 2008) and FPD. Several findings showed that PROP responsiveness positively associates with responsiveness to various oral stimuli. The higher density (Essick et al., 2003) and the increased functionality of FP due to gustin (trophic factor protein for growth and development of taste buds) gene polymorphism (Melis et al., 2013) in individuals highly responsive to PROP (Super

Taster) have been proposed as possible explanation to their general heightened oral responsiveness in comparison to other PROP taster groups (Non Taster and Medium Taster). However, other large-scale studies did not revealed a link between PROP taste responsiveness and FPD (Fischer et al., 2013) and or gustin gene polymorphism (Feeney & Hayes, 2014a).

Methods to quantify FP are all based on the inspection of tongue pictures, but a number of procedure variables are commonly adopted. Devices differing in image resolution capability are used; digital cameras and portable microscopes represent the most widely adopted ones. Tongue dyeing before picture acquisition is recommended to improve the visibility and recognizability of FP ((Nuessle et al., 2015), but procedure variants are reported in the type of dyeing agent and picture acquisition without dyeing is also documented. FP are unevenly distributed all over the anterior two-thirds of the tongue, with the highest density on the tongue tip, close to the midline. Restricted areas in the first 2 cm of the anterior tongue are generally considered for counting, however the positioning of the counting area (tip, side, distance from the tip), shape (circles, rectangles) and size can vary (from 6 to 1.6 cm^2) (Feeney & Hayes, 2014b; Masi et al., 2015). Furthermore, counting procedures on the whole tongue are also reported (Cattaneo et al., 2020; Sanyal et al., 2016).

In manual count method, the tongue picture is visually inspected by independent operators who identify and count FP. The picture modification in terms of magnification, brightness and saturation is often adopted to help FP identification. Specific operator training to FP identification and counting is recommended, the most widespread protocol is the Denver Papillae Protocol (Nuessle et al., 2015), that establishes a dichotomous criteria approach for the FP identification based on shape, size and elevation from the background of circular like elements visible on the picture. To improve the reliability of the measure, count validation is recommended to check for the degree of agreement across counters.

Automated count methods have been developed based on digital analysis of the tongue picture by specifically developed scripts and software (Cattaneo et al., 2020; Eldeghaidy et al., 2018; Piochi et al., 2017; Sanyal et al., 2016), but their prevalence beyond the self-implementers is still limited.

Inconsistency across finding on relationships between FP density and oral responsiveness may be at least partially due to the confounding effects of methodological variations in the procedure adopted for FPD assessment.

The present study aims at providing an updated picture of methods currently used for FPD assessment at European level, by describing the main procedure variables and exploring their possible effect on FP count. A survey was specifically developed and launched among European Sensory Science Society associates to collect information on methods and procedure variables and to share dataset on FP counting. Characteristics of sampled population (age, gender, smoking status) were also collected.

2. Material and methods

2.1 Survey and respondents

An online survey was run in spring 2022 on behalf of the European Sensory Science Society (E3S) network, though the coordination of the Working Group on Taste sensitivity of E3S. It included the following sections: equipment used to portray subjects' tongue (device and resolution); protocol adopted for the picture acquisition (tongue dyeing; type of colorant agent); region of the tongue, size and shape of the counting area; picture modification (magnification, brightness and saturation). Information on operators (number and training) and data validation (number of replicate and verification of the agreement among independent operators) were asked for manual count; software/script and eventual reference paper were asked in case of automated count.

Eight research units completed the survey and seven units agreed in sharing their dataset (Italy: University of Gastronomic Sciences IT1, University of Milan IT2, University of Florence IT3 and University of Cagliari IT4; Ireland: University College Dublin IR1; United Kingdom: University of Nottingham UK1 and University of Reading UK2). Inclusion criteria for the final dataset were to have available for each observation: gender, age (years old), raw count and computed FPD (expressed as FP/cm^2).

2.2 Data analysis

Descriptive statistics was applied to all variables. Mean values in the text are expressed with the standard error of the mean. The continuous variable age was converted into three classes: C1 (18-30 years old), C2 (31-45 y.o.) and C3 (>45 y.o.). Shapiro-Wilk test was applied to test the normality of the distribution FPD ($p<0.05$). Chi-squared tests were used to compare distributions for gender, age class, and smoking status in high-resolution and low-resolution datasets ($p<0.05$). One-way Analysis of Variance (ANOVA) models were applied to assess the effect of device resolution on FPD (fixed factor: device resolution, two levels HR and LR), separately evaluated in the whole population and in the C1 age class subset. The effect of procedure variables on FPD in C1 age class was assessed by one-way ANOVA models, independently run in HR and LR datasets. In HR C1 group, one-way ANOVA was applied to estimate the gender effect on FPD (fixed factor: gender). In LR sample, a 2-w ANOVA model with interaction was run to estimate the effects of gender and age on FPD (fixed factors: gender, age class, gender*age class).

All ANOVA model applied in the study were followed by Tukey HSD test ($p<0.05$). Analyses were conducted with XLSTAT Marketing (XLSTAT 2023.1.2, Addinsoft, New York, USA).

3. Results and Discussion

Final merged dataset included 1006 observations from FP manual count, contributed as follows by the different research units: 223 by IT1 (22%), 77 by IT2 (8%), 153 by IT3 (15%), 297 by IT4 (30%), 84 by IR1 (8%), 18 by UK1 (2%), 154 by UK2 (15%).

In the merged dataset, 59% observations were acquired from LR (digital microscope MicroCapture, version 2.0 for 20x-400x; Maplin Gadget Electronics, UK) and 41% from HR devices (Canon - model EOS D400 camera with lens EFS 55-250 mm; Canon, E05 700D with an EF-S 19–55 mm lens; Nikon 4DS). Main variables in procedures adopted to estimate FPD with HR and LR devices are shown in Table 1.

Table 1. Procedure variables and population characteristics of datasets from High- and Low-resolution devices.

		High-resolution (HR n=417)		Low-resolution (LR n=589)	
Variable	**Categories**	**n**	**%**	**n**	**%**
Tongue dye	No	18	4	0	0
	Yes	399	96	589	100
Dye agent	None	18	4	0	0
	Blue food dye (E133 or others)	315	76	589	100
	Methylene blue	84	20	0	0
Picture modification	No	333	80	136	23
	Yes	84	20	453	77
Area localization and dimension	One circle (0.6 cm diameter) left side middle	297	71	0	0
	Two circles (0.6 cm diameter) left and right 0.5 cm from tongue tip	102	24	453	77
	Two squares (1.0 cm^2 area) 0.5 cm from tongue tip	18	4	0	0
	One square 1.0 cm^2 2.0 cm from tongue tip	0	0	136	23
Training procedure	None	36	9	136	23
	The Denver Papillae Protocol	381	91	453	77
Independent operators	2	102	24	589	100
	3	315	76	0	0
Replicate	1	315	76	589	100
	2	102	24	0	0
Data validation	No	84	20	136	23
	Yes	333	80	453	77
Socio-demographic composition					
Gender	F	234	56	345	59
	M	183	44	244	41
Age class	C1	356	85	303	51
	C2	51	12	155	26
	C3	10	2	131	22
Smoking status	not smoker	284	68	492	84
	smoker	49	12	96	16
	*NA	84	20	1	0

Note: *NA= not available.

Tongue dyeing was adopted by the vast majority of respondents, mostly with blue food dye and in lower extent with Methylene blue. Picture modification (brightness and saturation) was largely applied in case of LR device use, while was much less used with HR device where the high quality of the image allows for FP identification on the original picture. FP are counted in small area (0.5-1.0 cm^2) on the tongue tip or in area close to the tongue tip, already proven to provide reliable measures of FPD and that highly correlate with the total FP number on the anterior tongue (Shahbake et al., 2005). In general, operators were trained adopting the Denver Papillae Protocol (DPP) in its original version (Nuessle et al., 2015) or with slight modifications, only a small percentage of data were collected by operators with no specific training (≤23%). Counting was performed by three or two operators and replicates per independent operator collected only in this latter case. Data validation was applied in the vast majority of observations (77-80%) and consisted in checking the agreement in FPD count across counters as indicated by the DPP or applying ANOVA models for checking for the absence of counter effect.

Automated count was only reported by four research units, each adopting different approaches, thus confirming the low penetration of this method and that its application is still restricted to the research team responsible for counting script/software development. The TongueSim software was developed and used by IR1 to count FP especially in small regions near the tongue tip (Sanyal et al., 2016). A Matlab script was applied by IT1 and IT3 to quantify FP with different diameters classes, in a blue stained rectangular area positioned at 0.5 from tongue tip (Piochi et al., 2017). UK1 used the method proposed by Eldeghaidy and colleagues (Eldeghaidy et al., 2018). based on a colour-based segmentation assisted by an algorithm. A high-performing computerized approach adopting a machine learning image processing method based on a convolutional neural network was adopted by IT2 (Cattaneo et al., 2020). Since no data were made available by research units, the comparison among different automated method was not possible.

In the whole population, the FPD from manual count ranged from 0.9 to 187.5, with a mean and median values of 38.1 and 33.6 FP/cm^2, respectively, and resulted well in line with the range of FPD variation reported in previous studies (Piochi et al., 2018).

The type of device resulted the largest procedure difference in manual count method, thus results from HR and LR were compared. Sampled populations in both HR and LR measurements appeared well balanced for gender and no differences in gender distribution were observed (chi-square observed=0.604; chi-square critical value=3.841; p=0.437). Age class distribution differed among HR and LR (chi-square observed=135.47; chi-square critical value=5.991; p<0.0001), with the older classes C2 and C3 being more frequent in LR than in HR. Distributions of counts from HR and LR tended to normality (Fig.1) with the characteristic distribution values slightly higher in HR than in LR (HR: mean=41.5; median= 35.4, 1st quartile limit =24.8; 3rd quartile limit=49.5, range=180.5; LR: mean=35.8, median=31.8, 1st quartile limit=21.1; 3rd quartile limit=45.1, range=131.1). The similarity in characteristic distribution values of HR and LR measurements indicates that the subject classification in low, medium and high FPD based on quartile limits, widely used to investigate relationships between FPD and oral responsiveness, would be similar. This allows for reasonable reliable comparisons among studies in which devices with different resolutions have been used for FPD assessment. The vast majority of measures were collected from not smokers (≥68% in HR, 84% in LR), thus this variable was no further considered.

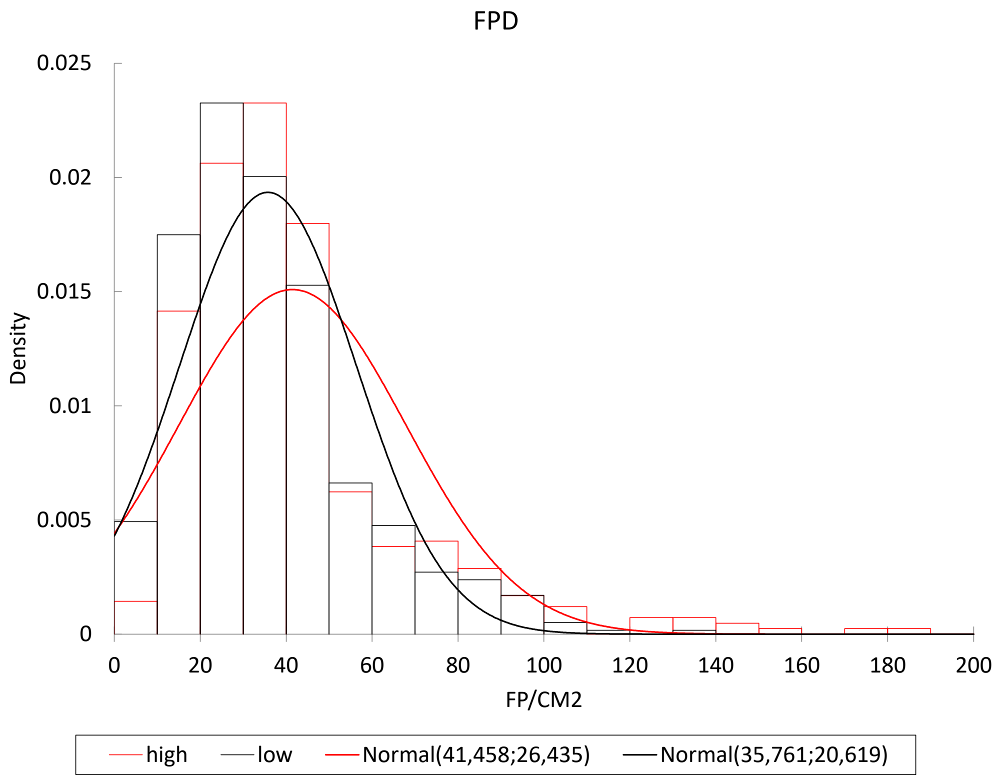


Figure 1. Distribution of FDP counts from measures with low-resolution LR (n=589; black) and high-resolution HR (n= 417; red) devices.

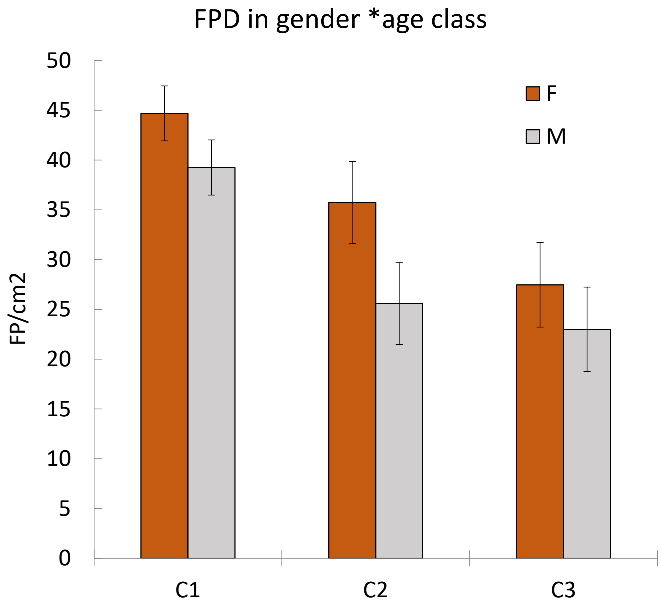
Mean FPD was significantly higher in HR than in LR (F= 14.71, p=0.000). This difference may be due to the better quality of tongue pictures in HR measurements which allow for the identification of a higher amount on FP. However, it is also possible that the younger age of HR sample contributes to this finding.

In order to compare FPD from HR and LR measurements and explore possible effects of procedure variables within each type of measurement, a subset population only consisting of the youngest class (C1:18-30 y.o.) was considered to avoid possible bias due to the negative age effect on FPD. HR and LR samples were comparable for gender composition (Chi-square Observed value=2.836, Chi-square Critical value=3.841, p-value=0.92). No significant differences were found in mean FPD value from HR and LR (F=0.387, p=0.537) in C1 group, thus supporting the age bias on FPD mean values comparison in the HR and LR whole sample. The effect of procedure variables was assessed in HR measurement; no significant differences were found between FPD values from procedures in which the picture was not modified and data were not validated, in comparison to counting performed on modified picture without validating data (p≤0.191). Procedures differed in the region and size of area considered for counting but the effect of this variable was not assessed due to the highly unbalanced sample.

FPD values collected with LR devices were significantly lower when the picture was modified, and data checked for operator agreement (mean value 33.0±1.1) in respect to values obtained from counting on unmodified pictures without validating data (mean value 63.5±1.7) (F=224.03, p<0.001). "False positive" identification of FP due to the low resolution of the unmodified picture and to the lack of the control for the agreement among operators might at least in part account for the observed results.

The ability of HR and LR procedures to capture the differences in FPD due to demographic factors in the sampled populations despite the variability in counting procedure was tested. No significant differences by gender were observed in C1 subsample of HR measurements (p=0.07), age effect was not tested due to low number of observations from the older age classes C2 and C3. In LR sample, a significant effect of gender (F=15.48, p<0.001) was found, with woman showing significantly higher FPD (38.6±1.0) than men (31.7±1.2) and a significant decrease of FPD by aging was observed (p<0.001) (Fig. 2). No significant gender*age interaction was found (p=0.366).

Figure 2. Effect of gender and age class in LR sample (n=589).



Results from both HR and LR datasets confirm that gender does not affect FPD in young adults (Dinnella et al., 2018). LR measures were sensitive to both gender and age effects confirming previous findings on the higher FDP in women and the negative effect of aging (Dinnella et al., 2018; Fischer et al., 2013).

Conclusions

A detailed picture of methods and procedures used across European research units for FPD assessment was obtained, showing that the manual count is the most popular method. The type of device was the largest procedure difference for manual count. However, measures form both devices were comparable and both able to capture variation in FPD due to demographic sample characteristics. Overall, measures from high resolution devices appear the best option to depict the inter-individual variability in FPD due to their stability, irrespectively to differences in the counting procedure. However, the similarity in characteristic values of FPD distributions allows for reasonably reliable comparisons among results from studies adopting low- or high-resolution devices for tongue picture acquisition.

Automated procedures are underutilized and, considering the advantages related to their adoption, an effort is envisaged to widen the accessibility to the script/software, as well as the implementation of commercial versions.

Credits

Maria Piochi: Conceptualization, Methodology, Investigation, Formal analysis, Data Curation, Visualization, Writing - Original Draft.

Melania Melis: Investigation, Writing - Original Draft.

Iole Tomassini Barbarossa, Luisa Torri, Monica Laureati, Ella Pagliarini, Emma Feeney, Rebecca Ford, Qian Yang, Lisa Methven: Investigation, Resources, Writing - Review & Editing.

Erminio Monteleone, Sara Spinelli: Conceptualization, Methodology, Writing - Review & Editing, Supervision,

Caterina Dinella: Conceptualization, Methodology, Investigation, Formal analysis Data Curation, Validation, Writing - Review & Editing, Supervision.

References

1. Cattaneo, C., Liu, J., Wang, C., Pagliarini, E., Sporring, J., & Bredie, W. L. P. (2020). Comparison of manual and machine learning image processing approaches to determine fungiform papillae on the tongue. *Scientific Reports*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-75678-2
2. Dinnella, C., Monteleone, E., Piochi, M., Spinelli, S., Prescott, J., Pierguidi, L., Gasperi, F., Laureati, M., Pagliarini, E., Predieri, S., Torri, L., Barbieri, S., Valli, E., Bianchi, P., Braghieri, A., Del Caro, A., Di Monaco, R., Favotto, S., & Moneta, E. (2018). Individual variation in PROP status, fungiform papillae density, and responsiveness to taste stimuli in a

large population sample. *Chemical Senses*, *43*(9), 697–710. https://doi.org/10.1093/chemse/bjy058

3. Eldeghaidy, S., Thomas, D., Skinner, M., Ford, R., Giesbrecht, T., Thomas, A., Hort, J., & Francis, S. (2018). An automated method to detect and quantify fungiform papillae in the human tongue: Validation and relationship to phenotypical differences in taste perception. *Physiology and Behavior*, *184*(October 2017), 226–234. https://doi.org/10.1016/j.physbeh.2017.12.003
4. Engelen, L., & Van Der Bilt, A. (2008). Oral physiology and texture perception of semisolids. *Journal of Texture Studies*, *39*(1), 83–113. https://doi.org/10.1111/j.1745-4603.2007.00132.x
5. Essick, G. K., Chopra, A., Guest, S., & McGlone, F. (2003). Lingual tactile acuity, taste perception, and the density and diameter of fungiform papillae in female subjects. *Physiology and Behavior*, *80*(2–3), 289–302. https://doi.org/10.1016/j.physbeh.2003.08.007
6. Feeney, E. L., & Hayes, J. E. (2014a). Exploring associations between taste perception, oral anatomy and polymorphisms in the carbonic anhydrase (gustin) gene CA6. *Physiology and Behavior*, *128*, 148–154. https://doi.org/10.1016/j.physbeh.2014.02.013
7. Feeney, E. L., & Hayes, J. E. (2014b). Regional differences in suprathreshold intensity for bitter and umami stimuli. *Chemosensory Perception*, *7*(3–4), 147–157. https://doi.org/10.1007/s12078-014-9166-3
8. Fischer, M. E., Cruickshanks, K. J., Schubert, C. R., Pinto, A., Klein, R., Pankratz, N., Pankow, J. S., & Huang, G. H. (2013). Factors related to fungiform papillae density: The beaver dam offspring study. *Chemical Senses*, *38*(8), 669–677. https://doi.org/10.1093/chemse/bjt033
9. Masi, C., Dinnella, C., Monteleone, E., & Prescott, J. (2015). The impact of individual variations in taste sensitivity on coffee perceptions and preferences. *Physiology and Behavior*, *138*, 219–226. https://doi.org/10.1016/j.physbeh.2014.10.031
10. Melis, M., Atzori, E., Cabras, S., Zonza, A., Calò, C., Muroni, P., Nieddu, M., Padiglia, A., Sogos, V., Tepper, B. J., & Tomassini Barbarossa, I. (2013). The Gustin (CA6) Gene Polymorphism, rs2274333 (A/G), as a Mechanistic Link between PROP Tasting and Fungiform Taste Papilla Density and Maintenance. *PLoS ONE*, *8*(9), 1–15. https://doi.org/10.1371/journal.pone.0074151
11. Miller, I. J., & Reedy, F. E. (1990). Variations in human taste bud density and taste intensity perception. In *Physiology and Behavior* (Vol. 47, Issue 6, pp. 1213–1219). https://doi.org/10.1016/0031-9384(90)90374-D
12. Mistretta, C. M., & Bradley, R. M. (2021). The fungiform papilla is a complex, multimodal, oral sensory organ. *Current Opinion in Physiology*, *20*, 165–173.

https://doi.org/10.1016/j.cophys.2021.01.012
13. Nachtsheim, R., & Schlich, E. (2013). The influence of 6-n-propylthiouracil bitterness, fungiform papilla count and saliva flow on the perception of pressure and fat. *Food Quality and Preference*, *29*(2), 137–145. https://doi.org/10.1016/j.foodqual.2013.03.011
14. Nuessle, T. M., Garneau, N. L., Sloan, M. M., & Santorico, S. A. (2015). Denver papillae protocol for objective analysis of fungiform papillae. *Journal of Visualized Experiments*, *2015*(100). https://doi.org/10.3791/52860
15. Piochi, M., Dinnella, C., Prescott, J., & Monteleone, E. (2018). Associations between human fungiform papillae and responsiveness to oral stimuli: Effects of individual variability, population characteristics, and methods for papillae quantification. *Chemical Senses*, *43*(5), 313–327. https://doi.org/10.1093/chemse/bjy015
16. Piochi, M., Monteleone, E., Torri, L., Masi, C., Almli, V. L., Wold, J. P., & Dinnella, C. (2017). Comparing manual counting to automated image analysis for the assessment of fungiform papillae density on human tongue. *Chemical Senses*, *42*(7), 553–561. https://doi.org/10.1093/chemse/bjx035
17. Sanyal, S., O'Brien, S. M., Hayes, J. E., & Feeney, E. L. (2016). TongueSim: Development of an automated method for rapid assessment of fungiform papillae density for taste research. *Chemical Senses*, *41*(4), 357–365. https://doi.org/10.1093/chemse/bjw008
18. Shahbake, M., Hutchinson, I., Laing, D. G., & Jinks, A. L. (2005). Rapid quantitative assessment of fungiform papillae density in the human tongue. *Brain Research*, *1052*(2), 196–201. https://doi.org/10.1016/j.brainres.2005.06.031
19. Tepper, B. J. (2008). Nutritional implications of genetic taste variation: The role of PROP sensitivity and other taste phenotypes. *Annual Review of Nutrition*, *28*, 367–388. https://doi.org/10.1146/annurev.nutr.28.061807.155458
20. Zhou, X., Yeomans, M., Thomas, A., Wilde, P., Linter, B., & Methven, L. (2021). Individual differences in oral tactile sensitivity and gustatory fatty acid sensitivity and their relationship with fungiform papillae density, mouth behaviour and texture perception of a food model varying in fat. *Food Quality and Preference*, *90*(November 2020), 104116. https://doi.org/10.1016/j.foodqual.2020.104116