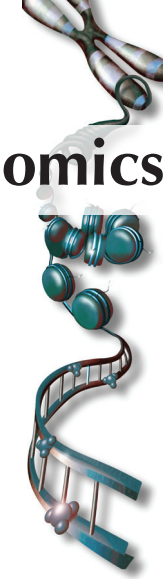
Review

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

Epigenomics



Epigenetic drift, epigenetic clocks and cancer risk

It is well-established that the DNA methylation landscape of normal cells undergoes a gradual modification with age, termed as 'epigenetic drift'. Here, we review the current state of knowledge of epigenetic drift and its potential role in cancer etiology. We propose a new terminology to help distinguish the different components of epigenetic drift, with the aim of clarifying the role of the epigenetic clock, mitotic clocks and active changes, which accumulate in response to environmental disease risk factors. We further highlight the growing evidence that epigenetic changes associated with cancer risk factors may play an important causal role in cancer development, and that monitoring these molecular changes in normal cells may offer novel risk prediction and disease prevention strategies.

First draft submitted: 4 January 2016; Accepted for publication: 17 February 2016; Published online: 22 April 2016

Keywords: aging • cancer • cancer risk • DNA methylation • epigenetic • epigenetic clock

Shijie C Zheng[‡,1,2], Martin Widschwendter[3] & Andrew E Teschendorff[*,‡,1,3,4]

[1]CAS Key Lab of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

[2]University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, China

[3]Department of Women's Cancer, University College London, 74 Huntley Street, London, WC1E 6AU, UK

[4]Statistical Cancer Genomics, UCL Cancer Institute, University College London, 72 Huntley Street, London, WC1E 6BT, UK

*Author for correspondence: a.teschendorff@ucl.ac.uk

‡Authors contributed equally

DNA methylation, a promising epigenetic cancer biomarker

Epigenetics can be defined as the study of mitotically heritable changes in gene regulation and cellular phenotype that cannot be explained by changes in DNA sequence [1]. Among the most important epigenetic modifications are those which affect DNA directly, mainly through covalent addition of a methyl ($-CH_3$) group at cytosines of CG dinucleotides (referred to commonly as 'CpGs'), although such DNA methylation (DNAm) can also occur in a non-CpG context [2]. Most CpGs in the human genome are methylated, occurring mainly in intergenic regions, repetitive elements and gene bodies. The unmethylated form occurs preferentially in the context of CpG islands (CGIs, regions of particular high CpG density), which colocalize with gene promoters [3]. CpG sites located in regions just outside CGIs (termed shores and shelves), or in distal regulatory elements, notably enhancers, exhibit the highest variability in DNAm [4,5]. DNAm can control gene expression, with promoter DNAm typically associated with gene silencing [3,6]. On the other hand, unmethylated gene promoters can associate with either active or inactive (poised) expression states, depending on the levels of nearby histone marks [7].

One of the most remarkable features of the DNAm landscape is that it gets reset during human embryogenesis, subsequently playing an essential role in development and tissue differentiation [6]. Specifically, DNAm in a differentiated cell of a given lineage is thought to play a critical role in irreversibly silencing genes that are not required for specification of that lineage [5]. It further plays a key role in determining enhancer function and transcription factor binding during development [5]. Once acquired, DNAm constitutes a metastable modification, which is maintained during cell division due to the action of DNA methyltransferase enzymes. However, the fidelity of the DNAm copying machinery is significantly lower than that of its DNA counterpart, which may result in

Future Medicine part of fsg

'epimutations' every time a cell divides [1]. The rate of such epimutations has been estimated to be as high as 10^{-5} per cytosine per cell division (cf. with a mutation rate of bases within CpG nucleotides of ~10^{-7} per cell division [8,9]), and may result in either loss of DNAm ('hypomethylation') at sites that are normally methylated, or in DNAm gains ('hypermethylation') at sites that are usually unmethylated [10].

Importantly, DNAm changes have been seen in a wide range of complex diseases, including cancer [1,11]. Specifically, two cancer hallmarks are hypermethylation of gene promoters, often affecting tumor suppressor genes and hypomethylation of intergenic regions [11]. Because of this, and because it constitutes a metastable, directly amplifiable, DNA-based mark (as opposed to, e.g., RNA 'snapshot' measurements, which are strongly time-dependent), DNAm offers great potential as a cancer biomarker [12–15]. Importantly, DNAm is also highly malleable, and has been shown to be influenced by many environmental exposures, including diet and levels of *in utero* nutrients [16–18]. Thus, DNAm represents not only an attractive biomarker for risk prediction and early detection of complex disease, but also offers to improve our understanding of the interface between environmental risk factors and disease phenotypes [1,18–19].

Epigenetic drift

Although age-associated DNAm changes affecting individual genes in normal tissue have long been observed [20,21], one of the first studies to explore this phenomenon beyond single genes was a 2005 study by Fraga *et al.* [22]. This study compared DNAm profiles of a number of monozygotic twins of different ages, and observed that while newborn twins exhibited effectively identical methylomes, adult twins showed divergent patterns with the level of divergence increasing with age. Due to the nonlongitudinal nature of this study, as well as limitations in sample size and genomic coverage, no apparent pattern of DNAm divergence was observed, with the loci exhibiting divergence within a twin pair not overlapping with the corresponding loci defined by another twin pair. Thus, the authors referred to the observed divergence in DNAm within twin pairs with the term 'epigenetic drift', to highlight the apparent stochastic nature of age-associated DNAm changes. However, in what follows we shall use a more general definition of epigenetic drift to encompass any type of age-associated DNAm change, be it of a stochastic nature or not.

Larger and higher genome-coverage studies subsequently confirmed that the DNAm landscape of normal cells changes substantially with age [23–27]. Several important novel insights were obtained from these studies. First, age-associated DNAm alterations do not happen randomly across the genome. For instance, it was observed that age-associated hypermethylation is more likely to happen at sites that carry bivalent [26] or PRC2 repressive marks [25], as defined in human embryonic or adult stem cells. By contrast, age-associated hypomethylation appears to target strong enhancers and active promoters [28]. Second, specific age-associated DNAm changes occur independently of tissue and cell-type, and this seems to be particularly true for age-associated hypermethylation, and specifically for the PRC2-enriched component [25,28–29]. For instance, an age-associated PRC2-marked 69 CpG DNAm signature derived in blood was shown to correlate with chronological age in other normal tissue types (e.g., lung and ovarian tissue) [25], and another age-related module enriched for PRC2 members was found to co-vary with age in brain and blood [30]. This cross-tissue independence not only demonstrates that a component of drift reflects an underlying universal mechanism, but also that drift in a complex tissue is not entirely the result of underlying alterations in cell type proportions. Indeed, this has been shown explicitly, as similar age-associated DNAm alterations are observed in different subsets of purified blood cells [26]. Third, age-associated DNAm alterations are also seen in adult stem cell populations, notably mesenchymal stem cells and hematopoietic stem/progenitor cells [25,31–33]. This supports the view that a component of epigenetic drift accrues in the underlying stem cell population of a given tissue, giving rise to the corresponding observed changes in the differentiated cells that make up the bulk of the tissue.

Horvath's epigenetic clock

The consistency and robustness of age-associated DNAm alterations across different tissue and cell types led to a number of studies to attempt predict the chronological age of an individual [34–37]. While some of these DNAm-based age predictors have been derived in specific tissues [34,35], Horvath derived a multitissue age predictor, which he then validated in a large number of independent datasets, encompassing in total over 8000 samples from over 50 different tissue and cell types [36]. This multitissue age predictor consists of 353 CpGs, and achieved a remarkable, clock-like accuracy on independent data with a median absolute deviation error of less than ±5 years. Although no direct comparison with other biological assays has yet been performed, it would appear that Horvath's 'epigenetic clock' may achieve substantially higher accuracies than those based on measuring telomere length or other molecular features such as T-cell DNA rearrangements [38–42]. Moreover, although studies

have found age-associated copy number [43,44], mutational [45] and gene expression [46] signatures, none of these have yet been developed into accurate age predictors, suggesting that these other types of molecular profiles may not be as relevant as DNAm for predicting chronological age.

DNAm-based age predictors containing far fewer than 353 marker CpGs have also been reported [34,35,37,42]. For instance, one study showed that DNAm values at only three CpG sites can accurately predict chronological age [42]. Another study indicated that 74% of the variation in chronological age can be predicted with as few as two CpG loci [35]. However, these age predictors have only been tested in specific tissues, or not extensively tested in other tissue types [36,47]. Another DNAm-based age predictor is the one developed by Hannum *et al.* [34], which trained its predictor on one of the largest whole blood datasets encompassing over 650 samples. While this 71 CpG age predictor also achieved high accuracy in independent blood datasets, recalibration of the predictor was necessary to achieve comparable accuracies in other tissue types [34].

There are a number of reasons why age predictors like that of Hannum *et al.* may not achieve accuracies comparable to those of Horvath's clock. First of all, none of the age predictors developed so far made explicit adjustment for age-associated changes in cell-type composition [48,49]. Since age-associated changes in tissue composition will vary from one tissue type to another, deriving an age predictor from one tissue type only, without correcting for changes in cell-type composition, will certainly 'bias' the predictor toward the tissue of origin. Because Horvath's epigenetic clock was trained on data from over 30 different tissue and cell types, this epigenetic clock is unlikely to have been confounded by tissue-specific age-associated changes in tissue composition. Second, an age predictor like that of Hannum *et al.* is likely to capture age-cumulative effects of specific endogenous and environmental factors, which are specific to blood-tissue and therefore not generalizable to other tissue types. In contrast, Horvath's clock, having been derived across so many different tissue types, is unlikely to have been confounded by these other tissue-specific effects. Third, tissue-specific DNAm levels could confound age predictors derived from one tissue type, when assessed in other tissue types, therefore, requiring recalibration [34]. A final reason could be that highly accurate quantification of chronological age from DNAm profiles may require a substantial number of CpGs, as is the case with Horvath's clock [36]. This would seem necessary if there is a substantial element of stochasticity underlying epigenetic drift [22]. Indeed, although it is clear that some genomic loci are more likely to undergo epigenetic drift than others, a more realistic picture is that of each CpG in the genome carrying an intrinsic probability of acquiring age-associated DNAm changes. Thus, the robustness of Horvath's clock stems in part from it measuring an aggregate level of absolute deviation in DNAm over a relatively large and specific set of 353 CpGs. In fact, the clock's accuracy only requires that a significant number of the 353 CpGs exhibit the expected DNAm deviations in a given sample, in order for the average deviation to then represent a meaningful number. Comparing two separate samples (e.g., different tissues from the same individual, or samples from identically aged individuals, e.g., twins), the specific subset of 353 CpGs that are altered in each sample may differ substantially. Thus, a highly accurate molecular clock is possible despite a level of underlying stochasticity, provided the clock is defined over a sufficient number of loci. Nevertheless, future high-coverage whole-genome bisulfite sequencing studies may pinpoint a few 'nonstochastic' loci, which, not unlike 'lighthouse beacons', keep track of chronological age with an accuracy comparable, or even exceeding that of Horvath's clock.

Significance & interpretation of the epigenetic clock

Although Horvath's epigenetic clock appears to provide, on average, a highly accurate measure of chronological age in seemingly healthy tissues, it is clear that for specific samples, large deviations/errors are also observed. This has led to the proposal that the deviation between DNAm age ('DNAm-age'), that is, the value predicted by the epigenetic clock, and chronological age may be informative of the true 'biological' age of a tissue [36]. Thus, the biological age of a tissue in an individual is not only a function of the person's chronological age, but also a function of other additional endogenous and exogenous factors, some of which may cause age acceleration, whereas others may cause age deceleration (see e.g., [19]).

A number of studies have explored whether biological age, specifically the difference between DNAm-age and chronological age, appears aggravated in tissues associated with disease phenotypes [50–54]. For instance, increased 'age acceleration', that is, tissues from individuals where DNAm-age is higher than the chronological age, has been observed in the human liver of obese individuals [50], in HIV-1-infected individuals [52] and in Down syndrome patients [53]. DNAm age in blood has also been shown to correlate with physical and cognitive fitness [54]. The significance of DNAm-age in the context of epithelial cancer is, however, less evident, since age acceleration is not observed across all cancer types [47,55]. The interpretability of DNAm-age

in affected tissues is also complicated due to potential confounding effects of the disease itself.

Other studies have, therefore, explored the possibility that Horvath's DNAm-age may be predictive of future disease risk. For instance, DNAm-age in blood has been found to be higher in men compared with women, a result which is consistent, in principle, with men's average lower longevity [34,36]. DNAm-age in the blood of postmenopausal women has been shown to correlate with the prospective risk of lung cancer [56]. A recent study also found DNAm-age in blood to be predictive of all-cause mortality [57]. Although not explicitly using DNAm-age, another study showed how DNAm of CpGs defining an age-associated DNAm signature in blood [25] always exhibited hypervariability in cervical normal cells, which 3 years later progressed to a high-grade cervical intraepithelial neoplasia [58]. Although all these results support the notion that DNAm-age could indicate disease risk for a number of different diseases, the reproducibility of these findings in independent cohorts still needs to be demonstrated, specially in those studies where progression of the DNAm changes could not be assessed.

A pressing unanswered question is the biological mechanism(s) underpinning the epigenetic clock. An initial attractive hypothesis would be that it constitutes a 'mitotic clock', measuring the number of cell divisions incurred by long-lived stem cells [59]. Under this model, incomplete maintenance of DNAm patterns by DNA methyltransferase enzymes (e.g. DNMT1) during DNA replication would lead to epimutations. This interpretation, however, cannot explain the ability of the clock to accurately predict chronological age across tissue types that differ so widely in their overall proliferation and turnover rates, including highly proliferative tissues such as colon and nonproliferative ones, such as brain [36]. Thus, while it is entirely plausible that components of other age-associated DNAm signatures may be mitotic in nature, as suggested by Beerman and Rossi [32] and Issa [59], the same does not appear to hold for Horvath's epigenetic clock. Instead, the epigenetic clock may reflect the indirect effects of the work performed by an epigenetic maintenance system, although at present, it is unclear what this epigenetic maintenance system may actually be [36].

Epigenetic drift & cancer risk factor DNAm signatures

Among risk factors for cancer, age is special, not only because it is the main risk factor for most cancer types, but because it indirectly captures the effects of age-associative cumulative exposure to exogenous and endogeneous risk factors. Thus, epigenetic drift may reflect molecular alterations caused by genetic and environmental risk factors. This in turn implies that if one wishes to study the effect a cancer risk factor may have on the DNA methylome, that adjustment for age is paramount. Over the last few years, many studies have explored the impact of major cancer risk factors on the DNA methylome of normal cells (Table 1). These studies include the effect of: HPV infection in normal cervical smears [58], smoking in blood and buccal tissue [60–65], *BRCA1* mutation in blood [66], sunlight (UV) exposure in skin [67,68], obesity in blood and adipose tissue [69–71], inflammation (inflammatory bowel disease) in colon tissue [72–74], alcohol intake in blood [75] and asbestos exposure in blood (Table 1) [23]. For other more specific chemical exposures, see [16].

Interestingly, if one focuses on the hypermethylated components of these cancer risk-factor DNAm signatures, one observes that PRC2/bivalently marked sites are often significantly enriched (Table 1). This is the case for age [25–27], HPV infection [58], smoking [62], obesity [69], *BRCA1* mutation [66] and inflammation [73]. The biological significance of this common PRC2 enrichment is, however, unclear. First of all, most of these enrichments have been established in relation to PRC2/bivalent marks, as determined in the human embryonic stem cell ground state, which is clearly not the most relevant one. Nevertheless, results have been shown to carry over to the corresponding PRC2/bivalent marks obtained in relevant adult stem cell populations, as for instance in the case of $CD133^+$ hematopoietic progenitor cells [25]. Second, most of the PRC2 targets represent transcription factors that are normally not expressed in the tissue of interest [81]. However, this does not exclude the possibility that a few key tissue-specific transcription factors are silenced through promoter hypermethylation. Therefore, it is entirely plausible that age-associated cumulative DNAm changes at PRC2 targets, which often encode key developmental and tissue-specific transcription factors, may result in deregulation of normal homeostasis, which is a key cancer hallmark [82–84].

A common cancer risk factor epigenetic signature?

The common enrichment for PRC2 sites among different cancer risk DNAm signatures also suggests that similar sites may be affected, irrespective of the risk factor. Although a comprehensive analysis of the overlap of such risk-factor DNAm signatures is still lacking, there are already some hints that signatures predictive for one risk factor may also be predictive for another. For instance, one study showed how an age-associated DNAm signature, involving PRC2 marked CpG sites that become hypermethylated with age, could discriminate cervical neoplasias from age-matched normal

Table 1. Table lists major cancer risk factors for which epigenome-wide association studies have been conducted in a number of tissue types.

Cancer risk factor	Normal tissue	Platform	PRC2/bivalent enrichment?	Ref.
Age	Blood (WB + purified)	27k and 450k	Y	[25,26]
	Colon	27k	Y	[76]
	Adipose	450k	Y	[69]
	Brain	27k	Y	[28,30,77]
	Kidney	27k	Y	[28]
	Muscle	27k	Y	[28]
	Buccal (saliva)	27k	Y	[29]
HPV	Cervix	27k	Y	[58]
BRCA1 mutation	Blood (WB)	27k	Y	[66]
Smoking	Blood (WB)	450k	Y	[60]
	Buccal	450k	Y	[62]
Obesity/BMI	Blood	CHARM and 450k	N	[70,71]
	Adipose	450k	Y	[69]
Alcohol	Blood (PBMCs)	27k	N	[75,78]
UV light	Skin	27k and 450k	N	[67,68]
EBV	Blood (B cell)	WGBS	Y	[79]
IBD	Intestine/colon	CGI agilent	Y	[80]†

We also indicate if a DNA methylation signature for the cancer risk factor was enriched for PRC2 or bivalently marked sites as determined in human embryonic stem cells. This enrichment is in the hypermethylated part of the signature since these PRC2/bivalent sites are normally unmethylated in the control samples. We also list some of the references where enrichment for PRC2/bivalent sites was Y, or those where findings were N.
†In mice.
450k/27k: Illumina Human Methylation 450k/27k beadchip; BMI: Body mass index; CGI: CpG island, EBV: Epstein–Barr virus; HPV: Human papilloma virus; IBD: Inflammatory bowl disease; N: Negative; PBMC: Peripheral blood mononuclear cells; WB: Whole blood; WGBS: Whole-genome bisulfite sequencing; UV: Ultraviolet, Y: Observed.

samples, suggesting that specific loci which undergo age-related DNAm changes do so also in response to HPV infection (the major risk factor for cervical cancer development) [25]. A more recent pan-cancer study compared a smoking-associated DNAm signature derived in buccal (epithelial) tissue with DNAm changes in cancer, and found that a DNAm-based smoking index (SMKI) constructed from this buccal DNAm signature, was highest in smoking-related lung cancers, but, surprisingly, also higher in every single cancer type compared with its respective normal tissue, including cancers which are not smoking-associated (e.g., endometrial cancer) [62]. That the SMKI is highest for smoking-associated lung cancer strongly supports the view that the smoking DNAm signature captures effects specific to the smoke carcinogens. On the other hand, that the SMKI is higher in every single cancer type compared with its corresponding normal tissue also suggests that a significant component of the smoking DNAm signature captures effects which are not specific to smoking. These additional nonspecific effects, therefore, suggest that other cancer risk factors (e.g., a high estrogen to progesterone ratio associated with obesity in the case of endometrial cancer [85]) may actively cause DNAm changes in normal tissue, which are similar to those induced by smoking. However, without an improved understanding of the biological mechanisms by which cancer risk factors may actively cause DNAm changes in normal cells, the existence of a common 'causal' cancer risk factor DNAm signature remains speculative.

A much more likely explanation as to why DNAm signatures associated with different cancer risk factors may overlap, or indeed why the smoking buccal DNAm signature reported in [62] is aggravated in all cancer types, is that these signatures contain a common 'mitotic clock' component (Figure 1), which would appear accelerated both in normal cells exposed to inflammation (e.g., buccal cells exposed to smoke carcinogens), as well as in highly proliferative cancer cells [59]. Interestingly, this 'mitotic clock' component also seems to be particularly well-defined at PRC2 sites, which are usually unmethylated in normal cells, but which would acquire stochastic hypermethylation

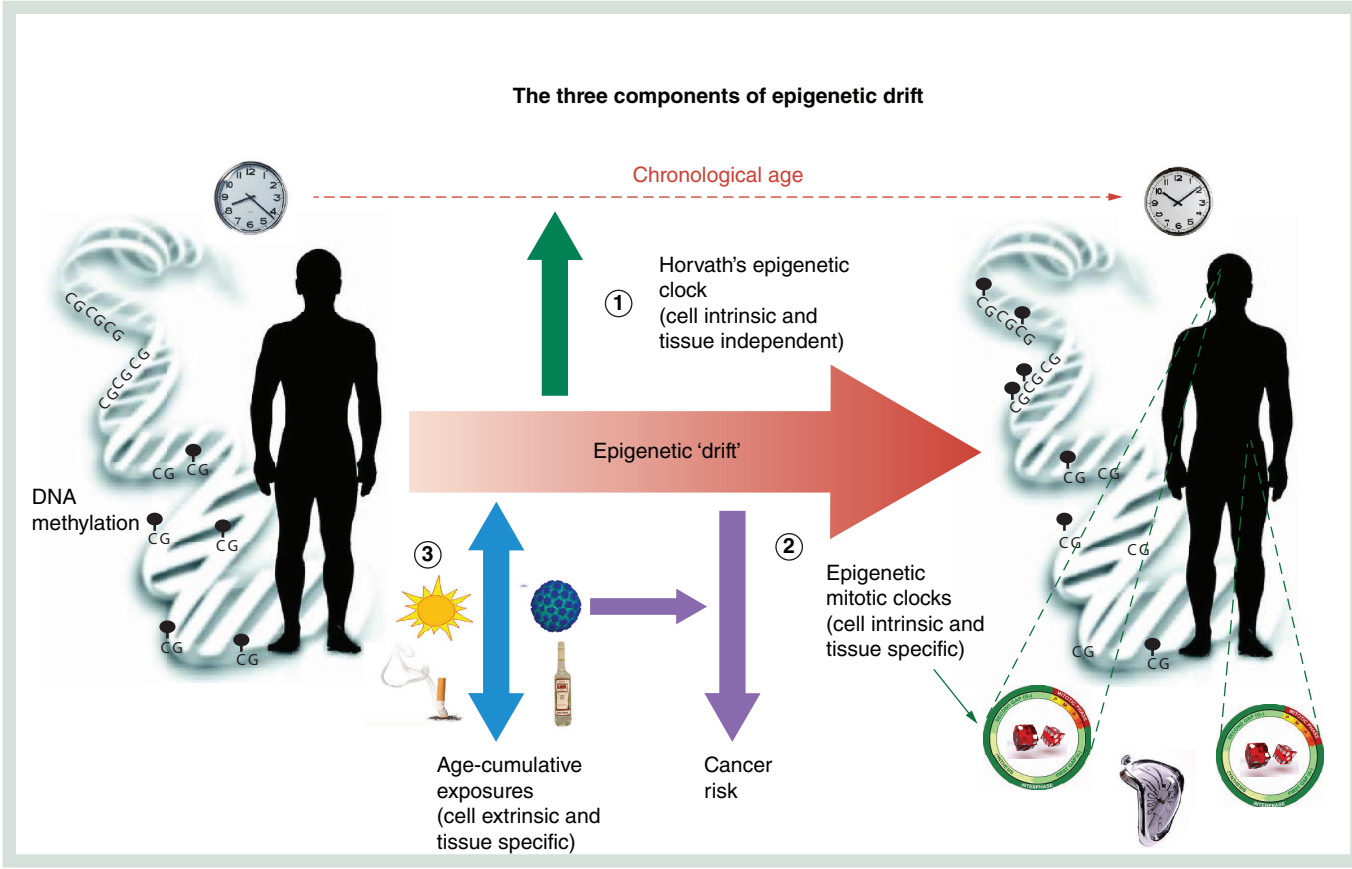


Figure 1. The three major components of epigenetic drift. We define epigenetic (DNAm) drift as any type of age-associated DNAm change. This epigenetic drift has a cell-intrinsic and tissue-type independent component, termed 'Horvath's epigenetic clock', which predicts chronological age with a remarkably high degree of accuracy. Another cell-intrinsic but tissue-type dependent component of DNAm drift is representing an 'epigenetic mitotic clock', which measures the cumulative number of cell divisions that the stem cell population of the tissue has undergone. The tick rate of this epigenetic mitotic clock may be influenced by endogenous (e.g., genetic risk factors) and exogenous (e.g., environmental risk factors) factors. These same cell-extrinsic factors may cause other types of active DNAm alterations, for instance, as seen for the AHRR gene in response to smoking or those that may be mediated by hormonal factors or viral infections.
DNAm: DNA methylation.

in a more proliferative state (Figure 2) [25,32,59,82]. This PRC2 enriched hypermethylation signature is also the one which is seen in aged stem cells and during hematopoietic ontogeny, further supporting a mitotic clock interpretation (Figure 2) [31–32,86–88].

DNAm-based predictors of cancer risk

Several studies have explored the possibility of using DNAm signatures in easily accessible nonepithelial tissues such as blood to predict the prospective risk of epithelial cancer. For instance, a number of studies have shown that the prospective risk of breast cancer can be predicted from whole blood DNAm profiles, yet the resulting AUCs or odd ratios are low (AUCs typically between 0.5 and 0.65), and therefore, only of marginal significance [66,89–91]. As mentioned earlier, DNAm-age in blood has also been shown to be predictive of lung cancer with a cancer incidence hazard ratio of 1.5 (p = 0.003) [56]. Focusing on a small set of inflammatory genes, which included *IL6*, *IFN*, *TLR2* and *ICAM1*, another recent study has shown how DNAm of these genes in blood could predict the prospective risk of prostate cancer with an incidence hazard ratio of approximately 1.5 [92]. Using four longitudinal cohorts, it has also been shown that DNAm-age in blood predicts all-cause mortality in later life, with age accelerations of 5 years or higher carrying a 16% increased mortality risk [57]. However, the influence of cancer-related mortality relative to mortality caused by other diseases in the context of DNAm-age needs to be more carefully assessed.

Other studies have shown how DNAm signatures derived in epithelial cells can predict the risk of neoplasia or invasive cancer. For instance, one study showed how DNAm variability in cytologically normal cervical smears, collected 3 years in advance of diagnosis, could

predict the prospective risk of a cervical intraepithelial neoplasia of grade 2 or higher, with an AUC of approximately 0.66 ($p < 0.05$) [58]. Another study showed how a smoking DNAm signature, as derived in buccal cells, and assessed in lung carcinomas *in situ* could predict the risk of progression to invasive lung cancer with an AUC of 0.88 [62]. While all of these results offer the exciting prospect of using DNAm in easily accessible

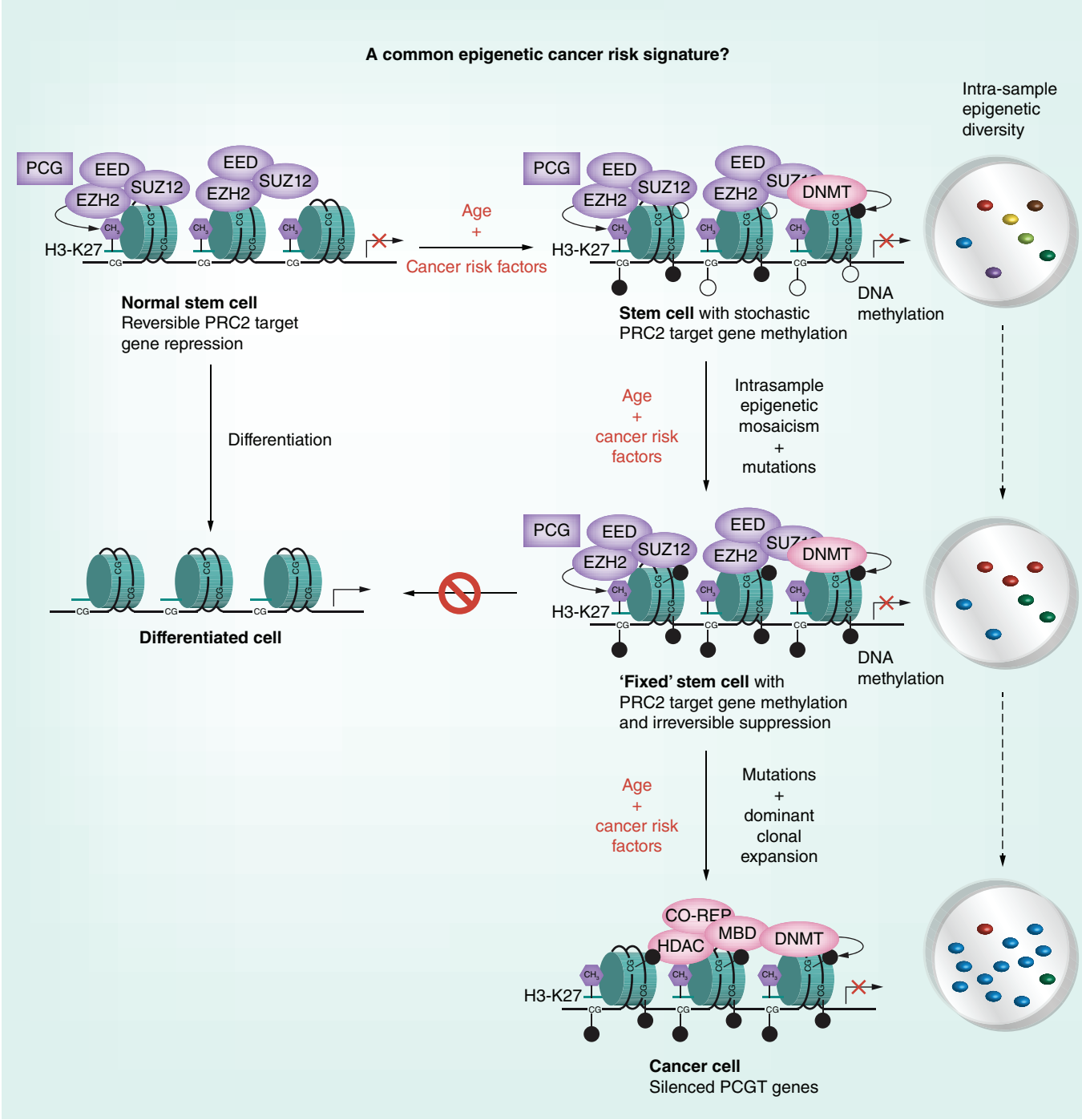


Figure 2. Epigenetic stem cell model of oncogenesis. Figure depicts how age-associated DNA methylation changes, which preferentially target genes marked by the PRC2 complex, accrue in underlying stem cells as a function of both age and exposure to cancer risk factors, leading at first to intrasample epigenetic nonclonal heterogeneity. Subsequent additional epigenetic and genetic changes can then give rise to epigenetic clonal mosaicism, from which subsequently a cancer clone can arise. The rate at which DNAm changes accrue at PRC2 targets will be determined by an epigenetic mitotic clock, with the rate influenced by endogenous and exogenous risk factors (see Figure 1).

tissues to predict cancer, larger studies will be needed to assess the potential for clinical application.

Epigenetic drift: is it functional & causal?

DNAm changes that correlate with age and other cancer risk factors and which can predict prospective cancer risk offer great biomarker potential. However, whether these associations are purely correlative or causative is still unclear. In the context of aging, age-associated DNAm changes may contribute to the well-known age-associated decline of stem cell function, and may underlie phenomena such as the well-known myeloid skewing of an aged hematopoietic system [93,94] and immunosenescence [95]. However, with the exception of a few genes, it has been extremely difficult to pinpoint age-associated alterations in gene expression or gene function, which are due to corresponding alterations at the DNAm level. Indeed, although there have been some reports of correlations between DNAm and mRNA expression in blood [34,96], the studies were either small [96] or did not use matched samples or did not correct for changes in blood cell-type composition [34]. Using two large but unmatched blood DNAm and mRNA expression datasets, and correcting for cellular heterogeneity, a recent study showed that one reason why there might not be a strong correlation is because age-associated DNAm changes appear to act by stabilizing pre-existing 'baseline' expression levels [97]. Thus, age-associated hypermethylation in blood is preferentially observed in promoters of genes that are normally not expressed in blood, and vice-versa, hypomethylation is observed in promoters of genes that are normally expressed [97]. Thus, if this result was to generalize to other tissue types, this suggests that most of the epigenetic drift is probably not functional. Given that epigenetic drift might already begin during embryogenesis [17,98] and that the rate of drift appears to be maximal before sexual maturity [99,100], one could speculate that most of the drift ought to be passive, as otherwise, it would bring forward the onset of complex diseases to coincide with the reproductive period, which would be highly undesirable and probably not selected for on evolutionary grounds [19]. Consistent with this, epigenetic drift also appears to target mostly peripheral nodes in protein interaction networks, avoiding essential and housekeeping genes, and specifically targeting transcription factors, most of which are not expressed in any given tissue type [101].

Thus, it is tantalizing to speculate that epigenetic drift generally does not affect gene function, but that it may occasionally 'hit' a key transcription factor, for instance, one that is critical for maintaining healthy homeostasis of a given tissue type, thus increasing cancer risk [94,102]. Recent work in the hematopoietic system supports this model, where DNAm alterations that are seen to accrue with age in blood and which may affect key lineage-specifying transcription factors, appear aggravated in myelodysplastic syndrome, with further increases in DNAm observed in acute myeloid leukemias [94]. Genetic mutations in key epigenetic regulators, which are seen as a function of age in pre-leukemic clonal expansions, and which can modify the normal DNAm landscape, provide further indirect support for such a model [103,104]. The WNT-signaling pathway, a key stem cell pathway, which is observed to undergo epigenetic deregulation with age [29,105–106], with DNAm-induced silencing of WNT-signaling antagonists potentially tipping the homeostatic balance toward increased self-renewal [106], provides further evidence for how drift could affect normal homeostasis.

Another more concrete example of how drift could increase cancer risk, and which may serve as a general paradigm for several other cancer types, is the one provided by *HAND2* in endometrial cancer [107]. Although the promoter of *HAND2* has not yet been conclusively shown to undergo hypermethylation with age in endometrial tissue, it does undergo hypermethylation with age in both blood [94,97] and in the colon of mice [24], suggesting that hypermethylation of its promoter with age is a tissue-wide phenomenon. This is significant, not only because age is a main risk factor for endometrial cancer, but because *HAND2* plays a critical role in mediating the tumor suppressive effects of progesterone, the main tumor suppressor pathway in this cancer type. In fact, a high body mass index, which is usually associated with a high estrogen to progesterone ratio, is the other main risk factor for endometrial cancer, and *HAND2* antagonizes the oncogenic activity of estrogen [107]. Given that promoter hypermethylation-induced silencing of *HAND2* is observed in normal tissue adjacent to complex atypical hyperplasias (CAH), in CAH itself, and in endometrial cancer [107], this supports the view that *HAND2* methylation is an early progressive event in endometrial carcinogenesis. Moreover, *HAND2* double knock-out mice develop CAH within weeks [107], suggesting that *HAND2* inactivation is a causal driver of endometrial cancer development. Thus, although *HAND2* promoter hypermethylation with age may be observed in many different tissue types, it is only the silencing in endometrial tissue that would be of particular functional consequence, increasing the risk of neoplastic transformation in that tissue.

Conclusion & future perspective

A deeper understanding of epigenetic drift, defined here as any age-associated DNAm alteration, will

require an improved characterization of its underlying components. As proposed here, epigenetic drift has at least three major components (Figure 1): a cell-intrinsic and tissue-type independent component, which allows highly accurate quantification of chronological age, as exemplified by Horvath's epigenetic clock; another cell-intrinsic, mitotic component whose tick rate will be tissue-dependent and influenced by endogenous (e.g., genetic risk factors) and exogenous (e.g., endocrine factors, inflammation) factors; and a cell-extrinsic, nonmitotic, active component, which will largely depend on tissue-type, and which also captures the age-cumulative effects of endogenous and exogenous risk factors.

Future studies will need to address a number of key questions to help shed further light on the nature of epigenetic drift and its components. Focusing on the epigenetic clock, it will be interesting to determine if other equally accurate clocks can be constructed and whether there is a minimum number of CpGs that are required to achieve such accuracy. Given that Horvath's clock is largely based on Illumina 27k probes, it would be interesting to see if more accurate clocks can be constructed based on the newer 450k/EPIC and eventually also whole genome bisulfite sequencing technologies. The degree of stochasticity in the epigenetic clock also needs to be assessed, for instance, by comparison of DNAm patterns at the 353 clock sites between different tissues (e.g., buccal and blood) from the same individual. Using longitudinal studies with multiple time points may also shed light on the temporal nature of DNAm changes at these specific sites [108]. The underlying biological mechanism underpinning Horvath's clock is another outstanding question, specially given that it does not seem to represent a mitotic clock. In particular, it will be of interest to explore if the 353 CpG sites making up Horvath's clock are located in special chromatin states, which are largely independent of tissue type.

Another key task for the future is the dissection of age-associated DNAm changes into those that are intrinsic to the aging process itself and those which reflect a cumulative exposure to environmental and lifestyle factors. To do this on human cohorts is an entirely nontrivial proposition, since properly adjusting for environmental and lifestyle factors is hard, specially given that we are still unaware of all factors that may impact on the epigenome and how this may vary across tissue types. Birth cohorts could help adjust for chronological age and thus help identify DNAm alterations, which are specific to environmental exposures [109]. Longitudinal studies profiling multiple tissues at multiple time points will be illuminating, in particular those based on twins [110]. An alternative approach to help identify DNAm alterations, which are intrinsic to aging, would be to perform studies on isogenic mice, where all mice are treated uniformly and kept under identical environmental conditions. It would be interesting to see if and how such changes vary according to tissue type, even within the same mice, and whether an analogous epigenetic clock for mice can be found. Similarly, the effect of a controlled environmental exposure on an isogenic mouse population could be studied to determine which factors accelerate or decelerate DNAm-age [111].

Of particular importance for cancer, is the component of epigenetic drift, which represents tissue-specific mitotic clocks, measuring the number of stem cell divisions in the tissue. As shown by Tomasetti and Vogelstein in the context of genetic mutations, such a mitotic clock may serve to predict the life-time risk of a given tissue-type turning cancerous [112,113]. Such a mitotic clock would accelerate in response to specific cancer risk factors, inflammation and endocrine factors being a few clear candidates [73–74,80,114–115], and thus may help explain the interindividual variation in cancer risk of a given tissue-type [15,59]. As argued earlier, PRC2 promoter sites may be particularly prone to acquisition of methylation marks during DNA replication, and consistent with this, hypermethylation of PRC2 sites appears to represent a universal DNAm signature of aging, preneoplastic lesions and cancer (Figure 2) [25]. Although an explicit link between an epigenetic mitotic clock and cancer risk still needs to be demonstrated, we have already seen that specific age and nonage related PRC2-enriched DNAm signatures in relevant epithelial cell types can predict the risk of certain cancers, including that of the cervix and lung. Therefore, we here propose that mitotic PRC2-enriched DNAm clocks, which correlate with the level of exposure to a generic cancer risk factor such as inflammation [73,74], may be particularly useful in the context of risk prediction or early detection (Figure 1 & 2). It follows from this that Horvath's epigenetic clock, which is not a mitotic clock, may not be that relevant for predicting the risk of a disease like cancer, which is universally characterized by an increased cell proliferation rate. Indeed, a highly optimized multitissue age predictor like Horvath's clock, which was trained over many tissue types with highly different mitotic rates, cannot be that informative of cancer risk, because it would not be able to capture the mitotic effects (e.g., inflammation, hormonal factors) that promote neoplastic transformation (Figure 1).

Apart from the silencing of key tissue-specific transcription factors, another related mechanism, which could link a mitotic PRC2-enriched DNAm clock to cancer risk, is through an increase in intrasample epi-

genetic mosaicism (Figure 2). Increased intrasample stochastic epigenetic variation could provide the very early seeds for carcinogenesis, facilitating the emergence of clonal expansions, which eventually lead to neoplastic transformation and cancer [73,116–121] (Figure 2). Indeed, recent studies have already shown that epigenetic clonal heterogeneity may play a key role in increasing the risk of neoplastic transformation [122] and in determining clinical outcome [123,124]. Thus, measures of intrasample clonal epigenetic heterogeneity may represent excellent general cancer biomarkers for early detection or risk prediction. It will, therefore, also be important to assess the relative level of stochastic versus nonstochastic epigenetic variation, which results from mitotic clocks operating at the stem cell level [10].

In relation to studies reporting DNAm signatures that predict the prospective risk of an epithelial cancer (including those which used the epigenetic clock), it is important to note that in most cases the predictions were obtained in a cell type (e.g., blood), which does not serve as the cell of origin for the cancer. Therefore, there is an urgent need to provide a mechanistic basis for these associations. One possibility is that DNAm changes in blood may represent an immune system defect, which could predispose individuals to the development of epithelial tumors like lung cancer [125]. Another possibility is that subtle alterations in blood tissue composition could be signaling an early response to the presence of preneoplastic cells in epithelial tissues. Such shifts in blood composition are certainly seen in patients with epithelial cancers [86,126], but how early in carcinogenesis such shifts might be detectable is currently unclear. Curiously, most cancer-risk predictive DNAm signatures in blood have not been consistently linked to immune-system related pathways [56,66,89–90], suggesting that there may be another basis for the observed associations. One appealing and exciting possibility is that DNAm changes associated with an exposure could be similar in different normal tissue types. That this may be so is supported by a study demonstrating that smoking induces similar DNAm alterations in buccal and blood tissue, although, as expected, the effect of smoking is far stronger in the cells directly exposed to the carcinogen (i.e., the buccals) [62]. For instance, the AHRR gene is similarly affected in both blood and buccal tissue [62], supporting the view that the observed hypomethylation at this gene locus represents an active response to the smoke carcinogen, a response which would be common across affected tissues. Likewise, age-associated DNAm signatures are generally valid across many different tissue types. HPV infection also seems to be associated with very similar DNAm changes in cervical and head and neck cancer [127]. It will, therefore, be extremely important to determine how DNAm changes associated with exposures vary according to tissue type. It will be equally important to assess which endogenous (e.g., genetic risk factors) and exogenous (exposure-related) factors cause similar and dissimilar DNAm changes in any given normal tissue type. For a given cancer type, these questions are key in order to then decide which easily accessible normal tissue might be suitable (if any) as a surrogate for developing risk prediction or early detection biomarkers.

The functional significance of epigenetic drift also needs further in-depth study. The lack of wide-spread *in-cis* associations between age-associated DNAm and mRNA expression changes does not mean that there might not be a more intricate *in-trans* association. Indeed, besides PRC2 members, binding sites of other key transcription factors like CTCF or those of the repressor NRSF/REST, become preferentially hypermethylated with age [97], strongly suggesting that specific regulatory networks, which support a certain 3D chromatin architecture, may become disrupted with age [128]. Thus, it will be specially interesting to investigate the patterns of epigenetic drift in relation to distal regulatory elements, including enhancers, as this may shed further light on how drift may impact on homeostasis in aged tissue, or to investigate the patterns of epigenetic drift in a multilayer setting, which also includes all major histone marks [129].

Finally, it will be important to see if there are other examples like *HAND2* in endometrial cancer. As we have seen, this gene undergoes age-associated hypermethylation in normal tissue, and inactivation appears to be also a causal driver of early endometrial cancer development. The example of *HAND2* is particularly enlightening, because *HAND2* methylation in endometrial tissue links together the two main epidemiological risk factors for endometrial cancer: age and obesity. We propose that this example may also serve as a more general paradigm linking age-associated DNAm-induced alterations in transcription factor activity to a modulation of the response to an exogenous cancer risk factor (in this case high estrogen levels), and therefore, to an increased cancer risk.

Financial & competing interests disclosure

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.

Executive summary

- DNA methylation (DNAm) drift in normal tissue reflects cell intrinsic and extrinsic mechanisms, some of which are tissue-independent, whereas others are tissue specific.
- The epigenetic clock describes cell-intrinsic age-associated DNAm alterations, which are tissue and cell-type independent, and which allows highly accurate prediction of the chronological age.
- Most of the epigenetic drift is nonfunctional, yet some of the drift may eventually affect the expression or binding affinity of transcription factors that are required for normal homeostasis.
- Epigenetic PRC2-enriched mitotic clock(s), which measure the cumulative rate of stem cell divisions in a tissue, and whose clock-rate may be affected by endogenous and exogenous factors, are of likely relevance for cancer-risk prediction.
- DNAm signatures associated with cancer risk factors and derived in easily accessible tissues such as blood and buccal tissue have been correlated with the prospective risk of epithelial neoplasia and invasive cancers.

References

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

1 Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* 465(7299), 721–727 (2010).

2 Lister R, Pelizzola M, Dowen RH *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271), 315–322 (2009).

3 Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev.* 25(10), 1010–1022 (2011).

4 Irizarry RA, Ladd-Acosta C, Wen B *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41(2), 178–186 (2009).

5 Ziller MJ, Gu HC, Muller F *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500(7463), 477–481 (2013).

6 Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447(7143), 425–432 (2007).

7 Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9(3), 215–216 (2012).

8 Campbell CD, Chong JX, Malig M *et al.* Estimating the human mutation rate using autozygosity in a founder population. *Nat. Genet.* 44(11), 1277–1281 (2012).

9 Horsthemke B. Epimutations in human disease. *Curr. Top. Microbiol. Immunol.* 310 45–59 (2006).

10 Landan G, Cohen NM, Mukamel Z *et al.* Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.* 44(11), 1207–1214 (2012).

11 Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* 7(1), 21–33 (2006).

12 Beck S, Rakyan VK. The methylome: approaches for global DNA methylation profiling. *Trends Genet.* 24(5), 231–237 (2008).

13 Beck S. Taking the measure of the methylome. *Nat. Biotechnol.* 28(10), 1026–1028 (2010).

14 Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* 12(8), 529–541 (2011).

15 Widschwendter M, Jones A, Teschendorff AE. Epigenetics makes its mark on women-specific cancers – an opportunity to redefine oncological approaches? *Gynecol. Oncol.* 128(1), 134–143 (2013).

16 Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* 13(2), 97–109 (2011).

17 Heijmans BT, Tobi EW, Stein AD *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl Acad. Sci. USA* 105(44), 17046–17049 (2008).

18 Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.* 8(4), 253–262 (2007).

19 Teschendorff AE, West J, Beck S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum. Mol. Genet.* 22(R1), R7–R15 (2013).

20 Ahuja N, Issa JP. Aging, methylation and cancer. *Histol. Histopathol.* 15(3), 835–842 (2000).

21 Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. Aging and DNA methylation in colorectal mucosa and cancer. *Cancer Res.* 58(23), 5489–5494 (1998).

22 Fraga MF, Ballestar E, Paz MF *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* 102(30), 10604–10609 (2005).

• **One of the first papers to demonstrate the phenomenon of epigenetic drift on a genome-wide scale.**

23 Christensen BC, Houseman EA, Marsit CJ *et al.* Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet.* 5(8), e1000602 (2009).

24 Maegawa S, Hinkal G, Kim HS *et al.* Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res.* 20(3), 332–340 (2010).

25 Teschendorff AE, Menon U, Gentry-Maharaj A *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* 20(4), 440–446 (2010).

• **One of the first papers to show that age-associated DNA methylation signatures are independent of tissue type, present in stem cell populations, and which become aggravated in cancer.**

26 Rakyan VK, Down TA, Maslau S *et al.* Human aging-associated DNA hypermethylation occurs preferentially at

bivalent chromatin domains. *Genome Res.* 20(4), 434–439 (2010).

27 Heyn H, Li N, Ferreira HJ *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc. Natl Acad. Sci. USA* 109(26), 10522–10527 (2012).

28 Day K, Waite LL, Thalacker-Mercer A *et al.* Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biol.* 14(9), R102 (2013).

29 West J, Beck S, Wang X, Teschendorff AE. An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways. *Sci. Rep.* 3, 1630 (2013).

30 Horvath S, Zhang Y, Langfelder P *et al.* Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* 13(10), R97 (2012).

31 Bork S, Pfister S, Witt H *et al.* DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. *Aging Cell* 9(1), 54–63 (2010).

32 Beerman I, Bock C, Garrison BS *et al.* Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. *Cell Stem Cell* 12(4), 413–425 (2013).

•• **Very important paper demonstrating DNA methylation alterations during hematopoietic ontogeny.**

33 Bocker MT, Hellwig I, Breiling A, Eckstein V, Ho AD, Lyko F. Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging. *Blood* 117(19), e182–e189 (2011).

34 Hannum G, Guinney J, Zhao L *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* 49(2), 359–367 (2013).

35 Bocklandt S, Lin W, Sehl ME *et al.* Epigenetic predictor of age. *PLoS ONE* 6(6), e14821 (2011).

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

•• **Landmark paper demonstrating that chronological age can be predicted with high accuracy across many independent tissue types.**

37 Koch CM, Wagner W. Epigenetic-aging-signature to determine age in different tissues. *Aging* 3(10), 1018–1027 (2011).

38 Aubert G, Lansdorp PM. Telomeres and aging. *Physiol. Rev.* 88(2), 557–579 (2008).

39 Meissner C, Ritz-Timme S. Molecular pathology and age estimation. *Forensic Sci. Int.* 203(1–3), 34–43 (2010).

40 Zubakov D, Liu F, Van Zelm MC *et al.* Estimating human age from T-cell DNA rearrangements. *Curr. Biol.* 20(22), R970–R971 (2010).

41 Lin J, Epel E, Blackburn E. Telomeres and lifestyle factors: roles in cellular aging. *Mutat. Res.* 730(1–2), 85–89 (2012).

42 Weidner CI, Lin Q, Koch CM *et al.* Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* 15(2), R24 (2014).

43 Laurie CC, Laurie CA, Rice K *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* 44(6), 642–650 (2012).

44 Jacobs KB, Yeager M, Zhou W *et al.* Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* 44(6), 651–658 (2012).

45 Alexandrov LB, Jones PH, Wedge DC *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* 47(12), 1402–1407 (2015).

46 De Magalhaes JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics* 25(7), 875–881 (2009).

47 Lin Q, Wagner W. Epigenetic aging signatures are coherently modified in cancer. *PLoS Genet.* 11(6), e1005334 (2015).

48 Teschendorff AE, Menon U, Gentry-Maharaj A *et al.* An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS ONE* 4(12), e8274 (2009).

49 Houseman EA, Accomando WP, Koestler DC *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13, 86 (2012).

50 Horvath S, Erhart W, Brosch M *et al.* Obesity accelerates epigenetic aging of human liver. *Proc. Natl Acad. Sci. USA* 111(43), 15538–15543 (2014).

51 Rickabaugh TM, Baxter RM, Sehl M *et al.* Acceleration of age-associated methylation patterns in HIV-1-infected adults. *PLoS ONE* 10(3), e0119201 (2015).

52 Horvath S, Levine AJ. HIV-1 infection accelerates age according to the epigenetic clock. *J. Infect. Dis.* 212(10), 1563–1573 (2015).

53 Horvath S, Garagnani P, Bacalini MG *et al.* Accelerated epigenetic aging in Down syndrome. *Aging Cell* 14(3), 491–495 (2015).

54 Marioni RE, Shah S, Mcrae AF *et al.* The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int. J. Epidemiol.* 44(4), 1388–1396 (2015).

55 Horvath S. Erratum to: DNA methylation age of human tissues and cell types. *Genome Biol.* 16, 96 (2015).

56 Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging* 7(9), 690–700 (2015).

57 Marioni RE, Shah S, Mcrae AF *et al.* DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* 16(1), 25 (2015).

• **Important paper demonstrating that Horvath's DNA methylation age measured in blood is associated with clinical phenotypes in large numbers of samples.**

58 Teschendorff AE, Jones A, Fiegl H *et al.* Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. *Genome Med.* 4(3), 24 (2012).

• **First paper to show that the prospective risk of an epithelial cancer can be predicted based on a DNA methylation signature in normal epithelial cells.**

59 Issa JP. Aging and epigenetic drift: a vicious cycle. *J. Clin. Invest.* 124(1), 24–29 (2014).

60 Zeilinger S, Kuhnel B, Klopp N *et al.* Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS ONE* 8(5), e63812 (2013).

61 Shenker NS, Ueland PM, Polidoro S *et al.* DNA methylation as a long-term biomarker of exposure to tobacco smoke. *Epidemiology* 24(5), 712–716 (2013).

62 Teschendorff AE, Yang Z, Wong A *et al.* Correlation of smoking-associated DNA methylation changes in buccal cells with DNA methylation changes in epithelial cancer. *JAMA Oncol.* 1(4), 476–485 (2015).

63 Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin. Epigenetics* 7, 113 (2015).

64 Zhang Y, Schottker B, Florath I *et al.* Smoking-associated DNA methylation biomarkers and their predictive value for all-cause and cardiovascular mortality. *Environ. Health Perspect.* 124(1), 67–74 (2015).

65 Zhang Y, Schottker B, Ordonez-Mena J *et al.* F2RL3 methylation, lung cancer incidence and mortality. *Int. J. Cancer* 137(7), 1739–1748 (2015).

66 Anjum S, Fourkala EO, Zikan M *et al.* A BRCA1-mutation associated DNA methylation signature in blood cells predicts sporadic breast cancer incidence and survival. *Genome Med.* 6(6), 47 (2014).

67 Gronniger E, Weber B, Heil O *et al.* Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genet.* 6(5), e1000971 (2010).

68 Vandiver AR, Irizarry RA, Hansen KD *et al.* Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. *Genome Biol.* 16(1), 80 (2015).

69 Ronn T, Volkov P, Gillberg L *et al.* Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. *Hum. Mol. Genet.* 24(13), 3792–3813 (2015).

70 Dick KJ, Nelson CP, Tsaprouni L *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* 383(9933), 1990–1998 (2014).

71 Feinberg AP, Irizarry RA, Fradin D *et al.* Personalized epigenomic signatures that are stable over time and covary with body mass index. *Sci. Transl. Med.* 2(49), 49ra67 (2010).

72 Hasler R, Feng Z, Backdahl L *et al.* A functional methylome map of ulcerative colitis. *Genome Res.* 22(11), 2130–2137 (2012).

73 Issa JP. Epigenetic variation and cellular Darwinism. *Nat. Genet.* 43(8), 724–726 (2011).

74 Issa JP, Ahuja N, Toyota M, Bronner MP, Brentnall TA. Accelerated age-related CpG island methylation in ulcerative colitis. *Cancer Res.* 61(9), 3573–3577 (2001).

75 Philibert RA, Plume JM, Gibbons FX, Brody GH, Beach SR. The impact of recent alcohol use on genome wide DNA methylation signatures. *Front. Genet.* 3 54 (2012).

76 Noreen F, Roosli M, Gaj P *et al.* Modulation of age- and cancer-associated DNA methylation change in the healthy colon by aspirin and lifestyle. *J. Natl Cancer Inst.* 106(7), pii: dju161 (2014) (Epub ahead of print).

77 Watson CT, Disanto G, Sandve GK, Breden F, Giovannoni G, Ramagopalan SV. Age-associated hyper-methylated regions in the human brain overlap with bivalent chromatin domains. *PLoS ONE* 7(9), e43840 (2012).

78 Lam LL, Emberly E, Fraser HB *et al.* Factors underlying variable DNA methylation in a human community cohort. *Proc. Natl Acad. Sci. USA* 109(Suppl. 2), 17253–17260 (2012).

79 Hansen KD, Sabunciyan S, Langmead B *et al.* Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. *Genome Res.* 24(2), 177–184 (2014).

80 Hahn MA, Hahn T, Lee DH *et al.* Methylation of polycomb target genes in intestinal cancer is mediated by inflammation. *Cancer Res.* 68(24), 10280–10289 (2008).

81 Lee TI, Jenner RG, Boyer LA *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125(2), 301–313 (2006).

82 Widschwendter M, Fiegl H, Egle D *et al.* Epigenetic stem cell signature in cancer. *Nat. Genet.* 39(2), 157–158 (2007).

83 Ohm JE, Mcgarvey KM, Yu X *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet.* 39(2), 237–242 (2007).

84 Schlesinger Y, Straussman R, Keshet I *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nat. Genet.* 39(2), 232–236 (2007).

85 Amant F, Moerman P, Neven P, Timmerman D, Van Limbergen E, Vergote I. Endometrial cancer. *Lancet* 366(9484), 491–505 (2005).

86 Udler MS, Ahmed S, Healey CS *et al.* Fine scale mapping of the breast cancer 16q12 locus. *Hum. Mol. Genet.* 19(12), 2507–2515 (2010).

87 Rando TA, Chang HY. Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock. *Cell* 148(1–2), 46–57 (2012).

88 Brack AS, Rando TA. Intrinsic changes and extrinsic influences of myogenic stem cell function during aging. *Stem Cell Rev.* 3(3), 226–237 (2007).

89 Xu Z, Bolick SC, Deroo LA, Weinberg CR, Sandler DP, Taylor JA. Epigenome-wide association study of breast cancer using prospectively collected sister study samples. *J. Natl Cancer Inst.* 105(10), 694–700 (2013).

90 Van Veldhoven K, Polidoro S, Baglietto L *et al.* Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. *Clin. Epigenetics* 7(1), 67 (2015).

91 Severi G, Southey MC, English DR *et al.* Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer. *Breast Cancer Res. Treat.* 148(3), 665–673 (2014).

92 Joyce BT, Gao T, Liu L *et al.* Longitudinal study of DNA methylation of inflammatory genes and cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 24(10), 1531–1538 (2015).

93 Beerman I, Bhattacharya D, Zandi S *et al.* Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc. Natl Acad. Sci. USA* 107(12), 5465–5470 (2010).

94 Maegawa S, Gough SM, Watanabe-Okochi N *et al.* Age-related epigenetic drift in the pathogenesis of MDS and AML. *Genome Res.* 24(4), 580–591 (2014).

95 Grolleau-Julius A, Ray D, Yung RL. The role of epigenetics in aging and autoimmunity. *Clin. Rev. Allergy Immunol.* 39(1), 42–50 (2010).

96 Steegenga WT, Boekschoten MV, Lute C *et al.* Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr.)* 36(3), 9648 (2014).

97 Yuan T, Jiao Y, De Jong S, Ophoff RA, Beck S, Teschendorff AE. An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. *PLoS Genet.* 11(2), e1004996 (2015).

98 Lee HS. Impact of maternal diet on the epigenome during *in utero* life and the developmental programming of diseases in childhood and adulthood. *Nutrients* 7(11), 9492–9507 (2015).

99 Martino D, Loke YJ, Gordon L *et al.* Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol.* 14(5), R42 (2013).

• **Important paper demonstrating that epigenetic drift can be measured in the first few months of life.**

100 Alisch RS, Barwick BG, Chopra P *et al.* Age-associated DNA methylation in pediatric populations. *Genome Res.* 22(4), 623–632 (2012).

101 West J, Widschwendter M, Teschendorff AE. Distinctive topology of age-associated epigenetic drift in the human interactome. *Proc. Natl Acad. Sci. USA* 110(35), 14138–14143 (2013).

102 Teschendorff AE. Epigenetic aging: insights from network biology. *Aging* 5(10), 719–720 (2013).

103 Genovese G, Kahler AK, Handsaker RE *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* 371(26), 2477–2487 (2014).

104 Adams PD, Jasper H, Rudolph KL. Aging-induced stem cell mutations as drivers for disease and cancer. *Cell Stem Cell* 16(6), 601–612 (2015).

105 Suzuki H, Toyota M, Carraway H *et al.* Frequent epigenetic inactivation of Wnt antagonist genes in breast cancer. *Br. J. Cancer* 98(6), 1147–1156 (2008).

106 Baylin SB, Ohm JE. Epigenetic gene silencing in cancer – a mechanism for early oncogenic pathway addiction? *Nat. Rev. Cancer* 6(2), 107–116 (2006).

•• **Seminal review paper, exposing the idea of DNA methylation alterations as early events in predisposing normal cells to carcinogenesis.**

107 Jones A, Teschendorff AE, Li Q *et al.* Role of DNA methylation and epigenetic silencing of HAND2 in endometrial cancer development. *PLoS Med.* 10(11), e1001551 (2013).

108 Tabassum R, Sivadas A, Agrawal V, Tian H, Arafat D, Gibson G. Omic personality: implications of stable transcript and methylation profiles for personalized medicine. *Genome Med.* 7(1), 88 (2015).

109 Wadsworth M, Kuh D, Richards M, Hardy R. Cohort profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *Int. J. Epidemiol.* 35(1), 49–54 (2006).

110 Bell JT, Spector TD. DNA methylation studies using twins: what are they telling us? *Genome Biol.* 13(10), 172 (2012).

111 Faulk C, Liu K, Barks A, Goodrich JM, Dolinoy DC. Longitudinal epigenetic drift in mice perinatally exposed to lead. *Epigenetics* 9(7), 934–941 (2014).

112 Tomasetti C, Vogelstein B, Parmigiani G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc. Natl Acad. Sci. USA* 110(6), 1999–2004 (2013).

113 Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347(6217), 78–81 (2015).

114 Hartnett L, Egan LJ. Inflammation, DNA methylation and colitis-associated cancer. *Carcinogenesis* 33(4), 723–731 (2012).

115 Suzuki H, Toyota M, Kondo Y, Shinomura Y. Inflammation-related aberrant patterns of DNA methylation: detection and role in epigenetic deregulation of cancer cell transcriptome. *Methods Mol. Biol.* 512 55–69 (2009).

116 Hansen KD, Timp W, Bravo HC *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* 43(8), 768–775 (2011).

117 Feinberg AP, Irizarry RA. Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc. Natl Acad. Sci. USA* 107(Suppl. 1), 1757–1764 (2010).

118 Feinberg AP. Epigenetic stochasticity, nuclear structure and cancer: the implications for medicine. *J. Int. Med.* 276(1), 5–11 (2014).

119 Wijetunga NA, Delahaye F, Zhao YM *et al.* The meta-epigenomic structure of purified human stem cell populations is defined at cis-regulatory sequences. *Nat. Commun.* 5, 5195 (2014).

120 Siegmund KD, Marjoram P, Woo YJ, Tavare S, Shibata D. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc. Natl Acad. Sci. USA* 106(12), 4828–4833 (2009).

121 Shibata D. Clonal diversity in tumor progression. *Nat. Genet.* 38(4), 402–403 (2006).

122 Teschendorff AE, Liu X, Caren H *et al.* The dynamics of DNA methylation covariation patterns in carcinogenesis. *PLoS Comput. Biol.* 10(7), e1003709 (2014).

123 Landau DA, Clement K, Ziller MJ *et al.* Locally disordered methylation forms the basis of intratumor methylome

variation in chronic lymphocytic leukemia. *Cancer Cell* 26(6), 813–825 (2014).

124 Mazor T, Pankov A, Johnson BE *et al.* DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. *Cancer Cell* 28(3), 307–317 (2015).

125 Issa JP. Age-related epigenetic changes and the immune system. *Clin. Immunol.* 109(1), 103–108 (2003).

126 Langevin SM, Houseman EA, Accomando WP *et al.* Leukocyte-adjusted epigenome-wide association studies of blood from solid tumor patients. *Epigenetics* 9(6), 884–895 (2014).

127 Lechner M, Fenton T, West J *et al.* Identification and functional validation of HPV-mediated hypermethylation in head and neck squamous cell carcinoma. *Genome Med.* 5(2), 15 (2013).

128 Xu Z, Taylor JA. Genome-wide age-related DNA methylation changes in blood and other tissues relate to histone modification, expression and cancer. *Carcinogenesis* 35(2), 356–364 (2014).

129 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539), 317–330 (2015).