BCPT Basic & Clinical Pharmacology & Toxicology

Basic & Clinical Pharmacology & Toxicology, 2017, **121**, 93–101 Doi: 10.1111/bcpt.12721

MiniReview

Salivary DNA Methylation Profiling: Aspects to Consider for Biomarker Identification

Sabine A. S. Langie[1,2], Matthieu Moisse[3], Ken Declerck[4], Gudrun Koppen[1], Lode Godderis[5,6], Wim Vanden Berghe[4], Stacy Drury[7,8] and Patrick De Boever[1,2]

[1]Environmental Risk and Health Unit, Flemish Institute for Technological Research (VITO), Mol, Belgium, [2]Faculty of Sciences, Hasselt University, Diepenbeek, Belgium, [3]VIB Vesalius Research Center, KU Leuven, Leuven, Belgium, [4]Laboratory of Protein Chemistry, Proteomics and Epigenetic Signaling, Department of Biomedical Sciences, University of Antwerp, Wilrijk, Belgium, [5]Centre Environment & Health, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium, [6]IDEWE, External Service for Prevention and Protection at Work, Heverlee, Belgium, [7]The Brain Institute, Tulane University, New Orleans, LA, USA and [8]Department of Psychiatry and Behavioral Science, Tulane University School of Medicine, New Orleans, LA, USA

(Received 1 September 2016; Accepted 22 November 2016)

Abstract: Is it not more comfortable to spit saliva in a tube than to be pricked with a needle to draw blood to analyse your health and disease risk? Many patients, study participants and (parents of) young children undoubtedly prefer non-invasive and convenient procedures. Such procedures increase compliance rates especially for longitudinal prospective studies. Saliva is an attractive biofluid providing good quality DNA to study epigenetic mechanisms underlying disease across development. In this MiniReview, we will describe the different applications of saliva in the field of epigenetics, focusing on genomewide methylation analysis. Advantages of the use of saliva and its comparability with blood will be discussed, as will the challenges in data processing and interpretation. Knowledge gaps will be identified and suggestions given on how to improve the analysis, making saliva 'the' biofluid of choice for future biomarker initiatives in many different epidemiological and public health studies.

The field of molecular biomarkers of health and disease is rapidly evolving, requiring extensive research and insight in the pathways involved in early disease development. Epigenetic factors defined as relatively stable and to some extent heritable changes that do not modify the DNA sequence provide important information and are critical regulators of gene and protein expression. The most studied epigenetic mechanisms are DNA methylation, histone modifications and non-coding microRNA (miRNA). The dynamic epigenome is thought to capture and encode environmental changes, serving as an important pathway for the interaction between genes and environment. Although we are only beginning to understand the importance of the epigenome in health and disease, it is clear that epigenetic factors are likely mediating factors in the development of complex diseases across the life course [1,2].

Environmental epidemiology and public health researchers often do not have access to specific target tissues related to complex diseases; as such, they must rely on peripheral biological sources for study. The majority of epigenetic and biomarker research to date has relied on peripheral blood as the main surrogate tissue. The need of trained personnel and logistics can make blood sampling challenging when doing decentralized investigations. In addition, blood sampling in vulnerable individuals, such as children, should be avoided as much as possible to reduce discomfort, increase participation and improve overall feasibility. Saliva can be a good alternative surrogate tissue as it is easier to collect and it is a good source of high-quality DNA for use in (epi)genomic studies [3–8]. Saliva has recently attracted much attention because it also contains a broad range of other diagnostically relevant molecules (i.e. microRNA, RNA, inflammation markers and antibodies).

Salivary research began with the study of salivary analytes to detect local mouth and throat diseases and expanded towards research into prediction/diagnosis of systemic diseases and health conditions [9,10]. For example, salivary cytokine profiles have been successfully used as biomarkers of respiratory and other immunological disorders in the field of clinical diagnostics [9,10]. Over the last decade, research examining the impact of stress on human health and disease has relied almost exclusively on the measurement of neuroendocrine markers, including cortisol, from saliva particularly after side-by-side studies demonstrated the equivalence of salivary and blood measurement [11]. The research field is now also focusing on the detection of time-sensitive and context-dependent epigenomic alterations in saliva, and it has embraced a wide range of biomolecular techniques, including bisulphite sequencing, methylation arrays, a range of PCR and qPCR-based techniques.

Author for correspondence: Dr. Sabine A. S. Langie, VITO NV, Boeretang 200, Mol 2400, Belgium (e-mail sabine.langie@vito.be).

One of the most important advantages of saliva over blood is the ease of collection, facilitating multiple collections in 1 day, or even over shorter time periods. Apart from the fact that it is more comfortable to take a saliva sample, repeated sampling is usually more practical than blood as (i) there is no need for the addition of an anticoagulant to avoid clotting upon collection and (ii) the risk of disease transmission that can occur when in contact or via needle prick is lower than with blood sampling [12]. In addition, no specially trained (para)medical personnel is required and, with appropriate instructions and attention to sample collection, an individual is able to self-collect multiple samples for clinical and research purposes. As with all biological samples, collection, storage and transport of saliva should be carried out with care to preserve sample integrity. The importance of correct sample collection procedures and the available different analytes present in saliva was recently discussed [9,10] and will therefore not be dealt with in this MiniReview. Briefly, it is advisable to carefully choose your saliva collection method (e.g. stimulated *versus* unstimulated and use of suction) depending on the envisioned downstream analysis.

microRNA expression studies in saliva [9,13,14] have increased in the past decade, while research into salivary histone modifications is still in its infancy with only few reports on salivary proteomics (including histones), mainly in the field of oral diagnostics [15,16]. In the current MiniReview, we will be focusing on DNA methylation profiling, data processing and interpretation. Our aim is to highlight the utility of saliva for DNA methylation research and biomarker identification.

Primer on the Methylome

DNA methylation occurs through the covalent binding of a methyl group to the DNA, catalysed by DNA methyltransferases (DNMTs), mainly at cytosine residues within CpG dinucleotides. In mammalian genomes, the CpG dinucleotide is under-represented and is mainly observed as dense clusters of CpG dinucleotides called 'CpG islands' [17]. About 60% of these islands overlap promoter regions and are largely unmethylated. It is generally believed that this status of hypomethylation promotes gene expression. Nonetheless, about 80% of all CpG sites are said to be methylated, mostly in repetitive sequences made up of transposable elements (incl. long terminal repeats (LTR)-retrotransposons, long and short interspersed nuclear elements (LINE and SINE, respectively) and DNA transposons), which are silenced by methylation to support genome stability and integrity [17]. The removal of 5mC is catalysed by the ten eleven translocation (TET) enzymes and involves, as an intermediate step, the oxidation of 5mC into 5-hydroxymethylcytosine (5hmC). It is generally believed that 5hmC may have a more specific role in regulating transcription, most probably reactivating gene expression, while 5mC might have additional roles in maintaining genomic integrity and transposon stability [17].

Altered global DNA methylation content is a feature of several diseases; its occurrence in cancer was first highlighted by Feinberg and Volgelstein [18]. An important feature of cancer development and progression is the dysregulation of DNA methylation patterns, characterized by the hypermethylation of specific genes concurrent with an overall decrease in the abundance of 5mC. In particular, hypomethylation can lead to overexpression of oncogenes [19,20]. Hypermethylation of the promoter regions of tumour suppressor and DNA repair genes causes gene silencing and contributes to tumorigenesis. Furthermore, studies with monozygotic twins show that environmental factors influence DNA methylation [21,22]. Age has also been shown to have a multifacet effect on DNA methylation [23,24]. Interestingly, the within-tissue epigenetic correlation was reported to be higher earlier in development and diverges more later in development. In this respect, the DNA methylation status of approximately 50 epigenetic clock-associated CpG sites allows to calculate accelerated or decelerated epigenetic ageing as compared to biological ageing [25–27]. Although there is evidence of changes in methylation across development, ageing and with different environmental exposures recent data suggest that more dynamic changes in methylation, even over the course of a social stressor, may occur although studies carefully accounting for cell type changes and other factors are needed to determine the validity of these changes in either blood or saliva [28]. For example, acute psychosocial stress was observed to increase DNA methylation of the stress-associated gene *OXTR* already 10 min. after applying the stressor, returning to initial pre-stress DNA methylation levels 90 min. after the stress was applied [29]. This work underlines the dynamic nature of DNA methylation profiles. Unternaehrer *et al.* demonstrated the effect of acute psychosocial stress on DNA methylation in blood. We propose the investigation into the epigenetic effects of stress using saliva as biofluid because multiple saliva samples can be collected in a short time period and do not cause any additional stress as may be the case with blood sampling.

Sample Quality

Given the increased interest to use saliva for (epi)genomics, a range of DNA collection kits (available from, e.g., DNA Genotek, Ottawa, Canada, Norgen Biotek Corp. Thorold, Canada and Isohelix, Kent, United Kingdom), stabilizing reagents and purification procedures have been developed [30]. The kits ensure nucleic acid stabilization and isolation of high-quality and high molecular weight DNA. An increasing number of reports confirm that the yield and quality of saliva is high, as assessed by purity and feasibility of downstream applications including sequencing, genotyping, PCR amplification and genomewide arrays [31–33]. In our hands, saliva yielded good quality DNA as determined by the UV absorption profiles (median 260/280 = 1.8, median 260/230 = 1.1, median yield 54.4 μg from 2 mL saliva), which is in line with previous reports [34,35]. The samples passed the different steps for genomewide DNA methylation analysis using the Illumina Methylation 450K BeadChip [36]. To further confirm that salivary DNA is of high quality, in our recent birth cohort study, we observed that saliva of 99 children, collected at 11 years of age, yielded high

amounts of good quality DNA (median yield 79.6 μg from 2 mL saliva with median 260/280 = 1.81 and median 260/230 = 1.11) [37].

An additional advantage of salivary DNA collection kits claimed by the manufacturers is the possibility of sample storage at room temperature for 5 years or more without any detectable DNA degradation [38]. Independent researchers performed similar tests and reported that saliva can be stored at temperatures up to 37°C for between 6 and 18 months without compromising the DNA quantity, quality and applicability for a range of different genetic analyses [32,39,40]. Consequently, individuals can collect their saliva at home and either transport it to their doctor or the research site, or mail the sample directly via standard postal systems to the research institute, clinic or hospital. Ng *et al.* [40] even reported that storage of saliva collection kits at air-conditioned room temperature (20°C) for 6 months or at −80°C for 6 months did not affect DNA quality (OD260/280 values were comparable) in real-time PCR experiments and genotyping fidelity remained undiminished.

Concern about the presence of bacterial DNA in saliva has been raised by some researchers, particularly as current quantification methodologies cannot distinguish between bacterial and human DNA, resulting in a potential overestimation of DNA quantity. For instance, Rylander-Rudqvist *et al.* [33] observed PicoGreen and UV absorbance measurements (detecting bacterial and human DNA) to overestimate the yield of human DNA from salivary cells by ~1.6-fold compared to real-time PCR measurements (detecting only human DNA). Nonetheless, significant correlations were seen between the various DNA quantification methods. Bacterial DNA also contains 5mC, and concerns about the presence of bacterial DNA leading to an overestimation of global genomic 5mC levels in human saliva DNA have been discussed. For instance, ~0.75% of all cytosines in the DNA of *E. coli* are methylated, compared to ~4% of all cytosine residues in the human genome [17,41]. To the best of our knowledge, there have been no reports indicating that bacterial DNA competes with human DNA in the hybridization steps of genomewide DNA methylation screening approaches. Abraham *et al.* [34] reported DNA from saliva and blood samples to be of comparable quality when genotyping using either TaqMan assays or genomewide chip arrays. Moreover, in a recent study [36], we observed all our saliva samples to pass quality controls and behave in the same way as blood samples taken from the same individuals. In addition, results from our Illumina arrays were confirmed by gene-specific bisulphite pyrosequencing, which is also an indication that the presence of bacterial DNA was probably not an issue when using proper collection kits for genomewide analyses [36]. One approach to address this would be to perform a series of experiments where bacterial DNA at varying concentrations was spiked into the reaction and assess changes in genomewide DNA methylation analysis. Alternatively, manufactures of genomewide methylation chips could build in probes tagged to bacterial DNA. Analyses of these specific probes could be used to determine the amount of bacterial DNA contamination and examined for overall differences across the array. Noteworthy, to minimize any bacterial interference, the manufactures of the above-mentioned saliva DNA collection kits add an antibacterial agent to the stabilizing fluid, which prevents the growth of bacteria. For instance, it has been shown that the saliva collected with the Oragene DNA collection kits contain only 11.8% bacterial DNA, while about 5–8 times higher amounts are obtained from mouthwashes (median bacterial DNA content 60%) or buccal swabs (median bacterial DNA content 90%) [42]. In an independent study, Rylander-Rudqvist *et al.* [33] showed 68% of the total DNA to be of human origin when using the Oragene DNA collection kit. Taken together, the ability to obtain significant amounts of high-quality and high molecular weight DNA from saliva samples, despite some manageable concerns related to bacterial contamination when coupled with the improved efficiency of collection, supports the use of salivary DNA as an alternative to blood DNA in molecular epidemiological studies.

Salivary DNA Methylation

A handful of studies have compared DNA methylation patterns in blood and saliva [6–8,36,43]. However, these studies differ in experimental set-up and type of sample (whole blood *versus* more homogenous isolated cell subpopulations) that was used for the comparison, making generalization difficult at this moment. When looking at genomewide methylation patterns, Smith *et al.* [6] observed the saliva methylome to be positively correlated with methylation in blood for 88.5% of the CpG sites studied on the Illumina 450K arrays. Thompson and colleagues [7] generated genomewide DNA methylation profiles of whole blood and saliva samples of healthy adults on an Illumina 27K platform, observing 1.8% of the probes to be differentially methylated when applying a Benjamini–Hochberg-adjusted *p*-value (p_{adj}) < 0.001 and DiffScore > |30| as cut-offs. The methylation difference score (DiffScore) for a CpG sites is a parameter that Illumina's Genome Studio Software provides in their output, which takes into account background noise and sample variability [44]. Langie *et al.* [36] applied the same cut-offs as Thompson *et al.* [7] (p_{adj} < 0.001 and DiffScore > |30|) to their Illumina 450K Beadchip data and also observed 1.8% of the CpG sites to be differentially methylated when considering the 27K probes. When using cut-offs of p_{adj} < 0.001 and $|\Delta\beta| > 0.2$, Langie *et al.* [36] showed 4% of the CpG sites on the Illumina 450K Beadchip to be differentially methylated. Overall, these data indicate that the majority of CpG sites were similarly methylated in blood and saliva.

The studies mentioned above were performed with adult volunteers. As DNA methylation patterns can be dependent on age and lifestyle, blood and saliva methylomes of young adolescents were recently compared. Langie *et al.* [37] recruited 11-year-old children, among which 20 with doctor-diagnosed respiratory allergy (having Phadiatop IgE ≥ 0.35 kU/L) and 20 healthy controls (no self-reported/diagnosed respiratory allergy, Phadiatop IgE < 0.35 kU/L). DNA was isolated from peripheral blood mononuclear cells (PBMC) (from 10 mL

blood) and saliva samples (2 mL in Oragene DNA OG-500 self-collection kit; DNA Genotek), followed by bisulphite treatment and 450K Illumina Beadchip analysis as described previously [36]. Ethical approval from the University Hospital in Antwerp (Belgian registration number B300201317947) and written informed consent from the children's parents was obtained prior to sample collection. Following quality filtering and data normalization, 471,560 individual CpG sites were included in downstream data analyses. When using the same cut-offs as Langie *et al.* [36] applied in their study on adults ($p_{adj} < 0.001$ and $|\Delta\beta| > 0.2$), 16,575 CpG sites (3.5%) were differentially methylated between saliva and PBMC of 11-year-old children. About 28% (4666) of these CpG sites were hypermethylated and 72% (11,909) hypomethylated in saliva DNA compared to PBMC. When considering only the methylation profiles of the 20 healthy 11-year-old controls, 16,735 CpG sites (3.5%) showed differential methylation in saliva *versus* PBMC; of which, 4630 (28%) were hyper- and 12,105 (72%) hypomethylated. Saliva and blood methylation patterns could be distinguished from each other, but the methylation status of about 96.5% of the CpG sites was highly comparable between PBMC and saliva (fig. 1).

A number of independent studies confirm that genomewide DNA methylation profiles of saliva are more than 90% comparable to blood, both in adults [6,7,36] and in adolescents [37]. However, based on analysis of global DNA methylation patterns, Godderis *et al.* could not completely confirm this. The authors compared global DNA methylation and hydroxymethylation (5hmC) in whole blood and saliva from 14 healthy volunteers [43]. Global DNA methylation was assessed as 5-methyl-cytidine (5mC) by LC–MS/MS and as the percentages of methylation of DNA repetitive elements LINE1 and Alu via bisulphite pyrosequencing. While methylation percentages were significantly lower in saliva samples compared to blood samples (e.g. for 5mC; 4.61 ± 0.80% *versus* 5.70 ± 0.22%, respectively), levels of 5hmC were significantly higher (0.036 ± 0.011% in saliva *versus* 0.027 ± 0.004% in blood). Levels of 5hmC and 5mC in saliva showed a significant positive correlation, which was not observed for blood. No significant correlations between saliva and blood samples were observed for global methylation levels (either as 5mC, LINE1 or Alu), nor for 5hmC. In contrast, Wu *et al.* [8] observed salivary methylation levels in the repetitive elements LINE1 and Sat2 to be significantly positively correlated with those in DNA from WBC fraction. Apart from sample size, the main differences between the study from Wu *et al.* and Godderis *et al.* are as follows: (i) the sex of the volunteers (all girls *versus* four males and 10 females, respectively), (ii) the age range (6–15 years *versus* 22–43 years), (iii) ethnicity (N-Americans with different family backgrounds *versus* Caucasians, respectively) and (iv) use of DNA from isolated WBC *versus* whole blood. As mentioned above, DNA methylation is a dynamic process that is easily affected by environmental factors and can differ between people from various age groups, physical fitness (exercise, oxidative stress), opposite sexes or different ethnicity [27,45,46]. In addition, the different observations by Wu *et al.* and Godderis *et al.* could be due to the different cell composition of the WBC fraction as compared to whole blood. Genomewide methylation analysis of whole blood *versus* subfractions has shown clear differences induced by the different cell lineages [47–50]. Moreover, tissue's cell composition can vary in response to various factors including age, gender genetic variations and health status (reviewed by ref. [12]).

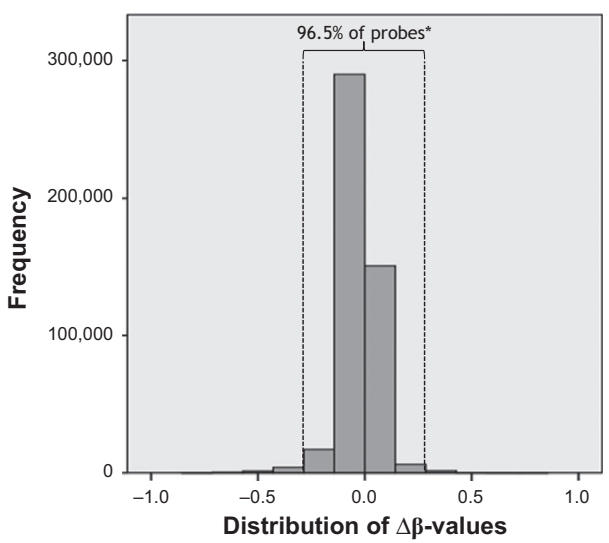

Fig. 1. The distribution of difference in DNA methylation (|Δβ|) between peripheral blood mononuclear cells and saliva from 11-year-old children. *The majority of the CpG sites (96.5%) show <20% difference in methylation between blood and saliva.

Accounting for Differences in Cell Composition

DNA methylation patterns vary between different tissues because specific phenotypic features are controlled by epigenetic marks and tissues are composed of different cell types that also can have different epigenetic features. Blood samples typically contain leucocytes (granulocytes, lymphocytes and monocytes), whereas saliva also contains (dead) exfoliated epithelial cells in addition to leucocytes [6,51]. Such differences in heterogeneity and viability of the cells can result in interindividual variation of the salivary as well as blood DNA methylation profiles and affect the outcome of genomewide analysis. Therefore, researchers need to be cautious when selecting biomarker candidates as they may simply reflect variable proportion of each cell type, when this aspect is not taken into consideration [52,53]. For example, varying cell composition may explain apparent age-associated differences [54] or affect differential methylation associated with inflammatory diseases [50]. Thus, cell type heterogeneity and external or internal factors affecting this heterogeneity may influence genomewide results.

Various methods have been developed to correct epigenomewide methylation data for differences in cell composition (reviewed by ref. [49]). The most widely applied method is the reference-based deconvolution method originally described by Houseman *et al.* [48]. This method permits the estimation of the proportion of various cell types within a sample based on

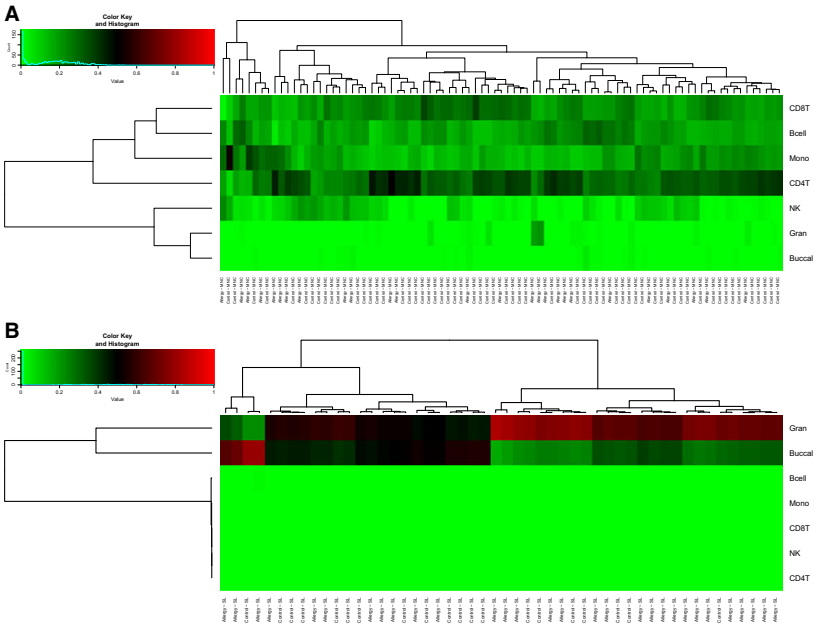

Fig. 2. Heatmaps illustrating the differences in heterogeneity between peripheral blood mononuclear cells (A) and saliva (B). Reference-based deconvolution, using (GEO GSE48472 and GEO GSE35069 as reference data sets, was applied to estimate the various cell types.

existing reference data sets, which are available from open-source databases such as the Gene Expression Omnibus (GEO) and the Genomic Data Commons Data Portal (specific for cancer genomic data sets, including The Cancer Genome Atlas (TCGA) data) [55]. For instance, for blood, several studies have analysed the methylation profile of the specific cell types present in whole blood [48–50], which can serve as reference data for correction of genomewide methylation analysis. On the contrary, detailed salivary cell composition and reference methylation profiling has not been performed systematically. Few studies have shown the presence of leucocytes, granulocytes, epithelial cells and bacterial cells in saliva [51,56], but as far as we know, none have quantified the various cell subtypes (e.g. CD8+ T-cell, CD4+ T-cell, B-cell, NK-cell, monocytes, granulocytes, buccal cells), nor studied the methylation profile of the salivary cell subtypes. To apply the Houseman deconvolution method on salivary genomewide DNA methylation data, reference methylomes from leucocyte subtypes (GEO GSE35069; [50]) were recently combined with buccal epithelial cells reference methylomes (GEO GSE46573; [57]) in a select number of studies [6,36]. When applying this, Houseman deconvolution method to the genomewide DNA methylation profiles of both salivary DNA and DNA from the PBMC fraction of 11-year-old children, saliva turned out to be less heterogenic compared to blood (fig. 2). However, there is a caveat when selecting reference data sets. This was observed in our recent methylome study [37] with saliva samples of 46 Flemish children aged 11 years, among which 26 with doctor-diagnosed respiratory allergy (having Phadiatop IgE $\geq$ 0.35 kU/L) and 20 healthy controls (no self-reported/diagnosed respiratory allergy, Phadiatop IgE $<$ 0.35 kU/L). To apply the reference-based deconvolution method described by Houseman [48], we identified two possible reference data sets for buccal cells: (i) GEO GSE46573 [57], including Illumina 450K methylation data from three replicates of the same buccal epithelial cell sample originating from a single male volunteer, and (ii) GEO GSE48472 [58], including Illumina 450K methylation data from buccal epithelial cells of five volunteers (three females/two males; aged 22–40 years). To estimate the relative proportions of each cell type in our saliva samples, we used either of the buccal reference data sets in combination with reference methylomes from leucocyte subtypes (GEO GSE35069). We observed that the choice of reference data set can significantly affect the estimated proportions of expected salivary cell types (fig. 3). Hence, it is possible that the choice of reference data has a significant impact on downstream data analysis and the discovery of differentially methylated probes. The availability of a comprehensive reference data set could help increase robustness of salivary methylation analysis. However, we are not aware of such a publicly available data set.

In the meantime, a solution for this issue could be the use of reference-free deconvolution methods (reviewed by ref. [49]). This is essentially the unsupervised adjustment of DNA methylation profiles for cell type distribution. Most of these reference-free methods give only limited information of underlying cell types, which is often hard to interpret. Houseman *et al.* [49] recently published a reference-free deconvolution approach that can partly overcome these shortcomings, providing both proportions of putative cell types based on their underlying methylomes as well as a way to evaluate to which extent the underlying profiles reflect specific cell types. To compare the application of the reference-based and reference-free method published by Houseman *et al.* on salivary DNA methylation profiles, we applied both methods on saliva

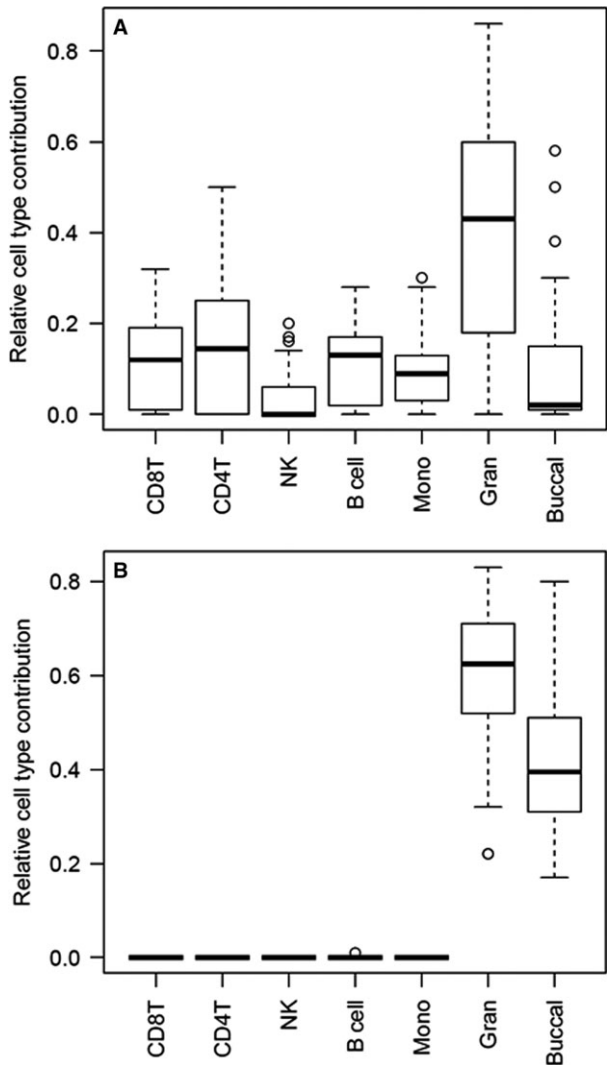

Fig. 3. The choice of reference data set can significantly affect the estimated proportions of buccal cells leucocytes. We applied the reference-based deconvolution method using reference methylomes from leucocyte subtypes (GEO GSE35069) in combination with either of the following two possible reference data sets for buccal cells: (A) GEO GSE46573, including Illumina 450K methylation data from three replicates of buccal epithelial cells from a single male volunteer [57], and (B) GEO GSE48472, including Illumina 450K methylation data from buccal epithelial cells of five volunteers [58].

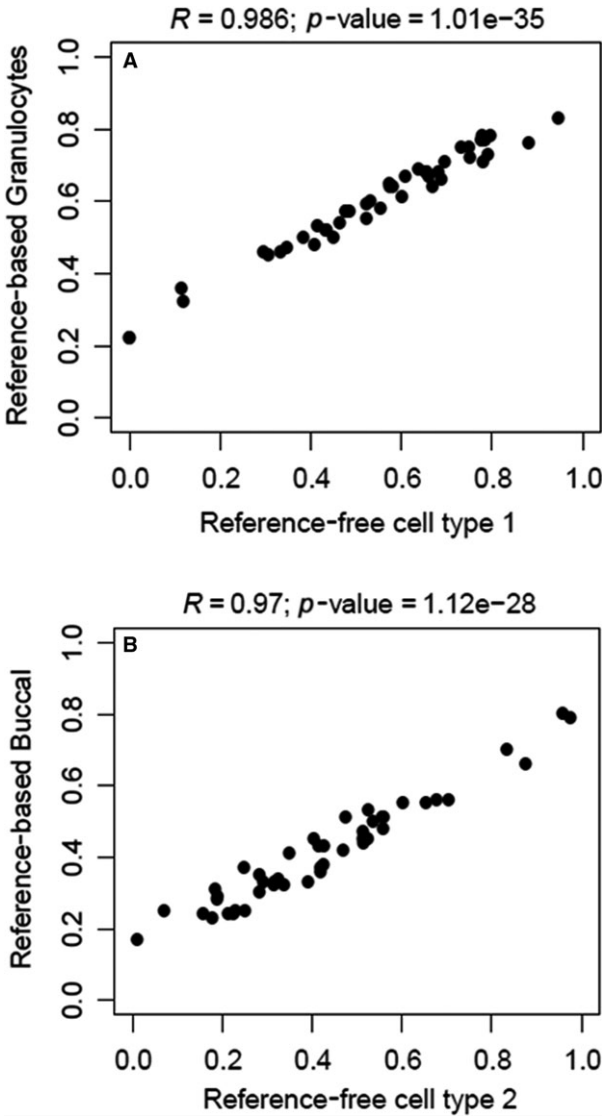

Fig. 4. Correlations between the main cell type constituents in saliva samples as estimated via the reference-free [49] and reference-based [48] (using GEO GSE35069 and GSE48472) approach. Positive correlations were observed between the contribution of reference-free cell type 1 and reference-based granulocytes (A), and between reference-free cell type 2 and reference-based buccal cells (B).

samples of 46 Flemish children aged 11 years. The reference-free method estimated an optimal of two underlying cell types in saliva, which is in accordance with the reference-based method (using GEO GSE35069 and GSE48472 as leucocyte and buccal reference methylome, respectively) where granulocytes and buccal epithelial cells were estimated to be the main cell type constituents in saliva samples. Furthermore, there were strong positive correlations between the estimated fractions of reference-free cell type 1 and reference-based granulocytes (fig. 4A; $R = 0.986$, $p < 0.001$), and between reference-free cell type 2 and reference-based buccal cells (fig. 4B; $R = 0.970$, $p < 0.001$). This suggests that the outcome of the reference-free and reference-based method is comparable in saliva samples of this particular study. It is the first time that this reference-free approach by Houseman *et al.* [49] has been applied to saliva samples and additional initiatives are warranted to confirm our observation.

Although it is still not possible to retrieve the exact biologically relevant cell types, with this new approach or any other unsupervised method, the reference-free approach can be valuable for biofluids and tissues that lack proper reference data. However, when data sets of specific tissue methylomes are available, we advise to adhere to the reference-based correction method. As such, we propose to determine cell populations in saliva aliquots and perform epigenomic profiling of specific salivary cell subpopulations to allow for better data correction and identification of robust epigenetic biomarkers.

Biomarkers of Disease

Ideally, epigenetic modifications should be studied in target tissue. Unfortunately, for most diseases, it is impossible to obtain target tissue in humans and most studies into epigenetic biomarkers for complex diseases have used blood samples. Although Wren *et al.* [9] indicated that blood and saliva molecular profiles overlap, in the end the utilization of one source over another should depend on careful consideration of target outcome, analytes and practical aspects of sample collection and processing. To date, no study has definitely confirmed that for all applications one source of DNA is superior over another.

In the field of psychiatric disorders, Smith *et al.* [6] compared saliva and blood methylomes with methylation patterns in different brain tissues. They observed the salivary methylome to be more similar to methylation patterns in each of the brain regions than methylation in blood and suggested that DNA methylation profiling using saliva may offer distinct opportunities for epidemiological and longitudinal studies of psychiatric traits. This may be due to the finding that a significant number of cells in saliva are of buccal epithelial origin, a tissue that is derived from the ectoderm – the same embryonic tissue from which the central nervous system develops. In agreement, correspondence between brain and saliva methylation profiles has been observed in young, healthy adult individuals as well as in clinical populations (reviewed by ref. [9]). Early adversity, characterized by chronic violence or neglect, has been reported to influence the methylation state of important neuropsychiatric gene targets. Using saliva samples from children (5–14 years), 2868 CpG sites were found to be differentially methylated and three differentially methylated stress/neurodevelopment-related genes were identified as significant predictors of depression (reviewed by ref. [9]). Illumina 450 K Human Methylation analysis of salivary DNA from children (7–12 years) diagnosed with ADHD revealed altered DNA methylation in *VIPR2* [59].

Recently, Langie *et al.* [36] used a case–control design to analyse and compare DNA methylation patterns in blood and saliva in individuals with respiratory allergies. When comparing respiratory allergy cases with healthy controls, 485 and 437 differentially methylated sites were identified in saliva and PBMC, respectively, of which 216 were in common and showed the same polarity in blood and saliva. Pyrosequencing analysis of three selected cg-sites confirmed the array data. In contrast, Godderis *et al.* [43] observed individuals with allergy to have slightly but significantly higher ($p = 0.042$) levels of hydroxymethylation ($0.029 \pm 0.002\%$, 0.027–0.032%) in DNA from blood compared to non-allergic participants ($0.025 \pm 0.004\%$; 0.021–0.031%), which was not observed in saliva.

Furthermore, salivary DNA has been used to identify differentially methylated genes in relation to systemic conditions, such as diabetes and a facioscapulohumeral muscular dystrophy [9]. In addition, salivary analysis has been shown to be a useful diagnostic tool in the field of cancer research [10,60,61]. Viet and Schmidt [62] used Illumina GoldenGate Methylation Arrays to study 807 cancer-associated genes in the saliva of patients with oral cancer and reported that methylation array analysis of saliva can produce a set of cancer-related genes that can be used as a composite biomarker for the early detection of oral cancer. Moreover, gene promoter methylation analysis of a test panel in salivary DNA was able to detect the early stages of head and neck squamous cell carcinoma [63,64]. Salivary DNA methylation analyses have also been applied for distant malignancies such as breast cancer [60,61]. Several salivary DNA methylation markers have been identified in breast cancer-related genes and were associated with risk factors for breast cancer development [65–67]. Overall, these data indicate that saliva is a useful source of DNA for detecting differential methylation marks in a non-invasive manner in vulnerable groups.

Table 1.
Overview of the advantages and the areas that require further research in relation to the use of saliva for DNA methylation studies.

Advantages of saliva	Areas of further research
Non-invasive collection	Evaluate the effect of the presence of bacterial DNA on global genome and genomewide DNA methylation analyses
Decentralized and multiple collections	Control probes on arrays to assess bacterial DNA contamination
Stable storage at room temperature	Full characterization of salivary cell composition
High yield and quality of high molecular weight DNA	Profiling of the methylome of salivary cell subtypes, to be used as reference data
Less heterogeneous compared to blood	
Can capture acute and chronic effects of exposures and stressors	
Useful for biomarker discovery	

Conclusion

We conclude that certain DNA methylation marks are comparable between blood and saliva, both at a gene-specific as well as at a global(hydroxyl)methylation level. Not surprisingly, there are also marked differences between both biofluids. Much attention has been focused on blood as a surrogate tissue, and this MiniReview highlights the potential of saliva as surrogate tissue for epigenetic biomarker screening.

Saliva offers key advantages for DNA methylation studies, particularly studies that enrol vulnerable populations or seek to collect repeated samples. Saliva is easy to collect, and there are few constraints to storage and processing. Several studies indicate that high-quality methylation profiles can be generated from saliva. Furthermore, differentially methylated CpG sites/gene regions were identified in salivary DNA by several independent studies, and they are promising candidate biomarkers. The vast increasing development of bioinformatics tools (including cell composition correction methods) has boosted epigenome-wide data analysis. However, there is still a need for extensive profiling of the salivary cell composition and proper reference data sets for each of the specific salivary cell subpopulations. Table 1 gives an overview of the advantages of the use of saliva for DNA methylation studies, as well as the areas that require further research. The use of different

non-invasive biospecimens, and saliva specifically, in epigenomics studies will provide new levels of insight in the molecular mechanisms through which environmental factors and interventions can alter an individual's risk of complex diseases.

Funding Statement

SL is the beneficiary of the Cefic-LRI Innovative Science Award 2013 and of a postdoctoral fellowship [12L5216N; http://www.fwo.be/] provided by The Research Foundation-Flanders (FWO) and the Flemish Institute of Technological Research (VITO). PDB is recipient of a Bill & Melinda Gates Foundation Grand Challenges Exploration Grant (OPP119403) in the field of saliva biomarker discovery, enabling collaboration with SD. WVB and KDK are supported by FWO grants G079614N and G059713N and BOF NOI/DOCPRO grants (UA). The work by LG was supported via an FWO-VITO PhD grant (grant number 11B1812N). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

1 Kalish JM, Jiang C, Bartolomei MS. Epigenetics and imprinting in human disease. Int J Dev Biol 2014;**58**:291–8.
2 Zoghbi HY, Beaudet AL. Epigenetics and human disease. Cold Spring Harb Perspect Biol 2016;**8**:a019497.
3 Goode MR, Cheong SY, Li N, Ray WC, Bartlett CW. Collection and extraction of saliva DNA for next generation sequencing. J Vis Exp 2014;Aug 27. doi: 10.3791/51697.
4 Koni AC, Scott RA, Wang G, Bailey ME, Peplies J, Bammann K *et al.* DNA yield and quality of saliva samples and suitability for large-scale epidemiological studies in children. Int J Obes (Lond) 2011;**35**(Suppl 1):S113–8.
5 Sebastian T, Cooney CG, Parker J, Qu P, Perov A, Golova JB *et al.* Integrated amplification microarray system in a lateral flow cell for warfarin genotyping from saliva. Clin Chim Acta 2014;**429**:198–205.
6 Smith AK, Kilaru V, Klengel T, Mercer KB, Bradley B, Conneely KN *et al.* DNA extracted from saliva for methylation studies of psychiatric traits: evidence tissue specificity and relatedness to brain. Am J Med Genet B Neuropsychiatr Genet 2015;**168B**:36–44.
7 Thompson TM, Sharfi D, Lee M, Yrigollen CM, Naumova OY, Grigorenko EL. Comparison of whole-genome DNA methylation patterns in whole blood, saliva, and lymphoblastoid cell lines. Behav Genet 2013;**43**:168–76.
8 Wu HC, Wang Q, Chung WK, Andrulis IL, Daly MB, John EM *et al.* Correlation of DNA methylation levels in blood and saliva DNA in young girls of the LEGACY Girls study. Epigenetics 2014;**9**:929–33.
9 Wren ME, Shirtcliff EA, Drury SS. Not all biofluids are created equal: chewing over salivary diagnostics and the epigenome. Clin Ther 2015;**37**:529–39.
10 Pfaffe T, Cooper-White J, Beyerlein P, Kostner K, Punyadeera C. Diagnostic potential of saliva: current state and future applications. Clin Chem 2011;**57**:675–87.
11 Kirschbaum C, Hellhammer DH. Salivary cortisol. In: Fink G (ed.). Encyclopedia of Stress. Academic Press, San Diego CA, 2000;379–83.
12 Zhang X, Kulasinghe A, Karim RS, Punyadeera C. Saliva diagnostics for oral diseases. In: Streckfus FC, (ed.). Advances in Salivary Diagnostics. Springer Berlin Heidelberg Berlin, Heidelberg, 2015; 131–56.
13 Vriens A, Nawrot TS, Saenen ND, Provost EB, Kicinski M, Lefebvre W *et al.* Recent exposure to ultrafine particles in school children alters miR-222 expression in the extracellular fraction of saliva. Environ Health 2016;**15**:80.
14 Salazar C, Nagadia R, Pandit P, Cooper-White J, Banerjee N, Dimitrova N *et al.* A novel saliva-based microRNA biomarker panel to detect head and neck cancers. Cell Oncol (Dordr) 2014;**37**:331–8.
15 Castagnola M, Inzitari R, Fanali C, Iavarone F, Vitali A, Desiderio C *et al.* The surprising composition of the salivary proteome of preterm human newborn. Mol Cell Proteomics 2011;**10**:M110.003 467.
16 Hong SW, Seo DG, Baik JE, Cho K, Yun CH, Han SH. Differential profiles of salivary proteins with affinity to *Streptococcus mutans* lipoteichoic acid in caries-free and caries-positive human subjects. Mol Oral Microbiol 2014;**29**:208–18.
17 Breiling A, Lyko F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. Epigenetics Chromatin 2015;**8**:24.
18 Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983;**301**:89–92.
19 Langie SA, Koppen G, Desaulniers D, Al-Mulla F, Al-Temaimi R, Amedei A *et al.* Causes of genome instability: the effect of low dose chemical exposures in modern society. Carcinogenesis 2015;**36**(Suppl 1):S61–88.
20 Ehrlich M, Lacey M. DNA hypomethylation and hemimethylation in cancer. Adv Exp Med Biol 2013;**754**:31–56.
21 Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL *et al.* Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet 2009;**5**:e1000602.
22 Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci USA 2005;**102**:10604–9.
23 Day K, Waite LL, Thalacker-Mercer A, West A, Bamman MM, Brooks JD *et al.* Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. Genome Biol 2013;**14**:R102.
24 Ong ML, Holbrook JD. Novel region discovery method for Infinium 450K DNA methylation data reveals changes associated with aging in muscle and neuronal pathways. Aging Cell 2014;**13**:142–55.
25 Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 2013;**49**:359–67.
26 Szarc vel Szic K, Declerck K, Vidakovic M, Vanden Berghe W. From inflammaging to healthy aging by dietary lifestyle choices: is epigenetics the key to personalized nutrition? Clin Epigenetics 2015;**7**:33.
27 Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. Genome Biol 2016;**17**:171.
28 Kumsta R, Hummel E, Chen FS, Heinrichs M. Epigenetic regulation of the oxytocin receptor gene: implications for behavioral neuroscience. Front Neurosci 2013;**7**:83.
29 Unternaehrer E, Luers P, Mill J, Dempster E, Meyer AH, Staehli S *et al.* Dynamic changes in DNA methylation of stress-associated genes (OXTR, BDNF) after acute psychosocial stress. Transl Psychiat 2012;**2**:e150.
30 Slowey PD. Saliva collection devices and diagnostic platforms. In: Streckfus FC, (ed.). Advances in Salivary Diagnostics. Springer Berlin Heidelberg Berlin, Heidelberg, 2015; 33–61.
31 Hansen TvO, 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 Biomark Prev 2007;**16**:2072–6.
32 Nunes AP, Oliveira IO, Santos BR, Millech C, Silva LP, Gonzalez DA *et al.* Quality of DNA extracted from saliva samples collected with the Oragene DNA self-collection kit. BMC Med Res Methodol 2012;**12**:65.

33 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**:1742–5.

34 Abraham JE, Maranian MJ, Spiteri I, Russell R, Ingle S, Luccarini C *et al.* Saliva samples are a viable alternative to blood samples as a source of DNA for high throughput genotyping. BMC Med Genomics 2012;**5**:19.

35 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**:319–26.

36 Langie SA, Szarc Vel Szic K, Declerck K, Traen S, Koppen G, Van Camp G *et al.* Whole-genome saliva and blood DNA methylation profiling in individuals with a respiratory allergy. PLoS One 2016;**11**:e0151109.

37 Langie S, Szic KSV, Van der Plas E, Declerck K, Moisse M, Koppen G *et al.* Environmental programming of respiratory allergy in childhood: the applicability of saliva. Allergy 2015;**70**: 236.

38 Iwasiow RM, Desbois A, Birnboim HC. Long-term stability of DNA from saliva samples stored in the Oragene® self-collection kit. DNA Genotek: 2011 Contract No.: PD-WP-005.

39 Anthonappa RP, King NM, Rabie AB. Evaluation of the long-term storage stability of saliva as a source of human DNA. Clin Oral Invest 2013;**17**:1719–25.

40 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**:81–5.

41 Marinus MG, Lobner-Olesen A. DNA methylation. EcoSal Plus 2014;**6**:doi:10.1128/ecosalplus.ESP-0003-2013.

42 Birnboim HC, Iwasiow RM, James CMP. Human genomic DNA content of saliva samples collected with the Oragene self-collection kit. DNA Genotek: 2008 Contract No.: PD-WP-011.

43 Godderis L, Schouteden C, Tabish A, Poels K, Hoet P, Baccarelli AA *et al.* Global methylation and hydroxymethylation in DNA from blood and saliva in healthy volunteers. BioMed Res Int 2015;**2015**:8.

44 Naumova OY, Lee M, Koposov R, Szyf M, Dozier M, Grigorenko EL. Differential patterns of whole-genome DNA methylation in institutionalized children and children raised by their biological parents. Dev Psychopathol 2012;**24**:143–55.

45 Marsit CJ. Influence of environmental exposure on human epigenetic regulation. J Exp Biol 2015;**218**:71–9.

46 Delatte B, Jeschke J, Defrance M, Bachman M, Creppe C, Calonne E *et al.* Genome-wide hydroxymethylcytosine pattern changes in response to oxidative stress. Sci Rep 2015;**5**:12714.

47 Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G *et al.* Heterogeneity in white blood cells has potential to confound DNA methylation measurements. PLoS One 2012;**7**:e46705.

48 Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012;**13**:86.

49 Houseman EA, Kile ML, Christiani DC, Ince TA, Kelsey KT, Marsit CJ. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. BMC Bioinformatics 2016;**17**:259.

50 Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D *et al.* Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One 2012;**7**:e41361.

51 Aps JK, Van den Maagdenberg K, Delanghe JR, Martens LC. Flow cytometry as a new method to quantify the cellular content of human saliva and its relation to gingivitis. Clin Chim Acta 2002;**321**:35–41.

52 Ollikainen M, Smith KR, Joo EJ, Ng HK, Andronikos R, Novakovic B *et al.* DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Hum Mol Genet 2010;**19**:4176–88.

53 Schneider E, Pliushch G, El HN, Galetzka D, Puhl A, Schorsch M *et al.* Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns. Nucleic Acids Res 2010;**38**:3880–90.

54 Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol 2014;**15**:1–9.

55 Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) 2015;**19**:A68–77.

56 Vidovic A, Vidovic Juras D, Vucicevic Boras V, Lukac J, Grubisic-Ilic M, Rak D *et al.* Determination of leucocyte subsets in human saliva by flow cytometry. Arch Oral Biol 2012;**57**: 577–83.

57 Lowe R, Gemma C, Beyan H, Hawa MI, Bazeos A, Leslie RD *et al.* Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. Epigenetics 2013;**8**:445–54.

58 Slieker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R *et al.* Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigenetics Chromatin 2013;**6**:26.

59 Wilmot B, Fry R, Smeester L, Musser ED, Mill J, Nigg JT. Methylomic analysis of salivary DNA in childhood ADHD identifies altered DNA methylation in VIPR2. J Child Psychol Psychiatry 2016;**57**:152–60.

60 Nagler RM. Saliva as a tool for oral cancer diagnosis and prognosis. Oral Oncol 2009;**45**:1006–10.

61 Delmonico L, Moreira Ados S, Franco MF, Esteves EB, Scherrer L, Gallo CV *et al.* CDKN2A (p14(ARF)/p16(INK4a)) and ATM promoter methylation in patients with impalpable breast lesions. Hum Pathol 2015;**46**:1540–7.

62 Viet CT, Schmidt BL. Methylation array analysis of preoperative and postoperative saliva DNA in oral cancer patients. Cancer Epidemiol Biomarkers Prev 2008;**17**:3603–11.

63 Ovchinnikov DA, Cooper MA, Pandit P, Coman WB, Cooper-White JJ, Keith P *et al.* Tumor-suppressor gene promoter hypermethylation in saliva of head and neck cancer patients. Transl Oncol 2012;**5**:321–6.

64 Ovchinnikov DA, Wan Y, Coman WB, Pandit P, Cooper-White JJ, Herman JG et al. DNA methylation at the novel CpG sites in the promoter of MED15/PCQAP gene as a biomarker for head and neck cancers. Biomark Insights 2014;**9**:53–60.

65 Bryan AD, Magnan RE, Hooper AE, Harlaar N, Hutchison KE. Physical activity and differential methylation of breast cancer genes assayed from saliva: a preliminary investigation. Ann Behav Med 2013;**45**:89–98.

66 Swift-Scanlan T, Smith CT, Bardowell SA, Boettiger CA. Comprehensive interrogation of CpG island methylation in the gene encoding COMT, a key estrogen and catecholamine regulator. BMC Med Genomics 2014;**7**:5.

67 Stueve TR, Wolff MS, Pajak A, Teitelbaum SL, Chen J. CYP19A1 promoter methylation in saliva associated with milestones of pubertal timing in urban girls. BMC Pediatr 2014;**14**:78.