INVITED REVIEW

Cell-type deconvolution from DNA methylation: a review of recent applications

Alexander J. Titus[1,2], Rachel M. Gallimore[2], Lucas A. Salas[2] and Brock C. Christensen[2,3,4,*]

[1]Program in Quantitative Biomedical Sciences, [2]Department of Epidemiology, [3]Department of Molecular and Systems Biology and [4]Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth College, Hanover, NH 03755, USA

*To whom correspondence should be addressed at: Department of Epidemiology, Geisel School of Medicine at Dartmouth College, 1 Medical Center Drive, Williamson Level 6, HB7650, Lebanon, NH 03766, USA. Tel: 603 6501828; Fax: 603 6501840; Email: brock.c.christensen@dartmouth.edu

Abstract

Recent advances in cell-type deconvolution approaches are adding to our understanding of the biology underlying disease development and progression. DNA methylation (DNAm) can be used as a biomarker of cell types, and through deconvolution approaches, to infer underlying cell type proportions. Cell-type deconvolution algorithms have two main categories: reference-based and reference-free. Reference-based algorithms are supervised methods that determine the underlying composition of cell types within a sample by leveraging differentially methylated regions (DMRs) specific to cell type, identified from DNAm measures of purified cell populations. Reference-free algorithms are unsupervised methods for use when cell-type specific DMRs are not available, allowing scientists to estimate putative cellular proportions or control for potential confounding from cell type. Reference-based deconvolution is typically applied to blood samples and has potentiated our understanding of the relation between immune profiles and disease by allowing estimation of immune cell proportions from archival DNA. Bioinformatic analyses using DNAm to infer immune cell proportions, part of a new field known as *Immunomethylomics*, provides a new direction for consideration in epigenome wide association studies (EWAS).

Introduction

Development of multisystem life requires progenitor cells to differentiate in a lineage-specific manner and lineage commitment is accompanied by heritable changes to the epigenome such as DNA methylation (DNAm) (1–3). DNAm is a stable covalent modification to cytosine that can occur in the context of a Cytosine-Guanine dinucleotide (CpG). Genome-scale measures of DNAm in samples derived from heterogeneous mixtures of cells, such as peripheral blood, include signals from all cells present (Fig. 1). Therefore, variation in cell-type proportions across samples has the potential to confound associations of DNAm with modeled outcomes.

Reference-based cell-type deconvolution uses cell-type specific differentially methylated regions (DMRs) to infer cell type proportions, allowing investigators to control for potential confounding by cell type when analysing the relation of DNAm with an outcome of interest. The need to adjust for variation in cell type proportions emerged when genome-scale DNAm arrays were introduced, allowing epigenome-wide association studies (EWAS). Large differences in DNAm observed in early EWAS comparing peripheral blood DNAm between cancer cases
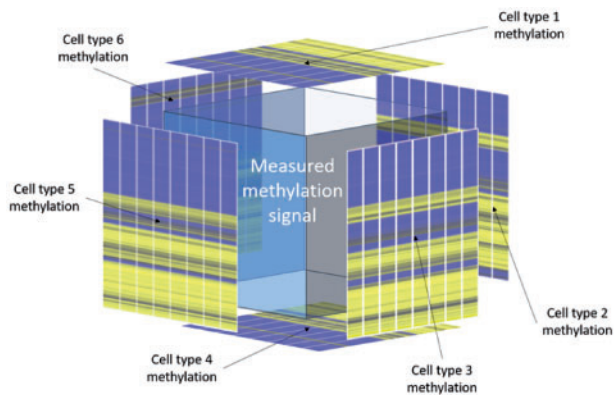
Figure 1. Measured DNA methylation is a composite of signals from specific cell types in a bio-specimen.

and controls were driven by different distributions of cell type proportions between cancer patients and controls. By including cell type proportions in regression models, a clearer understanding of independent DNAm alterations related to disease, exposure, or other outcome of interest is afforded. Importantly, inference of cell type proportions from DNAm also allows for tests of association between cell type proportions (or ratios) themselves with various outcomes, phenotypes, or exposures of interest across the spectrum of health and disease.

To date, epigenome-wide association studies (EWAS) that include adjustment for variation in cell type proportions have been conducted in a wide range of diseases including mental health conditions, cardiovascular disease, and cancers. Methods utilizing DNAm as surrogate estimates of immune cell proportions were initially developed in 2012 by Houseman *et al.* (4) and have gained increasing popularity. More recently, some investigators have begun to focus on analyses of immune cell type proportions and ratios with phenotypes/disease states, termed Immunomethylomics (5). Specifically, in the emerging era of cancer immunotherapy, having additional tools to investigate the relation of peripheral immune status with disease and its response to therapy has potential to accelerate effective application of precision medicine and may contribute to the development of novel therapeutic targets.

In this review, we focus on the most recent applications of reference-based and reference-free cell-type deconvolution with an emphasis on reference-based methods and comment on future directions in the field. Reference-based deconvolution methods depend on cell-type specific differentially methylated regions (DMRs) to infer individual cell type proportions within a sample, typically using constrained projection. Reference-free deconvolution methods do not require DMRs from purified cell types in the measured sample to estimate putative number and proportions of cell types. In both cases, cell type proportions can be included as covariates in regression models to adjust for potential confounding.

DNA Methylation Data

Much of the DNAm data used in EWAS are collected using Illumina BeadArray based technologies. The first version of this array was the GoldenGate (6) platform that included 1,505 CpG sites tracking to ~800 cancer-related genes. Later, more comprehensive genome-scale arrays were developed starting with the HumanMethylation27 (7) in 2009, then the HumanMethylation450,

(8) which was released in 2011, and the current platform, MethylationEPIC (9) array, was released in 2015. Established data processing methods exist for these array-based platforms (10–13). There are also extensive recommendations on controlling for population stratification (14), bias inflation (15), integrating data sources (16), and controlling for cell-type heterogeneity (17). Of the approximately 29 million CpG sites in the human geome (18), ~485,000 sites are measured by the Illumina HumanMethylation450 BeadArray, where the latest version, the MethylationEPIC array measures ~860,000 CpG sites. Most population studies of DNAm use array-based technologies due to cost considerations and advantages of platform standardization for data analysis and data sharing. For those studies using different technologies, such as whole-genome (WGBS) and reduced representation bisulfite sequencing (RRBS), cross platform mapping methods, *e.g. methyLiftover* (16) may provide additional opportunities for cell-type deconvolution applications.

Cell-type deconvolution is based on well-established definitions of cell types that are used in gold standard cell sorting methods, such as fluorescence-activated cell sorting (FACS), which requires sample preparation, antibody tagging, and flow cytometry. Several challenges with flow cytometry methods are the requirement of fresh blood, the need for larger volumes of substrate to isolate less prevalent cell types, and limited number of cell types measurable at one time. Once cell type DMR libraries are identified, cell-type deconvolution using DNAm only requires small amounts of DNA, which can be obtained from whole blood or buffy coat collected with different types of anticoagulant tubes, and archival samples can be used irrespective of storage conditions or freeze-thaw cycles (19). Cell-type deconvolution with DNAm can simultaneously measure a panel of immune types and importantly, as new DMRs for additional cell subtypes or activation states are identified, the DNAm data are largely forward compatible, allowing integration of past and future studies. Gene expression based deconvolution is also possible for cell-type resolution. Methods include CellMix (20) and CIBERSORT (21) for deconvolution of tissue expression profiles, as well as VoCAL (22) and ImmQuant (23) for inferring immune cell-type composition from expression data. These applications have shown promise. However, the number of transcripts can vary in a manner that is not as proportional to the number of cells (24). Gene expression is more likely to be influenced by copy number variation (25) than DNAm (26), and RNA is not as stable as DNA, all of which may increase error in cell type proportion estimates compared with methods using DNAm.

DNAm-based deconvolution relies on carefully selected cell-type DMRs using samples of sorted, purified cell populations from multiple subjects, where DNAm is typically measured using one of the above-mentioned arrays. Cell-type DMRs are invariant among subjects within cell type and often occur at regions known to correlate with expression of genes that are well-established as cell-type-specific. In addition, because of the fundamental role of DNAm in lineage commitment, the unidirectional nature of DNAm changes related to cell-type specification, and the approach to identify cell-type deconvolution libraries, cell-type DMRs do not vary with exposures. To date, immune cell-type DMR libraries have been generated from cells isolated from adults as well as fetal/newborn cord blood samples. The presence of nucleated erythrocytes in fetal/infant cord blood can confound cell-type deconvolution based on adult DMRs (27–29) and has prompted the development of an independent set of DMRs for deconvoluting cell types in cord blood samples (30,31).

In contrast with DMRs used for cell-type deconvolution, there has been recent interest in sets of CpG sites that predict phenotypic characteristics independent of cell type, such as age

(32) and mitotic clocks (33). In 2013, a library of CpG sites whose DNAm status can be used to estimate subject age was identified and validated by Steve Horvath (32). The approach to identify a cell and tissue-type agnostic set of age-related CpGs included ~8,000 samples across 82 data sets encompassing 51 healthy tissues and cell types. These sites were identified through Elastic Net Regression over 21,369 probes shared on the Illumina HumanMethylation27k and HumanMethylation450k arrays and introduced the concept of accelerated DNAm aging for subjects whose estimated DNAm age exceeded their chronological age. Subsequent publications have applied the age-related CpG sites to test the relation of DNAm age acceleration in health and disease (34–36). More recently, an extension of the idea of DNAm age, the epigenetic mitotic clock, was introduced by Yang and colleagues (33). We highlight the utility of cell-type agnostic CpG sets used to predict phenotypes as they have contributed to our understanding of DNAm in human biology, but note that these sets are distinct from cell-type specific DMRs.

Methylation-Based Deconvolution Algorithms

There are two classes of cell-type deconvolution algorithms: reference-based and reference-free. Reference-based deconvolution algorithms are supervised methods that take advantage of known cell-type-specific DMRs (37). The original reference-based algorithm using these DMRs was developed by Houseman *et al.* (38) in 2012 utilizing constrained projection/quadratic programming. The algorithm models sample DNAm as a weighted combination of the individual DNAm patterns of underlying cell types (4,39). The Houseman algorithm has been independently validated in adult whole blood samples (19,40) and until recently, it was the only reference-based algorithm available (21,41) Reference-free deconvolution algorithms do not require referent cell-type DMRs for inference. Instead, reference-free methods estimate putative cellular proportions or the proportion of variation due to unmeasured factors based on various unsupervised deconvolution methods. In 2014, the first two reference-free algorithms were proposed by Zou *et al.* (42) and Houseman *et al.* (43) and in 2016, Houseman *et al.* released a revised version of their method (RefFreeEWAS2.0) based on a form of non-negative matrix factorization (38).

Recently, multiple reference-based (4,21,41) and reference-free (38,42–47) deconvolution algorithms have emerged. In practical application, recent work (41) has shown that non-constrained methods (21,41) can outperform the popular Houseman deconvolution method under certain circumstances (4) but under most circumstances, the Houseman reference-based method implemented in the R package *minfi* (11) is the optimal method for reference-based deconvolution (48). In reference-free deconvolution, recent work (17,48) has shown that Surrogate Variable Analysis (SVA) (45) has the most robust sensitivity and specificity in the presence of high confounding. For a more detailed review of the mathematical (49) and theoretical principles behind various deconvolution approaches, see the recent work by Teschendorff *et al.* (50).

Applications of Deconvolution Algorithms

The impetus behind developing and refining deconvolution algorithms comes from the confounding due to cell-type specific DNAm profiles. These signals may impede the discovery of DNAm changes associated with outcomes of interest. The applicability of cell-type deconvolution algorithms is highlighted in selected recent work below, which illustrates the broad utility of these algorithms for identifying both the role of DNAm and immune cell-type shifts in human health and disease.

Breast cancer

Both reference-based and reference-free deconvolution approaches have been applied to study DNAm in breast cancer, including work on disease risk and tumor biology. For example, Yang *et al.* (51) recently demonstrated evidence that hypomethylation of *S100P* CpGs 3' of the TSS in circulating whole blood is significantly associated with breast cancer risk. The study included data from the HumanMethylation27k platform in >1,000 cases and controls from three independent retrospective cohort studies at different centers. By applying Houseman's reference-based deconvolution algorithm, (4) the authors demonstrated that decreased *S100P* methylation in circulating blood came from the estimated leukocyte subpopulations, excepting B cells. Methylation on the *S100P*-associated CpG sites was inversely correlated with expression of *S100P* in leukocytes and in tumor tissue.

Reference-free approaches to cell type adjustment to study normal breast tissue and breast tumor DNAm have also emerged. Applying the Houseman RefFreeEWAS2.0 (38) algorithm to DNAm data from 100 normal breast tissue samples donated by cancer-free women across a wide age range, Johnson *et al.* (52) tested the relationship between DNAm and known breast cancer risk factors including age, body mass index, and reproductive and family history variables. Johnson *et al.* identified and replicated DNAm changes significantly related to subject age, observed that the CpGs with age-related DNAm in normal tissue were significantly more likely to occur at breast-specific enhancer elements, and that these age-related DNAm changes are implicated in cancer. Another recent study that applied the Houseman RefFreeEWAS2.0 (38) algorithm to study breast cancer focused on early stage invasive breast tumors and identified shared alterations of DNAm relative to normal tissue across intrinsic molecular subtypes (53). Titus *et al.* identified and replicated associations in nineteen differentially methylated gene regions between tumor and normal tissue across all PAM50 (54)-intrinsic molecular subtypes (Basal-like, Luminal A, Luminal B, and Her2). Identification of shared epigenetic alterations in early stage disease across distinct molecular subtypes suggests some common biology underlying breast carcinogenesis irrespective of disease subtype. These examples of varied applications for deconvolution algorithms investigating DNAm in breast cancer demonstrate the utility of adjusting for potential confounding due to cell-type specific DNAm patterns.

Additional applications

There are broad applications for deconvolution of cell type with DNAm. Many recent studies utilizing such techniques extend to analyses in cancer types other than breast, such as glioma (5,55), colorectal (56), and lung (57,58) cancers. In colorectal cancer, Heiss *et al.* (56) extracted DNAm data from leukocytes and identified two CpG sites in the promoter region of *KIAA1549L* with significant differential methylation compared to controls. Then, using *KIAA1549L* promoter methylation in logistic regression models, they created prediction models for colon cancer with *c*-statistics accuracies of 0.69 in the screening setting and 0.73 in the clinical setting. In two recent lung cancer studies, Baglietto *et al.* (58) identified and replicated significant associations between pre-diagnostic peripheral blood DNAm and lung cancer risk in a nested case-control study. Similarly, Zhang *et al.*

(57) showed that peripheral blood DNAm at smoking-related and lung cancer-related genes is associated with lung cancer mortality. Zhang *et al.* also showed smoking-related DNAm is strongly predictive of lung cancer mortality, with an optimism-corrected *c*-index of 0.87. Many other studies investigating the association between smoking and altered methylation profiles have been similarly conducted (59–65).

Many studies have investigated the effects of environmental exposures (66–70) such as air pollution and smoking on the methylome. In 2016, Panni *et al.* (67) observed that fine particulate matter air pollution is associated with adverse health effects, in part through alterations in DNAm measured in peripheral blood. In a 2017 meta-analysis, Gruzieva *et al.* (66) also demonstrated that maternal exposure to NO_2 air pollution during pregnancy is associated with differential DNAm in offspring.

Indeed, the association of child health and development (71) and overall mental health (72–75) have also been the focus of many recent investigations of blood DNAm. In a 1,709 subject study, Walton *et al.* (72) demonstrated that differential methylation, measured in peripheral whole blood at birth, is strongly associated with attention-deficit/hyperactivity disorder (ADHD) trajectories in children. In the children, the same DNAm alterations identified at birth were no longer associated with ADHD by age seven, suggesting that alterations in DNAm play a role in neurodevelopment and maturation.

Studies involving body lipids (76–78), BMI and metabolism (79–86), and diabetes (87–89) have also employed cell-type deconvolution methods. Such methods have also been used to study inflammation (90–92) and physical limitations such as frailty (93) Cardiovascular diseases (94–99), as leading causes of mortality in the US, have been the focus of numerous recent analyses of blood DNAm. Smith *et al.* (95), using a meta-analysis of genome wide association studies and follow-up genotyping, identified a genetic variant on chromosome 5q22 that is associated with a 36% increased risk of death in patients with heart failure. The polymorphism was associated with DNAm alterations in peripheral whole blood that also associated with allergenic sensitization and expression of the cytokine TSLP in blood, identifying new factors associated with mortality in patients with heart failure.

Methylation studies utilizing cell-type deconvolution have also been shown to strongly predict all-cause mortality (100). In a study released in early 2017, Zhang *et al.* (100) conducted an EWAS in a population-based cohort and, using the Houseman reference-based deconvolution method (4), identified 11,063 CpG sites that were significantly related with all-cause mortality at a genome wide level (FDR < 0.05). In addition, these authors then used LASSO regression to select the 10 most informative CpGs in the set to develop a risk score for all-cause mortality where the following ORs for all-cause mortality comparing a
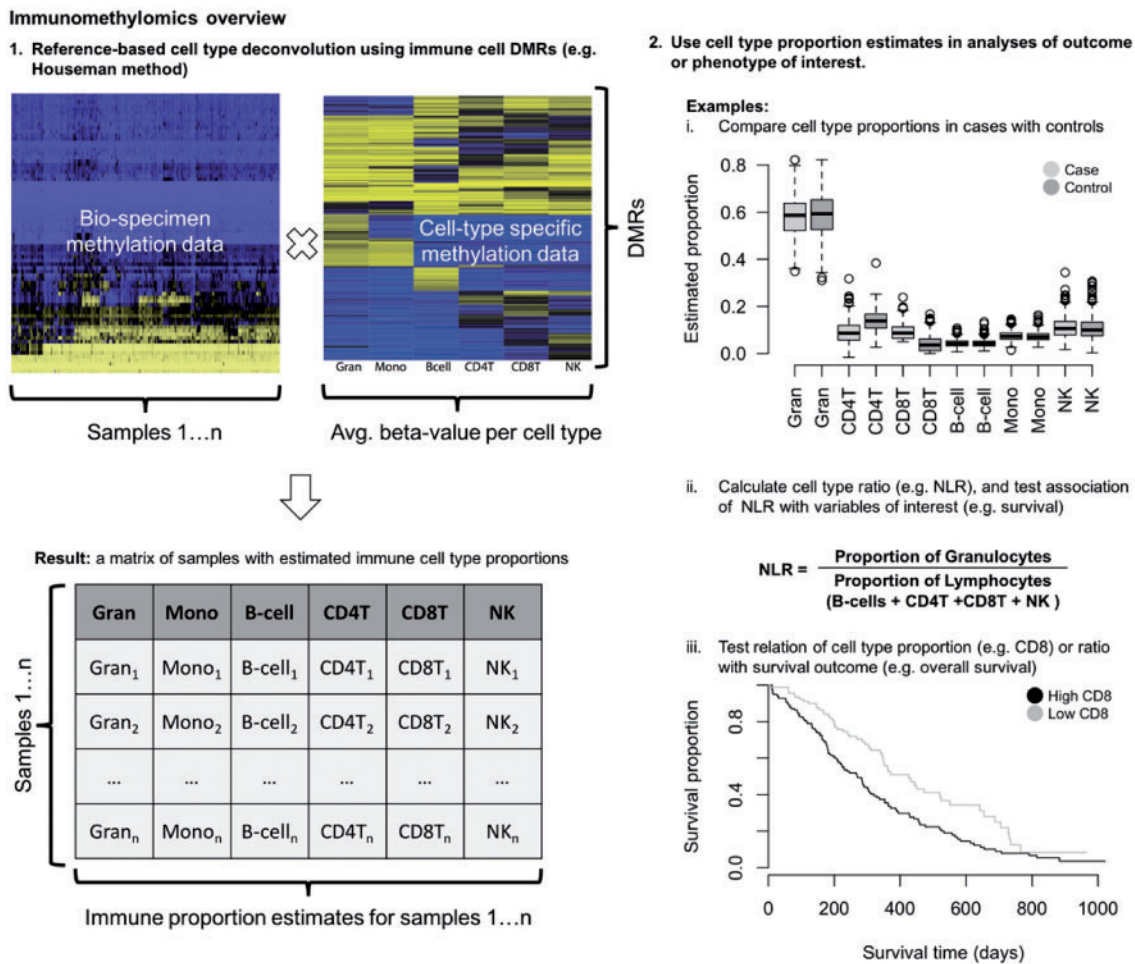

Figure 2. Overview of the *Immunomethylomics* workflow from 1) reference-based deconvolution and cell proportion estimates to 2) utilizing cell proportion estimates in analyses.

score of 0 were noted; a score of 1 (OR 2.16, CI 1.10-4.24), score ranging from 2-5 (OR 3.42, CI 1.81-6.46), and scores >5 (OR 7.36, CI 3.69-14.68). The risk score was found to be independent of cell-type DMRs (4) as well as the Horvath methylation age CpG loci (32). Ultimately, the goal of many of these studies is the development of novel biomarkers (101) of disease risk and prognosis that contribute to our understanding of biology and may aid in risk stratification. There are many applications beyond those listed here that benefit from methylation and cell-type deconvolution based analyses.

Immunomethylomics – DNA methylation analysis of the immune response

Traditional EWAS includes tests of association for methylation at 400,000 – 800,000 CpG loci with phenotype or disease status. The burden of correcting for multiple hypothesis testing in such settings leads to conservative cutoffs that may result in false negative associations, particularly when smaller sample sizes are used. An opportunity in this setting, particularly if immune regulation or response is related to or may mediate associations with the outcome of interest, is to first test the relation of inferred proportions of immune cell types with the outcome of interest. Inference of CD4+ T-cell, CD8+ T-cell, B-cell, NK, Granulocyte, and Monocyte proportions from DNAm using deconvolution has allowed researchers to directly use immune cell proportions or ratios in their analyses (Fig. 2). For example, the neutrophil-to-lymphocyte ratio (NLR) is a consistent prognostic factor for survival in cancer (102) and can be derived using cell type proportion estimates from DNAm-based deconvolution of peripheral blood (103). Koestler *et al.* (103) have shown that methylation-derived NLR (mdNLR) from referenced-based deconvolution is also a robust indicator of patient prognosis. In the paper published in early 2017, the authors demonstrated the utility of mdNLR as an indicator of prognosis in five studies: head and neck squamous cell carcinoma, bladder cancer, ovarian cancer, breast cancer, and a healthy aging study. In each instance, Koestler *et al.* showed that elevated mdNLR was indicative of disease or clinical characteristics of healthy patients. In another study of 72 patients with glioma, Wiencke *et al.* (5) showed that subjects with gliomas have elevated neutrophil-to-lymphocyte ratio scores and significantly decreased survival times (HR 2.02, 95% CI 1.11-3.69). The authors also demonstrated that mdNLR is an effective candidate methylation biomarker with applications in EWAS and clinical studies of glioma. In addition to the utility of DMRs for deconvolution of cell types, there are also opportunities to identify DMRs for activation states within cell types such as NK cell activation. For example, Wiencke et al. isolated naïve NK cells from six donors, performed *in vitro* activation treating with antibodies to the NKp46 and CD2 receptors, and profiled DNAm to identify DMRs for activated versus naïve NK cells (104). These initial studies demonstrate the utility of DNA-based methods to deconvolute cell type to investigate immunomodulation in human health and disease without the need for fresh or specially-preserved samples.

Future Directions

Since their introduction, reference-based algorithms such as the Houseman method have been optimized for more accurate cell-type estimation by either improving DMR selection and/or by improving the subset of data used from the reference libraries. Optimization (105) of such DMRs has been shown to explain additional variation in EWAS (106) and can be used in single step and multi-step (107) deconvolution algorithms. Accurate estimation of leukocyte subtypes from DNAm has opened the door for the development of novel methods to further investigate the relation of immune profiles with disease and outcomes. In conjunction with development of additional deconvolution methods and more complete DMR libraries for immune cell types, directions for future work include refinement to identify DMRs for further cell subtypes or cell activation states. Cell estimates provide an opportunity to independently test associations of immune cell proportions or ratios with the disease (with a limited number of statistical tests), before taking more traditional epigenome wide approaches. Similarly, classifications blocks for phenotypes derived from DNA methylation data such as DNA methylation age or the epigenetic mitotic clock offer an opportunity to test specific hypotheses before extensive feature-by-feature tests in an EWAS.

The potential application of deconvolution methods goes beyond DNAm in peripheral blood samples. As reference libraries of purified epithelial, mesenchymal, and progenitor cell subtypes emerge, it is expected that DMRs for reference-based deconvolution will be extended to solid tissue samples for use in pathologically normal and diseased subjects. In this way, deconvolution methods will improve our understanding of immune cell infiltration in pathogenesis of several diseases. In addition, reference-free and reference-based deconvolution approaches in solid tissues can assist researchers in modeling the relation of exogenous exposures with networks of cell-cell interactions in the tissue microenvironment in disease pathogenesis. Collectively, advances in the identification of DMR libraries for cell type and biologic phenotypes have broad potential to accelerate our understanding of immunomodulation, cell systems, and DNAm in human health and disease.

Conflict of Interest statement. None declared.

Funding

References

1. Khavari, D.A., Sen, G.L. and Rinn, J.L. (2010) DNA methylation and epigenetic control of cellular differentiation. *Cell Cycle*, **9**, 3880–3883.
2. Ji, H., Ehrlich, L.I.R., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K. *et al.* (2010) Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature*, **467**, 338–342.
3. Rönnerblad, M., Andersson, R., Olofsson, T., Douagi, I., Karimi, M., Lehmann, S.S., Hoof, I., de Hoon, M., Itoh, M., Nagao-Sato, S. *et al.* (2014) Analysis of the DNA methylome and transcriptome in granulopoiesis reveals timed changes and dynamic enhancer methylation. *Blood*, **123**, e79–e89.
4. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K. and Kelsey, K.T. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, **13**, 86.

5. Wiencke, J.K., Koestler, D.C., Salas, L.A., Wiemels, J.L., Roy, R.P., Hansen, H.M., Rice, T., McCoy, L.S., Bracci, P.M., Molinaro, A.M. *et al.* (2017) Immunomethylomic approach to explore the blood neutrophil lymphocyte ratio (NLR) in glioma survival. *Clin. Epigenetics*, **9**, 10.

6. Bibikova, M. and Fan, J.-B. (2009) GoldenGate assay for DNA methylation profiling. *Methods Mol. Biol.*, **507**, 149–163.

7. Bibikova, M., Le J, Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R. and Gunderson, K.L. (2009) Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics*, **1**, 177–200.

8. Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L. *et al.* (2011) High density DNA methylation array with single CpG site resolution. *Genomics*, **98**, 288–295.

9. Moran, S., Arribas, C. and Esteller, M. (2016) Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*, **8**, 389–399.

10. Wilhelm-Benartzi, C.S., Koestler, D.C., Karagas, M.R., Flanagan, J.M., Christensen, B.C., Kelsey, K.T., Marsit, C.J., Houseman, E.A. and Brown, R. (2013) Review of processing and analysis methods for DNA methylation array data. *Br. J. Cancer*, 1394–1402.

11. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D. and Irizarry, R.A. (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, **30**, 1363–1369.

12. Assenov, Y., Muller, F., Lutsik, P., Walter, J., Lengauer, T. and Bock, C. (2014) Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods*, **11**, 1138–1140.

13. Maksimovic, J., Phipson, B. and Oshlack, A. (2016) A cross-package Bioconductor workflow for analysing methylation array data. *F1000Research*, **5**, 1281.

14. Barfield, R.T., Almli, L.M., Kilaru, V., Smith, A.K., Mercer, K.B., Duncan, R., Klengel, T., Mehta, D., Binder, E.B., Epstein, M.P. *et al.* (2014) Accounting for Population Stratification in DNA Methylation Studies. *Genet. Epidemiol.*, 231–241.

15. van Iterson, M., van Zwet, E.W. and Heijmans, B.T. (2017) Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* **18**, 19.

16. Titus, A.J., Houseman, E.A., Johnson, K.C. and Christensen, B.C. (2016) MethyLiftover: cross-platform DNA methylation data integration. *Bioinformatics*, **32**, 2517–2519.

17. McGregor, K., Bernatsky, S., Colmegna, I., Hudson, M., Pastinen, T., Labbe, A. and Greenwood, C.M.T. (2016) An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. *Genome Biol.*, **17**, 84.

18. Lövkvist, C., Dodd, I.B., Sneppen, K. and Haerter, J.O. (2016) DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Res.*, **44**, 5123–5132.

19. Accomando, W.P., Wiencke, J.K., Houseman, E.A., Nelson, H.H. and Kelsey, K.T. (2014) Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biol.*, **15**, R50.

20. Gaujoux, R. and Seoighe, C. (2013) CellMix: a comprehensive toolbox for gene expression deconvolution. *Bioinformatics*, **29**, 2211–2212.

21. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M. and Alizadeh, A.A. (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat. Meth.*, **12**, 453–457.

22. Steuerman, Y. and Gat-Viks, I. (2016) Exploiting Gene-Expression Deconvolution to Probe the Genetics of the Immune System. *PLOS Comput. Biol.*, **12**, e1004856.

23. Frishberg, A., Brodt, A., Steuerman, Y. and Gat-Viks, I. (2016) ImmQuant: a user-friendly tool for inferring immune cell-type composition from gene-expression data. *Bioinformatics*, **32**, 3842–3843.

24. Tian, Q., Stepaniants, S.B., Mao, M., Weng, L., Feetham, M.C., Doyle, M.J., Yi, E.C., Dai, H., Thorsson, V., Eng, J. *et al.* (2004) Integrated genomic and proteomic analyses of gene expression in Mammalian cells. *Mol. Cell. Proteomics*, **3**, 960–969.

25. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C. *et al.* (2007) Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. *Science*, **315**, 848–853.

26. Houseman, E.A., Christensen, B.C., Karagas, M.R., Wrensch, M.R., Nelson, H.H., Wiemels, J.L., Zheng, S., Wiencke, J.K., Kelsey, K.T. and Marsit, C.J. (2009) Copy number variation has little impact on bead-array-based measures of DNA methylation. *Bioinformatics*, **25**, 1999–2005.

27. de Goede, O.M., Razzaghian, H.R., Price, E.M., Jones, M.J., Kobor, M.S., Robinson, W.P. and Lavoie, P.M. (2015) Nucleated red blood cells impact DNA methylation and expression analyses of cord blood hematopoietic cells. *Clin. Epigenetics*, **7**, 95.

28. de Goede, O.M., Lavoie, P.M. and Robinson, W.P. (2016) Characterizing the hypomethylated DNA methylation profile of nucleated red blood cells from cord blood. *Epigenomics*, **8**, 1481–1494.

29. Bakulski, K.M., Feinberg, J.I., Andrews, S.V., Yang, J., Brown, S., McKenney Witter, L.S., Walston, F., Feinberg, J. and Fallin, A.P. MD. DNA (2016) methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics*, **11**, 354–362.

30. Gervin, K., Page, C.M., Aass, H.C.D., Jansen, M.A., Fjeldstad, H.E., Andreassen, B.K., Duijts, L., van Meurs, J.B., van Zelm, M.C., Jaddoe, V.W. *et al.* (2016) Cell type specific DNA methylation in cord blood: A 450K-reference data set and cell count-based validation of estimated cell type composition. *Epigenetics*, **11**, 690–698.

31. Cardenas, A., Allard, C., Doyon, M., Houseman, E.A., Bakulski, K.M., Perron, P., Bouchard, L. and Hivert, M.-F. (2016) Validation of a DNA methylation reference panel for the estimation of nucleated cells types in cord blood. *Epigenetics*, 773–779.

32. Horvath, S.D.N.A. (2013) methylation age of human tissues and cell types. *Genome Biol.*, **14**, R115.

33. Yang, Z., Wong, A., Kuh, D., Paul, D.S., Rakyan, V.K., Leslie, R.D., Zheng, S.C., Widschwendter, M., Beck, S. and Teschendorff, A.E. (2016) Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol.*, **17**, 205.

34. Lu, A.T., Hannon, E., Levine, M.E., Crimmins, E.M., Lunnon, K., Mill, J., Geschwind, D.H. and Horvath, S. (2017) Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. *Nat. Commun.*, **8**, 15353.

35. Sehl, M.E., Henry, J.E., Storniolo, A.M., Ganz, P.A. and Horvath, S. (2017) DNA methylation age is elevated in breast tissue of healthy women. *Breast Cancer Res. Treat.*, 1–11.

36. Cole, J.H., Ritchie, S.J., Bastin, M.E., Valdes Hernandez, M.C., Munoz Maniega, S., Royle, N., Corley, J., Pattie, A., Harris,

S.E., Zhang, Q. *et al*. (2017) Brain age predicts mortality. *Mol. Psychiatry*, doi: 10.1038/mp.2017.62.

37. Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlen, S.-E., Greco, D., Soderhall, C., Scheynius, A. and Kere, J. (2012) Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One*, **7**, e41361.

38. Houseman, E.A., Kile, M.L., Christiani, D.C., Ince, T.A., Kelsey, K.T. and Marsit, C.J. (2016) Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *BMC Bioinformatics*, **17**, 259.

39. Wiencke, J.K., Accomando, W.P., Zheng, S., Patoka, J., Dou, X., Phillips, J.J., Hsuang, G., Christensen, B.C., Houseman, E.A., Koestler, D.C. *et al*. (2012) Epigenetic biomarkers of T-cells in human glioma. *Epigenetics*, 1391–1402.

40. Koestler, D.C., Christensen, B., Karagas, M.R., Marsit, C.J., Langevin, S.M., Kelsey, K.T., Wiencke, J.K. and Houseman, E.A. (2013) Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. *Epigenetics*, **8**, 816–826.

41. Teschendorff, A.E., Breeze, C.E., Zheng, S.C. and Beck, S.A. (2017) comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinformatics*, **18**, 105.

42. Zou, J., Lippert, C., Heckerman, D., Aryee, M. and Listgarten, J. (2014) Epigenome-wide association studies without the need for cell-type composition. *Nat. Meth*., **11**, 309–311.

43. Houseman, E.A., Molitor, J. and Marsit, C.J. (2014) Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics*, 1431–1439.

44. Rahmani, E., Zaitlen, N., Baran, Y., Eng, C., Hu, D., Galanter, J., Oh, S., Burchard, E.G., Eskin, E., Zou, J. *et al*. (2016) Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nat. Meth*., **13**, 443–445.

45. Leek, J.T. (2007) Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. *PLoS Genet*., **3**, 1724–1735.

46. Teschendorff, A.E., Zhuang, J. and Widschwendter, M. (2011) Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics*, **27**, 1496–1505.

47. Gagnon-Bartsch, J.A. and Speed, T.P. (2012) Using control genes to correct for unwanted variation in microarray data. *Biostatistics*, **13**, 539–552.

48. Kaushal, A., Zhang, H., Karmaus, W.J.J., Ray, M., Torres, M.A., Smith, A.K. and Wang, S.-L. (2017) Comparison of different cell type correction methods for genome-scale epigenetics studies. *BMC Bioinformatics*, **18**, 216.

49. Houseman, E.A., Kelsey, K.T., Wiencke, J.K. and Marsit, C.J. (2015) Cell-composition effects in the analysis of DNA methylation array data: a mathematical perspective. *BMC Bioinformatics*, **16**, 95.

50. Teschendorff, A.E. and Zheng, S.C. (2017) Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics*, **9**, 757–768.

51. Yang, R., Stöcker, S., Schott, S., Heil, J., Marme, F., Cuk, K., Chen, B., Golatta, M., Zhou, Y., Sutter, C. *et al*. (2017) The association between breast cancer and S100P methylation in peripheral blood by multicenter case–control studies. *Carcinogenesis*, **38**, 312–320.

52. Johnson, K.C., Houseman, E.A., King, J.E. and Christensen, B.C. (2017) Normal breast tissue DNA methylation differences at regulatory elements are associated with the cancer risk factor age. *Breast Cancer Res*., **19**, 81.

53. Titus, A.J., Way, G.P., Johnson, K.C. and Christensen, B.C. (2017) Deconvolution of DNA methylation identifies differentially methylated gene regions on 1p36 across breast cancer subtypes. *bioRxiv*, doi: https://doi.org/10.1101/115873.

54. Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z. *et al*. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol*., **27**, 1160–1167.

55. Johnson, K.C., Houseman, E.A., King, J.E., von Herrmann, K.M., Fadul, C.E. and Christensen, B.C. (2016) 5-Hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients. *Nat. Commun.*

56. Heiss, J.A. and Brenner, H. (2017) Epigenome-wide discovery and evaluation of leukocyte DNA methylation markers for the detection of colorectal cancer in a screening setting. *Clin. Epigenetics*, **9**, 24.

57. Zhang, Y., Breitling, L.P., Balavarca, Y., Holleczek, B., Schöttker, B. and Brenner, H. (2016) Comparison and combination of blood DNA methylation at smoking-associated genes and at lung cancer-related genes in prediction of lung cancer mortality. *Int. J. Cancer*, **139**, 2482–2492.

58. Baglietto, L., Ponzi, E., Haycock, P., Hodge, A., Bianca Assumma, M., Jung, C.-H., Chung, J., Fasanelli, F., Guida, F., Campanella, G. *et al*. (2017) DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. *Int. J. Cancer*, **140**, 50–61.

59. Rotroff, D.M., Joubert, B.R., Marvel, S.W., Haberg, S.E., Wu, M.C., Nilsen, R.M., Ueland, P.M., Nystad, W., London, S.J. and Motsinger-Reif, A. (2016) Maternal smoking impacts key biological pathways in newborns through epigenetic modification in Utero. *BMC Genomics*, **17**, 976.

60. Joehanes, R., Just, A.C., Marioni, R.E., Pilling, L.C., Reynolds, L.M., Mandaviya, P.R., Guan, W., Xu, T., Elks, C.E., Aslibekyan, S. *et al*. (2016) Epigenetic signatures of cigarette smoking. *Circ. Cardiovasc. Genet*., **9**, 436–447.

61. Gao, X., Mons, U., Zhang, Y., Breitling, L.P. and Brenner, H. (2016) DNA methylation changes in response to active smoking exposure are associated with leukocyte telomere length among older adults. *Eur. J. Epidemiol*., **31**, 1231–1241.

62. Zhang, Y., Florath, I., Saum, K.-U. and Brenner, H. (2016) Self-reported smoking, serum cotinine, and blood DNA methylation. *Environ. Res*., **146**, 395–403.

63. Georgiadis, P., Hebels, D.G., Valavanis, I., Liampa, I., Bergdahl, I.A., Johansson, A., Palli, D., Chadeau-Hyam, M., Chatziioannou, A., Jennen, D.G.J. *et al*. (2016) Omics for prediction of environmental health effects: Blood leukocyte-based cross-omic profiling reliably predicts diseases associated with tobacco smoking. *Sci. Rep*., **6**, 20544.

64. Rzehak, P., Saffery, R., Reischl, E., Covic, M., Wahl, S., Grote, V., Xhonneux, A., Langhendries, J.-P., Ferre, N., Closa-Monasterolo, R. *et al*. (2016) Maternal Smoking during Pregnancy and DNA-Methylation in Children at Age 5.5 Years: Epigenome-Wide-Analysis in the European Childhood Obesity Project (CHOP)-Study. *PLoS One*, **11**, e0155554.

65. Joubert, B.R., Felix, J.F., Yousefi, P., Bakulski, K.M., Just, A.C., Breton, C., Reese, S.E., Markunas, C.A., Richmond, R.C., Xu, C-J. *et al*. (2016) DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. *Am. J. Hum. Genet*., **98**, 680–696.

66. Gruzieva, O., Xu, C.-J., Breton, C.V., Annesi-Maesano, I., Antó, J.M., Auffray, C., Ballereau, S., Bellander, T., Bousquet, J., Bustamante, M. *et al*. (2017) Epigenome-wide

meta-analysis of methylation in children related to prenatal NO(2) air pollution exposure. *Environ. Health Perspect.*, **125**, 104–110.

67. Panni, T., Mehta, A.J., Schwartz, J.D., Baccarelli, A.A., Just, A.C., Wolf, K., Wahl, S., Cyrys, J., Kunze, S., Strauch, K. *et al.* (2016) Genome-wide analysis of DNA methylation and fine particulate matter air pollution in three study populations: KORA F3, KORA F4, and the normative aging study. *Environ. Health Perspect.*, **124**, 983–990.

68. Green, B.B., Karagas, M.R., Punshon, T., Jackson, B.P., Robbins, D.J., Houseman, E.A. and Marsit, C.J. (2016) Epigenome-wide assessment of DNA methylation in the placenta and arsenic exposure in the New Hampshire Birth Cohort Study (USA). *Environ. Health Perspect.*, **124**, 1253–1260.

69. Quach, A., Levine, M.E., Tanaka, T., Lu, A.T., Chen, B.H., Ferrucci, L., Ritz, B., Bandinelli, S., Neuhouser, M.L., Beasley, J.M. *et al.* (2017) Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany. NY)*, **9**, 419–437.

70. Cardenas, A., Rifas-Shiman, S.L., Agha, G., Hivert, M.-F., Litonjua, A.A., DeMeo, D.L., Lin, X., Amarasiriwardena, C.J., Oken, E., Gillman, M.W. *et al.* (2017) Persistent DNA methylation changes associated with prenatal mercury exposure and cognitive performance during childhood. *Sci. Rep.*, **7**, 288.

71. Xu, C.-J., Bonder, M.J., Soderhall, C., Bustamante, M., Baiz, N., Gehring, U., Jankipersadsing, S.A., van der Vlies, P., van Diemen, C.C., van Rijkom, B. *et al.* (2017) The emerging landscape of dynamic DNA methylation in early childhood. *BMC Genomics*, **18**, 25.

72. Walton, E., Pingault, J.-B., Cecil, C.A.M., Gaunt, T.R., Relton, C.L., Mill, J. and Barker, E.D. (2017) Epigenetic profiling of ADHD symptoms trajectories: a prospective, methylome-wide study. *Mol. Psychiatry*, **22**, 250–256.

73. Edvinsson, Å., Bränn, E., Hellgren, C., Freyhult, E., White, R., Kamali-Moghaddam, M., Olivier, J., Bergquist, J., Boström, A.E., Schiöth, H.B. *et al.* (2017) Lower inflammatory markers in women with antenatal depression brings the M1/M2 balance into focus from a new direction. *Psychoneuroendocrinology*, **80**, 15–25.

74. Hannon, E., Dempster, E., Viana, J., Burrage, J., Smith, A.R., Macdonald, R., St Clair, D., Mustard, C., Breen, G., Therman, S. *et al.* (2016) An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol.*, **17**, 176.

75. Clive, M.L., Boks, M.P., Vinkers, C.H., Osborne, L.M., Payne, J.L., Ressler, K.J., Smith, A.K., Wilcox, H.C. and Kaminsky, Z. (2016) Discovery and replication of a peripheral tissue DNA methylation biosignature to augment a suicide prediction model. *Clin. Epigenetics*, **8**, 113.

76. Dekkers, K.F., van Iterson, M., Slieker, R.C., Moed, M.H., Bonder, M.J., van Galen, M., Mei, H., Zhernakova, D.V., van den Berg, L.H., Deelen, J. *et al.* (2016) Blood lipids influence DNA methylation in circulating cells. *Genome Biol.*, **17**, 138.

77. Braun, K.V.E., Dhana, K., de Vries, P.S., Voortman, T., van Meurs, J.B.J., Uitterlinden, A.G., Hofman, A., Hu, F.B., Franco, O.H. and Dehghan, A. (2017) Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. *Clin. Epigenetics*, **9**, 15.

78. Sayols-Baixeras, S., Subirana, I., Lluis-Ganella, C., Civeira, F., Roquer, J., Do, A.N., Absher, D., Cenarro, A., Muñoz, D., Soriano-Tárraga, C. *et al.* (2016) Identification and validation of seven new loci showing differential DNA methylation related to serum lipid profile: an epigenome-wide approach. The REGICOR study. *Hum. Mol. Genet.*, **25**, 4556–4565.

79. Wilson, L.E., Harlid, S., Xu, Z., Sandler, D.P. and Taylor, J.A. (2017) An epigenome-wide study of body mass index and DNA methylation in blood using participants from the Sister Study cohort. *Int. J. Obes.*, **41**, 194–199.

80. Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W.R., Kunze, S., Tsai, P.-C., Ried, J.S., Zhang, W., Yang, Y. *et al.* (2017) Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature*, **541**, 81–86.

81. Ali, O., Cerjak, D., Kent, J.W., James, R., Blangero, J., Carless, M.A. and Zhang, Y. (2016) Methylation of SOCS3 is inversely associated with metabolic syndrome in an epigenome-wide association study of obesity. *Epigenetics*, **11**, 699–707.

82. Agha, G., Hajj, H., Rifas-Shiman, S.L., Just, A.C., Hivert, M.-F., Burris, H.H., Lin, X., Litonjua, A.A., Oken, E., DeMeo, D.L. *et al.* (2016) Birth weight-for-gestational age is associated with DNA methylation at birth and in childhood. *Clin. Epigenetics*, **8**, 118.

83. Kriebel, J., Herder, C., Rathmann, W., Wahl, S., Kunze, S., Molnos, S., Volkova, N., Schramm, K., Carstensen-Kirberg, M., Waldenberger, M. *et al.* (2016) Association between DNA Methylation in Whole Blood and Measures of Glucose Metabolism: KORA F4 Study. *PLoS One*, **11**, e0152314.

84. Mamtani, M., Kulkarni, H., Dyer, T.D., Göring, H.H.H., Neary, J.L., Cole, S.A., Kent, J.W., Kumar, S., Glahn, D.C., Mahaney, M.C. *et al.* (2016) Genome- and epigenome-wide association study of hypertriglyceridemic waist in Mexican American families. *Clin. Epigenetics*, **8**, 6.

85. Lin, X., Lim, I.Y., Wu, Y., Teh, A.L., Chen, L., Aris, I.M., Soh, S.E., Tint, M.T., MacIsaac, J.L., Morin, A.M. *et al.* (2017) Developmental pathways to adiposity begin before birth and are influenced by genotype, prenatal environment and epigenome. *BMC Med.*, **15**, 50.

86. Sharp, G.C., Salas, L.A., Monnereau, C., Allard, C., Yousefi, P., Everson, T., London, S.J., Felix, J.F. and Relton, C.L. Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: Findings from the Pregnancy and Childhood Epigenetics (PACE) consortium. *bioRxiv*. 2017.

87. Elliott, H.R., Shihab, H.A., Lockett, G.A., Holloway, J.W., McRae, A.F., Smith, G.D., Ring, S.M., Gaunt, T.R. and Relton, C.L. (2017) The role of DNA methylation in Type 2 diabetes aetiology – using genotype as a causal anchor. *Diabetes*, **66**, 1713–1722.

88. Soriano-Tárraga, C., Jiménez-Conde, J., Giralt-Steinhauer, E., Mola-Caminal, M., Vivanco-Hidalgo, R.M., Ois, A., Rodríguez-Campello, A., Cuadrado-Godia, E., Sayols-Baixeras, S., Elosua, R. *et al.* (2016) Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. *Hum. Mol. Genet*, **25**, 609–619.

89. Florath, I., Butterbach, K., Heiss, J., Bewerunge-Hudler, M., Zhang, Y., Schöttker, B. and Brenner, H. (2016) Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. *Diabetologia*, **59**, 130–138.

90. Ventham, N.T., Kennedy, N.A., Adams, A.T., Kalla, R., Heath, S., O'Leary, K.R., Drummond, H., Wilson, D.C., Gut, I.G., Nimmo, E.R. *et al.* (2016) Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate

genetic risk in inflammatory bowel disease. *Nat. Commun.*, **7**, 13507.

91. Meng, W., Zhu, Z., Jiang, X., Too, C.L., Uebe, S., Jagodic, M., Kockum, I., Murad, S., Ferrucci, L., Alfredsson, L. *et al.* (2017) DNA methylation mediates genotype and smoking interaction in the development of anti-citrullinated peptide antibody-positive rheumatoid arthritis. *Arthritis Res. Ther.*, **19**, 71.

92. Ligthart, S., Marzi, C., Aslibekyan, S., Mendelson, M.M., Conneely, K.N., Tanaka, T., Colicino, E., Waite, L.L., Joehanes, R., Guan, W. *et al.* (2016) DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.*, **17**, 255.

93. Breitling, L.P., Saum, K.-U., Perna, L., Schöttker, B., Holleczek, B. and Brenner, H. (2016) Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. *Clin. Epigenetics*, **8**, 21.

94. Lin, H., Yin, X., Xie, Z., Lunetta, K.L., Lubitz, S.A., Larson, M.G., Ko, D., Magnani, J.W., Mendelson, M.M., Liu, C. *et al.* (2017) Methylome-wide Association Study of Atrial Fibrillation in Framingham Heart Study. *Sci. Rep.*, **7**, 40377.

95. Smith, J.G., Felix, J.F., Morrison, A.C., Kalogeropoulos, A., Trompet, S., Wilk, J.B., Gidlöf, O., Wang, X., Morley, M., Mendelson, M. *et al.* (2016) Discovery of genetic variation on chromosome 5q22 associated with mortality in heart failure. *PLOS Genet.*, **12**, e1006034.

96. Hedman, Å.K., Mendelson, M.M., Marioni, R.E., Gustafsson, S., Joehanes, R., Irvin, M.R., Zhi, D., Sandling, J.K., Yao, C., Liu, C. *et al.* (2017) Epigenetic patterns in blood associated with lipid traits predict incident coronary heart disease events and are enriched for results from genome-wide association studies. *Circ. Cardiovasc. Genet.*, **10**.

97. Chatziioannou, A., Georgiadis, P., Hebels, D.G., Liampa, I., Valavanis, I., Bergdahl, I.A., Johansson, A., Palli, D., Chadeau-Hyam, M., Siskos, A.P. *et al.* (2017) Blood-based omic profiling supports female susceptibility to tobacco smoke-induced cardiovascular diseases. *Sci. Rep.*, **7**, 42870.

98. Rask-Andersen, M., Martinsson, D., Ahsan, M., Enroth, S., Ek, W.E., Gyllensten, U. and Johansson, Å. (2016) Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction. *Hum. Mol. Genet.*, **25**, 4739–4748.

99. Hedman, Å.K., Zilmer, M., Sundström, J., Lind, L. and Ingelsson, E.D.N.A. (2016) methylation patterns associated with oxidative stress in an ageing population. *BMC Med. Genomics*, **9**, 72.

100. Zhang, Y., Wilson, R., Heiss, J., Breitling, L.P., Saum, K.-U., Schöttker, B., Holleczek, B., Waldenberger, M., Peters, A. and Brenner, H. (2017) DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat. Commun.*, **8**, 14617.

101. Langie, S.A.S., Moisse, M., Declerck, K., Koppen, G., Godderis, L., Vanden Berghe, W., Drury, S. and De Boever, P. (2017) Salivary DNA Methylation Profiling: Aspects to Consider for Biomarker Identification. *Basic Clin. Pharmacol. Toxicol.*

102. Chen, J., Deng, Q., Pan, Y., He, B., Ying, H., Sun, H., Liu, X. and Wang, S. (2015) Prognostic value of neutrophil-to-lymphocyte ratio in breast cancer. *FEBS Open Bio.*, **5**, 502–507.

103. Koestler, D.C., Usset, J.L., Christensen, B.C., Marsit, C.J., Karagas, M.R., Kelsey, K.T. and Wiencke, J.K. (2016) DNA methylation-derived neutrophil-to-lymphocyte ratio: an epigenetic tool to explore cancer inflammation and outcomes. *Cancer Epidemiol. Biomarkers Prev.*, **26**, 328–338.

104. Wiencke, J.K., Butler, R., Hsuang, G., Eliot, M., Kim, S., Sepulveda, M.A., Siegel, D., Houseman, E.A. and Kelsey, K.T. (2016) The DNA methylation profile of activated human natural killer cells. *Epigenetics*, **11**, 363–380.

105. Koestler, D.C., Jones, M.J., Usset, J., Christensen, B.C., Butler, R.A., Kobor, M.S., Wiencke, J.K. and Kelsey, K.T. (2016) Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). *BMC Bioinformatics*, **17**, 1–21.

106. Kim, S., Eliot, M., Koestler, D.C., Houseman, E.A., Wetmur, J.G., Wiencke, J.K. and Kelsey, K.T. (2016) Enlarged leukocyte referent libraries can explain additional variance in blood-based epigenome-wide association studies. *Epigenomics*, **8**, 1185–1192.

107. Waite, L.L., Weaver, B., Day, K., Li, X., Roberts, K., Gibson, A.W., Edberg, J.C., Kimberly, R.P., Absher, D.M. and Tiwari, H.K. (2016) Estimation of Cell-Type Composition Including T and B Cell Subtypes for Whole Blood Methylation Microarray Data. *Front. Genet.*, **7**, 23.