




Short history of 5-methylcytosine: from discovery to clinical applications

Olga Taryma-Lesniak , Katarzyna Ewa Sokolowska ,
Tomasz Kazimierz Wojdacz 

Independent Clinical Epigenetics Laboratory, Pomeranian Medical University in Szczecin, Szczecin, Zachodniopomorskie, Poland

Correspondence to

Dr Tomasz Kazimierz Wojdacz, Independent Clinical Epigenetics Laboratory, Pomeranian Medical University in Szczecin, 71-252 Szczecin, Zachodniopomorskie, Poland; tomasz.wojdacz@pum.edu.pl

Received 4 July 2020

Revised 22 December 2020

Accepted 27 December 2020

Published Online First

11 January 2021

ABSTRACT

Covalent modifications of nucleotides in genetic material have been known from the beginning of the last century. Currently, one of those modifications referred to as DNA methylation, is impacting personalised medicine both as a treatment target and a biomarker source for clinical disease management. In this short review, we describe landmark discoveries that led to the elucidation of the DNA methylation importance in the cell's physiology and clarification of its role as one of the major processes in disease pathology. We also describe turning points in the development of methodologies to study this modification, which ultimately resulted in the development of in-vitro diagnostic kits targeting disease related DNA methylation changes as biomarkers.

COVALENT MODIFICATIONS OF GENETIC MATERIAL

In the 1940s, nucleotides which are the building blocks of the genetic material were shown to contain covalent modifications, such as methyl group at the carbon 5 position on the pyrimidine ring of cytosines referred to as DNA methylation.¹ In 1964, P R Srinivasan and E Borek, while studying the biological role of this phenomenon, suggested that this modification may act to protect DNA from enzymes (figure 1 represents a timeline schematic of the discoveries described in this short review).² In the subsequent years, DNA modifications were identified to play a central role in the Restriction and Modification system used by bacteria as a defence against foreign DNA.^{3 4} In this system, the nucleic acid with methyl groups is protected from cutting by endonucleases as opposed to non-modified nucleic acid. These discoveries led R Holliday and J E Pugh in 1975 to suggest that it is not only methylation sensitive endonucleases, there are also other proteins that can distinguish covalently modified nucleotides present in regulatory regions, such as gene promoters. Here, these proteins regulate the expression of genes with specific covalent modifications in the promoter.⁵ Also in 1975, but independently A D Riggs suggested that DNA methylation may take part in X chromosome inactivation in eukaryotes.⁶ Nevertheless, at that time none of the above authors had experimental evidence proving the claims they were putting forward.

Along with the discoveries leading to elaboration of the function of covalent modifications of genetic material, the term epigenetics was proposed to describe all mechanisms involved in the development of a phenotype from genotype. In 1979,

R Holliday suggested that “the specific methylation of bases such as adenine or cytosine could provide epigenetic switches in gene activity”⁷ and from that point on the field began to recognise covalent modifications of DNA as a mechanism of epigenetic gene expression regulation. Today, after years of research, several different covalent modifications of genetic material have been discovered.

The most frequent covalent modification of nucleic DNA in mammals is the addition of methyl group to the 5th atom in the 6-atom ring, counting counterclockwise from the NH nitrogen at the 6 o'clock position. In the human genome, methylated cytosines (5-methylcytosine (5-mC)) account for about 4% of all cytosines and are frequently referred to as the 5th base.⁸ In general terms, genes with methylated cytosines in the promoter are not transcriptionally active. However, a substantial volume of research was needed to elucidate the function of this modification in gene expression regulation and we will briefly review these discoveries below.

THE 5TH BASE

Methylated cytosine (5-methylcytosine (5-mC)) was synthesised in 1904 by H L Wheeler and T B Johnson, who also showed that 5-mC has very similar chemical properties to cytosine.⁹ The presence of 5-mC in nucleic acid was first described by T B Johnson and R D Coghill in 1925,¹⁰ who obtained a crystallised form of cytosine and 5-mC from hydrolysed nucleic acid of tubercle bacillus and were able to distinguish picrate salt crystals of cytosine from 5-cytosine under a microscope with polarised light. Despite the fact that the paper describing those findings was widely criticised and other researchers failed to reproduce those results,¹¹ this publication was significant as it accelerated the progress in the field. Further evidence for the presence of 5-mC in the genetic material was described in 1948 by R D Hotchkiss, who noticed an additional spot on paper chromatographs representing the 5th base, while studying the components of nucleic acids obtained from calf thymus.¹ R D Hotchkiss also named the 5th base ‘epi-cytosine’. Subsequently, G R Wyatt showed that apart from calf thymus, 5-mC is present in DNA of several other species, including: beef spleen, ram sperm, herring sperm, locusts (whole) and wheat germ.¹² The precise localisation of the 5-mC within the context of nucleotides in DNA was described in 1968 by P Grippo *et al.* This group, used DNase digest of the genetic material isolated from sea urchin embryos to show that the distribution of 5-mC in the DNA of this invertebrate



© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Taryma-Lesniak O, Sokolowska KE, Wojdacz TK. *J Clin Pathol* 2021;**74**:692–696.

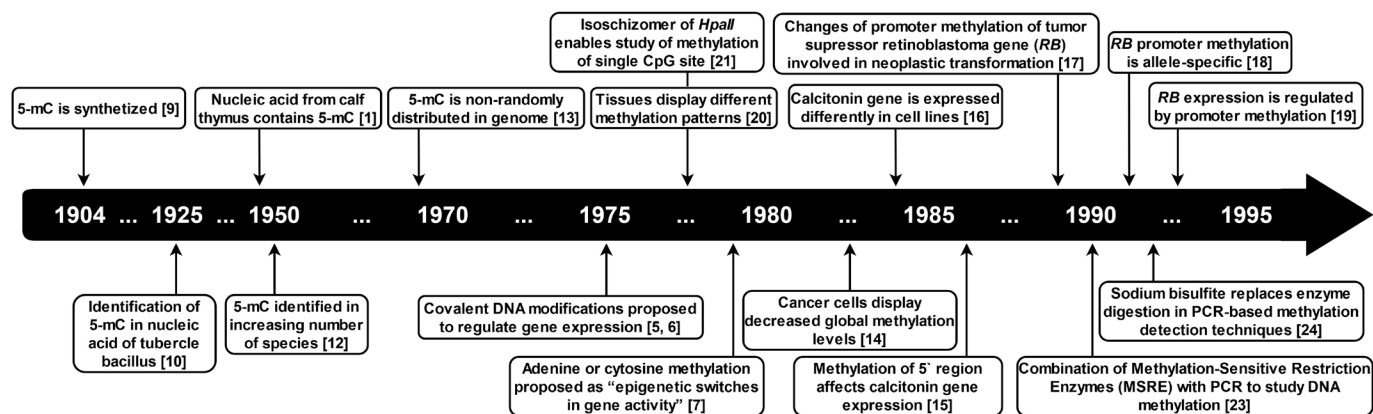


Figure 1 The timeline with schematic representation of discoveries described in this short review. 5mC, 5-methylcytosine.

is not random and 90% of 5-mC occurs within CpG dinucleotide separated from the digest.¹³

Subsequent analyses of the methylation within CpG sites showed that the methylation status of those dinucleotides differs between tissues and cells¹⁴ and CpG sites frequently group to the genomic regions with higher than expected concentration of CpG sites - CpG islands (CGI).¹⁵ With the advent of genome wide screening technologies both CGIs and single CpG sites outside of the CGIs were mapped to specific functional parts of the genomes, and 60% of human genes have been shown to have promoters embedded in CGI. This makes DNA methylation a potential mechanism of expression regulation for over half of the human genes.^{16 17}

DNA METHYLATION IN DISEASE

Studies of the significance of DNA methylation in disease started to rapidly develop from the publication by A P Feinberg and B Vogelstein in 1983. Using methylation-sensitive restriction enzymes, they showed that the global level of 5-mC in DNA from human cancer was significantly lower than in the DNA from the tissue that cancer originated from.¹⁸ Subsequently, S B Baylin *et al*, reported that the 5' region of the calcitonin gene had a different number of methylated CpG sites in different cell lines.¹⁹ Those findings, together with the results of earlier experiments, in which they showed that the same cell lines had different levels of calcitonin gene expression,²⁰ lead S B Baylin *et al* to suggest that methylation of a gene promoter regulates gene expression.¹⁹ The above discoveries paved the way for further studies on the role of DNA methylation in regulation of genes involved in oncogenesis.

The retinoblastoma suppressor gene (*RB*), considered at the time to be a tumour suppressor prototype, was the first gene for which the direct correlation between DNA methylation and gene expression was established. In 1989, V Greger *et al* while studying retinoblastoma tumours found that one of the patients, who was unilaterally affected and did not have familial history of retinoblastoma, harboured methylation of the *RB* gene promoter. With those observations, the authors speculated that hypermethylation of the gene could be involved in human neoplasia development and hypomethylation can result in spontaneous regression of tumours.²¹ Subsequently, in 1991, T Sakai *et al* confirmed V Greger's results in a study performed on a larger number of retinoblastoma patients and additionally were able to demonstrate allele-specific methylation of the *RB* gene. Moreover, as the patient they studied with allele-specific *RB* gene methylation, did not harbour any mutation within the gene, they speculated that methylation alone could lead to cancer development.²² However, the direct evidence that

hypermethylation of the *RB* gene promoter results in the gene inactivation was published in 1993 by N Ohtani-Fujita *et al*, who in the study performed on a neuroblastoma cell line showed that cells harbouring gene constructs with methylated *RB* promoter had different expression levels of the gene.²³ The above, were the pivotal discoveries that in principle initiated and accelerated studies of the involvement of the methylation dependent gene expression regulation in disease.

THE METHODOLOGY TO STUDY THE 5TH BASE

The functional studies of DNA methylation accelerated with the development of methodologies facilitating differentiation of patterns of DNA methylation between various nucleic acid sources. This was initiated by A P Bird and E M Southern, who in 1978 adapted the previously described bacterial restriction enzymes⁴ to study DNA methylation in eukaryotes. Using DNA from two different tissues of vertebrate *Xenopus laevis*, they showed that some enzymes are able to cleave the DNA at non-methylated CpG sites, while DNA with methyl group attached to cytosine at the same CpG sites were resistant to digestion. Thus, the digestion of DNA with this type of enzymes resulted in methylation-specific DNA fragmentation.²⁴ Another milestone in development of methodologies to study methylation patterns was discovery by C Waalwijk and R A Flavell, who also in 1978 showed that the enzyme *MspI* (an isoschizomer of *HpaII*), is capable of cleaving both unmethylated and methylated DNA, while *HpaII* cleaves only the unmethylated sites. This enabled the study of methylation status of single CpG dinucleotide.²⁵ Using those enzymes C Waalwijk and R A Flavell showed that various tissues differ in methylation status at specific CpG sites.²⁶ However, studies of DNA methylation biological function based on methylation-sensitive restriction enzymes digestion (MSRE) followed by Southern blot hybridisation, required large amounts of DNA, cannot detect methylation if a CpG site is methylated in only a few per cent of alleles in the sample and can only assess methylation within the enzyme's recognition sites. A significant improvement of this approach was the combination of MSRE with PCR first used in 1990 by J Singer-Sam *et al*. In this approach a CpG site is flanked with PCR primers and presence or absence of the PCR product from the amplification of the DNA digest indicates the methylation status of that site.²⁷ Although this combination increased the sensitivity of methylation detection, the limitation of the technology to assess only CpG sites within methylation-sensitive restriction enzymes recognition sequence, remained.

The breakthrough in development of technologies to study DNA methylation was the combination of sodium bisulfite modification

of DNA with PCR amplification.²⁸ Sodium bisulfite deaminates unmethylated cytosines to uracil, while methylated cytosines are resistant to that modification.²⁹ Thus, treatment of the DNA with sodium bisulfite preserves the methylation status of the cytosines in a sequence of interest during PCR amplification, as in the PCR product amplified from the bisulfite modified template all non-methylated cytosines are amplified as thymine (or uracil) and methylated ones as cytosines.

One of the first and a landmark method that takes advantage of bisulfite conversion was methylation-specific PCR published by J G Herman *et al* in 1996.³⁰ This method requires PCR primers designed to contain as many as possible CpG sites within primer. Those primers bind only to the template where CpG sites were methylated before bisulfite modification and amplify only fully methylated template. The disadvantage of this method is that another primer set needs to be designed and PCR performed to confirm non-methylated status of the screened locus. To overcome this limitation a second primer design approach was proposed where primers are targeted to the parts of the template devoid of CpG sites.^{31 32} That allows amplification of the locus of interest regardless of its methylation status and investigation of methylation status of that locus in post PCR manner with technologies such as sequencing or melting. A review of technologies for methylation studies is out of scope for this publication (more detailed information can be found in ³³).

The PCR based technologies are indispensable in the assessment of methylation within specific loci. However, with technological advances the assessment of methylation at the genome level has become available. Microarray technologies first delivered the possibility to screen for methylation at thousands of loci, for example, all the promoters in single experiment with the oligonucleotide resolution. Subsequently, with the advent of next generation sequencing (NGS) we are currently able to obtain global maps of methylation within whole genomes at single nucleotide resolution. It is worth mentioning the enormous progress in the development of NGS for the study of methylation patterns. With the use of NGS, within a few years we went from being able to study methylation of single molecules in the pool of the genomes extracted from a tissue or a

mix of tissues using ultra-deep bisulfite sequencing,^{34–36} to study methylomes of the single cell.^{37–39}

METHYLATION BIOMARKERS IN THE IN-VITRO DIAGNOSTIC SETTING

The disease-related methylation changes involving single CpG sites as well as a region containing a number of CpG sites (eg, promoter CGIs) can be considered methylation biomarkers. The diagnostic utility of methylation alterations at single CpG sites is still to be elucidated. Though, with increasing research data indicating that a methylation of a single CpG can interfere with the binding of, for example, SP1 transcription factor⁴⁰ or enhancers⁴¹ it is likely that methylation changes at single CpG sites will become clinically useful biomarkers. Currently however, the vast majority of the methylation biomarkers used in clinical practice are regional methylation changes in gene promoters that have been associated with clinical outcomes. The regional methylation changes are likely to remain the best methylation biomarker candidates. Mainly because, studies of the distribution of methylation across the genome show that the short distance between consecutive CpG sites is the best determinant of methylation status and the cytosines in the clusters of CpG sites (CpG islands) are characterised by uniform methylation status.⁴² Consequently, PCR-based methods are currently the most frequently used technologies in methylation biomarker screening in diagnostic laboratories. Those techniques are labour and cost effective, along with being relatively easy to implement in diagnostic setting, as opposed to the genome-wide methylation screening technologies which still are technically complex and require scientific knowledge for result interpretation. Regardless of the technological challenges in measuring methylation changes it is no longer a question if methylation biomarkers can significantly contribute to all the aspects of the personalised patient care, including: risk assessment, disease detection, clinical disease management and with the increasing number of the disease becoming chronic, monitoring for the relapse.^{43–45}

The discussion of specific in-vitro diagnostic tests currently used or at the advanced stages of clinical validation for diagnostic

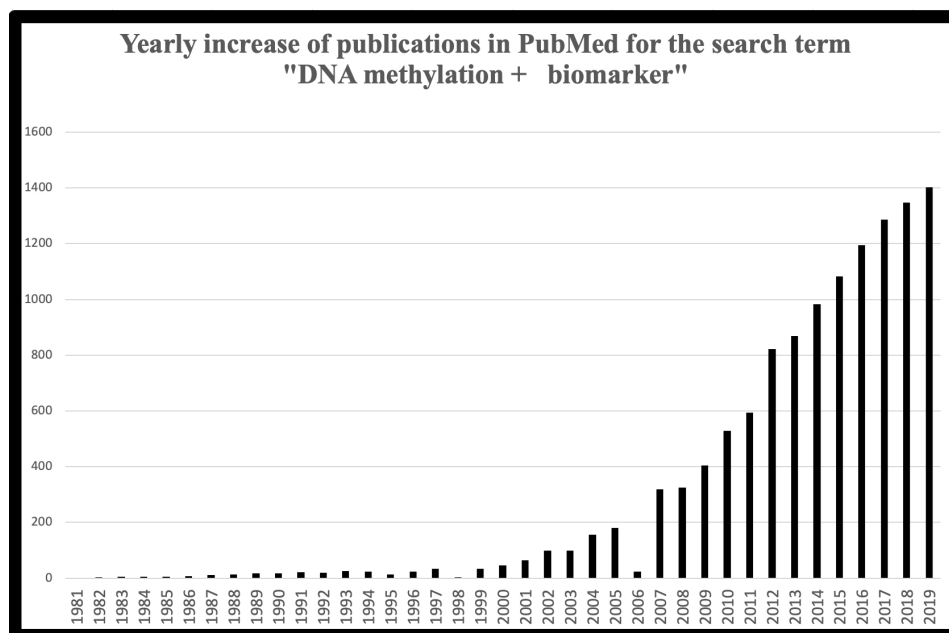


Figure 2 Increase in the number of publications per annum in the file of DNA methylation biomarkers (source PubMed).

use is out of scope for this short review, but the references to the studies reporting clinical validation data for the methylation biomarkers currently entering in-vitro diagnostics can be found in our recent review.⁴⁶

CURRENT STATE OF THE APPLICATION OF METHYLATION BIOMARKERS IN CLINICAL DISEASE MANAGEMENT

From the first evidence that methylation changes contribute to the disease phenotype, which also indicated that those changes can be used as biomarkers useful in clinical disease management, the number of the research publications indicating potential in-vitro diagnostic utility of disease-related methylation changes is growing exponentially (figure 2). Nevertheless, the use of disease-related methylation changes in in-vitro diagnostics is still lagging. This is mainly due to the fact that the introduction of biomarkers to in-vitro diagnostic testing is strictly regulated and requires systematic and comprehensive evidence of clinical validity for each biomarker test. This evidence can rarely be obtained in a research study, where potential disease-related biomarkers are discovered. Thus, the biggest challenge the field of methylation biomarkers is currently facing is to implement a systematic assessment of clinical validity of the large number of known disease-related methylation changes. Nevertheless, recent successes of the diagnostic test such as Cologuard (from Exact Sciences Corporation, USA) which is partly built on methylation biomarkers or Epi proLung (Epigenomics AG, Germany) which is solely built on methylation biomarkers, indicate that in the near future we will witness increasing use of methylation biomarkers in personalised medicine.

Take home messages

- ▶ Methylation of cytosine is the most common covalent modification of nucleotides in humans.
- ▶ DNA methylation regulates gene expression in healthy cells.
- ▶ Changes of gene methylation contribute to disease phenotype and can be used as biomarkers.
- ▶ Aberrant methylation is potentially reversible and thus can be targeted by therapy.

Handling editor Tahir S Pillay.

Contributors OTL and KES: performed the literature searches, drafted the manuscript. TKW: performed the literature searches, drafted the manuscript and coordinated editing of the manuscript.

Funding This work was financed by the Polish National Agency for Academic Exchange

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs

Olga Taryma-Lesniak <http://orcid.org/0000-0002-1161-5399>

Katarzyna Ewa Sokolowska <http://orcid.org/0000-0002-9250-5799>

Tomasz Kazimierz Wojdacz <http://orcid.org/0000-0001-6003-2432>

REFERENCES

- 1 HOTCHKISS RD. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem* 1948;175:315–32.
- 2 Srinivasan PR, BOREK E. Enzymatic alteration of nucleic acid structure. *Science* 1964;145:548–53.
- 3 Boyer HW. DNA restriction and modification mechanisms in bacteria. *Annu Rev Microbiol* 1971;25:153–76.
- 4 Meselson M, Yuan R, Heywood J. Restriction and modification of DNA. *Annu Rev Biochem* 1972;41:447–66.
- 5 Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. *Science* 1975;187:226–32.
- 6 Riggs AD. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* 1975;14:9–25.
- 7 Holliday R. A new theory of carcinogenesis. *Br J Cancer* 1979;40:513–22.
- 8 Breiling A, Lyko F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* 2015;8:24.
- 9 Wheeler HL. Researches on pyrimidine derivatives: 5-Methylcytosine. *Am Chem J* 1904;31:591–606.
- 10 Johnson TB. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle Bacillus. *J Am Chem Soc* 1925;47:2838–44.
- 11 Vischer E, Zamenhof S, CHARGAFF E. Microbial nucleic acids; the desoxyribose nucleic acids of avian tubercle bacilli and yeast. *J Biol Chem* 1949;177:429–38.
- 12 WYATT GR. Recognition and estimation of 5-methylcytosine in nucleic acids. *Biochem J* 1951;48:581–4.
- 13 Grippo P, Iaccarino M, Parisi E, et al. Methylation of DNA in developing sea urchin embryos. *J Mol Biol* 1968;36:195–208.
- 14 Ehrlich M, Gama-Sosa MA, Huang LH, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* 1982;10:2709–21.
- 15 Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *J Mol Biol* 1987;196:261–82.
- 16 Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* 2006;103:1412–7.
- 17 Jones PA. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat Rev Genet* 2012;13:484–92.
- 18 Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983;301:89–92.
- 19 Baylín SB, Höppner JW, de Bustros A, et al. DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. *Cancer Res* 1986;46:2917–22.
- 20 Nelkin BD, Rosenfeld KI, de Bustros A, et al. Structure and expression of a gene encoding human calcitonin and calcitonin gene related peptide. *Biochem Biophys Res Commun* 1984;123:648–55.
- 21 Greger V, Passarge E, Höppner W, et al. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum Genet* 1989;83:155–8.
- 22 Sakai T, Toguchida J, Ohtani N, et al. Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am J Hum Genet* 1991;48:880–8.
- 23 Ohtani-Fujita N, Fujita T, Aoike A, et al. CpG methylation inactivates the promoter activity of the human retinoblastoma tumor-suppressor gene. *Oncogene* 1993;8:1063–7.
- 24 Bird AP, Southern EM. Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from *Xenopus laevis*. *J Mol Biol* 1978;118:27–47.
- 25 Waalwijk C, Flavell RA, MspI Flavell RA. MspI, an isoschizomer of HpaII which cleaves both unmethylated and methylated HpaII sites. *Nucleic Acids Res* 1978;5:3231–6.
- 26 Waalwijk C, Flavell RA. Dna methylation at a CCGG sequence in the large intron of the rabbit beta-globin gene: tissue-specific variations. *Nucleic Acids Res* 1978;5:4631–42.
- 27 Singer-Sam J, Grant M, LeBon JM, et al. Use of a HpaII-polymerase chain reaction assay to study DNA methylation in the pgk-1 CpG island of mouse embryos at the time of X-chromosome inactivation. *Mol Cell Biol* 1990;10:4987–9.
- 28 Frommer M, McDonald LE, Millar DS, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* 1992;89:1827–31.
- 29 Wang RY, Gehrke CW, Ehrlich M. Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res* 1980;8:4777–90.
- 30 Herman JG, Graff JR, Myöhänen S, et al. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* 1996;93:9821–6.
- 31 Wojdacz TK, Hansen LL, Dobrovic A. A new approach to primer design for the control of PCR bias in methylation studies. *BMC Res Notes* 2008;1:54.
- 32 Clark SJ, Harrison J, Paul CL, et al. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* 1994;22:2990–7.
- 33 Kurdyukov S, Bullock M. *DNA methylation analysis: choosing the right method*, 2016.
- 34 Bock C, Tomazou EM, Brinkman AB, et al. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol* 2010;28:1106–14.
- 35 Harris RA, Wang T, Coarfa C, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol* 2010;28:1097–105.
- 36 Serre D, Lee BH, Ting AH. MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res* 2010;38:391–9.
- 37 Smallwood SA, Lee HJ, Angermueller C, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* 2014;11:817–20.
- 38 Karemaker ID, Vermeulen M. Single-cell DNA methylation profiling: technologies and biological applications. *Trends Biotechnol* 2018;36:952–65.

- 39 Gouil Q, Keniry A. Latest techniques to study DNA methylation. *Essays Biochem* 2019;63:639–48.
- 40 Douet V, Heller MB, Le Saux O. DNA methylation and Sp1 binding determine the tissue-specific transcriptional activity of the mouse Abcc6 promoter. *Biochem Biophys Res Commun* 2007;354:66–71.
- 41 Fleischer T, Tekpli X, Mathelier A, et al. DNA methylation at enhancers identifies distinct breast cancer lineages. *Nat Commun* 2017;8:1379.
- 42 Affinito O, Palumbo D, Fierro A, et al. Nucleotide distance influences co-methylation between nearby CpG sites. *Genomics* 2020;112:144–50.
- 43 Koch A, Joosten SC, Feng Z, et al. Analysis of DNA methylation in cancer: location revisited. *Nat Rev Clin Oncol* 2018;15:459–66.
- 44 Kim H, Wang X, Jin P. Developing DNA methylation-based diagnostic biomarkers. *J Genet Genomics* 2018;45:87–97.
- 45 Bailleux C, Lacroix L, Barranger E, et al. Using methylation signatures on cell-free DNA for early cancer detection: a new era in liquid biopsy? *Ann Oncol* 2020.1016/j.annonc.2020.03.307. [Epub ahead of print: 11 Apr 2020].
- 46 Taryma-Leśniak O, Sokolowska KE, Wojdacz TK. Current status of development of methylation biomarkers for in vitro diagnostic IVD applications. *Clin Epigenetics* 2020;12:100.