○ Discover

Review

An integrative framework and recommendations for the study of DNA methylation in the context of race and ethnicity

Meingold Hiu-ming Chan[1,2,3] ⬤ · Sarah M. Merrill[1,2,3] · Chaini Konwar[1,2,3] · Michael S. Kobor[1,2,3,4]

Abstract

Human social epigenomics research is critical to elucidate the intersection of social and genetic influences underlying racial and ethnic differences in health and development. However, this field faces major challenges in both methodology and interpretation with regard to disentangling confounded social and biological aspects of race and ethnicity. To address these challenges, we discuss how these constructs have been approached in the past and how to move forward in studying DNA methylation (DNAm), one of the best-characterized epigenetic marks in humans, in a responsible and appropriately nuanced manner. We highlight self-reported racial and ethnic identity as the primary measure in this field, and discuss its implications in DNAm research. Racial and ethnic identity reflects the biological embedding of an individual's sociocultural experience and environmental exposures in combination with the underlying genetic architecture of the human population (i.e., genetic ancestry). Our integrative framework demonstrates how to examine DNAm in the context of race and ethnicity, while considering both intrinsic factors—including genetic ancestry—and extrinsic factors—including structural and sociocultural environment and developmental niches—when focusing on early-life experience. We reviewed DNAm research in relation to health disparities given its relevance to race and ethnicity as social constructs. Here, we provide recommendations for the study of DNAm addressing racial and ethnic differences, such as explicitly acknowledging the self-reported nature of racial and ethnic identity, empirically examining the effects of genetic variants and accounting for genetic ancestry, and investigating race-related and culturally regulated environmental exposures and experiences.

1 Introduction

The burgeoning field of human social epigenomics is a critical avenue for studying how the external social environment can get "under the skin" and influence key biological systems, a process known as biological embedding, through gene regulation with implications for human health and development [1–4]. For example, research efforts have been devoted to unraveling the molecular mechanisms underlying the extensively documented health disparities across race and ethnicity, especially in the USA. In particular, there is increasing interest in DNA methylation (DNAm), a commonly interrogated epigenetic mark in human population studies, to explain racial and ethnic health disparities [5, 6]. However, by

---

✉ Sarah M. Merrill, Smm474@cornell.edu; ✉ Michael S. Kobor, michael.kobor@ubc.ca | [1]Department of Medical Genetics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada. [2]British Columbia Children's Hospital Research Institute, University of British Columbia, Vancouver, BC, Canada. [3]Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada. [4]Edwin S. H. Leong Healthy Aging Program, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada.

Check for updates

🍷 Springer

examining the intersection of genetic and environmental influences, this field has encountered an issue vital for both methodology and interpretation: how to disentangle the confounded social and biological aspects of constructs, such as race and ethnicity.

The field of epigenetics, like many other areas of biomedical research, has historically used race and ethnicity to categorize study participants [7]. However, both concepts are increasingly recognized not as biologically meaningful categories [8, 9] but as complex sociocultural constructs that may reflect the biological embedding or embodiment of an individual's social experience and environmental exposures in combination with their underlying genetic architecture. Given the history and inherent dangers of the misuse of biomedical science, including epigenomics, to reinforce racial stereotypes, robust investigations and careful interpretation of results related to race and ethnicity are required to avoid unintended consequences.

Building on progress in this field to date, this review discusses how race and ethnicity have been approached in the past and how to move forward taking account of these constructs in epigenetics research in a responsible and appropriately nuanced manner. We propose an integrative framework for studying DNAm in the context of race and ethnicity. This framework highlights and integrates specific aspects of both biological and environmental factors relevant to race and ethnicity in DNAm research, which can also be applied to other omics research. Given the interest in using DNAm to understand health disparities across race and ethnicity, we review the current state of research in this field based on this framework and discuss challenges and limitations in the field. We provide recommendations for conducting responsible and methodologically robust research in both the short and long term. Although this review is centered on DNAm, the overall framework can also be expanded and applied to other omics research, including gene expression and proteomics studies [3].

2 Race and ethnicity measures in biomedical research

Race and ethnicity are both social constructs which, while commonly considered synonymous, represent distinct concepts. Specifically, race is commonly defined by physical traits, including skin color and hair texture, while ethnicity is a more complex and multidimensional construct reflecting cultural and historical background, including shared language, norms, traditions, values, diet, and geographical origin [10, 11]. Both have long been used in biomedical research as proxies for genetic population differences, especially as genotype data are not always available, which could contribute to the propagation of harmful racial stereotypes [12]. However, as clearly illustrated in the literature, different racial and ethnic groups are not genetically discrete [11], with more genetic variation within than across groups (93%–95% vs. 3%–5%, respectively) [13]. Further, these groupings based on race and ethnicity may be neither clinically meaningful nor appropriate. For example, in a study examining drug responses of individuals from eight different ethnic groups, four clusters inferred from genetic markers did not necessarily correspond to the groupings based on ethnicity [14].

Importantly, both race and ethnicity are commonly assessed by self-identification. At times, these constructs are inferred by a third party, such as health care providers, without soliciting self-reporting [15], or no details are provided regarding how they were measured [16]. However, as biomedical studies have often assessed race and ethnicity based on self-reporting [10, 17], we use the term "racial and ethnic *identity*" throughout this review where appropriate. Very few studies relied on other measures reflecting social standing (e.g., other-perceived racial or ethnic groups) or cultural orientations. Often, these measures are assessed using forced multiple choice questions, which can also impact the accuracy of reporting in medical research settings depending on the choices available and whether multiple identities can be selected [18]. Historically, participants who do not clearly fit into available categories may choose "other" as an option, and are subsequently grouped together with many unrelated identities or may be excluded from the analysis altogether [19]. In 2000, the U.S. Census updated their race and ethnicity measures to allow selection of multiple options, and thus documented that 2.4% of the U.S. population self-identified as multiracial. The 2020 U.S. Census revealed a 276% increase in the number of individuals identified as multiracial from 9 to 33 million since 2010, illustrating the increasing racial diversity of the U.S. population [20]. This further reinforced the importance of clearly stating exactly what was measured, i.e., racial and ethnic identity, along with the way in which it was measured to allow appropriate interpretation.

Racial and ethnic identity is a powerful determinant of social experience and environmental exposures. In countries where such identity is associated with social standing and underscores health outcomes, these constructs are likely to capture experiences of social inequalities and structural disparities rather than reflecting true biological differences [16]. Measures of race and ethnicity in health research are more common in North America, and less so in Europe and Asia [11, 16]. Given the lack of data, therefore, health disparities across race and ethnicity may remain undocumented in some countries.

3 DNA methylation to explain racial and ethnic differences

Epigenetics, which sits at the interface of genes and the environment, offers a promising avenue to scrutinize how social and genetic factors together contribute to differential epigenetic signals across race and ethnicity. Specifically, DNAm has been increasingly appreciated and studied as a molecular mark pertinent to understanding racial and ethnic differences [5]. Typically, DNAm refers to the attachment of a methyl group to cytosine residues in the genome, especially at cytosine-phosphate-guanine (CpG) dinucleotides [21]. Similar to many other epigenetic modifications, DNAm has the potential to alter gene expression without changing the underlying DNA sequence. There is now strong evidence that the environment can contribute to differential DNAm, especially from studies in monozygotic twins with different environmental exposures exhibiting differences in epigenetic signatures, including DNAm and histone modifications [22]. While the relative impacts of genetics and environment are still under investigation, it is well established that DNAm is highly susceptible to the influences of both factors [23]. This malleability along with its mitotic heritability and stability over cellular generations make DNAm an ideal candidate for studying biological embedding, especially in the context of constructs such as race and ethnicity, which are simultaneously influenced by differences in genetics and sociostructural environments [21, 24].

Variations in DNAm have been explored to explain phenotypic differences across racial and ethnic groups, especially in the context of health disparities [6]. Substantial differences in a wide range of health risks and outcomes across race and ethnicity have long been recognized in North America, including disparities in rates of cancer [25, 26], chronic illnesses such as obesity [27], cardiovascular disease [28, 29], birth outcomes [30], and neurocognitive [31] and mental health conditions [32], and more recently infection and mortality associated with COVID-19 [33–35]. In addition to identifying health conditions with such racial and ethnic disparities, a growing body of research has attempted to elucidate the underlying factors and biological mechanisms [8]. Although much early research in the biomedical field assumed that genetic variants could explain health disparities across racial and ethnic identities, the recent discovery of a downsized human genome and the recurring issue of missing heritability suggested the importance of accounting for environmental effects [34, 36]. The global COVID-19 pandemic has further illustrated the need to understand health disparities through the lens of social determinants of health and to consider how societal contexts, such as living conditions, population density, social interactions, and access to/interactions with health care, can interact with processes within the body (e.g., immune response), which may exacerbate the risks of infection and mortality [34]. The need to expand research on social determinants of health and move beyond treating race and ethnicity solely as demographic characteristics has been strongly emphasized in social epidemiology and also discussed in the fields of social epigenomics [37] and other biomedical areas [38–40]. Given that health disparities are defined as "systematic, avoidable and unfair differences in health outcomes that can be observed between populations, between social groups within the same population or as a gradient across a population ranked by social position" (p. 28) [41], it is especially critical to examine how known sociostructural pressures may explain health disparities and influence DNAm in conjunction with other biological factors. In addition, the Developmental Origins of Health and Disease (DOHaD) hypothesis further posited that health disparities may originate from prenatal and early-life exposures, at times with racial underpinnings [42], resulting in predisposition of an individual to increased risk of developing diseases and poorer health outcomes over their life course [43]. A recent review identified a total of 49 epigenetics-related studies addressing racial and ethnic health disparities, with the majority focusing on DNAm, and documented increasing interest in this research area over the past several years [6]. Despite the burgeoning interest in this area, there has been limited discussion regarding how to examine DNAm in the context of racial and ethnic identity and related social constructs in a robust and nuanced manner while also considering the genetic backbone upon which DNAm is built [11].

4 The current review

Building on prior efforts [6, 11, 44], this review continues the ongoing discussion of how to study DNAm in relation to race and ethnicity in a methodologically robust and culturally sensitive manner. We first discuss how both genetic ancestry and structural/sociocultural factors are confounded with racial and ethnic identity in a general context. Second, we introduce an integrative framework with a DOHaD and sociocultural perspective to guide future DNAm research in the context of race and ethnicity considering both genetic and environmental factors along with other biological influences relevant to DNAm. Based on this framework, we review the current state of DNAm studies addressing racial and ethnic

health disparities, as this is a highly active and relevant field concerning racial and ethnic differences in DNAm, to identify existing limitations and present recommendations for future research in this field.

5 Interrelations of racial and ethnic identity, genetic ancestry, and sociocultural experience/ environment

In the context of biomedical research, it is crucial to acknowledge that race and ethnicity are typically self-identified measures, which are inherently confounded with two interrelated components: genetic ancestry and social experience/ environmental exposure (Fig. 1).

An individual's racial and ethnic identity is an internal construct determined by the complex interplay of extrinsic and intrinsic factors. Extrinsic factors, such as structural and sociocultural environmental factors, may have bidirectional relations with racial and ethnic identity—the environment in which one is raised can contribute to the development of identity, which can also be a powerful determinant of one's sociocultural environment and experience. As discussed previously, race and ethnicity as sociopolitical constructs can vary with time and their interpersonal and political contexts [45–48], such as immigration and political narratives around certain racial or ethnic groups. At the same time, an individual's own identification with particular racial and ethnic groups can also change over time, sometimes in reaction to changing political narratives and/or along with the process of identity development and feelings of group belonging. This is especially true for multiracial individuals, who are increasingly prevalent in the population. Even the recent increase in availability of genetic testing from commercial entities, such as *23 and Me* and *Ancestry.com*, has also introduced knowledge of one's genetic ancestry as a factor that may influence racial and ethnic identity, including the acceptance of these results [49]. In turn, the process of racial and ethnic identity development can influence an individual's perception of their contexts and subsequent potential positive and negative health impacts [45, 50, 51]. Therefore, given their sociocultural and structural contexts as well as their effects on environmental exposures, it is necessary to take identity development into consideration when interpreting any racial or ethnic differences, ideally through empirical investigation. In this review, we specifically focus on structural factors, such as racism and socioeconomic disadvantage, and culturally regulated sociocultural factors, such as cultural norms and customs.

Racial and ethnic identity is often used as a proxy of genetic relatedness, and specifically genetic ancestry, a biological construct capturing the differences in genetic architecture described by the ancestral population structure. Genetic ancestry has some, albeit incomplete, concordance with racial and ethnic identity; individuals with the same racial or ethnic identity can have different genetic ancestries, and conversely those with the same genetic ancestry can identify as belonging to different racial or ethnic groups [15]. Indeed, studies have shown that racial and ethnic identity does not always accurately reflect genetic ancestry [10, 14].

While the importance of genetic ancestry in explaining DNAm variation is now increasingly appreciated among researchers, a common understanding of this construct is derived from the continent of origin concept. Continental-level

Fig. 1 Interrelations between genetic ancestry, self-identified race and ethnicity, and structural/sociocultural environment. The first three constructs (blue box), which interact with the structural and sociocultural environment, have some albeit incomplete concordance and are at times used as proxies for each other. Figure modified from Lu et al. [16]
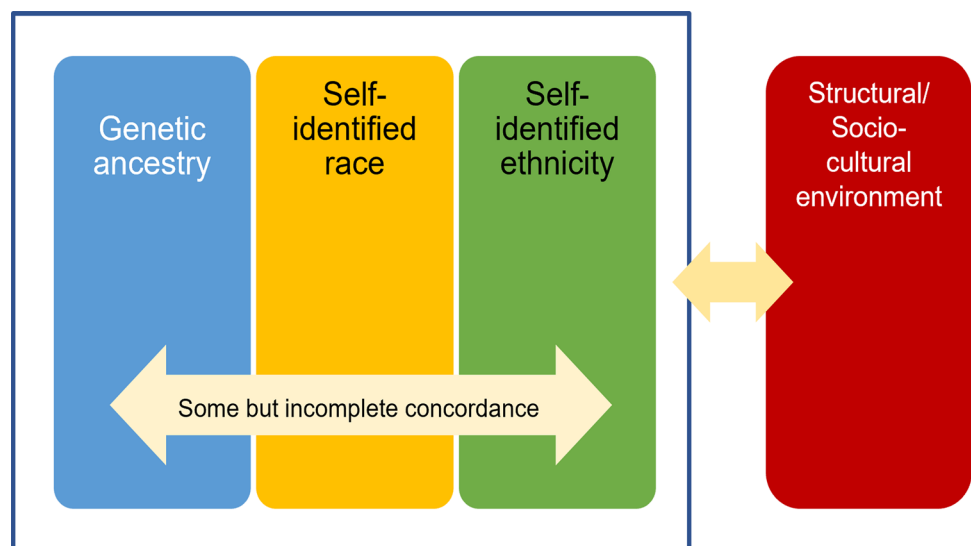
ancestry groupings have been shown to capture sizable population differences in human genetic variation corresponding to geographical regions or superpopulations largely representing the major continents of the world [52, 53]. However, these continental-level ancestry categories are often confounded with racial or ethnic identities, which therefore challenges their applicability. Furthermore, such clustering overlooks genetic diversity within a continent, which is inaccurate given that the highest degree of genetic variation is observed within Africa [54, 55].

In addition, continental-level ancestry groupings neglect global and continuous patterns of migration along with the resulting extent of admixture across superpopulations and continents, which is misleading. For example, modern Latin Americans represent a classic admixed group comprised of African, Native American, and European ancestral parent genomes [56]. In admixed individuals, the fraction of genomic ancestry unique to the ancestral source populations can be estimated using global genetic ancestry methods [57, 58]. Alternatively, estimates of genetic ancestry for each of the chromosomes can be inferred by locus-specific local genetic ancestry measures. Therefore, genetic ancestry has emerged as an appropriate measure to capture differences in human genetic variation, which is best described along a continuum rather than in discrete categories, and thus explicitly captures the impacts of human histories that have led to different forms of genetic diversity around the globe.

6 An integrative framework to study DNAm in the context of race and ethnicity

To facilitate nuanced, precise, and responsible social epigenomics studies, we propose an integrative framework to guide the investigation of DNAm in the context of race and ethnicity (Fig. 2). This framework outlines how genetic ancestry and structural/sociocultural environment are both linked to self-identified racial and ethnic identity, together contributing to DNAm. Echoing the agenda of conceptualizing human biology as social biology [34], we propose a framework integrating both intrinsic and extrinsic factors potentially underlying the relations of DNAm with race and ethnicity. These factors are interdependent and interact to connect elements within and outside the body [34]. To specify our conceptualization of the extrinsic environment surrounding individuals in early childhood and over the life course, we adopted the developmental niche model, a well-established cultural research framework for child development [59, 60].

6.1 Developmental Niche

The developmental niche model is an influential framework in the field of developmental science and cultural anthropology [59], and has been instrumental in guiding developmental research to study the cultural regulation of children's immediate environment responsible for shaping their long-term development. Another significant contribution of this model is its emphasis on the "immediacy of cultural forces in the environment of an individual" (p. 284) [61], which helps to conceptualize and operationalize culture as a factor with direct impacts on the construction of children's everyday experience [59]. Adopting the developmental niche model allows us to examine extrinsic early-life factors that may explain some of the epigenetic signals found in relation to race and ethnicity, such as signals found in certain racial groups but not others, in a systematic manner. This model also incorporates psychobiological concepts, such as child temperament and stress regulation, and highlights the interaction between children's intrinsic biological factors and extrinsic cultural factors, which can be meaningfully adopted in the study of DNAm.

6.2 Intrinsic factors

The operational definition of intrinsic factors for our framework includes those aspects arising from the self, both biological and psychological, for which there is evidence of associations with DNAm. Genetic ancestry and other biological factors, such as single nucleotide polymorphisms (SNPs) defined as genetic variations at the individual base pair level, tissue type, cell type heterogeneity, sex, and age, are all known to be drivers of variable and differential DNAm [24, 62]. As discussed above, self-identified race and ethnicity, referring to the internal psychological representation of two distinct but related group memberships, share some, albeit incomplete, concordance with genetic ancestry [14]. Other psychological traits, such as temperament (i.e., predisposed patterns of emotional and behavioral responses), are also associated with health outcomes and DNAm.

Many of these intrinsic factors simultaneously contribute to DNAm and shape the extrinsic environment. For example, psychological traits, such as temperament, are determined by genetics and can shape an individual's
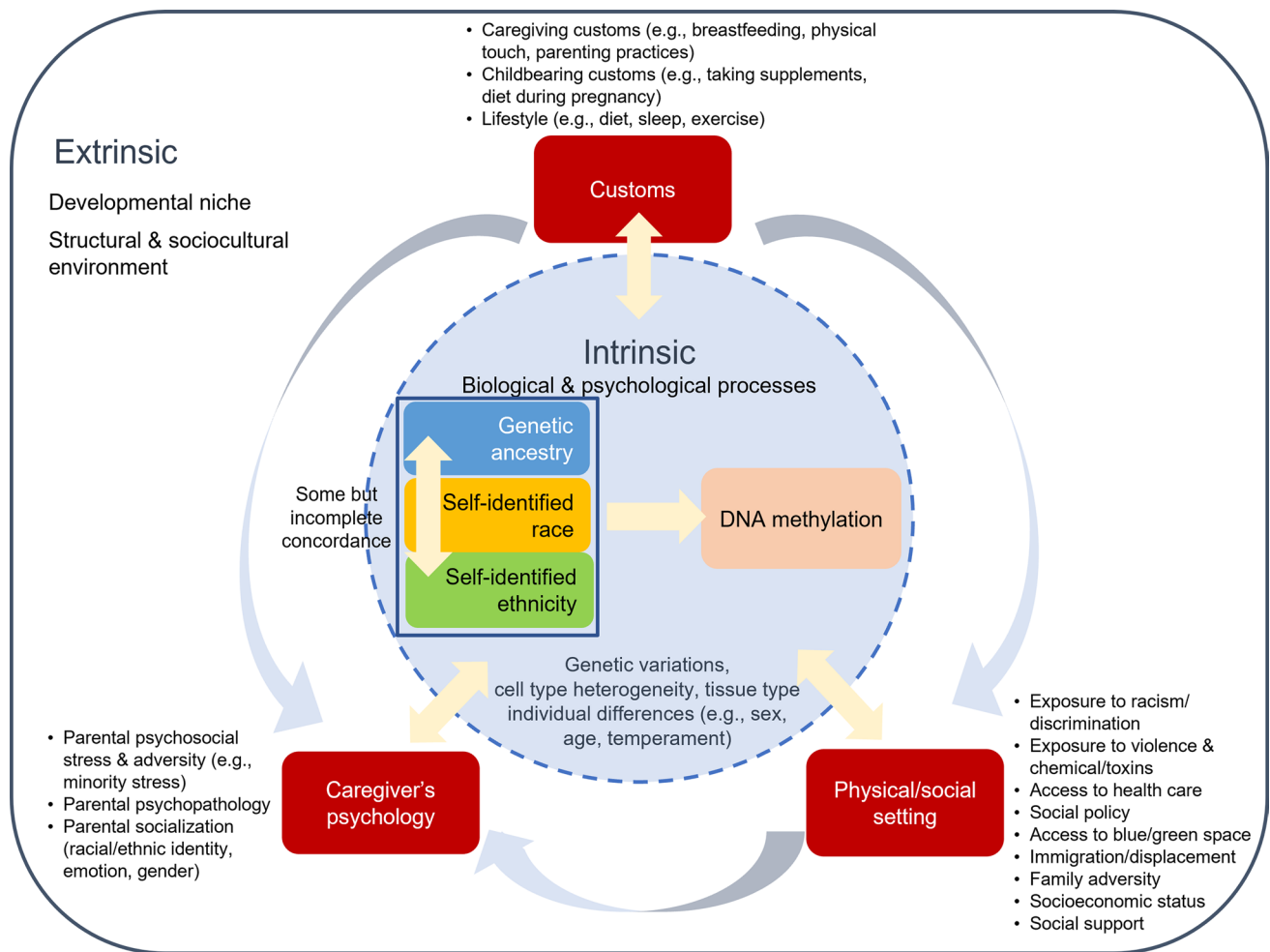
Fig. 2 Integrative framework illustrating how intrinsic and extrinsic environmental factors tied to self-identified race and ethnicity contribute to DNA methylation (DNAm), which has the potential to alter gene expression and influence downstream biological functions associated with health and diseases. The circle represents an individual, with all the components inside the circle as intrinsic factors. The three main constructs—genetic ancestry, self-identified race, and self-identified ethnicity—are highlighted by the colored boxes, with other intrinsic factors that may influence DNAm encompassed within the circle. The extrinsic part is made up of both developmental niche and other structural and sociocultural environmental factors shown in red boxes. The developmental niche surrounding the individual consists of three subsystems, i.e., physical/social settings, customs, and caregiver's psychology, which bring culture into their daily experience. Specific factors under each of the three subsystems of the developmental niche associated with DNAm and self-identified race and ethnicity are listed. The arrows between the three subsystems as well as those between the extrinsic and intrinsic factors represent their constant interactions and their contributions to the relations of race, ethnicity, and genetic ancestry with DNAm

susceptibility to their environment. Children with a fearful temperament are more likely to develop psychopathology when exposed to a stressful environment than their less fearful counterparts [63]. In addition, genetics, specifically the sex chromosomes, is associated with an individual's sex, which also informs their gender identity and determines their social experience based on cultural expectations of gender roles. Similarly, social experiences change over the life course associated with cellular aging (which can be captured by epigenetic aging), changes in cell type composition, and health status. Therefore, it is necessary to recognize the complex relations of intrinsic and extrinsic factors with DNAm.

6.3 Extrinsic factors

Our framework specifies ways in which to study the relations of race and ethnicity with DNAm in the context of extrinsic environments that are interrelated with the intrinsic elements, especially self-identified race and ethnicity. Extrinsic factors also have dynamic relations with other intrinsic factors, such as sex and age. Hence, the environment and experience are expected to change over the life course, and to differ across developmental stages and in relation to an individual's sex. We highlight early-life environmental factors that may be culturally regulated, which aligns with the DOHaD perspective. Guided by the developmental niche model, we focused on extrinsic factors under three subsystems that operate in concert and interact with each other to form the cultural context of child development: physical and social settings in which the individual lives; culturally regulated customs, such as caregiving and childbearing practices; and caretaker's psychology, such as feelings, goals, and beliefs [60]. Integrating these three subsystems in our framework, we identified environmental factors for which there is already some, albeit primarily correlational, evidence of their associations with DNAm that can be extended to the context of race and ethnicity, and underscored structural factors relevant to each system.

7 Current state of DNAm research in the context of race, ethnicity, and health

Here, we delineate how the intrinsic and extrinsic factors in the framework contribute to our understanding of DNAm and how they have been studied in the context of race and ethnicity. To illustrate each aspect of the integrative framework, we present an overview of the current state of DNAm research in race, ethnicity, and health using 49 studies that were carefully screened and selected to meet criteria of addressing racial and ethnic differences in human health and DNAm in a recent review paper [6]. In their original analyses of this set of studies, the authors took a critical view and evaluated whether these studies "implicitly reflect or reiterate some aspects of the environmentally driven template" (p. 10) [6]. In light of their longer historiographic premise and historical review, they identified a list of current issues, including the focus on negative exposure and pathology, limited research on reversibility of early-life effects, and racial typologies without considering heterogeneity within populations [6]. This previous paper provided a cautionary review of how epigenetic research could be misused to propagate stigmas and presented recommendations for a balanced approach to treat race in postgenomic-era research, including moving beyond using White as the norm and embracing plasticity, variation within populations, and reversibility in epigenetics research [6]. While the current review shared a similar goal of contributing to responsible practices of studying epigenetics in the context of race and ethnicity and examined the same set of papers as in the previous review, it made a distinct contribution by extending the discussion to evaluation of *methodologies and analyses* used in these studies and provided actionable recommendations for future research based on our evaluation.

Specifically, our review was guided by four questions: (1) Did the authors acknowledge how race and ethnicity were measured and describe what measures were used? (2) Did they measure genetic ancestry, and how was it considered in their analyses? (3) Did the authors empirically examine the effects of genetic variants on DNAm? and (4) Did the authors empirically examine any environmental effects on DNAm, and if so, what measures of environment/experience were considered in the analysis? Full details of each study, including information related to the four questions outlined above, the country of each corresponding author's research institute, the journal in which the study was published, technology used to measure DNAm and genotype, and age of the participants, can be found in Additional file 1: Table S1.

7.1 Intrinsic factors

7.1.1 Racial and ethnic identity

Less than half of the studies reviewed (22 of 49; 44.9%) explicitly discussed how race and ethnicity were measured (Fig. 3). As expected, the majority of those that explicitly reported these measures used self-reported identification (or parent/guardian-reported identification for participants younger than 18 years) [64, 65]. In rare cases, biological signatures were used to categorize participants. For example, one study categorized participants as indigenous Huichol based on the criteria established by the National Institute of Indigenous People of Mexico, as this group has a biological signature of

a specific blood type (type O and Rh+) [66]. Most of the studies did not adopt a critical view of race and ethnicity as poor proxies of human physiology [6, 67]. In addition, they did not acknowledge the malleability of racial and ethnic identity over time or based on context and experiences, unlike the genetic architecture [68, 69]. For example, none of the studies addressed or measured the processes of racial or ethnic identity development or sense of belonging [45] as potential influences on experience or biological embedding as measured by DNAm.

7.1.2 Genetic ancestry and genetics

The majority of the studies overlooked the effects of genetics on DNAm (Fig. 3). Arguably, however, differences in the underlying genetic backbone onto which methyl groups can be deposited are among the largest contributors to DNAm levels in human differential DNAm studies [24]. There are several potential mechanisms by which this can occur. For example, as DNAm occurs at cytosines in the context of CpG dinucleotides, a polymorphic site with a SNP in which the guanine is replaced by another nucleotide, such as thymine, would not be methylated. In addition, the presence of certain sequence variants before or after CpGs can make methylation at a given site more or less likely. Therefore, the majority of these SNPs strongly associated with CpG methylation levels, known as methylation quantitative trait loci (mQTLs), are located physically close (usually within 50 kb) to the CpGs they influence in the genome [70]. These SNPs close to CpG sites, referred to as *cis*-mQTLs, are thought to disrupt regular protein binding, leading to passive or active changes in DNAm either through steric hindrance with mechanisms of DNAm deposition or directly through interactions with the machinery involved in these mechanisms [71]. These mQTLs were estimated to account for approximately 20%–80% of DNAm variance [72–75], and a significant proportion of commonly characterized DNAm sites on array platforms are likely under genetic influence. For example, a recent report showed that approximately 37.9% of CpGs tested had an associated mQTL in at least one tissue, although many of these were tissue-specific [76]. Notably, genetic influences on DNAm variation were detected among different continental-level ancestries due to differences in the underlying DNA sequences across groups; however, many mQTLs have also been shown to exert their influence on DNAm within populations [70, 77].

Only 38.8% (19 of 49) of the reviewed studies empirically considered the effects of genetic variants on DNAm (Fig. 3). The majority of these studies (16 of 19, 84.2%) identified potential mQTLs in either a post hoc manner following discovery of CpGs associated with the variable of interest in an epigenome-wide association study (EWAS) [78, 79] or by a priori approaches to examine the effects of genetic variants on DNAm [80, 81]. Studies in which the samples lacked corresponding genotyping data leveraged existing mQTL databases (e.g., BIOS QTL browser, NCBI dbSNP database, http://www.mqtldb.org/) [82] or previous meta-analyses of empirical studies [27] to identify the influence of potential mQTLs on CpGs of interest. Alternatively, potential mQTLs were identified by examining the distribution of DNAm level

Fig. 3 Summary of the 49 studies included in the previous review based on whether they acknowledged how race and ethnicity were measured, whether they measured genetic ancestry and how was it considered in their analyses, whether they empirically examined the effects of genetic variants on DNAm, and whether they empirically examined the effects of environment/experience on DNAm. PCs, principal components
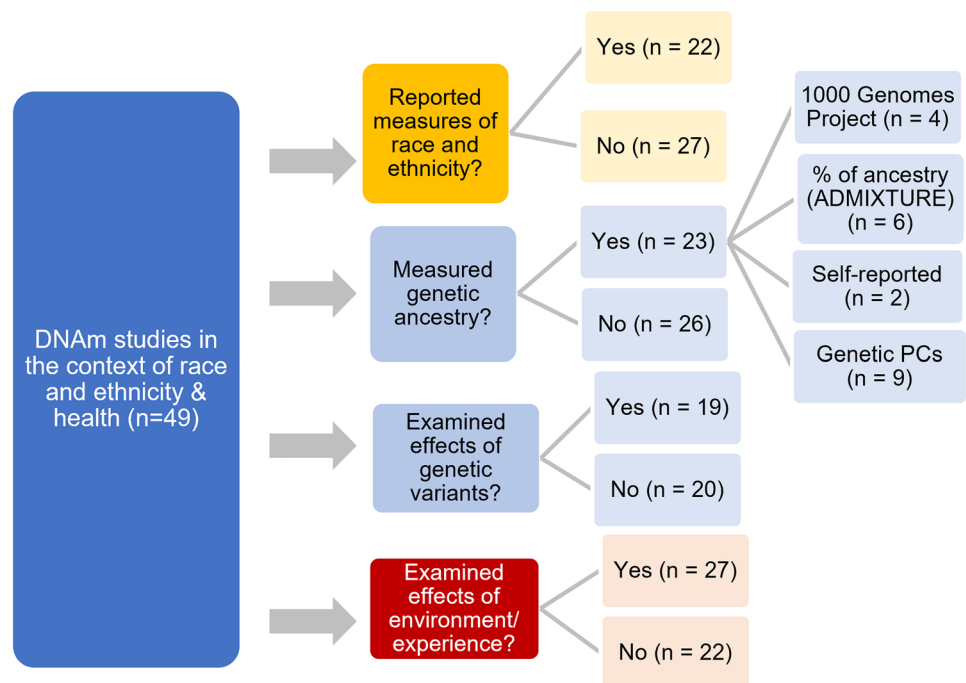
at specific CpGs, as a bimodal or trimodal distribution usually signifies the effects of genetic variants [83]. In other studies, sensitivity analyses were conducted after removing CpGs potentially associated with mQTLs to ensure that the findings were not driven by genetic influences [84, 85] or the effects of SNP mutations [86].

Taking a more holistic view with a data reduction approach rather than focusing on specific SNPs or CpGs, 46.9% (23 of 49) of the studies measured genetic ancestry. The majority of these studies (15 of 23, 65.2%) described genetic ancestry as a continuous variable, with representation of genetic heterogeneity in the sample populations using genetic principal components (PCs) or multidimensional scaling (MDS) coordinates, both commonly employed genomic data dimension reduction techniques. To account for the influence of genetic ancestry on DNAm variation, some studies included the genetic PCs or MDS coordinates in their statistical models [80, 87]. A few studies also acknowledged the admixed nature of the sampled populations and used methods to estimate local and global ancestry with ADMIXTURE [57] and PLINK [79, 85]. On the other hand, some studies used self-reported ancestry based on the continent of origin [88] or self-reported racial identity as a proxy for genetic ancestry [67], which as discussed in previous sections may be inappropriate and could make subsequent meta-analyses across studies difficult.

7.2 Extrinsic factors: Environment and experience

Individuals with different racial and ethnic identities may be exposed to unique developmental niches and sociostructural environments, which can contribute to differential DNAm both within and across race and ethnicity. Although the majority of the studies included in the review discussed environmental associations, most merely speculated on these as potential explanations without empirical investigation, with only 55% (27 of 49) empirically examining environmental exposures or experiences (Fig. 3). Only six of these studies treated an environmental exposure/experience as the main variable of interest [65, 89–93], whereas most included environmental factors as covariates or as post hoc explanations for racial and ethnic differences (Fig. 4a). Furthermore, only two studies explicitly examined whether environmental exposures are indeed different across racial and ethnic groups [67, 94]. Some studies verified whether these identity-associated CpGs were discovered in previous analyses of environmental effects when direct measures of environment were not available for their own samples [95]. We characterized the physical, psychosocial, and cultural aspects of environmental measures relevant to DNAm within the context of race and ethnicity (Fig. 4a), and illustrated how they can become biologically embedded through changes in DNAm and thus have long-lasting impacts on health [96].

7.2.1 Physical environment

There is a large body of evidence from correlational data for the association of differential DNAm with physical settings or built environments, including exposure to pollution, chemicals, or toxins, and ecological habitats (e.g., food deserts, availability of green space, neighborhood safety), which can differ across race and ethnicity [97–99]. These physical and built environments (e.g., neighborhoods and households) fall under the subsystem of physical settings, comprising the immediate environment with which children engage in their daily activities. In particular, the physical environments of some minoritized racial and ethnic groups are subject to structural inequity due to social policies, such as redlining, including residing in impoverished neighborhoods with higher likelihoods of exposure to tobacco smoke, allergens, chemicals, and toxins, especially via pollution and occupational hazards, and socioeconomic disadvantages leading to limited resources and poorer nutrition [100]. This is particularly salient in the USA and Canada where racial minorities are exposed to a range of disadvantages resulting from long histories of discrimination and mistreatment [30, 101, 102]. These exposures contribute to health disparities, such as the higher prevalence rates of asthma and allergies among racially minoritized groups [103].

7.2.2 Psychosocial environment

In addition to disadvantages within the physical environment, racially minoritized individuals also experience more psychosocial stressors, such as racism, discrimination, and acculturation stress, which shape social settings and experience in their developmental niche [104, 105]. The weathering hypothesis posits that health disparities among racially minoritized groups are consequences of these cumulative daily stressors, which cause "wear and tear" on health throughout the life course [100, 106]. Indeed, perceived racial discrimination and social strain often experienced by African Americans are associated with differential DNAm in genes implicated in inflammation [107] and accelerated epigenetic aging [102, 108], even in cases of outwardly displayed resilience [109]. Hence, beyond self-reported psychological states or behaviors, the underlying impact of stress on the health of racialized and ethnically marginalized groups can be reflected by
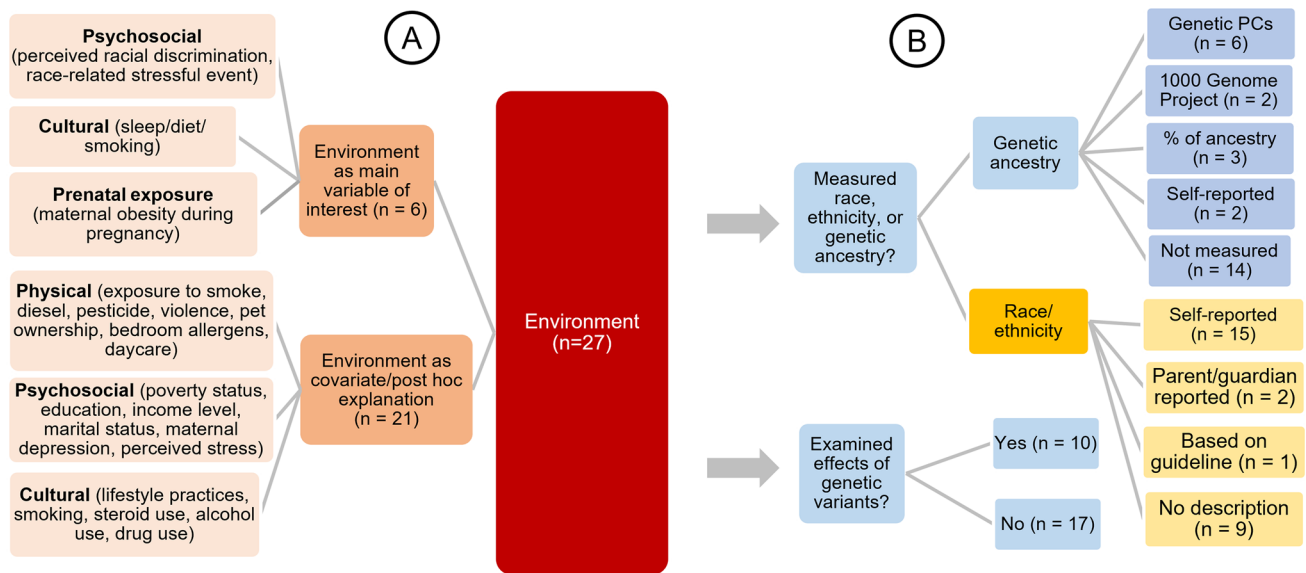
Fig. 4 Summary of 27 studies examining environmental factors with **A** categorization of environmental factors as physical, psychosocial, and cultural, and **B** measures of race and ethnicity/genetic ancestry and genetic variant effects

molecular markers, such as DNAm, although there are methodological limitations to measures of epigenetic aging due to the homogeneity of most DNAm data utilized for developing these tools.

A particularly severe source of psychosocial stress involving race and ethnicity occurs during immigration, especially for those experiencing displacement due to warfare or other sociopolitical crises. Cumulative stressors include, but are not limited to, leaving a familiar environment and social support, becoming a racial and/or ethnic minority in a new society, and the pressure of acculturation, which together represent sources of daily stress that may become biologically embedded. To our knowledge, there have been few studies of the link between differential DNAm and immigration-related stress. Two studies directly examining immigration-related stress demonstrated its association with epigenetic age [110] and differential DNAm at stress-related candidate genes, such as *SLC6A4* [111]. Other studies indirectly addressing this topic compared immigrants with non-immigrants. While some studies echoed the so-called "healthy immigrant" phenomenon referring to a health advantage of immigrants, as indicated by lower epigenetic age acceleration, compared to the native non-migrant population in the home country [112], others did not [84]. These reports showed that whether immigrants do indeed have better health is a nuanced issue, which has also been challenged by a longitudinal study showing convergence of health among immigrants and the native population in the host country after 10–20 years [113]. However, none of the studies we reviewed directly measured perceived immigration or acculturation stress to examine the underlying factors that may explain their findings.

These psychosocial stressors can be associated with the psychology of caregivers, including their beliefs, values, and feelings. For example, African American and Latino mothers, especially those who have experienced racism, are more likely to promote and discuss their children's racial or ethnic identities than White mothers [114, 115]. Caregiver's psychology, both pre- and postnatally, including mental health and experience of psychosocial stress, can become biologically embedded in their offspring through DNAm [30, 116]. While most DNAm studies considered maternal factors, such as smoking behaviors, health, and socioeconomic status, few studies have taken into account maternal psychology (e.g., depression) in the context of race and ethnicity [64]. It is also worth noting that very few studies have examined paternal behavioral and psychosocial factors, such as the father's role in caregiving, their mental health, and their influences on mothers via marital relationships or financial and social support, and their relations with DNAm [117–119]. Furthermore, the effects of race-associated adversity, such as racism and discrimination, are still largely understudied in epigenetics research [30].

7.2.3 Cultural environment

Cultural environments are constructed by culturally regulated customs commonly used and accepted by the members of the community of which group members may not even be conscious [60]. Customs are shaped by the psychology of caregivers,

including their beliefs concerning their children's needs and appropriate behaviors, and can be bidirectionally related to the physical and social settings. Cultural practices, particularly lifestyle differences, including drinking and smoking behaviors, sleep, diet, exercise, social networks, and customs surrounding elder care and childbearing and rearing (e.g., breastfeeding, physical touch, and parenting practices), have all been linked to health and development, potentially through adaptations to these cultural environments within physiological systems and the ensuing possibility of biological embedding as measured by DNAm [6, 120–122]. For example, caregiving behaviors, such as breastfeeding, vary considerably across racial and ethnic groups even after adjusting for socioeconomic and psychosocial factors. In particular, mothers from racially or ethnically minoritized groups in the USA who are less acculturated to mainstream practices are more likely to breastfeed [123, 124]. There is substantial evidence in the literature that both the practice and duration of breastfeeding are associated with children's DNAm of genes implicated in long-term health issues, such as obesity and diabetes [122, 125]. However, additional research regarding other relevant maternal factors related to the effects of breastfeeding, such as diet, is needed. Indeed, diet is another culturally regulated custom that has attracted considerable interest in epigenetics research. The growing field of nutritional epigenetics has demonstrated robust associations of diet quality and nutrition with DNAm, including in dietary clinical trials [126]. Specifically, the "Mediterranean diet," originating in Southern Italy and characterized by a high proportion of fish, similar to diets commonly found in Japanese culture, has been shown to reduce the rate of cardiovascular disease and to be linked to lower epigenetic age acceleration [127]. Similarly, there is emerging evidence for associations of traditional Indian and Chinese medicine with epigenetic modifications, including DNAm [128, 129]. Thus, the cultural environment comprised of culturally regulated practices and behaviors that vary across individuals with different racial or ethnic identities can contribute to differences in DNAm.

8 Considering intrinsic and extrinsic factors together

The relative effects of extrinsic and intrinsic factors can only be disentangled from each other when both are examined empirically. Among the 27 studies included in the review that examined environmental effects, only 10 also empirically investigated the effects of genetic variations on DNAm (Fig. 4b). In particular, two studies examined the associations between differential DNAm and extrinsic factors in conjunction with racial and ethnic identity and genetic effects. Unsurprisingly, both of these studies revealed strong genetic effects such that shared genetic ancestry explained 75% of the variance of DNAm associated with ethnicity [79] and 51% of race-related differentially methylated CpGs were associated with at least one mQTL [64]. However, DNAm of 314 of 916 CpGs associated with self-reported ethnicity remained significantly related to ethnicity even after adjusting for genetic ancestry [79], indicating remaining effects of ethnicity on DNAm that could not be explained by shared genetic ancestry. These studies also examined environmental exposures to shed light on factors capable of explaining differential DNAm associated with race or ethnicity beyond genetic effects. The results showed that some environmental factors known to vary by ethnicity or culture, such as maternal smoking during pregnancy, but not other exposures, such as maternal depression, were correlated with differential DNAm at CpGs whose associations with ethnicity could not be explained by ancestry alone, thus providing a possible environmental pathway for racial and ethnic differences [79]. These studies serve as exemplars to begin to untangle the relative effects of often-confounded genetic and environmental factors on differential DNAm associated with race and ethnicity.

In addition to considering both genetic and environmental effects together, it is also important to examine the interactions between genetics and environment, as recent studies indicated that DNAm at most CpGs can be best explained by gene-by-environment interaction models [23, 130]. Indeed, some studies showed that the relations between DNAm and lifestyle factors, such as cigarette smoking and sleep, varied according to racial and ethnic groups [80, 91]. However, these studies did not discuss their interpretation of these ethnicity-dependent findings and did not report how they accounted for genetic variations across groups. Therefore, it remains unclear whether these findings are a result of genetic effects or cultural specificity.

9 Application of the integrative framework in addressing race and ethnicity in future DNAm research

As illustrated by the review of the current state of DNAm studies in the context of race and ethnicity outlined above, limitations remain in the literature on DNAm in addressing racial and ethnic differences and their underlying drivers. Nonetheless, we recognize that social epigenomics is a rapidly growing field and authors of previous studies were likely to have used the most robust and appropriate methods and analyses given the knowledge and technology available

at the time. Building on the valuable lessons learned from previous studies and the integrative framework proposed here, we provide recommendations for how to apply the framework and conduct more nuanced and methodologically robust research in future (Table 1).

9.1 Application of the proposed integrative framework

The proposed integrative framework can be applied to future research to answer important questions about factors contributing to DNAm variation across racial and ethnic groups in a biologically and culturally sensitive manner. This framework can be used as a guide to which relevant factors should be considered at the stages of conceptualization, research design, data analysis, and interpretation. In particular, the framework illuminates the complex relations between intrinsic and extrinsic factors relevant to both DNAm and race and ethnicity, which should be taken into consideration together to avoid findings that are impossible to interpret and the reaching of inaccurate conclusions. Taking the study of racial disparities in birth outcomes as an example, an overemphasis on genetic differences as explanations may be used to propagate racist narratives. Hence, in addition to genetic variations and other intrinsic factors, it is necessary to investigate how extrinsic factors, such as structural and sociocultural factors (e.g., maternal exposure to racism and diet) [23], may drive some of the biological signals by themselves and/or through their interactions with genetic variations. Specifically with regard to extrinsic factors, our framework extends beyond physical and social factors that are more commonly studied in DNAm research and highlights cultural environmental factors, such as caregiver practices, that may regulate the developmental niches of individuals during early life and contribute to variations in DNAm. In the following sections, we provide actionable recommendations to uncover DNAm signals associated with race and ethnicity, while

Table 1 Short and long-term recommendations for addressing race and ethnicity in future DNAm research

Short-term recommendations	*Intrinsic factors: Self-identified race and ethnicity*
	1. Determine the theoretical and/or biological reasoning for including race and ethnicity in the purpose of the research
	2. Clearly define and describe how race and ethnicity are measured and conceptualized, and avoid the interchangeable use of these terms
	Intrinsic factor: Genetic ancestry
	3. Describe genetic ancestry along a continuum, and racial derived groupings or ethnicity categories or continent-level ancestry clusters should not be used as proxies for genetic ancestry
	4. Account for population structure, specifically genetic heterogeneity, when possible and appropriate based on the goals of the research. This may include stratification, interaction, or adjusting the model for genetic covariates
	(DNAm-based tools may be used to account for population stratification if genotype data are not available.)
	Intrinsic factors: Genetic variations
	5. Examine the effects of genetic variants
	(Use mQTL databases and/or plot the distribution of CpGs to identify potential mQTLs if genotype data are not available.)
	6. Acknowledge limitations of currently available data and tools
	Intrinsic & extrinsic factors
	7. Consider interactions between genes and environment as well as between culture and environment
	8. Leverage interdisciplinary collaborations and engage local stakeholders in the research process to consider relevant intrinsic and extrinsic factors
Long-term recommendations	*Extrinsic factors*
	1. Expand metadata collection to structural and sociocultural factors relevant to race and ethnicity by adopting interdisciplinary perspectives
	2. Extend research to examine more associations within the same individual over time, such as in the context of interventions and longitudinal designs
	3. Support and fund multidisciplinary and global teams to collect data and interpret results in diverse populations sensitively and appropriately
	Intrinsic factors
	4. Improve methodology for characterizing genetic heterogeneity in multiracial and multiethnic individuals
	5. Increase diversity of genetic ancestry and racial and ethnic groups in DNAm research and mQTL databases
	6. Update DNAm tools to include more genetic, racial, and ethnic diversity in their training and testing sets
	Intrinsic and extrinsic factors
	7. Compare individuals with different experiences or environments within the same genetic ancestry, including different ethnic and racial identities

minimizing the risks of observing signals originating from methodological issues, inherent confounding factors, and/or misinterpretation and inappropriate results.

9.2 Recommendations for addressing race and ethnicity in future DNAm research

In terms of intrinsic factors, most demographic variables, such as sex and age, are commonly collected and can be included in the analysis. However, some factors that should be taken into account, e.g., genetics, may not be available in existing cohorts for which data collection has already been completed, thus posing challenges to good practices in future studies. Despite the argument to withhold investigation of race and ethnicity in the context of DNAm until appropriate data become available, we contend that, as researchers, we have an obligation to fully utilize the valuable biological data provided by our study participants in the most responsible way possible. Here, we present short-term recommendations for handling existing data and suggest possible ways to overcome potential shortcomings inherent in the data, in addition to long-term goals in relation to collecting the types of data, especially those related to extrinsic factors, that will enable more appropriate and culturally sensitive investigation of DNAm in the context of race and ethnicity.

9.2.1 Short-term recommendations

The most important constructs among the intrinsic factors in the framework are self-identified race and ethnicity. However, more than half of the studies included in the review outlined above did not describe how race and ethnicity were measured. The first steps to achieve sensitivity and awareness of the complexity of issues related to race and ethnicity are to determine the research goals and how race and ethnicity are involved in the hypothesis [17]. Then, a clearly defined operational definition and clarification of any related terms used, explicitly acknowledging how race and/or ethnicity were measured in the study, and with a critical view regarding what the construct truly represents in a given case are critical for both contextual understanding and interpretation [6, 16, 17, 131, 132].

As discussed above and illustrated in our framework, genetic ancestry and genetic variants are important intrinsic factors closely related to race and ethnicity as well as DNAm, and it is important to account for genetics when possible and appropriate [11]. Furthermore, as the world becomes increasingly diverse and populations are becoming increasingly multiracial, researchers must recognize the complexity of the genetic structures of admixed populations and the increasing importance of describing and analyzing genetic ancestry along a continuous spectrum, both across diverse groups and within seemingly homogeneous groups. In addition, tools are currently available to estimate global fractions of genetic ancestry by individual and, when examining specific areas of the epigenome, local ancestry or the ancestry group from which a particular SNP is most likely to have originated. Together, these approaches provide multiple bioinformatics opportunities to analyze DNAm in the context of genetics in a both sociologically sensitive and biologically accurate manner. Understandably, genotype data may not be available for some cohorts and/or ethnicity information may be incomplete/missing, making it difficult to account for population differences in the study. While not without limitations, a number of DNAm-based tools that allow inference of population structure from DNAm data are available, including EPISTRUCTURE, PlaNET, and Barfield's SNP-based filtering [44, 134, 135]. In addition to these tools and in the absence of genotyping data, SNP-based DNAm effects can be identified by taking advantage of existing mQTL databases, while acknowledging the limitations of their Eurocentricity. When examining a population or tissue not represented in these databases, it is also possible to use the modality of DNAm (e.g., with the *nmodel* function in the R package ENmix) [136] to identify likely genetic influences, specifically characterizing multimodal CpGs, such as those with a trimodal pattern, indicating DNAm clustered around haplotypes [83]. Finally, CpGs of interest can be compared with those shown previously to be associated with racial and ethnic identity and/or genetic ancestry (see Additional file 2: Table S2 as a resource for future research).

A growing body of research has demonstrated the importance of considering the interactions between genes and environment in studies of DNAm [23], suggesting that individuals with certain genetic backgrounds may have greater susceptibility to some environmental exposures [79]. Similarly, cultural research has also documented social factors, such as perceived social support, which may differentially contribute to an individual's health depending on their cultural background [137]. Cultural specificity in the effects of environmental exposures/experiences on DNAm can be identified by examining the moderating effects of racial or ethnic identities on DNAm when individuals from different groups are exposed to the same environmental factors [91]. Future research should carefully consider multiple layers of interaction (e.g., gene-by-environment, racial and ethnic identity-by-environment), and examine the relative impacts of genetics and cultural background on an individual's susceptibility to certain extrinsic factors.

Several statistical frameworks can be applied to address the interconnectedness of these biopsychosocial constructs, such as: model adjustment, which mathematically accounts for the influence of genetic heterogeneity, however it is defined operationally in a given study, on the association between the variable of interest and DNAm [138]; stratification, which can be helpful in the case of distinct groups or otherwise unavoidable confounding factors [139]; mixed models, which are increasingly being used in the genomics space to address population stratification as an issue of kinship, with those of different ancestries being slightly less related than those of similar ancestries, and they may also be applied to DNAm [139]; interaction terms, also known as moderation with hypothesized directionality, can be used to identify CpGs with differing relations to the variable of interest depending on genetic ancestry [80, 91, 140]; and when temporally and theoretically appropriate, mediation models can be adopted to examine whether cultural differences in extrinsic factors, such as customs, have some explanatory power in the associations of race and ethnicity with DNAm, and are widely used in cross-cultural comparisons in psychological research [141]. However, many of these methods have larger statistical power requirements, and it is therefore highly recommended that they be coupled with biologically informed dimension reduction methods to reduce the multiple testing burden of EWASs, such as with the CoMeBack or coMET packages in R [142, 143]. As recommended above, when direct measures of pertinent environments are not available, researchers may also compare their discovered CpGs with race- or ethnicity-related CpGs identified in previous studies of environmental effects on DNAm to infer their relevance [95, 144].

Furthermore, as the study of race and ethnicity in epigenetics requires perspectives from both biological and social sciences, an interdisciplinary team with a range of expertise is necessary. This is also highlighted in other review papers on social epigenomics, including a previous paper reviewing the same set of studies [6, 133]. More importantly, it is crucial to ensure that local stakeholders of the racial and ethnic or cultural communities investigated are involved in the research process to ensure local contexts, values, and goals are preserved, respected, and incorporated into the study [15]. Inclusion of local researchers and other ways to ensure diversity in research teams are nicely described by the *Nature* guidelines on "Authorship: Inclusion and Ethics in Global Research," which should be used as a reference when planning a study. Taken together, these actions represent important steps to slowly regain the trust of minoritized communities in research.

9.2.2 Long-term recommendations

Our framework recommends considering both intrinsic (e.g., genetic) and extrinsic factors while acknowledging the challenge of untangling their effects. One possible solution is to examine individuals within the same genetic ancestry but with a diverse range of experience (e.g., immigration). Some studies compared individuals sharing the same genetic ancestry (i.e., African ancestry) but with different lifestyles and different habitats (e.g., hunter-gatherers or farmers) to assess the degree of intrapopulation variation in DNAm [84, 145]. Such study designs can help to decipher the relative contributions of racial and ethnic disparities in health outcomes [135]. To disentangle the impacts of sociostructural factors from genetic ancestry on DNAm, future studies should compare DNAm of individuals with the same genetic ancestry but different experience, for example, comparison of individuals of African ancestry residing in the USA with those in Africa and with fewer socioeconomic disadvantages.

To avoid the propagation of harmful stereotypes and unwanted misinterpretation of disparities as genetically determined, a more nuanced understanding of structural and sociocultural factors underlying racial and ethnic differences is required. Hence, researchers should expand metadata collection to include environmental exposures and experience relevant to race and ethnicity as denoted by our framework. Only about half of the studies reviewed here examined environmental factors, with very few focusing primarily on environmental effects. As suggested in previous reviews, further investigations of the associations of DNAm with stress and adversity associated with race, such as racism especially during prenatal exposure, are needed [6, 146]. Additional efforts are also required to develop and validate measures capable of capturing important experiences related to race and ethnicity, such as racial trauma and acculturation stress [100, 133].

Interdisciplinary teams consisting of experts in various fields in the biological and social sciences are more likely to have the expertise necessary to identify relevant sociocultural measures that should be included in data collection in studies of DNAm. Furthermore, substantive knowledge across disciplines can ensure that relevant questions and appropriate analyses are used to answer questions related to how sociocultural and structural environments contribute to DNAm in the context of race and ethnicity [133, 147].

It is important to acknowledge that most of the current empirical evidence on the relations between human DNAm and the sociocultural environment related to race and ethnicity remains correlational and does not imply causation. Although we have primarily discussed the impacts of environment and experience on DNAm, we cannot exclude the

possibility of bidirectional relations between these factors. Experiments with animal models have yielded some evidence supporting the effects of the social environment, such as maternal caregiving, on DNAm [148], while similar experiments remain impossible in humans for ethical reasons. Nonetheless, there is growing interest in studying DNAm in the context of interventions, and in performing longitudinal studies capable of illuminating how changes in social conditions and environment/experience may be linked to changes in DNAm over time. Specifically, examinations of the longitudinal associations of DNAm with developmental changes in racial and/or ethnic identity may elucidate some of these gene-by-environment effects within the psychosocial perspective. Longitudinal research guided by positive strengths-based approaches can identify changeable extrinsic factors, which can potentially help buffer the negative effects of race-associated stress and adversity on health and provide insight into the associated DNAm processes [100].

In addition, steps should be taken to increase the diversity within both samples and research teams in DNAm research [149]. The previous review reported that the majority of studies on DNAm related to race and ethnicity are still led by researchers based in the USA (77%) [6], including primarily White researchers, and available epigenetic data, both published and biobanked, are largely of populations of European ancestry [149]. The scientific community as a whole should endeavor to support diversity in both the participants entrusting us with their data and in the researchers designing and interpreting the studies [133].

Missing diversity in DNAm research has strong implications for the creation of databases, such as mQTL repositories, and the development of DNAm-based molecular tools, such as methylation risk score (MRS), a score reflective of an individual's epigenetic susceptibility to certain phenotypes based on the methylation state at CpGs previously identified as linked to the phenotype [150], proxy biomarkers, such as the DNAm-based interleukin-6 (IL-6) predictor [151], and the widely used epigenetic clocks [152]. Comparison across 13 epigenetic clocks, such as the two first-generation clocks (Horvath pan-tissue and Hannum clocks) and 11 second-generation clocks, including GrimAge and the newly developed Pace of Aging clock, showed inconsistent relations of epigenetic age acceleration with race and ethnicity [153]. These results may have been partly due to the small numbers of non-White samples in the training sets and different genetic influences on epigenetic profiles across race and ethnic groups. For example, there is a Eurocentric bias in the discovery cohorts (44%–85% European ancestry) for the three most commonly used epigenetic clocks, i.e., the Horvath pan-tissue clock, the Hannum clock, and the PhenoAge clock [152]. As some genetic variants differ in frequency across genetic ancestries, a DNAm tool developed primarily with data of European ancestry may not be applicable to individuals with non-European genetic ancestry. The Horvath pan-tissue clock is known to be more likely to overestimate the age of individuals who identify as Black Americans and underestimate the age of those who identify as Latin Americans, although whether this is a result of genetics, culture, or societal context, or all of these factors together, remains unclear [84, 152]. In addition, a genome-wide association study (GWAS) of the Horvath and Hannum clocks found substantial genetic associations with epigenetic age acceleration predicted by these clocks, indicating a strong likelihood of an influence of genetic heterogeneity [154]. Meanwhile, the DNAm-based IL-6 predictor was created with an exclusively European cohort [151], even though an allele associated with the density of soluble IL-6 receptors and circulating IL-6 levels has been found by admixture mapping to have frequencies of 4% in West Africans and 35% in European Americans [155]. Given the current limitations of epigenetic clocks and their popularity in both research and the public space, it is important to exercise caution and acknowledge their limitations when examining epigenetic age acceleration in racial and ethnic groups that are not typically included in the training sets. Therefore, future research should involve the recruitment and collection of data from groups with more diverse racial and ethnic backgrounds and genetic ancestries, and the creation of more diverse mQTL databases.

10 Conclusion

To move toward conceptualization of human biology as social biology, it is crucial for epigenomics research to take into consideration the sociocultural and structural components underlying self-identified racial and ethnic identity in a methodologically robust and culturally sensitive manner. Hence, our framework advocates a holistic and comprehensive perspective to consider both biological and social factors while incorporating different lenses to promote a productive conversation toward future DNAm research related to race and ethnicity. To facilitate culturally sensitive epigenetics research in the context of race and ethnicity, this paper provided a guiding framework to help recognize and identify relevant structural and sociocultural factors that can contribute to DNAm. At the same time, when addressing the social determinants of health through DNAm, a robust investigation of social impact must include analyses taking into

account the underlying genetic architecture upon which DNAm is built and the potential interactions between genes and the environment. With the expanding opportunities for cross-disciplinary collaborations and availability of increasing amounts of biosocial data, epigenetics research has tremendous potential to continue to elucidate health disparities across race and ethnicity.

Data availability and code availability No empirical data or code were used in this study. Information from previous studies reviewed in this paper are included in the Supplementary Material.

Declarations

Competing interests The authors declare no competing interests.

References

1. Slavich GM, Cole SW. The emerging field of human social genomics. Clin Psychol Sci. 2013;1:331–48. https://doi.org/10.1177/2167702613 478594.
2. Cole SW. Social regulation of human gene expression: mechanisms and implications for public health. Am J Public Health. 2013;103:S84-92. https://doi.org/10.2105/AJPH.2012.301183.
3. Cole SW. Human social genomics. PLOS Genet. 2014;10:e1004601. https://doi.org/10.1371/journal.pgen.1004601.
4. Hertzman C. The biological embedding of early experience and its effects on health in adulthood. Ann N Y Acad Sci. 1999;896:85–95. https://doi.org/10.1111/j.1749-6632.1999.tb08107.x.
5. Salas LA, Peres LC, Thayer ZM, Smith RW, Guo Y, Chung W, et al. A transdisciplinary approach to understand the epigenetic basis of race/ethnicity health disparities. Epigenomics. 2021;13:1761–70. https://doi.org/10.2217/epi-2020-0080.
6. Meloni M, Moll T, Issaka A, Kuzawa CW. A biosocial return to race? A cautionary view for the postgenomic era. Am J Hum Biol. 2022. https://doi.org/10.1002/ajhb.23742.
7. Cooper RS. A case study in the use of race and ethnicity in public health surveillance. Public Health Rep Wash DC. 1974;1994(109):46–52.
8. Williams DR, Sternthal M. Understanding racial-ethnic disparities in health: sociological contributions. J Health Soc Behav. 2010;51:S15-27. https://doi.org/10.1177/0022146510383838.
9. Deyrup A, Graves JL. Racial biology and medical misconceptions. N Engl J Med. 2022;386:501–3. https://doi.org/10.1056/NEJMp2116224.
10. Mersha TB, Abebe T. Self-reported race/ethnicity in the age of genomic research: Its potential impact on understanding health disparities. Hum Genomics. 2015;9:1–15. https://doi.org/10.1186/s40246-014-0023-x.
11. Rehkopf DH, Needham BL. The impact of race and ethnicity in the social epigenomic regulation of disease. Elsevier. 2020. https://doi.org/10.1016/B978-0-12-816843-1.00004-7.
12. Li A, Deyrup AT, Graves JL, Ross LF. Race in the reading: a study of problematic uses of race and ethnicity in a prominent pediatrics textbook. Acad Med. 2022;97:1521–7. https://doi.org/10.1097/ACM.0000000000004666.
13. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. Science. 2002;298:2381–5. https://doi.org/10.1126/science.1078311.
14. Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG, et al. Population genetic structure of variable drug response. Nat Genet. 2001;29:265–9. https://doi.org/10.1038/ng761.
15. Khan AT, Gogarten SM, McHugh CP, Stilp AM, Sofer T, Bowers ML, et al. Recommendations on the use and reporting of race, ethnicity, and ancestry in genetic research: Experiences from the NHLBI TOPMed program. Cell Genomics. 2022;2:100155. https://doi.org/10.1016/j.xgen.2022.100155.
16. Lu C, Ahmed R, Lamri A, Anand SS. Use of race, ethnicity, and ancestry data in health research. PLOS Glob Public Health. 2022;2:e0001060. https://doi.org/10.1371/journal.pgph.0001060.

17. Kaufman JS, Cooper RS. Commentary: considerations for use of racial/ethnic classification in etiologic research. Am J Epidemiol. 2001;154:291–8. https://doi.org/10.1093/aje/154.4.291.

18. Moscou S, Anderson MR, Kaplan JB, Valencia L. Validity of racial/ethnic classifications in medical records data: an exploratory study. Am J Public Health. 2003;93:1084–6. https://doi.org/10.2105/AJPH.93.7.1084.

19. Baker DW, Cameron KA, Feinglass J, Thompson JA, Georgas P, Foster S, et al. A system for rapidly and accurately collecting patients' race and ethnicity. Am J Public Health. 2006;96:532–7. https://doi.org/10.2105/AJPH.2005.062620.

20. US Census Bureau. 2020 Census illuminates racial and ethnic composition of the country. CensusGov. n.d. https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html. Accessed 5 Jan 2023.

21. Aristizabal MJ, Anreiter I, Halldorsdottir T, Odgers CL, McDade TW, Goldenberg A, et al. Biological embedding of experience: a primer on epigenetics. Proc Natl Acad Sci U S A. 2020;117:23261–9. https://doi.org/10.1073/pnas.1820838116.

22. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A. 2005;102:10604–9. https://doi.org/10.1073/pnas.0500398102.

23. Teh AL, Pan H, Chen L, Ong M-L, Dogra S, Wong J, et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. Genome Res. 2014;24:1064–74. https://doi.org/10.1101/gr.171439.113.

24. Jones MJ, Moore SR, Kobor MS. Principles and challenges of applying epigenetic epidemiology to psychology. Annu Rev Psychol. 2018;69:459–85. https://doi.org/10.1146/annurev-psych-122414-033653.

25. Chowdhury-Paulino IM, Ericsson C, Vince R Jr, Spratt DE, George DJ, Mucci LA. Racial disparities in prostate cancer among Black men: epidemiology and outcomes. Prostate Cancer Prostatic Dis. 2022;25:397–402. https://doi.org/10.1038/s41391-021-00451-z.

26. Prakash O, Hossain F, Danos D, Lassak A, Scribner R, Miele L. Racial disparities in triple negative breast cancer: a review of the role of biologic and non-biologic factors. Front Public Health. 2020. https://doi.org/10.3389/fpubh.2020.576964.

27. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou Y-H, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. Hum Mol Genet. 2015;24:4464–79. https://doi.org/10.1093/hmg/ddv161.

28. Kuzawa CW, Sweet E. Epigenetics and the embodiment of race: Developmental origins of US racial disparities in cardiovascular health. Am J Hum Biol. 2009;21:2–15. https://doi.org/10.1002/ajhb.20822.

29. Pepin ME, Ha C-M, Potter LA, Bakshi S, Barchue JP, Haj Asaad A, et al. Racial and socioeconomic disparity associates with differences in cardiac DNA methylation among men with end-stage heart failure. Am J Physiol-Heart Circ Physiol. 2021;320:H2066–79. https://doi.org/10.1152/ajpheart.00036.2021.

30. Conradt E, Carter SE, Crowell SE. Biological embedding of chronic stress across two generations within marginalized communities. Child Dev Perspect. 2020;14:208–14. https://doi.org/10.1111/cdep.12382.

31. Tajuddin SM, Hernandez DG, Chen BH, Noren Hooten N, Mode NA, Nalls MA, et al. Novel age-associated DNA methylation changes and epigenetic age acceleration in middle-aged African Americans and whites. Clin Epigenetics. 2019;11:119. https://doi.org/10.1186/s13148-019-0722-1.

32. Adams LM, Miller AB. Mechanisms of mental-health disparities among minoritized groups: how well are the top journals in clinical psychology representing this work? Clin Psychol Sci. 2022;10:387–416. https://doi.org/10.1177/21677026211026979.

33. Fortuna LR, Tolou-Shams M, Robles-Ramamurthy B, Porche MV. Inequity and the disproportionate impact of COVID-19 on communities of color in the United States: The need for a trauma-informed social justice response. Psychol Trauma Theory Res Pract Policy. 2020;12:443–5. https://doi.org/10.1037/tra0000889.

34. McDade TW, Harris KM. From society to cells and back again: new opportunities for discovery at the biosocial interface. Discov Soc Sci Health. 2022;2:4. https://doi.org/10.1007/s44155-022-00007-z.

35. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. JAMA. 2020;323:2466–7. https://doi.org/10.1001/jama.2020.8598.

36. Fine MJ, Ibrahim SA, Thomas SB. The Role of Race and Genetics in Health Disparities Research. Am J Public Health. 2005;95:2125–8. https://doi.org/10.2105/AJPH.2005.076588.

37. Mancilla VJ, Peeri NC, Silzer T, Basha R, Felini M, Jones HP, et al. Understanding the interplay between health disparities and epigenomics. Front Genet. 2020;11:1–14. https://doi.org/10.3389/fgene.2020.00903.

38. Peterson RE, Kuchenbaecker K, Walters RK, Chen C-Y, Popejoy AB, Periyasamy S, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. Cell. 2019;179:589–603. https://doi.org/10.1016/j.cell.2019.08.051.

39. Duggan CP, Kurpad A, Stanford FC, Sunguya B, Wells JC. Race, ethnicity, and racism in the nutrition literature: an update for 2020. Am J Clin Nutr. 2020;112:1409–14. https://doi.org/10.1093/ajcn/nqaa341.

40. Neufcourt L, Castagné R, Mabile L, Khalatbari-Soltani S, Delpierre C, Kelly-Irving M. Assessing how social exposures are integrated in exposome research: a scoping review. Environ Health Perspect. 2022;130:116001. https://doi.org/10.1289/EHP11015.

41. McCartney G, Popham F, McMaster R, Cumbers A. Defining health and health inequalities. Public Health. 2019;172:22–30. https://doi.org/10.1016/j.puhe.2019.03.023.

42. Singh G, Morrison J, Hoy W. DOHaD in Indigenous populations: DOHaD, epigenetics, equity and race. J Dev Orig Health Dis. 2019;10:63–4. https://doi.org/10.1017/S2040174419000023.

43. Fletcher JM, Schwarz H, Engelman M, Johnson NJ, Hakes J, Palloni A. Understanding Geographic Disparities in Mortality. Demography. 2023;10609710. https://doi.org/10.1215/00703370-10609710.

44. Barfield RT, Almli LM, Kilaru V, Smith AK, Mercer KB, Duncan R, et al. Accounting for population stratification in DNA methylation studies. Genet Epidemiol. 2014;38:231–41. https://doi.org/10.1002/gepi.21789.

45. Lee BA, Maghsoodi AH, Ruedas-Gracia N, Williams CD. Measures of racial-ethnic-cultural (REC) belonging: a systematic review of the literature. Identity. 2022;80:1–23. https://doi.org/10.1080/15283488.2022.2127725.

46. Burrow AL, Tubman JG, Montgomery MJ. Racial identity: toward an integrated developmental psychological perspective. Identity. 2006;6:317–39. https://doi.org/10.1207/s1532706xid0604_2.

47. Padilla JB, Vargas JH, Chavez HL. Influence of age on transracial foster adoptions and its relation to ethnic identity development. Adopt Q. 2010;13:50–73. https://doi.org/10.1080/10926751003662598.

48. Umaña-Taylor AJ, Quintana SM, Lee RM, Cross WE Jr, Rivas-Drake D, Schwartz SJ, et al. Ethnic and racial identity during adolescence and into young adulthood: an integrated conceptualization. Child Dev. 2014;85:21–39. https://doi.org/10.1111/cdev.12196.

49. Roth WD, Ivemark B. Genetic options: the impact of genetic ancestry testing on consumers' racial and ethnic identities. Am J Sociol. 2018;124:150–84. https://doi.org/10.1086/697487.

50. Yip T, Wang Y, Mootoo C, Mirpuri S. Moderating the association between discrimination and adjustment: a meta-analysis of ethnic/racial identity. Dev Psychol. 2019;55:1274–98. https://doi.org/10.1037/dev0000708.

51. Goodstein R, Ponterotto JG. Racial and ethnic identity: their relationship and their contribution to self-esteem. J Black Psychol. 1997;23:275–92. https://doi.org/10.1177/00957984970233009.

52. Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung H-C, et al. Genotype, haplotype and copy-number variation in worldwide human populations. Nature. 2008;451:998–1003. https://doi.org/10.1038/nature06742.

53. Yang H-C, Chen C-W, Lin Y-T, Chu S-K. Genetic ancestry plays a central role in population pharmacogenomics. Commun Biol. 2021;4:1–14. https://doi.org/10.1038/s42003-021-01681-6.

54. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, et al. The genetic structure and history of Africans and African Americans. Science. 2009;324:1035–44. https://doi.org/10.1126/science.1172257.

55. Gomez F, Hirbo J, Tishkoff SA. Genetic variation and adaptation in Africa: implications for human evolution and disease. Cold Spring Harb Perspect Biol. 2014;6:e008524. https://doi.org/10.1101/cshperspect.a008524.

56. Norris ET, Wang L, Conley AB, Rishishwar L, Mariño-Ramírez L, Valderrama-Aguirre A, et al. Genetic ancestry, admixture and health determinants in Latin America. BMC Genomics. 2018;19:861. https://doi.org/10.1186/s12864-018-5195-7.

57. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64. https://doi.org/10.1101/gr.094052.109.

58. Loh P-R, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, et al. Inferring admixture histories of human populations using linkage disequilibrium. Genetics. 2013;193:1233–54. https://doi.org/10.1534/genetics.112.147330.

59. Worthman CM. The ecology of human development: evolving models for cultural psychology. J Cross-Cult Psychol. 2010;41:546–62. https://doi.org/10.1177/0022022110362627.

60. Super CM, Harkness S. The developmental niche: a conceptualization at the interface of child and culture. Int J Behav Dev. 1986;9:545–69. https://doi.org/10.1177/016502548600900409.

61. Super CM, Harkness S. The environment as culture in developmental research. In: Friedman SL, Wachs TD, editors. Meas. Environ. Life Span Emerg. Methods Concepts. Washington: American Psychological Association; 1999. pp. 279–323. https://doi.org/10.1037/10317-010.

62. Bergstedt J, Azzou SAK, Tsuo K, Jaquaniello A, Urrutia A, Rotival M, et al. The immune factors driving DNA methylation variation in human blood. Nat Commun. 2022;13:5895. https://doi.org/10.1038/s41467-022-33511-6.

63. Belsky J, Pluess M. Beyond diathesis stress: differential susceptibility to environmental influences. Psychol Bull. 2009;135:885–908. https://doi.org/10.1037/a0017376.

64. McKennan C, Naughton K, Stanhope C, Kattan M, Connor GTO. Longitudinal data reveal strong genetic and weak non-genetic components of ethnicity-dependent blood DNA methylation levels. Epigenetics. 2021;16:662–76. https://doi.org/10.1080/15592294.2020.1817290.

65. Zhu H, Bhagatwala J, Huang Y, Pollock NK, Parikh S, Raed A, et al. Race/ethnicity-specific association of vitamin D and global DNA methylation: cross-sectional and interventional findings. PLoS ONE. 2016;11:e0152849. https://doi.org/10.1371/journal.pone.0152849.

66. Paredes-Céspedes DM, Rojas-García AE, Medina-Díaz IM, Ramos KS, Herrera-Moreno JF, Barrón-Vivanco BS, et al. Environmental and socio-cultural impacts on global DNA methylation in the indigenous Huichol population of Nayarit. Mexico Environ Sci Pollut Res. 2021;28:4472–87. https://doi.org/10.1007/s11356-020-10804-1.

67. Mozhui K, Smith AK, Tylavsky FA. Ancestry dependent DNA methylation and influence of maternal nutrition. PLoS ONE. 2015;10:e0118466. https://doi.org/10.1371/journal.pone.0118466.

68. Pauker K, Meyers C, Sanchez DT, Gaither SE, Young DM. A review of multiracial malleability: Identity, categorization, and shifting racial attitudes. Soc Personal Psychol Compass. 2018;12:e12392. https://doi.org/10.1111/spc3.12392.

69. Sanchez DT, Shih M, Garcia JA. Juggling multiple racial identities: malleable racial identification and psychological well-being. Cultur Divers Ethnic Minor Psychol. 2009;15:243–54. https://doi.org/10.1037/a0014373.

70. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, et al. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. BMC Genomics. 2014;15:145. https://doi.org/10.1186/1471-2164-15-145.

71. Villicaña S, Bell JT. Genetic impacts on DNA methylation: research findings and future perspectives. Genome Biol. 2021;22:127. https://doi.org/10.1186/s13059-021-02347-6.

72. Islam SA, Goodman SJ, MacIsaac JL, Obradović J, Barr RG, Boyce WT, et al. Integration of DNA methylation patterns and genetic variation in human pediatric tissues help inform EWAS design and interpretation. Epigenetics Chromatin. 2019;12:1–18. https://doi.org/10.1186/s13072-018-0245-6.

73. Cheung WA, Shao X, Morin A, Siroux V, Kwan T, Ge B, et al. Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. Genome Biol. 2017;18:50. https://doi.org/10.1186/s13059-017-1173-7.

74. Do C, Lang CF, Lin J, Darbary H, Krupska I, Gaba A, et al. Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. Am J Hum Genet. 2016;98:934–55. https://doi.org/10.1016/j.ajhg.2016.03.027.

75. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. 2011;12:R10. https://doi.org/10.1186/gb-2011-12-1-r10.

76. Oliva M, Demanelis K, Lu Y, Chernoff M, Jasmine F, Ahsan H, et al. DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. Nat Genet. 2022. https://doi.org/10.1038/s41588-022-01248-z.

77. Husquin LT, Rotival M, Fagny M, Quach H, Zidane N, McEwen LM, et al. Exploring the genetic basis of human population differences in DNA methylation and their causal impact on immune gene regulation. Genome Biol. 2018;19:222. https://doi.org/10.1186/s13059-018-1601-3.

78. Daca-Roszak P, Jaksik R, Paczkowska J, Witt M, Ziętkiewicz E. Discrimination between human populations using a small number of differentially methylated CpG sites: a preliminary study using lymphoblastoid cell lines and peripheral blood samples of European and Chinese origin. BMC Genomics. 2020;21:706. https://doi.org/10.1186/s12864-020-07092-x.

79. Galanter JM, Gignoux CR, Oh SS, Torgerson D, Pino-yanes M, Thakur N, et al. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. Elife. 2017. https://doi.org/10.7554/eLife.20532.

80. Barfield R, Wang H, Liu Y, Brody JA, Swenson B, Li R, et al. Epigenome-wide association analysis of daytime sleepiness in the Multi-Ethnic Study of Atherosclerosis reveals African-American-specific associations. Sleep. 2019;42:101. https://doi.org/10.1093/sleep/zsz101.

81. Jhun M-A, Mendelson M, Wilson R, Gondalia R, Joehanes R, Salfati E, et al. A multi-ethnic epigenome-wide association study of leukocyte DNA methylation and blood lipids. Nat Commun. 2021;12:3987. https://doi.org/10.1038/s41467-021-23899-y.

82. Tekola-Ayele F, Zeng X, Ouidir M, Workalemahu T, Zhang C, Delahaye F, et al. DNA methylation loci in placenta associated with birthweight and expression of genes relevant for early development and adult diseases. Clin Epigenetics. 2020;12:78. https://doi.org/10.1186/s13148-020-00873-x.

83. Zaghlool SB, Al-Shafai M, Al Muftah WA, Kumar P, Gieger C, Waldenberger M, et al. Mendelian inheritance of trimodal CpG methylation sites suggests distal cis-acting genetic effects. Clin Epigenetics. 2016;8:124. https://doi.org/10.1186/s13148-016-0295-1.

84. Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. Genome Biol. 2016;17:10–22. https://doi.org/10.1186/s13059-016-1030-0.

85. Workalemahu T, Shrestha D, Tajuddin SM, Tekola-Ayele F. Maternal cardiometabolic factors and genetic ancestry influence epigenetic aging of the placenta. J Dev Orig Health Dis. 2021;12:34–41. https://doi.org/10.1017/S2040174419000801.

86. Li J, Xu C, Lee HJ, Ren S, Zi X, Zhang Z, et al. A genomic and epigenomic atlas of prostate cancer in Asian populations. Nature. 2020;580:93–9. https://doi.org/10.1038/s41586-020-2135-x.

87. Adkins RM, Krushkal J, Tylavsky FA, Thomas F. Racial differences in gene-specific DNA methylation levels are present at birth. Birt Defects Res A Clin Mol Teratol. 2011;91:728–36. https://doi.org/10.1002/bdra.20770.

88. Pheiffer C, Willmer T, Dias S, Abrahams Y, Louw J, Goedecke JH. Ethnic and adipose depot specific associations between DNA methylation and metabolic risk. Front Genet. 2020. https://doi.org/10.3389/fgene.2020.00967.

89. Do WL, Whitsel EA, Costeira R, Masachs OM, Le Roy CI, Bell JT, et al. Epigenome-wide association study of diet quality in the Women's Health Initiative and TwinsUK cohort. Int J Epidemiol. 2021;50:675–84. https://doi.org/10.1093/ije/dyaa215.

90. Needham BL, Smith JA, Zhao W, Wang X, Mukherjee B, Kardia SLR, et al. Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: The multi-ethnic study of atherosclerosis. Epigenetics. 2015;10:958–69. https://doi.org/10.1080/15592294.2015.1085139.

91. Park SL, Patel YM, Loo LWM, Mullen DJ, Offringa IA, Maunakea A, et al. Association of internal smoking dose with blood DNA methylation in three racial/ethnic populations. Clin Epigenetics. 2018;10:110. https://doi.org/10.1186/s13148-018-0543-7.

92. Rizzo HE, Escaname EN, Alana NB, Lavender E, Gelfond J, Fernandez R, et al. Maternal diabetes and obesity influence the fetal epigenome in a largely Hispanic population. Clin Epigenetics. 2020;12:34. https://doi.org/10.1186/s13148-020-0824-9.

93. Tehranifar P, Wu H-C, McDonald JA, Jasmine F, Santella RM, Gurvich I, et al. Maternal cigarette smoking during pregnancy and offspring DNA methylation in midlife. Epigenetics. 2018;13:129–34. https://doi.org/10.1080/15592294.2017.1325065.

94. Chitrala KN, Hernandez DG, Nalls MA, Mode NA, Zonderman AB, Ezike N, et al. Race-specific alterations in DNA methylation among middle-aged African Americans and Whites with metabolic syndrome. Epigenetics. 2020;15:462–82. https://doi.org/10.1080/15592294.2019.1695340.

95. Giuliani C, Sazzini M, Bacalini MG, Pirazzini C, Marasco E, Fontanesi E, et al. Epigenetic variability across human populations: a focus on DNA methylation profiles of the KRTCAP3, MAD1L1 and BRSK2 genes. Genome Biol Evol. 2016;8:2760–73. https://doi.org/10.1093/gbe/evw186.

96. Boyce WT, Kobor MS. Development and the epigenome: the 'synapse' of gene–environment interplay. Dev Sci. 2015;18:1–23. https://doi.org/10.1111/desc.12282.

97. Mancilla VJ, Peeri NC, Silzer T, Basha R, Felini M, Jones HP, et al. Understanding the interplay between health disparities and epigenomics. Front Genet. 2020. https://doi.org/10.3389/fgene.2020.00903.

98. England-Mason G, Merrill SM, Gladish N, Moore SR, Giesbrecht GF, Letourneau N, et al. Prenatal exposure to phthalates and peripheral blood and buccal epithelial DNA methylation in infants: an epigenome-wide association study. Environ Int. 2022. https://doi.org/10.1016/j.envint.2022.107183.

99. Xu R, Li S, Li S, Wong EM, Southey MC, Hopper JL, et al. Residential surrounding greenness and DNA methylation: an epigenome-wide association study. Environ Int. 2021;154:106556. https://doi.org/10.1016/j.envint.2021.106556.

100. Carter SE, Gibbons FX, Beach SRH. Measuring the biological embedding of racial trauma among Black Americans utilizing the RDoC approach. Dev Psychopathol. 2021;33:1849–63. https://doi.org/10.1017/s0954579421001073.

101. Krieger N. Stormy weather: *race,* gene expression, and the science of health disparities. Am J Public Health. 2005;95:2155–60. https://doi.org/10.2105/AJPH.2005.067108.

102. Simons RL, Lei MK, Klopack E, Beach SRH, Gibbons FX, Philibert RA. The effects of social adversity, discrimination, and health risk behaviors on the accelerated aging of African Americans: further support for the weathering hypothesis. Soc Sci Med. 2021;282:113169. https://doi.org/10.1016/j.socscimed.2020.113169.

103. Clausing ES, Tomlinson CJ, Non AL. Epigenetics and social inequalities in asthma and allergy. J Allergy Clin Immunol. 2023. https://doi.org/10.1016/j.jaci.2023.01.032.

104. Sullivan S. The epigenome: on the transgenerational effects of racism. In: Sullivan S, editor. Physiol. Sexist Racist Oppression. Oxford University Press; 2015. https://doi.org/10.1093/acprof:oso/9780190250607.003.0004.

105. Mulligan CJ. Systemic racism can get under our skin and into our genes. Am J Phys Anthropol. 2021;175:399–405. https://doi.org/10.1002/ajpa.24290.

106. Geronimus AT, Hicken M, Keene D, Bound J. "Weathering" and age patterns of allostatic load scores among Blacks and Whites in the United States. Am J Public Health. 2006;96:826–33. https://doi.org/10.2105/AJPH.2004.060749.

107. de Mendoza VB, Huang Y, Crusto CA, Sun YV, Taylor JY. Perceived racial discrimination and DNA methylation among African American women in the InterGEN Study. Biol Res Nurs. 2018;20:145–52. https://doi.org/10.1177/1099800417748759.

108. Chen E, Miller GE, Yu T, Brody GH. The Great Recession and health risks in African American youth. Brain Behav Immun. 2016;53:234–41. https://doi.org/10.1016/j.bbi.2015.12.015.

109. Miller GE, Yu T, Chen E, Brody GH. Self-control forecasts better psychosocial outcomes but faster epigenetic aging in low-SES youth. Proc Natl Acad Sci U S A. 2015;112:10325–30. https://doi.org/10.1073/pnas.1505063112.

110. Clausing ES, Binder AM, Non AL. Epigenetic age associates with psychosocial stress and resilience in children of Latinx immigrants. Epigenomics. 2021;13:1677–99. https://doi.org/10.2217/epi-2019-0343.

111. Clausing ES, Non AL. Epigenetics as a mechanism of developmental embodiment of stress, resilience, and cardiometabolic risk across generations of latinx immigrant families. Front Psychiatry. 2021;12:696827. https://doi.org/10.3389/fpsyt.2021.696827.

112. Chilunga FP, Henneman P, Elliott HR, Cronjé HT, Walia GK, Meeks KAC, et al. Epigenetic-age acceleration in the emerging burden of cardiometabolic diseases among migrant and non-migrant African populations: a population-based cross-sectional RODAM substudy. Lancet Healthy Longev. 2021;2:e327–39. https://doi.org/10.1016/S2666-7568(21)00087-8.

113. Markides KS, Rote S. The healthy immigrant effect and aging in the United States and other Western countries. Gerontologist. 2019;59:205–14. https://doi.org/10.1093/geront/gny136.

114. Sanchez D, Smith LV, Adams W. The relationships among perceived discrimination, marianismo gender role attitudes, racial-ethnic socialization, coping styles, and mental health outcomes in Latina college students. J Lat Psychol. 2018;6:1–15. https://doi.org/10.1037/lat0000077.

115. Harding JF, Hughes DL, Way N. Racial/ethnic differences in mothers' socialization goals for their adolescents. Cultur Divers Ethnic Minor Psychol. 2017;23:281–90. https://doi.org/10.1037/cdp0000116.

116. Unternaehrer E, Meinlschmidt G. Psychosocial stress and DNA methylation. In: Spengler D, Binder E, editors. Epigenetics Neuroendocrinol. Clin. Focus Psychiatry. Cham: Springer International Publishing; 2016. pp. 227–61. https://doi.org/10.1007/978-3-319-29901-3_11.

117. Braun K, Champagne FA. Paternal influences on offspring development: behavioural and epigenetic pathways. J Neuroendocrinol. 2014;26:697–706. https://doi.org/10.1111/jne.12174.

118. Gettler LT, Lin DTS, Miegakanda V, Lew-Levy S, Eick GN, Snodgrass JJ, et al. Epigenetic aging in children from a small-scale farming society in The Congo Basin: Associations with child growth and family conflict. Dev Psychobiol. 2020;62:138–53. https://doi.org/10.1002/dev.21935.

119. Cimino S. DNA Methylation at the DAT Promoter and Risk for Psychopathology : Intergenerational Transmission between School-Age Youths and Their Parents in a Community Sample. Front Psychiatry. 2018;8:1–12. https://doi.org/10.3389/fpsyt.2017.00303.

120. Merrill SM, Gladish N, Fu MP, Moore SR, Konwar C, Giesbrecht GF, et al. Associations of peripheral blood DNA methylation and estimated monocyte proportion differences during infancy with toddler attachment style. Attach Hum Dev. 2021;10:1–30. https://doi.org/10.1080/14616734.2021.1938872.

121. Provenzi L, Brambilla M, Scotto G, Montirosso R, Borgatti R. Maternal caregiving and DNA methylation in human infants and children: systematic review. Genes Brain Behav. 2020. https://doi.org/10.1111/gbb.12616.

122. Hartwig FP, LoretdeMola C, Davies NM, Victora CG, Relton CL. Breastfeeding effects on DNA methylation in the offspring: a systematic literature review. PLoS ONE. 2017;12:e0173070. https://doi.org/10.1371/journal.pone.0173070.

123. Oyelana O, Kamanzi J, Richter S. A critical look at exclusive breastfeeding in Africa: Through the lens of diffusion of innovation theory. Int J Afr Nurs Sci. 2021;14:100267. https://doi.org/10.1016/j.ijans.2020.100267.

124. Kelly YJ, Watt RG, Nazroo JY. Racial/ethnic differences in breastfeeding initiation and continuation in the United Kingdom and comparison with findings in the United States. Pediatrics. 2006;118:e1428–35. https://doi.org/10.1542/peds.2006-0714.

125. Pauwels S, Symons L, Vanautgaerden E-L, Ghosh M, Duca RC, Bekaert B, et al. The influence of the duration of breastfeeding on the infant's metabolic epigenome. Nutrients. 2019;11:1408. https://doi.org/10.3390/nu11061408.

126. Arpón A, Milagro F, Razquin C, Corella D, Estruch R, Fitó M, et al. Impact of consuming extra-virgin olive oil or nuts within a Mediterranean diet on DNA methylation in peripheral white blood cells within the PREDIMED-Navarra randomized controlled trial: a role for dietary lipids. Nutrients. 2017;10:15. https://doi.org/10.3390/nu10010015.

127. Quach A, Levine ME, Tanaka T, Lu AT, Chen BH, Ferrucci L, et al. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. Aging. 2017;9:419–37. https://doi.org/10.18632/aging.101168.

128. Sharma H, Keith WR. Ayurveda and epigenetics. Medicina (Mex). 2020;56:687. https://doi.org/10.3390/medicina56120687.

129. Zhu D, Li A, Lv Y, Fan Q. Traditional Chinese Medicine: a class of potentially reliable epigenetic drugs. Front Pharmacol. 2022. https://doi.org/10.3389/fphar.2022.907031.

130. Czamara D, Eraslan G, Page CM, Lahti J, Lahti-Pulkkinen M, Hämäläinen E, et al. Integrated analysis of environmental and genetic influences on cord blood DNA methylation in new-borns. Nat Commun. 2019;10:2548. https://doi.org/10.1038/s41467-019-10461-0.

131. Lett E, Asabor E, Beltrán S, Cannon AM, Arah OA. Conceptualizing, Contextualizing, and Operationalizing Race in Quantitative Health Sciences Research. Ann Fam Med. 2022;20:157–63. https://doi.org/10.1370/afm.2792.

132. Why Nature is updating its advice to authors on reporting race or ethnicity. Nature. 2023;616:219. https://doi.org/10.1038/d41586-023-00973-7.

133. Non AL. Social epigenomics: are we at an impasse? Epigenomics. 2021;13:1747–59. https://doi.org/10.2217/epi-2020-0136.

134. Yuan V, Price EM, Del Gobbo G, Mostafavi S, Cox B, Binder AM, et al. Accurate ethnicity prediction from placental DNA methylation data. Epigenetics Chromatin. 2019;12:51. https://doi.org/10.1186/s13072-019-0296-3.

135. Rahmani E, Shenhav L, Schweiger R, Yousefi P, Huen K, Eskenazi B, et al. Genome-wide methylation data mirror ancestry information. Epigenetics Chromatin. 2017;10:1–12. https://doi.org/10.1186/s13072-016-0108-y.

136. Xu Z, Niu L, Taylor JA. The ENmix DNA methylation analysis pipeline for Illumina BeadChip and comparisons with seven other preprocessing pipelines. Clin Epigenetics. 2021;13:216. https://doi.org/10.1186/s13148-021-01207-1.

137. Shavitt S, Cho YI, Johnson TP, Jiang D, Holbrook A, Stavrakantonaki M. Culture moderates the relation between perceived stress, social support, and mental and physical health. J Cross-Cult Psychol. 2016;47:956–80. https://doi.org/10.1177/0022022116656132.

138. Kraemer HC. A source of false findings in published research studies: adjusting for covariates. JAMA Psychiat. 2015;72:961. https://doi.org/10.1001/jamapsychiatry.2015.1178.

139. Sul JH, Martin LS, Eskin E. Population structure in genetic studies: Confounding factors and mixed models. PLOS Genet. 2018;14:e1007309. https://doi.org/10.1371/journal.pgen.1007309.

140. Memon MA, Cheah JH, Ramayah T, Ting H, Chuah F, Cham TH. Moderation analysis: issues and guidelines. J Appl Struct Equ Model. 2019;3:1–9. https://doi.org/10.47263/JASEM.3(1)01.

141. Hughes C, Devine RT, Wang Z. Does parental mind-mindedness account for cross-cultural differences in preschoolers' theory of mind? Child Dev. 2018;89:1296–310. https://doi.org/10.1111/cdev.12746.

142. Martin TC, Yet I, Tsai P-C, Bell JT. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. BMC Bioinform. 2015;16:131. https://doi.org/10.1186/s12859-015-0568-2.

143. Gatev E, Gladish N, Mostafavi S, Kobor MS. CoMeBack: DNA methylation array data analysis for co-methylated regions. Bioinformatics. 2020;36:2675–83. https://doi.org/10.1093/bioinformatics/btaa049.

144. Philibert R, Beach SRH, Lei M-K, Gibbons FX, Gerrard M, Simons RL, et al. Array-based epigenetic aging indices may be racially biased. Genes. 2020;11:685. https://doi.org/10.3390/genes11060685.

145. Fagny M, Patin E, Macisaac JL, Rotival M, Flutre T, Jones MJ, et al. The epigenomic landscape of African rainforest hunter-gatherers and farmers. Nat Commun. 2015. https://doi.org/10.1038/ncomms10047.

146. Lapato DM, Moyer S, Olivares E, Amstadter AB, Kinser PA, Latendresse SJ, et al. Prospective longitudinal study of the pregnancy DNA methylome: the US Pregnancy, Race, Environment, Genes (PREG) study. BMJ Open. 2018;8:e019721. https://doi.org/10.1136/bmjopen-2017-019721.

147. Larson K, Russ SA, Kahn RS, Flores G, Goodman E, Cheng TL, et al. Health disparities: a life course health development perspective and future research directions. Springer. 2017. https://doi.org/10.1007/978-3-319-47143-3_20.

148. Weaver ICG, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, et al. Epigenetic programming by maternal behavior. Nat Neurosci. 2004;7:847–54. https://doi.org/10.1038/nn1276.

149. Breeze CE, Beck S, Berndt SI, Franceschini N. The missing diversity in human epigenomic studies. Nat Genet. 2022;54:737–9. https://doi.org/10.1038/s41588-022-01081-4.

150. Hüls A, Czamara D. Methodological challenges in constructing DNA methylation risk scores. Epigenetics. 2020;15:1–11. https://doi.org/10.1080/15592294.2019.1644879.

151. Stevenson AJ, Gadd DA, Hillary RF, McCartney DL, Campbell A, Walker RM, et al. Creating and validating a DNA methylation-based proxy for interleukin-6. J Gerontol Ser A. 2021;76:2284–92. https://doi.org/10.1093/gerona/glab046.

152. Raffington L, Belsky DW. Integrating DNA methylation measures of biological aging into social determinants of health research. Curr Environ Health Rep. 2022;9:196–210. https://doi.org/10.1007/s40572-022-00338-8.

153. Crimmins EM, Thyagarajan B, Levine ME, Weir DR, Faul J. Associations of age, sex, race/ethnicity, and education with 13 epigenetic clocks in a nationally representative US sample: The Health and Retirement Study. J Gerontol A Biol Sci Med Sci. 2021;76:1117–23. https://doi.org/10.1093/gerona/glab016.

154. Gibson J, Russ TC, Clarke T-K, Howard DM, Hillary RF, Evans KL, et al. A meta-analysis of genome-wide association studies of epigenetic age acceleration. PLoS Genet. 2019;15:e1008104. https://doi.org/10.1371/journal.pgen.1008104.

155. Reich D, Patterson N, Ramesh V, De Jager PL, McDonald GJ, Tandon A, et al. Admixture mapping of an allele affecting interleukin 6 soluble receptor and interleukin 6 levels. Am J Hum Genet. 2007;80:716–26. https://doi.org/10.1086/513206.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.