EPIGENOMICS
2025, VOL. 17, NO. 9, 599–610
https://doi.org/10.1080/17501911.2025.2500907

Taylor & Francis
Taylor & Francis Group

PERSPECTIVE

The predictive power of profiling the DNA methylome in human health and disease

Paraskevi Christofidou [a,b] and Christopher G. Bell [a,b]

[a]William Harvey Research Institute, Barts & The London Faculty of Medicine and Dentistry, Queen Mary University of London, London, UK; [b]QMUL Centre for Epigenetics, Queen Mary University of London, London, UK

ABSTRACT

Early and accurate diagnosis significantly improves the chances of disease survival. DNA methylation (5mC), the major DNA modification in the human genome, is now recognized as a biomarker of immense clinical potential. This is due to its ability to delineate precisely cell-type, quantitate both internal and external exposures, as well as tracking chronological and biological components of the aging process. Here, we survey the current state of DNA methylation as a biomarker and predictor of traits and disease. This includes Epigenome-wide association study (EWAS) findings that inform Methylation Risk Scores (MRS), EpiScore long-term estimators of plasma protein levels, and machine learning (ML) derived DNA methylation clocks. These all highlight the significant benefits of accessible peripheral blood DNA methylation as a surrogate measure. However, detailed DNA methylation biopsy analysis in real-time is also empowering pathological diagnosis. Furthermore, moving forward, in this multi-omic and biobank scale era, novel insights will be enabled by the amplified power of increasing sample sizes and data integration.

ARTICLE HISTORY
Received 9 January 2025
Accepted 28 April 2025

KEYWORDS
Epigenetics; epigenomics; DNA methylation; methylation risk scores; EWAS; epigenetic estimators; epigenetic clocks; biological ageing

1. Introduction

Accurate early disease prediction brings significant personal, societal and economic health benefits [1]. With the ability to gather multi-omic measures – including genomic, transcriptomic, and epigenomic – nowadays at an unprecedented scale, we have the great opportunity of identifying, in this mountain of information, disease markers of robust clinical utility. Epigenomic data comprising DNA modifications, post-translational histone tail alterations and histone variants, have enormous potential in this medical revolution and are set to become a major component in predicting, diagnosing and monitoring human disease. The possibilities include screening biomarkers in surrogate tissue to detailed biopsy analyses in pathogenic organs informing on functional abrogation [2].

Both inherited and external factors contribute to common disease risk. To construct a more complete representation, optimal personalized disease prediction and health care, therefore, require bringing together both genetic and environmental sources of data [3]. A significant advantage of epigenomic analysis, particularly DNA methylation, (DNAm) is its ability to quantitate long-term environmental influences, both external (*e.g.*, tobacco smoke [4]) and internal (*e.g.*, inflammatory levels, *i.e.*, C-Reactive protein [5]). Dynamic change due to exogenous and endogenous factors over time enables powerful insights into pathophysiology [6] and, as these epigenomic data are not static, they can be gathered longitudinally at multiple time points [7]. Personalized epigenomic measurements can indicate the initiation and progression of disease. However, the analysis of the epigenome comes with significant complexities discussed in this Perspective that currently remain obstacles in the implementation and integration of epigenomics in precision health [8].

Here, we highlight the range of applications that genome-wide analysis of the DNA methylome can be successfully applied to in assessing human health and pathology (see Table 1 and Figure 1), with contemporary examples, as well as pointing towards its future transformative role in clinical medicine.

2. DNA methylation array analysis

The advent of robust DNAm arrays empowered the high-throughput measure of a consistent and functionally enriched portion of the DNA methylome [21]. The latest Illumina iterations include the EPIC v2 that interrogates ~930k CpGs [22], and more recently, a higher sample throughput Methylation Screening Array (MSA) with ~270k prior association evidence enriched CpGs (described in a recent preprint [23]). DNA modifications carry the advantage of their long-term stability in archived DNA [24], availing them to yield further insights from deeply phenotyped cohorts gathered over previous decades. Epigenome-wide association studies (EWAS) of cytosine DNAm pinpoint individual differentially methylated positions or cytosines (DMP/Cs) or regions (DMRs), allowing the identification and construction of DNAm-based disease predictors. Whilst the analysis of these sensitive DNAm arrays requires meticulous

CONTACT Christopher G. Bell c.bell@qmul.ac.uk William Harvey Research Institute, Barts & The London Faculty of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK

Article highlights

- Peripheral blood-derived DNA methylation has significant potential as a biomarker with clinical utility.
- Methylation Risk Scores by quantitating environmental exposures may complement Polygenic Risk Scores in common disease prediction.
- Long-term trends in plasma protein levels may be captured by EpiScore DNA methylation estimators.
- DNA methylation EpiSignatures can aid in pediatric monogenic disease diagnosis.
- The evolution of DNA methylation clocks now enables the prediction of aging-related morbidity and mortality.
- Cell-free DNA modification has immense potential in monitoring both oncogenic processes as well as organ pathology.

quality control [21], and appropriate statistical thresholds [25], replicable results have been identified [26].

These data enable inferences regarding cell-type heterogeneity, integration with chromatin data to explore how the levels of the epigenomic apparatus work together, and synergistically with genomic and transcriptomic information. Furthermore, this scale of data can be applied to machine learning (ML) methods for novel interrogation [6], for example, Elastic Net Regression has been highly successful in reducing epigenome-wide data into parsimonious sets for DNAm clock construction [18].

3. Epidemiological DNA methylation estimates of the exposome

The strong signal of tobacco smoking that can be seen in peripheral blood-derived DNAm indicates the potential power of exposome quantification, even in the prenatal environment [27]. Smoking-associated DNAm findings can be further dissected into myeloid and lymphoid cell-type specific signatures [28]. Meta-analysis in ~15k individuals has shown that a significant proportion of the peripheral blood DNA methylome is modified by smoking exposure, including ~64k CpGs (~7.5%) of the EPIC v1 array [4]. Clearly illustrating the disease predictive potential, one of the strongest individual CpG findings [4], residing in the *AHRR* gene, shows predictive power as a significant biomarker for the future onset of lung cancer [29]. Interestingly, personal and prenatal tobacco exposure in the early life period is even potentially able to be apportioned [30].

Many DNAm biomarkers of environmental pollutants are being explored to enable their quantification, including arsenic [31], lead [32], and ambient fine particulate matter (PM2.5) [33]. These DNAm estimations can be used to investigate the possible impact of prenatal pollutant and chemical exposures during critical developmental periods [34]. Additionally, the accurate biomarker ability that comes with the stability of stored DNAm empowers environmental and gene–environment interaction studies in existing biobanked samples [35]. With the increasing scale of cohort sample size, the quantification of the exposome may enable the identification of novel disease etiological factors.

4. DNA methylation biomarkers of common disease

Peripheral blood DNAm biomarkers of common diseases have been identified across the full range of organ systems [2]. These data include DNAm-associations for cardiovascular disease (CVD) phenotypes, such as ischemic stroke [36], and coronary heart disease [37]. A recent prospective EWAS for acute coronary syndrome biomarkers, in a Chinese cohort, identified 26 replicated DMCs including those that are in or near known CVD genes (*EHBP1L1*, *PRDM16*, *PRKCZ*) [38]. These DNAm data improved prediction models compared to both traditional risk factors and polygenic risk scores (PRS). In a European study, combining DNAm estimators for multiple CVD-related phenotypes, including BMI, blood pressure (BP), fasting glucose, insulin, cholesterol, triglycerides, and coagulation markers into a predictor (DNAmCVDscore) enhanced prediction of CVD events in the short term, compared to traditional risk factors or algorithms (SCORE2) [39]. Furthermore, utilizing both transcriptomic as well as DNAm peripheral blood data, via an ML analysis, was also shown to be advantageous over either omic dataset alone for coronary heart disease prediction [40]. Thus, there is considerable potential in DNAm biomarkers for CVD and heart failure for risk prediction, diagnosis and prognosis, and therapeutic monitoring [41].

The interrelating components of the metabolic syndrome generate intersecting CpG biomarkers, as large-scale EWAS meta-analyses for BMI have identified consistent findings that overlap with hyperglycemic, triglyceride,

Table 1. DNA methylation informative biomarkers. This includes MRS: methylation risk score; ERS: epigenetic risk scores.

DNAm Marker	Example Uses	Disease Setting	References
Methylation Risk Score (MRS)/Epigenetic Risk Scores (ERS)	Disease Prediction	Polygenic Disease – Environmental component	[9,10]
EpiScores; DNA methylation estimators; Epigenetic Biomarker Proxies	Plasma Protein Estimators; Smoking and other Environmental Factors. Biochemical, Metabolome, Protein Plasma Estimators	Polygenic Disease/Environmental Quantifiers	[11–13]
Cell Type Deconvolution	Inflammatory measures, infiltration. Senescent cell quantification	Pathological Diagnosis	[14,15]
EpiSignatures	Adjunct to Genetic Diagnostics	Monogenic Developmental Neurodevelopmental	[16]
DNA methylation Clocks	Chronological Age, Biological Age, Mitotic Rate Count	Polygenic Ageing Related Disease, Cancer; Assess Ageing related interventions	[17,18]
'liquid biopsy;' cell free DNA methylation (cfDNAm)	Tumour of Unknown Primary, Relapse Monitoring	Cancer; Organ Damage	[19]
Deep Learning Integrated Histological	Real-time Pathological Analysis	Neurological Cancers	[20]

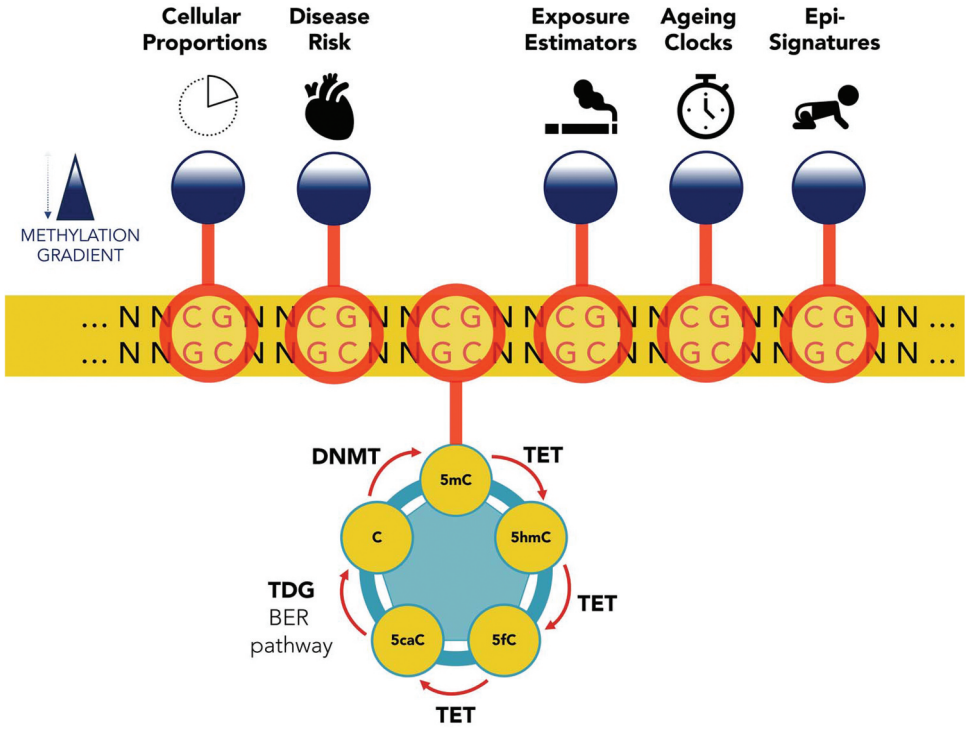

Figure 1. DNA methylation biomarkers. DNA methylation changes at specific CpG loci can be used as biomarkers for i) cell proportion deconvolution; ii) estimating disease risk via methylation risk scores (MRS); iii) environmental exposure estimators; iv) chronological and biological ageing clocks; v) diagnostic episignatures in genetic syndromes due to mutations in the epigenomic machinery genes. The DNA methylation cycle of active gain (via the DNMT enzyme; with DNMT3 being the *de novo* version) to 5mC; and loss through oxidation (via the TET enzyme) to 5hmC, 5fC and 5caC, followed by TDG excision, leaving an abasic site for the base excision repair (BER) pathway to install an unmodified Cytosine (C). Passive DNAm loss occurs through dilution when there is inaction of the maintenance DNMT (DNMT1) post replication.

and inflammatory EWAS results (*e.g.*, CpGs in/near *ABCG1*, *CPT1A*, *MYO5C*, *SREBF1*, *etc.*) [42]. In terms of Type 2 Diabetes (T2D), similar findings can be seen, as highlighted in a recent Native American study (*e.g.*, *ABCG1* and *SREBF1*) [43] and in Black and White participants from the Atherosclerosis Risk in Communities Study [44]. Furthermore, this blurring across these associated traits is also borne out in a T2D meta-EWAS in European cohorts with replication in a South-Asian cohort [45]. Of 64 replicated CpGs, only four remained significant after BMI correction and subsequently reduced to three after a further smoking adjustment (in/near *TXNIP*, *ABCG1*, and *CFL2*). However, quantification of these related components of the metabolic syndrome is still potentially informative for T2D precision medicine [46].

For other organ system diseases, identifying inflammatory blood-derived biomarkers may also lead to possible clinical utility. Respiratory diseases, including asthma, have been explored with pathognomonic increases in the proportion of eosinophils detected via DNAm [47] and these, along with immunoglobulin associations, enabled the construction of a DNAm-based biomarker [48].

Moreover, as large-scale DNAm array analyses become available, multiple common diseases can be queried simultaneously, *e.g.*, 19 common diseases in the ~18k Generation Scotland cohort [49]. This study identified 69 DMCs (58 novel) with the prevalence of four conditions (breast cancer, chronic kidney disease, ischemic heart disease, and T2D) and 64 DMCs (56 novel) with the incidence of two diseases (chronic obstructive pulmonary disease and T2D).

With the increasing power that comes as DNAm datasets expand in size, further epigenetic associations will be identified, as has been the case for Genome-wide association studies (GWAS).

Additional preliminary findings that may have prognostic utility in the future, if replicated, include *PTEN* DNAm in breast cancer aggressiveness [50] and recurrence of clear cell renal cell cancer via globally hypermethylated tumors [51]. Furthermore, whilst several diagnostic DNAm markers for colorectal cancer have been identified, with two approved, *i.e.*, plasma DNAm state of *SEPT9* [52] and the combination of both aberrant stool DNAm for *NDRG4* and *BMP3* with KRAS mutation status and fecal hemoglobin [53], some of these genes may also possess prognostic DNAm changes [54].

5. Methylation risk scores

The common genetic contribution to disease risk can be captured through polygenic risk scores (PRS) [55]. These scores have identified that individuals in the PRS high-risk tail can have a comparable disease risk to monogenic disease [56] – as well as modifying monogenic penetrance itself [57]. This conceptual use of GWAS results has been ported into EWAS with the construction of Methylation Risk Scores (MRS) or Methylation Profile Scores (MPS) from array-derived DNAm data [9]. MRS are generally constructed from the weighted sum of DNAm values for an individual of associated DMCs identified from trait or phenotype EWAS [58]. In an analysis

of electronic health records, MRS was able to impute more traits than PRS, as well as significantly improving this imputation for ~37% of laboratory tests [59]. A similarly named epigenetic risk score (ERS) was constructed from 144 alcohol-associated DMCs and revealed a significant association with hypertension phenotypes [10]. Thus, indicating utility for a risk measure that is acknowledged to be inaccurate in self-reported alcohol intake data.

In chronic kidney disease (CKD), an MRS was constructed in the US Hypertension Genetic Epidemiology Network (~6.9k individuals) and, in subsequent meta-analysis, was a significant predictor of prevalent CKD in both African American and non-African American individuals [60]. A cross-sectional pan-European blood-based multivariate MRS for mild cognitive impairment, independent of age and sex, was able to predict the prospective development of cognitive impairments in independent studies of Alzheimer's disease and Parkinson's disease [61].

Further refinement of these MRS/ERS is ongoing. This includes improvements through the addition of sub-threshold regional CpGs, by incorporating information on co-methylation patterns to inform the pruning of inter-correlated DNAm, as well as testing robustness across multi-ancestries [58].

DNAm predictors incorporating time-to-event data for T2D incidence over a 10-year risk period have been shown to augment known traditional risk factors derived from electronic health records in the powerful Generation Scotland cohort [62]. These studies are slowly building the evidence of the potential clinical utility of incorporating DNAm analysis in standard assessments.

6. Power of epigenomic cellular deconvolution

The epigenome is cell-type specific [63]; therefore, the precise data coming from DNAm arrays enable deconvolution of the proportion of major cell types present within heterogenous samples, *e.g.*, peripheral blood [14] or a tumor biopsy tissue [15]. This can be highly informative in EWAS, estimating cell proportion changes driven by pathological processes, such as inflammatory infiltration [64]. Deconvolution can be employed even within purified cell-types to delve further into cellular subsets, e.g., macrophage subclasses (M0, M1 & M2) [65]. Along with arrays, sequencing-based DNAm data can also be interrogated for deconvolution, including a recent Deep Learning (DL) model that analyses read-level DNAm patterns for tumor purity estimation (MethylBERT) [66].

Benchmarking of the methods used to perform these calculations has highlighted several factors that can affect accuracy [67]. These include the complexity of the reference, the method and number of markers chosen, and, critically, sequencing depth in sequencing-based tools. Innovative methods have been devised for the analysis of bulk tissues that lack reference DNAm information by incorporating tissue-specific single cell (sc) RNA-seq data [68].

7. Single-Cell DNA methylome analysis

The analysis of the DNA methylome at the single-cell (sc) level is now possible [69]. This is an advancing technology with successes so far with small sample numbers [70]. These data have also been ascertained in parallel in the same cell with transcriptomic data [71] as well as other various single-cell multi-omic combinations [72]. There are still significant limitations for DNA methylome analyses due to stochastic coverage that currently make accessibility analysis preferable to base-resolution DNAm. The degradation of DNA via the bisulfite reaction, technical complexity, as well as cost further contribute to this [73]. However, novel methods such as simultaneous sc-5mC and sc-5hmC via bisulfite-free chemical labeling reactions are also now achievable [74].

8. Plasma protein DNA methylation estimators (EpiScores)

Peripheral blood-derived DNAm estimators of circulating plasma protein levels have been constructed via protein measures derived from aptamer- and antibody-based platforms [11]. After accounting for known protein quantitative trait loci (pQTLs), these 'EpiScores' were able to explain between 1% and 58% of the variance of 109 protein measures. The additional power of these DNAm estimators resides in their ability to be potentially less influenced by short-term diurnal trends and provide more constant long-term data. In the context of cardiovascular disease prediction, a composite EpiScore based on 45 proteins was a significant predictor of CVD risk [75]. Additionally, DNAm estimators of C-Reactive Protein (CRP) levels have been found to be a stronger biomarker of chronic inflammation association than the direct serum measure itself [5], likely due to long-term trend capture. EpiScore CRP was also successful in identifying associations with cognitive aging [76]. DNAm estimators for the cardiovascular biomarkers plasma growth differentiation factor 15 (GDF15) and N-terminal pro-B-type natriuretic peptide (NT-proBNP) have shown robust associations with a range of traits, including T2D and poorer brain health [77]. This approach has been recently extended on through the generation of Epigenetic Biomarker Proxies (EBPs) in a MedRxiv preprint [13]. Clearly indicating their potential clinical relevance, of the 1,694 EBPs for biochemical, metabolomic, and proteomic levels that were identified, >62% had higher disease outcome odds or hazard ratios than their observed laboratory measures.

Blurring the line between MPS and EpiScores, metabolic EpiScores trained in the Generation Scotland cohort for BMI, body fat percentage, waist–hip ratio, and blood-based measures of glucose, high-density lipoprotein cholesterol, and total cholesterol also performed well in an ethnically distinct Singaporean dataset [78]. In addition, a recent alcohol consumption EpiScore was shown to be a good epidemiological tool, as significant associations with poorer global brain imaging metrics were observed that were not seen in self-reported data [79].

9. Blood-derived Episignatures aid the diagnosis of monogenic diseases

The diagnosis of pediatric monogenic disease has been considerably advanced via family trio exome sequencing – although there is clearly further potential in whole-genome sequencing (WGS) (*e.g.*, Genomics England [80]). Developmental disorders are enriched for mutations in the genes encoding the

Epigenomic Machinery: the readers, writers, and erasers of the Epigenome [81]. Thus, their abrogation contributes to conditions that make up the Mendelian Disorders of the Epigenetic Machinery (MDEM) [82]. As these genetic mutations lead to constitutive changes in the epigenome, they are observed across all tissue types, including blood, even if the central nervous system (CNS) is the major pathogenically affected organ. Due to the often inconclusive genetic findings for neurodevelopmental disorders, *e.g.*, variants of unknown significance within plausible pathogenic genes, the use of DNAm arrays has become a useful additional adjunct to exome and/or genome-sequencing in the diagnostic armamentarium [16]. Epigenetic signature profiles or 'Episignatures' improve variant interpretation and can reveal further phenotypic complexity [83].

Episignatures include MDEMs involving chromatin, or chromatinopathies, and at least 65 rare disorders exhibit a distinctive genome-wide DNAm profile [84]. Examples highlighting the utility of blood-derived episignatures include the diagnosis of a neurodevelopmental disorder due to *de novo* mutations within *MSL2*, which is a component of the H4K16 acetylator MSL complex [85]. Recently, a diagnostic blood-derived episignature classifier was also developed for the cluster of embryonic multiple malformations, which have no known cause, termed 'recurrent constellations of embryonic malformations' (RCEM) [86]. No previous diagnostic test was available. Another novel use beyond genetic syndromes has been the identification of an episignature of exposure to the antenatal teratogenic Valproic Acid [87]. This anti-epileptic is known to act as a histone deacetylase inhibitor. Quantification of exposure can enable the diagnosis of fetal valproate syndrome, as this can present with a clinical picture that overlaps several genetic syndromes.

Looking forward, there is significant diagnostic utility with nanopore sequencing [88]. For episignature detection, this has recently been evaluated, with the noted advantage of concurrent haplotyped genomic and epigenomic information [89].

10. DNA methylation clocks to capture biological ageing

The concept of 'biological' age attempts to capture an individual's deviation from the chronological age-specific population average for a biological marker or trait, or disease or mortality risk. Multi-omic methods of measuring 'biological' age are increasingly used to attempt to quantitate 'healthspan,' as well as biomarkers to test interventions that may slow 'biological' aging [90]. DNAm 'clocks' have been at the forefront of these aging biomarkers since the construction of the Hannum et al. [91] and Horvath clocks [92] in 2013, with the latter especially being a paradigm shift due to not only its accuracy across the life range but also pan-tissue ability. Subsequently, whilst these first generation DNAm clocks were designed to estimate chronological age, the recognition that their clock 'error' captured a component of 'biological' age as a risk of morbidity or mortality [93], led to the construction of further DNAm clocks focused on this disease prediction [94]. These phenotypic, or 2nd generation, DNAm 'clocks' were trained not only on age but also age-related biochemical measures (PhenoAge [95]) or plasma proteins and a smoking exposure estimator (GrimAge [96], GrimAge2 [97]). Taking advantage of a longitudinal epidemiological dataset from Dunedin, New Zealand, with assessments every 7 years, has enabled training of DNAm with the pace of aging, termed 3rd generation clocks, through the DunedinPoAm [98] and DunedinPACE measures [99]. A faster pace of this latter clock was recently described in preprint, with an increased likelihood of developing the Metabolic Syndrome over 7 years later [100].

Numerous additional specific DNAm clocks have subsequently been published, including a pediatric-specific analysis via buccal tissue (PedBE [101]), chronological accuracy for forensics [102–104], and focused on centenarians to aid validating claims of exceptional longevity [105]. A systematic archive of DNAm clock age-acceleration findings is being collated [106].

The cogs driving the consistent aging-related changes that facilitate DNAm clocks have revealed insights into the epigenomic machinery, and, how specific processes, if not adequately performed, lead to a stochastic loss or gain in DNAm. These random but directional alterations have been accurately termed a 'quazi-stochastic' mechanism [107]. This is because these aging changes are not random fluctuations throughout the entire DNA methylome. Instead, they occur in specific genomic compartments that are prone through cell replication to either increased DNAm (CpG dense promoters targeted by PRC2/bivalent domains) or decreased DNAm (sparse CpGs within partially DNA methylated domains – more frequently when surrounded by A or T bases) [108]. Furthermore, on top of these intrinsic (genomic) changes, extrinsic (multicellular) factors also significantly contribute [109]. Cell-type proportion shifts are a very strong component in the assumed environmental DNAm signal, and in aging this was recently estimated to be at least 35% of a DNAm clock's signal [109]. Even within the supposed pan-tissue signal, it is acknowledged that training bias as well as small changes in infiltrating NK cells and those between naive and activated T cells play a role [110]. Exploration of T-cell aging through a T-cell DNAm clock further confirmed that naïve T-cells appear young no matter the age of the organism [111], as opposed to effector memory/senescent T cells ($CD8^+CD28^-$). A DNAm clock was recently devised to minimize the influence of these immune cell changes (IntrinClock) [112].

The impact of leucocyte counts on DNAm clocks was further evaluated through Mendelian Randomisation (MR), revealing their causal effect on the age acceleration calculated with Hannum et al., PhenoAge and GrimAge clocks [113]. Another recent analysis in the Taiwan Biobank also investigated causality with respect to DNAm clock age acceleration [114]. By calculating a PRS for the 12-point Cardiovascular Heath Score and then employing this in an MR analysis, they identified a causal effect of the CVH score on GrimAge acceleration.

Clonal hematopoiesis becomes a readily detectable phenomenon in ~50% of those >80 years [115] and is a significant risk factor for disease, such as myeloproliferative neoplasia [116]. The expansion rate of clonal hematopoiesis is identified to be associated with accelerated Hannum *et al.*, PhenoAge and GrimAge clocks [117].

DNAm clocks have also been used as biomarkers of intrauterine lead exposure with effects in adolescence with first trimester levels positively associated with an increase in extrinsic epigenetic aging assessed via a modified Hannum *et al*. clock, where the blood cell count contributions are upweighted [118]. Additionally, over a three-year period, an exercise program was shown to slow phenotypic DNAm clocks (including PhenAge, GrimAge, GrimAge2 and DunedinPACE) [119]. Furthermore, in the same study Omega-3 was seen to have a slowing effect on these DNAm clocks, except GrimAge.

There is clear potential for specific disease or organ system bespoke 'clocks' to incorporate aging as well as specific aging-related disease DNAm estimators to improve their potential biomarker clinical utility [17]. As described in a recent preprint, this approach has been employed in the symphonyAge DNAm assessment, by compartmentalizing 11 different system effects [120]. Furthermore, there are clear benefits that will come as datasets expand. This is seen in the successful age prediction obtained via the increased training data size and implementation of an alternate approach to Elastic Net Regression, in the recent transformer-based deep neural network implemented in the CpG Pretrained Transformer (CpGPT), also in preprint [121].

At a functional level, the role that aging-related DNAm changes may play in influencing common disease-associated regulatory apparatus is an additional level to be explored [122]. This needs to be precisely interrogated in the pathogenic tissue relevant to the disease. One of the strongest and most consistent aging-related DNAm changes across tissue types is within the *ELOVL2* (ELOVL fatty acid elongase 2) locus [123]. Functional work has implicated these aging-related changes in a mouse model of retinal aging [124], as well as through the dysregulation of lipid metabolism in both mouse and human [125].

As cell-type proportional changes are a significant driver of aging-related changes in organ and tissue samples, several research groups have attempted to explore aging at the single cell level. However, as mentioned above, this resolution of analysis remains challenging, especially due to the sparsity and stochasticity of single-cell genome-wide data. Trapp *et al*. developed a statistical model called 'scAge' that was able to capture chronological age successfully across several cell-types, including hepatocytes, muscle and embryonic stem cells [126]. This work also identified selected heterogeneity in the aging progression of individual cells. By further integrating these data with single cell transcriptomics, they delineated co-regulated and stochastic CpG clusters [127]. A single-cell aging DNAm clock in blood and brain quantified that 39% and 12% of the aging signal, respectively, was due to proportional shifts in cell-types in these tissues [109]. Furthermore, analysis of neuronal and glial-specific DNAm clocks in Alzheimer's revealed that the latter had the strongest aging-related changes. This fits with the findings of cortical Alzheimer's disease EWAS that DNAm changes observed are predominately driven by the non-neuronal cells (*e.g*., microglia) [128].

11. Incorporating genetic risk with DNA methylation

Common disease traits are influenced by both inherited genetic factors as well as environmental influences. Therefore, accurate personalized prediction will require integrating these multifaceted measures together [3]. Incorporating genetic risk scores for disease, such as polygenic risk scores (PRS) with DNAm measures derived from the peripheral blood has revealed additive effects that vary drastically across differing traits. Obesity, assessed by BMI, benefits equally from both genetic and epigenetic predictors, with work from McCartney *et al*. estimating this to be 10.1% and 12.5%, respectively [129]. These values were predominately additive as the combined genetic and epigenetic BMI prediction was 19.7%. A strong DNAm BMI signature in blood can be found due to common overlapping signals from components of the metabolic syndrome, including hyperlipidemia, hyperglycemia, and inflammation [42]. Furthermore, illustrating the need to capture and incorporate non-genetic factors, recent work in the UK Biobank has shown that individuals with a high genetic BMI risk, calculated by PRS, were able to negate this and thus reduce their risk of obesity-related morbidities, if they were adherent to a healthy lifestyle [130]. By comparison, tobacco smoking's prediction is considerable, but almost exclusively via the exceeding strong peripheral blood DNAm signal (e.g., tobacco: 60.9% epigenetic, 2.8% genetic, 61.4% combined) [129].

12. Monitoring for disease

Multi-omic longitudinal monitoring has the potential to contribute significantly to precision medicine [7]. Intensive longitudinal monitoring for T2D was performed in one individual via peripheral DNA methylome and transcriptome assessment at ~6-week and ~3-week timepoints, respectively, over a 3-year period [131]. Hinting at the clinical possibilities, this revealed that DNAm changes were observed ~80–90 days prior to the first biochemical laboratory detectable glucose elevation. However, if these tests pass the threshold for clinical utility in the future, there are challenges that will need to be surmounted for translation into routine clinical practice, *i.e*., robust analysis, acceptable turn-around time, and cost [132].

Potential for cancer surveillance has also been documented. In cervical cancer screening, the addition of the cervical sample DNAm assessment of three genes (*DPP6*, *RALYL* and *GSX1*: the WID-qCIN test) in HPV^{+ve} women, along with HPV16/18 detection, resulted in better prediction and improved their triage compared to cytology alone [133].

Neurological decline is notoriously difficult to diagnose and predict. A cognitive DNAm EpiScore from peripheral blood was able to improve prediction by ~4% of the variance from PRS of cognitive functioning [134]. Thus, hinting that as study power improves these data may be able to augment and improve current testing methods.

Other examples of disease monitoring through the epigenome include the assessment of severe alcohol withdrawal syndrome, with DNAm peripheral blood changes in the *ZSCAN25* gene potentially having clinical utility, as well as for withdrawal-related seizures [135].

13. DNA methylation in Precision Pathological Diagnosis

Epigenetic changes are crucial hallmarks of cancer [136]. Additionally, epigenomic machinery genetic mutations are a common cancer driver. Furthermore, epigenetic monotherapies are approved in certain cancers [137], but these epigenomic modifiers also have possible utility in nonmalignant diseases as well, *e.g.*, cardiovascular disease [138].

There is significant potential for DNAm analysis to enhance the accuracy of cancer pathological diagnosis. With array-derived DNAm, a random forest classifier for CNS tumors showed improvements in diagnostic precision for ~12% of cases [139]. More recent work has attempted to explain the critical steps of this classifier in a human interpretable framework [140]. Additionally, a deep learning (DL) approach formulated a model that was then employed in the real-time assessment of the DNAm from rapid nanopore sequencing for neuropathology diagnostics [20]. A successful application involved Paediatric CNS tumor classification during surgery and delivered an immediate pathological report to the surgeon [20]. Also, via a naive Bayesian framework (MethyLYZR), sparse nanopore DNAm data were also able to diagnose CNS tumors [141]. Another study has taken the reverse approach and used DL to predict DNAm from histological slides [142]. Then, by incorporating the slide images, patient demographics as well as DL-predicted DNAm, the correct classification was possible, delineating CNS tumors into 10 major histopathological categories [142].

14. Cell-free methylated DNA 'liquid biopsy'

Cell-free DNA (cfDNA) analysis has enabled a paradigm shift in obstetrics with the ability to non-invasively diagnose the common gross chromosomal abnormalities, such as trisomies [143]. Additionally, in cancer, these 'liquid biopsies' of circulating tumor DNA (ctDNA) are being interrogated for diagnosis [144], but also for monitoring of minimal residual disease and relapse [145]. As well, assessing transplant rejection through analysis of donor-derived cfDNA shows promise [146]. Recent murine experimental evidence has indicated that Neutrophil-NET derived cfDNA may be highly informative in terms of risk of recurrent stroke [147].

These fragments of DNA within the circulating blood plasma derived from necrosed cells also retain their DNA modification state [148]. This empowers a tumor DNAm signature of the tissue of origin to be identified [19] – clearly useful in cancers of unknown primary. Additionally, these modifications may also enable the detection of a pre-disease state where therapeutic intervention would be actionable, *e.g.*, in preeclampsia maternal plasma derived cfDNA [149]. Experimentally, the real-time identification of sepsis-induced acute kidney injury has been shown through elevated kidney epithelial cell derived cfDNA, with its cell-type specific DNAm signature [150]. Also, predictive biomarker ability has been explored, including identifying a DNAm signature for brain metastasis development from lung cancer derived cfDNA cells [151]. The analysis of cfDNAm in cerebrospinal fluid (CSF) also shows significant promise in pediatric CNS tumors [152], as the biopsy of the brain brings clear risks and imaging struggles with tumor classification.

Additional information can be extracted from cfDNA, such as fragmentation profiles, which can be informative of epigenetic state, such as DNAm and nucleosome placement [153]. Furthermore, understanding how the DNAm status influences cfDNA shedding may enable increased sensitivity [154]. It is worth noting that cfDNAm is another large-scale data type where ML approaches are finding success [155].

15. Conclusion

Epigenomic DNAm data are already a well acknowledged and powerful diagnostic and predictive biomarker. From an unknown blood-derived DNA sample lying abandoned for decades at the bottom of a freezer, we can, for example, already accurately assess that individual's long-term inflammatory state through CRP estimators, their chronological and biological age, and how heavy a smoker they were at the time the sample was drawn.

With the increasing available study sizes – and further replication and meta-analysis that will be enabled from this – there is the potential to more accurately quantitate the plethora of environmental exposures (see preprint re the Generation Scotland data resource [156]). This may uncover unknown drivers of disease. These data may have further insights to bring as it is increasingly acknowledged the contributory role environmental exposures play in chronic inflammation and subsequently in cancer [157]. In this era of personalized genomic medicine, exploring the interaction of genetic variation with the environment will be vital.

Analytical improvements in the current DNAm assessment are ongoing. These include Bayesian approaches that incorporate probe correlations, genetic and cell heterogeneity effects [158]. As well, the imputation of individual CpG results from multiple correlated loci, e.g., through nearby CpGs co-methylation pattern. There is even the potential for these imputed association analyses to outperform the actual CpG result, as these more stable measures benefit from reducing individual erroneous outlier effects and can facilitate robust harmonization across array platforms (see preprint [159]).

Epigenomic analysis brings insight across the entire life course. It enables a molecular understanding of its critical role in cellular development and function, and in the definition of cell-type itself. In chronic diseases, aging biomarkers may be translated into clinically actionable insights [160]. Another important application is that of noninvasive surrogate DNAm biomarkers for inaccessible tissues, such as the visceral organs, brain, eye, *etc*.

16. Future perspective

Whilst here we have focused on DNAm, there will also very likely be significant uplift to come with the precise delineation of the rarer DNA modifications, such as hydroxymethylation (5hmC, see Figure 1), which standard bisulfite techniques are blind to. Techniques to analyze both 5mC and 5hmC within their haplotypic context will empower this

disentanglement (see preprint [161]). Further insights may also come from the additional oxidative derivative modifications of formylcytosine (5fC) [162] and carboxycytosine (5CaC) [163]. Along with the chromatin-related components of the epigenome – presenting distinctive strengths and weaknesses – analysis of these epigenomic mechanisms is becoming integral to the understanding of disease pathological processes. Both as diagnostic and predictive biomarkers and for precise molecular biopsies, they will become mainstays in clinical care.

Author contributions

Paraskevi Christofidou and Christopher G. Bell wrote and revised the manuscript,

Disclosure statement

The authors have no 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. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.No writing assistance was utilized in the production of this manuscript.

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

Funding

Paraskevi Christofidou is funded by Impetus Grants/Hevolution; Christopher G. Bell receives funding from Barts Charity, and Impetus Grants/Hevolution. The funders had no role in the preparation of the manuscript.

ORCID

Paraskevi Christofidou http://orcid.org/0000-0002-5809-6557
Christopher G. Bell http://orcid.org/0000-0003-4601-1242

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers

1. Fitzgerald RC, Antoniou AC, Fruk L, et al. The future of early cancer detection. Nat Med. 2022;28(4):666–677. doi: 10.1038/s41591-022-01746-x
2. Bell CG. Epigenomic insights into common human disease pathology. Cell Mol Life Sci. 2024;81(1):178. doi: 10.1007/s00018-024-05206-2
3. McCarthy M, Birney E. Personalized profiles for disease risk must capture all facets of health. Nature. 2021;597(7875):175–177. doi: 10.1038/d41586-021-02401-0
4. Hoang TT, Lee Y, McCartney DL, et al. Comprehensive evaluation of smoking exposures and their interactions on DNA methylation. EBioMedicine. 2024;100:104956. doi: 10.1016/j.ebiom.2023.104956
5. Hillary RF, Ng HK, McCartney DL, et al. Blood-based epigenome-wide analyses of chronic low-grade inflammation across diverse population cohorts. Cell Genom. 2024;4(5):100544. doi: 10.1016/j.xgen.2024.100544
•• **Illustrates the potential advantage of EpiScores in capturing long term trends.**
6. Yousefi PD, Suderman M, Langdon R, et al. DNA methylation-based predictors of health: applications and statistical considerations. Nat Rev Genet. 2022;23(6):369–383. doi: 10.1038/s41576-022-00465-w
•• **Details the clinical utility of DNAm episignature analysis for specific monogenic disease diagnoses.**
7. Schüssler-Fiorenza Rose SM, Contrepois K, Moneghetti KJ, et al. A longitudinal big data approach for precision health. Nat Med. 2019;25(5):792–804. doi: 10.1038/s41591-019-0414-6
8. Carter AC, Chang HY, Church G, et al. Challenges and recommendations for epigenomics in precision health. Nat Biotechnol. 2017;35(12):1128–1132. doi: 10.1038/nbt.4030
9. Nabais MF, Gadd DA, Hannon E, et al. An overview of DNA methylation-derived trait score methods and applications. Genome Biol. 2023;24(1):28. doi: 10.1186/s13059-023-02855-7
10. Bui H, Keshawarz A, Wang M, et al. Association analysis between an epigenetic alcohol risk score and blood pressure. Clin Epigenet. 2024;16(1):149. doi: 10.1186/s13148-024-01753-4
11. Gadd DA, Hillary RF, McCartney DL, et al. Epigenetic scores for the circulating proteome as tools for disease prediction. Elife. 2022;11:11. doi: 10.7554/eLife.71802
12. Bollepalli S, Korhonen T, Kaprio J, et al. EpiSmokEr: a robust classifier to determine smoking status from DNA methylation data. Epigenomics. 2019;11(13):1469–1486. doi: 10.2217/epi-2019-0206
13. Carreras-Gallo N, Chen Q, Balagué-Dobón L, et al. Leveraging DNA methylation to create epigenetic Biomarker proxies that inform clinical care: a new framework for precision medicine. medRxiv. 2024 Dec 06. p. 24318612.
14. Salas LA, Zhang Z, Koestler DC, et al. Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. Nat Commun. 2022;13(1):761. doi: 10.1038/s41467-021-27864-7
15. Chakravarthy A, Furness A, Joshi K, et al. Pan-cancer deconvolution of tumour composition using DNA methylation. Nat Commun. 2018;9(1):3220. doi: 10.1038/s41467-018-05570-1
16. Aref-Eshghi E, Kerkhof J, Pedro VP, et al. Evaluation of DNA methylation episignatures for diagnosis and phenotype correlations in 42 Mendelian neurodevelopmental disorders. Am J Hum Genet. 2020;106(3):356–370. doi: 10.1016/j.ajhg.2020.01.019
17. Bell CG, Lowe R, Adams PD, et al. DNA methylation aging clocks: challenges & recommendations. Genome Biol. 2019;20(1). doi: 10.1186/s13059-019-1824-y
18. Teschendorff AE, Horvath S. Epigenetic ageing clocks: statistical methods and emerging computational challenges. Nat Rev Genet. 2025;26(5):350–368. doi: 10.1038/s41576-024-00807-w
19. Moran S, Martinez-Cardus A, Sayols S, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol. 2016;17(10):1386–1395. doi: 10.1016/S1470-2045(16)30297-2
20. Vermeulen C, Pages-Gallego M, Kester L, et al. Ultra-fast deep-learned CNS tumour classification during surgery. Nature. 2023;622(7984):842–849. doi: 10.1038/s41586-023-06615-2
•• **Pioneers DNA methylation analysis enabling real time pathological diagnosis for intraoperative decision making.**
21. Teschendorff AE, Relton CL. Statistical and integrative system-level analysis of DNA methylation data. Nat Rev Genet. 2017;19(3):129–147. doi: 10.1038/nrg.2017.86
22. Noguera-Castells A, Garcia-Prieto CA, Alvarez-Errico D, et al. Validation of the new EPIC DNA methylation microarray (900K EPIC v2) for high-throughput profiling of the human DNA methylome. Epigenetics. 2023;18(1):2185742. doi: 10.1080/15592294.2023.2185742
23. Goldberg DC, Cloud C, Lee SM, et al. MSA: scalable DNA methylation screening BeadChip for high-throughput trait association studies. bioRxiv. 2024 May 17. p. 594606.
24. Gosselt HR, Griffioen PH, van Zelst BD, et al. Global DNA (hydroxy) methylation is stable over time under several storage conditions and temperatures. Epigenetics. 2021;16(1):45–53. doi: 10.1080/15592294.2020.1786318
25. Mansell G, Gorrie-Stone TJ, Bao Y, et al. Guidance for DNA methylation studies: statistical insights from the illumina EPIC array [journal article]. BMC Genomics. 2019;20(1):366. doi: 10.1186/s12864-019-5761-7
26. Xiong Z, Yang F, Li M, et al. EWAS open platform: integrated data, knowledge and toolkit for epigenome-wide association study. Nucleic Acids Res. 2022;50(D1):D1004–d1009. doi: 10.1093/nar/gkab972

27. Colwell ML, Townsel C, Petroff RL, et al. Epigenetics and the exposome: DNA methylation as a proxy for health impacts of prenatal environmental exposures. Exposome. 2023;3(1). doi: 10.1093/exposome/osad001
28. You C, Wu S, Zheng SC, et al. A cell-type deconvolution meta-analysis of whole blood EWAS reveals lineage-specific smoking-associated DNA methylation changes. Nat Commun. 2020;11(1):4779. doi: 10.1038/s41467-020-18618-y
29. Bojesen SE, Timpson N, Relton C, et al. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. Thorax. 2017;72(7):646–653. doi: 10.1136/thoraxjnl-2016-208789

•• **Illustrates future disease prediction for lung cancer from AHRR (cg05575921) blood-derived DNA methylation status.**

30. Richmond RC, Suderman M, Langdon R, et al. DNA methylation as a marker for prenatal smoke exposure in adults. Int J Epidemiol. 2018;47(4):1120–1130. doi: 10.1093/ije/dyy091
31. Domingo-Relloso A, Makhani K, Riffo-Campos AL, et al. Arsenic exposure, blood DNA methylation, and cardiovascular disease. Circ Res. 2022;131(2):e51–e69. doi: 10.1161/CIRCRESAHA.122.320991
32. Colicino E, Just A, Kioumourtzoglou MA, et al. Blood DNA methylation biomarkers of cumulative lead exposure in adults. J Expo Sci Environ Epidemiol. 2021;31(1):108–116. doi: 10.1038/s41370-019-0183-9
33. Wang C, Amini H, Xu Z, et al. Long-term exposure to ambient fine particulate components and leukocyte epigenome-wide DNA methylation in older men: the normative aging study. Environ Health. 2023;22(1):54. doi: 10.1186/s12940-023-01007-5
34. Bozack AK, Trasande L. Epigenetic rejuvenation: a journey backwards towards an epigenomic ground state. Epigenomics. 2024;17(1):1–9. doi: 10.1080/17501911.2024.2432851
35. Ladd-Acosta C, Fallin MD. DNA methylation signatures as biomarkers of prior environmental exposures. Curr Epidemiol Rep. 2019;6(1):1–13. doi: 10.1007/s40471-019-0178-z
36. Soriano-Tarraga C, Lazcano U, Giralt-Steinhauer E, et al. Identification of 20 novel loci associated with ischaemic stroke. Epigenome-Wide Assoc Study Epigenet. 2020;15(9):988–997. doi: 10.1080/15592294.2020.1746507
37. Agha G, Mendelson MM, Ward-Caviness CK, et al. Blood leukocyte DNA methylation predicts risk of future myocardial infarction and coronary heart disease. Circulation. 2019;140(8):645–657. doi: 10.1161/CIRCULATIONAHA.118.039357
38. Long P, Si J, Zhu Z, et al. Genome-wide DNA methylation profiling in blood reveals epigenetic signature of incident acute coronary syndrome. Nat Commun. 2024;15(1):7431. doi: 10.1038/s41467-024-51751-6
39. Cappozzo A, McCrory C, Robinson O, et al. A blood DNA methylation biomarker for predicting short-term risk of cardiovascular events. Clin Epigenet. 2022;14(1):121. doi: 10.1186/s13148-022-01341-4
40. Zhang X, Wang C, He D, et al. Identification of DNA methylation-regulated genes as potential biomarkers for coronary heart disease via machine learning in the Framingham heart study. Clin Epigenetics. 2022;14(1):122. doi: 10.1186/s13148-022-01343-2
41. Desiderio A, Pastorino M, Campitelli M, et al. DNA methylation in cardiovascular disease and heart failure: novel prediction models? Clin Epigenetics. 2024;16(1):115. doi: 10.1186/s13148-024-01722-x
42. Bell CG. The epigenomic analysis of human obesity. Obesity (Silver Spring). 2017;25(9):1471–1481. doi: 10.1002/oby.21909
43. Domingo-Relloso A, Gribble MO, Riffo-Campos AL, et al. Epigenetics of type 2 diabetes and diabetes-related outcomes in the strong heart study. Clin Epigenetics. 2022;14(1):177. doi: 10.1186/s13148-022-01392-7
44. Venkataraghavan S, Pankow JS, Boerwinkle E, et al. Epigenome-wide association study of incident type 2 diabetes in Black and White participants from the atherosclerosis risk in communities study. Diabetologia. 2025;68(4):815–834. doi: 10.1007/s00125-024-06352-9
45. Fraszczyk E, Spijkerman AMW, Zhang Y, et al. Epigenome-wide association study of incident type 2 diabetes: a meta-analysis of five prospective European cohorts. Diabetologia. 2022;65(5):763–776. doi: 10.1007/s00125-022-05652-2
46. Ling C, Bacos K, Rönn T. Epigenetics of type 2 diabetes mellitus and weight change — a tool for precision medicine? Nat Rev Endocrinol. 2022;18(7):433–448. doi: 10.1038/s41574-022-00671-w
47. Arathimos R, Suderman M, Sharp GC, et al. Epigenome-wide association study of asthma and wheeze in childhood and adolescence. Clin Epigenet. 2017;9(1):112. doi: 10.1186/s13148-017-0414-7
48. Van Asselt AJ, Beck JJ, Finnicum CT, et al. Epigenetic signatures of asthma: a comprehensive study of DNA methylation and clinical markers. Clin Epigenet. 2024;16(1):151. doi: 10.1186/s13148-024-01765-0
49. Hillary RF, McCartney DL, Smith HM, et al. Blood-based epigenome-wide analyses of 19 common disease states: a longitudinal, population-based linked cohort study of 18,413 Scottish individuals. PLOS Med. 2023;20(7):e1004247. doi: 10.1371/journal.pmed.1004247
50. Ramadan A, Hashim M, Abouzid A, et al. Clinical impact of PTEN methylation status as a prognostic marker for breast cancer. J Genet Eng Biotechnol. 2021;19(1):66. doi: 10.1186/s43141-021-00169-4
51. El Khoury LY, Fu S, Hlady RA, et al. Identification of DNA methylation signatures associated with poor outcome in lower-risk stage, size, grade and necrosis (SSIGN) score clear cell renal cell cancer. Clin Epigenet. 2021;13(1):12. doi: 10.1186/s13148-020-00998-z
52. Payne SR. From discovery to the clinic: the novel DNA methylation biomarker mSEPT9 for the detection of colorectal cancer in blood. Epigenomics. 2010;2(4):575–585. doi: 10.2217/epi.10.35
53. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med. 2014;370(14):1287–1297. doi: 10.1056/NEJMoa1311194
54. Müller D, Győrffy B. DNA methylation-based diagnostic, prognostic, and predictive biomarkers in colorectal cancer. Biochim Biophys Acta Rev Cancer. 2022;1877(3):188722. doi: 10.1016/j.bbcan.2022.188722
55. Kachuri L, Chatterjee N, Hirbo J, et al. Principles and methods for transferring polygenic risk scores across global populations. Nat Rev Genet. 2024;25(1):8–25. doi: 10.1038/s41576-023-00637-2
56. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219–1224. doi: 10.1038/s41588-018-0183-z
57. Fahed AC, Wang M, Homburger JR, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. Nat Commun. 2020;11(1):3635. doi: 10.1038/s41467-020-17374-3
58. Chen J, Gatev E, Everson T, et al. Pruning and thresholding approach for methylation risk scores in multi-ancestry populations. Epigenetics. 2023;18(1):2187172. doi: 10.1080/15592294.2023.2187172
59. Thompson M, Hill BL, Rakocz N, et al. Methylation risk scores are associated with a collection of phenotypes within electronic health record systems. npj Genom Med. 2022;7(1):50. doi: 10.1038/s41525-022-00320-1
60. Jones AC, Patki A, Srinivasasainagendra V, et al. A methylation risk score for chronic kidney disease: a HyperGEN study. Sci Rep. 2024;14(1):17757. doi: 10.1038/s41598-024-68470-z
61. Koetsier J, Cavill R, Reijnders R, et al. Blood-based multivariate methylation risk score for cognitive impairment and dementia. Alzheimers Dement. 2024;20(10):6682–6698. doi: 10.1002/alz.14061
62. Cheng Y, Gadd DA, Gieger C, et al. Development and validation of DNA methylation scores in two European cohorts augment 10-year risk prediction of type 2 diabetes. Nat Aging. 2023;3(4):450–458. doi: 10.1038/s43587-023-00391-4
63. Loyfer N, Magenheim J, Peretz A, et al. A DNA methylation atlas of normal human cell types. Nature. 2023;613(7943):355–364. doi: 10.1038/s41586-022-05580-6
64. 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. 2018;10(7):925–940. doi: 10.2217/epi-2018-0037
65. McErlean P, Bell CG, Hewitt RJ, et al. DNA methylome alterations are associated with airway macrophage differentiation and phenotype during lung fibrosis. Am J Respir Crit Care Med. 2021;204 (8):954–966. doi: 10.1164/rccm.202101-0004OC
66. Jeong Y, Gerhauser C, Sauter G, et al. MethylBERT enables read-level DNA methylation pattern identification and tumour deconvolution using a transformer-based model. Nat Commun. 2025;16(1):788. doi: 10.1038/s41467-025-55920-z
67. De Ridder K, Che H, Leroy K, et al. Benchmarking of methods for DNA methylome deconvolution. Nat Commun. 2024;15(1):4134. doi: 10.1038/s41467-024-48466-z
68. Teschendorff AE, Zhu T, Breeze CE, et al. EPISCORE: cell type deconvolution of bulk tissue DNA methylomes from single-cell RNA-Seq data. Genome Biol. 2020;21(1):221. doi: 10.1186/s13059-020-02126-9
69. Angermueller C, Lee HJ, Reik W, et al. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. Genome Biol. 2017;18(1):67. doi: 10.1186/s13059-017-1189-z
70. Nichols RV, Rylaarsdam LE, O'Connell BL, et al. Atlas-scale single-cell DNA methylation profiling with sciMetv3. Cell Genomics. 2025;5 (1):100726. doi: 10.1016/j.xgen.2024.100726
71. Angermueller C, Clark SJ, Lee HJ, et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods. 2016;13(3):229–232. doi: 10.1038/nmeth.3728
72. Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet. 2019;20(5):257–272. doi: 10.1038/s41576-019-0093-7
73. Vandereyken K, Sifrim A, Thienpont B, et al. Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet. 2023;24 (8):494–515. doi: 10.1038/s41576-023-00580-2
74. Bai D, Zhang X, Xiang H, et al. Simultaneous single-cell analysis of 5mC and 5hmC with SIMPLE-seq. Nat Biotechnol. 2025;43(1):85–96. doi: 10.1038/s41587-024-02148-9
75. Chybowska AD, Gadd DA, Cheng Y, et al. Epigenetic contributions to clinical risk prediction of cardiovascular disease. Circ Genom Precis Med. 2024;17(1):e004265. doi: 10.1161/CIRCGEN.123.004265
76. Conole ELS, Stevenson AJ, Munoz Maniega S, et al. DNA methylation and protein markers of chronic inflammation and their associations with brain and cognitive aging. Neurology. 2021;97(23): e2340–e2352. doi: 10.1212/WNL.0000000000012997
77. Gadd DA, Smith HM, Mullin D, et al. Dnam scores for serum GDF15 and NT-proBNP levels associate with a range of traits affecting the body and brain. Clin Epigenetics. 2024;16(1):124. doi: 10.1186/s13148-024-01734-7
78. Smith HM, Ng HK, Moodie JE, et al. DNA methylation-based predictors of metabolic traits in Scottish and Singaporean cohorts. Am J Hum Genet. 2025;112(1):106–115. doi: 10.1016/j.ajhg.2024.11.012
79. Bernabeu E, Chybowska AD, Kresovich JK, et al. Blood-based epigenome-wide association study and prediction of alcohol consumption. Clin Epigenetics. 2025;17(1):14. doi: 10.1186/s13148-025-01818-y
80. Smedley D, Smith KR, Martin A, et al. 100,000 genomes Pilot on rare-disease diagnosis in health care - preliminary report. N Engl J Med. 2021;385(20):1868–1880.
81. Bjornsson HT. The Mendelian disorders of the epigenetic machinery. Genome Res. 2015;25(10):1473–1481. doi: 10.1101/gr.190629.115
82. Fahrner JA, Bjornsson HT. Mendelian disorders of the epigenetic machinery: postnatal malleability and therapeutic prospects. Hum Mol Genet. 2019;28(R2):R254–R264. doi: 10.1093/hmg/ddz174
83. Levy MA, McConkey H, Kerkhof J, et al. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. HGG Adv. 2022;3(1):100075.
84. Trajkova S, Kerkhof J, Rossi Sebastiano M, et al. DNA methylation analysis in patients with neurodevelopmental disorders improves variant interpretation and reveals complexity. HGG Adv. 2024;5 (3):100309. doi: 10.1016/j.xhgg.2024.100309
85. Karayol R, Borroto MC, Haghshenas S, et al. MSL2 variants lead to a neurodevelopmental syndrome with lack of coordination, epilepsy, specific dysmorphisms, and a distinct episignature. Am J Hum Genet. 2024;111(7):1330–1351. doi: 10.1016/j.ajhg.2024.05.001
86. Haghshenas S, Karimi K, Stevenson RE, et al. Identification of a DNA methylation episignature for recurrent constellations of embryonic malformations. Am J Hum Genet. 2024;111(8):1643–1655. doi: 10.1016/j.ajhg.2024.07.005
87. Haghshenas S, Putoux A, Reilly J, et al. Discovery of DNA methylation signature in the peripheral blood of individuals with history of antenatal exposure to valproic acid. Genet Med. 2024;26 (10):101226. doi: 10.1016/j.gim.2024.101226
88. Liu T, Conesa A. Profiling the epigenome using long-read sequencing. Nat Genet. 2025;57(1):27–41. doi: 10.1038/s41588-024-02038-5
89. Geysens M, Huremagic B, Souche E, et al. Clinical evaluation of long-read sequencing-based episignature detection in developmental disorders. Genome Med. 2025;17(1):1. doi: 10.1186/s13073-024-01419-z
90. Moqri M, Herzog C, Poganik JR, et al. Validation of biomarkers of aging. Nat Med. 2024;30(2):360–372. doi: 10.1038/s41591-023-02784-9
91. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49(2):359–367. doi: 10.1016/j.molcel.2012.10.016
92. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115. doi: 10.1186/gb-2013-14-10-r115
93. Marioni RE, Shah S, McRae AF, et al. DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol. 2015;16(1):25. doi: 10.1186/s13059-015-0584-6
94. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018;19 (6):371–384. doi: 10.1038/s41576-018-0004-3
95. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY). 2018;10 (4):573–591. doi: 10.18632/aging.101414
96. Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY). 2019;11(2):303–327. doi: 10.18632/aging.101684
97. Lu AT, Binder AM, Zhang J, et al. DNA methylation GrimAge version 2. Aging (Albany NY). 2022;14(23):9484–9549. doi: 10.18632/aging.204434
98. Belsky DW, Caspi A, Arseneault L, et al. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. Elife. 2020;9. doi: 10.7554/eLife.54870
99. Belsky DW, Caspi A, Corcoran DL, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. Elife. 2022;11:11. doi: 10.7554/eLife.73420
100. Demuth I, Vetter VM, Homann J, et al. DunedinPACE predicts incident metabolic syndrome: cross-sectional and longitudinal data from the Berlin aging study II (BASE-II). medRxiv. 2024:2024.12.16.24319083.
101. McEwen LM, O'Donnell KJ, McGill MG, et al. The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. Proc Natl Acad Sci USA. 2020;117(38):23329–23335. doi: 10.1073/pnas.1820843116
102. Aliferi A, Sundaram S, Ballard D, et al. Combining current knowledge on DNA methylation-based age estimation towards the development of a superior forensic DNA intelligence tool. Forensic Sci Int Genet. 2022;57:102637. doi: 10.1016/j.fsigen.2021.102637
103. Ambroa-Conde A, Giron-Santamaria L, Mosquera-Miguel A, et al. Epigenetic age estimation in saliva and in buccal cells. Forensic Sci Int Genet. 2022;61:102770. doi: 10.1016/j.fsigen.2022.102770
104. Freire-Aradas A, Giron-Santamaria L, Mosquera-Miguel A, et al. A common epigenetic clock from childhood to old age. Forensic Sci Int Genet. 2022;60:102743. doi: 10.1016/j.fsigen.2022.102743
105. Dec E, Clement J, Cheng K, et al. Centenarian clocks: epigenetic clocks for validating claims of exceptional longevity. Geroscience. 2023;45(3):1817–1835. doi: 10.1007/s11357-023-00731-7
106. Chervova O, Panteleeva K, Chernysheva E, et al. Breaking new ground on human health and well-being with epigenetic clocks:

a systematic review and meta-analysis of epigenetic age acceleration associations. Ageing Res Rev. 2024;102:102552. doi: 10.1016/j.arr.2024.102552

107. Tong H, Dwaraka VB, Chen Q, et al. Quantifying the stochastic component of epigenetic aging. Nat Aging. 2024;4(6):886–901. doi: 10.1038/s43587-024-00600-8
- **Explorationof the mechanistic drivers of DNA methylation clock-related ageing changes.**

108. Bell CG. Quantifying stochasticity in the aging DNA methylome. Nat Aging. 2024;4(6):755–758. doi: 10.1038/s43587-024-00634-y
109. Tong H, Guo X, Jacques M, et al. Cell-type specific epigenetic clocks to quantify biological age at cell-type resolution. Aging (Albany NY). 2024;16(22):13452–13504. doi: 10.18632/aging.206184
110. Jonkman TH, Dekkers KF, Slieker RC, et al. Functional genomics analysis identifies T and NK cell activation as a driver of epigenetic clock progression. Genome Biol. 2022;23(1):24. doi: 10.1186/s13059-021-02585-8
- •• **Identifies the influence of T and NK cell activation in observed blood-derived DNA methylation ageing changes.**

111. Mi T, Soerens AG, Alli S, et al. Conserved epigenetic hallmarks of T cell aging during immunity and malignancy. Nat Aging. 2024;4(8):1053–1063. doi: 10.1038/s43587-024-00649-5
112. Tomusiak A, Floro A, Tiwari R, et al. Development of an epigenetic clock resistant to changes in immune cell composition. Commun Biol. 2024;7(1):934. doi: 10.1038/s42003-024-06609-4
113. Sun M, Yang H, Hu Y, et al. Differential white blood cell count and epigenetic clocks: a bidirectional Mendelian randomization study. Clin Epigenetics. 2024;16(1):118. doi: 10.1186/s13148-024-01717-8
114. Sung HL, Lin WY. Causal effects of cardiovascular health on five epigenetic clocks. Clin Epigenetics. 2024;16(1):134. doi: 10.1186/s13148-024-01752-5
115. Stacey SN, Zink F, Halldorsson GH, et al. Genetics and epidemiology of mutational barcode-defined clonal hematopoiesis. Nat Genet. 2023;55(12):2149–2159. doi: 10.1038/s41588-023-01555-z
116. Kar SP, Quiros PM, Gu M, et al. Genome-wide analyses of 200,453 individuals yield new insights into the causes and consequences of clonal hematopoiesis. Nat Genet. 2022;54(8):1155–1166. doi: 10.1038/s41588-022-01121-z
117. Mack TM, Raddatz MA, Pershad Y, et al. Epigenetic and proteomic signatures associate with clonal hematopoiesis expansion rate. Nat Aging. 2024;4(8):1043–1052. doi: 10.1038/s43587-024-00647-7
118. Halabicky OM, Tellez-Rojo MM, Goodrich JM, et al. Prenatal and childhood lead exposure is prospectively associated with biological markers of aging in adolescence. Sci Total Environ. 2024;913:169757. doi: 10.1016/j.scitotenv.2023.169757
119. Bischoff-Ferrari HA, Gangler S, Wieczorek M, et al. Individual and additive effects of vitamin D, omega-3 and exercise on DNA methylation clocks of biological aging in older adults from the DO-HEALTH trial. Nat Aging. 2025;5(3):376–385. doi: 10.1038/s43587-024-00793-y
120. Sehgal R, Markov Y, Qin C, et al. Systems age: a single blood methylation test to quantify aging heterogeneity across 11 physiological systems. bioRxiv. 2024:2023.07.13.548904.
121. de Lima Camillo LP, Sehgal R, Armstrong J, et al. CpGPT: a foundation Model for DNA methylation. bioRxiv. 2024:2024.10.24.619766.
122. Bell CG, Xia Y, Yuan W, et al. Novel regional age-associated DNA methylation changes within human common disease-associated loci. Genome Biol. 2016;17(1):193. doi: 10.1186/s13059-016-1051-8
123. Slieker RC, Relton CL, Gaunt TR, et al. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenet Chromatin. 2018;11(1):25. doi: 10.1186/s13072-018-0191-3
124. Chen D, Chao DL, Rocha L, et al. The lipid elongation enzyme ELOVL2 is a molecular regulator of aging in the retina. Aging Cell. 2020;19(2):e13100. doi: 10.1111/acel.13100
125. Li X, Wang J, Wang L, et al. Lipid metabolism dysfunction induced by age-dependent DNA methylation accelerates aging. Signal Transduct Target Ther. 2022;7(1):162. doi: 10.1038/s41392-022-00964-6
126. Trapp A, Kerepesi C, Gladyshev VN. Profiling epigenetic age in single cells. Nat Aging. 2021;1(12):1189–1201. doi: 10.1038/s43587-021-00134-3
127. Tarkhov AE, Lindstrom-Vautrin T, Zhang S, et al. Nature of epigenetic aging from a single-cell perspective. Nat Aging. 2024;4(6):854–870. doi: 10.1038/s43587-024-00616-0
128. Shireby G, Dempster EL, Policicchio S, et al. DNA methylation signatures of Alzheimer's disease neuropathology in the cortex are primarily driven by variation in non-neuronal cell-types. Nat Commun. 2022;13(1):5620. doi: 10.1038/s41467-022-33394-7
129. McCartney DL, Hillary RF, Stevenson AJ, et al. Epigenetic prediction of complex traits and death. Genome Biol. 2018;19(1):136. doi: 10.1186/s13059-018-1514-1
130. Kim MS, Shim I, Fahed AC, et al. Association of genetic risk, lifestyle, and their interaction with obesity and obesity-related morbidities. Cell Metab. 2024;36(7):1494–1503 e3. doi: 10.1016/j.cmet.2024.06.004
131. Chen R, Xia L, Tu K, et al. Longitudinal personal DNA methylome dynamics in a human with a chronic condition. Nat Med. 2018;24(12):1930–1939. doi: 10.1038/s41591-018-0237-x
132. Berdasco M, Esteller M. Clinical epigenetics: seizing opportunities for translation. Nat Rev Genet. 2019;20(2):109–127. doi: 10.1038/s41576-018-0074-2
133. Schreiberhuber L, Barrett JE, Wang J, et al. Cervical cancer screening using DNA methylation triage in a real-world population. Nat Med. 2024;30(8):2251–2257. doi: 10.1038/s41591-024-03014-6
134. McCartney DL, Hillary RF, Conole ELS, et al. Blood-based epigenome-wide analyses of cognitive abilities. Genome Biol. 2022;23(1):26. doi: 10.1186/s13059-021-02596-5
135. Andersen A, Milefchik E, Papworth E, et al. ZSCAN25 methylation predicts seizures and severe alcohol withdrawal syndrome. Epigenetics. 2024;19(1):2298057. doi: 10.1080/15592294.2023.2298057
136. Esteller M, Dawson MA, Kadoch C, et al. The epigenetic hallmarks of cancer. Cancer Discov. 2024;14(10):1783–1809. doi: 10.1158/2159-8290.CD-24-0296
137. Laisne M, Lupien M, Vallot C. Epigenomic heterogeneity as a source of tumour evolution. Nat Rev Cancer. 2024;25(1):7–26. doi: 10.1038/s41568-024-00757-9
138. Berjawi M, Bell CG. The therapeutic potential of epigenome-modifying drugs in Cardiometabolic disease. Curr Genet Med Rep. 2021;9(2):22–36. doi: 10.1007/s40142-021-00198-y
139. Capper D, Jones DTW, Sill M, et al. DNA methylation-based classification of central nervous system tumours. Nature. 2018;555(7697):469–474. doi: 10.1038/nature26000
- •• **DNA methylation derived data informing the pathological diagnosis of neurological cancers.**

140. Benfatto S, Sill M, Jones DTW, et al. Explainable artificial intelligence of DNA methylation-based brain tumor diagnostics. Nat Commun. 2025;16(1):1787. doi: 10.1038/s41467-025-57078-0
- • **This paper identifies a cfDNA methylation biomarker of pre-eclampia.**

141. Brandl B, Steiger M, Kubelt C, et al. Rapid brain tumor classification from sparse epigenomic data. Nat Med. 2025;31(3):840–848. doi: 10.1038/s41591-024-03435-3
142. Hoang DT, Shulman ED, Turakulov R, et al. Prediction of DNA methylation-based tumor types from histopathology in central nervous system tumors with deep learning. Nat Med. 2024;30(7):1952–1961. doi: 10.1038/s41591-024-02995-8
143. Norton ME, Phimister EG. Circulating cell-free DNA and Screening for trisomies. N Engl J Med. 2022;387(14):1322–1324. doi: 10.1056/NEJMcibr2209405
144. Cescon DW, Bratman SV, Chan SM, et al. Circulating tumor DNA and liquid biopsy in oncology. Nat Cancer. 2020;1(3):276–290. doi: 10.1038/s43018-020-0043-5
145. Widman AJ, Shah M, Frydendahl A, et al. Ultrasensitive plasma-based monitoring of tumor burden using machine-learning-guided signal enrichment. Nat Med. 2024;30(6):1655–1666. doi: 10.1038/s41591-024-03040-4
146. Aubert O, Ursule-Dufait C, Brousse R, et al. Cell-free DNA for the detection of kidney allograft rejection. Nat Med. 2024;30(8):2320–2327. doi: 10.1038/s41591-024-03087-3

147. Cao J, Roth S, Zhang S, et al. DNA-sensing inflammasomes cause recurrent atherosclerotic stroke. Nature. 2024;633(8029):433–441. doi: 10.1038/s41586-024-07803-4
148. Acton RJ, Bell CG. Cancer detection and tissue of origin determination with novel annotation and scoring of cell-free methylated DNA. AME Med J. 2017;2(8):110–110. doi: 10.21037/amj.2017.08.02
149. De Borre M, Che H, Yu Q, et al. Cell-free DNA methylome analysis for early preeclampsia prediction. Nat Med. 2023;29(9):2206–2215. doi: 10.1038/s41591-023-02510-5
150. You R, Quan X, Xia P, et al. A promising application of kidney-specific cell-free DNA methylation markers in real-time monitoring sepsis-induced acute kidney injury. Epigenetics. 2024;19(1):2408146. doi: 10.1080/15592294.2024.2408146
151. Zuccato JA, Mamatjan Y, Nassiri F, et al. Prediction of brain metastasis development with DNA methylation signatures. Nat Med. 2025;31(1):116–125. doi: 10.1038/s41591-024-03286-y
152. Cornelli L, Van Paemel R, Ferro dos Santos MR, et al. Diagnosis of pediatric central nervous system tumors using methylation profiling of cfDNA from cerebrospinal fluid. Clin Epigenet. 2024;16(1):87. doi: 10.1186/s13148-024-01696-w
153. An Y, Zhao X, Zhang Z, et al. DNA methylation analysis explores the molecular basis of plasma cell-free DNA fragmentation. Nat Commun. 2023;14(1):287. doi: 10.1038/s41467-023-35959-6
154. Pessei V, Macagno M, Mariella E, et al. DNA demethylation triggers cell free DNA release in colorectal cancer cells. Genome Med. 2024;16(1):118. doi: 10.1186/s13073-024-01386-5
155. Moser T, Kuhberger S, Lazzeri I, et al. Bridging biological cfDNA features and machine learning approaches. Trends Genet. 2023;39(4):285–307. doi: 10.1016/j.tig.2023.01.004
156. Walker RM, McCartney DL, Carr K, et al. Data resource profile: whole blood DNA methylation resource in generation Scotland (MeGS). medRxiv. 2024:2024.04.30.24306314.
157. Swanton C, Bernard E, Abbosh C, et al. Embracing cancer complexity: hallmarks of systemic disease. Cell. 2024;187(7):1589–1616. doi: 10.1016/j.cell.2024.02.009
158. Trejo Banos D, McCartney DL, Patxot M, et al. Bayesian reassessment of the epigenetic architecture of complex traits. Nat Commun. 2020;11(1):2865. doi: 10.1038/s41467-020-16520-1
159. Richmond A, Mur J, Harris SE, et al. Imputed DNA methylation outperforms measured loci associations with smoking and chronological age. bioRxiv. 2024:2024.09.05.611501. doi: 10.1101/2024.09.05.611501
160. Herzog CMS, Goeminne LJE, Poganik JR, et al. Challenges and recommendations for the translation of biomarkers of aging. Nat Aging. 2024;4:1372–1383. doi: 10.1038/s43587-024-00683-3
161. Puddu F, Johansson A, Modat A, et al. 5-methylcytosine and 5-hydroxymethylcytosine are synergistic biomarkers for early detection of colorectal cancer. bioRxiv. 2024:2024.10.30.621123.
162. Parasyraki E, Mallick M, Hatch V, et al. 5-formylcytosine is an activating epigenetic mark for RNA pol III during zygotic reprogramming. Cell. 2024;187(21):6088–6103 e18. doi: 10.1016/j.cell.2024.08.011
163. Zhao M, Zou G, Tang J, et al. Probe-labeled electrochemical approach for highly selective detection of 5-carboxycytosine in DNA. Anal Chim Acta. 2023;1273:341521. doi: 10.1016/j.aca.2023.341521