

Chapter 6

The Use of Saliva for Genetic and Epigenetic Research



Zsafia Nemoda

Abstract Minimally invasive sampling methods for collecting oral fluids or cells are preferred in genetic research involving children and in large-scale studies where the biological sample could be mailed to the laboratory at room temperature. Saliva is an easily accessible source of cells, containing both epithelial cells exfoliating from the oral mucosa, and leukocytes filtering from blood vessels. While it does not matter which type of somatic cells are used for genetic analyses, extreme caution and proper correction are necessary in DNA methylation analyses to address the technical heterogeneity in samples with different cell composition, since the majority of the epigenetic marks are tissue specific.

High-quantity and high-quality DNA can be obtained from whole saliva samples (collected by passive drooling) or from collection devices which usually absorb cells when collecting saliva. Although salivary DNA is a mix of human and bacterial DNA, due to the species-specific amplification step in the genotyping procedures, it can be used both in candidate gene analyses and genome-wide association studies. With a careful design and appropriate additional analytical steps, salivary DNA samples can be also successfully applied in epigenetic association studies. The technical recommendations for these studies are highlighted in this chapter.

Keywords Salivary DNA sample · Genetic polymorphism · DNA methylation · Cell composition · Tissue-specificity

6.1 Introduction

Over the past decades, the number of genetic studies using saliva samples has been steadily increasing (for a comprehensive review see Sun & Reichenberger, 2014). Since all normal somatic cells contain the same genetic sequence in our body,

Z. Nemoda (✉)

Department of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Budapest, Hungary

e-mail: nemoda.zsafia@med.semmelweis-univ.hu

analyzing genetic markers from any kind of cell other than gametes should provide the same information. There are some exceptions to this general rule, for example in genetic mosaicism or in the case of foreign human cells present in some individuals (such as the donor cells after transplantation or in twins exposed to fused circulation), and in precancerous cells where the original genomic DNA (gDNA) sequence is already altered. Epigeneticists also started to use salivary specimen, based on the assumption that saliva can be as good alternative source of DNA as blood in epigenetic association studies of non-blood-based diseases. However, there are still a few technical issues (due to tissue specificity and reversibility of epigenetic signals) to be considered in this area of research (see details at the last sections).

6.2 Using Saliva as a Minimally Invasive Genetic Sample

In hospital-based genetic laboratories, blood samples are the preferred DNA source for a couple of reasons: (1) reliable DNA yield with high quantity and quality, (2) more experience has been accumulated, (3) blood sampling can be easily combined with other (routine clinical) blood tests, and automatic sample processing is often available at hospital centers (reducing workload). However, in field studies where sample storage and processing cannot be easily achieved, or at participants' homes—where medical assistance for blood drawing is difficult to manage—saliva can serve as an ideal surrogate tissue for genetic analyses. Noninvasive sampling methods of collecting cells from the mouth are also preferred in research involving children, the elderly, and healthy nonclinical participants, especially in large-scale studies where biological samples could be mailed to the laboratory at ambient temperature. Table 6.1 summarizes the fields where saliva sampling is preferred. Since 1 ml of blood or saliva contains similar number of cells (~half a million), comparable amounts of DNA can be extracted from both (Sun & Reichenberger, 2014). However, blood samples have greater volume (typically ~8 ml) than saliva samples (maximum 4 ml, usually 1–2 ml, see Chap. 3). Therefore, the total DNA yield from blood samples is mostly higher.

Saliva is an easily accessible source of cells, containing both epithelial cells exfoliating from the oral mucosa and leukocytes filtrating from blood vessels, as it has been shown via microscopic observations (see schematic representation at Fig. 6.1 upper part). The approximate ratio of the different cells was demonstrated by genetic analyses of oral samples from patients after allogeneic blood stem cell transplantation using informative DNA markers to differentiate between the host's epithelial cells and donor's blood cells (Thiede, Prange-Krex, Freiberg-Richter, Bornhauser, & Ehninger, 2000). Nevertheless, a major drawback of oral biospecimens is that bacterial and fungal or viral contamination cannot be avoided, which can cause problems if bacterial deoxyribonuclease (DNase) enzymes degrade the human DNA. However, this issue can be resolved simply with proper design. For example, rinsing the mouth and avoid eating for at least half an hour prior to saliva collection reduces contamination. If the sample has to remain at room temperature

Table 6.1 Types of genetic analyses when saliva samples are preferred

Number of subjects	Type of analyses
Individual	<ul style="list-style-type: none"> ✓ At-home genetic tests (e.g., ancestry tests, familial genetic risk tests) ✓ Detection of genetic mutations in oral precancerous cells (sputum is used for detection of respiratory tract diseases) ~ Forensic tests (depends on the available tissue) ∅ Genetic tests of blood cells (e.g., leukemia) ∅ Clinical diagnosis of genetic disorders (lower failure rate is required)
Group-level	<ul style="list-style-type: none"> ✓ Field studies (when resources are limited), home visits (blood drawing is not easily manageable) ~ Genetic research conducted with children or the elderly in a clinical setting (additional blood drawing can be avoided by saliva sampling) ∅ Genetic research with adult patients in hospital settings (when blood is drawn anyway for other laboratory tests)
Population-based	<ul style="list-style-type: none"> ✓ Epidemiological studies, especially when samples are mailed ~ Biobanking samples of the general population (when higher consent rate is important, saliva can substitute blood samples) ∅ Genetic studies of blood-based (e.g., immunological) diseases

✓ Saliva is preferred (because of easier access or lower cost)

~ Saliva and blood-based samples are equally accepted

∅ Blood-based samples are preferred

for a longer time, a preservative can be added. There are several commercially available kits providing convenient protocols for saliva collection when the stabilizing solution is released into the sample after closing the device. This solution is a lysis buffer containing the ionic detergent sodium dodecyl sulfate at a proper pH and salt concentration that is optimal for DNA storage (i.e., removing divalent cations, which are required for DNase activity). Since all kinds of cells, either human or microbial, are disrupted by the lysis buffer, it also renders the sample nonhazardous, which might be important for complying with biosafety regulations.

6.3 Quantity and Quality of Salivary DNA Samples

If saliva is collected in a laboratory setting, the passive drooling method is advised (see Chap. 3), and bacterial growth can be avoided without adding any chemical by storing the sample at a temperature $-20\text{ }^{\circ}\text{C}$ or less. In this way, the salivary sample can be used for multiple assays (e.g., using the supernatant for hormone or other saliva-biomarker analyses, while using the cell pellet fraction for genetic analyses). Since DNA is a stable nucleic acid, repeated freeze-thaw cycles have a negligible effect on the DNA yield (Nemoda et al., 2011). Although several freeze-thaw cycles can damage the gDNA, this fragmentation rarely results in non-amplifiable sample (Digestion by bacterial DNase enzymes imposes a bigger risk if the sample is not stored properly). In addition, storage for extended time periods reduces the DNA yield but does not affect substantially the usability of gDNA in genotyping

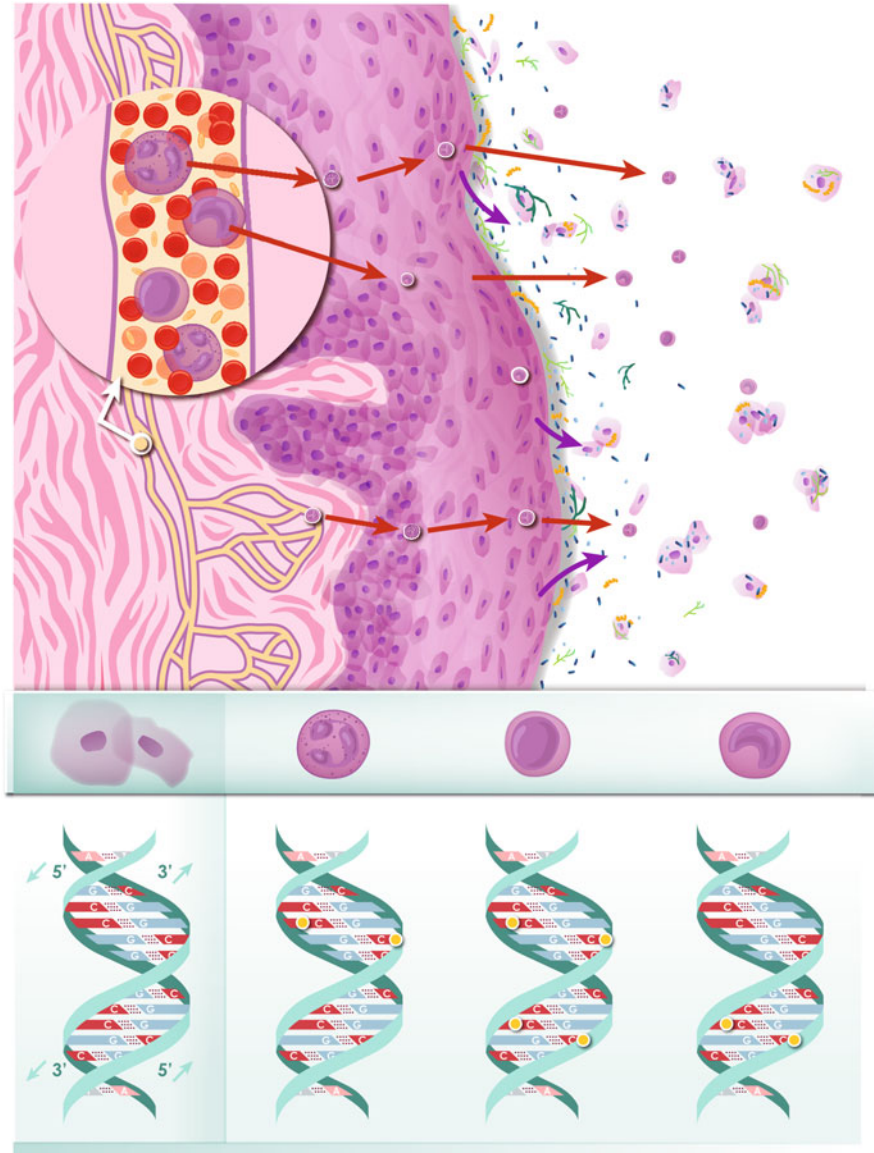


Fig. 6.1 Cell types and their possible epigenetic modifications in saliva. Top: Structure of the nonkeratinized buccal epithelium. Cells are shown in purple as they appear in the histological hematoxylin & eosin staining. The underlying collagen fibers are indicated by the pink stripes, and the capillary branches are represented in purple with yellow symbolizing plasma. A magnified section of a blood vessel demonstrates the different cell types in blood (beside the most abundant red blood cells and tiny thrombocytes, nucleus containing leukocytes are shown). To achieve their immunological functions, leukocytes can exit blood vessels (red arrows) and transverse through the epithelial cell layers. Hence the cell composition of saliva is quite heterogenous, which is shown on the right side. Based on microscopic observations and DNA methylation data of salivary samples, the two major cell types are buccal epithelial cells (large, pink cells) and granulocytes (smaller, purple cells with segmented nuclei). The different types of bacteria are represented by light blue

(Durdiaková, Kamodyová, Ostatníková, Vlková, & Celec, 2012). It is to note that routine DNA concentration measurements (using UV absorbance at 260 nm) are based only on the chemical properties of the nucleic acids. Thus, the integrity of high-molecular weight gDNA should be checked by other methods (e.g., gel electrophoresis, see Rethmeyer, Tan, Manzardo, Schroeder, & Butler, 2013) before the costly and time-consuming genetic analyses.

High-quantity and high-quality DNA can be obtained not only from whole saliva samples but also from the collection devices, that usually absorb cells present in saliva. Therefore, it is advisable to keep the collection device after the centrifugation of saliva if genetic analyses are also planned beside hormone measurements. In this way, the biological sample collection procedure can be simplified, which can be a crucial point in studies involving children. In terms of collection medium, if saliva is obtained via cotton or hydrocellulose absorbent device, most of the nucleic acid content can be recovered from the device after incubating it in cell lysis buffer (similar to DNA isolation starting from a cell pellet). However, if a synthetic swab is used, equal amounts of DNA could be isolated from the saliva filtrate and the collection device (Nemoda et al., 2011). Thus, preliminary analyses are advised to check the approximate DNA yield when a new protocol is planned with a collection device, especially when collecting saliva from infants, since much lower saliva volume can be obtained from them.

It is important to emphasize that the total amount of DNA obtained from saliva samples shows huge variability. In a review paper by Sun and Reichenberger (2014), the cited research groups report on average 20–40 µg DNA yield per ml saliva (ranging from 1 to 160 µg/ml). Larger DNA yield was observed when rubbing the tongue against the inside of the mouth before saliva collection (Nunes et al., 2012). Therefore, even though 0.1 ml of saliva is sufficient for genetic analyses (yielding approximately 1–2 µg DNA), researchers are advised to collect at least 1 ml saliva from children and adults, so that a larger number of analyses can be conducted (Usually one genotyping assay requires 10–20 ng of DNA). Better quality and higher amount of DNA (hence larger volume of saliva) is required for genome-wide association studies (GWAS), where thousands of gene variants are measured in parallel on microarrays (for a detailed review on GWAS see Stranger, Stahl, & Raj, 2011). Although the current methodologies can work with as low as 0.5 µg gDNA at



Fig. 6.1 (continued) dots, dark blue rods, and orange conglomerates, whereas fungi are illustrated by green filaments. These microorganisms live in the mucinous layer (shown in gray) and are often attached to the exfoliating epithelial cells. Therefore, genetic samples obtained from saliva contain both human and foreign DNA. Bottom: Epithelial cells (first panel) and the three main white blood cell types (granulocyte, lymphocyte, and monocyte), which can be present in saliva. The nuclei (shown in dark purple) are the source of gDNA. Epigenetic variations are shown at the bottom: The two parallel green ribbons represent the double stranded DNA with the sense sequence in 5'–3' direction. Bases are: A, adenine; C, cytosine; G, guanine; and T, thymine. Hydrogen bonds are denoted as dashed lines. Although the genetic sequence is the same in every somatic cell type of a healthy individual, epigenetic marks such as DNA methylation can show cell type-specific patterns, as illustrated by the yellow circles on the C bases

high-throughput, large-scale measurements, using DNA samples with concentration below 30–50 ng/ μ l is not recommended. Thus, the optimal protocol for saliva collection has to be selected according to the aims of each study so that enough good quality gDNA can be obtained for the genetic analyses while considering the lower end of the DNA yield range.

The quality of salivary DNA can also vary substantially depending on the DNA isolation technique (e.g., classical phenol-chloroform extraction versus silica membrane-based purification kit, Durdiaková et al., 2012). Interestingly, the ratio of human and microbial DNA can be also affected by the extraction method (Vesty, Biswas, Taylor, Gear, & Douglas, 2017). The pros and cons of the most frequently used DNA isolation methods are presented in Table 6.2. It has to be noted, that the preparatory step of cell lysis (to release DNA molecules in the sample) can also affect the quality of the DNA specimen. For example, using preloaded lysis buffer at saliva collection can be disadvantageous if the expected volume of saliva is not achieved (e.g., only 0.5 ml saliva is provided instead of the recommended 2 ml at an Oragene self-collection kit), because it can affect the efficiency of the molecular analyses (Pulford, Mosteller, Briley, Johansson, & Nelsen, 2013). Remaining chemicals, such as sodium dodecyl sulfate, phenol, or ethanol can inhibit or degrade the enzyme amplifying the DNA template in the polymerase chain reaction (PCR) (Rossen et al., 1992). While this problem could be easily detected in samples where the reaction is completely inhibited (i.e., not yielding sequence-specific amplicons), more subtle differences in amplification efficiencies could result in biases of quantitative measurements, like at DNA methylation analyses (Soriano-Tárraga et al., 2013).

The remaining organic compounds can also affect DNA concentration measurements using UV absorbance on a spectrophotometer, potentially resulting in overestimated (or confounded) DNA quantity and quality. Therefore, companies recommend using concentration measurements based on colorimetric reactions that can estimate the double stranded gDNA in salivary samples. However, this technique still does not provide precise information about the amount of useful human gDNA in a sample, since saliva always has a portion of foreign DNA due to its microbial content (even from healthy individuals). Importantly, the human/microbial DNA ratio can be estimated by real-time PCR technique using human-specific primers (for more details see methods by Nishita et al., 2009). Remarkably, varying portions of human DNA were reported by different research groups: Mean percentages of amplifiable human DNA varied between 40 and 80% in saliva samples, ranging from about 10 to 100% in most of the studies (see references by Sun & Reichenberger, 2014). Still, the human DNA yield can be kept on the higher end of this range by thoroughly rinsing the mouth with water 5–30 min before saliva collection. Notably, Hu et al. (2012) showed that salivary samples with at least 31% human-specific amplifiable DNA performed as well as blood-derived DNA samples.

In summary, oral cells can be used for various genetic analyses (see Table 6.1). Saliva can be easily collected in a broad age range, increasing the consent rate for providing biological sample, especially among healthy participants, which is a

Table 6.2 Techniques for genomic DNA isolation from saliva

Extraction method	Contamination problems ^a	Laboratory requirement	Cost	Most frequent applications Pros & Cons
Phenol-chloroform and isoamyl alcohol	Organic solvents (230 nm abs.): phenol, chloroform, alcohol	Equipment: centrifuge, fume hood Time: hours	\$	Classical laboratory experiments Pros: good for small DNA fragments & single stranded DNA Cons: working with hazardous solutions, higher chance of solvent contamination at low DNA yield samples
High-salt	Proteins (280 nm abs.), alcohol (230 nm abs.)	Equipment: centrifuge Time: hours	\$	Routine laboratory experiments Pros: cheap, relatively quick Cons: higher chance of impurity at low DNA yield samples
Silica membrane-based (solid phase)	Alcohol (230 nm abs.)	Equipment: centrifuge or vacuum manifold Time: <hour	\$\$	High-throughput, automatic processing (using plates); routine laboratory experiments (with columns in separate tubes) Pros: separation of high-molecular weight gDNA from small fragments Cons: different kits are required for the various DNA fragment sizes
Anion exchange resin based		Equipment: none (gravity) or centrifuge Time: <hour	\$\$	High-throughput, automatic processing & clinical or forensic diagnostics, when there is no need for longer DNA storage Pros: quick, efficient
Magnetic particles/beads based	Alcohol (230 nm abs.)	Equipment: magnetic rack (it can be included) Time: <hour	\$\$\$	High-throughput, automatic processing (in plates); molecular laboratory experiments (in tubes) Pros: good for small amount of DNA, as the recovery rate is high

The laboratory time and costs are presented only in relative amounts, as they can vary according to the preparatory steps (e.g., mechanical or enzymatic cell lysis procedure), number of samples, and other specific features of the genetic study. Note that alcohol contamination can affect almost every type of extraction method, since DNA is washed with 70% ethanol before the final dissolving step. Contamination can also arise from the preparatory steps, such as sodium dodecyl sulfate from the lysis buffer, which are not shown in the table

^aThe routine concentration measurement of DNA is based on the absorbance of the sample at 260 nm in a spectrophotometer, the absorbance at 230 nm and 280 nm shows the presence of organic compounds and proteins, respectively

crucial point in large-scale epidemiological studies (e.g., in follow-up analyses of specific birth cohorts). The processing of this noninvasively obtained biological specimen is similar to that of blood-based samples, yielding comparatively good quality DNA samples for a wide range of genetic analyses. Until recently, the majority of genetic and epigenetic studies used blood as a source of gDNA, hence saliva is often referred to as surrogate tissue in these studies. Although salivary DNA is a mix of human, bacterial, and fungal DNA (Fig. 6.1), due to the species-specific PCR amplification step in the genotyping procedures (for a visualized experiment see Lorenz, 2012), it can be readily used for genetic analyses. In the following sections, applications of salivary DNA samples in human genetic studies are discussed. These studies assess the sequence of human DNA which is present in every normal somatic cell (hence the source of cells does not matter in these genetic analyses). Measurements of malignant cells and DNA adducts used in oral cancer diagnostics are presented in Chap. 19, whereas Chap. 8 describes studies assessing salivary cell-free DNA (called as liquid biopsy in cancer diagnostics, see review by Siravegna, Marsoni, Siena, & Bardelli, 2017). For the usage of microbial DNA in salivary samples, see Chaps. 7 and 13.

6.4 Applications of Saliva in Genetic Analyses

There are two main types of genetic analyses where saliva samples are used: genetic tests on the individual level and genetic association studies that compare groups (see Table 6.1). Genetic tests aim to detect inherited risk factors for specific diseases (e.g., sickle cell disease, cystic fibrosis) helping diagnosis, whereas the current association studies try to reveal genetic susceptibility for developing common diseases (e.g., diabetes mellitus, hypertension, and Alzheimer's disease) by linking certain genetic variants to disease state or associated medical, physical, and psychological characteristics (e.g., blood sugar level, blood pressure, memory functions, respectively). In these studies, common gene variants—the so-called polymorphisms with allele frequencies higher than 5%—are the most often analyzed (for more information on human genetic topics, check NIH website: <https://ghr.nlm.nih.gov/primer>). Since individual genetic factors usually explain only a small portion of the heritability in complex (multifactorial) diseases, large numbers of study participants are needed to detect their modest effect (for an educational review, see Craig, 2008). Especially, GWAS require exceptionally large sample sizes, as they analyze thousands of polymorphisms in order to identify new genes which could be linked to the phenotypes of interest, without a priori hypotheses. In these large-scale epidemiological studies salivary samples became popular, since the consent rate for providing saliva is higher than for blood (Hansen, Simonsen, Nielsen, & Hundrup, 2007; Randell et al., 2016).

Genetic studies targeting children and the elderly, or with nonpersonal recruitment procedures (i.e., via mail or Internet) particularly benefit from the use of saliva. However, there are a few technical issues to consider when planning to recruit

participants providing this noninvasive biospecimen for a genetic study. For example, higher consent and return rates were reported from patients with pediatric Crohn's disease compared to controls in a pilot study by Kappelman et al. (2018): 75% of the contacted adolescent patients gave consent to their participation in the genetic study and returned saliva sample by mail, while only 44% of the sex and age matched controls gave their consent and saliva sample (expecting a gift card after the successful study enrollment). Another study investigated the effect of monetary incentive in donating biospecimen for genetic study: 43% of adult patients with inflammatory bowel disease participating in an internet-based survey gave salivary samples when 20\$ was offered, and only 26% mailed back the saliva collection kit when no compensation was offered (Randell et al., 2016). The age of the targeted population also matters, because older adults are more willing to donate saliva and send the home-collected kit by mail. In a UK study, 84% consent rate was reported among older individuals with a chronic disease compared to 59% of the contacted families with a sick child (Bhutta et al., 2013). Importantly, the prospective or retrospective nature of the recruitment procedure (i.e., calling families before or after the doctor's visit) can also impact consent rate. Where parental consent is necessary for the genetic study, contacting families before the doctor's visit in order to provide detailed information about the aims of the study is advised, and taking the saliva at the clinic personally would result in higher rate of sample donation (Bhutta et al., 2013). In sum, adult patients can be easily recruited for a genetic study via mail, especially if a telephone call is made by a specialist physician providing detailed information before the sample collection at home. In the recruitment procedure of healthy adults, a follow-up telephone call is also advised after sending the information via mail.

Nowadays, biobanks all over the world store various kinds of biological samples from patients with specific diseases, and often DNA samples (derived either from blood or saliva) of participants from the general population (see <https://biobanking.org/>). Besides the classical clinical case-control studies, geneticists investigate population-based cohorts at an increasing rate, since well-characterized subjects with data on thousands of genetic markers are a valuable research resource for association studies, such as the UK Biobank (see publications at <http://www.ukbiobank.ac.uk/genetic-publications/>). Large registries of patients and healthy individuals are building up in almost every country, where—for the sake of non-biased inclusion—saliva samples are also accepted for genetic analyses (see the All of Us Program at <https://allofus.nih.gov/>).

6.5 Comparison of Salivary and Blood Samples in Genetic Analyses

Following the spread of easily accessible and affordable genotyping methods, many research groups could try out saliva collection methods and compare the resulting DNA samples to the “gold standard” blood DNA. The simplest type of genetic variation is the Single Nucleotide Polymorphism (SNP) with only two types of alleles. It is also the most common type of genetic polymorphism (Genomes Project Consortium et al., 2015). For more details on the biological background of genetic variations see the NIH Biological Sciences Curriculum Study (2007). Since usually small fragments (100–200 base pairs) are amplified from the human gDNA during SNP genotyping (see Table 6.3), even degraded DNA samples can give reliable

Table 6.3 Types of genetic variants in the human genome and their analytic methods

Genetic variant	Variation length (bp)	Genotyping method	Amplicon length (bp)
SNP (Single Nucleotide Polymorphism), point-mutation	1	<i>Classical PCR & electrophoresis</i>	
		• Allele-specific amplification with two primer pairs	70–500
Small indel (insertion/deletion)	1–15	<i>Enzyme-based methods</i>	
		• Restriction fragment length polymorphism (RFLP)	100–500
		• Primer extension	60–150
		<i>Hybridization-based methods</i>	
		• Quantitative real-time-PCR (e.g., molecular beacon, Taqman assay)	100–150
		• SNP microarray (chip)	100–200
		<i>Sequencing methods</i>	
		• Sanger-sequencing	200–500
• Pyrosequencing	150–250		
• Next generation sequencing	100–200		
STR (Short Tandem Repeats), microsatellites	2–6	• PCR & capillary electrophoresis	100–500
		• Sequencing	100–500
VNTR (Variable Number of Tandem Repeats)	10–100	• PCR & electrophoresis	100–1000
		• Sequencing	100–500
Larger indel (insertion/deletion)	100–1000	• PCR & electrophoresis	200–1000
		• Sequencing	100–500
CNP (Copy Number Polymorphism)	1000–10,000	• Multiplex ligation-dependent probe amplification (MLPA)	70–500
Chromosomal microdeletion/microduplication	10,000–1,000,000	• Quantitative real-time-PCR	100–150
		• Multiple probes from SNP-chip	100–200

Note that chromosomal abnormalities (typically larger than a million base pairs) are not part of the list, as they are not diagnosed from saliva. Generally, these large genomic variants are determined with microscopic karyotyping or with fluorescence in situ hybridization (FISH) without PCR amplification, although current techniques apply amplification of cell-free DNA and next generation sequencing. For details on human genetic variations and their genotyping method see book chapter by Gonzalez-Bosquet and Chanock (2011)
bp base pair, PCR polymerase chain reaction

results, which is an important issue in forensic applications (reviewed by Sobrino, Brion, & Carracedo, 2005). As most of the genotyping methods include an amplification step with sequence-specific primer pair, mixed origin, low-concentration salivary DNA samples with a tiny amount (picograms) of human DNA can be used. Of course, the DNA input requirements are higher at high-throughput, multiplex PCR methods, where unbiased amplification should be achieved for multiple primer pairs.

With recent genome-wide analyses, precise estimates were gained for the accuracy of genetic analyses performed with salivary DNA samples. Approximately 99% (or higher) concordance rates have been reported with matched saliva and blood-derived DNA samples on high-density SNP-microarrays (Abraham et al., 2012; Bahlo et al., 2010; Gudiseva et al., 2016; Hu et al., 2012). Using similar, hybridization-based genotyping assays for the detection of larger Copy Number Variation (CNV), paired blood and saliva specimens were compared on chromosomal microarrays obtained from three different companies. Importantly, the bacterial content (ranging from 3 to 21%) of salivary DNA did not affect the genotyping quality of any platform used (Reiner et al., 2017), proving that saliva is a reliable alternative DNA source for genetic testing.

As for the genetic methodology, applying two types of SNP genotyping (SNP microarray and Taqman assays), Abraham et al. (2012) showed a high concordance rate (>99%) between paired blood and saliva samples in the genotype results. In addition, high-quality Sanger-sequencing could be produced from most saliva samples (Gudiseva et al., 2016). Failed genotyping was reported only for samples with DNA concentration below 10 ng/ μ l, highlighting the need for concentrating samples with low DNA yield. Fewer studies were published on saliva collection issues in connection with Short Tandem Repeats (STR) or Variable Number of Tandem Repeats (VNTR), where the length of the targeted genomic region can range from 100 to 1000 base pairs (Table 6.3). Genotyping performance of VNTRs was not influenced by either saliva or DNA sample characteristics (Nemoda et al., 2011; Nishita et al., 2009), but degradation of DNA samples can affect long-range PCR amplification (reviewed by Alaeddini, Walsh, & Abbas, 2010). In conclusion, saliva is as good a source of human cells and gDNA as blood, performing similarly in a wide range of genetic analyses. Although the human-specific portion is lower, and the risk of impurity and DNA degradation are higher in salivary samples, with a careful quality checkup step most of the problematic samples can be recognized and excluded from the analyses.

6.6 Applications of Saliva in Epigenetic Analyses

Following the disappointing results of the first wave of GWAS, the pursuit for the “missing heritability” prompted researchers to measure epigenetic variants in order to study the underlying biological mechanisms of gene–environment interactions (Manolio et al., 2009). Until recently, epigenetic analyses have been restricted to the

affected, disease-relevant tissues, since a substantial portion of the epigenetic marks is tissue specific. However, using the appropriate tissue for epigenetic association studies is often not feasible (e.g., having liver samples for metabolic diseases or brain samples for neurological disorders). Therefore, researchers started to use surrogate, easily accessible peripheral tissues, such as blood, saliva, or buccal cells. Previously, most of the epigenome-wide association studies (EWAS) used blood-derived DNA samples, although buccal cells could potentially serve as better surrogate tissue in non-blood-based diseases (Lowe et al., 2013). Importantly, saliva contains both buccal and blood cells (see Fig. 6.1); hence, it can serve as a good alternative source of gDNA in several EWAS. Although it is still questionable which peripheral tissue is more relevant for studying certain non-blood-based diseases or traits, the answers would be hopefully revealed by current bioinformatic analyses of epigenomic and transcriptomic datasets of various tissues, which are publicly available for researchers (e.g., Gene Expression Omnibus, GEO, <https://www.ncbi.nlm.nih.gov/geo/>).

Epigenetic marks create important information above the genetic sequence (*epi*—in Greek means on, above, over) which govern gene expression in multiple ways. These mechanisms are responsible for long-term regulation, switching on exclusively those genes which an individual cell requires (Almouzni & Cedar, 2016). Once the cell (and tissue) identity is established, epigenetic marks are transferred from the mother cell to the daughter cells during somatic cell divisions (contributing to the cellular memory). These marks include covalent modifications of the gDNA and the chromatin-associated histone proteins (such as acetylation, methylation, phosphorylation, and ubiquitination), controlling the accessibility of the chromatin structure. Importantly, there exists a reciprocal cross talk between these processes, therefore many studies measure only one type of epigenetic marks. Due to the stability of DNA, studying the chemical modifications of gDNA is one of the most popular epigenetic analyses, which will be discussed in this section. Analyses of other types of epigenetic mechanisms (such as chromatin structure and histone modifications) are more sensitive (requiring freshly frozen or processed samples), technically laborious, and expensive. Hence, mostly disease-specific tissues are studied by these detailed methods in cancer and infectious disease research; for these types of epigenetic analyses using saliva and oral tissues see Chaps. 9 and 19.

6.7 An Overview of Epigenetic Modifications on the DNA Molecule

The most frequent and most widely studied covalent modification in the human genome is the methylation of the cytosine base of a CpG dinucleotide which makes up about 1% of the genome (The letter “p” represents the phosphodiester bond between cytosine and guanine). The 5-methylcytosine (5mC) is often called as the fifth base of the DNA, present in 0.6–0.8% of all bases in the human genome,

depending on the developmental stage and tissue type. The majority of methylated CpG sites is located in repetitive sequences and confers repression on transposable elements (the so-called “junk” DNA). When DNA methylation occurs at crucial regulatory regions of protein-coding genes, such as promoters and enhancers, it usually correlates with gene silencing, especially at the so-called CpG islands (CpG-rich regions with high G & C base content). This transcription repression can be achieved by either directly inhibiting transcription factor binding or recruiting chromatin-modifying proteins (see review by Deaton & Bird, 2011). However, methylation at CpG islands located in gene bodies might result in the opposite effect by preventing aberrant transcription initiation events in order to guarantee the correct mRNA transcription (Neri et al., 2017). Therefore, it is important to note that an increase in DNA methylation level can either repress or enhance gene expression depending on the genetic position of the methylation change. Since the majority of previous analyses focused on promoter regions, the repressive feature of DNA methylation would be applied in the subsequent sections when discussing the role of this epigenetic process. The level of DNA methylation is usually expressed in percentages, although it is a binary signal (i.e., a certain CpG-site can be either methylated or non-methylated on the chromosome). It is due to the fact that the DNA content of a biological sample comes from a pool of cells whose DNA methylation patterns might differ substantially (see examples in Fig.6.1 lower part). Therefore, the percentage of DNA methylation is more representative of the proportion of cells in which certain CpG-sites have been methylated in order to shut down the transcription of the respective gene (according to the simplified model using the repressive feature of DNA methylation).

The 5mC can be further modified by hydroxylation catalyzed by a family of oxidases, the Ten Eleven Translocation (TET) enzymes (for a review, see Huang & Rao, 2014). Relatively high 5-hydroxymethylcytosine (5hmC) levels are reported in the brain, while other organs have various amount of this epigenetic modification (on average ~0.1% of all bases in the human genome, for details see Nestor et al., 2012). Growing evidence supports the hypothesis that 5hmC is an intermediate in active DNA demethylation processes, and it might also be involved in gene expression regulation (reviewed by Wu & Zhang, 2017). Additionally, 5hmC can be further oxidized to formyl- and carboxylcytosine, but these covalent cytosine modifications are observed at much lower rates in the genome of mature cells compared to 5mC and 5hmC (1000 and 10,000 times less, respectively). Hence, the formyl and carboxyl modifications are not measured routinely. It is noteworthy that stem cells can have a substantial level of these oxidized forms during organogenesis, as they are involved in DNA demethylation, a process yielding activation of genes, which is needed for acquiring new cell type-specific features.

It is important to mention that, until recently, the most commonly used techniques in epigenetic analyses were based on the classical bisulfite conversion of the gDNA, which cannot distinguish between 5mC and 5hmC (see Fig. 6.2c). Therefore, the more accurate terminology for 5mC and its oxidized forms together is “modified cytosine” when referring to previous DNA methylation array or pyrosequencing data

original gDNA sequence

non-modified 5' AGCCGGGCCGCT 3' 3' TCGGCCCGGCGA 5'	methylated (m) 5' AGCC ^m GGGCC ^m GCT 3' 3' TCGG ^m CCCGG ^m CGA 5'	hydroxymethylated (h) 5' AGCC ^h GGGCC ^h GCT 3' 3' TCGG ^h CCCGG ^h CGA 5'
---	---	--

A: Affinity enrichment can detect specific chemical modifications ✓

Detection method: - sequencing → signal intensity

- microarray → ratio of modified / total (input) DNA

B: Enzyme digestion uses the CCGG genomic sites

with glucosylation pretreatment 5mC & 5hmC can be differentiated ✓

<i>HpaII</i> ✂ 5' ..CCGG.. 3' <i>MspI</i> ✂ 3' ..GGCC.. 5'	<i>HpaII</i> ✘ 5' ..CC ^m GG.. 3' <i>MspI</i> ✂ 3' ..GG ^m CC.. 5'	<i>HpaII</i> ✘ 5' ..CC ^{glu} GG.. 3' <i>MspI</i> ✘ 3' ..GG ^{glu} CC.. 5'
---	---	---

C: Bisulphite treatment converts not-modified C into U

in PCR amplification U is replaced with T → detection of C / T ratio

converted, PCR-amplified sequence (sense & antisense direction separately)

originally non-modified 5' AGTGGGTGTT 3' 3' TTGGTTGGTGA 5'	originally methylated (m) 5' AGT ^m GGGT ^m GTT 3' ! 3' TTGG ^m CTTGG ^m CGA 5'	hydroxymethylated (h) 5' AGT ^h GGGT ^h GTT 3' 3' TTGG ^h CTTGG ^h CGA 5'
---	--	--

The 5mC & 5hmC can be differentiated with additional reactions:

- oxidative pretreatment with KRuO_4 → oxidative bisulfite sequencing (oxBS-Seq)
- glucosylation & TET-mediated oxidation → TET-assisted bisulfite sequencing (TAB-Seq)

Fig. 6.2 Techniques for detection of epigenetic modifications in human DNA. The non-modified cytosine base (C) can be methylated (m) in a CpG dinucleotide sequence (underlined), which can be further modified (hydroxymethylation: h). For details on the biochemical processes creating 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) see the overview of epigenetic modifications. (a) In the affinity enrichment methods, specific antibodies are applied in order to differentiate 5mC and 5hmC and also distinguish them from formyl- and carboxylcytosine. (b) Specific restriction enzymes can also differentiate covalently modified C from non-modified C. The scissors show when the CCGG sequence can be cut (∇ indicates cleavage site), whereas the red X indicates when an enzyme cannot cut the DNA. Both *HpaII* and *MspI* enzymes cleave the CCGG sequence with a non-modified CpG-site (left panel), but *HpaII* is blocked by any kind of chemical modification (other two panels). *MspI* can cut the CCGG sequence with 5mC or 5hmC in the middle, but its cleavage is blocked by glycosylated 5hmC (right panel). Therefore, the classical enzyme pair *HpaII* and *MspI* cleavage ratio gives information about the modified C (5mC & 5hmC together)/non-modified C level, but with a prior glycosylation (catalyzed by β-glucosyltransferase using UDP-glucose) the 5hmC can be differentiated from 5mC. (c) With the classical bisulfite treatment—used as a first step in bisulfite cloning, pyrosequencing, and at Infinium Human Methylation BeadChip arrays—the non-methylated C is converted to uracil (U is replaced by T in the PCR), whereas both 5mC and 5hmC stay as C. This way, the C/T ratio readout shows the proportion of modified C (5mC & 5hmC together)/non-modified C, which is an inaccurate measure for DNA methylation in certain tissues (indicated by a red exclamation mark). Note that after bisulfite conversion, the two parallel DNA strands are not complementary anymore, therefore one strand has to be selected for amplification and subsequent bisulfite sequencing (BS-seq). Additional chemical reactions can separate 5mC and 5hmC signals. For example, the oxidative pretreatment with potassium perruthenate (KRuO_4) changes 5hmC to 5-formylcytosine, which would be

where bisulfite converted samples were used. In newly developed methods, 5mC and 5hmC signals can be differentiated with additional chemical reactions (reviewed by Nestor, Reddington, Benson, & Meehan, 2014). On Fig. 6.2, green check marks indicate techniques which are specific enough to be used for accurate DNA methylation and hydroxymethylation analyses. However, most of the accumulated EWAS results obtained from blood samples can be still regarded relevant for DNA methylation data, because normal white blood cells have negligible 5hmC levels (around 0.02% of all bases, less than 5% of modified cytosines), hence one can argue that there is no need to distinguish it from the 5mC mark. There is less data on the different DNA modifications in oral mucosa, which can affect salivary 5mC and 5hmC levels (see the different cell types in saliva at Fig. 6.1). One study reported higher global 5hmC level in saliva compared to blood (0.036 vs. 0.027%, Godderis et al., 2015), but this is still in the range of <5% of DNA modifications. Although further studies with larger sample sizes are needed for more accurate estimates of salivary 5hmC levels, it seems that the proportion of 5hmC in saliva is similar to that in blood. In conclusion, the distinction between 5mC and 5hmC can be biologically important for certain cell types such as neurons and stem cells, but marginally important for others like leukocytes and buccal cells. Therefore, in the following section, only DNA methylation studies of saliva compared to blood samples would be discussed in detail.

6.8 Epigenetic Studies Measuring Environmental Effects in Salivary Samples

Besides its crucial role in embryogenesis, where innate developmental signals elicit epigenetic changes in a highly predictable pattern, DNA methylation is also responsive to the external environmental cues (Szyf & Bick, 2013). Growing evidence shows interindividual variation in every stage of life, even after birth in monozygotic twins. These differences between monozygotic twins potentially reflect the cumulative effects of environmental exposure (Tan, Christiansen, von Bornemann Hjelmberg, & Christensen, 2015). Analyzing monozygotic twin pairs is an important model in epigenetic studies, because it enables us to rule out the effects of



Fig. 6.2 (continued) changed to U at the subsequent bisulfite conversion step, then to T in the PCR amplification (similarly as non-methylated C is converted to U, then to T). This technique is called as oxidative bisulfite sequencing (oxBS-Seq, developed by Booth et al., 2013) where the readout of 5mC is achieved, whereas for 5hmC detection the comparison of the classical BS-seq and oxBS-Seq information is needed. Another alternative method is the TET-assisted bisulfite sequencing (TAB-seq, Yu et al., 2012), which involves a β -glucosyltransferase-mediated protection of 5hmC (The big glucose moiety protects this modified C from further chemical reactions), and subsequent oxidation with recombinant TET enzyme, which turns 5mC to 5-carboxylcytosine (5caC). The following bisulfite treatment and PCR amplification would change both originally non-methylated C and 5caC (derived from 5mC) into T, whereas 5hmC is read as C

genetic polymorphisms. While DNA methylation is a dynamic and reversible process, it is the most stable epigenetic mark as it is part of the covalent structure of the DNA itself; therefore, it has the potential to serve as a biomarker. The overall stability of DNA methylation patterns has been shown in longitudinal study samples (Forest et al., 2018), although at certain CpG-sites there could be significant age-related changes (Horvath et al., 2016). Consequently, the use of DNA methylation analyses in association studies has been increasing over the last decade (for brief summaries of recent epigenetic publications visit <https://www.whatisepigenetics.com/>). These epigenetic studies are often conducted on surrogate peripheral tissues, such as blood or saliva. Since saliva is easily accessible, and sampling is less invasive than blood, a relevant direction in epigenetic research is to establish if the changes in DNA methylation patterns of saliva samples are comparable to that of the target tissue (e.g., liver or brain).

Although many studies reported high correlations in methylation levels between blood-derived and salivary DNA samples (reviewed by Langie et al., 2017), one should not forget that the majority of CpG-sites are located within repetitive sequences with high methylation levels and at CpG islands of housekeeping genes with low methylation levels, resulting in little variability within or between individuals. Hence, these sequences are not ideal for association studies measuring correlations between DNA methylation level and environmental exposure. A recent epigenome sequencing analysis demonstrated that only about 10% of the human CpG-sites showed interindividual variability, representing 2 million out of the 26.8 million autosomal CpG-sites (Hachiya et al., 2017). However, less than 2% of the total CpG-sites have been analyzed with previous array-based techniques, which measured 27–450,000 sites. The presently available DNA methylation array (EPIC BeadChip analyzing more than 850,000 sites) has increased number of CpG-sites but it is still cancer research oriented, meaning that it covers most of the human genes; it does not focus on regions which are variable between healthy individuals, and could be informative for association studies.

When using surrogate tissue for their analyses, researchers are warned that only a minority of the variable CpG-sites show correlations between the DNA methylation levels of different tissues, as it was shown for example in paired blood and brain samples (Hannon, Lunnon, Schalkwyk, & Mill, 2015). Interestingly, DNA methylation profiles of saliva samples were more similar to publicly available data of brain samples, compared to that of whole blood samples (Smith et al., 2015). Comparative DNA methylation analyses of matched brain, blood, saliva, and buccal samples showed high overall correlation between brain and peripheral tissue ($r = 0.90$ for saliva-brain, $r = 0.86$ for blood-brain, $r = 0.85$ for buccal-brain) when assessing the average methylation level at each CpG-site in a group of 21 patients undergoing brain resection (Braun et al., 2019). However, the proportion of CpG-sites showing significantly similar DNA methylation levels between the target and surrogate tissue was the highest in blood samples (20.8% compared to 17.4% in buccal and 15.1% in saliva samples). The main conclusion of this study is that the similarity of DNA methylation patterns of different tissues highly depends on the actual chromosomal region. Researchers can check the degree of cross-tissue correlation of the analyzed

CpG-sites on the study website (Iowa Methylation Array Graphing for Experimental Comparison of Peripheral tissue & Gray matter, IMAGE-CpG, at <https://han-lab.org/methylation/default/imageCpG#>). As for blood-saliva comparisons, epigenome-wide array data of paired samples showed that 2–4% of the assayed CpG-sites were differentially methylated (Langie et al., 2017). Thus, selection of informative CpG-sites is highly recommended for EWAS to: (1) reduce the number of analyzed loci, which would be crucial for the detection of moderate-small effects given the available statistical methods; (2) focus only on those gene regions which are responsive to environmental stimuli (i.e., the variable CpG-sites); (3) select CpG-sites with good reported correlations of DNA methylation level between the surrogate and target tissue.

Based on animal and human epigenetic studies, DNA methylation mechanisms are proposed to be involved in recording early life experiences, thus influencing gene expression in order to fine-tune the activity of physiological systems. In particular, the prenatal environment, where the majority of epigenetic modifications are established, can have long-lasting effects on DNA methylation patterns. This has been shown in relation to both physical and psychosocial environmental exposure (see reviews by Marsit, 2015; Nemoda & Szyf, 2017). Based on previous epigenome-wide and targeted DNA methylation analyses, it is hypothesized that epigenetic changes involved in life-long responses to the intrauterine and early life environment are system-wide; hence, potentially detectable in multiple tissues. For example, after the pioneering animal studies, psychosocial stress evoked DNA methylation changes have been reported at the glucocorticoid receptor gene promoter in human studies using different tissues (reviewed by Turecki & Meaney, 2016). Increased methylation at the 1F promoter region of the glucocorticoid receptor gene was associated with childhood adversity in brain hippocampal samples of deceased adults (McGowan et al., 2009). It was also associated with prenatal exposure to maternal stress in newborns' cord blood, and in infants' salivary samples, although the affected CpG-sites varied (see meta-analysis by Palma-Gudiel, Cordova-Palomera, Eixarch, Deuschle, & Fananas, 2015). Therefore, saliva could be a suitable surrogate tissue in DNA methylation analyses, enabling measurement from an early age, even from early infancy.

However, other studies using blood or saliva to assess epigenetic changes caused by different intrauterine environment (e.g., birth weight discordant monozygotic twins) did not show significant differences in DNA methylation patterns of adult twin pairs (Souren et al., 2013; Tan et al., 2014). It has to be emphasized that these studies analyzed approximately 450,000 sites (using Illumina's Infinium HumanMethylation450 BeadChip array), without reducing the informative CpG-sites in their statistical analyses, as it was later suggested by Edgar, Jones, Robinson, and Kobor (2017) in their data reduction method, which lists more than 100,000 non-variable CpG-sites in both blood and buccal epithelial cells. Considering the limitations of current genome-wide studies assessing thousands of sites with potentially small individual effects, it is not surprising that none of the associations reached statistical significance. In addition, epigenetic changes triggered by early life adversity could be overshadowed by later environmental exposure (As of note, 34-

and 63-year-old adults were analyzed in the mentioned twin EWAS yielding no significant associations). Studying a younger age group and using a reduced number of CpG-sites, birth weight discordance was associated with within-pair differences of salivary DNA methylation at genes involved in neurodevelopment, as well as with differences in brain shape and size of the adolescent MZ twins (Casey et al., 2017). Using another approach to reveal biological processes, Zaghlool et al. (2018) analyzed intermediate molecular phenotypes, including blood, urinary, and salivary metabolite levels, and reported associations with DNA methylation levels at selected CpG-sites previously linked to diabetes mellitus, obesity, and smoking. Salivary tyramine metabolite, for instance, was associated with CpG-sites linked to smoking.

Finally, there are still a lot of technical issues that must be considered when using surrogate peripheral tissues for DNA methylation analyses. For example, although the type of somatic cells used for genetic analysis is irrelevant, tissue type variation and intraindividual differences in cell composition of non-sorted biological samples can hide authentic epigenetic differences. In addition, DNA methylation levels in blood and saliva samples can be affected by age, sex, and ethnicity (Horvath et al., 2016), and also by genetic variants (i.e., methylation quantitative trait loci, for details, see Do et al., 2017). Therefore, proper data processing is necessary to control for heterogeneity in samples—even when a cohort is homogenous and the biological sample type is the same throughout an epigenetic study—because different cell composition ratios can still substantially affect DNA methylation patterns. When using epigenome-wide arrays, the different proportion of leukocytes (mostly granulocytes) and buccal epithelial cells in salivary samples can be adjusted by reference-based or reference-free statistical methods (Langie et al., 2017). Cell ratios can also be assessed by measuring specific markers selected from cell type-specific CpG-lists in candidate gene analyses (Eipel et al., 2016). Lastly, caution should be taken when interpreting differences in DNA methylation levels, since current laboratory methods measuring epigenetic marks can be biased. This can be a common issue when using bisulfite-converted templates due to the different chemical properties of C- and T-rich DNA strands. However, this technical problem can be easily detected with internal controls and solved by suitable correction methods (see Moskalev et al., 2011). In conclusion, with careful design and appropriate additional analytical steps, salivary DNA samples can be successfully applied in epigenetic association studies.

6.9 Future Directions and Opportunities

Based on recent (often negative) findings of GWAS, it seems that individual genetic factors linked to complex diseases or traits explain only a small proportion of the inherited component of phenotypic variance. To improve the ability to detect moderate effects, researchers in medical genetic fields are aiming at: (1) increasing sample size in specific GWAS cohorts and pool samples in international consortia which would allow for conducting meta- and mega-analyses to identify genetic

variants with small effects; (2) applying more precise intermediate or endophenotypes, which are influenced by fewer genetic variants (Blanco-Gómez et al., 2016); (3) studying gene–gene and gene–environment interactions in order to reveal the “missing heritability” (Manolio et al., 2009). The use of saliva as a biospecimen seems valuable to these goals. Genetic studies over the last decade have shown that saliva is a reliable source to study inherited genetic variants present in every somatic cell of an individual. Moreover, because sequencing the coding gene regions (i.e., exome) or the whole genome in large patient cohorts is now a reality, studying rare genetic variants in the pathomechanisms of complex inheritance diseases became possible, thus supplementing present GWAS that measure common genetic polymorphisms. Importantly, the bioinformatic and statistical tools dealing with this enormous data have to be constantly updated in research laboratories. Fortunately, the research community is providing free program packages which can be easily applied (e.g., the R Project for Statistical Computing at <https://www.r-project.org/>).

In the wake of technical developments, the number of studies using saliva is likely to increase exponentially both in genetic and epigenetic analyses. However, caution is needed in DNA methylation studies due to numerous technical issues (Langie et al., 2017). In order to overcome the various biological and statistical challenges, improvement of bioinformatic analyses is continuously needed in this area of research. Fortunately, open access to publications has been increasing. These include databases of analytical procedures (e.g., European Bioinformatics Institute, <https://www.ebi.ac.uk/>) and genome-wide datasets (such as dbGAP, <https://www.ncbi.nlm.nih.gov/gap>), which are helping the research community to achieve scientific goals (see Complex Disease Epigenetics Group at <https://www.epigenomicslab.com/>). Similarly to genetic studies, the spread of sequencing methods could also widen the repertoire of analyzed CpG-sites in epigenetic association studies. Likewise, pooling different samples with the help of international consortia could facilitate the generalization of EWAS findings (Flanagan, 2015). Once the technical issues are controlled, epigenetic studies should offer great possibilities in disease prevention and management, as proposed by the developmental origins of health and disease (DOHaD) concept (Rosenfeld, 2015). This could be achieved, for instance, via longitudinal studies using epigenetic analyses at multiple time points starting with the in utero environment (assessed at birth), then in infancy and childhood, thus focusing on the most sensitive periods to adverse environmental effects. The potential outcomes of these studies would help intervention programs concentrate on specific time points. Furthermore, epigenetic changes are dynamic and could be modified by the reverse enzymatic processes later in life, potentially even in adulthood, hypothetically allowing epigenetic treatments. According to present theories of disease development, the early life environment can alter the genetically determined program to prepare the individual for the anticipated environment later in life (e.g., poor nutrition during early life would predict life-long undernutrition), prompting functional epigenetic changes. However, the adaptive responses may become maladaptive when there is an inconsistency between the anticipated and the real environments later in life, resulting in metabolic, cardiovascular, or mental

health problems (Gluckman, Hanson, Cooper, & Thornburg, 2008). Finally, although there are still lots of technical obstacles in clinical epigenetics (Aslibekyan, Claas, & Arnett, 2015), linking specific epigenetic alterations to disease-specific gene expression changes in the background of common diseases would pave the way for the development of targeted epigenetic treatments (Szyf, 2015).

Acknowledgments I thank Zsanett Botos-Hofbauer graphic designer for drawing Fig. 6.1, and my colleagues and collaborators—especially Melissa Nicolosi—for their helpful comments. I would also like to acknowledge my Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme (FP7-PEOPLE-2010-IOF N° 276107), which gave me the possibility to explore epigenetic techniques.

References

- Abraham, J. E., Maranian, M. J., Spiteri, I., Russell, R., Ingle, S., Luccarini, C., . . . Caldas, C. (2012). Saliva samples are a viable alternative to blood samples as a source of DNA for high throughput genotyping. *BMC Medical Genomics*, 5, 19. <https://doi.org/10.1186/1755-8794-5-19>
- Alaedddini, R., Walsh, S. J., & Abbas, A. (2010). Forensic implications of genetic analyses from degraded DNA—A review. *Forensic Science International Genetics*, 4(3), 148–157. <https://doi.org/10.1016/j.fsigen.2009.09.007>
- Almouzni, G., & Cedar, H. (2016). Maintenance of epigenetic information. *Cold Spring Harbor Perspectives in Biology*, 8(5), a019372. <https://doi.org/10.1101/cshperspect.a019372>
- Aslibekyan, S., Claas, S. A., & Arnett, D. K. (2015). Clinical applications of epigenetics in cardiovascular disease: The long road ahead. *Translational Research: The Journal of Laboratory and Clinical Medicine*, 165(1), 143–153. <https://doi.org/10.1016/j.trsl.2014.04.004>
- Bahlo, M., Stankovich, J., Danoy, P., Hickey, P. F., Taylor, B. V., Browning, S. R., . . . Rubio, J. P. (2010). Saliva-derived DNA performs well in large-scale, high-density single-nucleotide polymorphism microarray studies. *Cancer Epidemiology, Biomarkers & Prevention*, 19(3), 794–798. <https://doi.org/10.1158/1055-9965.EPI-09-0812>
- Bhutta, M. F., Hobson, L., Lambie, J., Scaman, E. S., Burton, M. J., Giele, H., . . . Furniss, D. (2013). Alternative recruitment strategies influence saliva sample return rates in community-based genetic association studies. *Annals of Human Genetics*, 77(3), 244–250. <https://doi.org/10.1111/ahg.12009>
- Blanco-Gómez, A., Castillo-Lluva, S., Del Mar Saez-Freire, M., Hontecillas-Prieto, L., Mao, J. H., Castellanos-Martin, A., & Perez-Losada, J. (2016). Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 38(7), 664–673. <https://doi.org/10.1002/bies.201600084>
- Booth, M. J., Ost, T. W., Beraldi, D., Bell, N. M., Branco, M. R., Reik, W., & Balasubramanian, S. (2013). Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nature Protocols*, 8(10), 1841–1851. <https://doi.org/10.1038/nprot.2013.115>
- Braun, P. R., Han, S., Hing, B., Nagahama, Y., Gaul, L. N., Heinzman, J. T., . . . Shinozaki, G. (2019). Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Translational Psychiatry*, 9, 47.
- Casey, K. F., Levesque, M. L., Szyf, M., Ismaylova, E., Verner, M. P., Suderman, M., . . . Booij, L. (2017). Birth weight discordance, DNA methylation, and cortical morphology of adolescent monozygotic twins. *Human Brain Mapping*, 38(4), 2037–2050. <https://doi.org/10.1002/hbm.23503>

- Craig, J. (2008). Complex diseases: Research and applications. *Nature Education*, 1(1), 184.
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, 25(10), 1010–1022. <https://doi.org/10.1101/gad.2037511>
- Do, C., Shearer, A., Suzuki, M., Terry, M. B., Gelernter, J., Grealley, J. M., & Tycko, B. (2017). Genetic-epigenetic interactions in cis: A major focus in the post-GWAS era. *Genome Biology*, 18(1), 120. <https://doi.org/10.1186/s13059-017-1250-y>
- Durdiaková, J., Kamodyová, N., Ostatníková, D., Vlková, B., & Celec, P. (2012). Comparison of different collection procedures and two methods for DNA isolation from saliva. *Clinical Chemistry and Laboratory Medicine*, 50(4), 643–647. <https://doi.org/10.1515/CCLM.2011.814>
- Edgar, R. D., Jones, M. J., Robinson, W. P., & Kobor, M. S. (2017). An empirically driven data reduction method on the human 450K methylation array to remove tissue specific non-variable CpGs. *Clinical Epigenetics*, 9, 11. <https://doi.org/10.1186/s13148-017-0320-z>
- Eipel, M., Mayer, F., Arent, T., Ferreira, M. R., Birkhofer, C., Gerstenmaier, U., ... Wagner, W. (2016). Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. *Aging*, 8(5), 1034–1048. <https://doi.org/10.18632/aging.100972>
- Flanagan, J. M. (2015). Epigenome-wide association studies (EWAS): Past, present, and future. *Methods in Molecular Biology*, 1238, 51–63. https://doi.org/10.1007/978-1-4939-1804-1_3
- Forest, M., O'Donnell, K. J., Voisin, G., Gaudreau, H., MacIsaac, J. L., McEwen, L. M., ... Greenwood, C. M. T. (2018). Agreement in DNA methylation levels from the Illumina 450K array across batches, tissues, and time. *Epigenetics*, 13(1), 19–32. <https://doi.org/10.1080/15592294.2017.1411443>
- Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., ... Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. <https://doi.org/10.1038/nature15393>
- Gluckman, P. D., Hanson, M. A., Cooper, C., & Thornburg, K. L. (2008). Effect of in utero and early-life conditions on adult health and disease. *The New England Journal of Medicine*, 359(1), 61–73. <https://doi.org/10.1056/NEJMra0708473>
- Godderis, L., Schouteden, C., Tabish, A., Poels, K., Hoet, P., Baccarelli, A. A., & Van Landuyt, K. (2015). Global methylation and hydroxymethylation in DNA from blood and saliva in healthy volunteers. *BioMed Research International*, 2015, 845041. <https://doi.org/10.1155/2015/845041>
- Gonzalez-Bosquet, J., & Chanock, S. J. (2011). Basic principles and laboratory analysis of genetic variation. *IARC Scientific Publications*, 163, 99–120.
- Gudiseva, H. V., Hansen, M., Gutierrez, L., Collins, D. W., He, J., Verkuil, L. D., ... O'Brien, J. M. (2016). Saliva DNA quality and genotyping efficiency in a predominantly elderly population. *BMC Medical Genomics*, 9, 17. <https://doi.org/10.1186/s12920-016-0172-y>
- Hachiya, T., Furukawa, R., Shiwa, Y., Ohmomo, H., Ono, K., Katsuoka, F., ... Shimizu, A. (2017). Genome-wide identification of inter-individually variable DNA methylation sites improves the efficacy of epigenetic association studies. *NPJ Genomic Medicine*, 2, 11. <https://doi.org/10.1038/s41525-017-0016-5>
- Hannon, E., Lunnon, K., Schalkwyk, L., & Mill, J. (2015). Interindividual methylomic variation across blood, cortex, and cerebellum: Implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics*, 10(11), 1024–1032. <https://doi.org/10.1080/15592294.2015.1100786>
- Hansen, T. V., Simonsen, M. K., Nielsen, F. C., & Hundrup, Y. A. (2007). 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 Epidemiology, Biomarkers & Prevention*, 16(10), 2072–2076. <https://doi.org/10.1158/1055-9965.EPI-07-0611>
- Horvath, S., Gurven, M., Levine, M. E., Trumble, B. C., Kaplan, H., Allayee, H., ... Assimes, T. L. (2016). An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biology*, 17(1), 171. <https://doi.org/10.1186/s13059-016-1030-0>

- Hu, Y., Ehli, E. A., Nelson, K., Bohlen, K., Lynch, C., Huizenga, P., . . . Davies, G. E. (2012). Genotyping performance between saliva and blood-derived genomic DNAs on the DMET array: A comparison. *PLoS One*, *7*(3), e33968. <https://doi.org/10.1371/journal.pone.0033968>
- Huang, Y., & Rao, A. (2014). Connections between TET proteins and aberrant DNA modification in cancer. *Trends in Genetics*, *30*(10), 464–474. <https://doi.org/10.1016/j.tig.2014.07.005>
- Kappelman, M. D., Lange, A., Randell, R. L., Basta, P. V., Sandler, R. S., Laugesen, K., . . . Erichsen, R. (2018). Feasibility of salivary DNA collection in a population-based case-control study: A pilot study of pediatric Crohn's disease. *Clinical Epidemiology*, *10*, 215–222. <https://doi.org/10.2147/CLEP.S143322>
- Langie, S. A. S., Moisse, M., Declerck, K., Koppen, G., Godderis, L., Vanden Berghe, W., . . . De Boever, P. (2017). Salivary DNA methylation profiling: Aspects to consider for biomarker identification. *Basic & Clinical Pharmacology & Toxicology*, *121*(Suppl 3), 93–101. <https://doi.org/10.1111/bcpt.12721>
- Lorenz, T. C. (2012). Polymerase chain reaction: Basic protocol plus troubleshooting and optimization strategies. *Journal of Visualized Experiments*, *63*, e3998. <https://doi.org/10.3791/3998>
- Lowe, R., Gemma, C., Beyan, H., Hawa, M. I., Bazeos, A., Leslie, R. D., . . . Ramagopalan, S. V. (2013). Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics*, *8*(4), 445–454. <https://doi.org/10.4161/epi.24362>
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., . . . Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747–753. <https://doi.org/10.1038/nature08494>
- Marsit, C. J. (2015). Influence of environmental exposure on human epigenetic regulation. *The Journal of Experimental Biology*, *218*(Pt 1), 71–79. <https://doi.org/10.1242/jeb.106971>
- McGowan, P. O., Sasaki, A., D'Alessio, A. C., Dymov, S., Labonte, B., Szyf, M., . . . Meaney, M. J. (2009). Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience*, *12*(3), 342–348. <https://doi.org/10.1038/nn.2270>
- Moskalev, E. A., Zavgorodnij, M. G., Majorova, S. P., Vorobjev, I. A., Jandaghi, P., Bure, I. V., & Hoheisel, J. D. (2011). Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression. *Nucleic Acids Research*, *39*(11), e77. <https://doi.org/10.1093/nar/gkr213>
- National Institutes of Health (Bethesda, MD, US) Biological Sciences Curriculum Study. (2007). NIH Curriculum Supplement Series [Internet] Understanding Human Genetic Variation. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK20363/>
- Nemoda, Z., Horvat-Gordon, M., Fortunato, C. K., Beltzer, E. K., Scholl, J. L., & Granger, D. A. (2011). Assessing genetic polymorphisms using DNA extracted from cells present in saliva samples. *BMC Medical Research Methodology*, *11*, 170. <https://doi.org/10.1186/1471-2288-11-170>
- Nemoda, Z., & Szyf, M. (2017). Epigenetic alterations and prenatal maternal depression. *Birth Defects Research*, *109*(12), 888–897. <https://doi.org/10.1002/bdr2.1081>
- Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., . . . Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, *543*(7643), 72–77. <https://doi.org/10.1038/nature21373>
- Nestor, C. E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., . . . Meehan, R. R. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Research*, *22*(3), 467–477. <https://doi.org/10.1101/gr.126417.111>
- Nestor, C. E., Reddington, J. P., Benson, M., & Meehan, R. R. (2014). Investigating 5-hydroxymethylcytosine (5hmC): The state of the art. *Methods in Molecular Biology*, *1094*, 243–258. https://doi.org/10.1007/978-1-62703-706-8_19
- Nishita, D. M., Jack, L. M., McElroy, M., McClure, J. B., Richards, J., Swan, G. E., & Bergen, A. W. (2009). Clinical trial participant characteristics and saliva and DNA metrics. *BMC Medical Research Methodology*, *9*, 71. <https://doi.org/10.1186/1471-2288-9-71>
- Nunes, A. P., Oliveira, I. O., Santos, B. R., Millech, C., Silva, L. P., Gonzalez, D. A., . . . Barros, F. C. (2012). Quality of DNA extracted from saliva samples collected with the Oragene DNA

- self-collection kit. *BMC Medical Research Methodology*, 12, 65. <https://doi.org/10.1186/1471-2288-12-65>
- Palma-Gudiel, H., Cordova-Palomera, A., Eixarch, E., Deuschle, M., & Fananas, L. (2015). Maternal psychosocial stress during pregnancy alters the epigenetic signature of the glucocorticoid receptor gene promoter in their offspring: A meta-analysis. *Epigenetics*, 10(10), 893–902. <https://doi.org/10.1080/15592294.2015.1088630>
- Pulford, D. J., Mosteller, M., Briley, J. D., Johansson, K. W., & Nelsen, A. J. (2013). Saliva sampling in global clinical studies: The impact of low sampling volume on performance of DNA in downstream genotyping experiments. *BMC Medical Genomics*, 6, 20. <https://doi.org/10.1186/1755-8794-6-20>
- Randell, R. L., Gulati, A. S., Cook, S. F., Martin, C. F., Chen, W., Jaeger, E. L., . . . Kappelman, M. D. (2016). Collecting biospecimens from an internet-based prospective cohort study of inflammatory bowel disease (CCFA partners): A feasibility study. *JMIR Research Protocols*, 5(1), e3. <https://doi.org/10.2196/resprot.5171>
- Reiner, J., Karger, L., Cohen, N., Mehta, L., Edelmann, L., & Scott, S. A. (2017). Chromosomal microarray detection of constitutional copy number variation using saliva DNA. *The Journal of Molecular Diagnostics*, 19(3), 397–403. <https://doi.org/10.1016/j.jmoldx.2016.11.006>
- Rethmeyer, J. A., Tan, X., Manzardo, A., Schroeder, S. R., & Butler, M. G. (2013). Comparison of biological specimens and DNA collection methods for PCR amplification and microarray analysis. *Clinical Chemistry and Laboratory Medicine*, 51(5), e79–e83. <https://doi.org/10.1515/cclm-2012-0429>
- Rosenfeld, C. S. (Ed.). (2015). *The epigenome and developmental origins of health and disease*. London: Academic. ISBN: 9780128016725.
- Rossen, L., Norskov, P., Holmstrom, K., & Rasmussen OF. (1992). Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions. *International Journal of Food Microbiology*, 17(1), 37–45.
- Siravegna, G., Marsoni, S., Siena, S., & Bardelli, A. (2017). Integrating liquid biopsies into the management of cancer. *Nature Reviews Clinical Oncology*, 14(9), 531–548. <https://doi.org/10.1038/nrclinonc.2017.14>
- Smith, A. K., Kilaru, V., Klengel, T., Mercer, K. B., Bradley, B., Conneely, K. N., . . . Binder, E. B. (2015). DNA extracted from saliva for methylation studies of psychiatric traits: Evidence tissue specificity and relatedness to brain. *American Journal of Medical Genetics Part B, Neuropsychiatric Genetics*, 168B(1), 36–44. <https://doi.org/10.1002/ajmg.b.32278>
- Sobrino, B., Brion, M., & Carracedo, A. (2005). SNPs in forensic genetics: A review on SNP typing methodologies. *Forensic Science International*, 154(2–3), 181–194. <https://doi.org/10.1016/j.forsciint.2004.10.020>
- Soriano-Tárraga, C., Jimenez-Conde, J., Giral-Steinhauer, E., Ois, A., Rodriguez-Campello, A., Cuadrado-Godia, E., . . . GeneStroke “The Spanish Stroke Genetics Consortium”. (2013). DNA isolation method is a source of global DNA methylation variability measured with LUMA. Experimental analysis and a systematic review. *PLoS One*, 8(4), e60750. <https://doi.org/10.1371/journal.pone.0060750>
- Souren, N. Y., Lutsik, P., Gasparoni, G., Tierling, S., Gries, J., Riemenschneider, M., . . . Walter, J. (2013). Adult monozygotic twins discordant for intra-uterine growth have indistinguishable genome-wide DNA methylation profiles. *Genome Biology*, 14(5), R44. <https://doi.org/10.1186/gb-2013-14-5-r44>
- Stranger, B. E., Stahl, E. A., & Raj, T. (2011). Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, 187(2), 367–383. <https://doi.org/10.1534/genetics.110.120907>
- Sun, F., & Reichenberger, E. J. (2014). Saliva as a source of genomic DNA for genetic studies: Review of current methods and applications. *Oral Health and Dental Management*, 13(2), 217–222.

- Szyf, M. (2015). Epigenetics, a key for unlocking complex CNS disorders? Therapeutic implications. *European Neuropsychopharmacology*, 25(5), 682–702. <https://doi.org/10.1016/j.euroneuro.2014.01.009>
- Szyf, M., & Bick, J. (2013). DNA methylation: A mechanism for embedding early life experiences in the genome. *Child Development*, 84(1), 49–57. <https://doi.org/10.1111/j.1467-8624.2012.01793.x>
- Tan, Q., Christiansen, L., von Bornemann Hjelmberg, J., & Christensen, K. (2015). Twin methodology in epigenetic studies. *The Journal of Experimental Biology*, 218(Pt 1), 134–139. <https://doi.org/10.1242/jeb.107151>
- Tan, Q., Frost, M., Heijmans, B. T., von Bornemann Hjelmberg, J., Tobi, E. W., Christensen, K., & Christiansen, L. (2014). Epigenetic signature of birth weight discordance in adult twins. *BMC Genomics*, 15, 1062. <https://doi.org/10.1186/1471-2164-15-1062>
- Thiede, C., Prange-Krex, G., Freiberg-Richter, J., Bornhauser, M., & Ehninger, G. (2000). Buccal swabs but not mouthwash samples can be used to obtain pretransplant DNA fingerprints from recipients of allogeneic bone marrow transplants. *Bone Marrow Transplantation*, 25(5), 575–577. <https://doi.org/10.1038/sj.bmt.1702170>
- Turecki, G., & Meaney, M. J. (2016). Effects of the social environment and stress on glucocorticoid receptor gene methylation: A systematic review. *Biological Psychiatry*, 79(2), 87–96. <https://doi.org/10.1016/j.biopsych.2014.11.022>
- Vesty, A., Biswas, K., Taylor, M. W., Gear, K., & Douglas, R. G. (2017). Evaluating the impact of DNA extraction method on the representation of human oral bacterial and fungal communities. *PLoS One*, 12(1), e0169877. <https://doi.org/10.1371/journal.pone.0169877>
- Wu, X., & Zhang, Y. (2017). TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nature Reviews Genetics*, 18(9), 517–534. <https://doi.org/10.1038/nrg.2017.33>
- Yu, M., Hon, G. C., Szulwach, K. E., Song, C. X., Jin, P., Ren, B., & He, C. (2012). Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nature Protocols*, 7(12), 2159–2170. <https://doi.org/10.1038/nprot.2012.137>
- Zaghlool, S. B., Mook-Kanamori, D. O., Kader, S., Stephan, N., Halama, A., Engelke, R., . . . Suhre, K. (2018). Deep molecular phenotypes link complex disorders and physiological insult to CpG methylation. *Human Molecular Genetics*, 27(6), 1106–1121. <https://doi.org/10.1093/hmg/ddy006>