DNA yield and quality of saliva samples and suitability for large scale epidemiological studies in children.

Anna Christina Koni[1], Robert A. Scott[1], Guan Wang[1], Mark E.S. Bailey[2], Jenny Peplies[3], Karin Bammann[4], Yannis P. Pitsiladis[1] on behalf of the IDEFICS Consortium

[1]Integrative and Systems Biology, Faculty of Biomedical and Life Sciences (FBLS), University of Glasgow, Glasgow, United Kingdom.

[2]Molecular Genetics, Faculty of Biomedical and Life Sciences (FBLS), University of Glasgow, Glasgow, United Kingdom.

[3]Division of Epidemiological Methods and Etiologic Research, Bremen Institute for Prevention Research and Social Medicine (BIPS), University of Bremen, Bremen, Germany.

[4]Division of Biometry and Data Management, Bremen Institute for Prevention Research and Social Medicine (BIPS), University of Bremen, Bremen, Germany.

Corresponding Author: Dr Yannis P Pitsiladis

Integrative and Systems Biology
Faculty of Biomedical and Life Sciences (FBLS)
West Medical Building
University of Glasgow
Glasgow, G12 8QQ
United Kingdom
Phone: +44 (0)141 330 3858
E-mail:Y.Pitsiladis@bio.gla.ac.uk

Note – this manuscript has been accepted by International Journal of Obesity for a Special Supplement edition, and should be published online before the end of 2010.

Abstract

Objective: To evaluate two saliva collection methods for DNA yield and quality as applied to a large, integrated, multi-centre, European project involving the collection of biological material from children. **Design:** Cross-sectional multi-centre comparative study in young children. **Methods:** Saliva samples were collected from 14,019 children aged 2-9 years from eight European countries participating in the IDEFICS (Identification and prevention of Dietary- and lifestyle-induced health EFfects In Children and infantS) study. This involved either the collection of 2 ml of saliva from children who were able to spit, or using a sponge to collect whole saliva and buccal mucosal cells from the inside of the mouth of younger children unable to spit. Samples were assembled centrally in each participating centre and subsequently dispatched for DNA extraction and biobanking to the University of Glasgow. A subgroup of 4,678 samples (approximately 33% of sampled individuals) was chosen for DNA extraction prior to genotyping. **Results:** The whole saliva collection method resulted in higher DNA yield than the sponge collection method (mean ± SD; Saliva: 20.95 ± 2.35 μg,; Sponge: 9.13 ± 2.25 μg; $P<0.001$). DNA quality as measured by A_{260}/A_{280} was similar for the two collection methods. A minimum genotype calling success rate of 95% showed that both methods provide good quality DNA for genotyping using TaqMan® allelic discrimination assays. **Conclusions**: Our results showed higher DNA yield from the whole saliva collection method compared to the assisted sponge collection. However, both collection methods provided DNA of sufficient quantity and quality for large-scale genetic epidemiological studies.

Keywords: saliva; sample collection; DNA extraction; DNA biobank

Introduction

Genetic epidemiology studies and clinical trials involving genetic association analysis are highly dependent on collecting, storing and distributing DNA of good quality from a representative sample of participants to examine genetic influences on treatment response and disease risk (1). Genotyping success rates depend on DNA quality and yield (2). The traditional method of collecting genetic material suitable for epidemiological studies and multiplex genotyping assays has been based on use of blood samples (3, 4), but this poses challenges, both financial and practical in large studies, particularly when children are investigated. Saliva is increasingly being collected in large studies because of its potential as diagnostic material (5) and has previously been shown to be a reliable source of human genomic DNA suitable for large genetic epidemiology studies (6). The non-invasive nature of this collection method makes it particularly suitable for children. Buccal swabs are convenient and relatively inexpensive compared to blood sampling (7). Saliva sampling has therefore become more popular (Table 1) as the methods for extraction of high quality DNA have developed and as costs have been reduced.

Saliva-based methods typically yield DNA of sufficient quantity and quality to carry out extensive genotyping (see Table 1). It would also appear that these convenient and non-invasive methods increase response rate considerably in epidemiologic studies. However, most previous studies using saliva-based methods have employed relatively small numbers of adult subjects and it therefore remains to be determined whether these methods are suitable for producing DNA of high quality and yield in larger studies, especially those involving young children. The IDEFICS (Identification and prevention of dietary- and lifestyle-induced health effects in children and infants) study is an integrated project funded by the 6th Framework Programme of the European

Commission and with a cohort size of 16,224 young children is one of the largest single studies to undertake saliva/DNA collection (8-13). Here we describe the design and methodological approaches used for DNA collection, extraction, biobanking and genotyping in the IDEFICS study based on the analysis of a subgroup of 4,678 samples selected from the full IDEFICS cohort.

Methods

Participants and saliva collection

Participants in the IDEFICS study included children from eight European countries: Spain, Estonia, Germany, Cyprus, Italy, Denmark, Hungary and Sweden. All applicable institutional and governmental approvals were obtained and all regulations concerning the ethical use of human volunteers were strictly adhered to during this research. Initially 31,543 children were contacted for the IDEFICS study, of whom 16,864 participated in the first assessment and 16,224 fulfilled the inclusion criteria (data available for age, gender, weight and height) (13). Of these 16,224 included subjects, 14,019 (86.4%) provided a saliva sample. The core characteristics of the included subjects from the various survey centers are presented in Table 2. Approximately 2 ml of saliva was collected from children who were able to provide a sputum sample (Oragene DNA Self Collection Kit, tube format OG-300; DNA Genotek Inc., Canada), while sponges (Oragene DNA Self Collection Kit, disc format OG-250 and CS-1 sponge accessory; DNA Genotek Inc., Canada) were used to soak up as much saliva as possible from the inside of the mouths of younger children unable to spit. Measurements in the IDEFICS study were undertaken by mobile field testing teams that visited the schools and nurseries of the children, or when children visited a testing centre. All samples were therefore collected under the supervision of trained personnel as subjects were typically too young

to follow the instructions of the manufacturer. Prior to sample collection, children were advised to rinse their mouths with drinking water and to wait at least 5 min before providing a saliva sample. When using the saliva collection tubes, children were advised to spit into the tube until saliva had been collected up to the level indicated on the collection container (approximately 2 ml). Sponge samples were collected by a trained individual and the sponges were cut into the collection disk container according to the instructions of the manufacturer. Once collected, the trained individual was responsible for covering the tube or disk by placing the cap securely and inverting the container repeatedly for approximately 10 sec to allow the saliva sample to mix well with the Oragene chemistry (DNA Genotek Inc., Canada). Samples from each country were stored at room temperature (approximately 10 to 15 weeks) and subsequently couriered to the central laboratory at the University of Glasgow (UGLW) for DNA extraction, biobanking and genotyping.

DNA processing/purification/extraction

Genomic DNA was extracted from a subgroup of 4,678 samples (all satisfied the following data availability selection criteria: parental questionnaire, height, weight, hip and waist circumferences, age, gender, birthplace and language spoken at home; see Table 2). Samples collected using the sponge method were available from 1,042 girls and 1,178 boys. Samples collected as whole saliva were available from 1,015 girls and 1,086 boys (357 samples were not recorded accurately regarding the saliva method, thus were excluded from the total number of 4,678. DNA from saliva collected in Oragene containers should be stable for at least five years at ambient temperature and resists degradation even when stored at temperatures as high as 50°C (17). Upon arrival at the central laboratory at UGLW, samples were logged using a barcode reader system and

stored at 4°C during processing. DNA was extracted using the protocol for manual purification of DNA from saliva advocated by the manufacturer (18) with minor adjustments to the protocol as detailed below. Prior to extraction, samples were incubated overnight at 50°C in an air incubator (Binder B28, BINDER GmbH, Tuttlingen, Germany). Following this, 500 µl of each sample was transferred into a 1.5 ml microcentrifuge tube and the remaining 1.5 ml sample was resealed in the original collection vessel and frozen at -20°C. Oragene DNA purifier (20 µl) was added to the microcentrifuge tube containing the sample, mixed by vortexing for a few seconds and then incubated on ice for 10 min. Following incubation, the sample mix was centrifuged using a microcentrifuge at room temperature for 10 min at 13000 rpm (15,000 × *g*). The supernatant was carefully transferred with a pipette into a fresh microcentrifuge tube, 500 µl of 100% ethanol at room temperature was added and the tube mixed by inverting approximately 10 times. The tube was then allowed to stand at room temperature for 10 min to precipitate the DNA followed by centrifugation at room temperature for 2 min at 13000 rpm. The supernatant was removed and discarded. Pellets were dried in an air incubator at 50°C for about 20 min and taken up in 500 µl of TE buffer (100 mM Tris, 10 mM EDTA, pH 8.0). Samples were vortexed momentarily and stored at room temperature overnight to encourage DNA dissolution. Extracted samples were then stored at -20°C until quantification. Samples were extracted in batches of 24 or 48, with each trained laboratory worker comfortably processing the extraction of 96 samples per day.

DNA Quantification

Aliquots (167 µl) of each of the 4,678 successfully extracted DNA samples were transferred to 2 ml deep well plates (Starlabs UK Ltd, Buckinghamshire, UK) and quantified using the Nanodrop Technologies Nanodrop® ND-8000 Spectrophotometer

(Wilmington, DE, USA) measuring 8 samples at a time using a multichannel pipette to transfer 1.5 µl undiluted sample to the sample platform. DNA concentrations were estimated from absorbance readings at 260nm (A_{260}) using a 1 O.D. unit = 50 µg/ml conversion factor. A_{260}/A_{280} ratios were also measured. Values for A_{260}/A_{280} ratio normally average approx. 1.8, with intact, high purity DNA usually having a ratio between 1.6 and 2.0. Ratios below approximately 1.6 indicate protein contamination and potentially reduced DNA stability and quality for polymerase chain reaction (PCR) amplification. Values above approx. 2.0 can indicate other small molecule/ionic contamination of the DNA solution. After a first round of quantification, DNA was diluted to a working concentration of 10 ng/µl in TE buffer in 2 ml 96 deep well plates and subsequent aliquots dispensed into 1 ml deep well plates (Starlabs UK Ltd, Buckinghamshire, UK) before storing the original samples in 2 ml plates at -20°C. The working samples were held at 4°C for several weeks during the genotyping analysis.

Genotyping

Genotyping was performed in all 4,678 samples for eight single nucleotide polymorphisms (SNPs) from the β2-adrenergic receptor (*ADRB2*) gene (two assays) and from the angiotensin I converting enzyme 1 (*ACE*) gene (6 assays) using Taqman® assays (Applied Biosystems, Warrington, UK). No DNA template controls were included on each plate. Genotype calls were made by the analysis software (StepOne™ v2.1; Applied Biosystems, Warrington, UK).

Statistical analysis

Analyses were carried out using SPSS software package, version 15.0 (SPSS Inc., Chicago, USA) and MINITAB 15.1.30 (Minitab Ltd., Coventry, UK). Distributions of

DNA yield and A_{260}/A_{280} ratio (by extraction method) were examined and normality assessed using the Anderson-Darling normality test. Non-normal distributions were transformed using the standard function (e.g. raw, log, inverse etc.) closest to optimal after Box-Cox analysis. Group differences for normally distributed (i.e. transformed) variables were tested using an independent t-test, with Welch's correction where variances were unequal (evaluated using Levene's test). For presentation purposes, mean values and other relevant statistics were back-transformed. Differences were considered statistically significant at $P<0.05$. Data are presented as mean ± SD/ or as median and interquartile range.

Results

DNA extraction yield and quality (by A_{260}/A_{280} ratio) were evaluated in the full sample set and by extraction method. Some samples (1% of the 14,019 samples taken) were dry (indicating leakage of the saliva and chemicals present in the Oragene kit) upon receipt at the central laboratory at University of Glasgow (UGLW) and/or were not properly labeled. Of the 4,678 samples selected for DNA extraction, 100% were successfully extracted. Investigation of the distribution of DNA yields from the successful extractions revealed 14 samples with improbable high yields (>1000 μg), 42 samples with zero or negative yield (based on negative A_{260} readings) and 13 samples with zero or negative A_{260}/A_{280} readings. These 'low yield' samples were excluded from further analysis. An upper yield cut-off for inclusion in the analysis was defined as 3 standard deviations (SD) from the mean. As a result, a further 94 samples were excluded, leaving a total of 4,584 samples with acceptable yield for final analysis. Yield and quality data are given for these included samples in Table 3. In 357 of the 4,584 samples, saliva collection method was

not recorded accurately. Therefore, these samples were excluded solely from the analysis of differences between saliva collection methods. The whole saliva collection method resulted in higher DNA yield than the sponge collection method (mean ± SD; Saliva: 20.95 ± 2.35 μg, Sponge: 9.13 ± 2.25 μg; P<0.001). DNA quality (as assessed by A_{260}/A_{280} ratio) did not differ between the two collection methods. Analysis of the relationship between DNA yield and quality (Figure 1, n=4,070) revealed that high A_{260}/A_{280} ratios were mainly observed in low yield samples prepared from sponges, while low ratios were predominantly observed in either high yield samples prepared from saliva or low yield samples prepared from sponges. Low yield samples from the sponge collection method appear to have a spread of A_{260}/A_{280} ratios without a peak around 1.8 suggesting that these samples were unlikely to be accurately measured and/or to be of useful quality. For the whole saliva collection method, on the other hand, low yield samples seemed to cluster around A_{260}/A_{280} ratios of about 1.6-2.0, probably representing DNA of sufficient quality to be usable in genotyping experiments.

PCR was performed on eight SNPs from the *ADRB2* and *ACE* genes. The proportion of samples in which genotypes were successfully called (the genotype-calling success rate) was 96-97% (N=4,678) (Table 4). To analyse the genotyping success rate in relation to DNA quality, we applied cutoffs at A_{260}/A_{280} ratios <1.5 and >2.1, and at DNA yields (expressed as natural log) of <0.5 or >5.0 (see Figure 1; these values were arbitrarily selected as being unusual in DNA of good quality made by the methods applied here, and as being the points that would create an 'outlier' dataset representing about 10% of the samples). In all, 427 samples fell outside the cutoffs and were used in further analysis as an 'outlier' dataset. Of these outliers, 83.1% were successfully genotyped in all 8 assays,

and 94.6% were successfully genotyped in at least 6 assays. The corresponding proportions for the non-outliers were 89.8% and 98.7%, respectively (see Table 5).

Discussion

These results demonstrate the effectiveness of the present methodological approaches used for DNA collection, extraction, biobanking and genotyping for use in large epidemiological studies in children such as those participating in the IDEFICS study. The overwhelming majority (4584 out of 4678 or 98%) of samples provided DNA of sufficiently high yield and quality (Table 3) for multiple genotyping assays as evidenced by the high genotyping success rates for both DNA collection methods (Table 4). The DNA yield differed between the two collection methods, with whole saliva yields being higher than yields from sponges (Table 3) and this is in line with previous reports in the literature (1,2,4,5,14-16,19-21). Total DNA yield from the 4 ml Oragene DNA/saliva solution fulfilled, on average, the specifications of the manufacturer of above 20 µg (22) for the whole saliva collection method (Table 3). However, considerable variability in DNA yield was evident from the data presented in Table 3, with a median and interquartile (IQ) range of 21.8 µg and 11.9-38.2 µg, respectively. In terms of the sponge saliva collection method, total DNA yield in the present study (Table 3) was somewhat lower (median 9.1 µg, IQ: 5.2-15.9 µg) than the specifications of the manufacturer (median 13.4 µg; 23) Again, considerable variability in DNA yield from the saliva sponge collection method was evident in the present study. Nevertheless, the genotyping success rate was high for both methods regardless of the lower yield when using the sponge method (Table 4) and in line with the success rate expected when using higher yield methods from blood and other biological material. The genotype-calling success rates (the average being 96-97% and the minimum, amongst the 8 SNPs tested being

95%) were well within expectations for the TaqMan assay approach to SNP genotyping. Genotyping failure can be due to many factors, including subtle variation in DNA quality that is not visible in the yield and ratio data, as well as factors operating during PCR setup. Interoperator variability may have contributed to the former. It should be noted that for the purposes of the present analysis we did not control for the effect of the age of the child on DNA yield. Whole saliva was generally obtained from children towards the higher end of the age range (typically ages 6-9 years) and it is likely that a larger amount of saliva was collected in these children (although instructions were to collect precisely 2 ml saliva samples), or that there was a higher DNA concentration in the saliva samples obtained in these older children. DNA quality, however, was very similar for the two collection methods. Looking at the plot of DNA yield vs quality in Figure 1 it was clear that samples with unusually high A_{260}/A_{280} ratios tended to have low yield and have been prepared from sponges, probably reflecting the low concentration of human cellular material and high levels of bacterial contamination in material collected on sponges. By contrast, low ratios were mainly observed in either high yield samples prepared from saliva or low yield samples prepared from sponges. Low ratios usually indicate protein contamination, which is often the case in low yield circumstances. Protein contamination in the high yield samples from saliva probably reflects high levels of bacterial contamination or possibly the presence of food particles in substantial quantities. Future analysis will address the question of whether yield is different per ml of saliva collected at different ages and whether genotype-calling success rate varies with collection method, DNA yield or A_{260}/A_{280} ratio. Saliva collection as a method for collecting genetic material is particularly applicable for children in the general population and has been found to be associated with significantly higher participation rates compared to collection methods utilizing venous blood (24). Interestingly, non-invasive sample collection

methods (i.e. use of urine and saliva) showed higher response rates than samples collected by venous blood and capillary blood in the multicentre IDEFICS study (56.6% and 90.1% for venous blood and saliva, respectively) (24). The good response rate amongst subjects, and high DNA extraction and genotyping success rates make collection of saliva viable as a method for collecting samples for extraction of genomic DNA in large scale multicentre studies in young children.

In summary, whole saliva and sponge collection methods provided DNA yields and quality sufficiently high for successful genotyping rates, making it suitable for large scale epidemiological studies. These results support the use of the Oragene Saliva collection kits in large scale studies of children. The fact that the present data comes from a large cohort of young children provides unique insight as no other study of this nature has been conducted in children to date.

Acknowledgements

The authors would like to acknowledge all members of the UGLW DNA extraction team for their excellent technical assistance in completing the DNA extraction process. This work was done as part of the IDEFICS Study (www.idefics.eu). We gratefully acknowledge the financial support of the European Community within the Sixth RTD Framework Programme Contract No. 016181 (FOOD). The information in this document reflects the author's view and is provided as is. No guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Tables and Figure

Figure 1. The relationship between DNA yield and quality from whole saliva and sponge samples.

A_{260} measurements were used to calculate DNA yield (μg) then natural log transformed for presentation and plotted against A_{260}/A_{280} ratio by collection method. To separate outliers (defined in the Results section) from the main cluster, four reference lines were applied. Vertical lines indicate A_{260}/A_{280} ratios of 1.5 and 2.1. Horizontal lines indicate in DNA yields of 0.5 and 5.

Table 1. Summary of previous studies using saliva collection methods.

Year	Population / Collection methods	Conclusions	Reference
2001	cytobrush: n=120, mouthwash: n=40,	Both methods are adequate for a wide range of PCR-based analysis. A single mouthwash sample provides larger amount and higher molecular weight DNA than two cytobrush samples.	(14)
2001	35 subjects-24 individuals (six men and 18 women) completed the study, Buccal cell, mouthwash	Buccal samples should be collected before brushing teeth and processed within 5 days of collection to maximize DNA yield.	(15)
2002	24 participants (45 years or older) – 22 completed the study (9 men and 13 women)	Cytobrush is cost effective in large scale studies, and yields sufficient quantity and quality of DNA for genotyping.	(16)
2006	Buccal cell, mouthwash 611 men 53-87 years Self-administered Oragene method	Oragene saliva DNA is of high quality and can be used as alternative to blood.	(2)
2006	10 subjects	Saliva is a viable alternative source of human genomic DNA for genetic epidemiological studies.	(4)
2006	5 males and 5 females Mouthwash and cheek cells using swabs.	Saliva samples provide a substantial increase in the amount of human DNA. Saliva samples can be obtained and transported under field conditions without refrigeration.	(19)
2007	300 nurses, 51-91 years, Oragene® DNA self-collection kit, Catch-All swabs, blood sample	Saliva sampling is a good alternative to blood sampling and would increase the response rate considerably in epidemiologic studies.	(20)
2007	17 adult volunteers: Oragene® DNA collection kit, cytobrush, foam swabs, oral rinse.	Both oral-rinse sample and whole-saliva sample provide sufficient DNA quantity and better quality DNA for genetic epidemiological studies than buccal swab and brush techniques.	(21)
2009	34 individuals: Buccal swabs and 1140 subjects: Catch-All swabs, Isohelix buccal	The buccal swabs are convenient and cost effective alternative to blood sampling. They provide DNA of sufficient quantity and quality for high-throughput SNP multiplex analysis.	(5)
2009	565 individuals, Oragene® Self Collection Kits	Demographic and behavioural characteristics of smoking cessation trial participants are associated with saliva and DNA metrics but not with the performance on genotyping.	(1)

Table 2. Children fulfilling the inclusion criteria and providing saliva samples

Country	n	% of total subjects[a]	Age Mean (SD)	boys % (N)
Italy	1952	86.8	6.1 (1.8)	51.6 (1007)
Estonia	1418	82.5	6.0 (2.0)	49.3 (699)
Cyprus	1679	70.5	6.1 (1.4)	51.3 (862)
Belgium	1524	79.1	5.7 (1.6)	51.0 (777)
Sweden	1601	88.5	5.9 (2.0)	51.2 (820)
Germany	1947	94.2	6.2 (1.8)	51.2 (997)
Hungary	2518	98.1	6.3 (1.8)	50.0 (1260)
Spain	1380	91.6	5.9 (1.7)	51.1 (705)
All	14019	86.4	6.1 (1.8)	50.8 (7127)

[a] proportion of subjects from each country (and the full cohort) that fulfilled the inclusion criteria who also gave saliva samples

Table 3. Comparison of DNA yield and quality by method of collection.

	DNA yield (μg)				A_{260}/A_{280} ratio		
Collection Method	n	Mean	SD	Median (IQ)	Mean	SD	Median (IQ)
Total sample	4584	14.36	2.83	13.88 (7.35-27.26)	1.78	0.27	1.80 (1.67-1.91)
Whole saliva	1962	20.95	2.35	21.76 (11.93-38.21)	1.77	0.18	1.79 (1.67-1.89)
Sponge	2108	9.13	2.25	9.24 (5.23-15.90)	1.80	0.20	1.82 (1.69-1.93)

IQ: Limits of the interquartile range

Table 4. Genotyping success rates in saliva samples.

Gene	*ADRB2*		*ACE*					
SNP	rs1042713	rs1042714	rs4351	rs4362	rs4329	rs4295	rs4353	rs4311
% successful genotype calls	96	96	96	97	97	96	96	97

Table 5. Genotyping success rate for the outliers.

	Number of SNP assays in which each sample resulted in successful genotype calls			
	8	7	6	0-5
Number (%) of outliers (total n=427)	355 (83.1)	38 (8.9)	11 (2.6)	23 (5.4)
% of non-outlier samples (total n=3,643)	89.8	7.4	1.5	1.3

The table shows the number and percentage of samples giving successful genotype calls for the indicated number of SNP assays, i.e. 355 outlier samples were called in all 8 assays, whereas 23 outlier samples were successful in five assays or fewer. Percentage of non-outliers giving successful genotype calls are also given for comparison.

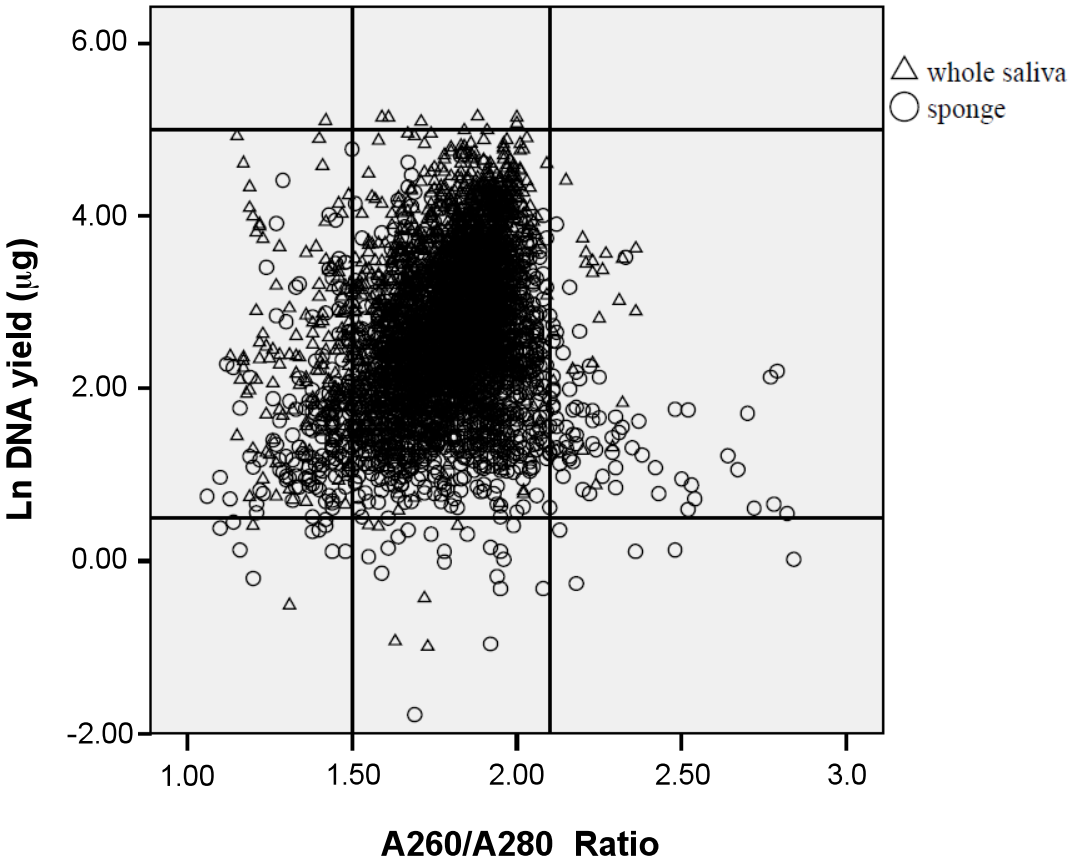

Figure 1. The relationship between DNA yield and quality from whole saliva and sponge samples

References

1. Nishita DM, Jack LM, McElroy M, McClure JB, Richards J, Swan GE*, et al.* Clinical trial participant characteristics and saliva and DNA metrics. *BMC Med Res Methodol* 2009; **9:** 71.
2. Rylander-Rudqvist T, Hakansson N, Tybring G, Wolk A. Quality and quantity of saliva DNA obtained from the self-administrated oragene method--a pilot study on the cohort of Swedish men. *Cancer Epidemiol Biomarkers Prev* 2006; **15:** 1742-1745.
3. Philibert RA, Zadorozhnyaya O, Beach SR, Brody GH. Comparison of the genotyping results using DNA obtained from blood and saliva. *Psychiatr Genet* 2008; **18:** 275-281.
4. Matheson LA, Duong TT, Rosenberg AM, Yeung RS. Assessment of sample collection and storage methods for multicenter immunologic research in children. *J Immunol Methods* 2008; **339:** 82-89.
5. Streckfus CF, Bigler LR. Saliva as a diagnostic fluid. *Oral Dis* 2002; **8:** 69-76.
6. Ng DP, Koh D, Choo S, Chia KS. Saliva as a viable alternative source of human genomic DNA in genetic epidemiology. *Clin Chim Acta* 2006; **367:** 81-85.
7. McMichael GL, Gibson CS, O'Callaghan ME, Goldwater PN, Dekker GA, Haan EA*, et al.* DNA from buccal swabs suitable for high-throughput SNP multiplex analysis. *J Biomol Tech* 2009; **20:** 232-235.
8. Ahrens W, Bammann K, de Henauw S, Halford J, Palou A, Pigeot I*, et al.* Understanding and preventing childhood obesity and related disorders--IDEFICS: a European multilevel epidemiological approach. *Nutr Metab Cardiovasc Dis* 2006; **16:** 302-308.
9. Haerens L, De Bourdeaudhuij I, Eiben G, Lauria F, Bel S, Keimer K*, et al.* Formative research to develop the IDEFICS physical activity intervention component: findings from focus groups with children and parents. *J Phys Act Health*; **7:** 246-256.
10. Espana-Romero V, Artero EG, Santaliestra-Pasias AM, Gutierrez A, Castillo MJ, Ruiz JR. Hand span influences optimal grip span in boys and girls aged 6 to 12 years. *J Hand Surg Am* 2008; **33:** 378-384.
11. Bammann K, Peplies J, *et al*. Assessment of diet, physical activity and biological , social and environmental factors in a multi-centre European project on diet- and lifestyle- related disorders in children (IDEFICS). *Journal of Public Health* 2006; **15:** 279-289.
12. Pigeot I, Barba G, Chadjigeorgiou C, de Henauw S, Kourides Y, Lissner L*, et al.* Prevalence and determinants of childhood overweight and obesity in European

countries: pooled analysis of the existing surveys within the IDEFICS Consortium. *Int J Obes (Lond)* 2009; **33:** 1103-1110.

13. Ahrens W, Pigeot I. The IDEFICS study: design, participation and compliance with collection of biological samples. Submitted to *Int J Obes*.
14. Garcia-Closas M, Egan KM, Abruzzo J, Newcomb PA, Titus-Ernstoff L, Franklin T*, et al.* Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. *Cancer Epidemiol Biomarkers Prev* 2001; **10:** 687-696.
15. Feigelson HS, Rodriguez C, Robertson AS, Jacobs EJ, Calle EE, Reid YA*, et al.* Determinants of DNA yield and quality from buccal cell samples collected with mouthwash. *Cancer Epidemiol Biomarkers Prev* 2001; **10:** 1005-1008.
16. King IB, Satia-Abouta J, Thornquist MD, Bigler J, Patterson RE, Kristal AR*, et al.* Buccal cell DNA yield, quality, and collection costs: comparison of methods for large-scale studies. *Cancer Epidemiol Biomarkers Prev* 2002; **11:** 1130-1133.
17. http://www.dnagenotek.com/pdf_files/PDPR012_LongTermStorage.pdf
18. http://www.dnagenotek.com/DNA_Genotek_Industry_AR_SCA_P.html
19. Quinque D, Kittler R, Kayser M, Stoneking M, Nasidze I. Evaluation of saliva as a source of human DNA for population and association studies. *Anal Biochem* 2006; **353:** 272-277.
20. Hansen TV, Simonsen MK, Nielsen FC, Hundrup YA. Collection of blood, saliva, and buccal cell samples in a pilot study on the Danish nurse cohort: comparison of the response rate and quality of genomic DNA. *Cancer Epidemiol Biomarkers Prev* 2007; **16:** 2072-2076.
21. Rogers NL, Cole SA, Lan HC, Crossa A, Demerath EW. New saliva DNA collection method compared to buccal cell collection techniques for epidemiological studies. *Am J Hum Biol* 2007; **19:** 319-326.
22. Maximizing DNA yield with Oragene® DNA http://www.dnagenotek.com
23. http://www.dnagenotek.com/DNA_Genotek_Product_Oragene_DNA_C_Overview.html.
24. Peplies J, Fraterman A, Scott R, Russo P, Bammann K. Quality management for the collection of biological samples in multicentre studies. *Eur J Epidemiol*. 2010 [Epub ahead of print]