

Analysis of DNA modifications in aging research

Dustin R. Masser · Niran Hadad · Hunter Porter ·
Michael B. Stout · Archana Unnikrishnan ·
David R. Stanford · Willard M. Freeman 

Received: 27 November 2017 / Accepted: 5 January 2018 / Published online: 11 January 2018
© American Aging Association 2018

Abstract As geroscience research extends into the role of epigenetics in aging and age-related disease, researchers are being confronted with unfamiliar molecular techniques and data analysis methods that can be difficult to integrate into their work. In this review, we focus on the analysis of DNA modifications, namely cytosine methylation and hydroxymethylation, through next-generation sequencing methods. While older techniques for modification analysis performed relative quantitation across regions of the genome or examined average

genome levels, these analyses lack the desired specificity, rigor, and genomic coverage to firmly establish the nature of genomic methylation patterns and their response to aging. With recent methodological advances, such as whole genome bisulfite sequencing (WGBS), bisulfite oligonucleotide capture sequencing (BOCS), and bisulfite amplicon sequencing (BSAS), cytosine modifications can now be readily analyzed with base-specific, absolute quantitation at both cytosine-guanine dinucleotide (CG) and non-CG sites throughout the genome or within specific regions of interest by next-generation sequencing. Additional advances, such as oxidative bisulfite conversion to differentiate methylation from hydroxymethylation and analysis of limited input/single-cells, have great promise for continuing to expand epigenomic capabilities. This review provides a background on DNA modifications, the current state-of-the-art for sequencing methods, bioinformatics tools for converting these large data sets into biological insights, and perspectives on future directions for the field.

D. R. Masser · N. Hadad · H. Porter · M. B. Stout ·
A. Unnikrishnan · D. R. Stanford · W. M. Freeman (✉)
Reynolds Oklahoma Center on Aging, University of Oklahoma
Health Sciences Center, Oklahoma City, OK, USA
e-mail: wfreeman@ouhsc.edu

D. R. Masser · D. R. Stanford · W. M. Freeman
Department of Physiology, University of Oklahoma Health
Sciences Center, Oklahoma City, OK, USA

D. R. Masser · N. Hadad · H. Porter · W. M. Freeman
Oklahoma Nathan Shock Center for Aging, University of
Oklahoma Health Sciences Center, Oklahoma City, OK, USA

N. Hadad · H. Porter · D. R. Stanford · W. M. Freeman
Oklahoma Center for Neuroscience, University of Oklahoma
Health Sciences Center, Oklahoma City, OK, USA

M. B. Stout · W. M. Freeman
Department of Nutritional Sciences, University of Oklahoma
Health Sciences Center, Oklahoma City, OK, USA

A. Unnikrishnan
Department of Geriatric Medicine, University of Oklahoma Health
Sciences Center, Oklahoma City, OK, USA

Keywords Epigenetics · Methods · DNA methylation

DNA modifications and aging

Epigenetic changes as a regulator of the aging process have been an area of research interest for a number of decades (Waddington 1940), even prior to the discovery of DNA methylation (Hotchkiss 1948). The roles of DNA methylation in regulation of genome organization and gene expression have progressed from solely being

a mechanism of permanent gene inactivation to one in which DNA methylation is also a dynamic gene regulator (Schubeler 2015), though the degree of dynamism is a matter of debate (Bestor et al. 2015). The concept of non-sequence-based changes to the genome and how their accumulation with advancing age may adversely affect lifespan and cellular function have obvious relevance to geroscience research (Lopez-Otin et al. 2013; Kennedy et al. 2014). Early efforts sought to examine total levels of DNA methylation and resulted in a theory of genomic hypomethylation (decreased methylation levels across the genome) with aging (Vanyushin et al. 1973; Wilson and Jones 1983). While the hypomethylation hypothesis is often referenced as dogma (Zampieri et al. 2015; Sen et al. 2016), the general consensus has shifted to rejecting this hypothesis based on data from modern quantitative techniques (Unnikrishnan et al. 2017a). Thus, the field has moved toward identifying specific genomic sites and regions with differential methylation in animal models (Maegawa et al. 2010) and humans (Rakyan et al. 2010). A number of recent reports have provided the first comprehensive analyses of altered DNA methylation with aging in a range of mouse tissues (Cole et al. 2017; Hahn et al. 2017; Masser et al. 2017b; Petkovich et al. 2017; Stubbs et al. 2017). Human studies have also advanced by examining age-affected tissues (e.g., Zykovich et al. 2014) and through the development of a variety of chronological aging “clocks” (Hannum et al. 2013; Horvath 2013; Weidner et al. 2014; Chen et al. 2016). While these clocks have reproducible predictive validity for chronological age and potentially mortality (Marioni et al. 2015), their relevance to biological aging processes remains to be determined. This is particularly evident given the counterintuitive results from recent reports regarding the relationship between “methylation age” and tissue function (Marioni et al. 2016; Kim et al. 2017; Kozlenkov et al. 2017; Simpkin et al. 2017). Overall, there is clearly a need for epigenetic studies of aging. This review is focused on the epigenetic domain of DNA modification—namely cytosine modifications. Given the demonstrated importance of the genetics of aging (Jeck et al. 2012), a similar effort is needed to understand how the epigenome changes in response to the various stimuli encountered throughout the lifecycle. Further review of the current state of scientific findings in the epigenomics of aging is beyond the scope of this review, but several recent publications provide an overview of findings in different models and organ systems

(Valdes et al. 2013; Benayoun et al. 2015; Jones et al. 2015; Sen et al. 2016). The purpose of this review is to examine the technical approaches that can be used in these studies to best meet individual experimental goals and educate geroscience researchers in designing and interpreting epigenomic studies.

Nature and regulation of DNA modifications

Epigenetic mechanisms are classically defined as chromatin structure and DNA modifications (Allis and Jenuwein 2016). While chromatin structure and histone modifications are an area of intense interest for aging research (Benayoun et al. 2015), the focus of this review is on DNA modifications due to their regulation of chromatin (Martinowich et al. 2003), the long-lasting nature of DNA modifications (Bird 2002), and because these modifications to the genome can be passed to daughter cells (Holliday 2006). The primary DNA modifications are methylation (mC) and hydroxymethylation (hmC) of cytosine at the 5' position of the pyrimidine ring (Globisch et al. 2010). Hydroxymethylation is especially abundant in brain tissues as compared to other tissues/organs (Kriaucionis and Heintz 2009; Lister et al. 2013). Much rarer modifications of formylcytosine (fC) and carboxylcytosine (caC) exist as well (He et al. 2011; Ito et al. 2011) (Fig. 1a). Non-cytosine modifications, though rarer still, are also beginning to be investigated (Yao et al. 2017).

DNA methyltransferases (DNMTs) maintain methylation patterns in CG contexts during cell division (DNMT1) and de novo methylate cytosines (DNMT3a and DNMT3b) in response to stimuli (Okano et al. 1999). Tet methylcytosine dioxygenases (TETs) oxidize mC into hmC, as well as fC and caC before the cycle is completed back to unmodified C (He et al. 2011) (Fig. 1b). This final stage of the cycle can also occur through thymine DNA glycosylase (TDG) and base excision repair (BER) (Maiti and Drohat 2011). Modifications occur on cytosines in both CG and CH (H being C, A, or T) dinucleotide contexts. Importantly, mCG is mirrored across the DNA strands because of the palindrome nature of the CG dinucleotide motif. On the other hand mCH is, by its nature, strand-specific (Fig. 1c) (He and Ecker 2015). Historically, CG methylation has been the focus of investigation. However, methylation of

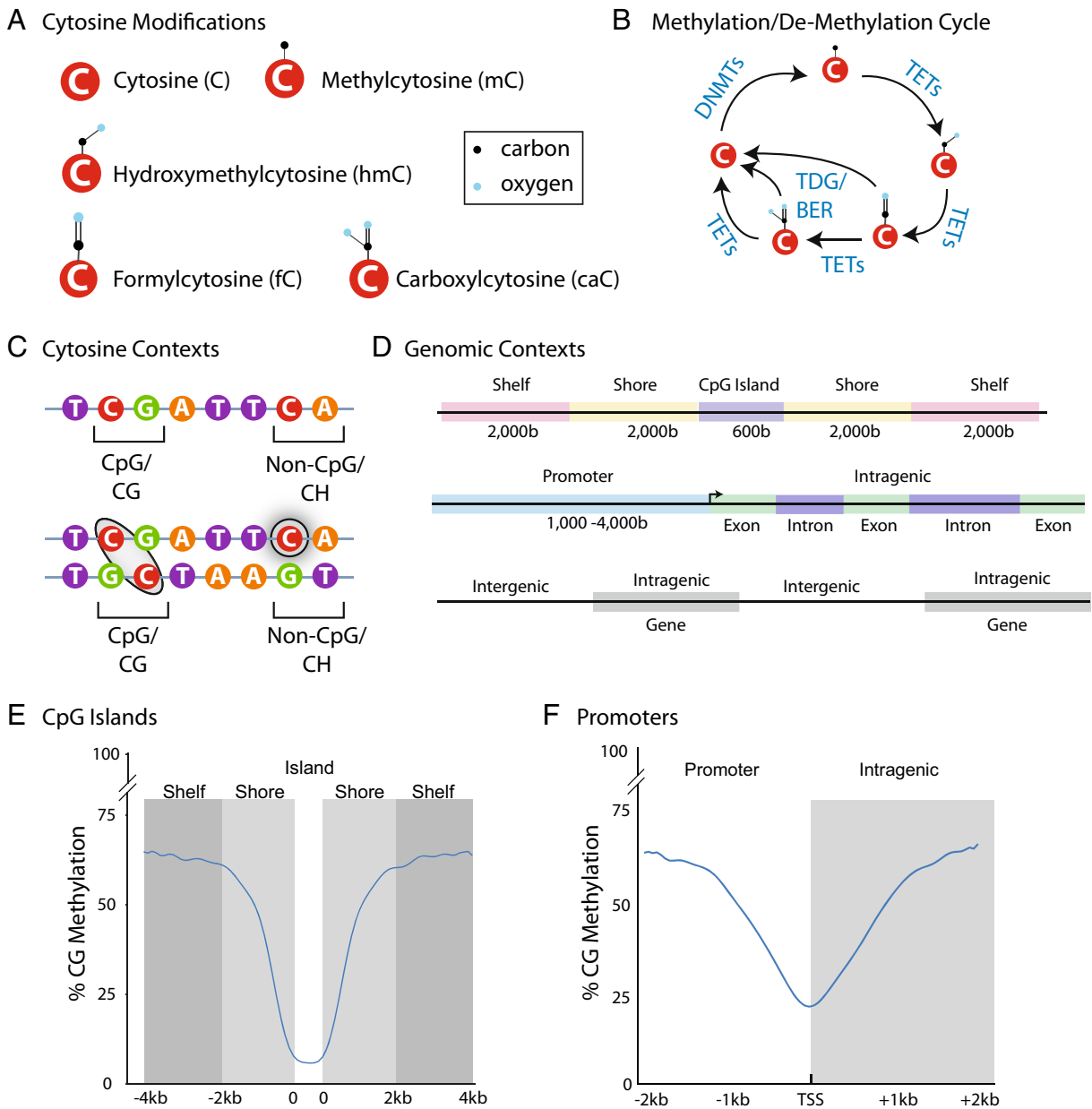


Fig. 1 Fundamental principles of DNA modifications. **a** Cytosine bases exist in an unmodified form and with methyl, hydroxymethyl, formyl, and carboxy additions at the five positions of the pyrimidine ring. **b** Over the last decade, advances in the understanding of modification regulation have characterized the cycle of modification addition and oxidation. DNA methyltransferases (DNMTs) add methyl groups in which Tet methylcytosine dioxygenases (TETs) sequentially oxidize modifications back to an unmodified cytosine or include base excision repair (BER) through thymine DNA glycosylase (TDG). **c** Cytosines are present in palindromic CG contexts with methylation paralleled between the two strands while CH (where H is A, C, or T) modifications are

by nature strand specific. **d** Changes in DNA modifications can occur across different forms of genomic elements with differing functional outcomes. CpG islands are areas of high CG density that are flanked by shore and shelf regions upstream and downstream. Similarly, methylation of promoter regions has traditionally been a focus but intragenic regions, either in exons or introns are an area of growing interest and play a role in gene expression regulation. **e** CG methylation is typically low over CG islands but higher in shores and shelves. **f** CG methylation across promoters often varies greatly across a relatively narrow region with lowest methylation typically observed around the transcription start site (TSS), even in un-expressed genes

CH sites and hydroxymethylation of CG sites are prevalent in several organ systems, particularly the central nervous system (Lister et al. 2013; Kinde et al. 2015). The high levels of hmCG and mCH in the brain stand in contrast to other organs (Nestor et al. 2012) and suggest the possibility that the epigenomics of the aging brain is substantially different from other tissues (Masser et al. 2017a).

CG dinucleotides occur far less frequently (<1% of dinucleotide pairs) than any other dinucleotide pairs (Bird 1980; Karlin and Mrázek 1997) and are clustered together in regions termed CG islands (Lander et al. 2001). Definitions for the regions around CG islands have also been established (Fig. 1d). The context of cytosine modification—promoter vs intergenic, intergenic vs intragenic—can result in different functional outcomes. For example, mCG in promoter regions typically represses gene expression whereas promoter hmCG levels are greater in highly expressed genes (Chapman et al. 2015). Within gene bodies, both mCG and hmCH are positively associated with gene expression (Lister et al. 2013; Lou et al. 2014). Promoter mCH is inversely correlated to gene expression and gene body mCH varies in its relationship to gene expression, being either repressive or associated with increased expression (Lister et al. 2013; Kinde et al. 2015; Lister and Mukamel 2015). The relationship of hmCH to gene expression is unknown, although it does appear that hmCH levels are extremely low (Hadad et al. 2016).

Although mC and hmC are recognized by common DNA binding proteins, they also have modification-specific binders (Spruijt et al. 2013), thereby providing a potential mechanism for the differential effects on gene expression between mC and hmC (Xu et al. 2011). Additionally, TET enzymes may prefer CG sites (Hu et al. 2013) which raises the intriguing possibility that mCH may be resistant to oxidation and elimination. Our understanding of the effects of mC/hmC on gene expression and genome organization with aging is in its infancy. A critical task for the field is to generate the data needed to test how, or whether, these general concepts apply to epigenomics of aging. From a geroscience perspective, it is essential that we clarify and expand our understanding of aging-associated alterations of epigenetic processes to determine their contribution to health and disease.

Historical methods for analysis of DNA modifications

The focus of this review is on current, next-generation sequencing (NGS) approaches to analyzing DNA modifications that provide base-specific, absolute quantitation. The differentiating factor of the historical methods from the current generation of NGS methods is that now absolute quantitation of modifications can be performed in a base-specific manner for either the whole genome or at least significant portions of it. However, a brief perspective of the approaches that lead to the current NGS methods is worthwhile. As the epigenomic field developed, the first assays were for total levels of cytosine modifications in the genome. High-performance liquid chromatography (HPLC) was one of the first methods to detect global methylation levels through column chromatography of hydrolyzed DNA (Reddy and Reddy 1990). Samples were then compared to a set of standards with known methylation levels (Reddy and Reddy 1990; Fuke et al. 2004). Updates to this method include replacing UV detection with mass spectrometry (Song et al. 2005). For more simple assays, enzyme-linked immunosorbent assays (ELISA) use antibodies that are specific for mC or hmC with an *in vitro* generated standard curve to measure the global levels of methylation in a given sample (Kalani et al. 2014). Challenges with this method are that many of the standards used in ELISAs contain different CG densities than mammalian genomes, resulting in inaccurate quantitation of methylation levels (Hadad et al. 2016) and modifications at CG and CH sites cannot be differentiated. The same antibodies developed for ELISAs can be used for immunohistochemistry (IHC) (Almeida et al. 2012), but great care must be taken as IHC is generally not quantitative and spurious conclusions can be drawn from simple densitometry of tissue sections. Another commonly proposed surrogate measure for total genomic methylation levels is pyrosequencing of repetitive elements (Yang et al. 2004). Repetitive elements constitute a large portion of the genome (de Koning et al. 2011), and analysis by relatively simple pyrosequencing approaches is a potentially facile approach to estimating genome-wide methylation. While methylation of repetitive elements is of biological interest in and of themselves, care should be taken to not simplistically assume they are an accurate marker of genome-wide methylation. Methylation levels vary greatly over the genome (Lister et al. 2013), and there is, in fact, very little evidence to support a

strong relationship between repeat element methylation level and other genomic regions (Blueprint consortium 2016; Hadad et al. 2016; Crary-Dooley et al. 2017).

All of these traditional methylation analysis approaches have been applied to a number of aging research studies but with varying degrees of limitations. Many of the findings have not been recapitulated across studies (see Unnikrishnan et al. 2017a), especially given the large magnitude of changes often presented. At times, the technical limitations of these methods have resulted in nonsensical findings like the sum of methylation and hydroxymethylation being greater than 100% (Mei et al. 2015). This argues for better analytical methods and analysis procedures that can reproducibly quantify DNA modifications. Preventing progress to more quantitative methods were the barriers of the cost of generating large amounts of sequencing data and the difficulty in differentiating modified and unmodified cytosines.

Affinity enrichment NGS

With the advent of high throughput sequencing (Shendure and Ji 2008), now commonly known as NGS, generating large amounts of accurate sequencing data become economical. This led initially to development of affinity-based methods in which modified DNA was “pulled” out from unmethylated DNA by immunoprecipitation in a manner similar to chromatin immunoprecipitation sequencing (ChIP-Seq) (Furey 2012). Methylated DNA immunoprecipitation (meDIP-Seq) (Weber et al. 2005; Jorgensen et al. 2006), methyl-binding domain sequencing (MBD-Seq) (Rauch and Pfeifer 2005; Serre et al. 2010), and a number of variations of these techniques (see Lister and Ecker 2009; Laird 2010) were developed. However, the limitation to any affinity-based approach is that the specific methylation sites (or hydroxymethylation; Tan et al. 2013) cannot be determined, only a general genomic region of methylation. While this was a significant advance, the primary limitations of these approaches, relative quantitation across regions and lack of base-specific data analyses, diminish the insight that could be garnered from such approaches.

Base-specificity of DNA modifications

Base-specificity in DNA modification analysis is critical for four principle reasons: (1) levels of DNA

modifications demonstrate large differences over narrow genomic regions, (2) age-related changes appear to occur in a base-specific manner, (3) base-specificity is required to differentiate CG from CH methylation, and (4) with base-specific absolute quantitation, meta-analysis and comparisons between studies can be readily performed. Starting with the first point, it has become apparent across a large number of studies that methylation levels differ dramatically over relatively short genomic distances (e.g., Edgar et al. 2014; Zhao et al. 2014; Schultz et al. 2015). This is especially true at transcription start sites (TSSs) and at CpG islands which tend to be minimally methylated. Methylation levels then rise dramatically upstream and downstream of these genomic features. For example, taking base-specific whole genome bisulfite sequencing data from mouse prefrontal cortex, nadirs of methylation are evident at CpG islands and TSSs (Fig. 1e, f). Therefore, averages over regions that are 1–10 kb wide provide limited insight into the true pattern of methylation. Furthermore, a number of reports have described base-specific changes in methylation with aging (Maegawa et al. 2010; Mangold et al. 2017a; Petkovich et al. 2017; Stubbs et al. 2017). Moreover, given the growing importance of examining methylation at both CG and CH sites, site-specific quantitation is required to differentiate between these two. We have previously demonstrated in the mouse hippocampus, for example, the majority of changes in methylation with aging are in the CH context, which cannot be differentiated from CG by meDIP-Seq and MBD-Seq techniques outlined above (Masser et al. 2017a). Lastly, absolute quantitation in a base-specific manner allows data from multiple studies to be combined to increase power, test reproducibility, identify biomarkers, and compare between species (Horvath 2013; Masser et al. 2017a; Petkovich et al. 2017; Stubbs et al. 2017). Taken together, these points provide strong justification that to fully understand differences in DNA modifications with aging, base-specific resolution is required.

The concept of examining base-specific methylation levels is not new. In the past, this has generally been performed on small genomic regions (< 100 bp) at specific genomic loci through bisulfite conversion and pyrosequencing or Sanger sequencing. Bisulfite conversion works through deamination of unmodified cytosines to uracils which are then copied as thymines during subsequent amplification (Frommer et al. 1992; Clark et al. 1994). Methylated and hydroxymethylated cytosines are

protected from this conversion. Thus, modification status is turned into a base difference that can easily be read by sequencing (Fig. 2a). The methylation level in the sample is determined by collecting a number of sequencing reads over the region and counting the number of cytosines and thymines at a specific site. Methylation level equals $C/C + T$ (Fig. 2b). Sequencing was traditionally performed by pyrosequencing (Dupont et al. 2004) or standard Sanger sequencing (Parrish et al. 2012). These methods have been used in a wide range of studies and perform

comparably (Reed et al. 2010). However, they provide coverage over only a small region, usually just a few cytosines. Additionally, pyrosequencing requires multiple sequencing primers and Sanger sequencing traditionally required cloning of PCR products, a laborious task (Zhang et al. 2009). Nevertheless, these techniques have been used to great effect to generate base-specific analysis of selected regions and changes in methylation, but not hydroxymethylation, with aging (Noer et al. 2007; Maegawa et al. 2010; King et al. 2012).

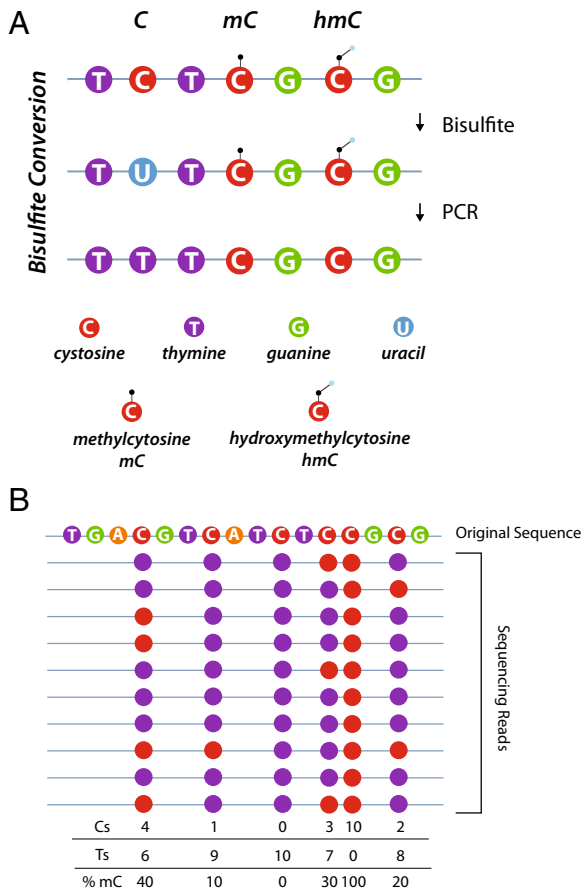
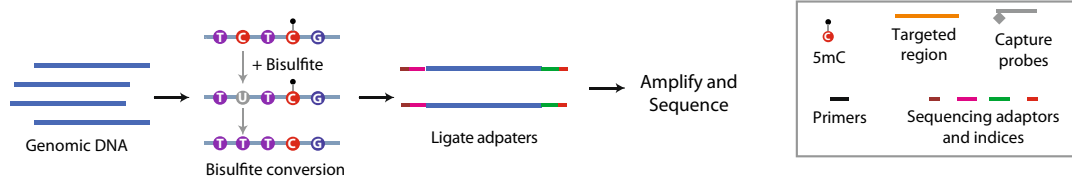


Fig. 2 Bisulfite conversion and quantitation of mC. **a** Bisulfite sequencing approaches are based on the concept that methylation, or hydroxymethylation, prevents the deamination of cytosine to uracil by bisulfite. With PCR amplification, uracils are copied as thymines. The difference in methylation, or hydroxymethylation, status of a cytosine can then be read through sequencing as a base difference (C vs T). **b** Quantitation of methylation level is achieved by sequencing over a genomic region many times. The number of times a cytosine in the reference sequence is read as a thymine or cytosine is counted. The methylation level is then computed for each cytosine as $C/C + T$ for that site. The greater the number of times that site is sequenced, the more accurate the methylation quantitation

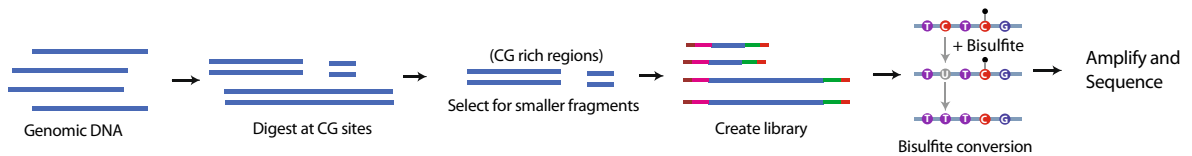
Bisulfite sequencing methods

As stated above, developments in NGS allowed rapid and efficient generation of massive amounts of high-quality sequencing data and bisulfite conversion approaches make base-specific quantitation possible. In a natural progression, these techniques were combined into a new generation of methods. Whole-genome bisulfite sequencing (WGBS) was first introduced to quantify total genome methylation levels in the plant *Arabidopsis* (Cokus et al. 2008; Lister et al. 2008) but was rapidly adapted for use in mammalian genomes (Lister et al. 2009). For WGBS, genomic DNA libraries are created and subsequently bisulfite converted, sequenced, and mapped back to the reference genome. Subsequent versions of WGBS perform library construction after bisulfite conversion (Miura et al. 2012; Khanna et al. 2013) (Fig. 3a) or use alternative transposase-mediated library construction (Adey and Shendure 2012). The benefits of WGBS are obvious; it provides the most comprehensive, base-specific, absolute quantitation of DNA methylation at potentially all cytosines in a genome. The challenge with WGBS is the amount of sequencing data needed. Humans and standard rodent laboratory animals have genomes of ~ 3 billion bases (3 Gb) per DNA stand. Sufficient sequencing depth (5–15 \times) over any given base is needed to accurately quantify methylation (Ziller et al. 2015) and thus across the whole genome requires >25 Gb of sequencing per sample, after accounting for losses due to quality control and alignment efficiency. This makes WGBS studies very large undertakings and not reaching needed sequencing depths can impair identification of statistically significant group differences. NGS costs have come down in recent years, but the depth requirements as detailed by the scientific community have increased. Therefore,

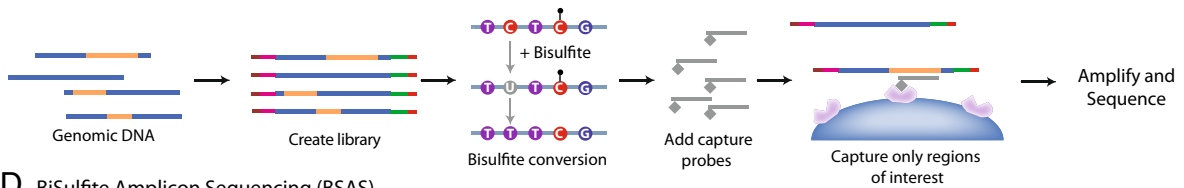
A Whole genome bisulfite sequencing (WGBS)



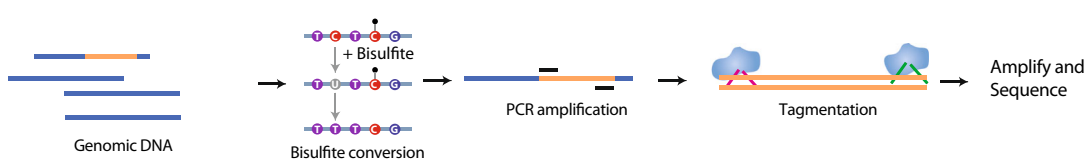
B Reduced Representation Bisulfite Sequencing (RRBS)



C Bisulfite Oligonucleotide-capture sequencing (BOCS)



D Bisulfite Amplicon Sequencing (BSAS)



E Methylation Microarray

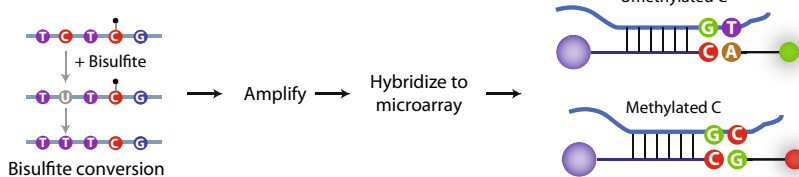


Fig. 3 Methods for base-specific cytosine modification quantitation. **a** Whole-genome bisulfite sequencing (WGBS) in general principle consists of bisulfite modification of genomic DNA and then creation of a sequencing library. Variants may switch the order of bisulfite conversion and library preparation. This approach gives quantitation across the genome but requires very large amounts of sequencing. **b** A long-standing approach to the analysis of large portions of the genome, but without the expense of WGBS is reduced representation bisulfite sequencing (RRBS). Genomic DNA is digested at CG sites, and then, the resulting smaller molecular weight fragments are isolated. These fragments are made into a sequencing library and bisulfite converted prior to sequencing. This approach focuses the data on CG rich regions of the genome. **c** Bisulfite oligonucleotide capture sequencing (BOCS) is analogous to exome sequencing techniques. In this approach, a whole genome library is made and then bisulfite converted. Genomic regions of interest are then captured with

oligonucleotide probes greatly enriching for regions of interest and thereby decreasing the amount of sequencing required. **d** Often, analysis of a specific genomic loci is desired, e.g., a gene of interest. With bisulfite amplicon sequencing (BSAS), the specific regions are amplified from bisulfite converted DNA and then made into a sequencing library. In effect, the PCR amplification greatly enriches for the region of interest. **e** While not a sequencing approach, one of the most common approaches to methylation quantitation is with a microarray format. This is a modification of SNP microarrays where probes are designed to detect a converted (T) or unconverted (C) at a CG site. Genomic DNA is bisulfite converted and amplified prior to hybridization to a microarray where the two different “alleles” have different reported colors. This is a high throughput approach, but a microarray must be available for the CG sites of interest and the species being examined. Currently, only human microarrays are commercially available

more sequencing is required to meet expectations of the community for the validity of the quantitative data. These factors make WGBS an expensive approach and the analytical methods required to process the data daunting, especially when including detailed analysis of all cytosine contexts (CG and CH), which can increase the number of analyzed cytosines by orders of magnitude. One approach to decreasing the scale of WGBS is to perform low coverage sequencing and eliminate base-specific analyses. With this approach, patterns of methylation can be examined by collapsing genomic elements, e.g., promoters, CpG islands (Hadad et al. 2016). This is very useful for getting initial insights into the total levels and general patterns of DNA modifications but goes against the goal of determining base-specific differences with aging. Thus, a number of approaches have been developed to target sequencing to only a portion of the genome and thus greatly reduce the amount of sequencing required.

Reduced representation bisulfite sequencing—RRBS/eRRBS

The first sequencing approach to reducing the amount of the genome covered utilized the use of restriction enzymes. This was built on prior work that used combinations of methyl-sensitive and insensitive restriction enzymes (Cedar et al. 1979; Reilly et al. 1982). Reduced representation bisulfite sequencing (RRBS) uses the restriction enzyme *MspI* to cleave CCGG motifs (Meissner et al. 2005) (Fig. 3b). Libraries are then made and only lower weight fragments (40–200 bp) (Gu et al. 2011) are selected. This enriches for CG-rich regions like CpG islands, and promoter regions containing CpG islands. However, coverage over other genome elements like CpG island shores and some enhancer elements is limited (Gu et al. 2011). With refinements to the method, enhanced RRBS (ERRBS) coverage in some of the less CG dense regions was improved (Akalın et al. 2012b; Garrett-Bakelman et al. 2015). The primary limitation to these methods is the dependence on the location of restriction sites in the genome and inability to tune coverage to regions of interest. Additionally, the size selection step is critical to which genomic regions are analyzed which is why there are often very limited overlapping sites between different RRBS datasets (Stubbs et al. 2017).

Oligonucleotide capture

Selecting regions of interest for sequencing is not a new concept. For a number of years, what is popularly termed exome sequencing has been used in genetics to focus sequencing on protein coding regions of the genome (Mamanova et al. 2010). An analogous approach can be applied to DNA modification analyses. Combining bisulfite sequencing with oligonucleotide capture offers the potential to focus analyses on genomic regions of interest while maintaining base-specific, absolute quantitation (Fig. 3c). We term these approaches as bisulfite oligonucleotide capture sequencing (BOCS). Several very similar approaches have been developed for assessing the human (Wang et al. 2011; Ivanov et al. 2013; Allum et al. 2015; Li et al. 2015), mouse (Hing et al. 2015; Li et al. 2015), and rat (Masser et al. 2016) genomes. In all of these approaches, capture of targeted regions is by complimentary oligonucleotide probes. Slight differences in the oligonucleotides and whether capture is performed before or after bisulfite conversion are the main differences between these methods. These approaches, which provide base-specific absolute quantitation over ~80–200 Mb of relevant genomic regulatory regions, are poised to be a widely employed approach for aging research studies. Recently, we have described the use of this approach for geroscience research (Hadad et al. 2017a; Masser et al. 2017c).

Generation of the oligonucleotide probes by an individual lab would be impractical and a variety of commercially available BOCS approaches are available. Roche, Agilent, and Illumina offer probesets for humans and mice. Regardless of the specific capture approach/supplier used, the ability to tune the regions of the genome being examined (Ziller et al. 2016) and the relatively straightforward nature of building probesets for other species (Masser et al. 2016) make this a promising approach for the future.

Amplicon sequencing

While both whole genome and BOCS approaches are well suited to discovery studies, often experimental goals only require analysis of a specific genomic locus or set of loci. Highly focused sequencing decreases the cost per sample and increases throughput. Additionally, very high sequencing depth is easily achieved when only small amounts of (< 1 Mb) of the genome are being tested.

For hypothesis testing experiments, several approaches have been developed that still take advantage of the power of bisulfite conversion and NGS. To achieve this, PCR primers can be designed to amplify bisulfite converted DNA, and the resultant amplicons are used to generate sequencing libraries. We term these approaches bisulfite amplicon sequencing (BSAS) (Masser et al. 2013; Masser et al. 2015) (Fig. 3d). Several other BSAS variations include using restriction enzymes (Varley and Mitra 2010), use of microdroplets for the PCR (Komori et al. 2011), further evolutions of the library preparation (Bernstein et al. 2015; Bhat et al. 2016), and the addition of analyzing hydroxymethylation (Chen et al. 2017). These approaches provide base-specific absolute quantitation in up to 96 samples of 1–10 kb targeted regions, making this method highly valuable for hypothesis-driven aging research. Additionally, this approach is much less expensive compared to whole genome or genome-wide capture techniques making it accessible to more laboratories. Recent aging studies have employed this technique demonstrating its usefulness in any tissue or cells with multiple targets of interest (Franzen et al. 2017; Mangold et al. 2017a; Unnikrishnan et al. 2017b).

The primary technical hurdle to amplicon based approaches is the design and validation of the PCR primers. They must be designed against bisulfite converted genomic sequence and generally avoid having CG sites in the primer binding region (Masser et al. 2015). In some cases, such as highly rich CG regions, primer design can be difficult and may preclude analysis of a specific locus.

Methylation microarrays

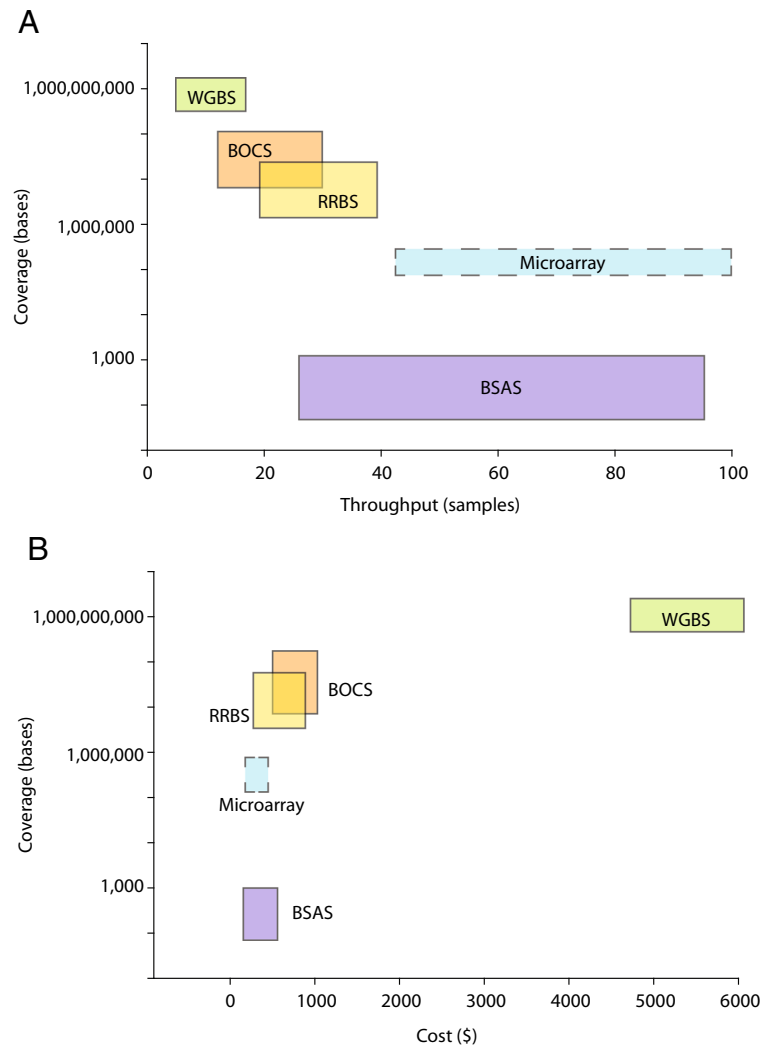
The primary alternate approach to methylation analysis that does not use NGS is high density microarrays (Bibikova et al. 2009; Bibikova et al. 2011). These microarrays are adaptations of SNP microarrays widely used in the genetics field, coupled with bisulfite conversion. This method works through bisulfite conversion of the sample DNA and hybridization to oligonucleotide probes on a microarray slide. The detection probe used is either complimentary to C—if the base was methylated and protected from bisulfite conversion, or complimentary to A—when the cytosine was unmodified and not protected from bisulfite conversion (Fig. 3e). Different fluorophores are conjugated to the different detection probes allowing

discrimination. Microarray slides are scanned and automated software computes a beta (β) value equivalent to methylation percentage. Recently, a new generation of these microarrays which cover 850,000 CG sites in the human genome have been released (Moran et al. 2016), although there are questions about the reproducibility of this platform as compared to prior microarrays (Logue et al. 2017). The advantages of this method are the relatively simple workflow, comparatively low cost per sample, and automated data analysis. This method is also amenable to high throughput analysis of large numbers of human samples (Chen et al. 2016). The primary limitations of methylation microarrays are that they are available only from one company (Illumina) and only for the human genome. As well, these arrays have been designed only against CG sites and cover far fewer sites than can be obtained through WGBS or BOCS. Nonetheless, this has been a valuable method for clinical geroscience studies and has seen wide utility in the field (Christensen et al. 2009).

Comparison of sequencing methods

As a researcher new to the study of DNA modifications, it would be entirely reasonable to be left overwhelmed by the variety and complexity of all the methods now available for DNA methylation analysis. Analytical performance of some of these methods has been compared (Harris et al. 2010; Blueprint_consortium 2016). Among focused analyses, bisulfite amplicon sequencing approaches performed better than pyrosequencing or methylation microarrays (Blueprint_consortium 2016). In-depth comparisons of WGBS, BOCS, and RRBS have not been performed although this is desperately needed by the field. Both BOCS (Li et al. 2015; Masser et al. 2016) and RRBS (Gu et al. 2011) have been compared in a limited manner to WGBS, though more in-depth comparisons are needed. It has been suggested that post-bisulfite library construction approaches may be better (Olova et al. 2017). If all bisulfite sequencing methods are considered technically sound, the choice of methods is a balance of the degree of genome coverage needed scientifically and practical reality of the resources available. Considering genomic coverage requirements, sample throughput, and resources available (Fig. 4), the current range of technologies can meet most needs.

Fig. 4 Sample throughput and cost as a function of genomic coverage. Experimental design for DNA modification analysis requires consideration of a range of scientific and practical factors. **a** Genomic coverage as a function of sample throughput. With greater coverage of the genome, as a general principle, sample throughput decreases. **b** Cost per sample increases as a function of genomic coverage. Costs vary on the number of samples performed and the specific circumstances of a study, but generally, the more genome analyzed the greater the cost as more sequencing is performed. Note: costs are estimated at a 15 \times sequencing depth and the microarray technology is denoted with dashed lines as the calculations of genomic coverage are inherently different as this is not a sequencing approach



Bioinformatics

So far, we have considered the methods for collection of methylation data, but this is only the first part of any study. Analysis of bisulfite sequencing data and the subsequent tertiary analysis are a complex topic. While a full discussion is beyond the scope of this review, recent publications provide greater detail (Bock 2012; Krueger et al. 2012; Baubec and Akalin 2016; Shafi et al. 2017). Nonetheless, the basic steps of sequencing data analysis consist of the following: quality control, alignment, quantitation, differential methylation determination, and tertiary analysis where patterns of methylation are integrated with other forms of annotation such as genomic features, other epigenomic data (e.g.,

chromatin and enhancers), gene expression, and pathways.

Quality control uses many of the same tools available for genomic and transcriptomic studies and consist of filtering out low-quality sequencing reads, contaminating adaptor sequences, etc. The principle quality control step specific to bisulfite sequencing is positive controls for complete bisulfite conversion of unmethylated cytosines to uracils. Estimation of conversion efficiencies was previously performed using non-CG sites as it was assumed that non-CG methylation did not exist, and therefore, any methylation signal at non-CG sites was failed conversion (Sun et al. 2015). This is clearly unwise given our current understanding of non-CG methylation as a common and biological relevant

phenomenon. Therefore, the current best practice for conversion controls is exogenous unmethylated spike-in sequences such as PhiX that are known to be unmethylated. When included at the beginning of workflow for each sample, conversion rates can be calculated and corrections can be applied on a sample-by-sample basis.

Alignment of bisulfite sequencing data also differs from standard sequencing alignments. Bisulfite conversion has created an ambiguous base at cytosines that can now be either C or T. Bismark is generally the most frequently used aligner for bisulfite sequencing data (Krueger and Andrews 2011), though other sequence aligners have been reported including BSMAP (Xi and Li 2009), BS Seeker (Chen et al. 2010), and others (Bock 2012). Bisulfite sequencing alignment is computationally intensive and can take days and weeks of processing time. Some new approaches seek to utilize computer hardware differently to greatly speed bisulfite sequencing alignments (Klus et al. 2012; Koster and Rahmann 2014; Manconi et al. 2014). With quality controlled and aligned sequencing data, quantitation of methylation levels is the next stage of analysis. It should be noted that sometimes an approach of “calling” whether methylation is present or not is employed at this stage. We have found this to be counter-productive as it oversimplifies and creates missing data. Rather, treating methylation as a continuous variable from 0 to 1 is the most useful approach to employ.

Most studies are performed to determine differential methylation between conditions. The primary decisions at this stage are whether to analyze differential methylation in a base-specific or regional fashion and whether or not to examine CH methylation. Based on our experience in brain aging research (Hadad et al. 2016; Hadad et al. 2017c; Mangold et al. 2017b; Masser et al. 2017a), we would advocate for base-specific and inclusion of CH analysis. To provide visualization of the advantage of base-specific analysis, we include an example comparing methylation between two different brain regions at an MHCI promoter by BSAS (Mangold et al. 2017a). This promoter region contains a number of CG and CH sites, as well as a range of transcription factor motifs (Fig. 5a). Examining CG methylation with base resolution a pattern of lower methylation nearer the TSS is evident with significantly higher methylation in the cortex at CGs in the promoter region than in the cerebellum (Fig. 5b). Similarly, differences in CH methylation between cortex and cerebellum throughout the

promoter region and into the intragenic regions are evident (Fig. 5c). Clearly, not examining CH data would have missed a number of differences in the methylation pattern between these brain regions. If quantitation is performed in a regional manner (1 kb region), higher cortical CG methylation (Fig. 5d) and higher cerebellar CH methylation (Fig. 5e) are evident. However, all resolutions as to where the specific differences are and their relationship to introns, exons, promoters, and transcription factors (Fig. 5a) are lost. It seems reasonable that if sufficient sequence depth is achieved, that base-specific analysis is superior to region-based analyses. The primary reason to perform regional analyses is when sequencing depth is limited and combining data across a region provides for more data.

For base-specific quantitation, a number of different software packages are available including MethyKit (Akalin et al. 2012a), DSS (Park and Wu 2016), BSsmooth (Hansen et al. 2012), and Limma (Pacheco et al. 2011). In-lab developed R or Python scripts are also commonly performed, requiring development of programming skills. This does require commitment of time to training but opens up many new possibilities in data analysis (Tippmann 2015). Whether as a base-specific or regional analyses, the outcome is a listing of differentially methylated cytosines (DMCs) or regions (DMRs). Interpreting the genomic localization and potentially relationships to gene expression is the next step of analysis—typically referred to a tertiary analysis.

Tertiary analysis requires a review of its own but is an important aspect of experimental design and output. For a brief summary, it is best to think of integration with genomic feature data, other epigenetic data, and gene expression data. For genomic features, this often takes the form of determining if changes are more or less likely than chance to occur in a certain feature such as a promoter, intron, and CG island. A number of examples exist from recent studies (Cole et al. 2017; Dozmorov 2017; Hadad et al. 2017b; Masser et al. 2017a; Wang et al. 2017). For integration with other epigenomic data, the process is similar but the comparison is with other epigenomic data such as ChIP-Seq. This requires data from the same tissue and ideally from the same sex and age. Most directly, this can be viewed as comparing two sets of genomic locations to determine where they overlap and if there are statistically significant over- or under-representations (Dozmorov 2017).

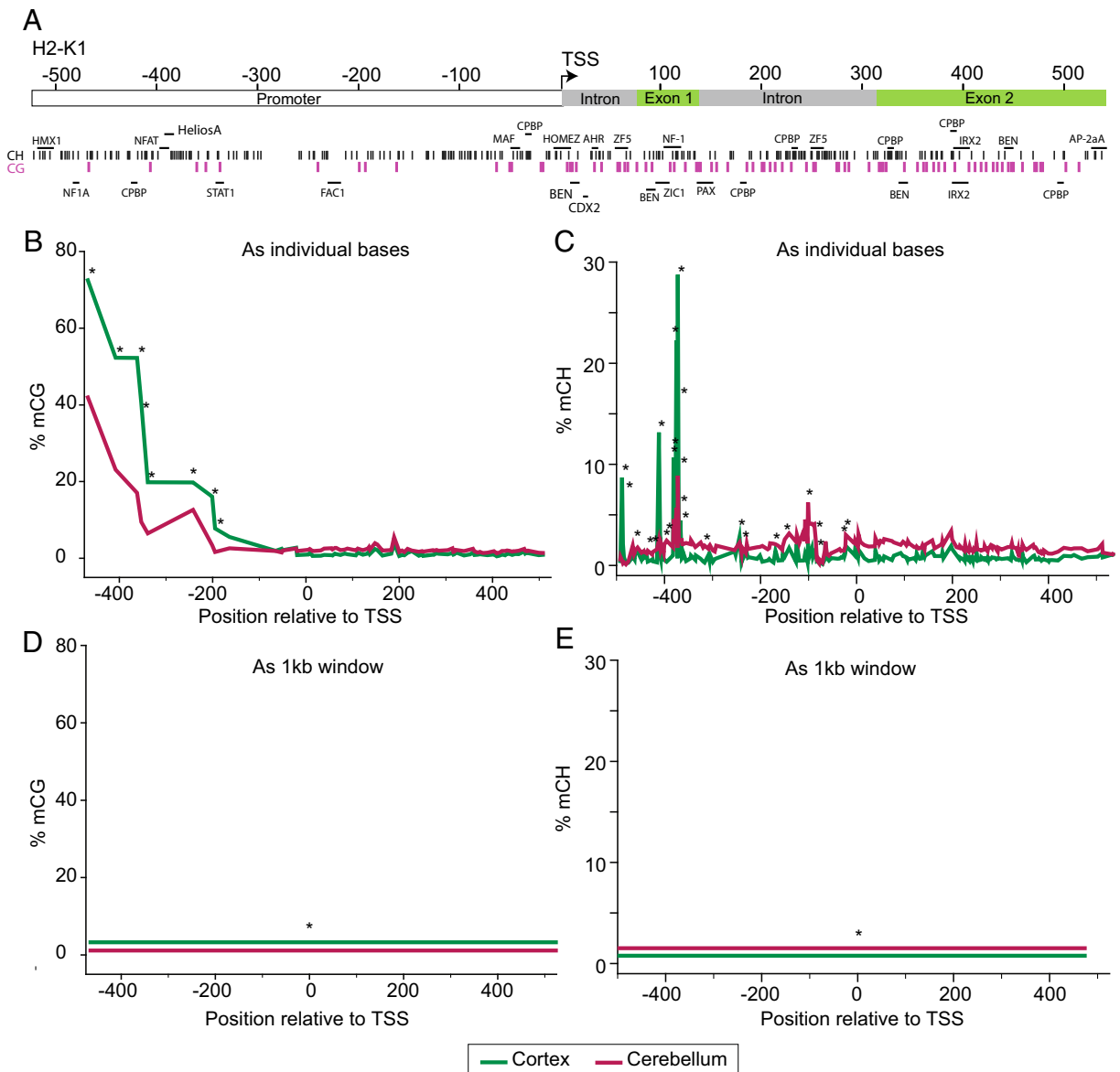


Fig. 5 Base-specific versus region-based quantitation. For computational simplicity, methylation analysis is sometimes performed as regions. This signal compression comes with distinct disadvantages. **a** In this example from Mangold et al. (2017a), a region ± 500 bp of the mouse H2-K1 promoter containing a number of genomic features (promoter, intron, exon, and transcription factor sites) and CG (purple ticks) and CH sites (black ticks) was examined by BSAS. CG (**b**) and CH (**c**) methylation demonstrated

distinct, base-specific, profiles across the region examined. Differences in methylation levels between the cerebral cortex and cerebellum were found at a number of sites ($*p < 0.05$, *t* test, B-H MTC, $n = 8/\text{group}$). If these values are compressed into a region analysis across the 1 kb, differences are still evident in CG (**d**) and CH (**e**) methylation but the magnitude of differences are compressed and the specificity for the differentially methylated sites to genomic features is lost

Comparing methylation with gene expression possesses a unique problem given the base-specific data to compare to transcript count data. The field is working on packages for aiding in this analysis but accepted solutions have not yet been developed.

Suffice it to say that simple correlations of region methylation to transcript count may not fully reveal the epigenetic control of a region. The best available approach is to utilize data visualization such as by methylPipe (Kishore et al. 2015).

Other modifications

With the discovery of the abundance and specific functional roles for alternate cytosine modifications (for reviews, see Moore et al. 2013; Shi et al. 2017), new methods have emerged in order to distinguish these modifications (hmC, fmC, caC) from mC (Raiber et al. 2017). hmC is of the greatest interest given its prevalence and potential regulation with aging (Szulwach et al. 2011). Because traditional bisulfite sequencing cannot differentiate between mC and hmC, methylation in a bisulfite sequencing experiment is really the sum of methylation and hydroxymethylation. When methods that distinguish mC from hmC are used up to a quarter of what was previously assumed to be methylation is actually hydroxymethylation depending on the tissue (Fig. 6a). This is further supported by the base-specific

regulation of mC and hmC over small genomic regions (Fig. 6b).

There are new methods that have been developed to distinguish between mC and hmC including Ox-BS (Booth et al. 2014) and TAB-seq (Yu et al. 2012). Oxidative bisulfite sequencing (Ox-BS) includes an oxidation step to convert hmC to fmC (Fig. 6c) which is then subject to bisulfite conversion. Parallel runs of standard bisulfite sequencing and oxidative bisulfite sequencing respectively give combined mC + hmC and mC only data, respectively. hmC is then determined by subtraction. Importantly, Ox-BS as a conversion chemistry has been adapted to analysis of focused genomic regions (Chen et al. 2017). TAB-Seq uses glycosylation of 5-hmC and then bisulfite conversion and sequencing DNA libraries to quantify methylation levels (Booth et al. 2012; Yu et al. 2012; Booth et al.

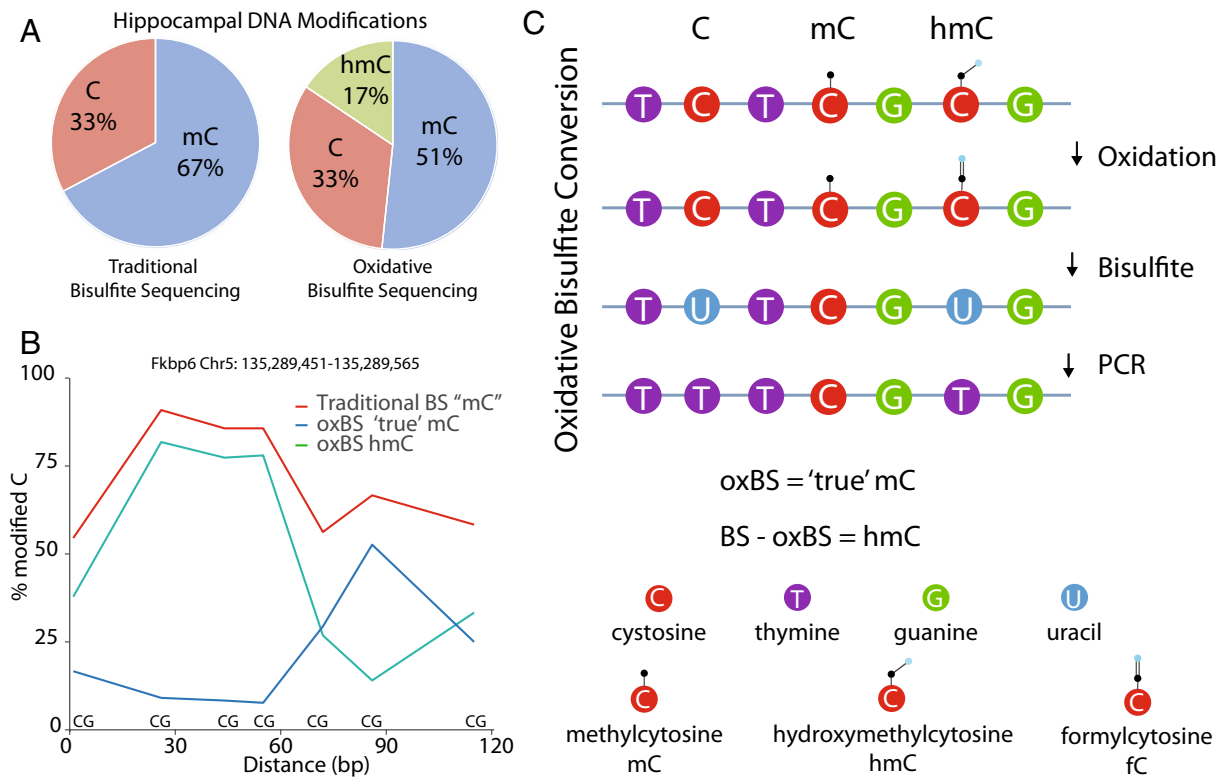


Fig. 6 Hydroxymethylation analysis. Most DNA modification analysis studies use bisulfite sequencing approaches that do not differentiate between methylation and hydroxymethylation. However, hydroxymethylation is abundant in a number of tissues. For example, using data from Hadad et al. (2016) who examined the mouse hippocampus by oxidative bisulfite sequencing that differentiates between methylation and hydroxymethylation, it is clear (a) that one fourth of what would normally be called as methylation is in fact hydroxymethylation. Further demonstrating this

point, within a relatively small region (~120 bp of Fkbp6; b), the levels of methylation and hydroxymethylation are clearly regulated in a base-specific fashion. Hydroxymethylation can be differentiated from methylation by an oxidation step (c) that deprotects hydroxymethylation and allows a specific methylation quantitation. When combined with bisulfite sequencing, a subtractive approach can be used to quantify methylation and hydroxymethylation individually

2013). RedBS-seq was developed to measure levels of fmC through reducing the fmC base back to hmC in order for quantify by sequencing, much like the Ox-BS method (Booth et al. 2014). caC can be quantified using CAB-seq, where EDC is used to chemically modify the 5-caC to prevent it from converting during bisulfite conversion, and levels quantified by sequencing (Lu et al. 2013). An alternative is methylation assisted bisulfite sequencing (MAB-Seq) which can be used to determine, in a base-by-base fashion combined caC and fC (Neri et al. 2016).

Single cell/cell-type analyses

While the field has experienced rapid advances in epigenomic analysis techniques, method development continues, especially in the context of analyzing single cells, single cell types, or very small samples. Several groups including those lead by Reik (Angermueller et al. 2016; Clark et al. 2016; Clark et al. 2017), Vijg (Gravina et al. 2015; Gravina et al. 2016; Yu et al. 2017), Adey (Mulqueen et al. 2017), and Eckardt (Luo et al. 2017) have also demonstrated that methylation patterns can be generated from single cells. However, the principle caveat to these methods is that only a small amount of the genome is analyzed (often around 5%) and the same genomic regions are not analyzed in each cell, making cell-to-cell comparisons of specific loci difficult. Clearly, there is a great need for these methods but they require additional development to provide comparisons of individual single cells. Currently, the best option for dealing with cellular heterogeneity in complex tissues is through existing methods for isolation of specific cell types that can then be analyzed (Mo et al. 2015; Mo et al. 2016).

An alternate approach to the question of cellular heterogeneity is through the use of algorithms that “correct” for cellular heterogeneity. This can either be based on reference patterns of cell-type specific methylation (Houseman et al. 2012) or without reference cell data (Zou et al. 2014). Comparisons of these methods have been performed (Kaushal et al. 2017; Teschendorff and Relton 2017) with recent data casting doubt on the accuracy of these correction methods (Zheng et al. 2017). Given the still limited data on the DNA modification profiles of different cell types, and their response to aging the accuracy of these methods, it needs to be

fully validated before they can see widespread usage in aging studies.

Alternatives to conversion-based sequencing

As discussed above, the current standard for analysis of DNA modifications is to use chemical conversion approaches to change modification status to a base difference that can be read with current next-generation sequencing approaches. This has strengths and weaknesses including controlling for conversion efficiency and accurately aligning altered sequences to their reference genome. An alternative approach may be to use nanopore sequencers to directly read different modifications from the DNA (Burgess 2017; Schatz 2017). Nanopore sequencing uses pores through which nucleic acid strands are “pulled” and the ionic pattern reveals the nucleotide sequence, including modifications (Jain et al. 2017; Shendure et al. 2017). These technologies are still in the developmental phase but show great promise (Rand et al. 2017). With continued technical development, they may be employed to aging research making direct analysis of DNA modification patterns from un-manipulated DNA possible. An exciting possibility of this single-molecule approach would be that individual cells could potentially be interrogated, without the need for extensive amplification (and the issues this can cause) but this still requires additional method development (Simpson et al. 2017).

Future directions

Employing the above described technologies offers great promise to understand the nature and patterns of DNA modifications as well as how they respond to aging. In addition to performing these discovery studies, the greatest challenge for the field is transitioning from much needed epigenome phenotyping studies to interventional studies that seek to prevent or cause age-related patterns of DNA modifications to determine the function. A major difficulty is that simplistic manipulation of DNMT or TET enzyme activity, similar to those typically performed in “mechanistic” studies through pharmacological or genetic manipulation, does not seem appropriate here. As demonstrated by studies described above, changes in DNA modification are a complex set of hyper- and hypomethylation at

specific genomic loci across the genome. The functional role of these patterns will almost certainly not be mechanistically unraveled by simply increasing or decreasing overall enzymatic activity and/or expression. An alternative approach would be to enlist the current generation of genome editing tools (Kim and Kim 2014) to modify the epigenome at specific genomic locations. Recent reviews provide a good overview of this potential paradigm shift (Voigt and Reinberg 2013; Thakore et al. 2016; Stricker et al. 2017; Willyard 2017). Briefly, the concept is to use tools like CRISPR (Vojta et al. 2016), zinc fingers (Kungulovski et al. 2015) or transcription activator-like effector (TALES) (Maeder et al. 2013) to drive fusion proteins with DNMTs or TETs to a specific location or locations in the genome to alter modification status. This offers great promise to move epigenomic patterns to more mechanistic investigations (Stricker et al. 2017). These approaches still require more development and the concerns about the specificity of these targeting mechanisms will need to be addressed. Nonetheless, a pathway from discovery to intervention is now feasible and should be able to, in time, address many of the fundamental questions in epigenomics of aging.

Conclusion

This review is intended to provide an overview and primer on analysis of DNA modifications in geroscience studies, especially for those new to critically analyzing or performing these studies. With previous hypotheses of the epigenetic contribution to aging already being refuted (Unnikrishnan et al. 2017a) and new era of understanding the epigenetics of aging is likely upon us (Lopez-Leon and Goya 2017). With a full understanding of how these new analytical techniques work, their strengths and limitation, and possible future directions, geroscientists will be equipped to understand and participate in these studies.

Acknowledgements The authors thank Donald Dunn for assistance with figure generation.

Funding information This work was supported the Donald W. Reynolds Foundation, the Oklahoma Nathan Shock Center of Excellence in the Biology of Aging Targeted DNA Methylation and Mitochondrial Heteroplasmy Core (P30AG050911), the National Institute on Aging (R01AG026607, F31AG038285,

T32AG052363, K99AG051661), the National Eye Institute (R01EY021716, R21EY024520, T32EY023202), and the Oklahoma Center for Advancement of Science and Technology (HR14-174).

References

- Adey A, Shendure J (2012) Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. *Genome Res* 22:1139–1143
- Akalin A et al (2012a) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* 13:R87
- Akalin A et al (2012b) Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet* 8:e1002781
- Allis CD, Jenuwein T (2016) The molecular hallmarks of epigenetic control. *Nat Rev Genet* 17:487–500
- Allum F et al (2015) Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. *Nat Commun* 6:7211
- Almeida RD et al (2012) Semi-quantitative immunohistochemical detection of 5-hydroxymethyl-cytosine reveals conservation of its tissue distribution between amphibians and mammals. *Epigenetics* 7:137–140
- Angermueller C et al (2016) Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* 13:229–232
- Baubec T, Akalin A (2016) Genome-wide analysis of DNA methylation patterns by high-throughput sequencing. In: Aransay AM, Lavín Trueba JL (eds) *Field guidelines for genetic experimental designs in high-throughput sequencing*. Springer International Publishing, Cham, pp 197–221
- Benayoun BA et al (2015) Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat Rev Mol Cell Biol* 16:593–610
- Bernstein DL et al (2015) The BisPCR(2) method for targeted bisulfite sequencing. *Epigenetics Chromatin* 8:27
- Bestor TH et al (2015) Notes on the role of dynamic DNA methylation in mammalian development. *Proc Natl Acad Sci U S A* 112:6796–6799
- Bhat S et al (2016) DNA methylation detection at single base resolution using targeted next generation bisulfite sequencing and cross validation using capillary sequencing. *Gene* 594:259–267
- Bibikova M et al (2009) Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics* 1:177–200
- Bibikova M et al (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98:288–295
- Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* 8:1499–1504
- Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16:6–21
- Blueprint consortium (2016) Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nat Biotechnol* 34:726–737
- Bock C (2012) Analysing and interpreting DNA methylation data. *Nat Rev Genet* 13:705–719

- Booth MJ et al (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336:934–937
- Booth MJ et al (2013) Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat Protoc* 8:1841–1851
- Booth MJ et al (2014) Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat Chem* 6:435–440
- Burgess DJ (2017) Epigenetics: rich pore methods for DNA methylation detection. *Nat Rev Genet* 18:209
- Cedar H et al (1979) Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic Acids Res* 6:2125–2132
- Chapman CG et al (2015) TET-catalyzed 5-hydroxymethylcytosine regulates gene expression in differentiating colonocytes and colon cancer. *Sci Rep* 5:17568
- Chen PY et al (2010) BS Seeker: precise mapping for bisulfite sequencing. *BMC Bioinformatics* 11:203
- Chen BH et al (2016) DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* 8:1844–1865
- Chen GG et al (2017) Medium throughput bisulfite sequencing for accurate detection of 5-methylcytosine and 5-hydroxymethylcytosine. *BMC Genomics* 18:96
- Christensen BC et al (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 5:e1000602
- Clark SJ et al (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* 22:2990–2997
- Clark SJ et al (2016) Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol* 17:72
- Clark SJ et al (2017) Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat Protoc* 12:534–547
- Cokus SJ et al (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452:215–219
- Cole JJ et al (2017) Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol* 18:58
- Crary-Dooley FK et al (2017) A comparison of existing global DNA methylation assays to low-coverage whole-genome bisulfite sequencing for epidemiological studies. *Epigenetics* 12:206–214
- de Koning AP et al (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 7:e1002384
- Dozmorov MG (2017) Epigenomic annotation-based interpretation of genomic data: from enrichment analysis to machine learning. *Bioinformatics* 33:3323–3330
- Dupont JM et al (2004) De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal Biochem* 333:119–127
- Edgar R et al (2014) Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. *Epigenetics Chromatin* 7:28
- Franzen J et al (2017) Senescence-associated DNA methylation is stochastically acquired in subpopulations of mesenchymal stem cells. *Aging Cell* 16:183–191
- Frommer M et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* 89:1827–1831
- Fuke C et al (2004) Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann Hum Genet* 68:196–204
- Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* 13:840–852
- Garrett-Bakelman FE et al. (2015) Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. *J Vis Exp*:e52246
- Globisch D et al (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 5:e15367
- Gravina S et al (2015) Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. *Nucleic Acids Res* 43:e93
- Gravina S et al (2016) Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol* 17:150
- Gu H et al (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc* 6:468–481
- Hadad N et al (2016) Absence of genomic hypomethylation or regulation of cytosine-modifying enzymes with aging in male and female mice. *Epigenetics Chromatin* 9:30
- Hadad N et al (2017a) Caloric-restriction attenuates age-associated alterations in CG and non-CG methylation in the old brain. *BioRxiv*. <https://doi.org/10.1101/175810>
- Hadad N et al. (2017b) Caloric restriction mitigates age-associated hippocampal differential CG and non-CG methylation. *bioRxiv*
- Hadad N et al (2017c) Caloric-restriction attenuates age-associated alterations in CG and non-CG methylation in the old brain. *BioRxiv*: <https://doi.org/10.1101/175810>
- Hahn O et al (2017) Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. *Genome Biol* 18:56
- Hannum G et al (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 49:359–367
- Hansen KD et al (2012) BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* 13:R83
- Harris RA et al (2010) Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol* 28:1097–1105
- He Y, Ecker JR (2015) Non-CG methylation in the human genome. *Annu Rev Genomics Hum Genet* 16:55–77
- He YF et al (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333:1303–1307
- Hing B et al (2015) Adaptation of the targeted capture methyl-Seq platform for the mouse genome identifies novel tissue-specific DNA methylation patterns of genes involved in neurodevelopment. *Epigenetics* 10:581–596
- Holliday R (2006) Epigenetics: a historical overview. *Epigenetics* 1:76–80

- Horvath S (2013) DNA methylation age of human tissues and cell types. *Genome Biol* 14:R115
- Hotchkiss RD (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem* 175:315–332
- Houseman EA et al (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13:86
- Hu L et al (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* 155:1545–1555
- Ito S et al (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333:1300–1303
- Ivanov M et al (2013) In-solution hybrid capture of bisulfite-converted DNA for targeted bisulfite sequencing of 174 ADME genes. *Nucleic Acids Res* 41:e72
- Jain M et al. (2017) Nanopore sequencing and assembly of a human genome with ultra-long reads. *bioRxiv*
- Jeck WR et al (2012) Review: a meta-analysis of GWAS and age-associated diseases. *Aging Cell* 11:727–731
- Jones MJ et al. (2015) DNA methylation and healthy human aging. *Aging Cell*
- Jorgensen HF et al (2006) Engineering a high-affinity methyl-CpG-binding protein. *Nucleic Acids Res* 34:e96
- Kalani A et al (2014) Nutri-epigenetics ameliorates blood-brain barrier damage and neurodegeneration in hyperhomocysteinemia: role of folic acid. *J Mol Neurosci* 52:202–215
- Karlin S, Mrázek J (1997) Compositional differences within and between eukaryotic genomes. *Proc Natl Acad Sci* 94:10227–10232
- Kaushal A et al (2017) Comparison of different cell type correction methods for genome-scale epigenetics studies. *BMC Bioinformatics* 18:216
- Kennedy BK et al (2014) GeroScience: linking aging to chronic disease. *Cell* 159:709–713
- Khanna A et al (2013) EpiGnome[trade] methyl-Seq kit: a novel post-bisulfite conversion library prep method for methylation analysis. *Nat Methods* 10
- Kim H, Kim JS (2014) A guide to genome engineering with programmable nucleases. *Nat Rev Genet* 15:321–334
- Kim S et al (2017) The frailty index outperforms DNA methylation age and its derivatives as an indicator of biological age. *GeroScience* 39:83–92
- Kinde B et al (2015) Reading the unique DNA methylation landscape of the brain: non-CpG methylation, hydroxymethylation, and MeCP2. *Proc Natl Acad Sci U S A* 112:6800–6806
- King GD et al (2012) Promoter methylation and age-related downregulation of Klotho in rhesus monkey. *Age (Dordr)* 34:1405–1419
- Kishore K et al (2015) methylPipe and compEpiTools: a suite of R packages for the integrative analysis of epigenomics data. *BMC Bioinformatics* 16:313
- Klus P et al (2012) BarraCUDA—a fast short read sequence aligner using graphics processing units. *BMC Res Notes* 5:27
- Komori HK et al (2011) Application of microdroplet PCR for large-scale targeted bisulfite sequencing. *Genome Res* 21:1738–1745
- Koster J, Rahmann S (2014) Massively parallel read mapping on GPUs with the q-group index and PEANUT. *PeerJ* 2:e606
- Kozlenkov A et al (2017) DNA methylation profiling of human prefrontal cortex neurons in heroin users shows significant difference between genomic contexts of hyper- and hypomethylation and a younger epigenetic age. *Genes (Basel)* 8
- Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324:929–930
- Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for bisulfite-Seq applications. *Bioinformatics* 27:1571–1572
- Krueger F et al (2012) DNA methylome analysis using short bisulfite sequencing data. *Nat Methods* 9:145–151
- Kungulovski G et al (2015) Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. *Epigenetics Chromatin* 8:12
- Laird PW (2010) Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 11:191–203
- Lander ES et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- Li Q et al (2015) Post-conversion targeted capture of modified cytosines in mammalian and plant genomes. *Nucleic Acids Res* 43:e81
- Lister R, Ecker JR (2009) Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res* 19:959–966
- Lister R, Mukamel EA (2015) Turning over DNA methylation in the mind. *Front Neurosci* 9:252
- Lister R et al (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133:523–536
- Lister R et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322
- Lister R et al (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* 341:1237905
- Logue MW et al. (2017) The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. *Epigenomics*
- Lopez-Leon M, Goya RG (2017) The emerging view of aging as a reversible epigenetic process. *Gerontology* 63:426–431
- Lopez-Otin C et al (2013) The hallmarks of aging. *Cell* 153:1194–1217
- Lou S et al (2014) Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. *Genome Biol* 15:408
- Lu X et al (2013) Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *J Am Chem Soc* 135:9315–9317
- Luo C et al (2017) Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* 357:600–604
- Maeder ML et al (2013) Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* 31:1137–1142
- Maegawa S et al (2010) Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res* 20:332–340
- Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286:35334–35338

- Mamanova L et al (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7:111–118
- Manconi A et al (2014) GPU-BSM: a GPU-based tool to map bisulfite-treated reads. *PLoS One* 9:e97277
- Mangold CA et al (2017a) CNS-wide sexually dimorphic induction of the major histocompatibility complex 1 pathway with aging. *J Gerontol A Biol Sci Med Sci* 72:16–29
- Mangold CA et al (2017b) Sexually divergent induction of microglial-associated neuroinflammation with hippocampal aging. *J Neuroinflammation* 14:141
- Marioni RE et al (2015) DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* 16:25
- Marioni RE et al. (2016) The epigenetic clock and telomere length are independently associated with chronological age and mortality. *Int J Epidemiol*
- Martinowich K et al (2003) DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science* 302:890–893
- Masser DR et al (2013) Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenetics Chromatin* 6:33
- Masser DR et al. (2015) Targeted DNA methylation analysis by next-generation sequencing. *J Vis Exp*
- Masser DR et al (2016) Bisulfite oligonucleotide-capture sequencing for targeted base- and strand-specific absolute 5-methylcytosine quantitation. *Age (Dordr)* 38:49
- Masser DR et al. (2017a) Sexually divergent DNA methylation patterns with hippocampal aging. *Aging Cell*
- Masser DR et al. (2017b) Sexually divergent DNA methylation programs with hippocampal aging. *bioRxiv* 161752
- Masser DR et al. (2017c) Sexually divergent DNA methylation patterns with hippocampal aging. *Aging Cell*:Preprint from *BioRxv*. doi: <https://doi.org/10.1101/161752>
- Mei Y et al (2015) Aging-associated formaldehyde-induced norepinephrine deficiency contributes to age-related memory decline. *Aging Cell* 14:659–668
- Meissner A et al (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 33:5868–5877
- Miura F et al (2012) Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res* 40:e136
- Mo A et al (2015) Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* 86:1369–1384
- Mo A et al (2016) Epigenomic landscapes of retinal rods and cones. *elife* 5:e11613
- Moore LD et al (2013) DNA methylation and its basic function. *Neuropsychopharmacology* 38:23–38
- Moran S et al (2016) Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* 8:389–399
- Mulqueen RM et al. (2017) Scalable and efficient single-cell DNA methylation sequencing by combinatorial indexing. *bioRxiv*
- Neri F et al (2016) Methylation-assisted bisulfite sequencing to simultaneously map 5fC and 5caC on a genome-wide scale for DNA demethylation analysis. *Nat Protoc* 11:1191–1205
- Nestor CE et al (2012) Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res* 22:467–477
- Noer A et al (2007) Dynamics of adipogenic promoter DNA methylation during clonal culture of human adipose stem cells to senescence. *BMC Cell Biol* 8:18
- Okano M et al (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247–257
- Olova N et al. (2017) Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *bioRxiv*
- Pacheco SE et al (2011) Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. *PLoS One* 6:e20280
- Park Y, Wu H (2016) Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* 32:1446–1453
- Parrish RR et al. (2012) Direct bisulfite sequencing for examination of DNA methylation with gene and nucleotide resolution from brain tissues. *Curr Protoc Neurosci Chapter 7:Unit 7 24*
- Petkovich DA et al (2017) Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell Metab* 25:954–960 e956
- Raiber E-A et al (2017) Mapping and elucidating the function of modified bases in DNA. *Nat Rev Chem* 1:0069
- Rakyan VK et al (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20:434–439
- Rand AC et al (2017) Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* 14:411–413
- Rauch T, Pfeifer GP (2005) Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Lab Investig* 85:1172–1180
- Reddy PM, Reddy PR (1990) Effect of prolactin on DNA methylation in the liver and kidney of rat. *Mol Cell Biochem* 95:43–47
- Reed K et al (2010) Comparison of bisulfite sequencing PCR with pyrosequencing for measuring differences in DNA methylation. *Anal Biochem* 397:96–106
- Reilly JG et al (1982) DNA methylation in mouse cells in culture as measured by restriction enzymes. *Biochim Biophys Acta* 697:53–59
- Schatz MC (2017) Nanopore sequencing meets epigenetics. *Nat Methods* 14:347–348
- Schubeler D (2015) Function and information content of DNA methylation. *Nature* 517:321–326
- Schultz MD et al (2015) Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523:212–216
- Sen P et al (2016) Epigenetic mechanisms of longevity and aging. *Cell* 166:822–839
- Serre D et al (2010) MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res* 38:391–399
- Shafi A et al. (2017) A survey of the approaches for identifying differential methylation using bisulfite sequencing data. *Brief Bioinform*
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26:1135–1145
- Shendure J et al (2017) DNA sequencing at 40: past, present and future. *Nature* 550:345–353

- Shi DQ et al (2017) New insights into 5hmC DNA modification: generation, distribution and function. *Front Genet* 8:100
- Simpkin AJ et al (2017) Are objective measures of physical capability related to accelerated epigenetic age? Findings from a British birth cohort. *BMJ Open* 7:e016708
- Simpson JT et al (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* 14:407–410
- Song L et al (2005) Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. *Anal Chem* 77:504–510
- Spruijt CG et al (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152:1146–1159
- Stricker SH et al (2017) From profiles to function in epigenomics. *Nat Rev Genet* 18:51–66
- Stubbs TM et al (2017) Multi-tissue DNA methylation age predictor in mouse. *Genome Biol* 18:68
- Sun Z et al (2015) Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis. *Epigenomics* 7:813–828
- Szulwach KE et al (2011) 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci* 14:1607–1616
- Tan L et al (2013) Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic Acids Res* 41:e84
- Teschendorff AE, Relton CL (2017) Statistical and integrative system-level analysis of DNA methylation data. *Nat Rev Genet*
- Thakore PI et al (2016) Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat Methods* 13:127–137
- Tippmann S (2015) Programming tools: adventures with R. *Nature* 517:109–110
- Unnikrishnan A et al. (2017a) Revisiting the genomic hypomethylation hypothesis of aging. *Annals of the New York Academy of Science* In press
- Unnikrishnan A et al (2017b) Role of DNA methylation in the dietary restriction mediated cellular memory. *Geroscience* 39:331–345
- Valdes AM et al (2013) Omics technologies and the study of human ageing. *Nat Rev Genet* 14:601–607
- Vanyushin BF et al (1973) The 5-methylcytosine in DNA of rats. Tissue and age specificity and the changes induced by hydrocortisone and other agents. *Gerontologia* 19:138–152
- Varley KE, Mitra RD (2010) Bisulfite patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. *Genome Res* 20:1279–1287
- Voigt P, Reinberg D (2013) Epigenome editing. *Nat Biotechnol* 31:1097–1099
- Vojta A et al (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* 44:5615–5628
- Waddington CH (1940) *Organisers & genes*. The University Press, Cambridge
- Wang J et al (2011) High resolution profiling of human exon methylation by liquid hybridization capture-based bisulfite sequencing. *BMC Genomics* 12:597
- Wang T et al (2017) Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol* 18:57
- Weber M et al (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37:853–862
- Weidner CI et al (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol* 15:R24
- Willyard C (2017) The epigenome editors: how tools such as CRISPR offer new details about epigenetics. *Nat Med* 23:900–903
- Wilson VL, Jones PA (1983) DNA methylation decreases in aging but not in immortal cells. *Science* 220:1055–1057
- Xi Y, Li W (2009) BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinformatics* 10:232
- Xu Y et al (2011) Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell* 42:451–464
- Yang AS et al (2004) A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 32:e38
- Yao B et al (2017) DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. *Nat Commun* 8:1122
- Yu M et al (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149:1368–1380
- Yu B et al. (2017) Genome-wide, single-cell DNA methylomics reveals increased non-CpG methylation during human oocyte maturation. *Stem Cell Rep*
- Zampieri M et al (2015) Reconfiguration of DNA methylation in aging. *Mech Ageing Dev* 151:60–70
- Zhang Y et al (2009) DNA methylation analysis by bisulfite conversion, cloning, and sequencing of individual clones. *Methods Mol Biol* 507:177–187
- Zhao L et al (2014) The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. *Genome Res* 24:1296–1307
- Zheng SC et al (2017) Correcting for cell-type heterogeneity in epigenome-wide association studies: revisiting previous analyses. *Nat Methods* 14:216–217
- Ziller MJ et al (2015) Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat Methods* 12:230–232 231 p following 232
- Ziller MJ et al (2016) Targeted bisulfite sequencing of the dynamic DNA methylome. *Epigenetics Chromatin* 9:55
- Zou J et al (2014) Epigenome-wide association studies without the need for cell-type composition. *Nat Methods* 11:309–311
- Zykovich A et al (2014) Genome-wide DNA methylation changes with age in disease-free human skeletal muscle. *Aging Cell* 13:360–366