SPECIAL FOCUS ISSUE | Epigenomics and Health Disparities

Perspective

For reprint orders, please contact: reprints@futuremedicine.com

Epigenomics

A transdisciplinary approach to understand the epigenetic basis of race/ethnicity health disparities

Lucas A Salas*[1], Lauren C Peres[2], Zaneta M Thayer[3], Rick WA Smith[3,4], Yichen Guo[5], Wonil Chung[6,7,8], Jiahui Si[8,9] & Liming Liang[7,8,10]

[1]Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Lebanon, NH 03756, USA
[2]Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL 33612, USA
[3]Department of Anthropology, Dartmouth College, Hanover, NH 03755, USA
[4]The William H. Neukom Institute for Computational Science, Dartmouth College, Hanover, NH 03755, USA
[5]Merck & Co., Inc, Boston, MA 02115, USA
[6]Department of Statistics & Actuarial Science, Soongsil University, Seoul, 06478, Korea
[7]Program in Genetic Epidemiology & Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[8]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[9]Department of Biostatistics & Epidemiology, Peking University School of Public Health, Beijing, 100191, China
[10]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
*Author for correspondence: lucas.a.salas@dartmouth.edu

Health disparities correspond to differences in disease burden and mortality among socially defined population groups. Such disparities may emerge according to race/ethnicity, socioeconomic status and a variety of other social contexts, and are documented for a wide range of diseases. Here, we provide a transdisciplinary perspective on the contribution of epigenetics to the understanding of health disparities, with a special emphasis on disparities across socially defined racial/ethnic groups. Scientists in the fields of biological anthropology, bioinformatics and molecular epidemiology provide a summary of theoretical, statistical and practical considerations for conducting epigenetic health disparities research, and provide examples of successful applications from cancer research using this approach.

First draft submitted: 25 February 2020; Accepted for publication: 7 April 2020; Published online: 10 March 2021

Keywords: anthropology • cancer health disparities • computational biology • DNA methylation • epigenomics • ethnic groups • health status disparities • molecular epidemiology

The goal of this manuscript is to provide a transdisciplinary perspective on the contribution of epigenetics to the understanding of health disparities among socially defined racial/ethnic groups, drawing on expertise from scientists in the fields of biological anthropology, bioinformatics and molecular epidemiology. Herein, we summarize theoretical, statistical and practical elements of investigating racial/ethnic health disparities using epigenetics and provide examples of successful applications from cancer research using this approach. This work will aid in shaping how researchers understand and therefore approach problems of health disparities when incorporating epigenetic data.

What is epigenetics?

Epigenetics is the study of heritable phenotypic variations that do not involve changes in the DNA sequence, often involving control for gene activity and expression [1,2]. Examples that produce such changes are histone and chromatin modification, and DNA methylation. DNA methylation is the best-studied type of epigenetic modification in humans, playing a critical role in the regulation of gene expression. This primarily occurs by reducing the transcription of genes at the promoter and enhancer regions, although more complex regulatory mechanisms have been described in other gene contexts [109 3,4]. DNA methylation is one of the key players in cellular differentiation, providing cell identity [5]. Genetic variation and environmental factors can affect the cell subpopulation selection through epigenetic adaptation, and under extreme pressures, can alter cell differentiation

Future Medicine

and lead to abnormal phenotypes, including cancer cells [6–8]. This plasticity makes epigenetic processes lie at the interface of the environment and transcriptional control.

Using epigenetics to understand the development of health disparities

Health disparities are differences in disease risk between populations groups, and oftentimes, are accompanied by a higher than expected mortality burden. Health disparities are common across a variety of social contexts and are documented for a wide range of morbidities that occur across the life course. Such disparities may emerge according to race. Race is defined here as the social construct of human variability based on perceptions of biological differences (e.g., skin color or other aspects of physical appearance). While race categories do not reflect genetically distinct groups, experiences of racism and structural violence can adversely impact the biology and health of racialized minorities. Similarly, health disparities may also emerge according to ethnicity. Ethnicity, according to Mersha *et al.*, refers to a 'multidimensional construct reflecting biological factors, geographical origins, historical influences, as well as shared customs, beliefs, and traditions among populations that may or may not have a common genetic origin' [9]. As such, each 'race' or 'ethnicity' may include multiple subgroups (e.g., Hispanic/Latinx are comprised of Cubans, Panamanians, Ecuadorians, Argentinians, etc.).

Given that racial/ethnic groups are social categories that do not necessarily align with underlying patterns of genetic variation [10], genetic factors alone are insufficient to explain how racial/ethnic health disparities emerge. It is therefore critical to evaluate how differences in experience and environment shape health outcomes. In fact, it is now becoming clear that environmental differences are important in shaping more complex phenotypes that are of interest in public health regarding racial/ethnic inequalities, such as low birth weight, preterm birth, asthma, cancer and cardiovascular disease. Epigenetic studies may help to understand how differences in environmental experience translate into differences in phenotype. Although a limited number of studies have directly connected DNA methylation to health disparities, a few studies have reported intriguing results. Socioeconomic status [11], as well as factors that vary according to socioeconomic statuses such as psychosocial stress exposure [12,13] and diet [14,15], have been associated with variation in DNA methylation.

Early life experiences may be particularly important for shaping health disparities. The Developmental Origins of Health and Disease hypothesis suggests that a mother's experience during her life and her pregnancy may shape the epigenome and future health trajectory of her infant [16,17]. Moreover, different environmental exposures during pregnancy (maternal lifestyles, diseases and exposures to environmental toxicants) have been associated with alterations in DNA methylation in the placenta or the umbilical cord blood of the newborn [18–21]. However, there has been criticism of the overemphasis on maternal effects relative to paternal effects in predicting health outcomes via epigenetic mechanisms [22]. While it has been less frequently investigated, growing evidence suggests that paternal environmental experience can also affect offspring through epigenetic processes. For example, paternal obesity in the peripartum period is associated with significant differences in offspring methylation at imprinted genes important for regulating growth and development [23]. These findings suggest that socially patterned exposure to stressors in both parents could potentially affect offspring health via changes in the offspring epigenome.

Among historically marginalized communities, the ancestral experience of trauma (i.e., historical trauma) shapes disparities in later health generations [24–26]. In addition to being related to the intergenerational effects described above, the health impacts of historical trauma likely also reflects within-generation epigenetic impacts of environmental conditions shaped by ancestral experience [27,28]. For example, forced displacement of ancestors increases the likelihood that members of the contemporary generation experience poverty and therefore associated health sequelae. Likewise, the parental experience of trauma could shape both the intrauterine environment and patterns of parental care, both of which affect the developing epigenome of offspring [27,29]. Therefore, historical trauma should be considered as an additional conceptual model for explaining observed health disparities.

Statistical methodology for studying the epigenetic basis of health disparities

Studies investigating the DNA methylation basis of health disparities have generally employed global DNA methylation, targeted gene methylation, single variant methylation or network-based analysis [30–32]. With the rapid development of high-throughput technologies in recent years, population-based epigenome-wide association studies at a single-nucleotide resolution became a popular approach utilized in epigenetic studies. The common supervised selection strategy is to select CpG sites affecting phenotypic differences, noted as differentially methylated cytosines [33]. Differentially methylated cytosines can be selected based on the absolute difference in mean beta values or test statistics from a t-test, Wilcoxon test or multivariable regression model. Another selection approach is

to consider differential variance in methylation between two traits, noted as differentially variable cytosines, using, for instance, the Bartlett test or the Levene's test [34,35]. In genomics, variance-based selection of SNPs approaches could be used to prioritize those SNPs for subsequent gene–gene and gene–environment testing [36]. Although some tools have been developed for this goal [37], epigenome-wide approaches for epigenome–environment and genome–epigenome interactions are infrequently studied beyond locus-specific interactions due to computational burden and concerns over model assumptions when using untargeted approaches. On the other hand, the unsupervised selection procedure is to rank and filter CpG sites by variance, aiming for the selection of the most variably methylated cytosines [38]. The most variably methylated cytosines generally represent various levels of DNA methylation and may contain those driven by SNPs or cell heterogeneity (e.g., different immune cell populations in blood/saliva or immune-cell infiltration in solid tissues). Studies have also employed the most variably methylated cytosines approach to filter CpGs before single variant methylation tests to reduce the burden of multiple hypothesis testing.

Differences at any individual site may be small; however, if these differences are persistent across a region or a certain group of genes, statistical power to detect them may be greater. Several methods have been developed to identify sets of neighboring CpGs sites that are correlated with each other, known as differentially methylated regions, and link them with traits of interest, including DMRcate [39], bump hunting [40] and the A-clustering method [41]. Other methods aiming to build gene co-methylation networks have also been proposed. For example, weighted gene co-methylation network analysis aims to describe the correlation patterns among genes across microarray samples, find clusters of highly correlated genes, and relate such clusters to a phenotype of interest via enrichment analysis or network eigengenes (the top principal component of genes in the cluster) [42].

Challenges & opportunities in statistical methodology

Although advances in epigenetic studies are expected to help understand racial/ethnic health disparities, there are notable challenges and limitations to consider. Epigenetic studies are potentially impacted by a range of confounding factors, including but not limited to population genetic patterns, cell-type, environmental confounders related to ethnicity and sample processing batch [43]. Population stratification is another critical source of confounding for studies including heterogeneous populations. DNA methylation signatures of target tissue (e.g., saliva, whole blood, placenta, adipose and tumors) are an average of cell type-specific methylation levels. Hence, the cell-type proportion is generally related to the measured DNA methylation levels, and in many cases, is also associated with race/ethnicity [44] and traits of interest [45]. Various statistical methods have been proposed to adjust for this potential bias. Statistical models for cell-type deconvolution are classified into three categories called reference-based [46], reference-free [47] and semi-reference-free [48–51], the last of which alleviates some of the problems of both reference-based and reference-free methods. The choice of the appropriate method for cell-type deconvolution mainly depends on the availability of a proper reference database for DNA methylation of the cell types involved [52–54]. Some other methods developed specifically for methylation data, or for general purposes, can be used to control for all unmeasured confounding, including; surrogate variable analysis (SVA) [55], independent SVA [56], smartSVA [57], remove unwanted variation [58,59] and principal component analysis [60]. Several of these methods have been used to adapt reference-free approaches and semi-reference-free approaches and have been reviewed and compared elsewhere [2,61]. However, residual confounding may still be possible after adjustment. In statistical genetics, genomic inflation represents the excess of false positives in genomic analyses. In epigenome-wide association studies, the genomic inflation factor calculation and the quantile-quantile plots have been used to quantify the excess inflation in statistics, however, in most of these analyses, genomic inflation is not corrected. The application of genomic control correction has shown to be ineffective due to the small differences detected in epigenome-wide association studies, with only a few methods that have been adapted specifically for DNA methylation analyses but are still not widely used in the field [62]. Finally, several biomarkers derived from DNA methylation information have been developed which could offer global measures of epigenetic drift related to various phenotypic variations of interest. Among those, we have age acceleration using the DNA methylation age measures [63,64], fetal cell of origin [65], inference of multiple retrotransposons using epigenome-wide information [66] or global methylation changes [67]. The selection of specific methods should be adapted to the specific hypothesis being tested.

Insights & applications from cancer research

Locus specific changes and differentially methylated cytosines related to race/ethnicity have been identified among cancer biologists interested in health disparities and have been extensively reviewed in the past [30,68–70]. Beyond the locus-specific promoter changes, epigenome-wide association studies have continuously reported variation in

DNA methylation patterns between different populations such as Europeans, Hispanics, Africans and Asians [71]. Determining how these race/ethnicity variations are associated with disease outcomes will further help to understand health disparities. Another approach is to describe demographic and environmental factor-associated and disease-associated differentially methylated cytosines in different race/ethnicity groups.

Cancer health disparities research in the USA is largely focused on race/ethnicity, where cancer incidence is highest among African–Americans or Blacks, followed by non-Hispanic Whites, Hispanics and Asian/Pacific Islanders [72]. Generally, African–Americans also have the highest mortality rates and worse survival outcomes in comparison with all other race/ethnicity groups. The disparity gap between whites and African–Americans for cancer incidence and mortality has narrowed over time, but there is still a notable 14% difference in the mortality rate [72]. While African–Americans are disproportionately affected overall, other race/ethnicity groups (e.g., Hispanics, Asian/Pacific Islanders) have a greater cancer burden or worse survival for certain cancers. For example, Hispanics have a higher incidence of infection-associated cancers (e.g., liver, stomach and cervical cancer) [73]. Cancer is also more prevalent among socially, economically or environmentally disadvantaged populations. Higher cancer incidence and mortality rates, as well as lower survival, are experienced by cancer patients with low educational attainment or residents of impoverished neighborhoods compared with more educated individuals and residents of affluent areas [74]. As the socioeconomic status and race/ethnicity are inextricably linked to one another, it is often difficult to disentangle their independent effects on cancer disparities.

The underlying causes of cancer health disparities are complex and multifactorial. While a portion of the disease burden is due to the marginalization of minority populations, disease susceptibility is a combination of population isolation, genetic burden and selection of specific phenotypes that are advantageous for certain environments [75]. One example is the trends in cancer subtype susceptibility for certain race/ethnicity groups. Skin cancer distributions differ across race/ethnicity as the risk of squamous cell carcinoma is higher in Eurasian descendants and anecdotally in African populations with albinism [76]. Triple-negative breast cancer and aggressive prostate cancer are much more frequent in African–Americans compared with other racial/ethnic groups in the USA. Even after accounting for healthcare access and other social factors, African–Americans with these subtypes have a worse prognosis compared with white–Americans. Interestingly, Hispanic cancer patients have better outcomes than African–Americans despite similar sociodemographic characteristics, also known as the 'Hispanic paradox' [77], while at the same time, this group is still adversely affected by other health outcomes, such as infectious diseases, disabilities and diabetes compared with non-Hispanic whites. As neighborhood socioeconomic status has been shown to contribute to survival disparities in Black and Hispanic cancer patients, but not Asian/Pacific Islanders [78], neighborhood socioeconomic status does not represent the only source of variability contributing to health disparities for these groups.

Considerable strides have been made in cancer research to investigate the link between DNA methylation and cancer health disparities, primarily for the most common cancers. In breast cancer, several studies have identified differentially methylated loci when comparing tumors from African–American and European–American women [79–84], with the most differences observed in women with estrogen receptor (ER) negative tumors and younger women. Another study uncovered seven genes hypermethylated in Korean versus European women, which again, was particularly seen among ER and progesterone receptor (PR) negative tumors and women aged $\leq$50 years [85]. Similarly, work in prostate cancer found several CpG sites that are differentially methylated among tumors from Black versus white men [86–90], with studies consistently implicating *CD44* and *GSTP1* [86,87,90]. Besides breast and prostate cancer, the literature investigating the epigenetic basis of race/ethnicity disparities in other cancer types is fairly sparse [91–97], especially for rare cancers where challenges arise due to the limited number of cases in existing studies, particularly within minority or underserved populations. Many additional studies in the literature investigate DNA methylation of cancer patients within a specific race/ethnicity but do not compare with other race/ethnicity groups. While these studies uncover the unique epigenetic alterations within different populations, they do not provide a comparison group to elucidate a potential racial/ethnic disparity and are not discussed herein.

Challenges in epigenetic health disparities research

Race/ethnicity is the most common disparity investigated in cancer epigenetics as well as other disease disparities, with the majority of studies comparing African–American/Black and European–American/White populations. Little emphasis has been placed on other race/ethnicity groups (e.g., Hispanics, Asian/Pacific Islanders, American–Indians/Alaskan Natives), although these under-represented groups have notable disparities for many chronic and

acute diseases and are a growing proportion of the U.S. population. Moreover, any racial/ethnic categorization encompasses very heterogeneous populations. For instance, Hispanics and Asian/Pacific Islanders represent a variety of ethnic subgroups (e.g., Cuban, Mexican, Filipino and Chinese) that have different experiences and risk profiles. These racial/ethnic groups are often understudied due to inadequate sample sizes or the under-representation of these populations within any one study. Adding to this challenge, there is typically a lack of studies with available biospecimens to conduct epigenetic disparities research.

Promoter methylation of candidate genes or a preselected panel of genes has been the approach used most often to measure DNA methylation. The array-based methodology offers an agnostic approach to move beyond a single gene or gene promoter. However, the majority of studies using arrays to quantify DNA methylation levels have used the Illumina 27 or 450K array. Both are now obsolete after the introduction of the MethylationEPIC (or 850K) array, which provides comprehensive genome-wide coverage and captures additional enhancer and intergenic regions of the genome that were not included on the older versions of the array. As genome-wide association studies note the importance of noncoding regions of the genome in disease susceptibility, the EPIC array will be able to shed light on whether this is also true for epigenetic alterations. DNA methylation sequencing (e.g., reduced representation bisulphite sequencing or whole genome bisulphite sequencing) has been used in a few health disparities studies [98], however, this technology cost is higher than the microarrays and depending on the biospecimen, the genome coverage may not be consistent.

Conclusion

Epigenetic markers have shown several interesting associations that could be driving health disparities from a biological perspective. When investigating the association between racially/ethnically different epigenetic variations and disease outcomes in cross-sectional settings, determining causality is challenging [99]. Prospective follow-up studies among racially/ethnically heterogeneous populations would allow researchers to identify methylation changes involved in different pathways preceding disease onset. Using a Mendelian randomization approach to integrate genotype and epigenetic data may also prove useful in determining causality [100]. However, there are several limitations of this approach including but not limited to low statistical power, population stratification generating spurious genetic variants, re-introduced confounding through pleiotropy and linkage disequilibrium with multiple causal genetic variants of the epigenetic variation [100].

The use of machine learning is of great interest in disease prediction and classification. For example, the elastic net, a penalized regression model, has been applied in predicting human age with DNA methylation data in the USA [63], Chinese and multiracial/multiethnic populations [101,102]. Integrating the epigenome with other types of -omics data such as the genome, transcriptome, proteome, metabolome and the microbiome has the potential to unlock the 'black box' in health disparities. Although current technologies are still facing challenges, researchers have found intriguing results [103,104]. Future enhanced bioinformatics and analytical tools [105] will enable a more comprehensive analysis of human observational and interventional studies in a systematic way.

Future perspective

Some studies have devised an integrative approach including a comprehensive analysis of the social and environmental exposures of specific race/ethnicity associated epigenetic changes [32,106]. Newer longitudinal cohorts are trying to recruit more diverse populations representing minorities that were not included in traditional cohort studies [107]. Transdisciplinary approaches to understand the roots of health disparities are required to improve the outcomes of minorities and marginalized populations. From the genetic point of view, researchers are moving beyond self-reported race/ethnicity to the use of ancestry informative genetic markers. In highly admixed populations, such as Latin–Americans or African–American populations, ancestry informative genetic markers reveal a different layer of information about population migration and in some cases, clusters of disease susceptibility that may not be associated when using only self-reported race/ethnicity [9]. Ancestry information will provide broad geographically relevant population information (population migration and inbreeding); however, so far, the utility of genetic markers has been limited for interventions [108]. Epigenetics, on the other hand, provide a unique opportunity to fully integrate the genetic, social and environmental contributors to health disparities, while offering a potential for intervention.

Executive summary

Using epigenetics to understand the development of health disparities

- Health disparities reflect differences in morbidity and mortality among socially defined categories, including racial/ethnic groups.
- Experiences of racism and structural violence can adversely impact the biology and health of racialized minorities.
- Early life trauma and historical trauma are disproportionately experienced by socially disadvantaged groups and could contribute to health disparities via epigenetic changes.

Statistical methodology for studying the epigenetic basis of health disparities

- Studies investigating the DNA methylation basis of health disparities have generally employed methods at different resolutions. Population-based epigenome-wide association studies analyzing single-nucleotides became a popular approach.
- Other techniques employing differential variability or differentially methylated regions are being more widely used.
- Epigenetic studies are potentially impacted by a range of confounding factors, including but not limited to population genetic patterns, cell-type, environmental confounders related to ethnicity, and sample processing batch.

Insights & applications from cancer research

- Considerable strides have been made in cancer research to investigate the link between DNA methylation and cancer health disparities, primarily for the most common cancers (breast and prostate cancer) in US African–American populations.
- There is still limited information available for other race/ethnic groups in the US, with very heterogeneous populations.
- Determining causality in cross-sectional settings is challenging. Prospective follow-up studies among racially/ethnically heterogeneous populations and techniques as mendelian randomization will identify methylation changes involved in different pathways preceding disease onset.

Financial & competing interest disclosure

L Salas is supported by grants of the CDMRP/Department of Defense (W81XWH-20-1-0778) and the National Institute of Health-NIGMS (P20GM1044168299). L Peres is supported by NIH-NCI R00CA218681. Z Thayer is supported by the Wenner-Gren Fellowship #9687. R Smith is supported through a postdoctoral Neukom Fellows Program. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

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

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Murrell A, Rakyan VK, Beck S. From genome to epigenome. *Hum. Mol. Genet.* 14 Spec No(SPEC. ISS. 1), R3–R10 (2005).
2. Teschendorff AE, Relton CL. Statistical and integrative system-level analysis of DNA methylation data. *Nat. Rev. Genet.* 19(3), 129–147 (2018).
3. Spainhour JC, Lim HS, Yi SV, Qiu P. Correlation patterns between DNA methylation and gene expression in the Cancer Genome Atlas. *Cancer Inform.* 18, page identifier: 1176935119828776 (2019).
4. Murray SC, Haenni S, Howe FS *et al.* Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Res.* 43(16), 7823–7837 (2015).
5. Varley KE, Gertz J, Bowling KM *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* 23(3), 555–567 (2013).
6. Pacchierotti F, Spanò M. Environmental impact on DNA methylation in the germline: state of the art and gaps of knowledge. *Biomed Res. Int.* 2015(7), 123484 (2015).
7. Cortessis VK, Thomas DC, Joan Levine A *et al.* Environmental epigenetics: prospects for studying epigenetic mediation of exposure-response relationships. *Hum. Genet.* 131(10), 1565–1589 (2012).
8. Herceg Z. Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. *Mutagenesis* 22(2), 91–103 (2007).

9. Mersha TB, Abebe T. Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities. *Hum. Genomics.* 9(1), 1 (2015).
10. Fuentes A, Ackermann RR, Athreya S *et al.* AAPA Statement on Race and Racism. *Am. J. Phys. Anthropol.* 169(3), 400–402 (2019).

•• **Most recent, comprehensive anthropology statement regarding 'race'.**

11. McDade TW, Ryan C, Jones MJ *et al.* Social and physical environments early in development predict DNA methylation of inflammatory genes in young adulthood. *Proc. Natl Acad. Sci. USA* 114(29), 7611–7616 (2017).

• **Early life environments, which differ according to race/ethnicity within the USA and many other cultural contexts, predict differences in methylation at genes important for health.**

12. Non AL, Hollister BM, Humphreys KL *et al.* DNA methylation at stress-related genes is associated with exposure to early life institutionalization. *Am. J. Phys. Anthropol.* 161(1), 84–93 (2016).
13. Mulligan CJ, D'Errico NC, Stees J, Hughes DA. Methylation changes at NR3C1 in newborns associate with maternal prenatal stress exposure and newborn birth weight. *Epigenetics* 7(8), 853–857 (2012).
14. Mosher MJ, Schanfield MS. Epigenetic mechanisms as an archive of ancestral dietary history of populations: the premise, proposal and pilot. *J. Archaeol. Sci. Rep.* 5, 689–699 (2016).
15. Mosher MJ, Melton PE, Stapleton P, Schanfield MS, Crawford MH. Patterns of DNA methylation across the leptin core promoter in four diverse Asian and North American populations. *Hum. Biol.* 88(2), 121–135 (2016).
16. Gluckman PD, Hanson MA, Cooper C, Thornburg KL. Effect of in utero and early-life conditions on adult health and disease. *N. Engl. J. Med.* 359(1), 61–73 (2008).
17. Barker DJP. Sir Richard Doll lecture. developmental origins of chronic disease. *Public Health* 126(3), 185–189 (2012).
18. Joubert BR, Felix JF, Yousefi P *et al.* DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. *Am. J. Hum. Genet.* 98(4), 680–696 (2016).
19. Sharp GC, Salas LA, Monnereau C *et al.* Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: findings from the pregnancy and childhood epigenetics (PACE) consortium. *Hum. Mol. Genet.* 26(20), 4067–4085 (2017).
20. Vilahur N, Bustamante M, Morales E *et al.* Prenatal exposure to mixtures of xenoestrogens and genome-wide DNA methylation in human placenta. *Epigenomics* 8(1), 43–54 (2016).
21. Morales E, Vilahur N, Salas LA *et al.* Genome-wide DNA methylation study in human placenta identifies novel loci associated with maternal smoking during pregnancy. *Int. J. Epidemiol.* 45(5), 1644–1655 (2016).
22. Richardson S. Maternal bodies in the postgenomic order: gender and the explanatory landscape of epigenetics. In: *Postgenomics: Perspectives on Biology after the Genome*. Richardson SS, Stevens H (Eds). Duke University Press, Durham, London, UK 210–231 (2015).
23. Soubry A, Murphy SK, Wang F *et al.* Newborns of obese parents have altered DNA methylation patterns at imprinted genes. *Int. J. Obes. (Lond).* 39(4), 650–657 (2015).
24. Gone JP, Hartmann WE, Pomerville A, Wendt DC, Klem SH, Burrage RL. The impact of historical trauma on health outcomes for indigenous populations in the USA and Canada: a systematic review. *Am. Psychol.* 74(1), 20–35 (2019).
25. Lehrner A, Yehuda R. Cultural trauma and epigenetic inheritance. *Dev. Psychopathol.* 30(5), 1763–1777 (2018).
26. Yehuda R, Lehrner A. Intergenerational transmission of trauma effects: putative role of epigenetic mechanisms. *World Psychiatry* 17(3), 243–257 (2018).
27. Conching AKS, Thayer Z. Biological pathways for historical trauma to affect health: a conceptual model focusing on epigenetic modifications. *Soc. Sci. Med.* 230(March), 74–82 (2019).

• **Conceptual model describing how historical trauma can influence health via epigenetic mechanisms.**

28. van Otterdijk SD, Michels KB. Transgenerational epigenetic inheritance in mammals: how good is the evidence? *FASEB J.* 30(19), 1–9 (2016).
29. Aristizabal MJ, Anreiter I, Halldorsdottir T *et al.* Biological embedding of experience: a primer on epigenetics. *Proc. Natl Acad. Sci. USA* (117(38), 23261–23269 (2019).
30. Argentieri MA, Nagarajan S, Seddighzadeh B, Baccarelli AA, Shields AE. Epigenetic pathways in human disease: the impact of DNA methylation on stress-related pathogenesis and current challenges in biomarker development. *EBioMedicine* 18, 327–350 (2017).
31. Thayer ZM, Kuzawa CW. Biological memories of past environments: epigenetic pathways to health disparities. *Epigenetics* 6(7), 798–803 (2011).
32. Lerner L, Winn R, Hulbert A. Lung cancer early detection and health disparities: the intersection of epigenetics and ethnicity. *J. Thorac. Dis.* 10(4), 2498–2507 (2018).
33. Du P, Zhang X, Huang C-C *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 11, 587 (2010).
34. Teschendorff AE, Widschwendter M. Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. *Bioinformatics* 28(11), 1487–1494 (2012).

35. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* 32(2), 286–288 (2016).
36. Paré G, Cook NR, Ridker PM, Chasman DI. On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. *PLoS Genet.* 6(6), e1000981 (2010).
37. Pan H, Holbrook JD, Karnani N, Kwoh CK. Gene, Environment and Methylation (GEM): a tool suite to efficiently navigate large scale epigenome wide association studies and integrate genotype and interaction between genotype and environment. *BMC Bioinformatics* 17(1), 299 (2016).
38. van 't Veer LJ, Dai H, van de Vijver MJ *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871), 530–536 (2002).
39. Peters TJ, Buckley MJ, Statham AL *et al. De novo* identification of differentially methylated regions in the human genome. *Epigenetics Chromatin.* 8(6), 6 (2015).
40. Jaffe AE, Murakami P, Lee H *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int. J. Epidemiol.* 41(1), 200–209 (2012).
41. Sofer T, Schifano ED, Hoppin Ja, Hou L, Baccarelli AA. A-clustering: a novel method for the detection of co-regulated methylation regions, and regions associated with exposure. *Bioinformatics* 29(22), 2884–2891 (2013).
42. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559 (2008).
43. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat. Rev. Genet.* 14(8), 585–594 (2013).

• **The authors discuss the potential confounding and analytical considerations in population-based epigenetic studies.**

44. Freedman DS, Gates L, Flanders WD *et al.* Black/white differences in leukocyte subpopulations in men. *Int. J. Epidemiol.* 26(4), 757–764 (1997).
45. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* 11(3), 191–203 (2010).
46. Houseman EA, Accomando WP, Koestler DC *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13(1), 86 (2012).
47. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* 30(10), 1431–1439 (2014).
48. Houseman EA, Kile ML, Christiani DC, Ince TA, Kelsey KT, Marsit CJ. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *BMC Bioinformatics* 17(1), 259 (2016).
49. Onuchic V, Hartmaier RJ, Boone DN *et al.* Epigenomic deconvolution of breast tumors reveals metabolic coupling between constituent cell types. *Cell Rep.* 17(8), 2075–2086 (2016).
50. Lutsik P, Slawski M, Gasparoni G, Vedeneev N, Hein M, Walter J. MeDeCom: discovery and quantification of latent components of heterogeneous methylomes. *Genome Biol.* 18(1), 55 (2017).
51. Rahmani E, Schweiger R, Shenhav L *et al.* BayesCCE: a Bayesian framework for estimating cell-type composition from DNA methylation without the need for methylation reference. *Genome Biol.* 19(1), 141 (2018).
52. Salas LA, Koestler DC, Butler RA *et al.* An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol.* 19(1), 64 (2018).
53. Gervin K, Salas LA, Bakulski KM *et al.* Systematic evaluation and validation of reference and library selection methods for deconvolution of cord blood DNA methylation data. *Clin. Epigenetics* 11(1), 125 (2019).
54. Zheng SC, Webster AP, Dong D *et al.* A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. *Epigenomics* 10(7), 925–940 (2018).
55. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 3(9), 1724–1735 (2007).
56. Teschendorff AE, Zhuang J, Widschwendter M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics* 27(11), 1496–1505 (2011).
57. Chen J, Behnam E, Huang J *et al.* Fast and robust adjustment of cell mixtures in epigenome-wide association studies with SmartSVA. *BMC Genomics* 18(1), 1–13 (2017).

•• **A fast and reliable approach to control the unmeasured confounding in epigenetic studies.**

58. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. *Biostatistics* 13(3), 539–552 (2012).
59. Maksimovic J, Gagnon-Bartsch JA, Speed TP, Oshlack A. Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. *Nucleic Acids Res.* 43(16), e106 (2015).
60. Reese SE, Archer KJ, Therneau TM *et al.* A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. *Bioinformatics* 29(22), 2877–2883 (2013).

61. Brägelmann J, Lorenzo Bermejo J. A comparative analysis of cell-type adjustment methods for epigenome-wide association studies based on simulated and real data sets. *Brief. Bioinform.* 20(6), 2055–2065 (2019).

62. van Iterson M, van Zwet EW, BIOS Consortium, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* 18(1), 19 (2017).

63. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 14(10), R115 (2013).

64. Gibson J, Russ TC, Clarke T-K *et al.* A meta-analysis of genome-wide association studies of epigenetic age acceleration. *PLoS Genet.* 15(11), e1008104 (2019).

65. Salas LA, Wiencke JK, Koestler DC, Zhang Z, Christensen BC, Kelsey KT. Tracing human stem cell lineage during development using DNA methylation. *Genome Res.* 28(9), 1285–1295 (2018).

66. Zheng Y, Joyce BT, Liu L *et al.* Prediction of genome-wide DNA methylation in repetitive elements. *Nucleic Acids Res.* 45(15), 8697–8711 (2017).

67. Salas LA, Johnson KC, Koestler DC, O'Sullivan DE, Christensen BC. Integrative epigenetic and genetic pan-cancer somatic alteration portraits. *Epigenetics* 12(7), 561–574 (2017).

68. Vick AD, Burris HH. Epigenetics and health disparities. *Curr. Epidemiol. Rep.* 4(1), 31–37 (2017).

69. Ahmad A, Azim S, Zubair H *et al.* Epigenetic basis of cancer health disparities: looking beyond genetic differences. *Biochim. Biophys. Acta Rev. Cancer* 1868(1), 16–28 (2017).

70. Saini G, Ogden A, McCullough LE, Torres M, Rida P, Aneja R. Disadvantaged neighborhoods and racial disparity in breast cancer outcomes: the biological link. *Cancer Causes Control* 30(7), 677–686 (2019).

71. Kader F, Ghai M. DNA methylation-based variation between human populations. *Mol. Genet. Genomics* 292(1), 5–35 (2017).

72. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J. Clin.* 69(1), 7–34 (2019).

73. Miller KD, Goding Sauer A, Ortiz AP *et al.* Cancer statistics for Hispanics/Latinos, 2018. *CA Cancer J. Clin.* 68(6), 425–445 (2018).

74. Singh GK, Jemal A. Socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the United States, 1950–2014: over six decades of changing patterns and widening inequalities. *J. Environ. Public Health* 2017, 2819372 (2017).

•• **A large comprehensive analysis that not only investigates cancer disparities in terms of race/ethnicity but also socioeconomic disadvantage, showing that Blacks and residents of deprived neighborhoods have the worst survival.**

75. Özdemir BC, Dotto G-P. Racial differences in cancer susceptibility and survival: more than the color of the skin? *Trends Cancer* 3(3), 181–197 (2017).

76. Greaves M. Was skin cancer a selective force for black pigmentation in early hominin evolution? *Proc. Biol. Sci.* 281(1781), 20132955 (2014).

77. Valles SA. The challenges of choosing and explaining a phenomenon in epidemiological research on the "Hispanic Paradox". *Theor. Med. Bioeth.* 37(2), 129–148 (2016).

78. Ellis L, Canchola AJ, Spiegel D, Ladabaum U, Haile R, Gomez SL. Racial and ethnic disparities in cancer survival: the contribution of tumor, sociodemographic, institutional, and neighborhood characteristics. *J. Clin. Oncol.* 36(1), 25–33 (2018).

79. Espinal AC, Buas MF, Wang D *et al.* FOXA1 hypermethylation: link between parity and ER-negative breast cancer in African American women? *Breast Cancer Res. Treat.* 166(2), 559–568 (2017).

80. Huo D, Hu H, Rhie SK *et al.* Comparison of breast cancer molecular features and survival by African and European ancestry in the Cancer Genome Atlas. *JAMA Oncol.* 3(12), 1654–1662 (2017).

81. Conway K, Edmiston SN, Tse C-K *et al.* Racial variation in breast tumor promoter methylation in the Carolina Breast Cancer Study. *Cancer Epidemiol. Biomarkers Prev.* 24(6), 921–930 (2015).

82. Ambrosone CB, Young AC, Sucheston LE *et al.* Genome-wide methylation patterns provide insight into differences in breast tumor biology between American women of African and European ancestry. *Oncotarget* 5(1), 237–248 (2014).

• **First genome-wide methylation study addressing breast cancer disparities.**

83. Mehrotra J, Ganpat MM, Kanaan Y *et al.* Estrogen receptor/progesterone receptor-negative breast cancers of young African-American women have a higher frequency of methylation of multiple genes than those of Caucasian women. *Clin. Cancer Res.* 10(6), 2052–2057 (2004).

84. Wang S, Dorsey TH, Terunuma A, Kittles RA, Ambs S, Kwabi-Addo B. Relationship between tumor DNA methylation status and patient characteristics in African-American and European-American women with breast cancer. *PLoS ONE* 7(5), e37928 (2012).

85. Ji SL, Fackler MJ, Wei WT *et al.* Quantitative promoter hypermethylation profiles of ductal carcinoma *in situ* in north american and korean women: potential applications for diagnosis. *Cancer Biol. Ther.* 7(9), 1400–1408 (2008).

86. Enokida H, Shiina H, Urakami S *et al.* Ethnic group-related differences in CpG hypermethylation of the GSTP1 gene promoter among African-American, Caucasian and Asian patients with prostate cancer. *Int. J. Cancer* 116(2), 174–181 (2005).

87. Woodson K, Hayes R, Wideroff L, Villaruz L, Tangrea J. Hypermethylation of GSTP1, CD44, and E-cadherin genes in prostate cancer among US Blacks and Whites. *Prostate* 55(3), 199–205 (2003).

88. Devaney JM, Wang S, Furbert-Harris P *et al.* Genome-wide differentially methylated genes in prostate cancer tissues from African-American and Caucasian men. *Epigenetics* 10(4), 319–328 (2015).
89. Das PM, Ramachandran K, Vanwert J *et al.* Methylation mediated silencing of TMS1/ASC gene in prostate cancer. *Mol. Cancer* 5, 28 (2006).
90. Woodson K, Hanson J, Tangrea J. A survey of gene-specific methylation in human prostate cancer among black and white men. *Cancer Lett.* 205(2), 181–188 (2004).
91. Busch EL, Galanko JA, Sandler RS, Goel A, Keku TO. Lifestyle factors, colorectal tumor methylation, and survival among African Americans and European Americans. *Sci. Rep.* 8(1), 1–7 (2018).
92. Wang X, Ji P, Zhang Y *et al.* Aberrant DNA Methylation: implications in racial health disparity. *PLoS ONE* 11(4), e0153125 (2016).
93. Guerrero-Preston R, Lawson F, Rodriguez-Torres S *et al.* JAK3 variant, immune signatures, DNA methylation, and social determinants linked to survival racial disparities in head and neck cancer patients. *Cancer Prev. Res.* 12(4), 255–270 (2019).
94. Chen S, Zhou K, Yang L, Ding G, Li H. Racial differences in esophageal squamous cell carcinoma: incidence and molecular features. *Biomed Res. Int.* 2017, 1204082 (2017).
95. Geng X, Pu W, Tan Y *et al.* Quantitative assessment of the diagnostic role of FHIT promoter methylation in non-small cell lung cancer. *Oncotarget* 8(4), 6845–6856 (2017).
96. Piyathilake CJ, Henao O, Frost AR *et al.* Race- and age-dependent alterations in global methylation of DNA in squamous cell carcinoma of the lung (United States). *Cancer Causes Control* 14(1), 37–42 (2003).
97. Li G, Liu Y, Yin H *et al.* E-cadherin gene promoter hypermethylation may contribute to the risk of bladder cancer among Asian populations. *Gene* 534(1), 48–53 (2014).
98. Vantaku V, Amara C, Piyarathna D *et al.* DNA methylation patterns in bladder tumors of African American patients point to distinct alterations in xenobiotic metabolism. *Carcinogenesis* 40(11), 1332–1340 (2019).
99. Walton E, Relton CL, Caramaschi D. Using openly accessible resources to strengthen causal inference in epigenetic epidemiology of neurodevelopment and mental health. *Genes (Basel).* 10(3), e193 (2019).
100. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int. J. Epidemiol.* 41(1), 161–176 (2012).
101. Li J, Zhu X, Yu K *et al.* Exposure to polycyclic aromatic hydrocarbons and accelerated DNA methylation aging. *Environ. Health Perspect.* 126(6), 067005 (2018).
102. Horvath S, Gurven M, Levine ME *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol.* 17(1), 171 (2016).
103. Ramazzotti D, Lal A, Wang B, Batzoglou S, Sidow A. Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. *Nat. Commun.* 9(1), 4453 (2018).

•• **Integrates that multiomic data have the potential to unlock the 'black box' of health disparities. Studies have reported intriguing results.**

104. Hu FB. Metabolic profiling of diabetes: from black-box epidemiology to systems epidemiology. *Clin. Chem.* 57(9), 1224–1226 (2011).
105. Levy JJ, Titus AJ, Salas LA, Christensen BC. PyMethylProcess – convenient high-throughput preprocessing workflow for DNA methylation data. *Bioinformatics* 35(24), 5379–5381 (2019).
106. Watson KS, Hulbert A, Henderson V *et al.* Lung cancer screening and epigenetics in African Americans: the role of the socioecological framework. *Front. Oncol.* 9(MAR), 87 (2019).
107. All of Us Research Program Investigators, Denny JC, Rutter JL *et al.* The "All of Us" research program. *N. Engl. J. Med.* 381(7), 668–676 (2019).
108. Yuan J, Hu Z, Mahal BA *et al.* Integrated analysis of genetic ancestry and genomic alterations across cancers. *Cancer Cell* 34(4), 549–560.e9 (2018).
109. Wanger J. Computational approaches for the study of gene expression, genetic and epigenetic variation in human. [*PhD Thesis*].. McGill University, Quebec, Canada (2014).