



Epigenetics and the burden of noncommunicable disease: a paucity of research in Africa

Epidemiological evidence suggests that an adverse *in utero* environment is associated with an increased risk for developing adult onset diseases. The molecular mechanisms for susceptibility to chronic noncommunicable diseases are not fully understood, although recent research has proposed that epigenetic modifications play an important role in fetal programming. Genetic and environmental factors contribute to interindividual and spatiotemporal tissue-specific methylation patterns. Although the diverse environments and high genetic diversity of African populations provide unparalleled potential to investigate the effects of environmental change on the epigenetic profile in humans, only a small percentage of genomic and epigenetic studies have focused on populations from this continent. This emphasizes the need to build capacity in Africa for research that leads to an understanding of the association between genetic, epigenetic and environmental risk factors for noncommunicable diseases on the continent.

Keywords: Africa • DNA methylation • epigenetics • epigenome-wide association studies

Noncommunicable disease in Africa

The burden of noncommunicable diseases (NCDs) is increasing globally and is currently the leading cause of death and disease burden worldwide. These NCDs not only pose a major challenge for healthcare systems but also to affected individuals [1]. Cardiovascular disease (CVD), cancer, chronic lung disease and diabetes result in more than 30 million deaths annually worldwide [1,2]. Previously regarded as diseases of high-income countries, NCDs are increasing at an alarming rate in low- and middle-income countries [3]. This is mainly due to demographic transitions and changing lifestyles of populations associated with urbanization. In Africa it is estimated that 40% of the population live in urban areas but by 2030, this number will exceed 50% as Africa ceases to be a predominantly rural continent. This transformation is evident by the high urban growth rate of 3.6% in sub-Saharan Africa, double the world average. This rapid urbanization has been accompanied by sig-

nificant shifts in the health patterns of South Africans, thus increasing the prevalence of NCDs [4]. Although public health in African countries has focused primarily on communicable diseases such as HIV/AIDS, malaria and tuberculosis, NCDs are now major sources of morbidity and mortality and are projected to overtake infectious diseases by 2030 [1]. NCDs are to some extent preventable by modifying risk factors such as physical inactivity, high blood cholesterol, high blood pressure, obesity, unhealthy diet and inappropriate use of tobacco and alcohol [3,4]. These lifestyle related risk factors result in various long-term disease processes that lead to high mortality rates attributable to stroke, heart attacks, cancers, obstructive lung diseases and many others [4]. However, research also highlights the importance of *in utero* and early life environment to an increased susceptibility to developing NCDs later in life. Given the burden of NCD in many African countries where healthcare resources are scarce, it is important to identify pre-

Angela Hobbs¹
& Michèle Ramsay^{*2}

¹Division of Human Genetics, National Health Laboratory Service & School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

²Division of Human Genetics, National Health Laboratory Service & School of Pathology, Faculty of Health Sciences & the Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa

*Author for correspondence:
Tel.: +27 11 717 6635
michele.ramsay@wits.ac.za

natal risk factors in an attempt to curb this growing epidemic [3,5].

The Developmental Origins of Health and Disease paradigm (DOHaD) is a multidisciplinary field that examines how 'environmental factors acting during the phase of developmental plasticity interact with genotypic variation to change the capacity of the organism to cope with its environment in later life' [6]. Since Barker proposed this model by showing that a low birth weight is associated with an increased risk of CVD later in life, this concept has been extensively studied and given rise to complementary hypotheses [7]. The thrifty-phenotype hypothesis proposes that an early life exposure to malnutrition regulates genes to operate in a manner suited for a nutritionally sub-optimal environment. Such an adaptation may improve short-term chances of survival but may be deleterious to long-term health given a more resourced environment [3,8]. Fetal metabolic programming is a concept first suggested by Barker and Hales in the early 1990s. They hypothesized that fetal and perinatal events, such as maternal over- and undernutrition, were central to determining the risk of developing chronic metabolic diseases, such as diabetes, CVD and obesity, in adulthood [9]. The theory of a mismatch between a nutritionally adverse *in utero* environment and an increased susceptibility of developing adult onset disease is of particular interest in low- and middle-income African countries. In these countries, the standard of living increases substantially within one generation [3]. People are born into nutrient scarce rural environments and move to urban areas where they are exposed to NCD risk factors (smoking, drinking, less active, unhealthy diet). Although malnutrition and starvation remain an unresolved problem in Africa, where approximately 27–51% of women of reproductive age are underweight (WHO), the prevalence of NCDs in urban areas is on the rise. The growing burden of disease in these developing economies may be underpinned by early life exposure to limited nutrition [3,4]. There is substantial evidence that supports the hypothesis that increased susceptibility to NCD is influenced by prenatal exposures [10], and some of the most striking examples of developmental plasticity have come from studies on the effects of famine.

Maternal malnutrition in humans

A study which demonstrates the long-lasting effects of an early adverse nutritional environment on health and disease susceptibility is the Dutch Hunger Winter Cohort [11]. This cohort comprises both men and women who were exposed *in utero* to the Dutch famine of 1944–1945 [11,12]. Individuals exposed to under nutrition during early gestation exhibited an increased

risk for metabolic, cardiovascular and other complex diseases [13]. In rural Gambia, there is seasonal variation in the availability of micronutrients with an alternation between the dry season (when food is plentiful) and the wet season (when there is less food available and therefore poorer nutrition). This dramatic seasonal fluctuation in nutritional status has an important impact on pregnant women. During the nutritionally poor rainy season, a high incidence of deficiencies in several essential micronutrients has been observed in pregnant women [14]. This seasonal deficiency is associated with an increased incidence of low-birth-weight babies, as well as childhood morbidity and mortality [14,15]. In 1967, a civil war broke out in Nigeria after the Igbo people of the south-eastern provinces declared their independence as the Republic of Biafra. The war was a culmination of ethnic, economic and religious tensions among the various peoples of Nigeria. Disapproving of the separation, Nigerian forces pushed the Biafra people into a small enclave and cut off food supply. This resulted in an extensive famine among the Igbo people [16]. The Biafran famine cohort consists of 1339 Igbo individuals who were born between 1965 and 1973, i.e. before, during and after the Biafran famine. Hult *et al.* studied the risks for hypertension and glucose intolerance 40 years after fetal exposure to the famine. They found that fetal and infant undernutrition is associated with significantly increased risk in 40-year-old Nigerians. This study highlights that the prevention of undernutrition during pregnancy and in infancy must become a high priority in health, education and economic agendas [3,16].

There have been several mechanisms proposed to explain the associations between *in utero* environment and adult health in human populations [17], however recent research has highlighted epigenetics as the 'link between genome and environment' capable of modulating gene transcription [18]. Reflecting on an evolutionary perspective of DOHaD, Hanson and Gluckman proposed the fetal origin hypothesis which states that fetal reprogramming, induced by *in utero* exposures, is a short-term adaptation to the anticipated environment in order to maximize the survival of the individual. These adaptations are believed to occur partly through epigenetic changes [3,19].

Epigenetic effects

Epigenetic modifications are considered to be relatively stable during development and transmissible from one cell generation to the next (mitotic inheritance). It is believed that epigenetic marks can also be transmitted down organismal generations [20], a mode of inheritance referred to as transgenerational epigenetic inheritance. Some of the earliest evidence for transgen-

erational epigenetic inheritance comes from studies in plants. One of the oldest examples involves a change in flower symmetry from bilateral to radial in *Linaria vulgaris*. This change can be explained by a change in DNA methylation rather than DNA sequence. The phenotype of the flower correlates with the degree of DNA methylation at the promoter of the *Lyc* gene, and the presence or absence of DNA methylation at the promoter correlates with its silent or active state, respectively [21]. In order for transgenerational epigenetic inheritance to occur, the reprogramming of epigenetic modifications that take place in the primordial germ cells and in the zygote, to ensure the totipotency of cells in the early embryo, must be bypassed [21]. Epigenetic modifications are influenced by several factors including age, the environment/lifestyle and disease state. New and ongoing research is uncovering the role of epigenetics in a variety of NCDs. DNA methylation, the most prominently studied epigenetic modification, is dynamic across an individual's lifetime and is affected by environmental influences. DNA methylation has been implicated in developmental disorders and cancer [22] and is of great interest to researchers as a potential link between genome, environment and disease [23,24]. This epigenetic modification is crucial for normal mammalian development. Methylation patterns are established during embryogenesis in a spatiotemporal manner and remain mostly unchanged in adult cells. Modifications to the environment during early development can lead to permanent changes in the pattern of epigenetic modifications (Figure 1) [25]. An impressive example for fetal origins of adult disease is gestational diabetes mellitus (GDM). GDM is referred to as a carbohydrate intolerance that develops during pregnancy. In GDM pregnancies the balance between increased insulin resistance and maternal insulin production is disturbed, resulting in maternal hyperglycemia, fetal hyperinsulemia and consequently fetal overnutrition [13]. There have been a number of genome-wide and candidate gene DNA methylation studies designed to highlight the effects that overnutrition *in utero* (caused from GDM) has on the DNA methylation levels in the fetus. In 2013, Ruchat *et al.* reported on the impact of GDM exposure on offspring DNA methylation levels across the genome in placental and cord blood samples [26]. Genome wide DNA methylation was measured using the Infinium HumanMethylation 450 Beadchip and an Ingenuity Pathway Analysis was done to identify metabolic pathways epigenetically affected by GDM. They observed that a large number of genes in the placenta and cord blood are differentially methylated between samples exposed or not exposed to GDM and these genes are predominantly involved in metabolic disease path-

ways. They also observed a correlation between the level of DNA methylation at 326 genes in the placenta and 117 genes in cord blood with newborn weight. Together, these findings suggest that GDM exposure does have an impact on the newborn's epigenome. Since DNA methylation is mitotically fairly stable, the DNA methylation adaptations that occur in certain genes in response to GDM may have a profound long-term phenotypic effect. This could also explain why newborns exposed to adverse intrauterine environments such as GDM have an increased risk of developing chronic diseases later in life [26]. These studies have the potential to provide new insights into disease pathogenesis as well as provide disease predictive biomarkers. Despite the success of such epigenome-wide association studies (EWAS) in identifying common disease-associated loci in humans, a large proportion of disease causality remains unexplained [27]. Although EWAS studies have become more feasible, the candidate gene approach is still valuable and relevant, specifically in a resource poor environment and continues to prove successful in identifying genes involved in disease progression [9]. This approach avoids the multiple testing problems and allows one to detect significant changes of small effect size [13]. In two separate studies, Bouchard *et al.* examined whether a newborn exposed to a GDM intrauterine environment displayed an altered DNA methylation profile in two well-known obesity and diabetes candidate genes, leptin (*LEP*) and adiponectin (*ADIPOQ*) [28,29]. The most significant finding from the earlier study was that DNA methylation levels in the leptin gene promoter region, in placental tissue only, was directly correlated with glucose levels in GDM women. In both the placental and cord blood tissue, the DNA methylation level in the leptin gene was associated with a decrease in leptin gene mRNA levels. In the more recent study, they demonstrated that DNA methylation levels were lower in the promoter of *ADIPOQ* on the fetal side of the placenta when the mother's blood glucose concentration was high during the second trimester of pregnancy. They also found that the lower DNA methylation levels were associated with higher levels of circulating adiponectin in the mother throughout pregnancy. The results from both studies showed that the DNA methylation profiles of these two genes were deregulated by exposure to maternal hyperglycemia. In a more recent study, the DNA methylation profiles of 14 metabolic programming candidate genes were analyzed in cord blood and placental samples [30]. The maternally imprinted *MEST* gene and nonimprinted glucocorticoid receptor (*NR3C1*) gene showed a significant decrease in the DNA methylation levels in GDM samples when compared with controls in both tissue types. The highest

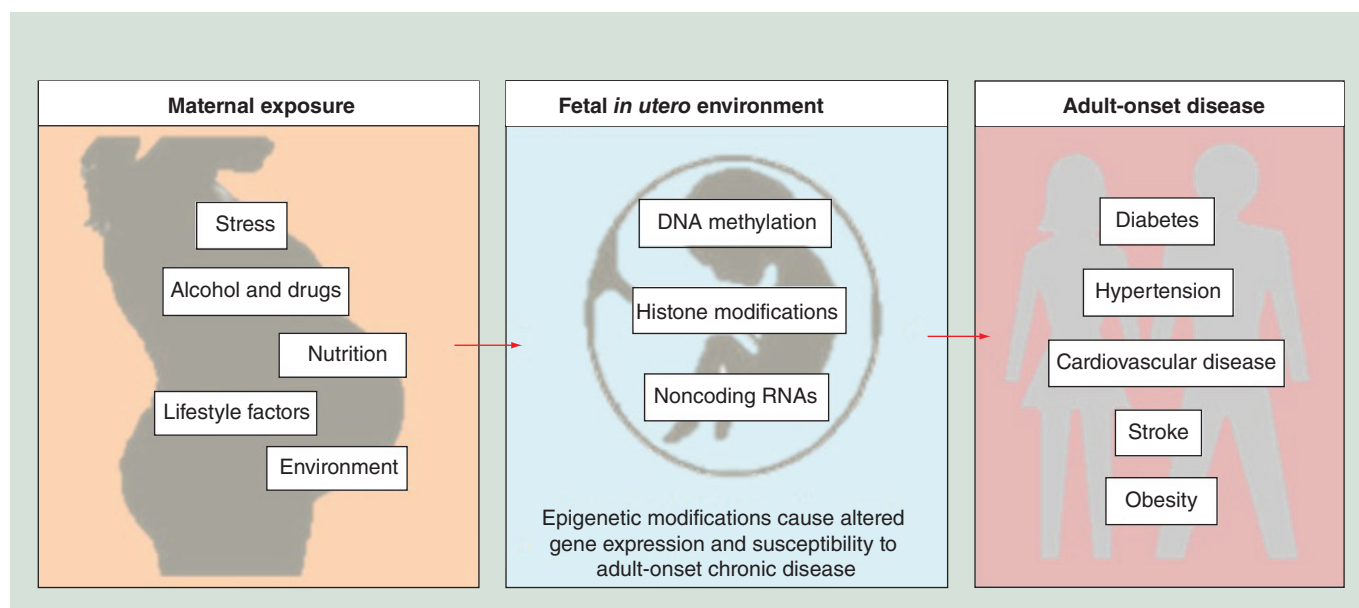


Figure 1. The link between an adverse *in utero* environment and epigenetic modifications that take place in the developing fetus and may have lifelong consequences.

DNA methylation difference between groups reached 7.2% in the placenta, which is likely to have a physiological effect [9,26].

Epigenetic effects of the maternal nutrition environment on offspring in animals

DNA methylation has been the topic of research in both animal models and human populations. Animal studies have demonstrated that nutritional factors can modify the epigenome of the developing offspring [11]. In a study done by Plagemann *et al.*, epigenetic changes in the hypothalamic proopiomelanocortin (*Pomc*) and insulin receptor (*Insr*) genes were associated with neonatal over feeding of rats [31]. The level of DNA methylation was directly dependent on the amount of glucose given to the rats. The DNA methylation pattern and expression levels of the glucocorticoid receptor (*Nr3c1*) and peroxisome proliferator-activated receptor α (*Ppara*) genes, was altered in the liver and brain tissue of pregnant rats that were fed a protein strict diet [32,33]. One of the most impressive animal model examples of the epigenetic effects of maternal nutrition on the fetus is the viable yellow agouti (A^{vy}) mouse model, in which coat color variation is correlated to epigenetic marks established early in development [34]. This mouse model has been used to investigate the impacts of nutritional and environmental influences on the fetal epigenome [34]. The agouti gene is present in all mammals and functions in the determination of coat color. In order to create a metastable epiallele in the mouse that can be switched on or off during early development, a transposable intracistern-

nal A particle (IAP) element was inserted upstream of the Agouti gene transcription start site. The degree of methylation that occurs at this IAP element correlates inversely with Agouti gene expression and hence the phenotype of the mouse. When the degree of methylation at the epiallele was increased by adding methyl donors to the mother's diet [21], the majority of the offspring appeared healthy and displayed a brown coat. This brown coat color or wild-type phenotype is the result of hypermethylation at the IAP element which suppresses Agouti gene expression. Hypomethylation of the IAP element increases Agouti expression and results in a yellow coat phenotype which is correlated with the susceptibility to metabolic diseases, cancers and obesity [11].

Epigenetic effects of the maternal nutrition environment on offspring in humans

Individuals who were exposed *in utero* to the Dutch Hunger Winter famine had, almost 60 years later, less DNA methylation at the imprinted insulin growth factor 2 (*IGF2*) gene when compared with their same-sex, unexposed siblings [12]. In 2009, Tobi *et al.* investigated the methylation levels of 15 genes implicated in growth and metabolic disease in exposed and unexposed individuals from the Dutch Hunger Winter Cohort. They observed that the methylation of the *INSIGF* was lower and methylation of *IL10*, *LEP*, *ABCA1*, *GNASAS* and *MEG3* was higher among exposed individuals when compared with their unexposed same-sex siblings. They also observed a significant interaction between the sex of the individuals and level of methylation in *INSIGF*,

LEP and *GNASAS*. Not only did exposed individuals have significantly altered levels of DNA methylation in a number of genes, they also had a higher incidence of chronic diseases such as obesity, CVD and diabetes when compared with unexposed siblings. The findings from these studies support the hypothesis that DNA methylation changes that occur in certain genes may be the result of prenatal famine exposure and that the level of methylation change depends on the sex of the exposed individual [35]. In 2010, Waterland *et al.* showed that the seasonal variation in periconceptional nutrition induces significant increases in DNA methylation at certain metastable epialleles (ME) of individuals conceived during Gambia's rainy season [14]. Some of the observed DNA methylation changes were shown to persist into early infancy [15], highlighting the permanent effect of periconceptional environment on a human epigenome [14]. These annual oscillations in nutrient availability and substrate utilization have long been known to affect fetal growth and development. To test this hypothesis in the Gambian population, Dominguez-Salas *et al.* (2014) observed the influence a mother's periconceptional dietary intake and plasma concentrations of key methyl-donor pathway substrates (methionin [MET], choline [CHOL], betain [BET]), cofactors (folate [FOL], vitamins B2, B6, B12, active B12 [ACTB12]) and intermediary metabolites (dimethyl glycine [DMG], *S*-adenosylmethionine [SAM], *S*-adenosylhomocysteine [SAH], homocysteine [HCY], cysteine [CYS]) has on their infant's DNA CpG methylation at six previously described MEs. After measuring the concentration of these 13 plasma biomarkers at the time of conception, they observed significant seasonal differences in the concentrations of eight of these biomarkers. The maternal periconceptional concentrations of FOL, B2, MET, BET and SAM were higher during the nutritionally poor rainy season and ACTB12, DMG, HCY and SAH were lower. They then tested whether these nutritionally driven seasonal differences in maternal periconceptional metabolism has an effect on the methylation profile of six known MEs. They found that methylation at these six MEs in the peripheral blood lymphocytes of infants conceived in the rainy season was consistently higher. This illustrates that maternal nutrition at the time of conception can influence the methylation patterns of MEs in offspring [36]. In 2012, Khulan *et al.* aimed to demonstrate that periconceptional maternal micronutrient supplementation might affect fetal genome-wide methylation within gene promoters. They obtained cord blood samples from offspring of Gambian mothers taking micronutrient supplementations or placebos during the pre- and periconceptional period. The global methylation patterns showed a sig-

nificant association between micronutrient supplementation and changes in DNA methylation of CpG loci. These significant changes in the epigenome in cord blood DNA were also present in infant blood DNA samples taken at 9 months, proving that a majority of these changes are persistent. These results not only highlight the importance of micronutrient supplementation during the rainy seasons and around the time of conception in Gambia but also support the idea that the nutritional environment in which a fetus develops can influence the epigenetic programming of gene activity later in life [15]. It is unknown whether any epigenetic analysis has been done on the Biafra famine cohort.

Limitations to epigenetic studies

Germline genetic variation is present and unaltered in virtually every cell in an individual whereas the epigenetic profiles (such as DNA methylation) are subject to temporal and developmental dynamics and are also influenced by environmental factors [27]. Because DNA methylation patterns are cell type and tissue specific [11,37], obtaining readily available and informative tissue is challenging. Ideally, researchers should assess epigenetic modification in tissues that contribute to the phenotype they are studying. This is difficult in epidemiological studies conducted in humans as obtaining disease relevant tissue is often invasive and/or impossible [22,38]. For this reason, surrogate tissues, most commonly whole blood, are used [39]. The assumption is that the cell types present in the surrogate tissue will reflect epigenetic modifications found in the target tissue, or at least yield biomarkers that, although not directly causative of the disease, can still be used for predictive and/or diagnostic purposes as well as provide new pathophysiological insights into the disease [27,40]. Such approaches will not explore epigenetic differences that exist between different cells and tissues [41]. Regardless of whether a target or surrogate tissue is used, the cause-and-effect relationship between epigenetic changes and disease phenotype is not a simple one. The causal link between variation in epigenetic modifications observed between an affected (case) and unaffected (control) individual and the disease (or predisposition to developing the disease) is difficult to determine and changes may simply reflect differences in cellular composition between the disease and non-disease state [27,38,42]. The disease itself could induce epigenetic changes in an individual so it is important to distinguish between causal and noncausal associations [38]. It is therefore necessary to take into account cellular heterogeneity in whole blood by using a *post hoc* bioinformatic solution for confounding cell-type

bias in EWAS [22,43]. Another challenge facing EWAS studies is sample size. The optimal size for an epigenomic study is not yet known, however, it is assumed that if the effects are moderate to small, ideally thousands of participants should be included in an EWAS study to avoid type one errors [38]. Moreover, power analysis, which is used to compute the minimum sample size for a study to give statistically significant results, is almost meaningless without realistic estimates of the degree of possible epigenetic difference between cases and controls [38]. Careful phenotyping can improve statistical power especially in studies with small-to-moderate sample sizes [44].

Epigenetic data from African countries is limited, leading to an under-representation of the epigenetic diversity in African populations. To highlight the international disparity in epigenetic NCD studies, the number of papers focusing on DNA methylation in stroke, CVD, diabetes and cancers in specific population groups (African, European, American and Asian) was quantified in the following manner using Google scholar. Search terms included 'DNA methylation' in combination with 'Africa' and 'stroke.' Each condition (stroke, CVD, diabetes and cancers) was limited to being a major subject matter. The search was repeated changing 'Africa' to 'USA' 'China' and 'Europe.' The advanced search on Google scholar was used to limit each search to those articles involving humans, having an abstract and written in English (Figure 2). Limitations of epigenetic studies that are unique to the African continent include lower levels of funding, poor infrastructure and health systems as well as a small number of trained scientists [5,45]. EWAS studies could play a pivotal role in identifying disease-associated biomarkers for screening high-risk populations and in developing strategies for disease control and treatment. The cost involved in conducting such large-scale research projects is currently high, prohibiting this type of research in countries where funding is scarce. One of the big-

gest challenges to this scale of research in the African continent is poor or unavailable health infrastructure. Hospital records tend to be poor and/or nonexistent and for many individuals, accessibility to a healthcare facility is limited. Many laboratories lack state-of-the-art facilities and equipment as well as a critical mass of skilled scientists. Ethics approval and recruitment for large-scale genomic and epigenomic studies requiring broad consenting for the sharing of data and biological material also pose challenges, as do language and cultural barriers [45–47]. On a more positive note, African populations harbor increased genetic diversity and rapidly changing environments (e.g., through urbanization) that provide a unique opportunity to explore genome–environment–epigenome relationships.

The effect of fixed genetic variation on DNA methylation

When interpreting the role of epigenetic variation in a complex disease, it is important to consider both fixed genetic variation and environmental factors. It is well known that environmental factors are a source of epigenetic variation observed between individuals [48], and there is a large amount of epidemiological data that links disease risk directly to the *in utero* environment which affects the epigenome through stable epigenetic modifications, and these resulting epigenetic marks alter the physiology to influence later disease risk [48,49]. Recent studies have highlighted that a large proportion of DNA methylation variability across individuals and populations is also attributable to underlying genetic variability [48–50]. McRae *et al.* found that the majority of the similarity in DNA methylation levels between relatives is due to genetic effects which means that approximately 20% of DNA methylation variation between individuals in a population is due to sequence-based DNA variants (SNPs) that are not located within the CpG sites. These identified SNPs, whose genotypes correlate with levels of DNA methylation, are termed methylation quantitative trait loci, or meQTLs [49]. Therefore the interindividual variation in DNA methylation is, in part, a result of nucleotide polymorphisms [48]. To determine whether these meQTLs varied with ancestral population, developmental stage and tissue types, Smith *et al.* analyzed genome-wide DNA methylation and SNP data from seven cohorts that varied by ancestry (African–American vs Caucasian), developmental stage (neonate vs adults) and tissue type (blood vs brain tissue). They found similarities in genotype-dependent DNA methylation across this diverse range of subject characteristics and tissues [51]. Bell *et al.* utilized the Illumina Human Methylation 27 Bead Chip to map associations between SNPs and methylation

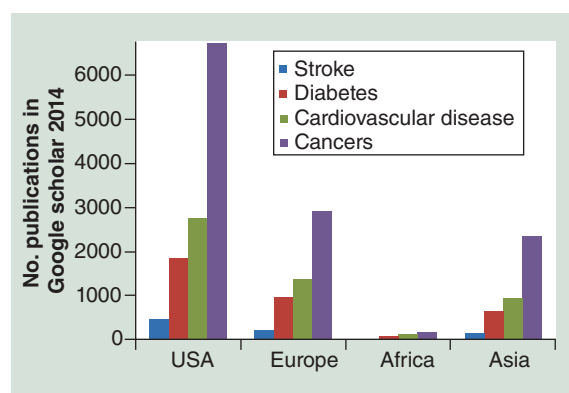


Figure 2. Number of publications based on epigenetic studies by noncommunicable disease.

levels at 22,290 CpG dinucleotides in lymphoblastoid cell lines (LCLs). They found 180 CpG sites associated with nearby SNPs [52]. In a similar study using the same DNA methylation platform, Gibbs *et al.* studied samples from four human brain regions in 150 individuals and reported hundreds of SNP-associated CpG sites in each brain sample, with meQTLs typically located very close to the associated CpG site [53]. Although many DNA methylation studies have examined genome-wide DNA methylation data through array-based or whole-genome bisulphate sequencing technologies, they have not established population-epigenetic principles to guide design, efficient statistics or interpretation [52]. Liu *et al.* examined the observed correlation in DNA methylation data, generated from an Illumina HumanMethylation450 array, on whole blood from 247 healthy European individuals, focusing particularly on the CpG sites with the largest methylation variation (vCpGs). They observed that the clustering of correlated DNA methylation at these vCpGs was similar to that of linkage disequilibrium correlation in genetic SNP variation but for shorter distances. Furthermore, they aimed to determine the relationship between these contiguous methylation clusters and genotype at a population level by examining whether methylation correlation in clusters is driven by genetic variants. In doing so, they identified that some distinct methylation clusters were associated with the same underlying SNP cluster but were not spatially defined as a single cluster because uncorrelated vCpGs existed between them. These potentially noncontiguous genetically controlled methylation clusters were termed GeMes. Given that SNPs were found to control GeMes and that clusters of methylated CpGs are smaller than LD blocks, it would be of interest to incorporate CpG-methylation data and GWAS-SNP data to better understand the functional effects of genome-wide associations [54].

Although the influences of prenatal environment on future disease risk are intensively studied, it is important to address the degree to which the environmental influences are moderated by genotype. Therefore any study aiming to explore the role of DNA methylation variation in a complex disease needs to carry out a parallel analysis of underlying genetic variation. These findings highlight the importance of understanding the genetic diversity of target populations. Africa is an important region in which to study human genetic diversity because of its complex population history and the dramatic variation in climate, diet and exposure to infectious diseases all of which result in high levels of genetic and phenotypic variation within the continent [55]. African populations are characterized by greater levels of genetic diversity when compared with

non-African populations and although Africa is critical for understanding modern human origins and genetic risk factors for disease, only a small percentage of genomic studies have focused on African populations due to a number of challenges (Table 1). Genetic studies of human disease are more challenging to perform in heterogeneous populations with higher genetic diversity. In 2009, Tishkoff *et al.*, found that the genetic structure of Africans traces to 14 ancestral population clusters that broadly correlate with ethnicity and culture or language. Given the high levels of substructure and admixture between genetically distinct populations in Africa, even within small geographic regions, the study highlighted the heterogeneity of the African populations. Therefore, there is no single African population that is representative of the diversity present in Africa [56]. Studying patterns of genetic diversity, and how they correlate with epigenetic signatures, among the multitude of ethnically diverse African populations will shed light on the questions regarding the genetic basis of phenotypic variation. However, despite the important contributions that studies of the African populations can make, these populations are vastly understudied. By comparing populations in Africa to populations of African ancestry outside Africa, such as the African-Americans, it may be possible to begin to distinguish the roles of environmental and genetic risk factors that contribute to complex diseases. There have been a number of genome-wide epigenetic studies designed to observe DNA methylation variation in African-American cohorts [57–60].

Population specificity & DNA methylation

Recent advances in high-throughput technologies for measuring quantitative locus specific and genome-wide DNA methylation have provided an opportunity to characterize methylation patterns in the context of human genome variation. DNA methylation can differ in diseases and cell types, or even between monozygotic twins, but while most research has focused on the potential for variation at the cellular level, relatively little research has been done to examine how epigenetic variation affects humans at the population level. DNA methylation patterns are important for establishing cell, tissue and organism phenotypes, but very little is known about their contribution to natural human variation [61]. As the scale of DNA methylation association studies approaches that of genome-wide association studies, issues such as population stratification have to be addressed. In 2012, Fraser *et al.* measured the DNA methylation near the transcription start sites of over 14,000 genes in 180 cell lines derived from 30 northern European family trios (mother, father and child) as well as from 30 Yoruba (west African) family trios. The

| Table 1. Challenges facing noncommunicable disease studies in Africa. | |
|---|--|
| Challenge | Possible solutions |
| Health infrastructure | |
| Poor or unavailable health infrastructure Fragmented and underfinanced healthcare systems that have not been set up for chronic disease management Hospital records poor or nonexistent Accessibility to healthcare is limited Lack of diagnostic and treatment capacity Weak referral systems | Major investments in policy, capacity building and infrastructure development in order to improve patient management as well as disease surveillance, control and prevention Develop broad partnerships/collaborations between local healthcare systems and research institutions, international organizations, the pharmaceutical industry, international organizations as well as national governments in developing countries Financial support necessary for transport of individuals to primary healthcare facilities |
| Resources | |
| Cost of large studies is expensive Infrastructure: Lack of adequate laboratories; equipment and supplies Scientific skills are scarce Data analysis expertise low | Highlight the importance of capacity building in African countries Design mentorship and training programs for African healthcare workers and scientists Identify a strategy to raise funding for large-scale studies |
| Socio-economic factors | |
| Poor basic education in many regions Language barriers High levels of poverty and under nutrition | Increase basic health education through mass media and strengthening public awareness Initiate information technology-based educational intervention programme studies – even the most poverty stricken individuals have cell phones (PROTECT Africa Study) |

data demonstrated a wide range of within population variability in the methylation of individual CpG sites. In addition to the variation within each population, they observed that a third of the genes they studied showed differences in the DNA methylation patterns between the populations. Although the methylation at over a thousand of the analyzed CpG sites is heritable, the heritability also seems to differ between populations, suggesting extensive divergence in the genetic control of DNA methylation [62]. These results suggest that DNA methylation is highly divergent between populations and this is due to a combination of genetic factors and complex gene–environment interactions. In 2013, Heyn *et al.* generated genome-wide DNA methylation profiles of three distinct human populations (namely Caucasian–American, African–American and Han Chinese–American) using the HumanMethylation450 BeadChip. They found that 439 CpG sites were significantly differentially methylated. Although many of these identified DNA methylation differences that distinguish the three major human ethnic groups were shown to be associated with underlying genetic changes (highlighting a direct relationship between the genetic and epigenetic modification), about a third stood alone as epigenetic markers and CpG methylation

quantitative trait loci associated with natural variation in the human population [61]. These data help to explain phenotype diversity beyond that which is attributable to genomic variation. There is mounting evidence that on a genome-wide scale there is substantial ethnic variation in DNA methylation, only some of which is associated with genetic changes. It is therefore likely that African populations have both genetic and environmental influences on the methylome and it will be a challenge to tease these apart in the context of chronic disease studies.

In 2014, Kwabi-Addo *et al.* conducted a genome-wide large-scale analysis of DNA methylation changes in prostate tissue from seven unaffected and three prostate cancer patients from African–American participants versus eight unaffected and three prostate cancer patients from European–American participants. Using the Human Methylation 450 Bead Chip they observed significant global DNA methylation differences in prostate tissue when comparing African–Americans to European–Americans. This finding provides a mechanistic explanation for the disease disparity [63]. In 2012, Nieminen *et al.* found distinct genetic and epigenetic signatures in colorectal cancers that were dependent on the ethnicity (Egyptian vs Finnish)

of the patients. They screened 14 tumor suppressor genes for aberrant promoter region methylation in 69 Egyptian and 80 Finnish individuals with colorectal cancer. Of these 14 genes, four (*MSS*, *CDKN2B*, *p15* and *TSGMP*) displayed a significant increased level of promoter region methylation when comparing Egyptian patients to Finnish patients. They concluded that the possible effects of environmental exposures in carcinogenesis occur through DNA methylation. This observation could have important implications for prevention, diagnosis, prognosis and treatment of the disease [64]. Although these ethnic differences in human DNA methylation in the above mentioned studies have been shown for a number of CpG sites, the genome-wide patterns and extent of these differences are largely unknown [62]. These small but extensive epigenetic differences observed between populations are most likely the result of both genetic [23] and environmental factors [62]. Adkins *et al.* found that, at birth, there are significantly different DNA methylation levels between African-Americans and Caucasians at a subset of CpG dinucleotides. The DNA methylation levels at 26,485 autosomal CpGs were assayed in 201 newborns (107 African-American and 94 Caucasian) and nonparametric analyses were performed to examine the relationship between these methylation levels and maternal parity and age, newborn gestational age, newborn gender and newborn race. They found that 13.7% (3623) of the autosomal CpGs exhibited significantly different levels of DNA methylation between African-Americans and Caucasians and that 2% of autosomal CpGs had significantly different DNA methylation levels between male and female newborns. Cancer pathway genes, including those involved in four cancers (pancreatic, prostate, bladder and melanoma) with substantial ethnic differences in incidence, were highly represented among the genes containing significant ethnic-divergent CpGs [65].

Environmental effects on DNA methylation variation

With regard to the environment, population specific environmental factors such as socio-economic status, infections and lifestyle may also contribute to differences in DNA methylation. Socio-economic status (SES) is a measure of an individual's or family's economic and social position based on education, income and occupation, and has long been a strong predictor of health. In many African countries, there exist inequalities in socio-economic status and the burden of infectious and NCD is greater among individuals of lower SES [66]. In a preliminary study, Borghol *et al.* revealed an association between exposure to a low SES in childhood and differential DNA methylation

in adulthood [67]. Forty adult males from the 1958 British Birth Cohort were selected according to their SES at childhood and in mid adulthood. The SES scores were determined using specific criteria, dividing the men into the most disadvantaged and the least disadvantaged. Genome-wide DNA methylation in the blood of these men was performed using MeDIP (methylated DNA immunoprecipitation). The DNA methylation profiles in adult blood indicated greater association with childhood SES than with adult SES. In a separate study, McGuinness *et al.*, investigated the relationship between SES and DNA methylation in a subset of individuals from the pSoBid cohort. This cohort is characterized by an extreme socio-economic and health gradient. They observed global DNA hypomethylation in the individuals classified as the 'most socio-economically deprived.' They also found an association between global DNA methylation and biomarkers of CVD and inflammation, after adjustment for socio-economic factors. Both these studies showed that there is an association between epigenetic modifications and SES. This relationship has direct implications for population health and is reflected in further associations between global DNA methylation content and emerging biomarkers of NCD [68]. Another example of an environmental stress factor that influences epigenetic modifications in humans is that of bacterial infections, a common cause of illness and death in Africa. Studies have shown that bacteria have the ability to change the chromatin structure and transcriptional activity of their host cells. This is achieved through the induction of epigenetic modifications, such as DNA methylation. These bacterial induced epigenetic modifications may affect the host cell function either to promote host defense or to allow pathogen persistence [69].

Conclusion & future perspective

De Vries and Pepper highlighted the importance of the development of genomics research 'In Africa, for Africa, by Africans.' We wish to adapt this statement to emphasize both the need for efficient collaboration and the potential global impact of studying African populations in Africa. This highlights the need to build capacity in Africa for research that leads to an understanding of the association between genetic, epigenetic and environmental risk factors for NCD on the continent [70]. Scientific interest in genomics studies in Africa is on the rise [70]. There are a few large research initiatives which specifically aim to understand the genomic and molecular underpinning of health in African populations. One such initiative is the Human Heredity and Health in Africa (H3Africa) Consortium that aims to facilitate a current research approach to

study the genomic and environmental determinants of common diseases on the continent [71]. This initiative aims to contribute to the development of the necessary expertise among African scientists, and to establish networks of African investigators. H3Africa also aims to create a network of laboratories that are well-equipped to conduct research focused on the complex interplay between environmental and genetic factors which determine disease susceptibility and drug responses in African populations [70]. It is important that genomic and epigenomic research include populations from the poorer parts of the world, including Africa [70]. Although there are challenges facing this type of research in many African countries, the information obtained on the genomes of African populations would have a positive influence on the health of individuals with African ancestry. The limited research infrastructure, including laboratories, biorepositories and databases, coupled with inadequate funding and other resources have hampered African scientists from carrying out high-quality research that focusses primarily on NCD. Building sustainable, comprehensive and multidisciplinary programs relevant to Afri-

can research is essential. Although advancing genomic research on the African continent is challenging, time consuming and cost intensive, the potential benefit of this research to improving health in Africa is unlimited. As the world moves toward epigenome-wide association studies to examine the relationship between epigenetic modifications and disease, it is important that African countries keep up with these improved technologies for high-throughput research. The use of epigenetics in a diagnostic setting and for the purpose of informing treatment is in its infancy. Implementing such tests in an African setting would require specific validation and would be challenging in many developing countries. Correct diagnosis is not only critical in a clinical setting, but also for meaningful research and eventually for determining patient outcomes. One way to improve diagnostic testing in African countries may be to introduce twinning with institutions in the more developed regions. Populations are not only genetically diverse but also have unique characteristics at an epigenetic level. Africans are exposed to many adverse environmental influences (such as poor nutrition, infection, low SES) which potentially trigger popu-

Executive summary

Noncommunicable disease in Africa

- Previously regarded as diseases of high-income countries, noncommunicable diseases (NCDs) are increasing at an alarming rate in low- and middle-income countries. This is mainly due to demographic transitions and changing lifestyles of populations associated with urbanization.
- Although public health in African countries has focused primarily on communicable diseases such as HIV/AIDS, malaria and tuberculosis, NCDs are now major sources of morbidity and mortality and are projected to overtake infectious diseases by 2030.
- Research highlights the importance of *in utero* and early life environment to an increased susceptibility to developing NCDs later in life.
- Given the burden of NCD in many African countries where healthcare resources are scarce, it is important to identify prenatal risk factors in an attempt to curb this growing epidemic.

Epigenetic effects

- Epigenetic modifications are influenced by several factors including age, sex, the environment/lifestyle and disease state.
- New and ongoing research is uncovering the role of epigenetics in a variety of NCDs.
- Epigenetic data from African countries is limited, leading to an under-representation of the epigenetic diversity in African populations.

The effect of fixed genetic variation on DNA methylation

- When interpreting the role of epigenetic variation in a complex disease, it is important to consider both fixed genetic variation and environmental factors.
- Recent studies have highlighted that a large proportion of DNA methylation variability across individuals and populations is also attributable to underlying genetic variability.
- Most research has focused on the potential for variation at the cellular level, relatively little research has been done to examine how epigenetic variation affects humans at the population level.

Conclusion & future perspective

- There is a need to build capacity in Africa for research that leads to an understanding of the association between genetic, epigenetic and environmental risk factors for NCD on the continent.
- African populations harbor increased genetic diversity and rapidly changing environments (e.g., through urbanization) that provide a unique opportunity to explore genome–environment–epigenome relationships.
- There is a need to identify biomarkers through a systematic comparison of fetal epigenomes using genome-wide methylation arrays concentrating on African populations.

lation-specific epigenetic changes that contribute to a population's epigenetic diversity. For this reason, epigenetic studies are not readily transferrable from one population to the next and it is necessary to identify disease-specific profiles, obtain knowledge regarding the exposure, the modifiable factors and their effects in order to develop population appropriate interventions.

Disclaimer

M Ramsay is a South African Research Chair in Genomics and Bioinformatics of African Populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology and administered by the National Research Foundation of South Africa (NRF). A Hobbs is supported by an NIH Fogarty Training Fellowship. The content of the review is solely the responsibility of the authors and does not necessarily represent the official views of the Fogarty International Centre, the NIH or the NRF.

References

Papers of special note have been highlighted as:

• of interest; •• of considerable interest

- Holmes MD, Dalal S, Volmink J *et al.* Non-communicable diseases in sub-Saharan Africa: the case for cohort studies. *PLoS Med.* 7(5), e1000244 (2010).
- Kim HC, Oh SM. Noncommunicable diseases: current status of major modifiable risk factors in Korea. *J. Prev. Med. Public Health* 46(4), 165–172 (2013).
- Ueda P. *Famine in Nigeria and vitamin D in Sweden: two early life exposures and their relation to cardiovascular risk in adulthood.* PhD, Karolinska Institutet, Stockholm, Sweden (2013).
- Puoane TT L, Sanders D, Parker W. Chronic non-communicable diseases. In: *South African Health Review 2008.* Barron P, Roma-Reardon J (Eds). Health Systems Trust, Durban, South Africa (2008).
- Consortium HA. Research capacity. Enabling the genomic revolution in Africa. *Science* 344(6190), 1346–1348 (2014).
- Uauy R, Kain J, Corvalan C. How can the Developmental Origins of Health and Disease (DOHaD) hypothesis contribute to improving health in developing countries? *Am. J. Clin. Nutr.* 94(6 Suppl.), S1759–S1764 (2011).
- Barker DJ, Winter PD, Osmond C, Margetts B, Simmonds SJ. Weight in infancy and death from ischaemic heart disease. *Lancet* 2(8663), 577–580 (1989).
- Hales CN, Barker DJ. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia* 35(7), 595–601 (1992).
- Bouchard L. Epigenetics and fetal metabolic programming: a call for integrated research on larger cohorts. *Diabetes* 62(4), 1026–1028 (2013).
- Wang G, Walker SO, Hong X, Bartell TR, Wang X. Epigenetics and early life origins of chronic noncommunicable diseases. *J. Adolesc. Health* 52(2 Suppl. 2), S14–S21 (2013).

Open access

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Financial & competing interests disclosure

The authors wish to acknowledge support from the National Health Laboratory Service and the Sydney Brenner Institute of Molecular Bioscience, as well as a training grant from the Fogarty International Centre (1D43TW008330) (NIH) which supported A Hobbs. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

- Lehnen H, Zechner U, Haaf T. Epigenetics of gestational diabetes mellitus and offspring health: the time for action is in early stages of life. *Mol. Hum. Reprod.* 19(7), 415–422 (2013).
- **Describes the epigenetics of gestational diabetes by focusing particularly on the role the environment (maternal nutrition) has on the developing fetus and consequences of this adverse environment for later life.**
- Heijmans BT, Tobi EW, Stein AD *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl Acad. Sci. USA* 105(44), 17046–17049 (2008).
- El Hajj N, Schneider E, Lehnen H, Haaf T. Epigenetics and life-long consequences of an adverse nutritional and diabetic intrauterine environment. *Reproduction* 148(6), R111–R120 (2014).
- **Gives an excellent overview of fetal overnutrition and how an individuals metabolic phenotype in life depends on intrauterine nutrition. The authors also discuss a link to epigenetics.**
- Waterland RA, Kellermayer R, Laritsky E *et al.* Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet.* 6(12), e1001252 (2010).
- Khulan B, Cooper WN, Skinner BM *et al.* Periconceptional maternal micronutrient supplementation is associated with widespread gender related changes in the epigenome: a study of a unique resource in the Gambia. *Hum. Mol. Genet.* 21(9), 2086–2101 (2012).
- Hult M, Tornhammar P, Ueda P *et al.* Hypertension, diabetes and overweight: looming legacies of the Biafran famine. *PLoS ONE* 5(10), e13582 (2010).
- Rinaudo P, Wang E. Fetal programming and metabolic syndrome. *Annu. Rev. Physiol.* 74, 107–130 (2012).
- Koukoura O, Sifakis S, Spandidos DA. DNA methylation in the human placenta and fetal growth (review). *Mol. Med. Rep.* 5(4), 883–889 (2012).

- 19 Gluckman PD, Hanson MA. The developmental origins of the metabolic syndrome. *Trends Endocrinol. Metab.* 15(4), 183–187 (2004).
- 20 Fernandez-Morera JL, Rodriguez-Rodero S, Menendez-Torre E, Fraga MF. The possible role of epigenetics in gestational diabetes: cause, consequence, or both. *Obstet. Gynecol. Int.* 2010, 605163 (2010).
- 21 Daxinger L, Whitelaw E. Transgenerational epigenetic inheritance: more questions than answers. *Genome Res.* 20(12), 1623–1628 (2010).
- 22 Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* 15(2), R31 (2014).
- 23 Barfield RT, Almlil LM, Kilaru V *et al.* Accounting for population stratification in DNA methylation studies. *Genet. Epidemiol.* 38(3), 231–241 (2014).
- **Proposes several approaches to correct for population stratification using principal components (PCs) from different subsets of genome-wide methylation data. Issues such as population stratification will need to be addressed as the scale of DNA methylation association studies approaches that of genome-wide association studies.**
- 24 Relton CL, Groom A, St Pourcain B *et al.* DNA methylation patterns in cord blood DNA and body size in childhood. *PLoS ONE* 7(3), e31821 (2012).
- 25 Bollati V, Baccarelli A. Environmental epigenetics. *Heredity (Edinb.)* 105(1), 105–112 (2010).
- 26 Ruchat SM, Houde AA, Voisin G *et al.* Gestational diabetes mellitus epigenetically affects genes predominantly involved in metabolic diseases. *Epigenetics* 8(9), 935–943 (2013).
- 27 Lowe R, Rakyan VK. Correcting for cell-type composition bias in epigenome-wide association studies. *Genome Med.* 6(3), 23 (2014).
- 28 Bouchard L, Hivert MF, Guay SP, St-Pierre J, Perron P, Brisson D. Placental adiponectin gene DNA methylation levels are associated with mothers' blood glucose concentration. *Diabetes* 61(5), 1272–1280 (2012).
- 29 Bouchard L, Thibault S, Guay SP *et al.* Leptin gene epigenetic adaptation to impaired glucose metabolism during pregnancy. *Diabetes Care* 33(11), 2436–2441 (2010).
- 30 El Hajj N, Pliushch G, Schneider E *et al.* Metabolic programming of MEST DNA methylation by intrauterine exposure to gestational diabetes mellitus. *Diabetes* 62(4), 1320–1328 (2013).
- 31 Plagemann A, Harder T, Brunns M *et al.* Hypothalamic proopiomelanocortin promoter methylation becomes altered by early overfeeding: an epigenetic model of obesity and the metabolic syndrome. *J. Physiol.* 587(Pt 20), 4963–4976 (2009).
- 32 Lillycrop KA, Phillips ES, Jackson AA, Hanson MA, Burdge GC. Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *J. Nutr.* 135(6), 1382–1386 (2005).
- 33 Slater-Jefferies JL, Lillycrop KA, Townsend PA *et al.* Feeding a protein-restricted diet during pregnancy induces altered epigenetic regulation of peroxisomal proliferator-activated receptor-alpha in the heart of the offspring. *J. Dev. Orig. Health Dis.* 2(4), 250–255 (2011).
- 34 Dolinoy DC. The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. *Nutr. Rev.* 66(Suppl. 1), S7–S11 (2008).
- 35 Tobi EW, Lumey LH, Talens RP *et al.* DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum. Mol. Genet.* 18(21), 4046–4053 (2009).
- 36 Dominguez-Salas P, Moore SE, Baker MS, Bergen AW. Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nat. Commun.* 5(3746), 1–7 (2014).
- 37 Barouki R, Gluckman PD, Grandjean P, Hanson M, Heindel JJ. Developmental origins of non-communicable disease: implications for research and public health. *Environ. Health* 11, 42 (2012).
- 38 Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* 465(7299), 721–727 (2010).
- **Describes the role epigenetic modifications, such as DNA methylation and histone modification, play in the development of common complex diseases. They highlight that the plasticity of the epigenome is what allows it to react to cues from the external and internal environments and that such induced epigenetic changes can be solidified and propagated during cell division, resulting in permanent maintenance of the acquired phenotype.**
- 39 Paul DS, Beck S. Advances in epigenome-wide association studies for common diseases. *Trends Mol. Med.* 20(10), 541–543 (2014).
- 40 Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* 12(8), 529–541 (2011).
- 41 Heijmans BT, Mill J. Commentary: the seven plagues of epigenetic epidemiology. *Int. J. Epidemiol.* 41(1), 74–78 (2012).
- **Highlights the downfalls and limitations of epigenetic studies that researchers need to take into consideration when planning such a study, independent on population group.**
- 42 Verma M. Epigenome-wide association studies (EWAS) in cancer. *Curr. Genomics* 13(4), 308–313 (2012).
- 43 Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J. Epigenome-wide association studies without the need for cell-type composition. *Nat. Methods* 11(3), 309–311 (2014).
- 44 Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* 15(5), 335–346 (2014).
- 45 Ramsay M. Africa: continent of genome contrasts with implications for biomedical research and health. *FEBS Lett.* 586(18), 2813–2819 (2012).
- 46 Jallo N, Lyon DE, Kinser PA, Kelly DL, Menzies V, Jackson-Cook C. Recruiting for epigenetic research: facilitating the informed consent process. *Nurs. Res. Pract.* 935740 (2013).

- 47 Ramsay M, De Vries J, Soodyall H, Norris SA, Sankoh O. Ethical issues in genomic research on the African continent: experiences and challenges to ethics review committees. *Hum. Genomics* 8(1), 15 (2014).
- 48 Teh AL, Pan H, Chen L *et al.* The effect of genotype and *in utero* environment on interindividual variation in neonate DNA methylomes. *Genome Res.* 24(7), 1064–1074 (2014).
- 49 McRae AF, Powell JE, Henders AK *et al.* Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol.* 15(5), R73 (2014).
- 50 Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol.* 15(2), R37 (2014).
- 51 Smith AK, Kilaru V, Kocak M *et al.* Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage and tissue type. *BMC Genomics* 15, 145 (2014).
- 52 Bell JT, Pai AA, Pickrell JK *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* 12(1), R10 (2011).
- 53 Gibbs JR, Van Der Brug MP, Hernandez DG *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* 6(5), e1000952 (2010).
- 54 Liu Y, Li X, Aryee MJ *et al.* GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease. *Am. Soc. Hum. Genet.* 94, 485–495 (2014).
- 55 Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.* 9, 403–433 (2008).
- 56 Tishkoff SA, Reed FA, Friedlaender FR *et al.* The genetic structure and history of Africans and African Americans. *Science* 324(5930), 1035–1044 (2009).
- 57 Ambrosone CB, Young AC, Sucheston LE *et al.* Genome-wide methylation patterns provide insight into differences in breast tumor biology between American women of African and European ancestry. *Oncotarget* 5(1), 237–248 (2013).
- 58 Ashktorab H, Daremipouran M, Goel A *et al.* DNA methylome profiling identifies novel methylated genes in African American patients with colorectal neoplasia. *Epigenetics* 9(4), 503–512 (2014).
- 59 Dogan MV, Shields B, Cutrona C *et al.* The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. *BMC Genomics* 15, 151 (2014).
- 60 Wang S, Dorsey T, Terunuma A, Kittles R, Ambs S, Kwabi-Addo B. Relationship between tumor DNA methylation status and patient characteristics in African-American and European-American women with breast cancer *PLoS ONE* 7(5), e37928 (2012).
- 61 Heyn H, Moran S, Hernando-Herraez I *et al.* DNA methylation contributes to natural human variation. *Genome Res.* 23(9), 1363–1372 (2013).
- 62 Fraser HB, Lam LL, Neumann SM, Kobor MS. Population-specificity of human DNA methylation. *Genome Biol.* 13(2), R8 (2012).
- **Reports on ethnic differences in human DNA methylation. They show that at a number of CpG sites the DNA methylation is highly divergent between populations, and that this divergence may be due in large part to a combination of differences in allele frequencies and complex epistasis or gene–environment interactions.**
- 63 Kwabi-Addo BW, S and Devaney, J. Epigenome-wide profiling identified significant differences in DNA methylation between African-American and European-American men with prostate cancer. *Cancer Res.* (74), 388 (2014).
- 64 Nieminen TT, Shoman S, Eissa S, Peltomaki P, Abdel-Rahman WM. Distinct genetic and epigenetic signatures of colorectal cancers according to ethnic origin. *Cancer Epidemiol. Biomarkers Prev.* 21(1), 202–211 (2012).
- 65 Adkins RM, Krushkal J, Tylovsky FA *et al.* Racial differences in gene specific DNA methylation Levels are present at birth. *Birth Defects Res. A. Clin. Mol. Teratol.* 91(8), 728–736 (2011).
- 66 Ataguba JE, Akazili J, McIntyre D. Socioeconomic-related health inequality in South Africa: evidence from General Household Surveys. *Int. J. Equity Health* 10(1), 48 (2011).
- 67 Borghol N, Suderman M, Mcardle W *et al.* Associations with early-life socio-economic position in adult DNA methylation. *Int. J. Epidemiol.* 41(1), 62–74 (2012).
- 68 Mcguinness D, Mcglynn LM, Johnson PC *et al.* Socio-economic status is associated with epigenetic differences in the pSoBid cohort. *Int. J. Epidemiol.* 41(1), 151–160 (2012).
- 69 Bierne H, Hamon M, Cossart P. Epigenetics and bacterial infections. *Cold Spring Harb. Perspect. Med.* 2(12), a010272 (2012).
- 70 De Vries J, Pepper M. Genomic sovereignty and the African promise: mining the African genome for the benefit of Africa. *J. Med. Ethics* 38(8), 474–478 (2012).
- 71 H3Africa – Human hereditary and health in Africa. www.h3africa.org