

# 3rd Progression Review

The Epigenomics of Human Ageing

Academic Unit - Human Development and Health

Supervisors - Chris Bell, Karen Lillycrop and Cyrus Cooper

*Richard J. Acton*

*2020-01-19*



# Contents

<b>Abstract</b>	<b>5</b>
<b>Acknowledgements</b>	<b>7</b>
<b>I Introduction</b>	<b>9</b>
<b>1 The Epigenomics of Human Ageing</b>	<b>11</b>
1.1 Epigenomics - Overview . . . . .	11
1.2 DNA Methylation . . . . .	12
1.3 Ageing . . . . .	20
<b>II Methods</b>	<b>37</b>
<b>2 Methods</b>	<b>39</b>
2.1 DNA methylation assays . . . . .	39
2.2 Discussion . . . . .	46
<b>III Results 1</b>	<b>49</b>
<b>3 Assessment of Umbilical Cord Tissue DNA Methylation with Respect to Bone Health, Environmental Factors and Ageing</b>	<b>51</b>
3.1 Abstract . . . . .	51
3.2 Introduction . . . . .	51
3.3 Methods . . . . .	53
3.4 Results . . . . .	71
3.5 Discussion . . . . .	76
<b>IV Results 2</b>	<b>79</b>
<b>4 The Genomic Loci of Specific Human tRNA Genes Exhibit Ageing-Related DNA Hypermethylation</b>	<b>81</b>
4.1 Abstract . . . . .	83

---

4.2 Introduction . . . . .	84
4.3 Results . . . . .	86
4.4 Discussion . . . . .	101
4.5 Methods . . . . .	103
Supplementary materials . . . . .	108
4.6 References . . . . .	116
<b>V Results 3</b>	<b>117</b>
<b>5 Alu Clock Paper Draft Introduction</b>	<b>119</b>
<b>6 Final Words</b>	<b>121</b>

# Abstract

Aims:

1. Identify Age related changes in DNA methylation in regions of the genome characterised in the TwinsUK MeDIP-seq dataset and poorly covered by or covered in small samples by other technologies.
  - a. The tRNA genes, following up on the previous finding of age-related DNA hypermethylation at tRNA-iMet-CAT-1-4 [Bell et al., 2016]. tRNA genes have a core role in cellular metabolism, many emerging regulatory functions both structural and as signalling molecules. tRNA genes also interface with many systems the modulation of which impact longevity, making any age-related changes in their epigenetic state potentially very consequential.
  - b. Alu repeat elements, a primate specific family of SINEs present in over 1 million copies in the human genome. We aim to construct a DNA methylation clock using only these elements. We reason that age acceleration based on this clock will yield different information about biological ageing than from previous clocks as the effects of DNA methylation changes differ between repressed repetitive elements and regulatory sequences.
2. Identify epigenetic associations with bone health outcomes and of vitamin D supplementation during pregnancy as an intervention to improve bone outcomes [Harvey et al., 2012, Cooper et al., 2016]. Provide some ground work for mechanistic studies to extend understanding what systems influence bone health by identifying what genomic features change epigenetically with bone traits and interventions intended to affect these traits.



# Acknowledgements

I would like to acknowledge Nevena Krstic for her work extracting the DNA for the MAVIDOS EPIC array analysis and Dr Millie Parsons for her assistance with the MAVIDOS sample metadata, as well as Dr Beth Curtis and Professor Nick Harvey for their assistance with framing the research questions for the vitamin D and bone development outcomes work with the MAVIDOS samples. I would like to acknowledge Nikki Graham for her assistance identifying suitable samples for targeted bisulfite sequencing work and her work extracting DNA from those samples. I acknowledge the use of the IRIDIS High-Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work. Finally, I would like to acknowledge the extensive help and support of my primary supervisor Dr Chris Bell.



# **Part I**

# **Introduction**



# Chapter 1

## The Epigenomics of Human Ageing

### 1.1 Epigenomics - Overview

Epigenetics generally refers to modifications to DNA and chromatin which do not affect the primary sequence of DNA bases, but which are to varying degrees stable and heritable. The term derives from epigenesis and genetics originating with Conrad Waddington [Waddington, 2012]. Epigenesis refers to the idea that organisms develop through the progressive differentiation of cells from the egg into adult tissues. As modern genetics revealed that every cell contained a complete copy of the genome which was differentially utilised by the cells of adult organisms, the terms were fused to reflect the study of how this occurs. The usage has evolved a little further since molecular biology began to elucidate the mechanisms involved in this process and the term now frequently refers to the study of these mechanisms and their effects in less explicitly developmental contexts. Arthur Riggs et al. considered that heritability should be a criterion for a mark to be considered epigenetic, but this excludes many phenomena now commonly referred to using the term [Russo et al., 1996]. Requiring heritability results in a further definitional dispute over degrees of heritability; mitotic, meiotic, transgenerational and to what degree of fidelity? Adrian Bird proposed the definition: “the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states.” as a useful compromise [Bird, 2007]. Epigenomics refers to the totality of the epigenetic modifications present in a particular cell, tissue type or genome.

Whilst an organism can generally be thought of as having a single genome, with exceptions such as Somatic mutations and Chimerism, it will have at least as many epigenomes as it has cell types. There are on the order of  $10^{13}$  cells in the human body [Bianconi et al., 2013]. Estimates of the number of cell types vary with resolution at which one deems cells functionally distinct [HumanCellAtlas, 2017]. At this point in time, however, there is no clear definition of what constitutes a distinct cell type, indeed they are beginning to be defined by the distinct patterns of gene expression and epigenetic modifications they exhibit [CellSystemsVoices, 2017]. Consequently, the number epigenomes that could be considered distinct will likely ultimately approximate to the number of epigenomes which can usefully distinguish between sub-populations of cells. A consortium has been established to produce a Human Cell Atlas [Regev et al., 2017] which aims to define all human cell types, and a cell type ontology [CellOntology, 2017] exists.

There are a number of epigenetic modifications which can be subdivided into four broad categories:

#### 1. DNA modifications

DNA methylation primarily takes place on Cytosine residues. Cytosine methylation occurs principally in an mCpG sequence context but can also occur in mCpH, mCHG and mCHH

(H=A/C/T) contexts; particularly in cells of the nervous system [Guo et al., 2014] and embryonic stem cells where as many as 25% of all cytosines can be methylated in non-canonical contexts. In contrast to methylation in a differentiated non-neuronal tissue such as foetal lung fibroblasts which is 99.98% in CpG context [Lister et al., 2009], [Schultz et al., 2015]. 5-methylcytosine (5mc) can be oxidised to produce another modified DNA base 5-hydroxymethylcytosine (5hmC). 5hmC may act as an intermediate to demethylation and potentially has regulatory functions in its own right, reviewed in [Pfeifer et al., 2013]. 5hmC can be oxidised further to formylcytosine (5fC) and carboxylcytosine (5caC). The extent to which these modifications are stable and functional is still being explored. Other DNA bases can be methylated such as N6-methyladenine, but this occurs at a substantially lower frequency than 5mC and much less is known about their potential functions [Wu et al., 2016].

## 2. Histone tail modifications

Histone proteins form disk-shaped octamers around which ~150bp of DNA can be wrapped to form a nucleosome. The ‘tails’ are generally the N termini of the histone proteins, outside of the core globular domains, which protrude from the nucleosome structure. Histone tails can be subject to post-translational modification which commonly takes place at lysine residues, other residues are subject to modifications but lysines are among the best characterised. Examples of modifications include methylations with between one and three methyl groups per lysine, Acetylations, Ubiquitylation, SUMOylations, and phosphorylations (Reviewed in Bannister and Kouzarides [2011]). Many of these marks can be generally classified as permissive or repressive but are frequently found in combinations of marks of opposing effect, rendering the interpretation of the ‘Histone Code’ extremely challenging [Voigt et al., 2012]. One approach to interpreting chromatin state is segmentation. Segmentation makes use of pattern discovery techniques to divide the genome into discrete sections assigning these sections to a set of categories which can then be compared to existing annotations to provide functional labels [Hoffman et al., 2012, Carrillo-de Santa-Pau et al. [2017]]. When successful the pattern discovery algorithm independently recapitulates our existing ontology of functional elements and hopefully provides new insights by for example highlighting regions not previously considered to be in a given functional category. This is a “top-down” approach making use of the “sum” of available chromatin state data rather than trying to interpret the significance of a single type of mark.

## 3. Histone Variants

Alternatives to the canonical histones can be substituted in the nucleosome, examples of such alternative histones include H2A.Z and H3.3. Alternative histones can alter chromatin structure and dynamics by altering nucleosome stability, binding different factors, and presenting a different substrate for histone modifications, thus altering their pattern. (reviewed in Weber and Henikoff [2014].)

## 4. Some non-coding RNAs (ncRNA)

A subset of ncRNAs, particularly those which persist in close association with chromatin can fall within the definition of epigenetic features. (reviewed in Kaikkonen et al. [2011].) An example of such a long non-coding RNA (lncRNA) is Xist which is a key regulator of X inactivation [Cerase et al., 2015].

# 1.2 DNA Methylation

## 1.2.1 Structure and Context

DNA methylation is the most well-studied epigenetic modification [Jenuwein, 2001]. Specifically CpG methylation, the addition of a Methyl group to the 5 carbon of a cytosine base (figure 1.1), in a CpG dinucleotide context (figure 1.2). (hereafter ‘DNA Methylation’ or ‘DNAm’ refers to 5mC CpG methylation unless stated otherwise.)

DNA methylation is stable and relatively easy to characterise from biological samples by comparison with other epigenetic marks, the methylation status of ancient DNA has even been characterised, see for example: DNAm profiles have been produced from DNA extracted from mammoth and neandertal remains [Briggs et al., 2010] and from native American remains ranging in age from 230 to 4500 years [Smith et al., 2015]. 5mC is quite chemically stable such that for most tissue sample preservation techniques for which DNA can be extracted from the sample DNAm can also be characterised. DNA extracted with normal laboratory methods can be assayed for DNAm. Histone modification status, on the other hand, is less robust against environmental stresses on samples and often requires considerably more laborious sample preparation to characterise [Bauden et al., 2017]. DNA methylation also offers the appearance of relative simplicity in comparison to the complex picture of the ‘histone code’ giving the impression that its useful interpretation may be more readily tractable. However, interaction between DNAm and histone modifications is well documented [Hashimshony et al., 2003, Estève et al. [2006], Rose and Klose [2014]] consequently, a fuller understanding of DNAm will likely require an understanding histone modifications and their potential to feedback on and thus affect DNAm.

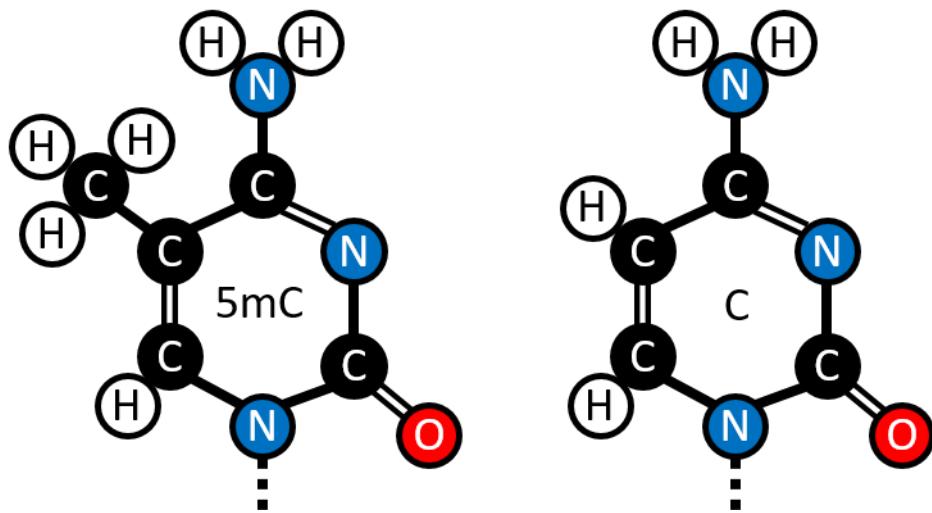


Figure 1.1: The structure of 5-methylcytosine contrasted with Cytosine. (Figure created by the Author.)

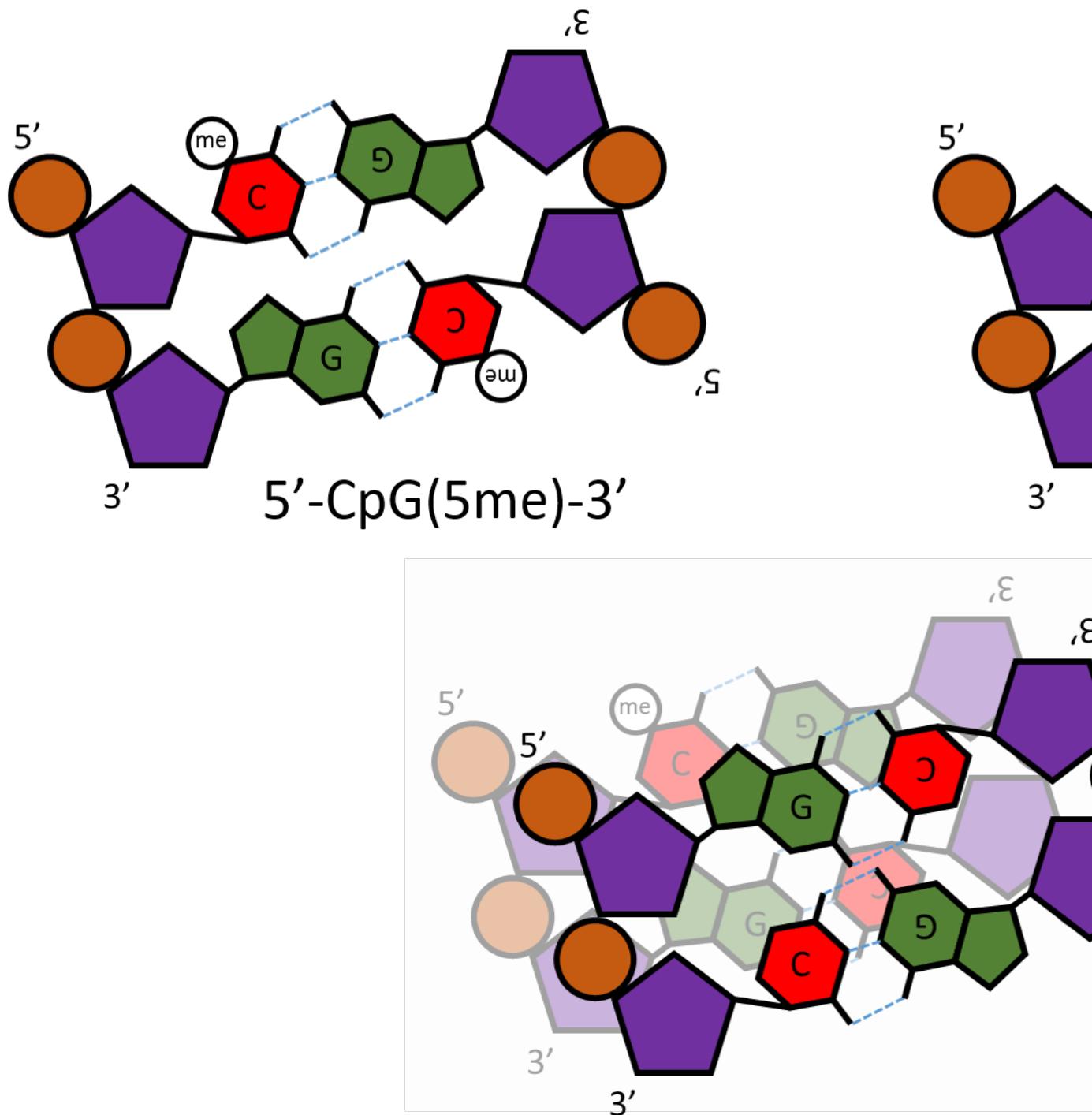


Figure 1.2: A schematic representation showing a Methylated **CpG** dinucleotide in its structural context. This is contrasted with a **GpC** dinucleotide, which cannot be superimposed on a **CpG**. upper left: methylated cytosine residues in a **CpG** dinucleotide, showing 5mC is 5' of G on both DNA strands. upper right: unmethylated cytosine residues in a **GpC** dinucleotide, showing G is 5' of C on both DNA strands. lower: superimposition of a **GpC** and **CpG** dinucleotide illustrating the mirrored nature of these structures and the impossibility of superimposing them through rotation. (Figure created by the Author.)

### 1.2.2 Distribution and Global Trends

There are 28,299,634 CpGs [Luo et al., 2014] in the hg19 assembly of the human genome [Lander et al., 2001]. Given that the GC content of the human genome is 42% the prior probability of getting a CpG dinucleotide is:  $0.21 \times 0.21 = 0.0441$ , 4.41%. CpGs represent  $\sim 1.8\%$  of the dinucleotides in the human genome ( $28,299,634 \div (3.23 \times 10^9 \div 2) \approx 0.0175$ ). Making them  $\sim 2.5$  fold less frequent than would be expected *a priori*. Methylated cytosines are prone to deamination to thymines, resulting in mismatch lesions [Duncan and Miller, 1980]. This increased mutagenic potential generally means they are selected against, accounting for at least some part of their under-representation in the genome. The three major classes of repeat elements SINEs, LINEs and LTRs contain some 46% of all CpG sites, with a further 5% in other repetitive elements [Luo et al., 2014].

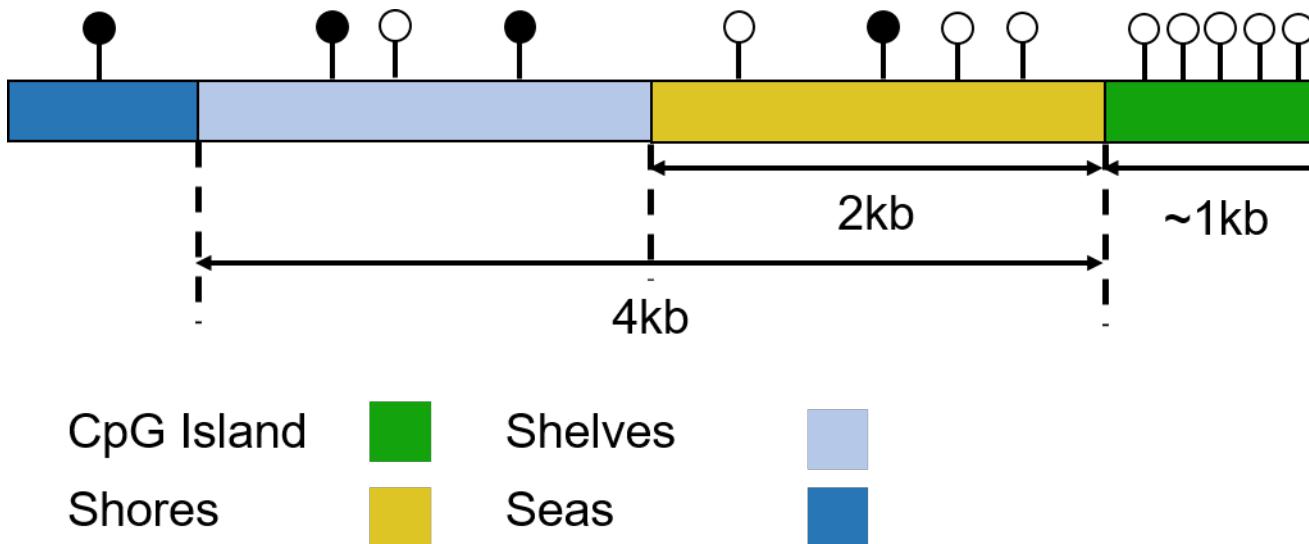


Figure 1.3: Schematic representation of CpG Islands, shores, shelves and seas. CpG density declines from shores to seas and methylation increases (CpG density and methylation proportion not to scale). (Figure created by the Author.)

CpGs are not uniformly distributed in the genome, they occur at higher frequency in some regions. “CpG islands” or CGIs are regions of high CpG density. The total number of CpGs in the UCSC repeat masked CGI annotation list of 28,691 CGIs is 1,990,729. Therefore, these CpGs comprise  $\sim 7.0\%$  of the total number of CpGs in the genome. The mean percentage of the sequence of these CGIs that is comprised of CpG dinucleotides is  $\sim 18.5\%$  and their mean length is 761bp. Irizarry et al. found that the 2kb regions flanking CpG islands which they termed “CpG island shores” exhibited greater tissue-specific differential methylation than the islands themselves [Irizarry et al., 2009]. This nomenclature has subsequently been expanded further with “CpG island shelves” which are 2kb - 4kb from the CGIs, and “seas” referring to the rest of the genome, see figure 1.3.

CGIs overlap the promoters of  $\sim 70\%$  of genes [Saxonov et al., 2006], this, however, leaves roughly half of all CGIs as “orphans” not associated with a known transcription site. These orphan CGIs frequently constitute cell-type specific enhancers [Bell and Vertino, 2017] and alternate promoters [Illingworth et al., 2010],[Maunakea et al., 2010]. They are found in both intergenic and intragenic regions. CpGs are distributed quite sparsely through the genome occurring at low density in “seas” and at increasing density in CGIs and their flanking regions, see figure 1.3. Early work showed 70-80% of CpGs are constitutively methylated [Ehrlich et al., 1982] (reviewed Bird [2002]) Stadler et al. produced a more detailed picture of the distribution of CpG methylation in mouse embryonic stem cells (ESCs) [Stadler et al., 2011] see figure 1.4. It is a

characteristic of CGIs that they are generally unmethylated.

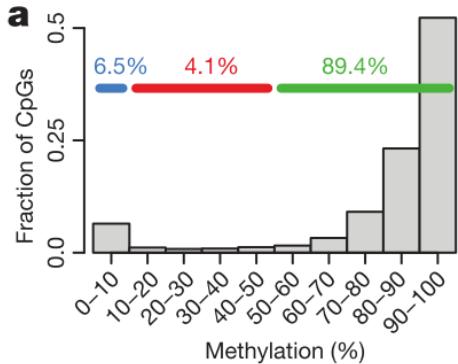


Figure 1.4: The fraction of CpGs with a given methylation level. Data from Whole-Genome Bisulfite Sequencing (WGBS) of Mouse embryonic stem cells. (Figure reproduced from Stadler et al. [2011] figure 1a)

The regulatory role played by DNA methylation varies with context and the scale at which it is examined. For example methylation at a single CpG in a transcription factor binding site can determine whether or not the factor can bind at that site. There are methylation-sensitive transcription factors which can have their affinity for DNA either increased or decreased by methylation [Yin et al., 2017, Zuo et al. [2017]]. In addition, larger scale changes in DNAm can through interactions with chromatin modifiers alter DNA compaction and more general accessibility of DNA regions for binding, and/or impacting on their topological organisation. DNAm also varies with nucleosome occupancy with lower methylation in linker sequences than on nucleosome-associated DNA [Kelly et al., 2012, Huff and Zilberman [2014]]. The oxidation products of 5mC can also affect chromatin organisation, 5-Formylcytosine can impact nucleosome positioning through covalently binding to histones [Raiber et al., 2018]. This raises the question if some DNA methylation occurs specifically to act as an intermediate step for regulatory mechanisms which utilise its oxidation products.

### 1.2.3 Pathways of DNA methylation and demethylation

CpG methylation is produced and maintained by DNA methyltransferase (DNMT) enzymes. All the DNMT enzymes use S-adenosylmethionine as the source of the methyl donor group. DNMTs form a covalent intermediate between a conserved cysteine residue and the target base, through a nucleophilic attack on the C6 position in the cytosine ring. This is followed by the transfer of the S-adenosylmethionine methyl group to C5, and deprotonation of the C5 to reform the double bond between C5 and C6; which is mediated by a base provided by the enzyme (Figure 1.5).

DNMT1 is associated with the replication machinery [Vertino et al., 2002] and reproduces the methylation state of the parent strand on the daughter strand during replication. The largest of the DNMT family DNMT1 contains a ‘replication foci targeting sequence’ (RFTS) domain required for its targeting to replication forks. DNMT1 specialises in recognising hemimethylated DNA and methylating the unmethylated C in a palindromic CpG dinucleotide site, the UHRF1 protein assists in the recognition of these sequences [Bostick et al., 2007]. DNMT3a and DNMT3b are responsible for *de novo* DNA methylation along with DNMT3L a catalytically inactive, but DNA-binding subunit [Jia et al., 2007]. The location of *de novo* methylation by the DNMT3s is influenced by a number of factors including the Chromatin state and other DNA binding factors [Lyko, 2017].

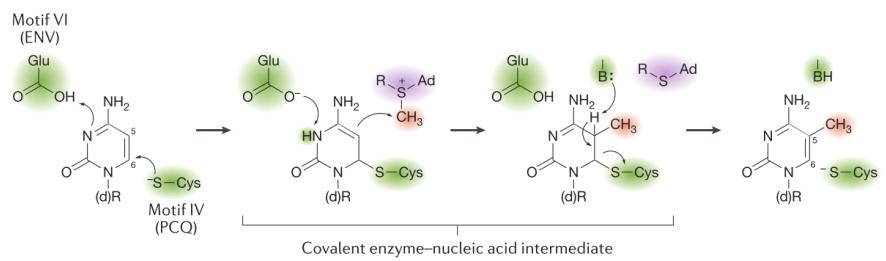


Figure 1.5: Motif VI (ENV/PCQ) refers to the conserved catalytic motif in DNMTs. ‘B:’ Represents the enzyme supplied base (Figure reproduced from the review of DNMT biology by Lyko [2017] (fig. 1b))

DNMT2 is unlike the other member of the DNMT family in that it targets an RNA substrate. DNMT2 is a tRNA methyltransferase which methylates a specific subset of tRNA genes, mostly Asp isoacceptors, at a site adjacent to the anticodon which protects them from endonucleolytic cleavage under stress conditions. Unprotected tRNAs produce fragments which compete with small interfering RNAs interfering with their signalling pathways.

The mechanisms of the demethylation of DNA were more recently characterised than those which govern its methylation. They are more complex and less well understood. A distinction is drawn between active and passive demethylation, in passive demethylation, 5mC bases are diluted out in the process of DNA replication. In the leading model of active demethylation, they are oxidised one or more times by an enzyme from the TET (Ten-Eleven-translocase) family. They are then either passively removed by DNA replication or actively removed by a DNA glycosylase (thymine DNA glycosylase TGD) to create an apurinic site which is restored to a C by the Base Excision Repair (BER) pathway. This cycle of cytosine methylation and demethylation is illustrated in Figure 1.6 [Wu and Zhang, 2017].

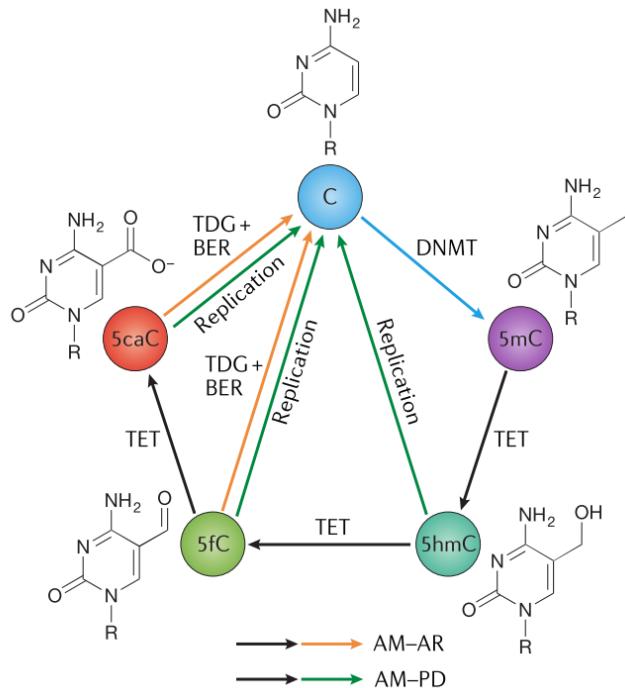


Figure 1.6: BER = Bases Excision Repair; AM = Active Modification; AR = Active Removal, PD = Passive Dilution (Figure reproduced from the review of TET mediated active demethylation by Wu and Zhang [2017] (fig. 1a))

### 1.2.4 Maintenance and Fidelity

DNA methylation is highly dependent on the underlying DNA sequence. Sequence features of particular importance to determining DNA methylation status are; transcription factor and other DNA binding protein recognition motifs, and CpG density [Lienert et al., 2011],[Ziller et al., 2013]. Alterations in underlying DNA sequence such as SNPs and copy number variants (CNVs) can have a significant impact on methylation level and the susceptibility of the methylation level to change. In addition CNVs can result in dosage effects on measures of DNAm, causing regions to appear, respectively, substantially more or less methylated when fewer or greater copies than expected are present [Boks et al., 2009], [Kerkel et al., 2008],[Schalkwyk et al., 2010],[Shoemaker et al., 2010], and [Bell et al., 2018]. Despite the strong influence of sequence on methylation, global CpG methylation exhibits change over developmental time as illustrated in figure 1.7 and exhibits tissue-specific changes in distribution and amount [Ziller et al., 2013].

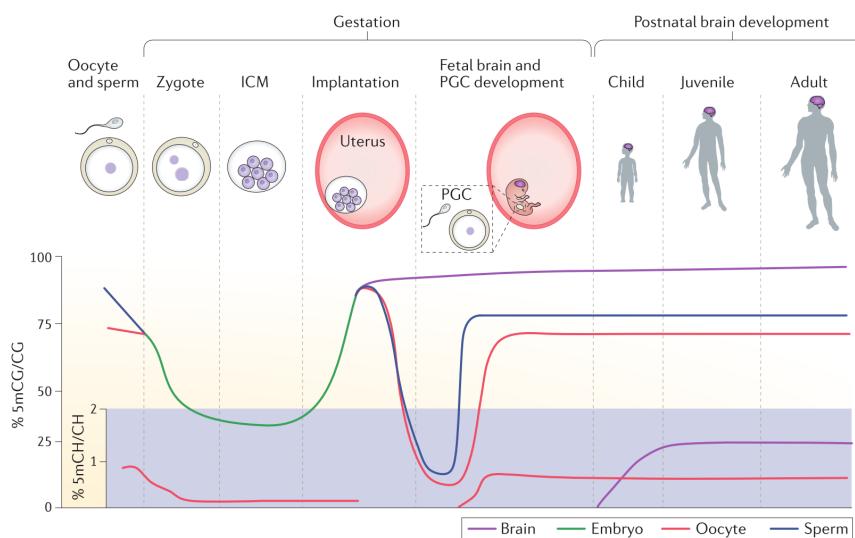


Figure 1.7: 5mC levels over developmental time. 5mCH or 5-methylcytosine-(A, T or C) levels are represented on a separate (purple) Axis from the main 5mCG axis. PGC = Primordial Germ Cell. (Figure reproduced from Ciernia and LaSalle [2016] figure 3)

In mitosis, DNA methylation is inherited by daughter cells with an error rate on the order of  $1 \times 10^{-3}$  per site per generation [Ushijima et al., 2005], several orders of magnitude less than that of DNA replication (error rate of  $1 \times 10^{-7}$  -  $1 \times 10^{-8}$  [Kunkel, 2004]). The fidelity of DNA methylation copying can be assayed by Hairpin-Bisulfite PCR (Polymerase Chain Reaction) [Laird et al., 2004]. Laird et al. looked at two alleles of a portion of the CpG island from the human *FMR1* gene in uncultured lymphocytes, one hypermethylated and one hypomethylated. In the hypermethylated allele, they found that 96% of sites methylated in the parent strand remained methylated in the daughter strand and 86% of unmethylated sites remained unmethylated. By contrast in the hypomethylated allele, there were no methylated sites to be retained and >99% of unmethylated sites remained so following replication. Laird et al.'s work suggests a higher degree of overall methylation fidelity for hypomethylated DNA and a propensity for unmethylated sites in hypermethylated DNA to become methylated.

Change in DNAm levels over time (divisions) can be modelled using the differential equations [Pfeifer et al., 1990], which predict that a fully methylated site and a fully unmethylated site will converge on an equilibrium level. This level is determined by the probability of maintenance of the methylation state and of *de novo* methylation for a given locus. This stochastic model of DNAm is in agreement with experimental findings [Laird et al., 2004, Riggs and Xiong [2004]], Figure 1.8.

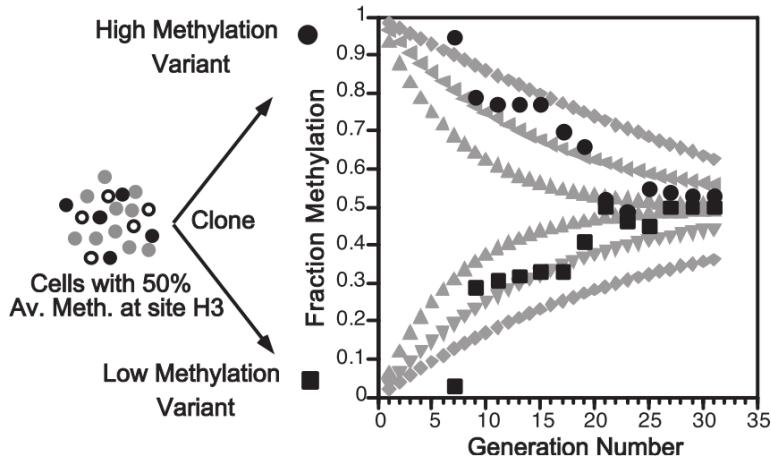


Figure 1.8:  $E_m$  = probability of methylation maintenance,  $E_d$  = probability of *de novo* methylation.  $M$  &  $U$  = the number of methylated and unmethylated molecules at specific CpG sites respectively. Modelling (Grey points)  $E_m, E_d$  values of 0.90, 0.10; 0.95, 0.05; 0.97, 0.03 (outermost to innermost, two curves for each  $E_m, E_d$  set, one starting at  $M = 1, U = 0$  and one a  $M = 0, U = 1$ ). Experimental data (Black points) from the *HpaII* locus (site H3) in 17 clones of mouse cell line BML-2 which has a known methylation level of 50%. (Reproduced from [Riggs and Xiong, 2004])

Jenkinson et al. [2017] used an information-theoretic model, modelling DNAm as a binary communications channel using a 1-dimensional Ising model from the field of statistical physics. This permitted them to examine properties of DNAm not accessible to conventional means of analysis which typically capture the mean methylation level and perhaps the variability for a given locus. When considering methylation fidelity using this lens the maintenance of a given methylation state can be seen as an information processing task which requires the consumption of free energy in order to reduce the probability of error in transmission of that methylation state. Thus Relative Dissipated Energy (RDE) can serve as a measure of the work expended by a cell at a given locus in order to preserve the current methylation state of that locus.

Zhao et al [Zhao et al., 2014] applied Hairpin-Bisulfite PCR genome-wide in mouse embryonic stem cells (ESCs). They also found high degrees of methylation fidelity in hypomethylated regions such as CGIs and Promoters, as well as a high degree of fidelity in sites bound by transcription factors. This is in agreement with the findings of Jenkinson et al. [2017] who noted that entropy (methylation stochasticity) was lower and more variable in CGIs and TSS (transcription start sites).

Methylation inheritance fidelity is reduced in cancer [Ushijima et al., 2005] and increases with differentiation [Zhao et al., 2014]. The RDE (relative dissipated energy) at CGIs and TSSs is higher in differentiated tissues such as the brain, implying low entropy, and lower in embryonic stem cells, implying greater entropy [Jenkinson et al., 2017] (see Figure 1.9). In addition, the correlation between CpG sites increases in cancer [Jenkinson et al., 2017], suggesting reduced higher level regulatory control and tendency to fall back on lower level feedback mechanisms.

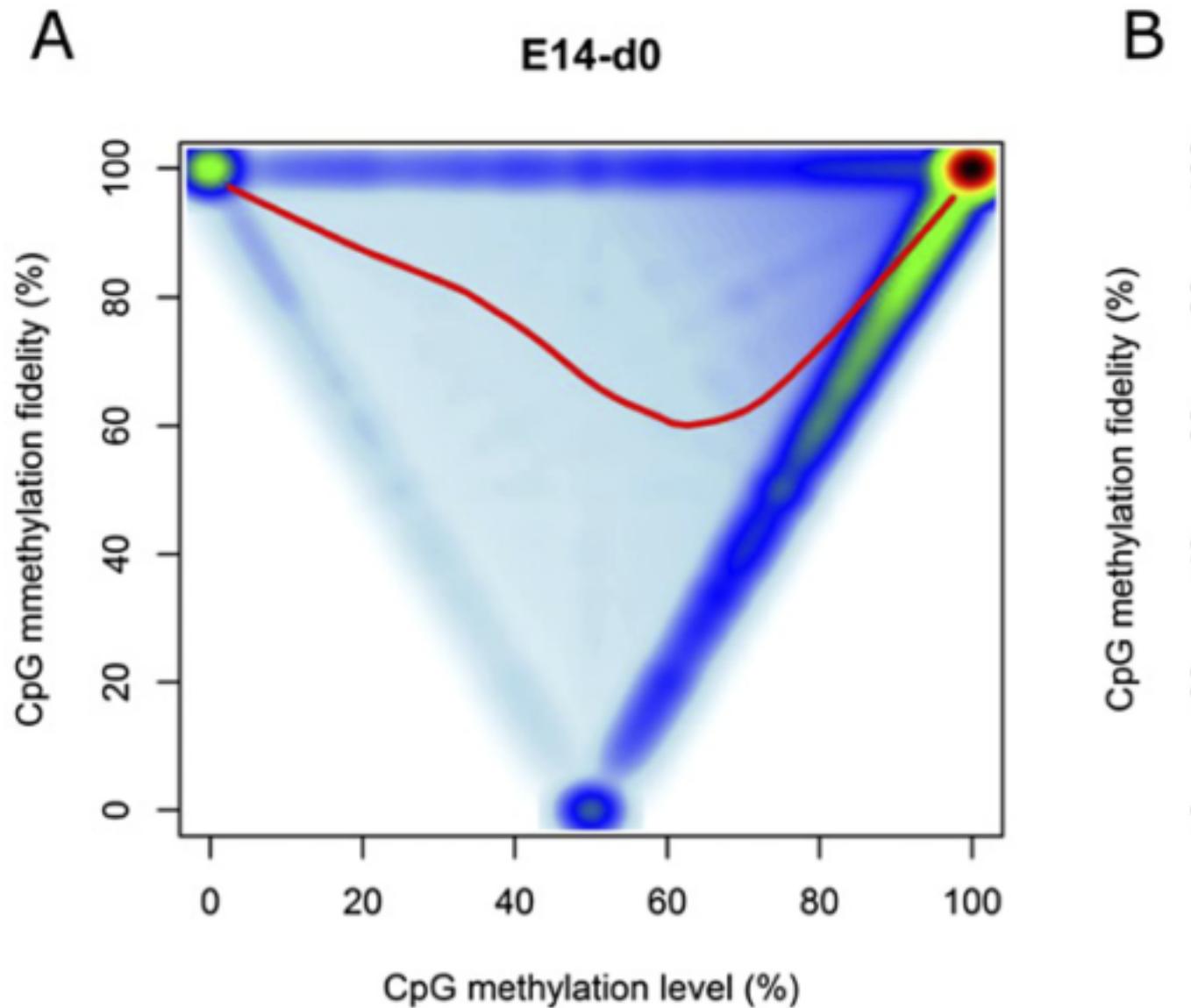


Figure 1.9: Bimodal Distribution of DNA methylation level and fidelity. Methylation fidelity exhibits a bimodal distribution with most and least methylated regions exhibiting the highest fidelity. Additionally, fidelity for methylation levels of 10-50% is considerably better than fidelity for levels of 50-90%. These data are from the mouse cell line ES-E14TG2a which is capable of self-renewal when cultured with Leukaemia Inhibitory Factor (LIF) and spontaneous differentiation upon removal of LIF (denoted as E14-d0 at day 0 and E14-d6 at day 6 after the withdrawal of LIF, respectively). Fidelity represents the percentage of symmetrically methylated or unmethylated CpG dyads for a given position as determined by Hairpin-Bisulfite PCR. (Reproduced from Zhao et al. [2014] (figure 3 a and b).)

## 1.3 Ageing

### 1.3.1 Hallmarks

Ageing is characterised by a progressive deterioration of physiological integrity with time, resulting in increased risk of mortality and morbidity. These changes at the organismal level have

their origins in changes at the molecular and cellular scales. López-Otín et al. [2013] identified nine hallmarks of ageing in their 2013 review. These hallmarks are:

1. Genomic instability
2. Telomere attrition
3. **Epigenetic alterations**
4. Loss of proteostasis
5. Deregulated nutrient sensing
6. Mitochondrial dysfunction
7. Cellular senescence
8. Stem cell exhaustion
9. Altered intercellular communication

This review focuses on the 3rd hallmark, epigenetic alterations, specifically the role of DNA methylation in ageing. Booth and Brunet [2016] argue that Epigenetic dysregulation is a hub through which many of the hallmarks of ageing are mediated. Gene regulation as governed by epigenetic factors is a key module in the complex network of systems maintaining tissue homeostasis and is highly connected to other modules such that the deterioration of the epigenetic control of gene expression module can snowball into the dysregulation of other modules (Figure 1.10).

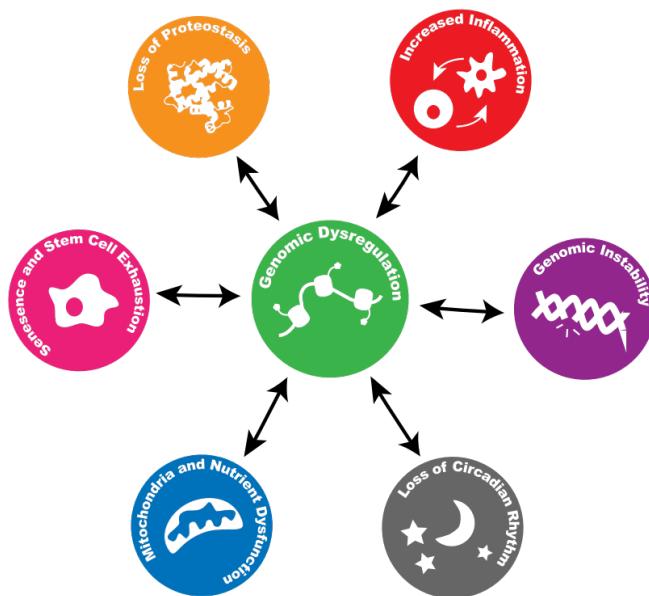


Figure 1.10: Epigenetic Changes as a hub for the hallmarks of ageing. (reproduced from [Booth and Brunet, 2016], figure 3)

### 1.3.2 DNA Methylation and Ageing

Early work on DNA methylation and ageing used biochemical, chromatographic and radiolabelling techniques to assay global changes in the amount of 5mC. These studies found differences in 5mC composition with cell type [Ehrlich et al., 1982] but did not see changes with age. Other work around that time however found differences in 5mC levels with the age of cells in culture [Wilson and Jones, 1983]. Wilson et al. noted that DNA methylation decreased across several tens of generations of cell lines in culture, but not in immortal cell lines. Immortal cell lines had lower absolute levels of 5mC to start with but remained constant over time. Wilson et

al. also cite earlier work by Romanov and Vanyushin [1981] and Berdyshev et al. [1967] which documented decreases in 5mC with age in cows and salmon, respectively.

Later work by Wilson et al. [1987] in mice *P. leucopus* and *M. musculus* found decreases in 5mC with age and that the rate of decline was less in the longer-lived *P. leucopus* than in *M. musculus*. Interestingly, a recent study using different methods by Cole et al. found no global differences between young and old mice in short and long-lived strains, however, their other observations would seem to corroborate these initial trends. For the assayed sites long-lived mice exhibited 10x more hypermethylation than wild-type (WT), and WT mice had 3x more Differentially Methylated Regions significantly associated with age (DMRs/aDMRs) than did long-lived mice. Notably, the WT and long-lived mice shared many of the same aDMRs which differed in their degree of methylation as opposed to affecting different sites in the genome [Cole et al., 2017]. The lack of apparent global changes may be due to the biases of the reduced representation bisulfite sequencing (RRBS) method used. RRBS uses a restriction enzyme based approach to enrich for regions with high GC content such as CGIs which tend to have low levels of methylation, and consequently may not be sensitive to loss of methylation in generally hypermethylated regions which could contribute to a global trend [Meissner et al., 2005].

Wilson et al. also noted that the mitotic index of tissues did not relate to the loss of DNA methylation with age in tension with their earlier observations *in vitro*. The persistent loss of 5mC over time and the dramatic changes in methylation seen in cancer cells lead Wilson et al. to suggest that dysregulation of DNA methylation may have a substantial role to play in the age dependency of cancer risk and ageing more generally, a possibility which has subsequently been further explored by others, reviewed by Feinberg and Tycko [2004]. It has been suggested that “Epimutations” may be able to substitute for mutations in the multi-hit model of carcinogenesis [Knudson, 1971]. For example; hypermethylation of the *BRCA1* promoter [Esteller et al., 2000], Or, the development of Wilms’ tumour due to Beckwith–Wiedemann syndrome, a disorder arising from loss of imprinting of the gene encoding insulin-like growth factor 2 (IGF-2) leading to a double dose of IGF-2 protein [Feinberg, 2018].

Methods which allowed the examination of changes in DNA methylation at known loci in the genome permitted a more nuanced picture of changes in DNA methylation over time to develop. Fraga et al. [2005] introduced the concept of “epigenetic drift” being the divergence of DNA methylation and other epigenetic modification patterns with time. Fraga et al. looked at the divergence in DNA methylation along with global Histone H3 and H4 acetylation patterns between monozygotic (MZ) twins over a wide range of ages. They observed that older twins had greater epigenetic differences with time, in DNA methylation as well as H3, H4 Acetylation. Fraga et al. also noted that the divergence in epigenetic state was greater in twins who had lived longer apart and had different medical histories. Figure 1.11 is a useful visual encapsulation of Fraga et al.’s DNA methylation results. Slieker et al. [2016] identified 6366 CpGs whose methylation variability increased with age using the Illumina 450k array platform on whole-blood from 3295 individuals, both the increase in variability with age and the increasing divergence of twins support a narrative of epigenetic dysregulation with age.

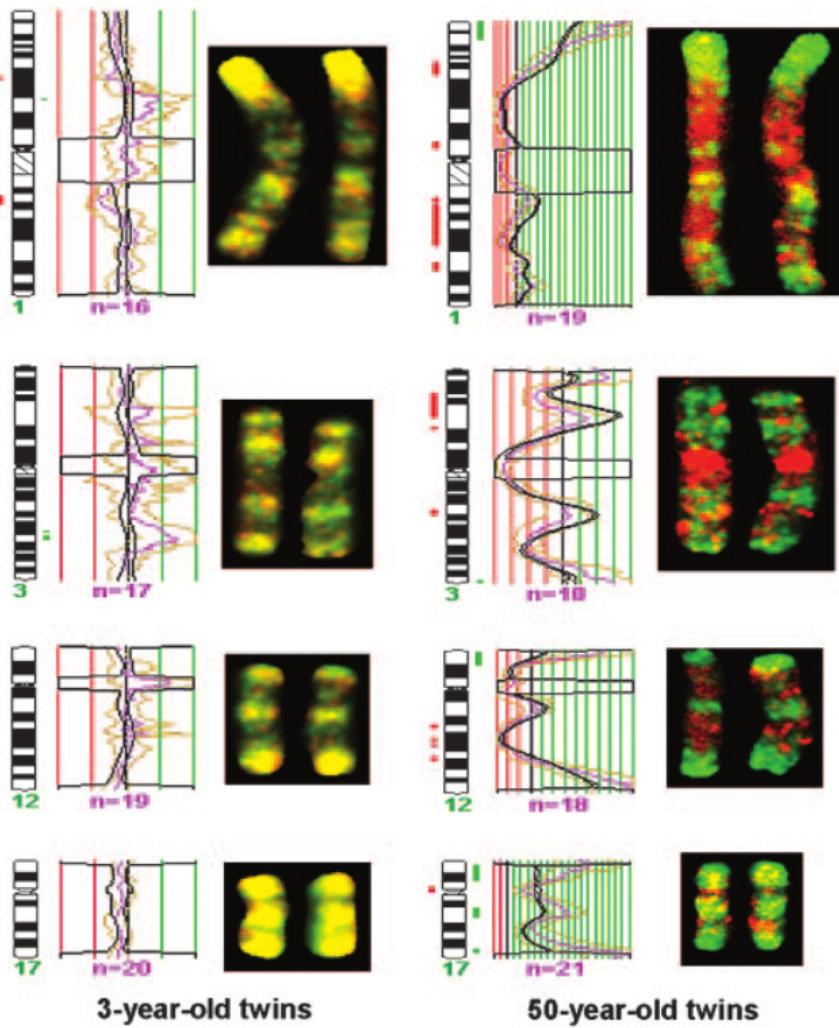


Figure 1.11: Mapping Chromosomal Regions of Differential DNA Methylation. Examples from Chromosomes 1, 3, 12 & 17 are shown for a 3 and 50-year-old twin pair. Methylation status is shown by competitive hybridization of AIMS (amplification of inter-methylated sites) products to metaphase chromosomes. Green and Red signals indicate hypermethylation and hypomethylation events between twins, Yellow indicates little difference. Red and Green blocks adjacent to ideograms indicate areas of significant DNA methylation change. (Reproduced from [Fraga et al., 2005] (figure 3).)

### 1.3.3 Age-Related Differential Methylation and Epigenetic Clocks

With the advent of array-based techniques which permitted the measurement of the methylation status of specific CpGs placed strategically throughout the genome, researchers were able to undertake a more fine-grained analysis of the changes in DNA methylation with age. This has allowed the prediction of chronological age from a metric of “DNA methylation age” (DNAm age), and interpretation of the differences in predicted and chronological age in terms of the pathophysiology of ageing.

The Illumina Golden gate array [Bibikova, 2006] with 1505 probes targeted to cancer-related genes was the first of these. Bjornsson et al. observed DNAm changes with age as well as noting substantial intra-individual differences. In contrast earlier work by Eckhardt et al. [2006] and Ehrlich et al. [1982] found no age-related changes using averages across individuals. Bjornsson [2008] also found that the intra-individual differences in differential methylation with ageing were

highly heritable by use of familial clustering. Work by Boks et al. [2009] also using the golden gate array corroborated Bjornsson et al.'s findings in monozygotic twins.

The Bjornsson and Boks studies used the golden gate array on peripheral blood samples, whereas Christensen et al. [2009] used the golden gate array to look for age-related changes in DNAm in several tissue types. CpG sites that are differentially methylated with age when searching across tissues were frequently specific to a small number of tissue types. In addition, CpGs in CGIs tended to be hypermethylated with age and *vice versa*. Christensen et al. also found that the methylation profiles for different tissues were highly predictive of tissue type.

Rakyan et al. [2010] looked at age-related changes in DNAm using the Illumina 27k array [Bibikova et al., 2009] with whole blood samples in a discovery set. They replicated their findings in sorted cell fractions to see if they could capture changes in DNAm that could be attributed to changes in blood cell-type composition over time, which they did not find. Teschendorff et al. [2010] found that promoters of targets of the polycomb group proteins (PCGTs) were more likely than non-PCGTs to become methylated with age. Repression of PCGTs is required for Stem cells to differentiate. PCGTs whose methylation status was associated with age were associated with pre-neoplastic conditions in a large cohort using the 27k array with blood and epithelial cell samples.

Bocklandt et al. [2011] also using the 27k array but with saliva samples created the first “epigenetic clock” used to predict the chronological age of donors based on the DNAm landscape of their cells, the mean error was 5.2 years. Bocklandt et al. were able to achieve similar predictive accuracy with as few as 3 CpG sites in their model. Koch and Wagner [2011] used publicly available 27k array datasets from a variety of different tissues to train a model using a different statistical method but only achieved an error of 11 years. They did, however, identify a number of CpGs also identified by Bocklandt et al. Bell et al. [2012b] used 27k array data to look for correlations between differential methylation and age-related phenotypes as well as chronological age. They found that few age related phenotypes were correlated with differential methylation, only 5 CpG sites were identified for the 16 age-related phenotypes examined. Whereas 490 significant CpGs were associated with chronological age. Many of the CpGs they identified persisted across tissue types and replicated in a second cohort, several had been identified previously by Rakyan et al. and Bocklandt et al.

Heyn et al. [2012] examined the DNA methylation status of a newborn and a centenarian in CD4+ T cells using whole genome bisulfite sequencing (WGBS), along with a group of newborns and nonagenarians using the Illumina 450k array [Bibikova et al., 2011]. Heyn et al. observed a global decrease in methylation from newborn to centenarian, as well as an intermediate level of methylation on a sample of intermediate age (see figure 1.12). This observation was replicated in 450k data. In addition, Heyn et al. noted that adjacent CpGs normally exhibit a substantial degree of correlation in methylation status and that they were less well correlated with their neighbours with increasing age. Interestingly Jenkinson et al. [2017] found that correlation among nearby CpGs increased in cancer tissues. Modelling work done by Affinito et al. [2016] agrees with this correlation between neighbouring CpGs under physiological conditions, which is particularly pronounced in CpG dense regions like CGIs. The greater physical proximity of CpGs in CpG dense regions means that greater correlation is expected. Due to the fact that these CpGs are more likely to be affected by the same proteins and regulatory features than CpGs with greater distance separating them [Haerter et al., 2014]. Garagnani et al. used 450k array data from a cohort of 64 subjects aged 9-83 to identify those CpGs most well correlated with age as had been done with previous array technologies [Garagnani et al., 2012].

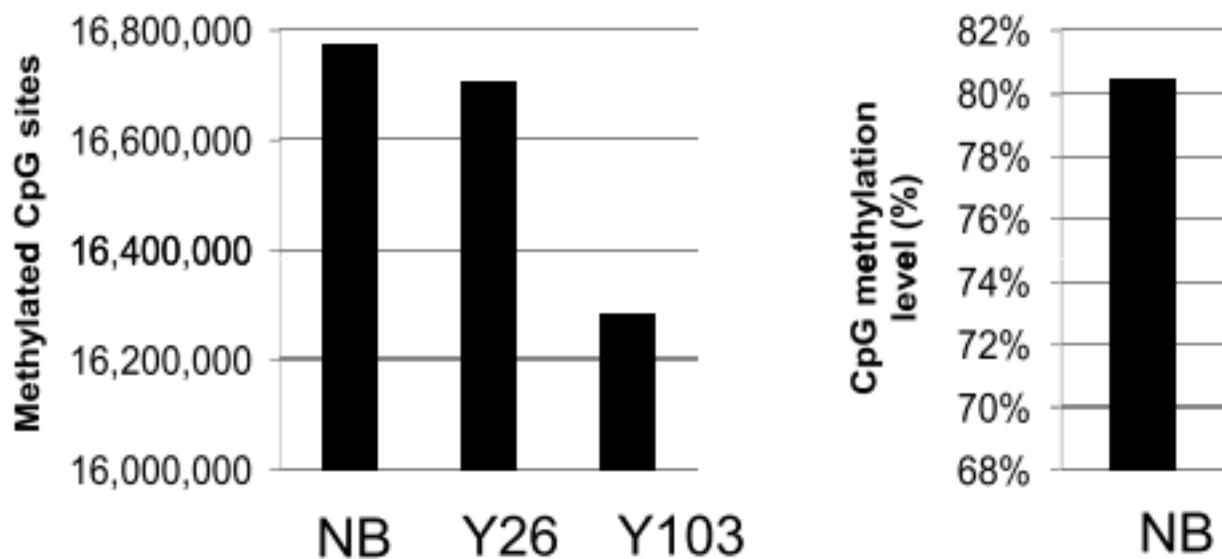


Figure 1.12: Global Hypomethylation with Age. NB = Newborn, Y26 = 26 year old, Y103 = 103 year old. (Reproduced from Heyn et al. [2012] figure 1 b.)

In January of 2013 Hannum et al. [Hannum et al., 2013] published a study using the Illumina 450k array and whole blood samples for 656 participants aged 19-101. Their optimised model was able to predict age with an error of 3.9 years. In December of the same year, Horvath [Horvath, 2013] published a study using multiple datasets, including Hannum's data, some produced on the 27k and some the 450k array. Both used 'elastic net' penalised multivariate regression models to identify CpGs which collectively provide the best predictor of DNAm age rather than the CpGs which are individually most highly correlated with age [Friedman et al., 2010].

Florath et al. [2014] identified some additional age-associated CpGs from two large cohorts totalling N=898 and an N=67 8yr follow-up longitudinal cohort. More than 3/4 of the CpG sites they identified began as hypomethylated and increased in methylation with age, a proportion likely skewed by the large number of 450k probes in CGIs and other typically low methylation regions. Bacalini et al. [2015] performed a meta-analysis of existing DNAm age datasets and employed a "region-centric" approach to try to identify loci larger than single CpGs which they anticipate will be more biologically meaningful than lone CpGs. They found that their approach increased the number of common features identified using the Hannum et al. and Heyn et al. datasets. Zaghloul et al. [2015] performed 450k array DNAm age study in Qatari population they identified 12/88, 23/490 and 102/162 of the CpGs found by Bocklandt et al, Bell et al. and Florath et al. respectively. Zaghloul et al. also found that Horvath's age predictor had an error of 3.7 years in their dataset. This is in agreement with existing findings and indicates that ethnicity has no major effects on the DNAm age signature. Benton et al. [2017] looked at changes in DNA methylation with age in a genetically isolated population on Norfolk Island with similar results to previous studies but identifying some novel age-associated CpGs.

Most DNAm relationships with age noted in previous studies have been linear, Johnson et al. [2017] used the 450k array in peripheral blood samples. They identified 21 CpGs whose DNAm changes at a rate that changes with age from an initial pool of 27,723 CpGs which were differentially methylated with age. Two sites exhibited an increasing rate of increase in DNAm with age, and 18 sites a decreasing rate of increase.

### 1.3.4 Genetic Influences on DNA Methylation

Epigenetic variation falls on a continuum of genetic influence that can be summarised by three categories:

1. **pure** - DNA sequence has no predictive value for epigenetic state.
2. **facilitated** - DNA sequence biases epigenetic state.
3. **obligatory** - DNA sequence permits exact prediction of epigenetic state.

Regional methylation state is strongly influenced by genotype, by single nucleotide polymorphisms (SNPs) [Smith et al., 2014] and by structural variants [Bell et al., 2018]. An example of an obligatory effect on methylation is a point mutation at a CpG site, a C to T transition precludes methylation at that site in future. The effect of structural variants on methylation can be hard to determine as changes in sequence dosage often lead to measurement artefacts. Efforts have been made to correct for the influence of genetic factors in EWAS, when searching for purely epigenetic effects but the potential interaction of the somatic mutations known to accumulate with age (Figure 1.13) and the changes in DNAm with age remain largely unexplored. This is of particular relevance to DNAm as the profile of the types of mutation which accumulate with age (Figure 1.14) distinctly favours C to T transition mutations which can disrupt CpG dinucleotides [Jaiswal et al., 2014].

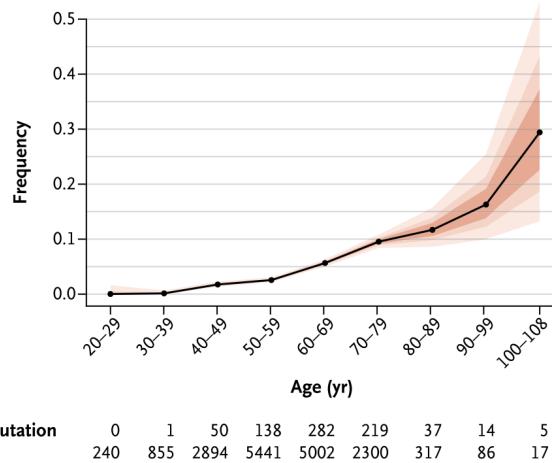


Figure 1.13: Somatic mutations increase with Age. (Reproduced from [Jaiswal et al., 2014], figure 1)

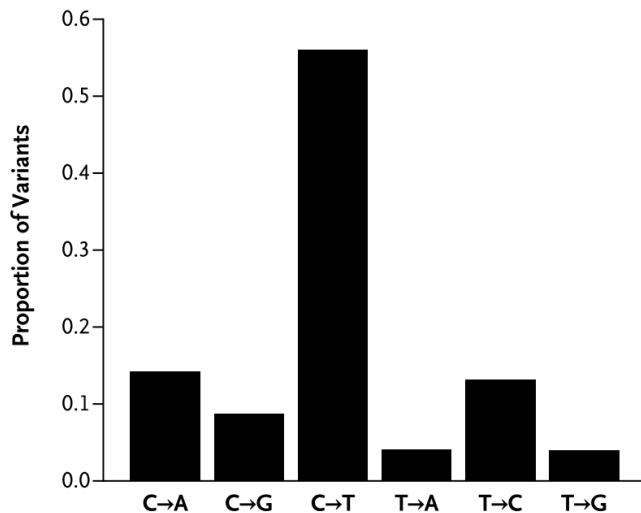


Figure 1.14: C to T transitions are the most common substitutions that occur with age. (Reproduced from [Jaiswal et al., 2014], figure2 c)

### 1.3.5 DNA Age as a Biomarker of Ageing

We have seen that DNAm can be a fairly accurate predictor of age, however, there is more to being a good biomarker of ageing than predictive accuracy. Weidner et al. [2014] conducted a study with the specific intent of identifying CpG sites that would serve as the best biomarkers of ageing. Below is a list of specific criteria for a high-quality biomarker of ageing laid out by Johnson [2006].

1. It must predict the rate of ageing. In other words, it would tell exactly where a person is in their total lifespan. It must be a better predictor of lifespan than chronological age.
2. It must monitor a basic process that underlies the ageing process, not the effects of disease.
3. It must be able to be tested repeatedly without harming the person. For example, a blood test or an imaging technique.
4. It must be something that works in humans and in laboratory animals, such as mice. This is so that it can be tested in lab animals before being validated in humans.

We have seen that DNA methylation clocks can be good predictors of chronological age, we will revisit the question of whether or not they are superior to chronological age in their ability to predict lifespan below. Bell et al. [2012b] found DNAm did not generally correlate well with ageing phenotypes but did correlate well with chronological age and Teschendorff et al. [2010] found that their age-associated DNAm signature remained fairly constant across several disease states including ovarian cancer and type 1 diabetes. Suggesting that DNAm age prediction does indeed: “monitor a basic process that underlies the ageing process, not the effects of disease”, meeting criterion number 2. Several of the DNAm clocks discussed so far have been based on whole peripheral blood samples meeting criterion number 3. DNAm age clocks have been shown to work in Chimpanzees [Horvath, 2013], Dogs/Wolves [Thompson et al., 2017], Mice [Stubbs et al., 2017], naked mole rats, rhesus macaques, humpback whales [Lowe et al., 2018] and are likely to work in other mammalian model organisms. It should be noted that Horvath’s Human clock uses array data and that the Dog/wolf and mouse clocks use RRBS. Model organisms such as *C. elegans* which essentially lack DNA methylation would not, however, be likely to be tractable for DNAm based age prediction. Consequently DNAm age prediction at least partially meets criterion number 4.

Is DNA methylation a better predictor of lifespan than chronological age? Both Hannum and Horvath identified the difference between DNAm age and chronological age or  $\Delta_{Age}$  as a potential

indicator of disproportionate biological ageing. Marioni et al. [2015] explicitly set forth to test the value of DNAm age as a predictor of mortality. Adjusting for Age, Sex, childhood IQ, education, social class, hypertension, diabetes, cardiovascular disease, and *APOE* e4 status a  $\Delta_{Age}$  of +5 is associated with a 16% increased mortality risk (See figure 1.15). A longitudinal study of twins found 3.2 fold increase in the risk of dying first per 5yr  $\Delta_{Age}$  within twin pairs [Christiansen et al., 2016]. A meta-analysis of DNAm age predictors by Chen et al. [2016] also found that measures of age acceleration based on Hannum and Horvath clocks were superior predictors of mortality than chronological age before and after correction for various potentially confounding factors. Chen et al. also noted that correction for blood cell composition improved predictive power. Indicating that DNAm age is indeed superior to chronological age as a predictor of lifespan meeting criterion number 1 for an ageing biomarker.

A new DNAm age clock explicitly designed to capture phenotypic age and outperform age acceleration as a biomarker of ageing has recently been produced by Levine et al. [2018]. The ‘PhenoAge’ metric was created by selecting nine biomarkers of ageing from 42 possible metrics using a proportional hazards penalised regression model and combining these with chronological age. (The nine biomarkers are: Albumin, Creatinine, Serum glucose, C-reactive protein, Lymphocyte percent, Mean red cell volume, Red cell distribution width, Alkaline phosphatase, White blood cell count.) DNAm data was then regressed against PhenoAge using elastic-net regression to produce a DNAm based PhenoAge predictor which made use of 513 CpGs. The PhenoAge predictor outperformed the Horvath and Hannum clocks at predicting all-cause mortality, comorbidities, coronary heart disease risk, and measures of physical functioning. 41 of the 513 CpGs in the PhenoAge clock were present in the original 353 CpG sites used in the Horvath clock.

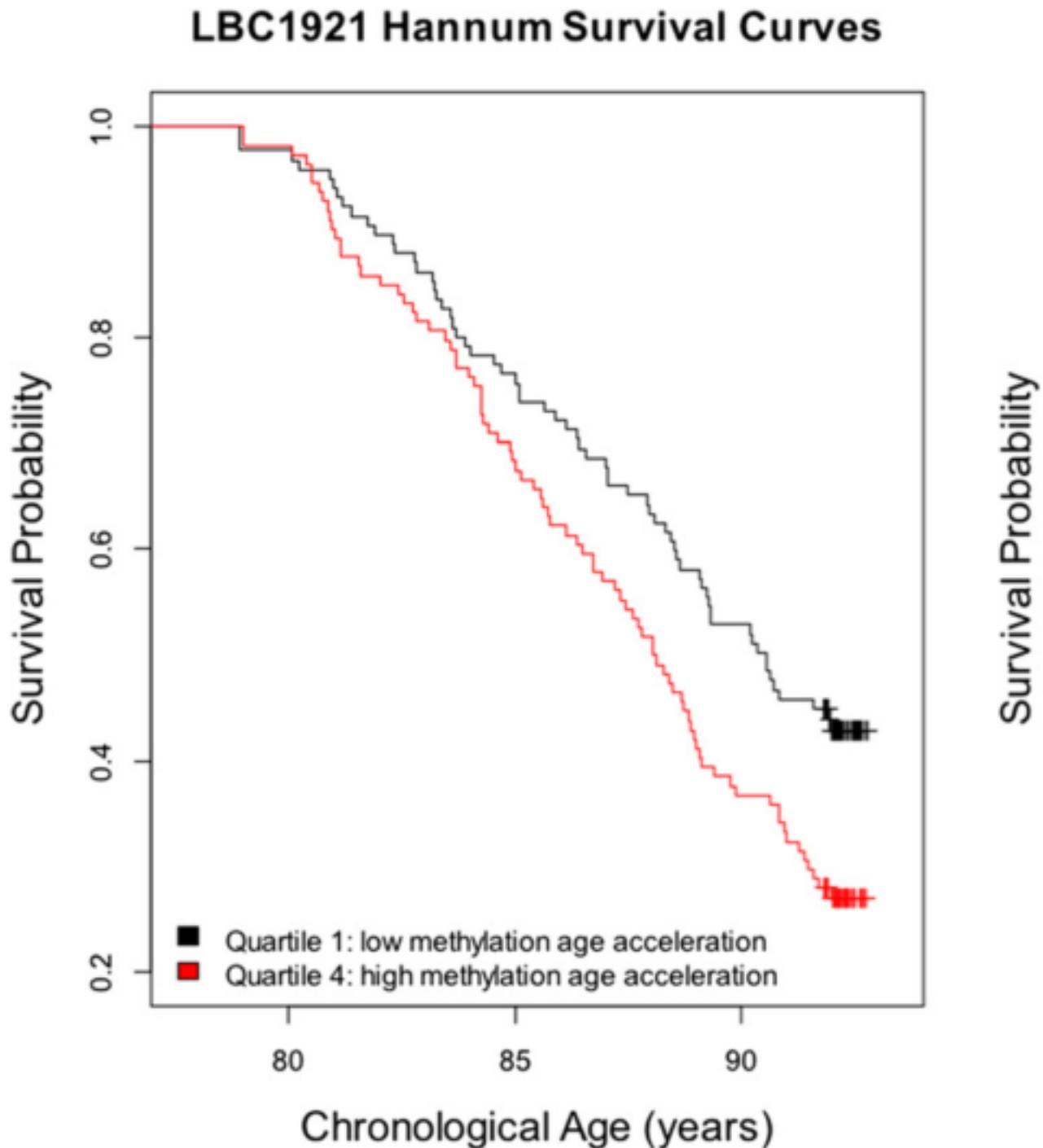


Figure 1.15: Survival probability by quartiles of  $\Delta_{Age}$  in LBC 1921 adjusted for sex, and chronological age. LBC = Lothian Birth Cohort. Using the Hannum and Horvath predictors' values for  $\Delta_{Age}$ . (Reproduced from Marioni et al. [2015] figure 3.)

#### 1.3.6 Discussion and Analysis of the Findings of DNAm Age Studies

Interestingly as illustrated in figure 1.16 CpGs whose methylation status has been identified as correlating with age in various studies show relatively little overlap, this overlap would be

reduced a little further if I had also required a consistent direction of change with age in this search. Despite this, changes in DNAm with age are probably the most robust and reproducible large-scale epigenetic change yet captured. This lack of agreement in absolute terms, however, presents a challenge about how to interpret DNAm changes with age.

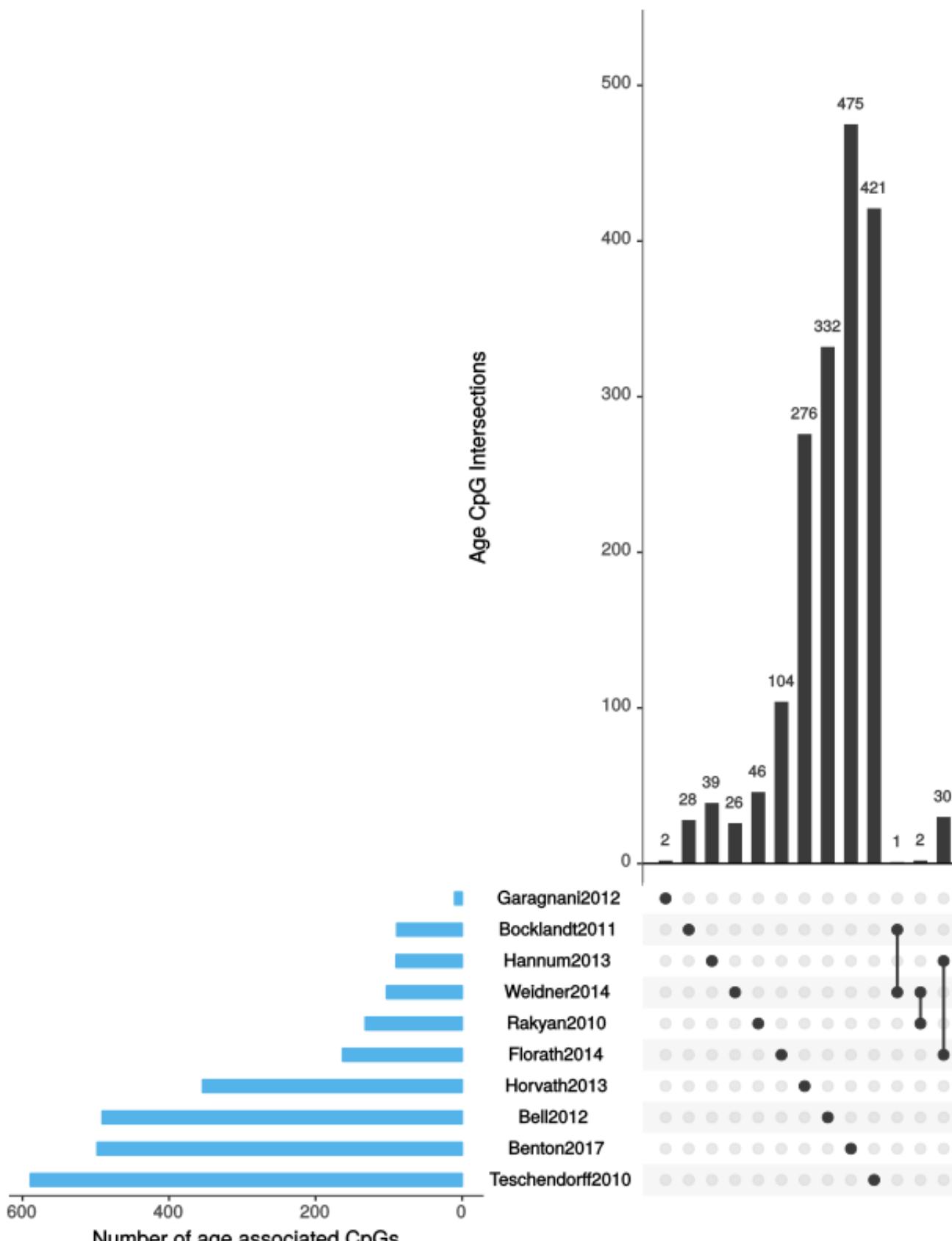


Figure 1.16: Overlap of CpG sites identified in various DNA methylation ageing models, set intersections with no contents are not included. Data from: [Rakyan et al., 2010],[Teschendorff et al., 2010],[Bocklandt et al., 2011],[Garagnani et al., 2012],[Bell et al., 2012b],[Hannum et al., 2013],[Horvath, 2013],[Florath et al., 2014],[Weidner et al., 2014],[Benton et al., 2017]. NB some of the datasets shown here are **not independent**, data used to identify aDMRs in some of these studies was also used in others. These data also **do not use consistent cut-off thresholds or check for a consistent direction of change** when including loci in the list of top age associated hits. (Created with UeSatP [Liu et al., 2014].)

The result of the various DNA methylation studies display some seemingly contradictory trends. For example there is the global loss of methylation with age [Wilson et al., 1987],[Fraga et al., 2005],[Heyn et al., 2012] and yet the majority of the highly significant age predicting CpGs in clocks are undergoing hypermethylation by a ratio of very roughly 2:1 [Rakyan et al., 2010], [Bocklandt et al., 2011], [Koch and Wagner, 2011], [Garagnani et al., 2012], [Florath et al., 2014] (see figure 1.17). Interestingly when expanded to look at all age-associated CpGs and not just the most highly correlated CpGs a slightly different picture emerges. Teschendorff et al. [2010] noted in their original study and Zhang et al. [2017] noted in a subsequent analysis, if more loosely age correlated sites are included hypomethylation with age is more prevalent. Zhang et al. [2017] states roughly 60% of sites are hypomethylated and 40% hypermethylated, they also observed that all of the highly age-predictive CpGs which overlapped between the studies they examined were hypomethylated.

It is worth noting that the technologies used to measure the DNA methylation introduce their own biases into global changes in DNA methylation levels. The more Global measures such as WGBS and biochemical measures seem to favour hypomethylation with age [Wilson et al., 1987],[Fraga et al., 2005],[Heyn et al., 2012]. The arrays seem to have agreed more with these global measures as they have increased in size. RRBS also seems to skew in favour of hypermethylation with age [Cole et al., 2017]. These more narrow technologies are biased towards promoters and CGIs with their higher CpG density and lower baseline methylation levels making the prior probability of observing a methylation event greater than demethylation by dint of having a dearth of methylated sites to start with.

Sites which are hypomethylated as their “ground state” are more likely to be in CGIs than sites which are hypermethylated as their “ground state”, and thus hypomethylated sites are more likely to be variable in their methylation state and subject to tight regulatory control than are hypermethylated sites. Expanding the pool to look at CpG sites more poorly correlated with age is likely to introduce more sites whose hypomethylation with age is due to increased stochasticity with age. However from the background information, one would expect that hypermethylated sites in CpG dense regions like CGIs whose methylation level declines with age would make good quality indicators. Given that CpGs in CGIs have a high prior probability of being hypomethylated, it follows that one would be more likely to observe a hypermethylation event by chance. Consequently, a CpG in a CGI that exhibits hypomethylation with age is less likely to be a result of noise than its converse. This may explain the higher reproducibility of these hypomethylated sites.

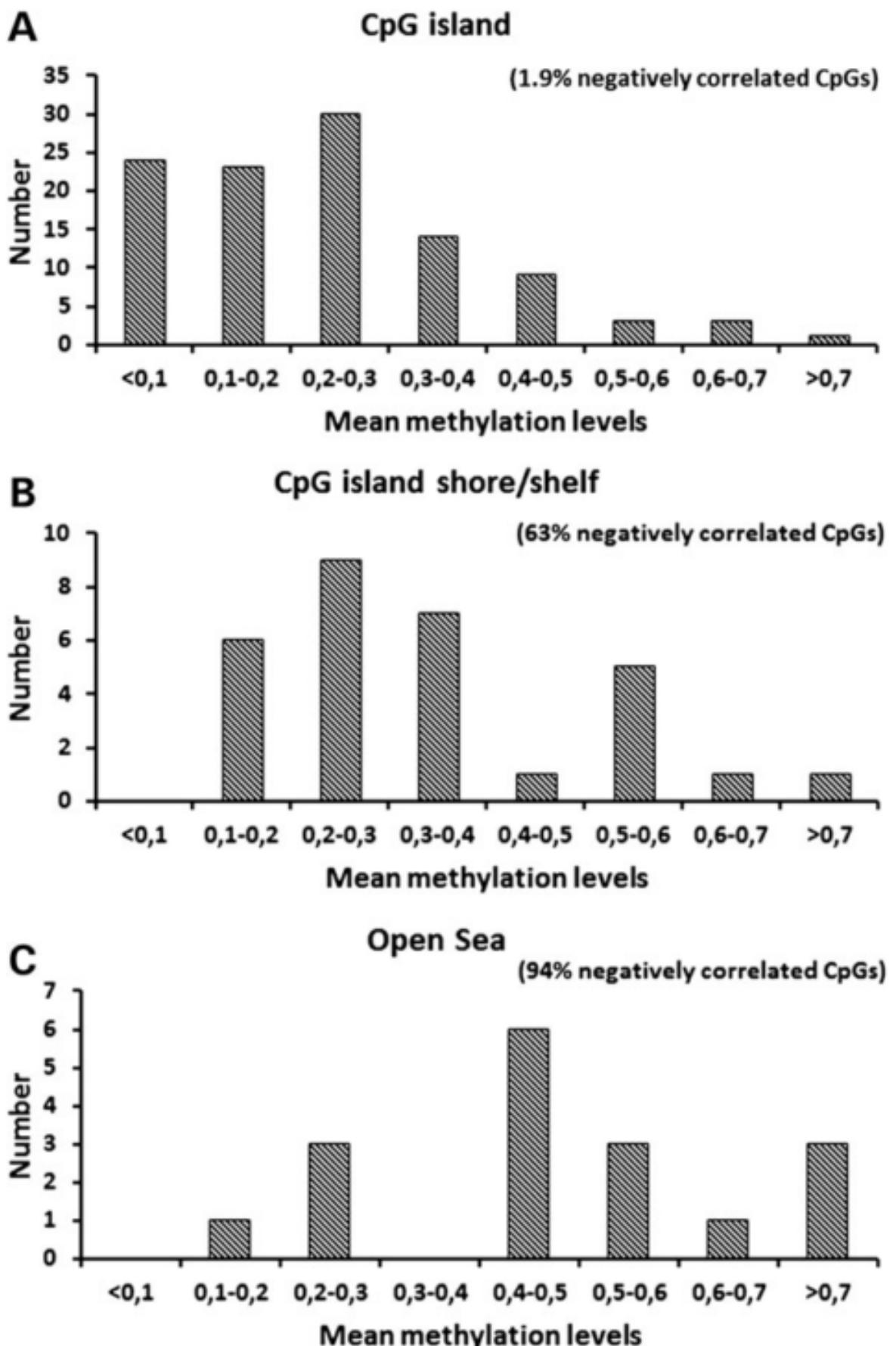


Figure 1.17: Mean methylation level by genomic region showing proportions negatively correlated with age, i.e. become hypomethylated as age increases. (Reproduced from Florath et al. [2014] figure4 a, b, c.)

In a review of DNA methylation and ageing Jones et al. [2015] drew a distinction between epigenetic drift changes and epigenetic clock changes with age wherein they defined drift as “the collection of DNA methylation changes that are associated with age within an individual but are not common across individuals.”, which is slightly different from the definition used by Fraga et al. [2005]. Clock changes are defined as “those sites that are associated with age across individuals and can thus in some cases be used to predict chronological age”. Drift constitutes stochastic change over time not driven by some process which strongly biases the resulting changes in DNA methylation in direction or to specific sites. Clock changes are the converse, some underlying process biases DNA methylation changes in direction and/or position.

This picture is complicated by the tissue specificity of DNA methylation changes. Whilst it is evidently the case that multi-tissue DNA methylation age prediction is possible, it also seems to be the case that an age predictor trained on samples from a specific tissue will outperform one trained on multiple or other tissues when predicting the age of a sample of matching tissue type. This is supported by the out-performance of the Horvath predictor by the Hannum predictor in blood samples [Marioni et al., 2015] and differentially predicted ages of tissues such as breast noted by Horvath [2013]. DNA methylation was highly correlated with age for a variety of tissue-specific sites in many studies, suggesting that the definition of clock sites is frequently tissue dependent. This may also be true of drift changes, stochastic changes may occur at sites in one tissue that would be tightly controlled in another. One could restrict the definition of clock changes further to only those which are present in all/most tissue types to make a multi-tissue predictor.

Apparent changes in the DNA methylation state of a mixed cell population could be due not to global changes in that population but rather its composition, blood exhibits changes in the cellular composition with age [Rimmelé et al., 2014]. As haematopoietic stem cells age their numbers increase and their regenerative potential declines. Their output becomes skewed towards the myeloid lineage but produces poorer quality cells. The number of naive T-cells decreases with age, memory and effector T-cells accumulate as to memory B cells limiting the diversity in the cell-types and contributing to immunosenescence (reviewed in [Geiger et al., 2013]). Cellular composition of tissues can have a substantial impact on epigenome-wide association studies (EWAS), especially when dealing with small effect sizes [Jaffe and Irizarry, 2014]. Thus shifts in cell composition such as those in blood during ageing can be substantial sources of confounding when searching for cell-intrinsic changes in DNA. Some attempts to assay and/or correct for cell type composition have been made, for example, Rakyan et al. [2010] isolated  $CD4^+$  T-cells and found that they had a 60% overlap with age-related CpGs in whole blood. Chen et al. [2016] used the method developed by Houseman et al. [2012] to correct for cell-type composition and found that it did improve the performance of the Horvath and Hannum age predictors. The changes were significant but the effect of this correction was relatively modest. Whilst cell composition has an effect, DNA methylation definitely seems to be capturing something beyond shifts in cell-type composition. However, the age acceleration computed with Hannum’s method, which is trained on blood data only, was more strongly correlated with cell-type composition than Horvath’s method, indicating that extrinsic changes in cell-type composition likely have some impact on the model.

Jenkinson et al. [2017] took an information-theoretic approach to DNA methylation maintenance, modelling the binary methylation state of CpGs as a noisy communications channel. Permitting them to compute the information capacity (CAP), methylation entropy (ENT), and the relative dissipated energy (RDE) of these channels. High capacity channels represent reliable transmission of methylation state, consistent with high RDE as more energy is needed to ensure reliability and low entropy as methylation is more ordered at sites under tighter control than at less highly controlled sites. They found that the Transcription start sites (TSS) exhibited high levels of RDE and high information capacity as well as low levels of entropy these values trended in the opposite direction as they moved outwards to CGIs, shores and seas. This information-theoretic framing provides a useful lens through which to view drift and clock loci.

It could be hypothesised that as a cell differentiates those sites the methylation status of which are most important for successful differentiation will be the focus of the expenditure of cellular resources to set and maintain their methylation state, and will thus exhibit high capacity and low entropy. Whatever mechanism is focusing the resources of the cell on the maintenance of

the methylation at these sites may be responsible for the tissue-specific biases in DNAm changes over time i.e. their targets will be clock sites. Those sites not involved in the maintenance of the specific differentiated state would be a lower priority for the expenditure of cellular resources and thus display lower information capacity and higher entropy i.e. be drift sites. This would also lead one to predict that multi-tissue age predicting loci would be enriched for factors common to the maintenance of a differentiated state. The polycomb and Trithorax group proteins are very broadly evolutionarily conserved and play a key role in developmental cell fate choice and particularly in long-term stable maintenance of epigenetic state [Schuettengruber et al., 2007]. Consistent with this hypothesis Teschendorff et al. [2010] observed that PCGTs are enriched among CpGs associated with age and Levine et al. [2018] also noted that their 513 clock CpGs were enriched for PCGTs. This is also complementary to the idea that DNAm dysregulation with age promotes carcinogenesis by stabilising stem cell character but with reduced capacity for differentiation as proposed by Teschendorff et al. [2010] and Rakyan et al. [2010].

In his paper [Horvath, 2013] Horvath proposed that DNAm age is a measure of the work done by an “Epigenetic Maintenance System” (EMS) which maintains epigenetic stability over time. Considering work in its physical sense of the integral of power over time, Horvath considered power, the rate of change in the use of energy by the EMS as the ‘tick-rate’ of the epigenetic clock. He proposed that during development power would be high due to the high level of stress on the system and that power would drop to a consistent level once maturity was reached. Horvath laid out four predictions of this EMS model and argued that there was evidence in support of each of them. The predictions were:

1. Cancer tissue should show signs of accelerated age, reflecting the protective actions of the EMS
2. Many mitogens, genomic aberrations, and oncogenes, which trigger the response of the EMS, should be associated with accelerated DNAm age.
3. High age acceleration of cancer tissue should be associated with fewer somatic mutations given the protective role of the EMS
4. Mutations in TP53 should be associated with a lower age acceleration of cancer tissue if one further assumes that p53 signalling helps trigger the EMS

I would propose that a fifth prediction of this model is that DNA methylation fidelity should be elevated during development and differentiation as a reflection of the greater amount of work done by the EMS during this period. This is borne out by the results of Zhao et al. [Zhao et al., 2014] who noted a global increase in DNA methylation fidelity during the early differentiation of mouse embryonic stems cells (see figure 1.9).

If Horvath’s Concept of an EMS is accurate and Jenkinson et al.’s RDE captures the energy expended by the cell on the maintenance of the methylation state of a locus, one would expect Horvath’s concept of an EMS to tie in quite nicely with Jenkinson et al.’s methylation channel model. With the Relative dissipated energy (RDE) for a CpG channel corresponding to the work done by Horvath’s EMS at a given locus. Jenkinson et al. did indeed observe high levels of RDE in stem cells and a decline of RDE with age. Interestingly, however, some cancers exhibited decreased RDE, CAP and increased entropy which is consistent with an accelerated ageing profile but seemingly less so with elevated activity of an EMS. It is possible that whilst the distribution of RDE at various loci in the genome shifts giving a lower mean RDE that the total amount of energy expended is increasing but becoming less well targeted such that whilst there is increased energy flux through the EMS it is not being properly directed. There may also be a plateau effect wherein the energy flux through the EMS is maxed out and cannot keep up with demand from many loci. One would also predict more rapid equilibration of DNAm levels at loci with high RDEs, which could be examined by looking at the second derivatives for the stochastic model of DNA methylation described by Pfeifer et al. [1990] using the experimental systems Laird et al. [2004] and Riggs and Xiong [2004] used which supported the original model.

Jenkinson et al. [2017] found a global increase in entropy with age, but not with cell passage in culture suggesting an increase in entropy is associated with epigenetic age independent of mitotic age. This observation is in agreement with the finding that epigenetic state becomes

more stochastic and diverges with age [Fraga et al., 2005] & [Slieker et al., 2016], indeed genes whose methylation were most divergent with age were enriched for ageing associations. They also observed a loss of entropic sensitivity with age. Entropic sensitivity is an indicator of how plastic DNAm state is to extrinsic effects. Jenkinson et al. noted that there is a general loss of phenotypic plasticity with age, but cited no specific instance of loss of DNAm plasticity to environmental effects with age. Hahn et al. [2017] provided a possible example of this effect when they reported that dietary restriction caused fewer differences in the methylation state of older than younger mice.

### 1.3.7 Conclusions

DNAm age is a high-quality biomarker of ageing which can be tailored to the particular requirements of the context, be it estimating the chronological age of individuals, measuring age acceleration/deceleration of individuals relative to their chronological age or estimating age disparity of tissues. There is predictably a cost/accuracy trade-off. DNAm age is to some degree capturing “biological age” and likely not being driven by changes due to specific ageing related diseases as illustrated by lack of relationships with specific pathologies and phenotypes of ageing noted by Bell et al. [2012b]. The PhenoAge clock is a better predictor of morbidity and mortality than age acceleration but may not be a better predictor of underlying biological age if the risk it is capturing is driven more by disease than that captured by age acceleration. PhenoAge is strongly associated with comorbidity count, more so than either the Hannun or Horvath clocks, interestingly zero comorbidities are associated with a negative PhenoAge score suggesting some of the PhenoAge signal is derived from disease status. Disease may, of course, interact with biological age causing an acceleration thereof, making them difficult to disentangle. Neither Age acceleration nor PhenoAge perfectly captures biological age but both are highly informative on the subject

Global DNA methylation decreases with age, DNA methylation becomes more disordered and less plastic to environmental influences with age. A set of loci, not necessarily individual CpGs but rather some functional unit of methylation exhibit consistent predictable changes with age. Which loci are in this set varies with tissue type; there is, however, a subset which seems to be common to most tissues. The preponderance of these loci are in areas of high to moderate CpG density and become hypermethylated with age. As the high frequency of CpGs and the hypomethylated “ground state” of CGIs would lead one to expect a bias towards being located in CGIs and becoming hypermethylated it is unclear if changes in DNA methylation over time are biased towards regions of high CpG density and hypermethylation at a level that is greater or less than one would predict with the prior information.

For other reviews of DNA methylation and ageing which have informed this work see: [Jung and Pfeifer, 2015], [Jones et al., 2015] and [Horvath and Raj, 2018].

## **Part II**

# **Methods**



# Chapter 2

## Methods

### 2.1 DNA methylation assays

#### 2.1.1 Overview

There are numerous DNA methylation assays. Some are designed to assay global DNAm levels producing a single global measurement of the amount of 5mC present in a sample, others assay DNAm in a manner traceable to a specific genomic locus. These methods can be further subdivided into targeted methods where the loci to be examined are known in advance, and those where the loci covered are quasi-randomly sampled. Both the targeted and untargeted methods vary in their granularity, from individual CpGs to large regions of the genome, on the order of megabases. These assays also vary in their coverage of the genome from locus-specific approaches looking at small numbers of individual loci to whole genome methods providing methylation information on all ~28 million CpG sites. Between these two extremes are ‘epigenome-wide’ methods which are focused with varying degrees of specificity on genomic regions of interest.

These DNA methylation assays generally exploit one of four following things:

1. **Methylation sensitive restriction digestion** A variety of methods exploit enzymes which differentially cut methylated and unmethylated DNA. An example of such enzymes is: HpaII and MspI which are isoschizomers for the sequence: 5'-C<sup>^</sup>CGG-3' but only HpaII methylation sensitive and unable to cut the sequence if the internal cytosine is methylated. The methods which use such enzymes include: Restriction landmark genomic scanning, which gives a roughly megabase resolution indicators of DNAm levels by 2D electrophoresis [Hatada et al., 1991]. MRE-seq, which makes use of methylation-sensitive restriction digests to enrich for unmethylated DNA [Maunakea et al., 2010]. This is then sequenced to provide the genomic location of this unmethylated signal. Reduced Representation Bisulfite Sequencing (RRBS) which using methylation sensitive restriction digestion to enrich for unmethylated sequences which are then subject to bisulfite sequencing, (see point 3 below).
2. **DNA methylation-sensitive binding of DNA by antibodies or other proteins which bind methylated DNA** Anti-5mC antibodies were first used to isolate methylated DNA in 1985 and were subsequently paired with array-based technologies to assay DNAm levels at specific loci in 2005 [Weber et al., 2005, Harrison and Parle-McDermott [2011]]. This was followed by MeDIP-seq (methylated DNA immuno-precipitation and sequencing) [Down et al., 2008, Lienhard et al. [2014]], which enriches methylated DNA that is then sequenced and the number of reads mapping to a locus is indicative of the relative methylation level. MBD-seq uses a recombinant Methyl Binding Domain (MBD) protein to enrich for methylated double-stranded DNA, prior to sequencing [Serre et al., 2010]. Inferring the absolute methylation levels form the number of reads mapping to a locus enriched by one of these pull-down methods is complicated by variation in CpG density

in the genome and the fact that CpG dense regions tend to have low methylation levels, and thus tend not be pulled down by antibodies binding 5mC or MBD proteins. The best results for estimating absolute methylation levels with these methods come from pairing them with a complementary method to enrich for unmethylated sites such as MRE-seq (described above) [Li et al., 2015]. Though relative methylation levels remain effective in identifying differentially methylated regions when using these methods alone [Bock et al., 2010, Maunakea et al. [2010]].

3. **Sodium bisulfite conversion of unmethylated cytosines to uracil** Conversion of cytosine to uracil changes the base complementary to this site from G to A [Hayatsu et al., 1970]. This basic chemistry of the conversion process is illustrated in Figure 2.1. This conversion can be detected with a variety of other technologies from targeted bisulfite PCR to examine a small number of selected loci, to the Illumina bead chip methylation arrays or WGBS. Targeted bisulfite PCR can be performed on a larger number of loci with microfluidic multiplexing such as that provided by the Fluidigm access array [Adamowicz et al., 2018]. It is worth noting that bisulfite conversion cannot differentiate between methylated and hydroxymethylated cytosines. Because 5hmC represents a small fraction of modified bases compared to 5mC many analyses have made the working assumption that unconverted bases are methylated, but it is beginning to be recognised that distinguishing between the two may be biologically important especially in tissues where 5hmC is more prevalent such as neurons. A variant of bisulfite conversion, oxidative bisulfite conversion, exists which can permit 5hmC to be distinguished from 5mC [Skvortsova et al., 2017].

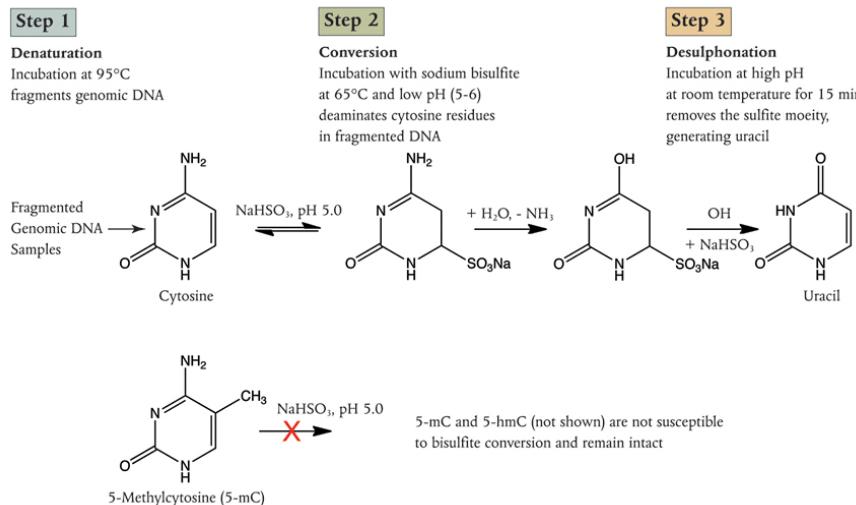


Figure 2.1: Sodium Bisulfite conversion of C to U, 5mC and 5hmC are unaffected. (Image Credit New England Biolabs).

4. **Physical differences in the methylated base** These are exploited by the not yet widely used ‘3rd generation’ sequencing technologies such as nanopore and single-molecule real-time sequencing. These methods respectively exploit the effects of modified bases on ion flow through the nanopore and impact on incorporation time of new bases whilst in the polymerase [Rhoads and Au, 2015, Simpson et al. [2017]].

## 2.1.2 Methods used in this work

### 2.1.2.1 MeDIP-seq

MeDIP-seq uses a monoclonal anti-5mC antibody to bind denatured fragmented genomic DNA at methylated CpG sites, this antibody-bound fraction of DNA is isolated and sequenced [Down

et al., 2008], Figure 2.2 outlines the workflow. Unlike bisulfite conversion approaches this method permits 5mC to be differentiated from 5hmC. The methylation level across the CpG sites in the region of the genome to which the resultant sequencing reads map can subsequently be estimated by counting the number of reads and accounting for the CpG density with software tools such as the MEDIPS R package [Lienhard et al., 2014].

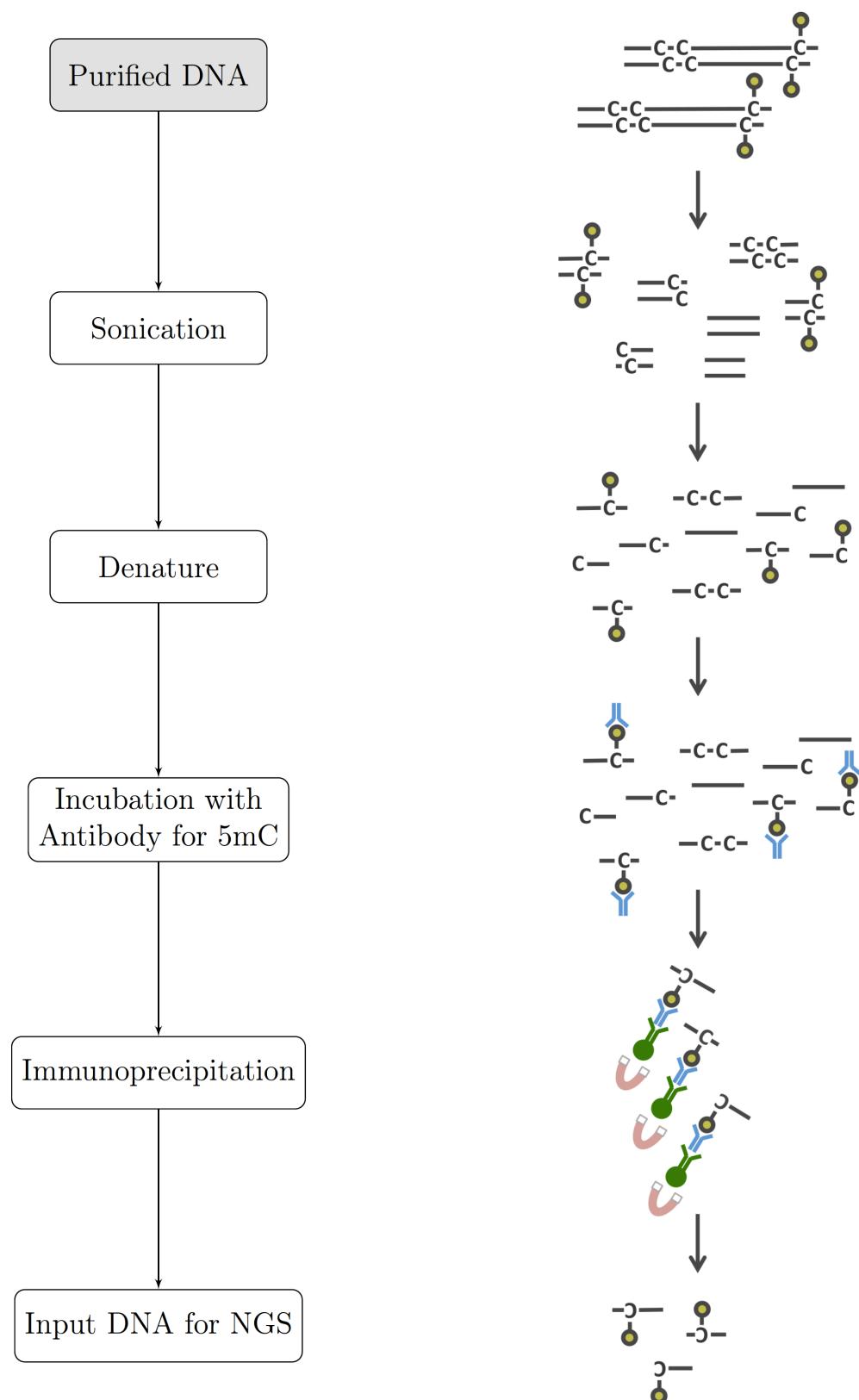


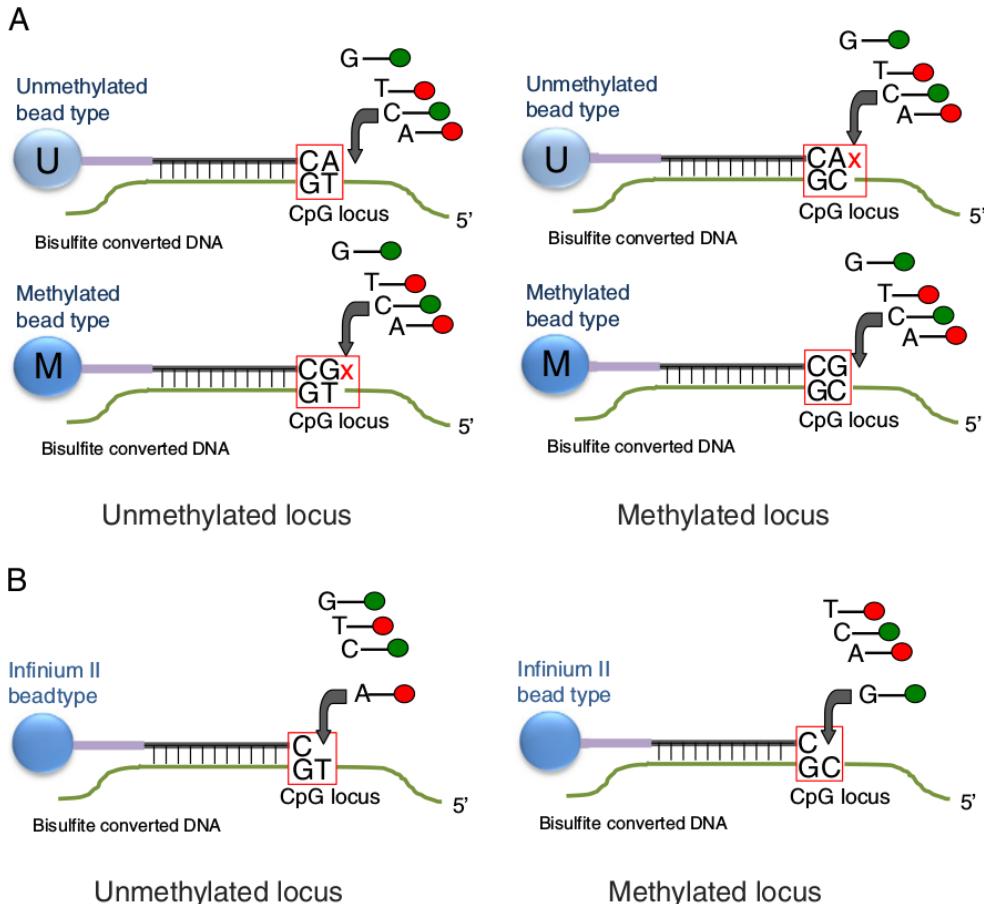
Figure 2.2: Graphical Summary of the MeDIP-seq process.

### 2.1.2.2 Bisulfite Conversion methods

#### 2.1.2.2.1 Illumina Methylation arrays

The Illumina methylation bead chip arrays make use of 50 bp long probes with sequences corresponding to the loci of interest in the genome. These probes are attached by their 5' ends to beads which are stuck to the surface of a glass slide, several beads represent each locus at a single spot on the slide. The positions of the bead spots are known and used to identify the probes. The CpG of interest is located at the 3' end of the probes. Samples are prepared for the array assay by fragmentation, bisulfite conversion and amplification. These sample fragments anneal to complementary probes on the array and unbound DNA is washed away. A single DNA polymerase-mediated elongation step is carried out to permit a fluorescently labelled base to be incorporated at the end of each probe. C and G are tagged with one fluorophore and A and T with another. There are two probe designs one which makes use of two distinct bead types for each locus, the type I probes, and one which makes use of a single type of bead at each locus, the type II. A scanner excites the two fluorophore channels with a laser and their respective fluorescence intensities at each spot are recorded with an imaging system. The ‘colour’ of the fluorophore incorporated signifies the methylation state and the intensity of the proportion of sites which are in that state. The logic of which ‘colour’ corresponds to methylated or unmethylated varies with the probe type and Figure 2.3 illustrates and explains this in depth.

These arrays have now undergone several iterations [Bibikova et al., 2009, Bibikova et al. [2011], Sandoval et al. [2011]]. The current array probe design was preceded by the ‘golden gate’ array which used a variant on the SNP probe design with methylation-specific PCR. This array contained ~1.5k probes covering 371 genes with a strong focus of cancer-related genes [Bibikova, 2006]. The first array using the probe design outlined above was the ‘Infinium’ array which contained 27k type I probes focused on the promoter regions for 14,475 consensus coding genes with 110 miRNA promoters [Bibikova et al., 2009]. The 450k array which had ~480k mixed type I and II probes covering 21,231 RefSeq genes, 26,658 CpG islands, 80,538 predicted enhancer regions and a variety of other features, including the MHC region [Bibikova et al., 2011]. The ‘EPIC’ array has ~850k probes including >90% of those on the 450k with greatly expanded coverage of loci with more dynamic methylation states than promoters and CGIs which tend to have relatively low methylation variability aiming to capture more functional methylation variation at regions such as enhancers [Sandoval et al., 2011, Ziller et al. [2013]]. The ‘EPIC’ array also contains both Type I and Type II probes, with many of the new sites being type II. The type I probes have greater dynamic range than the type II but take up twice as much space on the array so there is some trade-off between maximising the number of sites covered and the quality of data at those sites.



**Figure 2.3: Illumina Methylation Bead Chip Array Probe Designs** The probes on the array are 50bp in length. In the type I probes unlike the type II probes the base that is incorporated (or not) in the elongation step is the base that follows the CpG site. The probes on the array are bound by the amplification products derived from the bisulfite converted DNA not direct conversion products. Thus, Cs in the sample DNA represent methylated positions and Ts unmethylated positions, as Ts take the place of Us created by bisulfite conversion during the amplification process. **Type I:** Panel A shows the type I probe design. Each locus is represented on two beads, a methylated and an unmethylated bead. *Methylated Sample:* When sample DNA binds a methylated probe the G at the second position in the CpG on the methylated probe will be complementary to the C marking the location of the 5mC. Extension will then be able to occur incorporating a fluorescently labeled base complementary to the base just 5' of the 5mC position in the probe. On the unmethylated probe the C marking the position of the 5mC will not be complementary to the A base in the probe and thus extension cannot occur. No fluorescent base will be incorporated at the unmethylated probe. *Unmethylated Sample:* When sample DNA binds the methylated probe the T marking the position of the unmethylated C is not complementary to the G at the second position in the CpG site on the probe. Thus, no fluorescent base is incorporated. Whereas, on the unmethylated probe the T marking the position of the unmethylated C will be complementary to the A at the second position in the CpG site on the probe. Thus, a fluorescently labelled base can be incorporated. **Type II:** Panel B shows the Type II probe design. For the Type II probes only one probe represents each locus. In the Type II probes the incorporated base is at the second position in the CpG site. A fluorescently labelled A is always incorporated opposite a T in the sample DNA marking the position of an unmethylated C, and a fluorescently labelled G is always incorporated opposite a C marking the position of an unconverted 5mC in the sample DNA. For the Type II probes, in contrast to the Type I probes, methylation is always signalled in the green channel and unmethylation in the red channel. (figure reproduced from [Bibikova et al., 2011].)

Methylation at each site on the array is commonly reported as a  $\beta$  value, which correspond to the proportion of the sample DNA which was methylated at that site.

$$\frac{\beta = \text{intensity}[M]}{\text{intensity}[U] + \text{intensity}[M]}$$

Where  $M$  = methylated and  $U$  = unmethylated.

### 2.1.2.2.2 Sequencing Methods

#### 2.1.2.2.2.1 Whole Genome

Whole Genome Bisulfite Sequencing (WGBS) theoretically permits all ~28 million CpGs in the genome to have their methylation state assayed (excluding unmappable and other problematic regions) and thus represents a gold standard for characterising the methylome. However, WGBS is highly inefficient with 70-80% of the reads generated providing no information about methylation as many reads map to loci without CpG sites [Ziller et al., 2013]. In addition, to reliably call DNA methylation levels at all sites with a dynamic range equivalent to other popular technologies such as the Illumina array would require reads depths on the order of 85x [Libertini et al., 2016]. Given that read depth is semi-randomly distributed this means that it will likely not always be possible to get reliable calls at sites in which investigators are interested, without using what are a present prohibitively expensive read depths.

WGBS is very similar in principle to ordinary whole genome sequencing but preceded by a bisulfite conversion step as illustrated in Figure 2.1 [Bock et al., 2010]. However, bisulfite conversion has several implications for downstream processing. Bisulfite library prep can be either directional, sequencing reads correspond to a bisulfite-converted version of either the forward or reverse strand, or non-directional in which sequencing reads correspond to bisulfite-converted versions of either strand giving a total of four bisulfite converted sequences with the strand of origin unknown. If a library is directional there are only two bisulfite converted sequences.

Mapping is usually done with customised software wrapper to implementations of existing alignment tools. Bismark is a popular example of such as wrapper which uses the bowtie alignment tool and calls site methylation levels [Tran et al., 2014]. Alignment is performed with *in silico* bisulfite converted versions of the genome and reads, with G to A conversions for reverse strand reads. This requires four parallel instances of bowtie for forward and reverse G to A and C to T conversions, which are combined to determine the unique best alignment. (Figure 2.4 panel A). Bisulfite conversion results in loss of sequence complexity as many Cs effectively become Ts meaning short reads can be challenging to map. Bismark produces methylation calls either combined or by strand and which can be filtered by methylation context CpG, CHG or CHH.

Information theoretic analysis techniques which exploit the correlation of methylation levels between neighbouring CpG sites [Haerter et al., 2014, Affinito et al. [2016]] to call methylation levels in small regions have recently been developed. They also provide additional insight into the variability of methylation at a given site and reliability with which the methylation state will be maintained even in relatively low coverage WGBS data [Jenkinson et al., 2018].

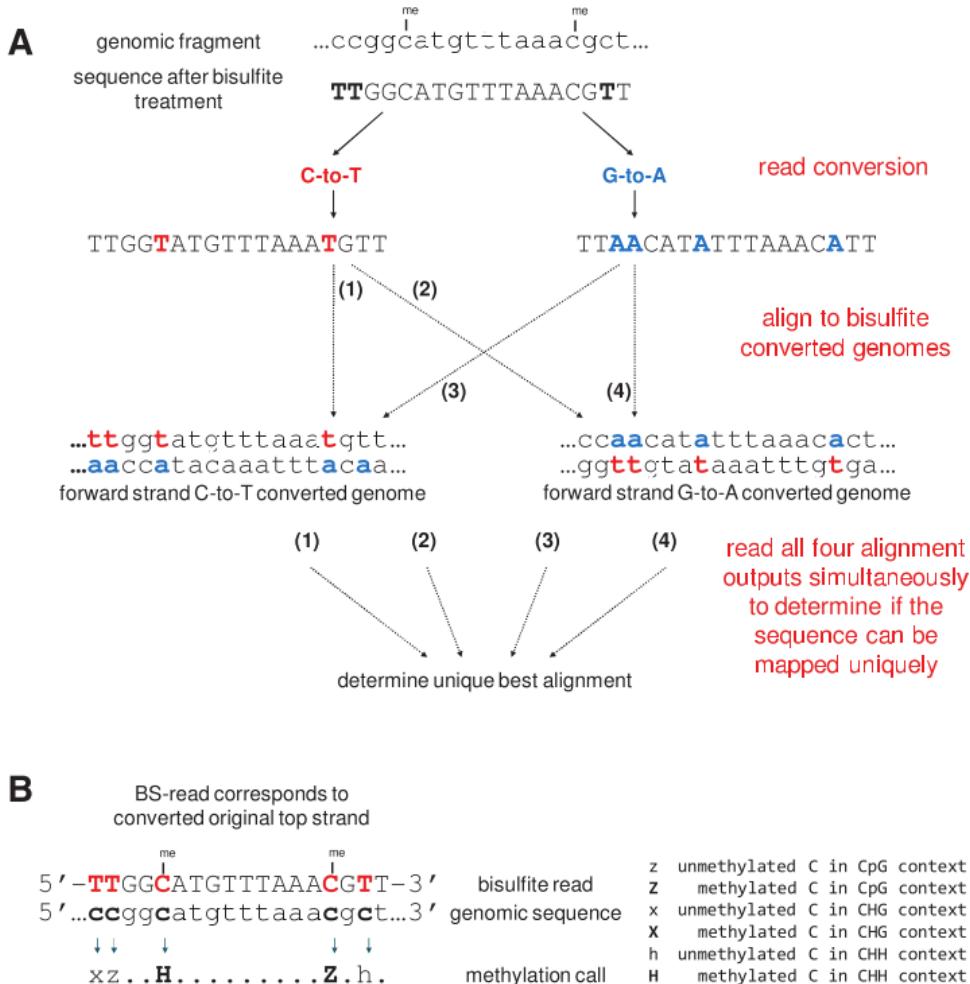


Figure 2.4: [Krueger and Andrews, 2011]

#### 2.1.2.2.2 Targeted

Targeted bisulfite sequencing can be carried out with a variety of methods for isolating the sequences of interest. Among them is bisulfite PCR, where the genome is bisulfite converted and PCR carried out with specially designed primers complementary to the regions flanking bisulfite converted sequence of interest. Tools for designing such primers are available, for example: ‘methPrimer’ [Li and Dahiya, 2002]. Subsequent analysis is similar to that for WGBS, the PCR products are sequenced. The resulting reads aligned and methylation levels called with a tool such as Bismark.

## 2.2 Discussion

In summary, there are a wide range of DNA methylation assays and analytical techniques available, each with strengths and weaknesses. Thus, the choice of assay and analysis will need to be informed by the biological question being asked. For example, the 450k array would not be well suited to study the methylation of repetitive elements as its coverage in these regions is low. WGBS may be prohibitively expensive and MeDIP-seq may represent a good compromise, as it would provide good coverage of these regions and permit accurate detection of differential methylation. With the caveat that MeDIP-seq can be highly sensitive to genetic variation affecting CpG numbers [Okitsu and Hsieh, 2015]. This approach could potentially be improved by

coupling MeDIP-seq with MRE-seq, as in conjunction they would provide an improved ability to discern absolute methylation levels compared to MeDIP-seq alone [Clark et al., 2012, Bock et al. [2010]]. Unlike the 450k array MeDIP-seq captures regional methylation levels, whilst this may seem like a disadvantage due to lower resolution it can result in a more accurate and biologically relevant picture of regional methylation especially in highly methylated regions. The 450k array systematically underestimates the methylation level in highly methylated regions [Clark et al., 2012], this may well be because it samples only single sites, and highly methylated regions are often not fully methylated. Modeling work by Affinito et al. [2016] suggests that once a threshold level of DNA methylation is reached in a region further methylation becomes less likely, meaning that sampling a single site is likely to produce a systematic underestimate. MeDIP-seq can also cover much more of the methylome than does the 450k array. The 450k array captures ~1.6% of the 28 million CpGs in the genome. At near saturation coverage of ~40 million reads per sample MeDIP-seq can cover ~60% of all CpGs with at least 1 read, this is almost all methylated CpGs. At half that total read count MeDIP-seq still covers ~50% of CpGs at  $\geq 1\times$  and ~20% at  $\geq 5\times$  [Taiwo et al., 2012].



# **Part III**

# **Results 1**



# **Chapter 3**

## **Assessment of Umbilical Cord Tissue DNA Methylation with Respect to Bone Health, Environmental Factors and Ageing**

### **3.1 Abstract**

Maternal vitamin D levels during pregnancy have previously been shown to impact on the methylation status of specific genes associated with vitamin D signalling and metabolism, such as *RXRA*. Long-term bone health outcomes have also been associated with both maternal vitamin D during pregnancy and methylation of the *CDKN2A* gene. To attempt to identify other loci the epigenetic regulation of which is pertinent to the impact of maternal vitamin D levels on long-term bone health Epigenome-Wide Association Studies (EWAS) were performed with DNA methylation data from the Illumina EPIC & 450k methylation bead chip arrays (n=137 & n=60, respectively). The impact of intervention vs placebo status in a trial of vitamin D supplementation in pregnancy, neonatal bone mineral content, Maternal Vitamin D levels (34wks) and change in maternal Vitamin D levels from 11-34wks, on methylation were examined. No loci were identified as changing significantly in methylation status in any of these EWAS, however, the power of this study was such that changes with small to medium effect sizes may have gone undetected. It is our intention to extend this analysis with additional samples to increase the power. We also intend to make use of this data to assess epigenetic estimates of biological and gestational age.

### **3.2 Introduction**

#### **3.2.1 Background**

Epigenetic states are a product of both genetics and environment, they exist on a continuum from complete or obligatory genetic influence through facultative variation influenced by genetics, to the purely epigenetic subject entirely to environmental factors. This picture is complicated by the observation that genetics can not merely influence epigenetic state but also the extent to which that state will vary depending on environmental factors [Holland et al., 2016]. Epigenome-wide association studies have been performed for many traits including age (Section 1.3.3), smoking

status [Dogan et al., 2017], exposure to atmospheric particulate matter [Dai2016], and obesity [Wahl et al., 2017]. Some robust and reproducible epigenetic changes have been identified with EWAS, particularly for traits such as age and smoking. Unlike genome-wide association studies (GWAS) associations found by EWAS can be due to reverse causation. That is to say, a genetic variant associated with a disease is very unlike to have been caused by the disease state whereas an epigenetic association can be the consequence of a disease state. Whilst this can complicate the dissection of disease aetiology it can also be a boon in unpicking the biology that follows on from a particular disease or environmental exposure by revealing what biological networks are affected.

Work in rat models has demonstrated the ability of maternal nutritional exposure during pregnancy to influence the methylation and expression of genes in their offspring [Burdge et al., 2007b, Lillycrop et al. [2008]] (reviewed in [Burdge et al., 2007a]). Consequently, the methylation status of target genes of interest for vitamin D and bone development related pathways have been examined.

Retinoid-X-receptor-alpha (RXRA), which forms a heterodimer with the vitamin D receptor is essential for the nuclear action of  $1,25(OH)_2$  – Vitamin D. DNA methylation at four of ten CpG sites in the *RXRA* gene were significantly ( $p \leq 0.05$ ) lower in the umbilical cord of offspring from cholecalciferol supplemented mothers compared to placebo with a mean difference in methylation of -1.98% ( $n=447$ , 95% CI -3.65 to -0.32,  $p=0.01$ ). One CpG site in the gene is related to estimated free  $25(OH)D$  levels [Harvey et al., 2014].

Methylation of cyclin-dependent kinase inhibitor 2A (CDKN2A) in umbilical cord tissue has been implicated in bone cell activity mediating skeletal development and homeostasis. The *CDKN2A* locus has quite complex biology encoding two cell cycle inhibitors *p14<sup>ARF</sup>* and *p16<sup>INK4a</sup>* as well as the long non-coding RNA *ANRIL* which inhibits *p16<sup>INK4a</sup>*. Targeted analysis of nine CpGs in a 300bp stretch of the *ANRIL* promoter region of *CDKN2A* was carried out. Methylation at several of these sites showed an inverse correlation with bone size, mineral content, and mineral density at age four years [Curtis et al., 2017].

The MAVIDOS trial is a randomised, double-blind, placebo-controlled trial of vitamin D supplementation in pregnancy focused on bone development outcomes [Harvey et al., 2012, Cooper et al. [2016]]. Motivated by findings that low levels of maternal  $25(OH)$ -Vitamin D during gestation have been associated with reduced bone mass, extending to peak bone mass [Zhu et al., 2014] and altered femoral morphology; leading to increased risk of osteoporosis in later life [Mahon et al., 2010, Viljakainen et al. [2010], Viljakainen et al. [2011]].  $1,25(OH)_2$  – Vitamin D has been implicated in the control of Plasma membrane  $Ca^{2+}$ -ATPases (PMCA). The expression of the *PMCA3* gene in the placenta has been associated with umbilical cord calcium concentration and intrauterine accrual of bone mineral content [Kip and Strehler, 2004]. This provides part of a candidate mechanism linking vitamin D to bone outcomes.

### 3.2.2 EWAS for Maternal Vitamin D Status and Offspring Bone Outcomes

This work sought to identify additional loci with changes in DNAm associated with bone development outcomes and vitamin D, beyond those identified in previous work targeted to specific loci. An ‘epigenome-wide’ approach examining ~850k sites in the genome was carried out using the Illumina EPIC methylation bead chip array on umbilical cord tissue samples from participants in the MAVIDOS study ( $n=140$ ). Results were compared with additional samples run on the Illumina 450k array ( $n=60$ ), which as the name suggests covers only ~450k sites.

Epigenome-Wide Association Studies (EWAS) were performed for four variables of interest: intervention vs placebo status, neonatal bone mineral content(g) as determined by dual-energy X-ray absorptiometry (DXA) scans, Maternal Vitamin D levels (34wks) and change in maternal Vitamin D levels from 11 to 34wks. Subsequent analyses planned with this data include: Testing the effectiveness of epigenetic predictors of gestational age developed in cord blood [Bohlin et al., 2016, Knight et al. [2016]] in cord tissue samples. Potentially testing if ‘phenotypic age’

as predicted by DNA methylation [Levine et al., 2018] is related to gestational age, unlike DNA methylation age [Simpkin et al., 2016, Simpkin et al. [2017a]].

## 3.3 Methods

Analysis of the Illumina EPIC and 450k methylation arrays was carried out in the R language using the ‘meffil’ package [Min et al., 2017], which is capable of performing functional normalisation in a more memory efficient fashion than alternatives such as ‘minfi’.

### 3.3.1 Quality Control (QC)

#### 3.3.1.1 Whole array QC

##### 3.3.1.1.1 EPIC arrays

The predicted sex of the samples generated using sex chromosome probe intensities was checked against that in the sample annotation and two mismatches were found. These were MAVIDOS IDs 206 and 63 and the associated arrays were excluded from further analysis, (Figure 3.1).

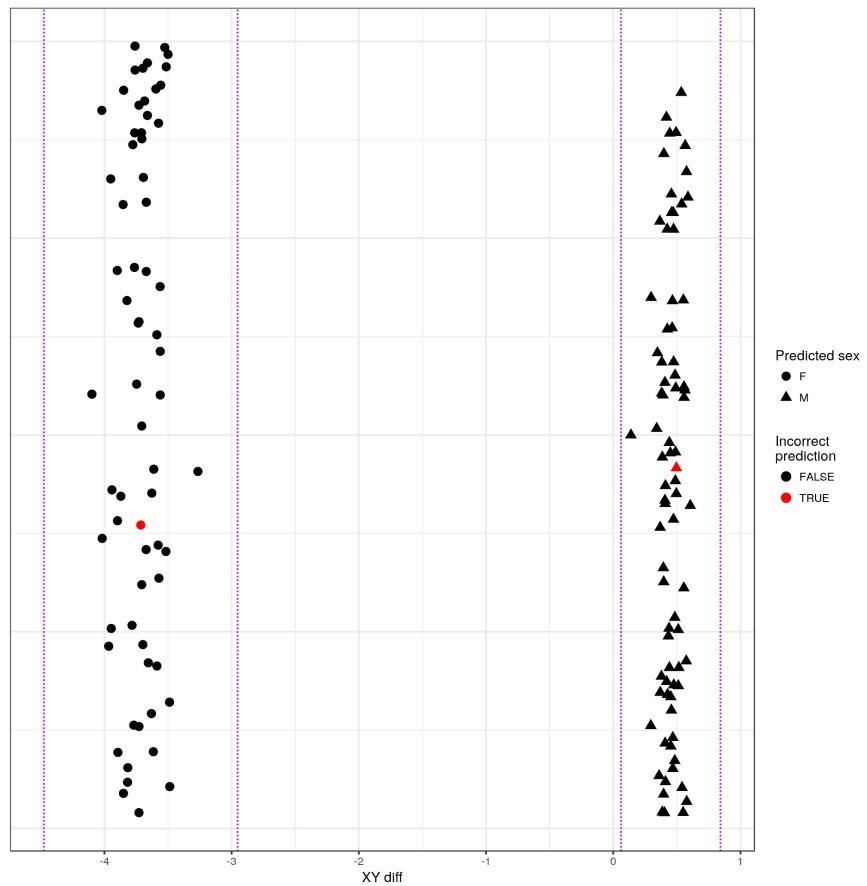


Figure 3.1: Predicted sex of each sample based on the sex chromosome copy numbers inferred from probe intensities for the EPIC array data. Mismatches between the predicted sex and that asserted in the sample annotation metadata are shown in red. Two predicted sex values differ from their annotations. Plot generated by meffil QC report.

The dataset also contained four samples for which there were two technical replicates, only the first replicate from each was used (144 arrays run for 140 individuals). Array 201516310023

(MAVIDOS ID 95) was excluded as its median methylated signal was more than  $3\sigma$  from the expected value, (Figure 3.2). No samples were excluded for having a higher than expected proportion of undetected probes (proportion of probes with detection p-value  $> 0.01$  is  $> 0.1$ ) (Figure 3.3). No samples were excluded for having a high proportion of probes with low bead counts (proportion of probes with bead number  $< 3$  is  $> 0.1$ ), (Figure 3.4).

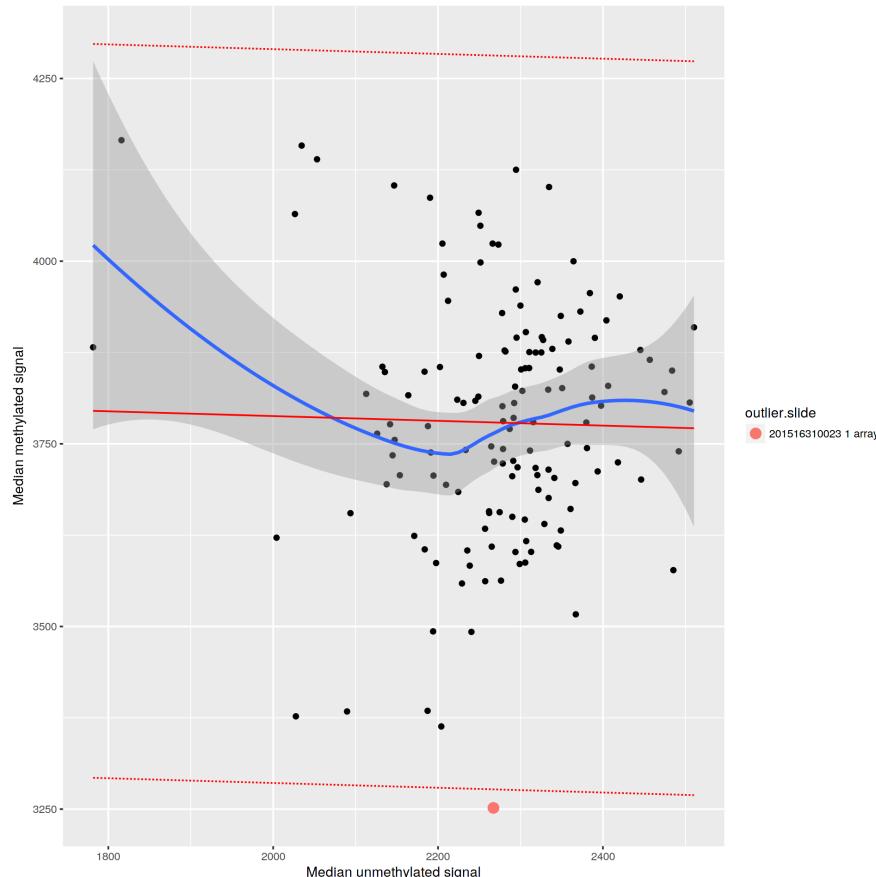


Figure 3.2: Median methylated signal vs unmethylated signal per sample for the EPIC array data, solid red line indicates linear regression of median methylated signal vs median unmethylated signal with dotted red lines representing  $3\sigma$  from the expected mean. Samples outside the expected range are indicated in the legend. Plot generated by meffil QC report.

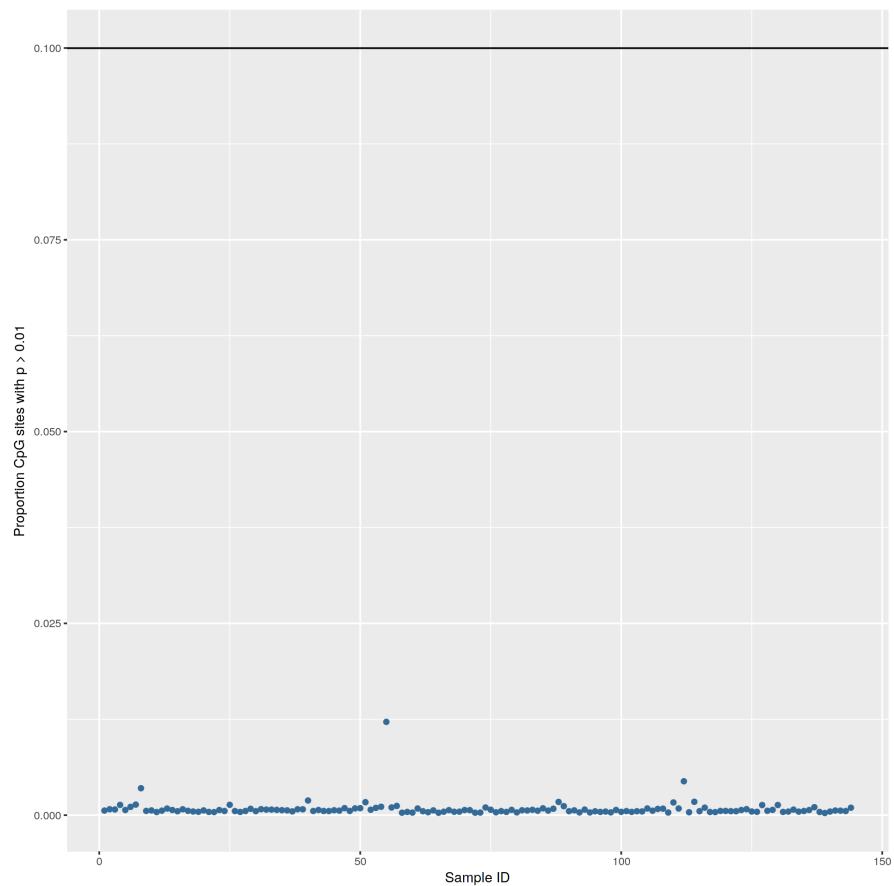


Figure 3.3: Proportion of probes with detection p-values  $>0.01$  by sample for the EPIC array data. Black line indicates the exclusion threshold of 0.1. Plot generated by meffil QC report.

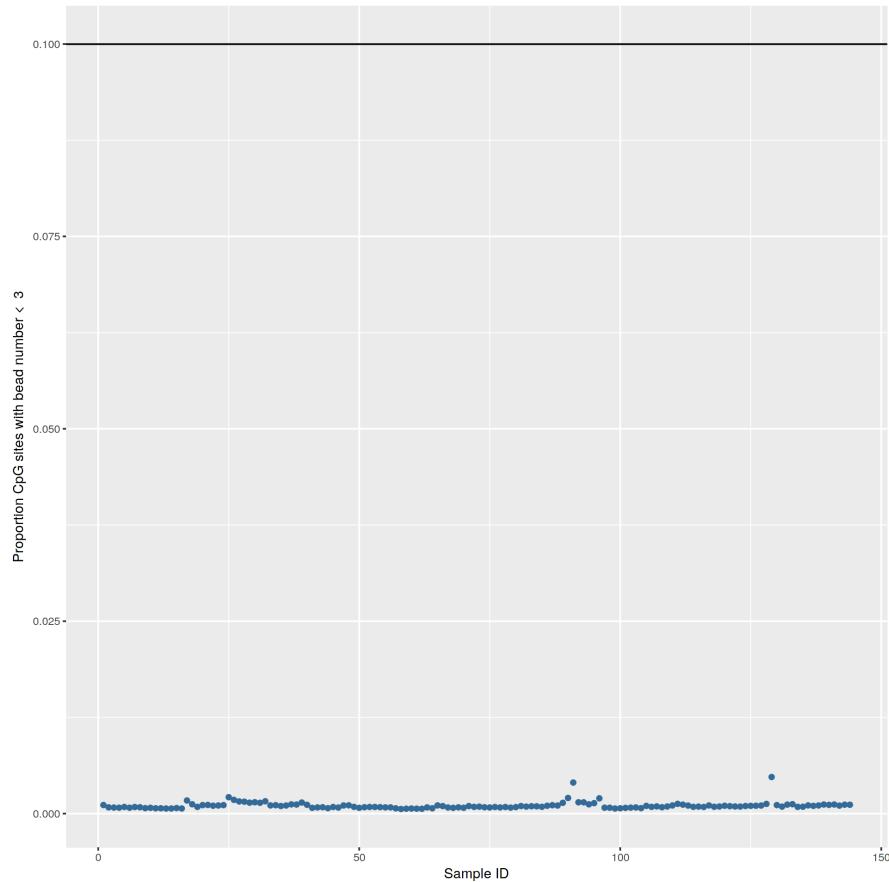


Figure 3.4: Proportion of probes with a bead count of  $< 3$  by sample for the EPIC array data. Black line indicates the exclusion threshold of 0.1. Plot generated by meffil QC report.

### 3.3.1.1.2 450k Arrays

There were no mismatches between predicted and annotated sex (Figure 3.5). There were not any samples with outliers in their methylated / unmethylated probe proportions (Figure 3.6). No samples were excluded for having a higher than expected proportion of undetected probes (proportion of probes with detection p-value  $> 0.01$  is  $> 0.1$ ), (Figure 3.7). No samples were excluded for having a high proportion of probes with low bead counts (proportion of probes with bead number  $< 3$  is  $> 0.1$ ), (Figure 3.8).

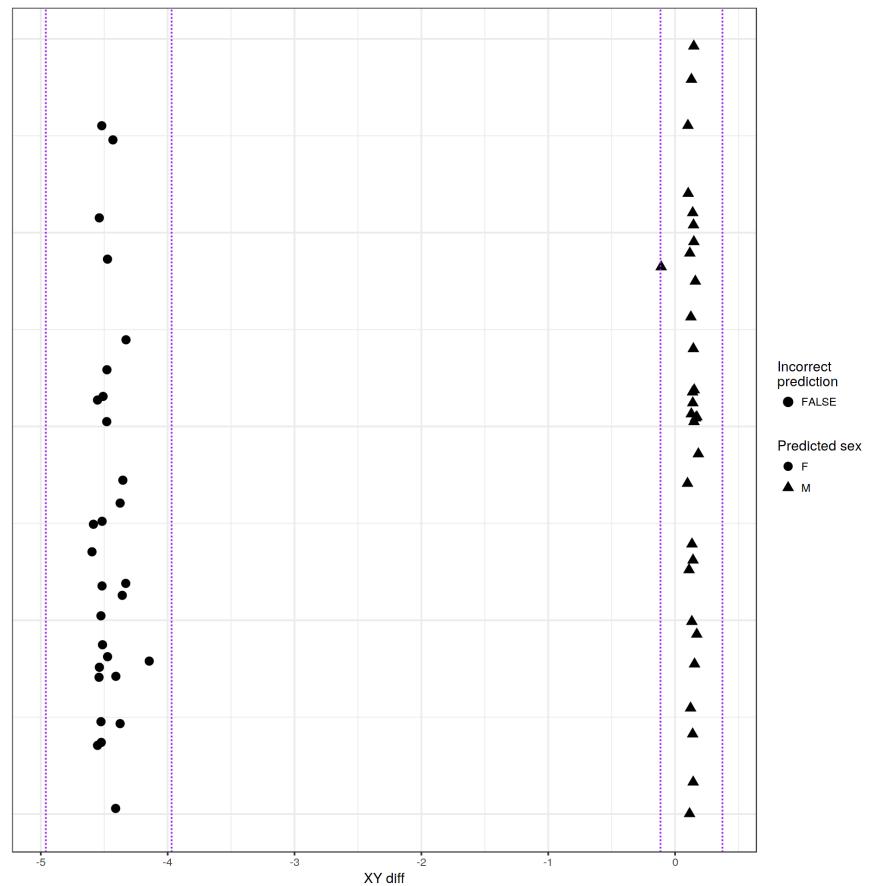


Figure 3.5: Predicted sex of each sample based on the sex chromosome copy numbers inferred from probe intensities for the 450k array data. No predicted sex values differ from their annotations. Plot generated by meffil QC report.

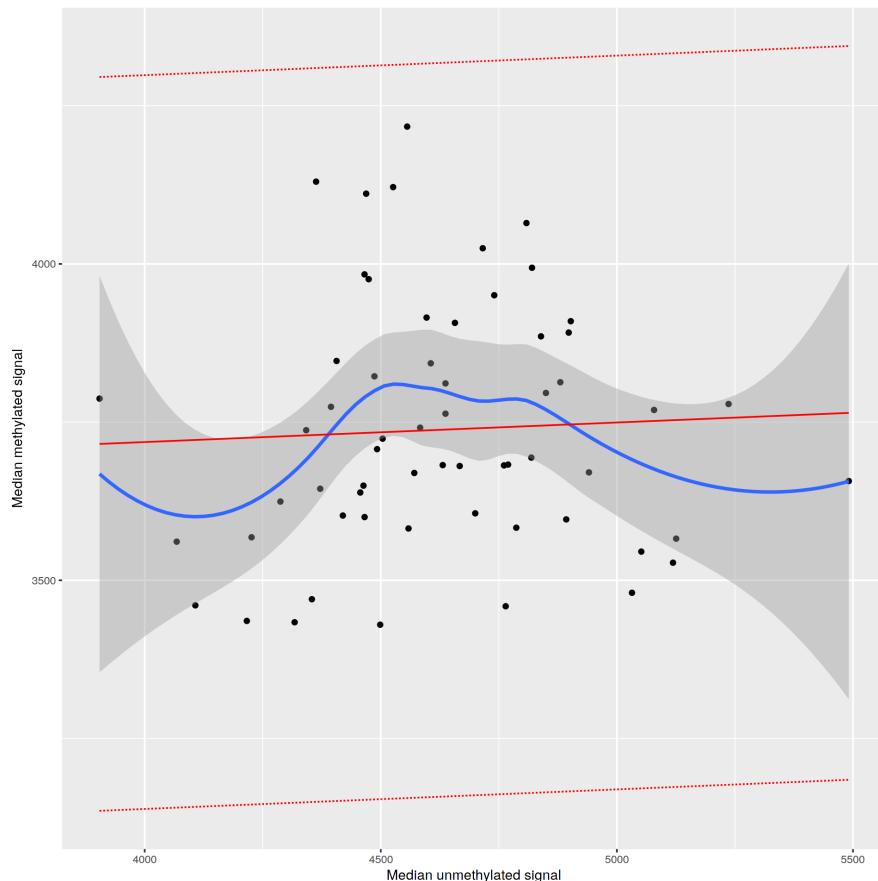


Figure 3.6: Median methylated signal vs unmethylated signal per sample for the 450k array data, solid red line indicates linear regression of median methylated signal vs median unmethylated signal with dotted red lines representing  $3\sigma$  from the expected mean. Samples outside the expected range would be indicated in the legend. Plot generated by meffil QC report.

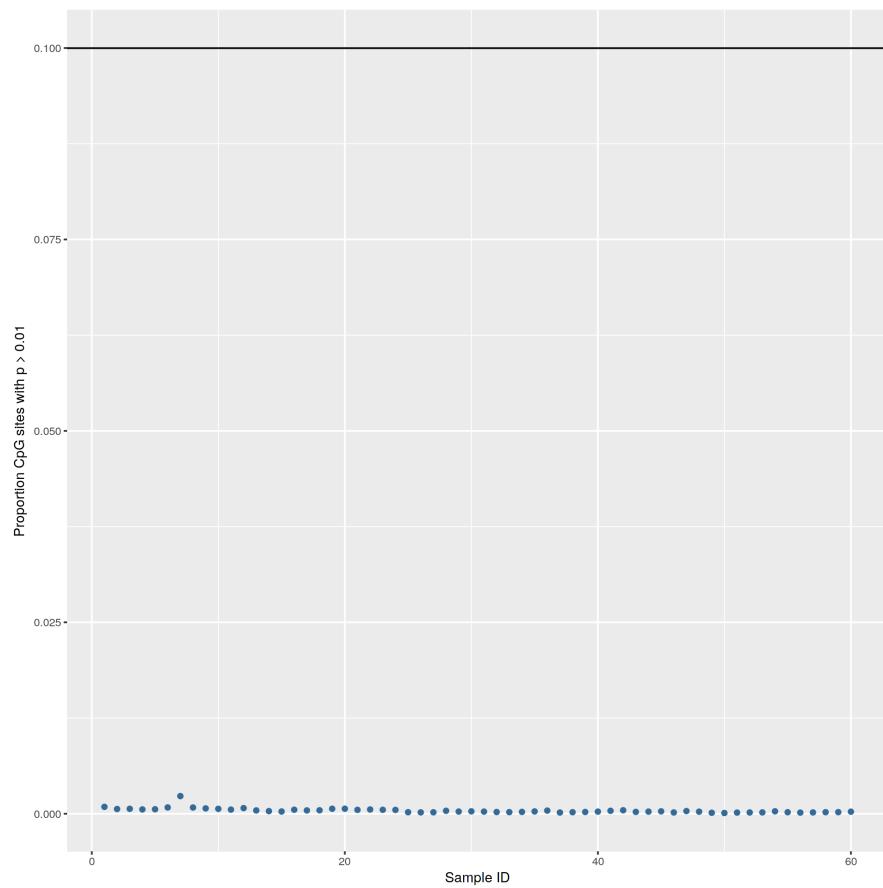


Figure 3.7: Proportion of probes with detection p-values  $>0.01$  by sample for the 450k array data. Black line indicates the exclusion threshold of 0.1. Plot generated by meffil QC report.

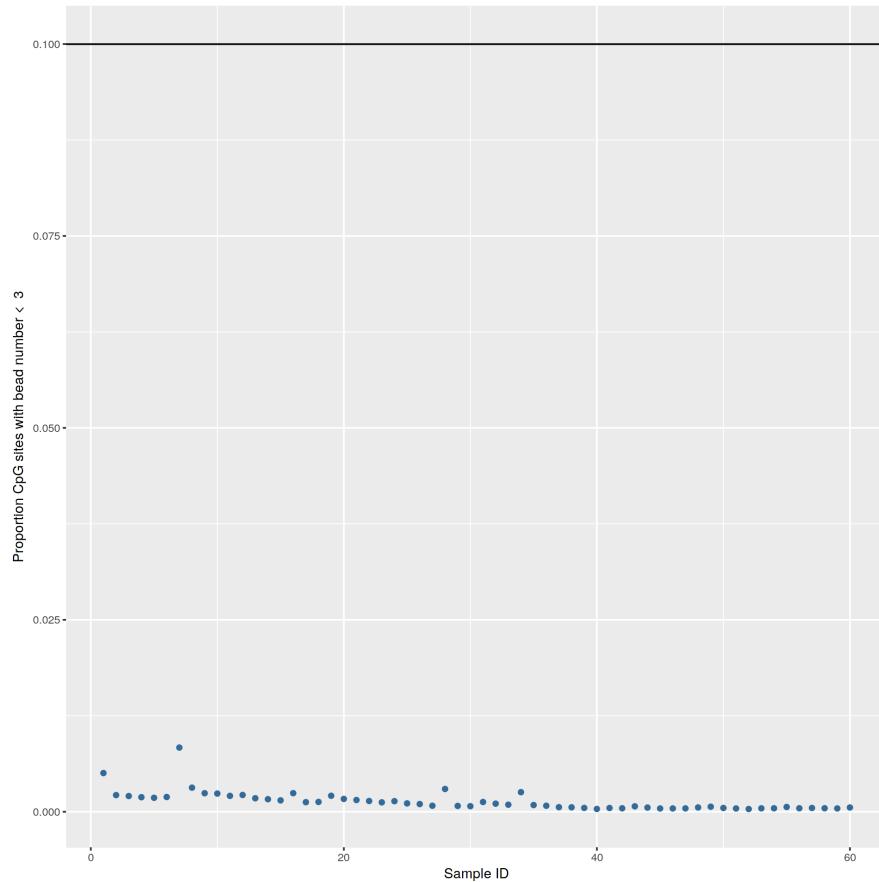


Figure 3.8: Proportion of probes with a bead count of  $< 3$  by sample for the 450k array data. Black line indicates the exclusion threshold of 0.1. Plot generated by meffil QC report.

### 3.3.1.2 Probe QC

#### 3.3.1.2.1 Probe QC - EPIC Arrays

##### 3.3.1.2.1.1 Probe Quality

There were no outliers within the control probes Figure 3.9. 1,626 probes were excluded for having high background signal in a high proportion of samples (proportion of samples with detection p-value  $> 0.01$  is  $> 0.1$ ), (Figure 3.10). 162 probes were excluded for having low bead count in a high proportion of samples (proportion of samples with bead number  $< 3$  is  $> 0.1$ ), (Figure 3.11). Probes with poor technical quality were excluded from the analysis prior to functional normalisation.

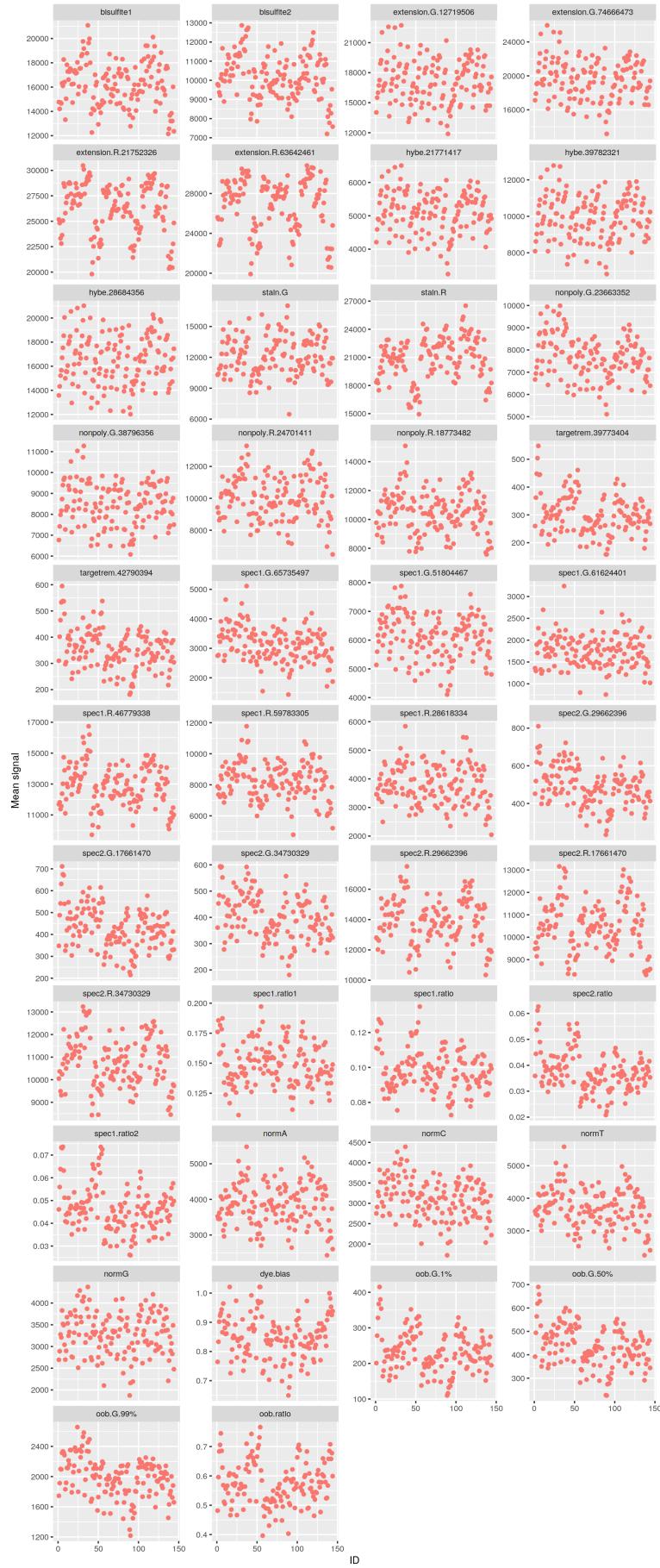


Figure 3.9: Control probe signal by sample for each summary group for the EPIC data. Outliers would be circled in black. Plot generated by meffil QC report.

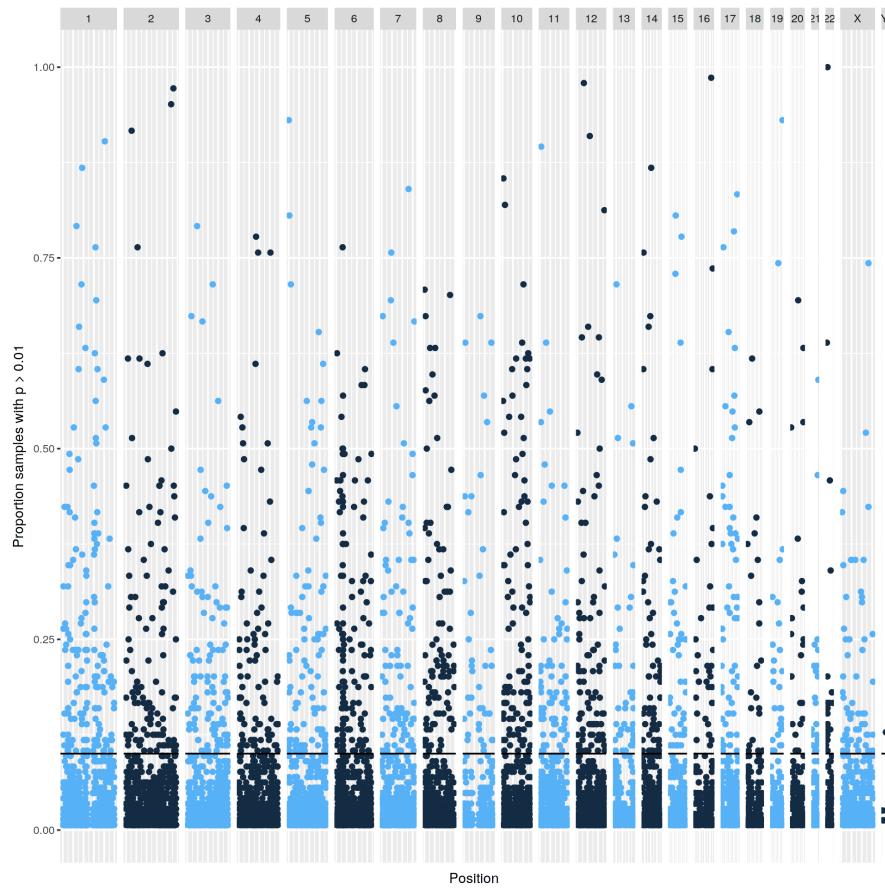


Figure 3.10: Undetectable probes across samples for EPIC data. Manhattan plot showing proportion of samples (y) in which a given probe (x) is not distinguishable from background noise, i.e. a detection p-value of  $> 0.01$ . Black line indicates the exclusion threshold of 0.1. Plot generated by meffil QC report.

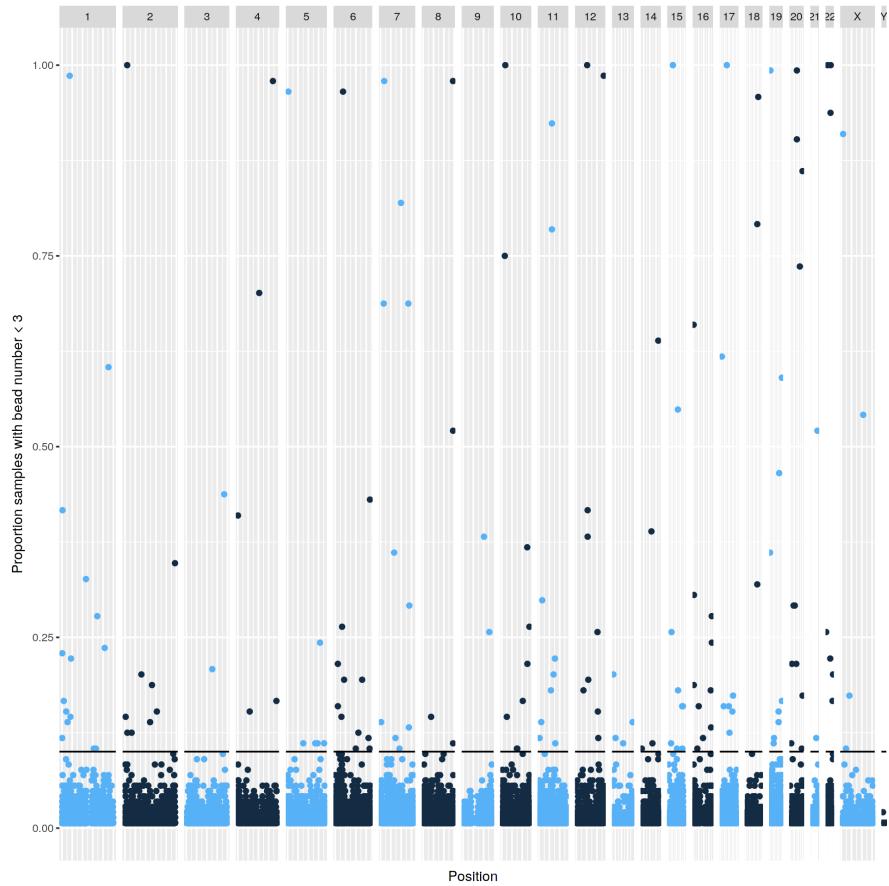


Figure 3.11: Low bead count probes across samples for EPIC data. Manhattan plot showing the proportion of samples (y) in which a given probe (x) has a bead count of  $<3$ . Black line indicates the exclusion threshold of 0.1. Plot generated by meffil QC report.

#### 3.3.1.2.1.2 Multi-mapping and Genetically Confounded Probes

Probes from different locations in the genome with similar sequences, especially following the reduction in sequence complexity associated with bisulfite conversion, can be cross-reactive on the arrays leading to erroneous signals and are thus commonly excluded from analysis. DNAm is often strongly influenced by genetic factors, the effect is especially pronounced when variants alter the sequence at CpG sites themselves as the site can then no longer be methylated. Thus, sites that have common genetic variants at the probe sites are also excluded from the analysis, as well as some sites whose methylation is known to be under strong genetic influence by common genetic variation.

43,254 probes on the EPIC array have been identified as multi-mapping, DNA binding to the probes may be derived from other locations in the genome invalidating these probes as a measure of methylation at their intended loci [Pidsley et al., 2016]. 12,510 probes were identified as having genetic variants at the CpG locus they are intending to assay, this can produce misleading results as mutant bases can resemble the products of bisulfite conversion [Pidsley et al., 2016]. 1,812 probes overlapping regions exhibiting haplotype-specific methylation associated with common non-SNP genetic variants (CNVs, Indels, STRs) and regional in-phase clustering of CpG-SNPs [Bell et al., 2018]. These were excluded from subsequent analysis after functional normalisation. This, including the poor quality probes, is a total of 57,396 unique probes excluded from the analysis (~6.62% of the total number of probes)

In order to identify any potential additional sources of genetic confounding, we looked for probes with methylation values which cluster into distinct groups using the ‘gap hunting’ method developed by Andrews et al. [2016]. Such distinct clusters of methylation can arise from genetic

variants which influence methylation levels being present in homozygous and heterozygous forms in the study population, see Figure 3.12. As the sensitivity and specificity of ‘gap hunting’ is limited, it is the advice of the authors not to exclude probes flagged by ‘gaphunter()’ prior to performing EWAS. It is instead advised to check if any of the results appear in this list after the fact and proceed with caution if they are. Gap hunter identified a further 77,398 probes (8.9% of all probes) which might be subject to genetic confounding.

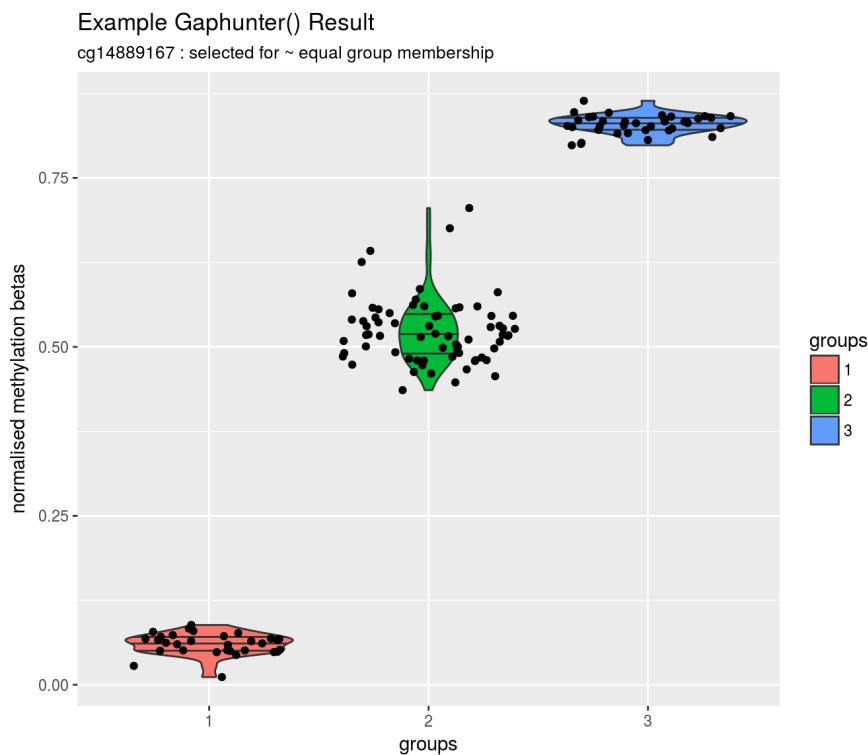


Figure 3.12: An example of the DNAm distribution for a result from the ‘gaphunter()’ function. This is an example chosen to best exemplify the sort of result which is strongly suggestive of a genetic variant with an impact on methylation status acting on this site. It is unrepresentative of typical results from ‘gaphunter()’ in that the groups have a relatively even membership, many results have a small number of individuals in one or more groups making it hard to distinguish methylation outliers caused by rarer genetic variants from those with other causes.

### 3.3.1.2.2 Probe QC - 450k Arrays

#### 3.3.1.2.2.1 Probe Quality

There was one sample (MAVIDOS ID 2183) with an outlier within the control probes, the ‘stain.R’ probe, (Figure 3.13). 509 probes were excluded for having high background signal in a high proportion of samples (proportion of samples with detection p-value > 0.01 is > 0.1), (Figure 3.14). 1037 probes were excluded for having low bead count in a high proportion of samples (proportion of samples with bead number < 3 is > 0.1) (Figure 3.15).

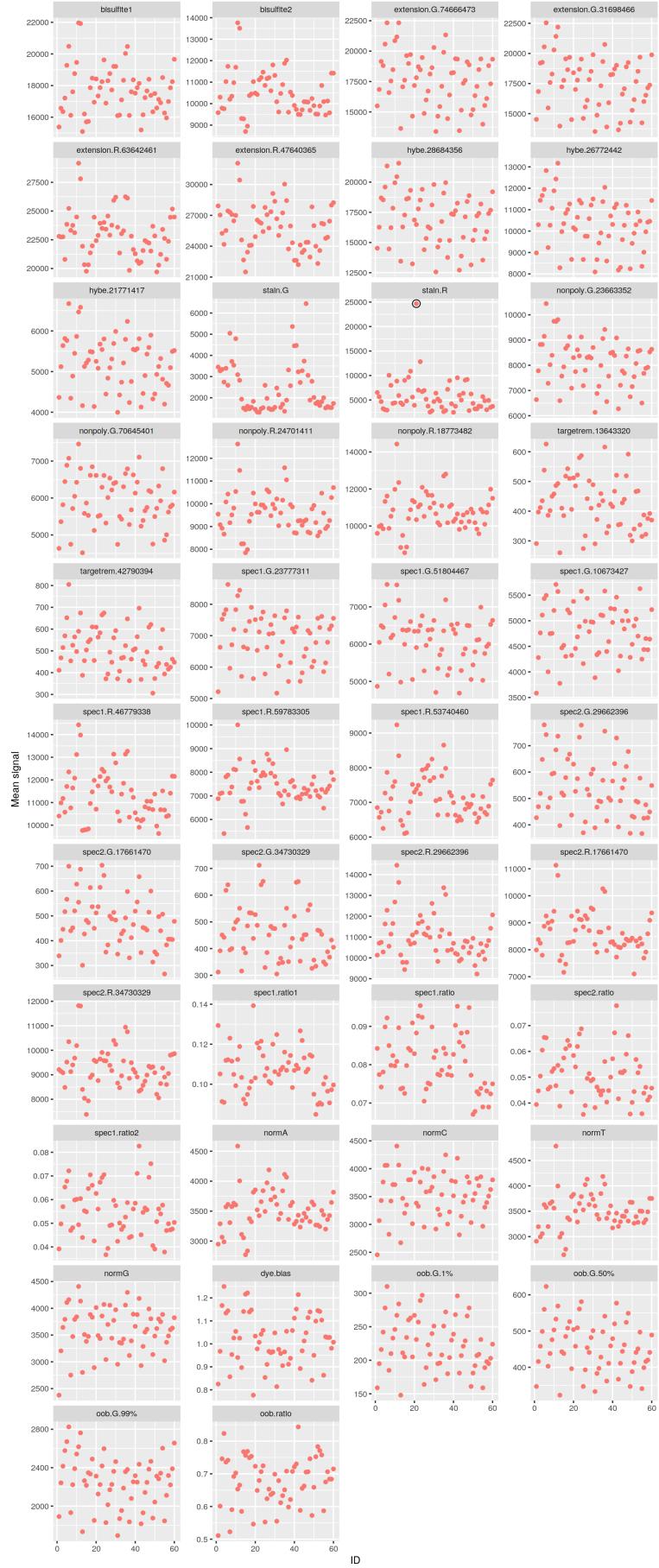


Figure 3.13: Control probe signal by sample for each summary group for the 450k data. Outliers would be circled in black. Plot generated by meffil QC report.

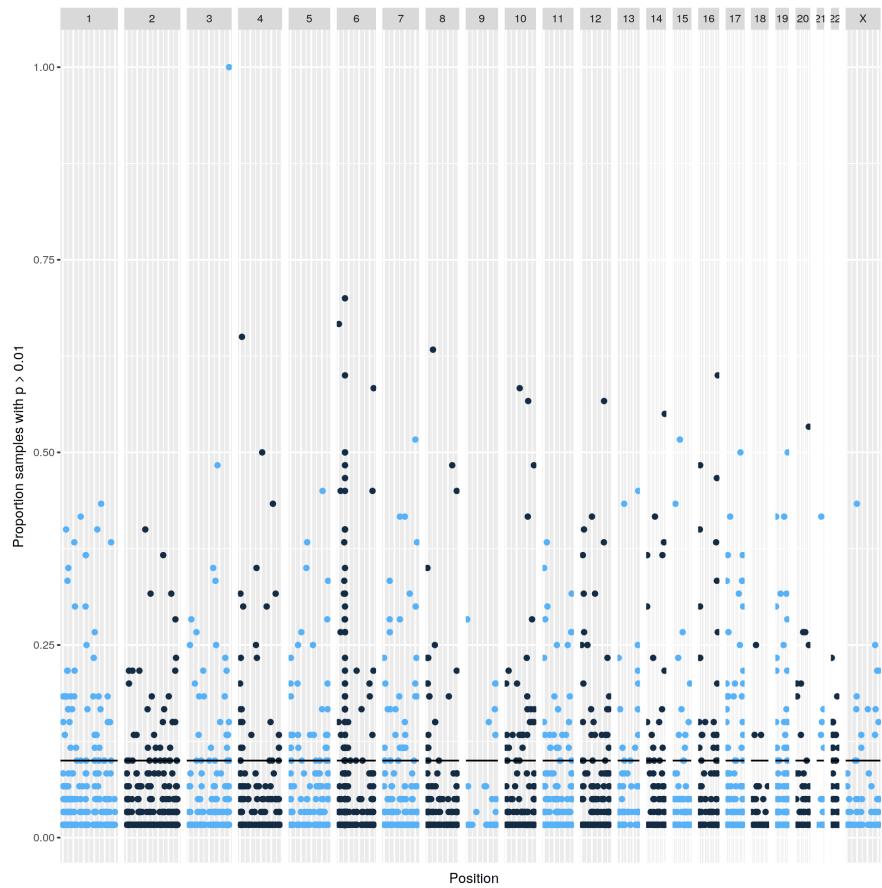


Figure 3.14: Undetectable probes across samples for 450k data. Manhattan plot showing proportion of samples (y) in which a given probe (x) is not distinguishable from background noise, i.e. a detection p-value of  $> 0.01$ . Black line indicates the exclusion threshold of 0.1. Plot generated by meffil QC report.

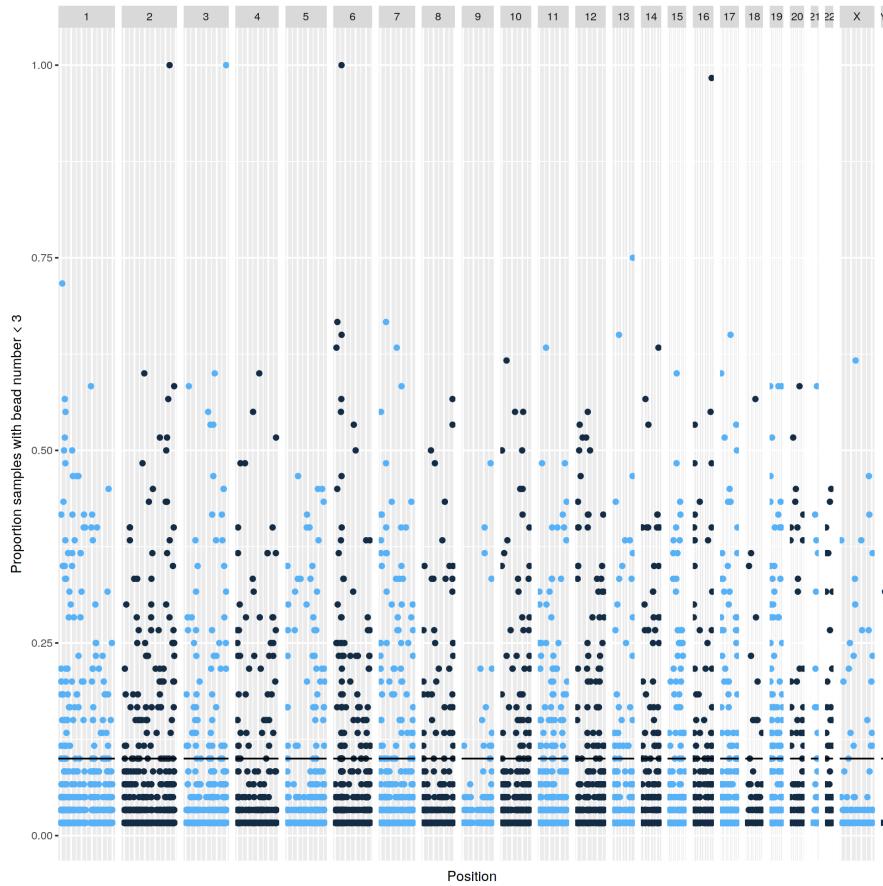


Figure 3.15: Low bead count probes across samples for 450k data. Manhattan plot showing the proportion of samples (y) in which a given probe (x) has a bead count of  $< 3$ . Black line indicates the exclusion threshold of 0.1. Plot generated by meffil QC report.

### 3.3.1.2.2.2 Multi-mapping and Genetically Confounded Probes

Zhou et al. [2016] provided a list of probes which they recommend ‘masking’ from the 450k array due to, multi-mapping issues, genetic variants overlapping the CpG sites and other factors which may render results from these probes problematic. All probes on this list were excluded from the analysis following functional normalisation, leaving a total of 418,632 probes for subsequent analysis.

## 3.3.2 Functional Normalisation

Functional normalisation is an approach to removing unwanted variation associated with ‘batch’ effects such as the date on which a sample was analysed or which slide a sample is on developed by Fortin et al. [2014]. This noise in the data masks the signal associated with the underlying biological effect of interest.

Functional normalisation makes use of control probes on the arrays which are designed to capture only technical variation as surrogates for the sources of unwanted variation. The control probes are processed into 42 summary measures, and principal components analysis is performed on the control probe summary matrices for all samples. The top  $m$  principal components (PCs) are used as the surrogates for technical variation going forward. The number,  $m$ , of PCs from the control probe summary matrices used is informed by the amount of residual variation remaining after normalisation. Picking the number of PCs which correspond to the last steep drop in residual variation is the approach recommended by the implementers of meffil, and Fortin et al.

[2014] recommend an  $m$  of 2 as performing well across a variety of analyses. (See supplementary material from [Fortin et al., 2014] for details of the control probe summary process.)

The process used by functional normalisation is a variant on quantile normalisation. Instead of forcing the empirical marginal distributions of the samples to be identical at each site across arrays it constructs quantile function which only removes variation arising from surrogates for batch variation to produce the normalised values. This approach is effective even when comparing samples with large global differences in methylation levels such as between normal and cancer samples, but cannot overcome high degrees of confounding [Fortin et al., 2014].

An  $m$  of 6 was chosen for the EPIC arrays as this value produced the last steep drop in residual variation, see Figure 3.16. An  $m$  of 6 was chosen for the 450k arrays as this value produced the last steep drop in residual variation, see Figure 3.17.

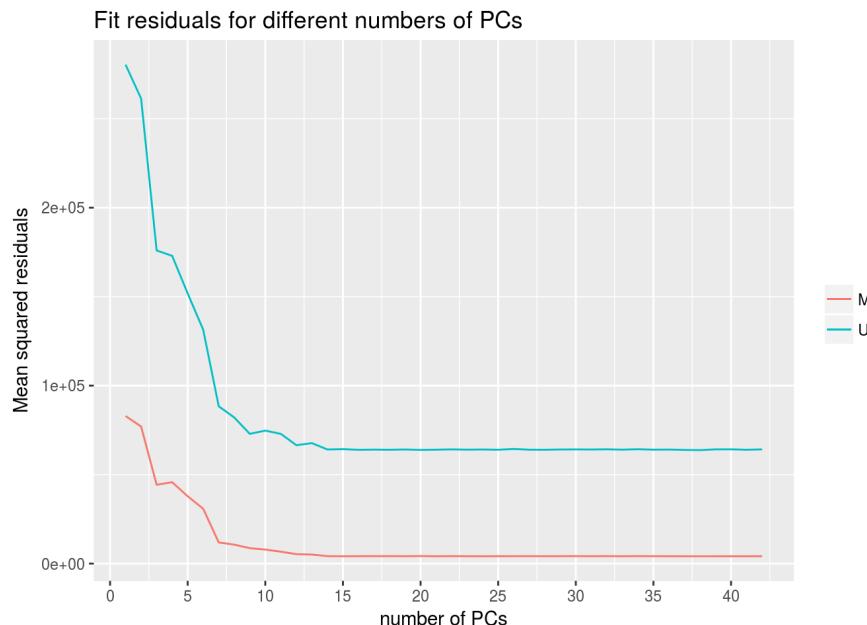


Figure 3.16: Residual variation remaining after functional normalisation of the top 20,000 most variable probes with  $m$  PCs from the control probe summary matrices for the EPIC array samples ( $n=137$ ), for M = methylated and U = unmethylated probes.

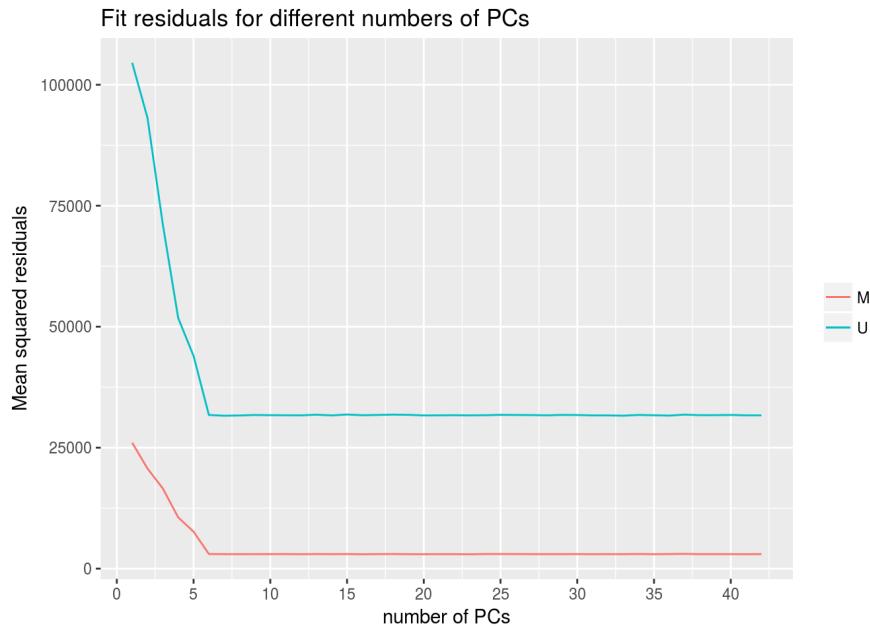


Figure 3.17: Residual variation remaining after functional normalisation of the top 20,000 most variable probes with  $m$  PCs from the control probe summary matrices for the 450k array samples ( $n=60$ ), for M = methylated and U = unmethylated probes.

### 3.3.3 EWAS Models

#### 3.3.3.1 Cell-Type Correction

Blood cell composition is a known confounder in epigenetic studies [Jaffe and Irizarry, 2014, Houseman et al. [2012]]. Observed variation DNA methylation can be cell-type intrinsic, changes in DNAm not driven by changes in the cell-type composition of the tissue, or cell-type extrinsic, due to changes in the cell-type composition of the population sampled (Figure 3.18). An established approach for addressing this potential source of confounding it is to add terms to the regression model which reflect the cell type composition of each sample. Cell type composition can be ascertained through three main approaches: Direct cell count data; predicted cell-count based on comparing experimental methylation data to methylation data from reference samples of the tissue in question [Houseman et al., 2012]; or reference free approaches which make use of mathematical methods to identify sources of confounding variation such as cell-type heterogeneity. Whilst there are several reference samples available for cord blood [Cardenas et al., 2016, de Goede et al. [2015], Bakulski et al. [2016], Gervin et al. [2016]], none were found for cord tissue samples, consequently, we made use of reference free methods for cell type correction. We performed EWAS using Surrogate Variable Analysis (SVA) [Leek and Storey, 2007] and Independent surrogate variable analysis (iSVA) [Teschendorff et al., 2011], the results displayed here are those generated using SVA as this method was recommended based on comparisons of the performance of various cell-type heterogeneity correction methods [McGregor et al., 2016, Teschendorff and Zheng [2017]].

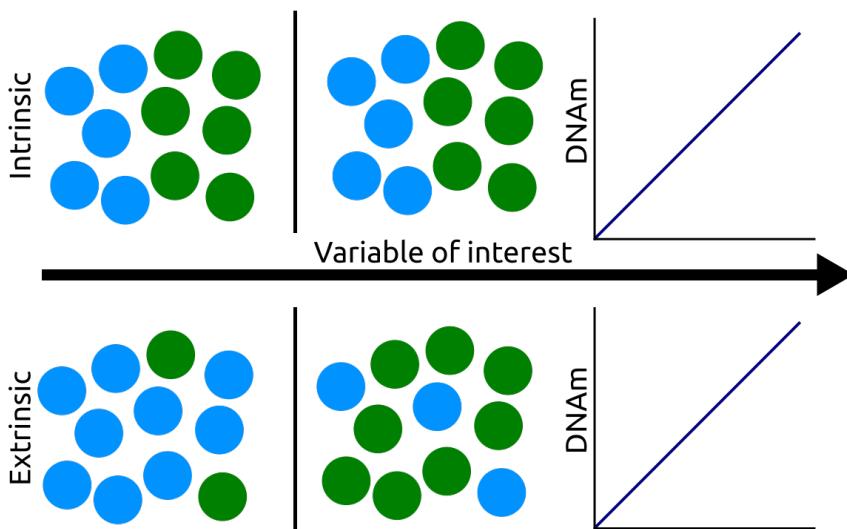


Figure 3.18: Diagrammatic representation of DNAm change arising from extrinsic, cell-type composition change and intrinsic, DNAm of cells changes without the proportion of cell-types in the population changing. These two modes of change are of course not mutually exclusive and both can be occurring.

### 3.3.3.2 Covariates Included in the Models

By default EWAS in meffil are run with four different models:

1. No covariates, attempting to predict methylation with the variable of interest.
2. All covariates, attempting to predict methylation with the variable of interest plus a user-supplied list of covariates.
3. SVs + all covariates, attempting to predict methylation with user-supplied covariates and surrogate variables generated from SVA.
4. iSVs + all covariates, attempting to predict methylation with user-supplied covariates and independent surrogate variables generated from iSVA.

In all the EWAS run here age at which bone measurements were made and sex were supplied as covariates.

Running EWAS with multiple models permits the effects of adding the various covariates on the results of the analysis to be seen. We observed no major differences in the results when using models with and without covariates so have reported here the results of model 3 above. That is to say, the covariates used were the surrogate variables produced from SVA, Age (days) and Sex.

### 3.3.4 Concordance of EPIC and 450k EWAS Results

Concordance between the top ranking probes by p-value with SVA correction was determined using the code which can be found in Section xxx: concordanceCode.

## 3.4 Results

### 3.4.1 Summary of Results

DNAm at none of the probes were significantly associated with the any of the four variables of interest for which EWAS were performed in either the EPIC or 450k datasets. The concordance between the probes with the top-ranked p-values in common between the EPIC and 450k data was at the level of chance. 3 of the EPIC arrays were excluded from the analysis for failing quality control leaving an n=137.

### 3.4.2 EWASs

#### 3.4.2.1 Neonatal Bone Mineral Content

No probes fell below the Bonferroni corrected significance threshold for an association between DNAm at that locus and neonatal bone mineral content, Figure 3.19.

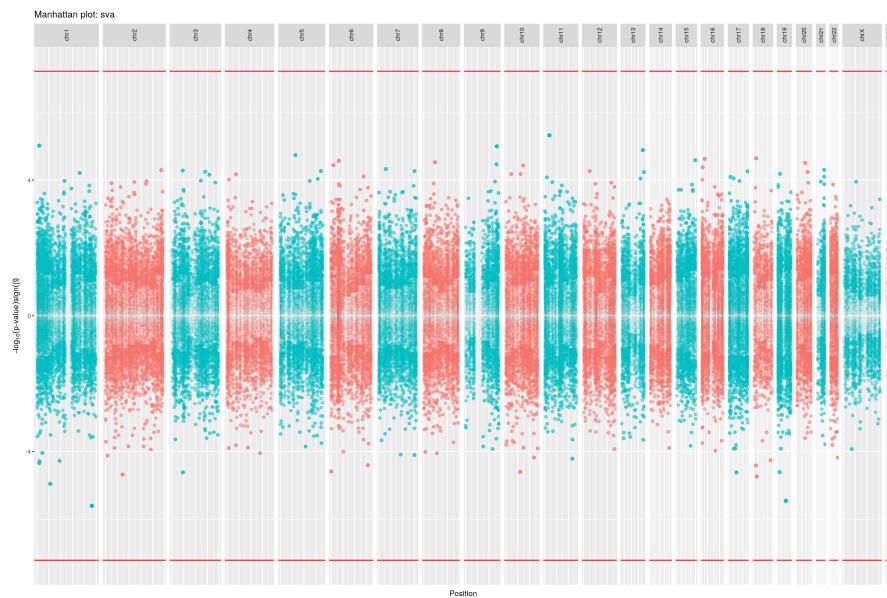


Figure 3.19: Results of EWAS for neonatal bone mineral content with SVA model. Bidirectional Manhattan plot on which  $-\log_{10}(p\text{-value})$  is plotted on the y axis and the sign of this value represents the direction of change. Size and transparency of points increases with  $-\log_{10}(p\text{-value})$  such that the most significant probes are represented the largest and least translucent points. x axis represent chromosomes and position thereupon. Red line indicates the significance threshold of  $6.18 \times 10^{-8}$  ( $0.05 \div 808,585$ )

#### 3.4.2.2 Intervention / Placebo

No probes fell below the Bonferroni corrected significance threshold for an association between DNAm at that locus and Intervention/placebo group status, Figure 3.20.

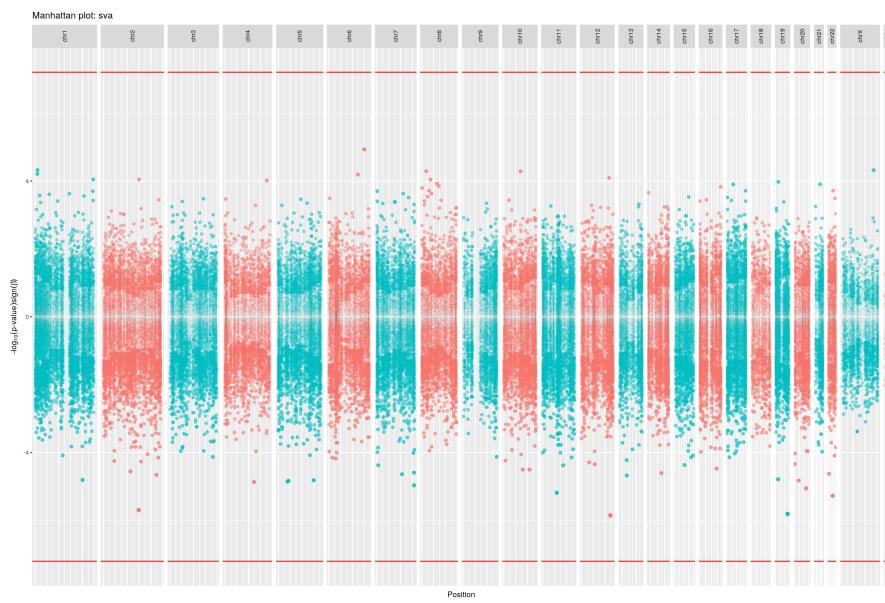


Figure 3.20: Results of EWAS for intervention/placebo group status with SVA model. Bidirectional Manhattan plot on which  $-\log_{10}(p\text{-value})$  is plotted on the y axis and the sign of this value represents the direction of change. Size and transparency of points increases with  $-\log_{10}(p\text{-value})$  such that the most significant probes are represented the largest and least translucent points. x axis represent chromosomes and position thereupon. Red line indicates the significance threshold of  $6.18 \times 10^{-8}$  ( $0.05 \div 808,585$ )

### 3.4.2.3 Maternal Vitamin D (34wks)

Maternal Vitamin D levels remain substantially overlapping between intervention and placebo groups at 34wks, see Figure 3.21. Thus maternal vitamin D at 34wks may prove a more useful variable to model than intervention/placebo status. No probes fell below the Bonferroni corrected significance threshold for an association between DNAm at that locus and maternal vitamin D levels at 34wks gestation, (Figure 3.22).

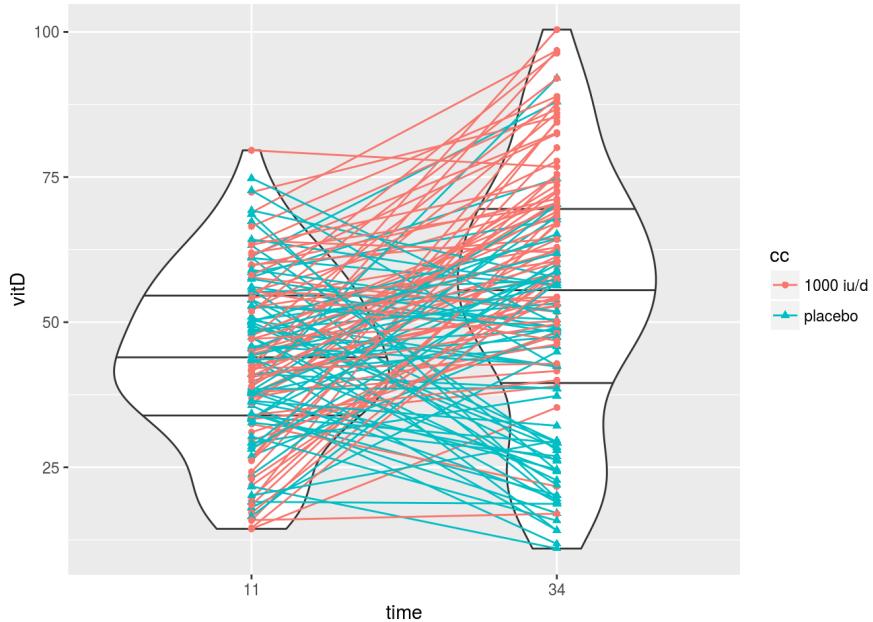


Figure 3.21: Maternal Vitamin D levels at 11 and 34wks gestation, supplemtation with 1000 IU/day cholecalciferol began at week 14. Each participant is shown at both time points linked by a line to indicate the direction of change. The violin plots indicate the density of the distribution of vitamin D values at each time point with the 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> quantiles indicated with horizontal black lines. The colour indicates Intervention (Red) / Placebo (Blue) group

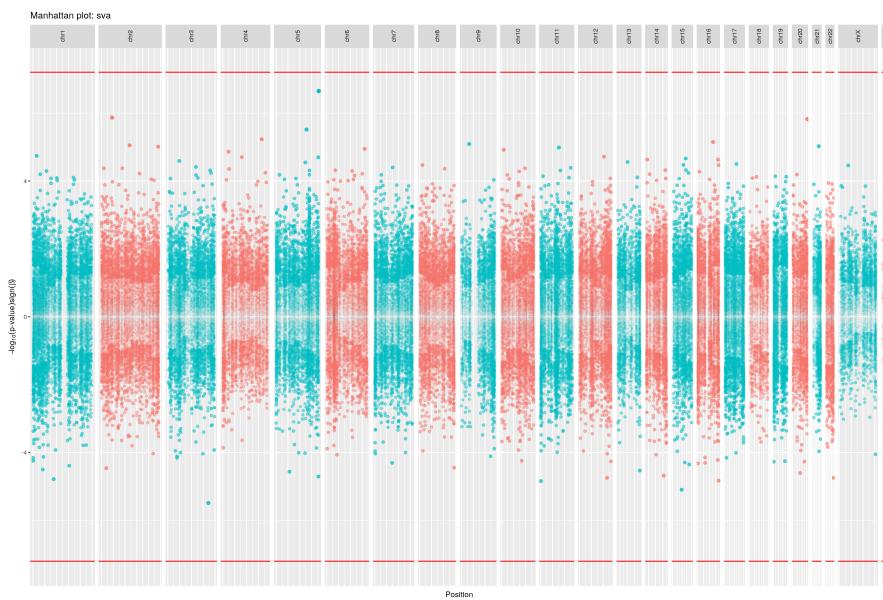


Figure 3.22: Results of EWAS for maternal vitamin D levels at 34wks gestation with SVA model. Bidirectional Manhattan plot on which  $-\log_{10}(p\text{-value})$  is plotted on the y axis and the sign of this value represents the direction of change. Size and transparency of points increases with  $-\log_{10}(p\text{-value})$  such that the most significant probes are represented the largest and least translucent points. x axis represent chromosomes and position thereupon. Red line indicates the significance threshold of  $6.18 \times 10^{-8}$  ( $0.05 \div 808,585$ )

### 3.4.2.4 Change in Maternal Vitamin D

No probes fell below the Bonferroni corrected significance threshold for an association between DNAm at that locus and change in maternal vitamin D from 11 to 34wks gestation, Figure 3.23.

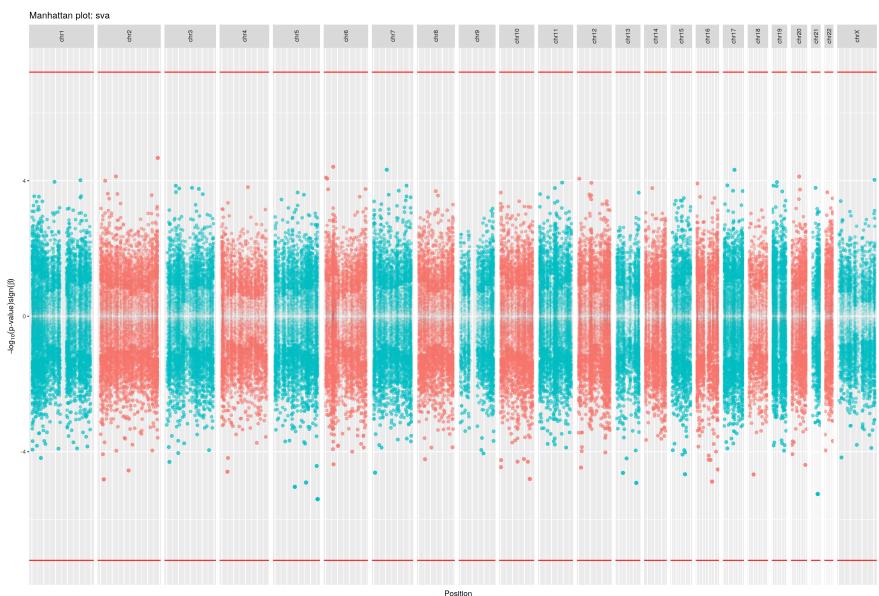


Figure 3.23: Results of EWAS for Change in maternal vitamin D levels from 11 to 34wks gestation with SVA model. Bidirectional Manhattan plot on which  $-\log_{10}(p\text{-value})$  is plotted on the y axis and the sign of this value represents the direction of change. Size and transparency of points increases with  $-\log_{10}(p\text{-value})$  such that the most significant probes are represented the largest and least translucent points. x axis represent chromosomes and position thereupon. Red line indicates the significance threshold of  $6.18 \times 10^{-8}$  ( $0.05 \div 808,585$ )

### 3.4.3 Concordance of EPIC and 850k EWAS results

In order to ascertain if the results from the EWAS in the 450k and EPIC arrays were producing similar sets of probes in the top ranking positions when ordered by p-value the concordance (% overlap) between the top k probes, where  $k = 1 \dots 100,000$  was calculated. Only probes in common between the two arrays were used and k was incremented in steps of 50.

Concordance between the rankings of probes would suggest that the EWAS may be capturing a ‘real’ signal that is simply below the significance threshold with the statistical sensitivity/power that is available in this dataset. The concordance between the EPIC and 450k datasets (Figures 3.24 & 3.25) appears to be at roughly the level expected by chance. This does not lend support to the possibility that there are associations between the variables of interest and DNAm that are beneath the current sensitivity of the study, it does not, however, rule this out. In the absence of any probes above the significance threshold and with poor concordance between the 450k and EPIC array p-value rankings no further analyses such as gene set enrichment and differentially methylated region (DMR) calling have been carried out at this time.

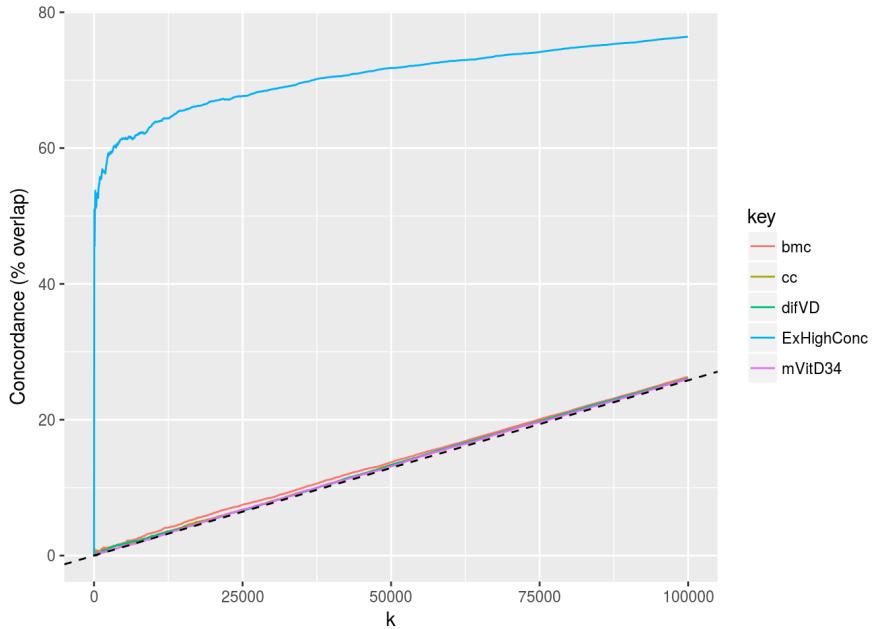


Figure 3.24: Concordance between the top 100,000 probes in common between the EWASs run on the EPIC ( $n=137$ ) and 405k ( $n=60$ ) data sets. bmc = bone mineral content, cc = Intervention / Placebo, difVD = Chance in Vitamin D from 11 to 34wks, ExHighConc = Example of High Concordance generated using SVA vs iSVA results for the 450k intervention/placebo EWAS. Dotted line denotes concordance expected by chance (intersects 50% at 387,511, the number of shared probes)

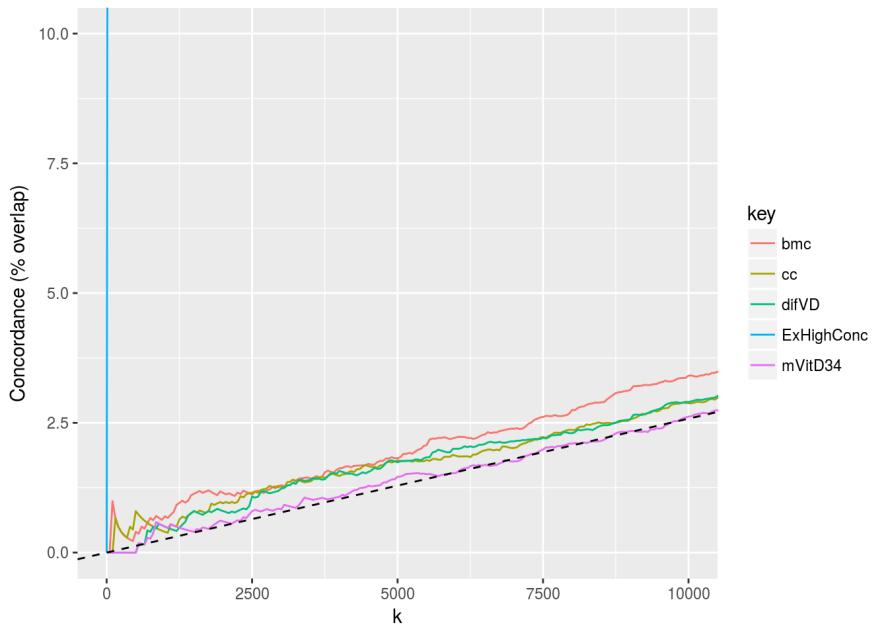


Figure 3.25: Concordance between the top 10,000 probes in common between the EWASs run on the EPIC ( $n=137$ ) and 405k ( $n=60$ ) data sets. bmc = bone mineral content, cc = Intervention / Placebo, difVD = Chance in Vitamin D from 11 to 34wks, ExHighConc = Example of High Concordance generated using SVA vs iSVA results for the 450k intervention/placebo EWAS. Dotted line denotes concordance expected by chance (intersects 50% at 387,511, the number of shared probes)

## 3.5 Discussion

We have not been able to detect any changes in DNA methylation associated with, neonatal bone mineral content, maternal vitamin D at 34wks gestation, change in maternal vitamin D from 11 to 34wks gestation or intervention/placebo group membership. However, at this juncture, we lack the necessary statistical power to rule out the presence of such changes, at effect sizes which may be biologically relevant. EWAS have identified biologically relevant changes associated with environmental exposures in DNAm with magnitudes of less than a single percentage point, and percentage changes in the low single digits are not uncommon in EWAS [Breton et al., 2017]. To achieve 80% power for the 450k array to detect moderate effect sizes (Cohen's  $f^2 = 0.15$ ) requires  $>300$  samples, to do so with small effect sizes (Cohen's  $f^2 = 0.02$ ) requires  $>2500$  samples, the additional multiple testing correction introduced when using the  $\sim 850k$  sites on the EPIC array increases these sample size requirements still further [Cohen, 1988].

Previous studies have found relationships between bone outcomes in childhood and later life rather than in neonates, we may see an association between DNAm and bone outcomes at subsequent time points.

The EPIC array covers many regions not present on the 450k array, in particular, EPIC covers many additional enhancer regions. The lack of concordance between the EPIC and 450k data applies only to those sites that are present on both arrays. Thus, there may be results on the 850k array which are at present false negatives due to the lack of statistical power to detect them but which would have consistently highly ranked p-values were we able to compare them to an additional set of EPIC data.

### 3.5.1 Future Work

We are intending to repeat this analysis with additional samples to increase the power. In addition, we may be getting data on the cord blood vitamin D levels corresponding to our samples, as the EWAS for maternal vitamin D levels at 34wks produced some of the lowest p-values we reason that a more direct measure of vitamin D levels for the sample in question may yield stronger results. We intend to alter the model construction for subsequent analyses to bring the covariates included in line with those used in other analyses performed on MAVIDOS data, for example including gestational age. We also intend to combine the EPIC and 450k array data and perform an analysis with all the available data to maximise the power.

At present we likely have insufficient power to perform EWAS on subsets, e.g. only births in a particular season as seasonal is known to impact Vitamin D levels [Sachs et al., 2013]. It may also be instructive to compare the concordance of results of the four EPIC array for which there are technical replicates to see if there is agreement between these replicates, despite the small size of this group. It would also be interesting to check if any of the CpGs in *CDKN2A* identified as correlating with bone outcomes are present on the arrays and possibly to perform a power analysis to ascertain if we would have predicted that we would be able to replicate these findings with this data.

The effect of maternal smoking on cord blood methylation is a very strong and well established EWAS result [Joubert et al., 2016, Breton et al. [2017]]. We have maternal smoking data so it would be interesting to see if we can recapitulate the smoking EWAS result in cord tissue.

#### 3.5.1.1 DNA Methylation Age analyses

Epigenetic clocks for gestational age have been developed [Bohlin et al., 2016, Knight et al. [2016]] but their estimates did not correlate as well with gestational age in a replication set [Simpkin et al., 2017b]. This was potentially due to a bias towards pre-term infants in the initial training set which was quite small. They were initially trained on cord blood and not cord tissue samples, we intend to assess their effectiveness at predicting gestational age in our cord tissue dataset. DNAm age estimates at birth with Horvath's clocks have minimal correlations

with gestational age at delivery [Simpkin et al., 2016, Simpkin et al. [2017a]], we also intend to attempt to replicate this finding. Once the PhenoAge calculator becomes available we also intend to assess if PhenoAge is similarly uncorrelated with gestational age [Levine et al., 2018]. Another potentially interesting question is if there is any association with any of the epigenetic age estimates and bone outcomes.



# **Part IV**

# **Results 2**



## Chapter 4

# The Genomic Loci of Specific Human tRNA Genes Exhibit Ageing-Related DNA Hypermethylation

*Richard J. Acton<sup>1,2,3</sup>, Wei Yuan<sup>4,5</sup>, Fei Gao<sup>6</sup>, Yudong Xia<sup>6</sup>, Emma Bourne<sup>7</sup>, Eva Wozniak<sup>7</sup>, Jordana Bell<sup>4</sup>, Karen Lillycrop<sup>3</sup>, Jun Wang<sup>6</sup>, Elaine Dennison<sup>2</sup>, Nicholas Harvey<sup>2</sup>, Charles A. Mein<sup>7</sup>, Tim D. Spector<sup>4</sup>, Pirro G. Hysi<sup>4</sup>, Cyrus Cooper<sup>2</sup>, Christopher G. Bell<sup>1</sup>*

1. William Harvey Research Institute, Barts & The London School of Medicine and Dentistry, Charterhouse Square, Queen Mary University of London, London, U.K.
2. MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, U.K.
3. Human Development and Health, Institute of Developmental Sciences, University of Southampton, Southampton, U.K.
4. Department of Twin Research & Genetic Epidemiology, St Thomas Hospital, King's College London, London, U.K.
5. Institute of Cancer Research, Sutton, U.K.
6. BGI-Shenzhen, Shenzhen, China
7. Barts & The London Genome Centre, Blizard Institute, Barts & The London School of Medicine and Dentistry, Queen Mary University of London, London, U.K.

2020-01-19



HOME | ABOUT | SUBMIT | ALERTS / RSS | CHANNELS

Search



Advanced Search

New Results

Comment on this paper

## The Genomic Loci of Specific Human tRNA Genes Exhibit Ageing-Related DNA Hypermethylation

Richard J. Acton, Wei Yuan, Fei Gao, Yudong Xia, Emma Bourne, Eva Wozniak, Jordana Bell, Karen Lillycrop, Jun Wang, Elaine Dennison, Nicholas Harvey, Charles A. Mein, Tim D. Spector, Pirro G. Hysi, Cyrus Cooper, Christopher G. Bell

**doi:** <https://doi.org/10.1101/870352>

This article is a preprint and has not been certified by peer review [what does this mean?].

Figure 4.1: <https://doi.org/10.1101/870352>

*bioRxiv* [Acton et al., 2019]

Under Consideration at Nature Communications

## 4.1 Abstract

Understanding how the epigenome deteriorates with age and subsequently impacts on biological function may bring unique insights to ageing-related disease mechanisms. As a central cellular apparatus, tRNAs are fundamental to the information flow from DNA to proteins. Whilst only being transcribed from ~46kb (<0.002%) of the human genome, their transcripts are the second most abundant in the cell. Furthermore, it is now increasingly recognised that tRNAs and their fragments also have complex regulatory functions. In both their core translational and additional regulatory roles, tRNAs are intimately involved in the control of metabolic processes known to affect ageing. Experimentally DNA methylation can alter tRNA expression, but little is known about the genomic DNA methylation state of tRNAs.

Here, we find that the human genomic tRNA loci (610 tRNA genes termed the tRNAome) are enriched for ageing-related DNA hypermethylation. We initially identified DNA hypermethylation of 44 and 21 specific tRNA genes, at study-wide ( $p < 4.34 \times 10^{-9}$ ) and genome-wide ( $p < 4.34 \times 10^{-9}$ ) significance, respectively, in 4,350 MeDIP-seq peripheral blood DNA methylomes (16 - 82 years). This starkly contrasted with 0 hypomethylated at both these significance levels. Further analysing the 21 genome-wide results, we found 3 of these tRNAs to be independent of major changes in cell-type composition (tRNA-iMet-CAT-1-4, tRNA-Ser-AGA-2-6, tRNA-Ile-AAT-4-1). We also excluded the ageing-related changes being due to the inherent CpG density of the tRNAome by permutation analysis (1,000x, Empirical p-value  $< 1 \times 10^{-3}$ ). We additionally explored 79 tRNA loci in an independent cohort using Fluidigm deep targeted bisulfite-sequencing of pooled DNA (n=190) across a range of 4 timepoints (aged ~4, ~28, ~63, ~78 years). This revealed these ageing changes to be specific to particular isodecoder copies of these tRNA (tRNAs coding for the same amino acid but with sequence body differences) and included replication of 2 of the 3 genome-wide tRNAs. Additionally, this isodecoder-specificity may indicate the potential for regulatory fragment changes with age.

In this study we provide the first comprehensive evaluation at the genomic DNA methylation state of the human tRNAome, revealing a discreet and strongly directional hypermethylation with advancing age.

## 4.2 Introduction

Ageing is implicated as a risk factor in multiple chronic diseases [Partridge et al., 2018]. Understanding how the ageing process leads to deteriorating biological function is now a major research focus, with hopes to increase the human ‘healthspan’ and ameliorate the extensive physical, social and economic costs of these ageing-related disorders [Campisi et al., 2019]. Epigenetic processes, which influence or can inform us about cell-type specific gene expression, are altered with age and are, furthermore, one of the fundamental hallmarks of this progression [López-Otín et al., 2013; Booth and Brunet, 2016].

DNA methylation (DNAm) is the most common epigenetic modification of DNA and age-associated changes in this mark in mammalian tissues have been recognised for decades [Wilson and Jones, 1983]. In fact, these alterations in DNAm with age are extensive with thousands of loci across the genome affected. Many represent ‘drift’ arising from the imperfect maintenance of the epigenetic state [Fraga et al., 2005]. However, specific genomic regions show distinct directional changes, with loss of DNA methylation in repetitive or transposable elements [Chuong et al., 2017], as well as gains in certain promoters, including the targets of polycomb repressor complex [Teschendorff et al., 2010] and bivalent domains [Rakyan et al., 2010]. These observations with the advent of high-throughput DNAm arrays also permitted the identification of individual CpG sites that exhibit consistent changes with age, enabling the construction of predictors of chronological age known as epigenetic or DNAm ‘clocks’ [Hannum et al., 2013; Horvath, 2013; Weidner et al., 2014; Bell et al., 2019]. Additionally, it was observed that ‘acceleration’ of this DNAm-derived measure is a biomarker of ‘biological’ ageing due to associations with morbidity and mortality (Reviewed in [Horvath and Raj, 2018] & [Field et al., 2018]). In a previous investigation of ageing-related DNAm changes within common disease-associated GWAS regions, we identified hypermethylation of the specific transfer RNA gene, tRNA-iMet-CAT-1-4 [Bell et al., 2016]. The initiator methionine tRNA possesses certain unique properties, including its capacity to be rate limiting for translation [Rideout et al., 2012], association with the translation initiation factor eIF2 [Kolitz and Lorsch, 2010], and ability to impact the expression of other tRNA genes [Pavon-Eternod et al., 2013].

tRNAs are evolutionarily ancient [Eigen et al., 1989] and fundamental in the translation process for all domains of life. This translation machinery and the regulation of protein synthesis are controlled by conserved signalling pathways shown to be modifiable in longevity and ageing interventions [Tavernarakis, 2008]. Additionally, beyond their core role in the information flow from DNA to protein sequence, tRNAs can fragment into numerous tRNA-derived small RNAs (tsRNAs) [Pliatsika et al., 2018] with signalling and regulatory functions [Schimmel, 2018; Lee et al., 2009; Li et al., 2018; Xu et al., 2017]. tsRNA abundance has been linked to locus specific tRNA gene expression, in some cases independent of mature tRNA levels [Torres et al., 2019].

The 610 annotated tRNA genes of the human tRNAome (gtRNAdb [Chan and Lowe, 2009]) cover <46 kb (including introns) which represents <0.002% of the human genome [Parisien et al., 2013]. Yet these genes produce the second most abundant RNA species next to ribosomal RNA [Lodish et al., 2000] and are required for the production of all proteins. tRNA genes are transcribed by RNA polymerase III (polIII) [Schramm, 2002] and have internal type II polIII promoters [Canella et al., 2010]. DNAm is able to repress the expression of tRNA genes experimentally [Besser et al., 1990] but may also represent co-ordination with the local repressive chromatin state [Varshney et al., 2015]. Transcription is also repressed by the highly conserved polIII specific transcription factor Maf1 [Murawski et al., 1994; Pluta et al., 2001], whose activity is modulated by the Target of Rapamycin Kinase Complex 1 (TORC1) [Mange et al., 2017]. TORC1 is a highly conserved hub for signals that modulate ageing [Kennedy and Lamming, 2016].

tRNAs as well as tsRNAs are integral to the regulation of protein synthesis and stress response, two processes known to be major modulators of ageing. Metabolic processes are also recognised to modulate the age estimates of DNAm clocks [Nwanaji-Enwerem et al., 2018]. Partial inhibition of translation increases lifespan in multiple model organisms [Hansen et al., 2007] and PolIII inhibition increases longevity acting downstream of TORC1 [Filer et al., 2017]. Furthermore, certain tsRNAs circulating in serum can be modulated by ageing and caloric restriction [Dhahbi et al., 2013].

We directly investigated ageing-related changes in the epigenetic DNA methylation state of the entire tRNAome, facilitated by the availability of a large-scale MeDIP-seq dataset. Arrays poorly cover this portion of genome, with even the latest EPIC (850k) arrays only covering <15% of the tRNA genes, with robust probes, and in total only ~4.7% of all the tRNAome CpGs [Zhou et al., 2016]. tRNA genes sit at the heart not only of the core biological process of translation but at a nexus of signalling networks operating in several different paradigms, from small RNA signalling to large scale chromatin organisation [Van Bortle et al., 2017]. In summary, tRNA biology, protein synthesis, nutrient sensing, stress response and ageing are all intimately interlinked. In this study, we have identified tRNAome DNA hypermethylation and independently replicated this newly described ageing-related observation.

## 4.3 Results

### 4.3.1 DNA Methylation of Specific tRNA Gene Loci Changes with Age

Due to tRNAs critical role in translation and evidence of their modulation in ageing and longevity-related pathways, we interrogated these genes for evidence of ageing-related epigenetic changes. Our discovery set was a large-scale peripheral blood-derived DNA methylome dataset comprising of 4350 samples (see Figure 4.2). This sequencing-based dataset had been generated by Methylated DNA Immunoprecipitation (MeDIP-seq) [Down et al., 2008], which relies on the enrichment of methylated fragments of 200-500 bp to give a regional DNAm assessment (500 bp semi-overlapping windows, see Methods). In total the human tRNAome is comprised of 610 tRNAs (gtRNAdb)(see Figure 4.3), though only 492 are autosomal and do not reside in blacklisted regions of the genome [Amemiya et al., 2019]. Due to the small size of these tRNAs (60-86bp, median 73bp, excluding introns which are present in ~30 tRNAs with sizes from 10-99bp, median 19bp), this fragment-based method enabled a robust examination of the epigenetic state of these highly similar sequences. This was supported by a mappability assessment. The median mappability score density for the tRNAome was 0.90 for 50mers when considering tRNA genes  $\pm 500\text{bp}$  reflecting the regional nature of the MeDIP-seq assay. In contrast the 50mer mappability density is 0.68 for tRNA genes alone representative of the mappability of reads generated using a technique such as whole-genome bisulfite sequencing (see Figures 4.16 & 4.17).

We identified 21 genome-wide significant and 44 study-wide significant results ( $p < 4.34 \times 10^{-9}$  and  $8.36 \times 10^{-5}$ , respectively, via linear regression see Methods (batch corrected  $n=4350$ ). Study-wide significance was calculated conservatively for all 598 autosomal tRNAs. There was a strong directional trend with all results at both significance levels being due to increases in DNA methylation. Age-related changes in cell type proportion are strong in heterogeneous peripheral blood, and include a myeloid skew, loss of naive T cells and increases in senescent cells [Geiger et al., 2013]. A subset of 3 genome-wide and 16 study-wide significant hypermethylation results remained significant even after correcting for potential cell-type changes by including lymphocytes, monocytes, neutrophils and eosinophil cell count data ( $n=3001$ , Listed in Table 4.1, Red in Figure 4.6). tRNA-iMet-CAT-1-4 is located on chromosome 6. tRNA-Ile-AAT-4-1 and tRNA-Ser-AGA-2-6 are neighbours and are located on chromosome 17 within the 3' UTR of *CTC1* (CST Telomere Replication Complex Component 1). Going forward we refer to these most robustly corrected sets of 3 and 16 tRNA genes as the genome-wide and study-wide significant tRNA genes respectively.

Due to the related nature of these twin samples, we also analysed these data in two subsets of  $n=1198$  & 1206 by selecting one twin from each pair into the separate sets. This analysis also included correction for Batch and Blood Cell counts. Whilst in these smaller datasets no tRNAs were genome-wide significant, 5 and 7 tRNA genes, respectively, reached study-wide significance. In these sets 5/5 and 6/7 of these were present in 16 study-wide significant tRNA genes.

Furthermore, we examined a subset of samples with longitudinal data ( $n=658$  methylomes from 329 individuals, median age difference 7.6 yrs). At the nominal significance threshold ( $p < 0.05$ ) this yielded a split of 41 hypermethylating tRNA genes and 22 hypomethylating tRNA genes. Of these hypermethylated tRNAs, 2 are in the previously identified genome-wide significant set of 3 (with tRNA-iMet-CAT-1-4 ranked 3rd by p-value) and 9 are in the study-wide significant set of 16.

DNA Methylation	Blood			Other Tissues
	Discovery	Validation	Replication	Tissue Specificity
	Method: MeDIP-Seq tRNAs: 598 N = 4,350 Ages: 19 - 82 yrs Source: Twins UK	Method: 450k array tRNAs: 158 N = 587 Ages: 18 - 81 yrs Source: Twins UK	Method: Targeted Bisulfite Sequencing tRNAs: 79 N = 190 in 8 pools Ages: 4 - 80 yrs Source: MAVIDOS / Hertfordshire	Method: 27k/450k array tRNAs: 43-115 N = 733 Ages: 0 - 90 yrs Source: TCGA/GDC/GEO
				19 Tissues matched Normal and Tumour, 11 Fetal

Figure 4.2: Study Structure. tRNAs differentially methylated with age initially identified in MeDIP-seq, validated (where covered in 450k array) and replicated in targeted bisulfite sequencing of pooled samples. Tissue specificity of these effects was explored in TCGA and foetal tissue data.

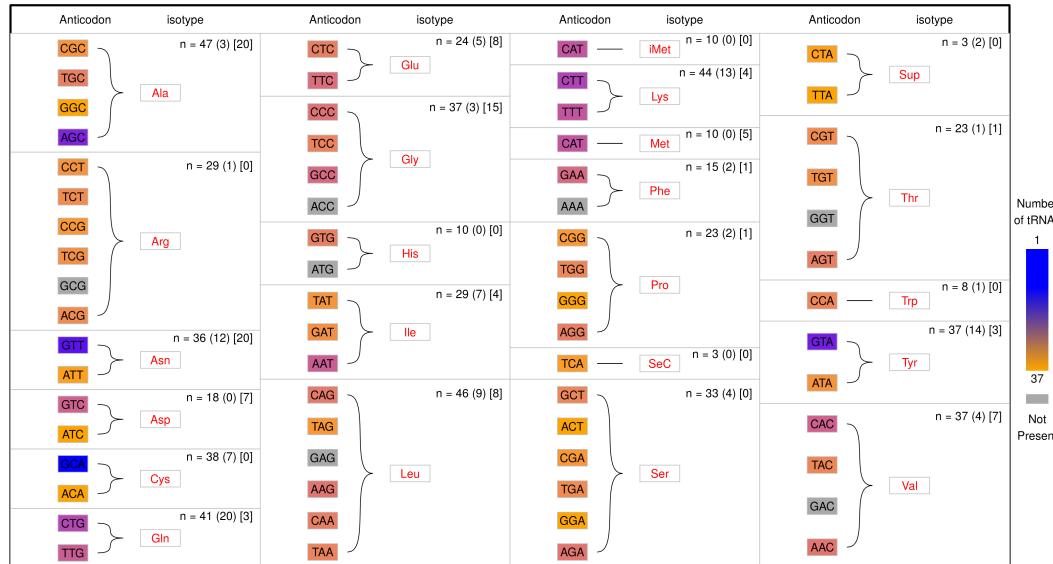


Figure 4.3: The genetic code as represented in the human tRNAome. The triplet genetic code leads to the incorporation of specific amino acids into an elongating protein via corresponding tRNAs. n is the number of tRNA genes which encode a given amino acid, the number in parentheses is how many of those may be pseudogenes based on their tRNAscan score [Lowe and Chan, 2016], and the number in square braces is the number in blacklisted regions [Amemiya et al., 2019]. There are a total of 610 tRNAs in GtRNAdb [Chan and Lowe, 2009], 116 of which are potential pseudogenes, and 107 are in blacklisted regions [Amemiya et al., 2019]. Notably 7 of the 61 non-STOP codons are missing from the human tRNAome therefore these codons are handled by wobble base matching (*e.g.* GCG Arg, ACC Gly). Also of note are the suppressor and selenocysteine tRNAs. The 20 methionine tRNAs are split equally between initiator methionine and internally incorporating methionine tRNAs, which are structurally distinct. There are also 23 nuclear encoded mitochondrial tRNAs.

### 4.3.2 tRNA Genes are Enriched for Age Related DNA Hypermethylation

Whilst ageing changes are pervasive throughout the DNA methylome, we identified a strong enrichment for this to occur within the discrete tRNAome (Fisher's Exact Test  $p = 1.05 \times 10^{-27}$ ) (see Figure 4.4). This is still significant if the 6 of the study-wide significant 16 tRNAs that overlap polycomb or bivalent regions are excluded ( $p = 4.66 \times 10^{-15}$ )

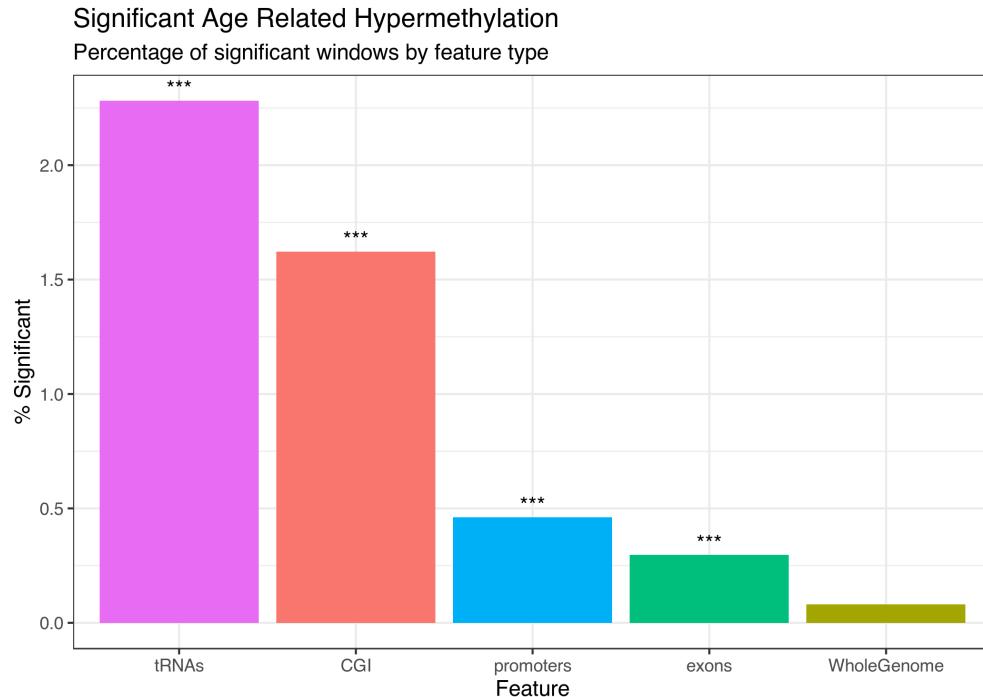


Figure 4.4: Age-related DNA hypermethylation Genomic Region Enrichment. ( $n = 3001$ , tRNAs are enriched compared to the genomic background, Fisher's Exact Test  $p < 1.05 \times 10^{-27}$ , Blood cell-type and batch corrected).

CpG density itself is known to have a clear impact on the potential for variability of the DNA methylome as well as ageing-related changes [Ziller et al., 2013, Christensen et al., 2009]. To assess whether this hypermethylation finding was being merely driven by the inherent CpG density of the tRNAome, we performed a CpG density matched permutation analysis (1,000X, see Methods). This supported the specific nature of these age-related DNAm changes within the functional tRNAome (Empirical p-value  $< 1.0 \times 10^{-3}$ , Figure 4.5). As a point of comparison for this genomic functional unit, we also performed the same permutations for the known age-related changes in the promoters of genes that are polycomb group targets [Teschendorff et al., 2010] and those with a bivalent chromatin state [Rakyan et al., 2010]. We were able to reproduce the enrichment of the polycomb group targets and bivalent regions (Empirical p-value  $< 1.0 \times 10^{-3}$ ) in our dataset.

Table 4.1: Significantly Hypermethylating tRNAs in blood cell-type and batch corrected model MeDIP-seq. ‘Slope’ corresponds to the beta value for methylation in the linear model.

tRNA	Location	MeDIP-seq p-value	BiS-seq p-value	BiS-seq slope
<b>tRNA-iMet-CAT-1-4</b>	chr6	2.83e-11	9.35e-04	4.54
<b>tRNA-Ile-AAT-4-1</b>	chr17	3.03e-10	6.88e-04	-0.745
<b>tRNA-Ser-AGA-2-6</b>	chr17	1.16e-10	4.28e-2	0.623

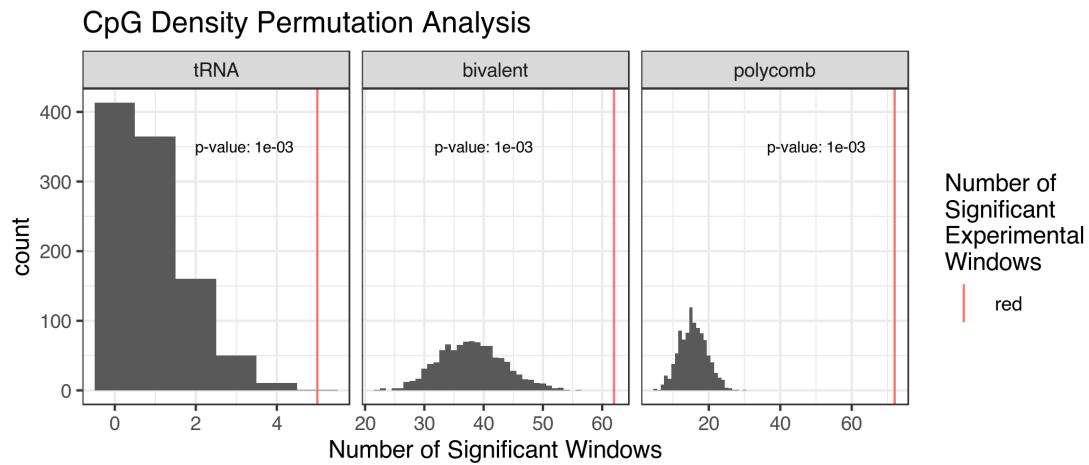


Figure 4.5: CpG Density Genome-wide Permutation Analysis. Each permutation represented a random set of windows matching the CpG density of the functional unit (bivalent domains, polycomb group target promoters & the tRNAome). These are subsequently assessed for significant age-related DNAm changes (see Methods). The red line is the observed number of significant loci.

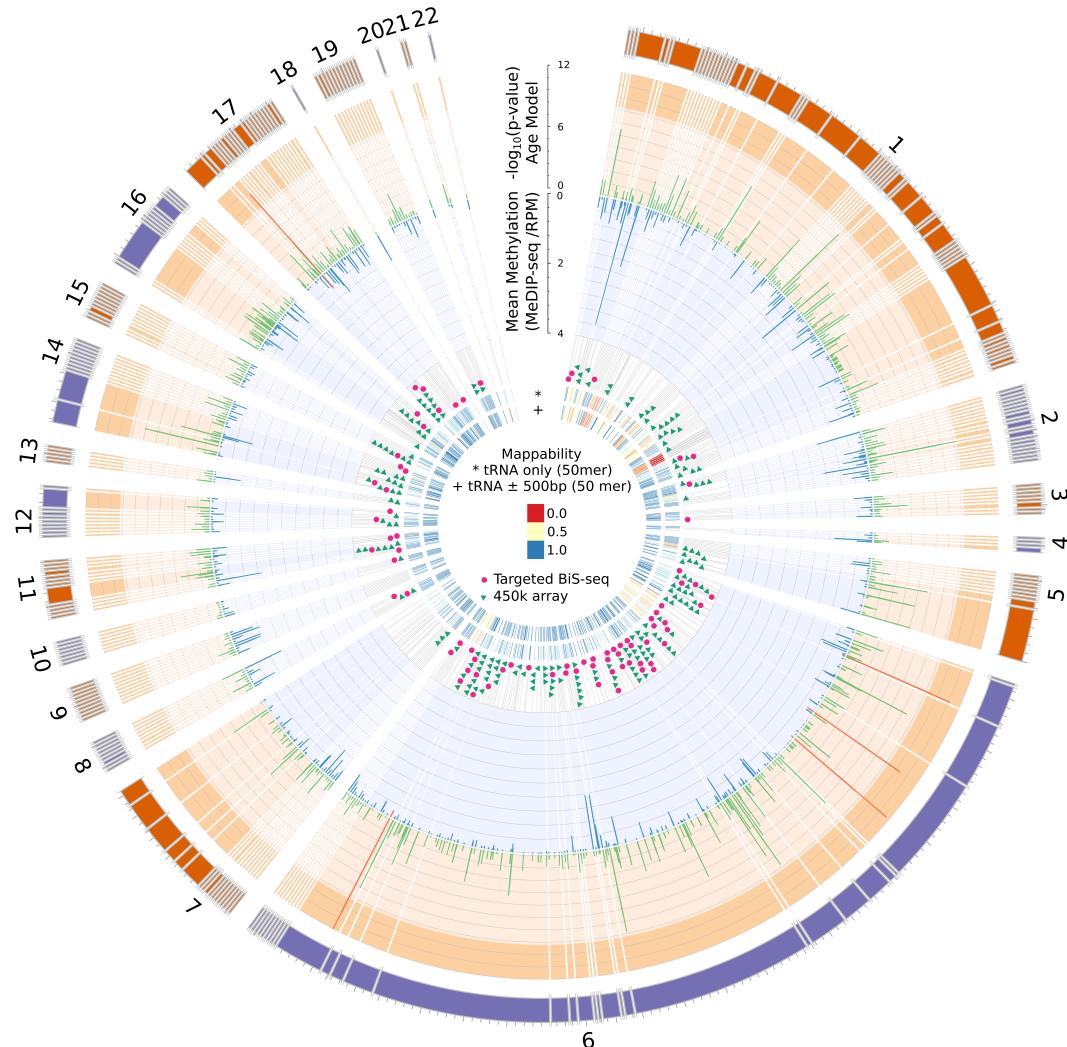


Figure 4.6: Human tRNAome overview. From the outside in: Chromosome ideograms scaled by the number of tRNA genes (total = 598), as excludes chromosome X (10), Y (0) and contig chr1\_gl000192\_random (2; see Methods). tRNA genes within 20kbp of one another are grouped with breaks inserted between these clusters. Radial grey lines represent the location of tRNA genes in the genome.  $-\log_{10}(p\text{-value})$  for the blood cell-type and batch corrected age model are shown for each window overlapping a tRNA gene in green. Mean methylation across all samples (n=3001) in RPM (reads per million base pairs) is shown in blue. Genome-wide significant cell-type & batch corrected ( $p < 4.34 \times 10^{-9}$ ) tRNAs show in red. The 158 Loci covered by 213 probes on the 450k array which directly overlap a tRNA gene are shown with green triangles. The 84 loci targeted for bisulfite sequencing in this study are indicated in magenta. Mappability score density is computed as the area under the encode mappability tracks [Derrien et al., 2012] over the length of the region.

#### 4.3.2.1 Age-related tRNAome DNA Hypermethylation is even observed in one Newborn versus one Centenarian

We examined an available Whole Genome Bisulphite sequencing (WGBS) dataset from Heyn *et al.* [Heyn et al., 2012] (see Methods). These data consisted of blood-derived DNA WGBS in one newborn child and one 26 year old, and centenarian (103 years). In their analysis, the centenarian was found to have more hypomethylated CpGs than the neonate across all genomic compartments, including promoters, exonic, intronic, and intergenic regions. However, even in this examination of 3 individuals of 3 different age in the 55% of tRNA that possessed coverage, we observed DNA hypermethylation with age among the study-wide significant tRNA hypermethylators. The centenarian was significantly more methylated in this set of tRNAs than the neonate (Wilcoxon rank sum test, 6.14% increase (95% CI -Inf - 4.31),  $W = 717$ ,  $p = 6.14 \times 10^{-4}$ , see Figure 4.7).

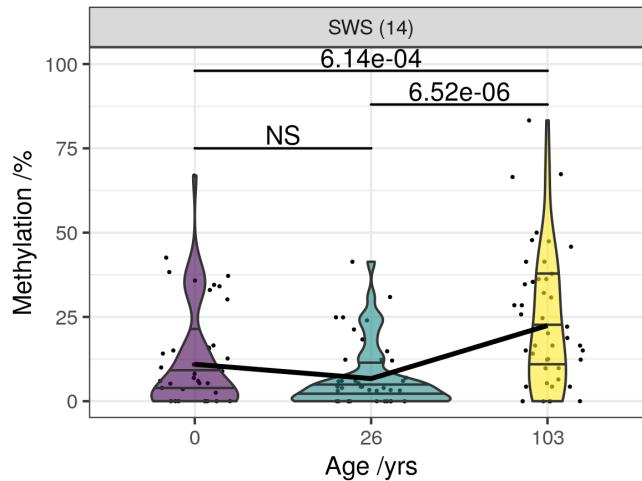


Figure 4.7: Whole Genome Bisulfite Sequencing Data in a newborn, as adult and a centenarian. Each point represents the methylation level at an individual CpG within a tRNA gene. Numbers in parentheses indicate the number of tRNA genes for which methylation data is shown. SWS: Available study-wide significant tRNA genes.

#### 4.3.2.2 Age-related Changes Independently Replicated with Targeted Bisulfite Sequencing

In order to further robustly support these-ageing related changes, we attempted to replicate these findings ourselves in an independent ageing dataset. Furthermore, we employed a different technology Targeted bisulfite (BiS) sequencing to also further validate the MeDIP-seq-derived results. These data provide individual CpG resolution to identify what may be driving the regional DNAm changes observed.

We performed this targeted BiS-seq in blood-derived DNA from 8 pools of age-matched individuals at 4 time-points (~4, ~28, ~63, ~78 years) from a total of 190 individuals, as detailed in Table 4.2. A total of 79 tRNA loci generated reliable results post-QC (see Methods). These tRNAs covered a total of 458 CpGs with a median of 6 CpGs per tRNA (range 1-9). Median Coverage per site across pools, technical replicates and batches was 679 reads (mean 5902).

Firstly, we ran the 8 Pooled samples on the Illumina EPIC (850k) array to confirm that our pooling approach was applicable for DNAm ageing-related evaluation. This showed an  $R^2 = 0.98$  between pool mean chronological age and Horvath clock DNAm predicted age [Horvath, 2013] (see Figure 4.8). Therefore, this confirmed the utility of this novel pooling approach. We also used these array derived data to estimate the major blood cell proportions for each of these pools with the Houseman algorithm [Houseman et al., 2012].

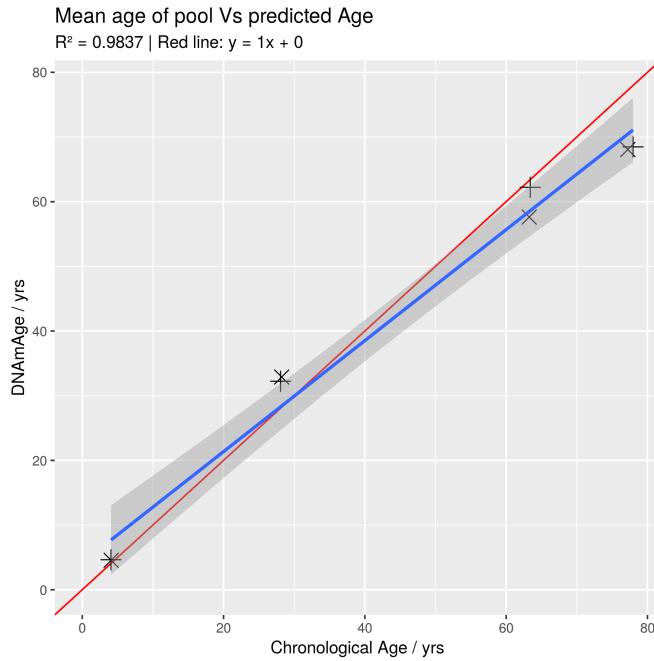


Figure 4.8: Chronological Age compared with DNAm Age estimated with Horvath's 2013 multi-tissue clock [Horvath, 2013] based on EPIC array data for the 8 pooled samples used in the Targeted Bisulfite sequencing (See Table 4.2 for pool details).

We noted that individual tRNA loci exhibiting age-related changes in DNAm had duplicate or isodecoder (same anticodon but body sequence variation) sequences in the genome, which despite exact or near sequence identity did not show similar changes. tRNA-iMet-CAT-1-4 for instance is 1 of 8 identical copies in the genome and was the only locus that showed significant changes. The results of pairwise differential methylation tests between age groups for the 6 top tRNAs from the MeDIP-seq models are listed in Table 4.3.

Of the 3 top hits in MeDIP-seq, tRNA-iMet-CAT-1-4 (Figure 4.10c) and tRNA-Ser-AGA-2-6 (Figure 4.10i) exceeded nominal significance ( $p$ -values =  $9.35 \times 10^{-4}$  &  $4.28 \times 10^{-2}$ , respectively). However, tRNA-Ile-AAT-4-1 (Figure 4.10n) showed a nominal decrease in DNAm with age. tRNA tRNA-Leu-TAG-2-1 from the study-wide significant set also showed nominally significant hypermethylation with age (Figure 4.10u). Also, four of the individual CpGs in tRNA-iMet-CAT-1-4 exhibited nominally significant increases in DNAm with Age (Figure 4.9).

Table 4.2: Summary information on participants in each pool.

Pool	Mean Age	Sex	Min Age	Max Age	n
Pool 1	4.07	Male	3.99	4.38	20
Pool 2	4.09	Female	3.99	4.36	20
Pool 3	28.07	Female	25.87	29.80	25
Pool 4	28.23	Female	26.05	30.01	25
Pool 5	63.40	Female	62.80	63.80	25
Pool 6	63.26	Female	62.70	63.70	25
Pool 7	77.96	Female	75.50	80.50	25
Pool 8	77.22	Female	74.40	80.10	25

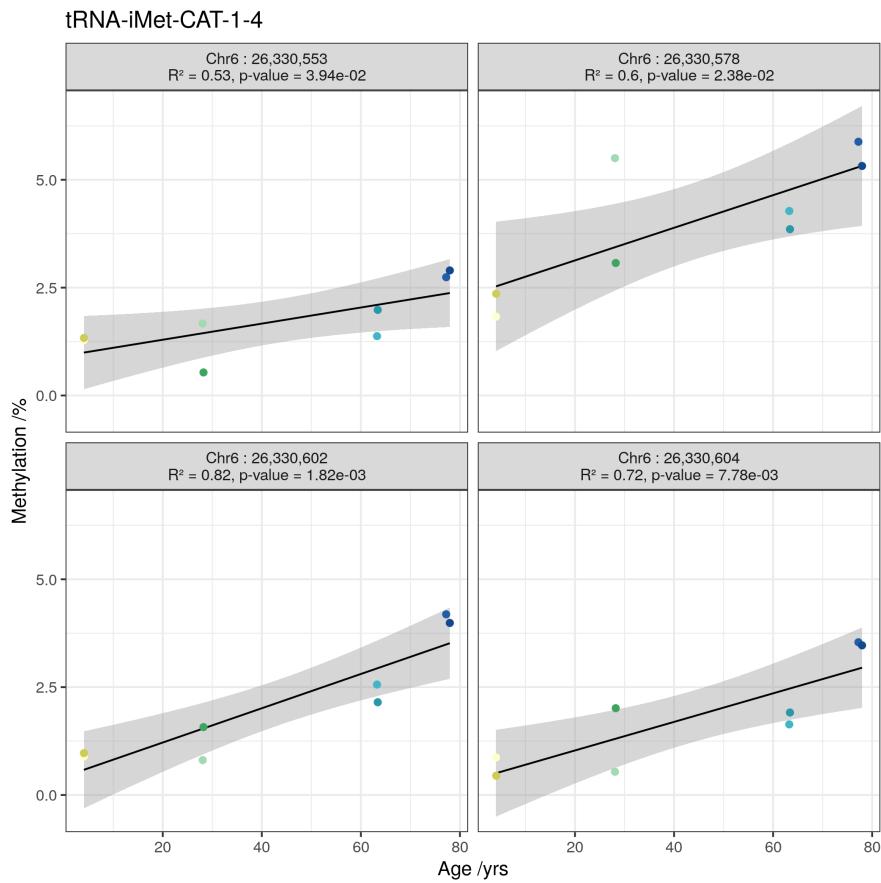
Figure 4.9: Individual CpG methylation increases (nominally significant  $p < 0.05$ ) in tRNA-iMet-CAT-1-4.

Table 4.3: Pairwise Differences in Methylation between Age groups by tRNA. p-values are for pairwise methylation differences (see Methods)[Müller et al., 2019].

tRNA	num.	CpGs	comparison	p-value	delta
tRNA-Ile-AAT-4-1	8	4	vs. 28	1.518e-01	-0.2
		4	vs. 63	1.774e-01	-0.234
		4	vs. 78	3.060e-01	0.0113
		28	vs. 63	7.152e-01	-0.0334
		28	vs. 78	1.553e-01	0.212
		63	vs. 78	2.057e-01	0.245
tRNA-iMet-CAT-1-4	5	4	vs. 28	8.403e-02	0.0116
		4	vs. 63	1.716e-01	0.0125
		4	vs. 78	1.997e-04*	0.0368
		28	vs. 63	3.943e-01	0.000869
		28	vs. 78	1.724e-02*	0.0252
tRNA-Ser-AGA-2-6	9	4	vs. 78	6.224e-02	0.0243
		4	vs. 28	4.222e-01	0.0573
		4	vs. 63	3.968e-01	0.0274
		28	vs. 63	1.095e-01	-0.0299
		28	vs. 78	2.126e-01	-0.015
		63	vs. 78	2.201e-01	0.0149

#### 4.3.2.2.1 Select Duplicates & Isodecoders of Hypermethylating tRNA loci remain unchanged

We targeted a selection of these duplicate and isodecoder loci for bisulfite sequencing in order to confirm that the identified DNAm changes are specific to a given locus and not general to related tRNAs. Examining the tRNA-iMet-CAT-1 family, only the previously identified 1-4 version confirmed significant hypermethylation (not 1-2, 1-3 or 1-5)(Figure 4.10a-e). Likewise the tRNA-Ser-AGA-2-6 version was supported compared to 2-1,2-4 and 2-5(Figure 4.10f-j)).

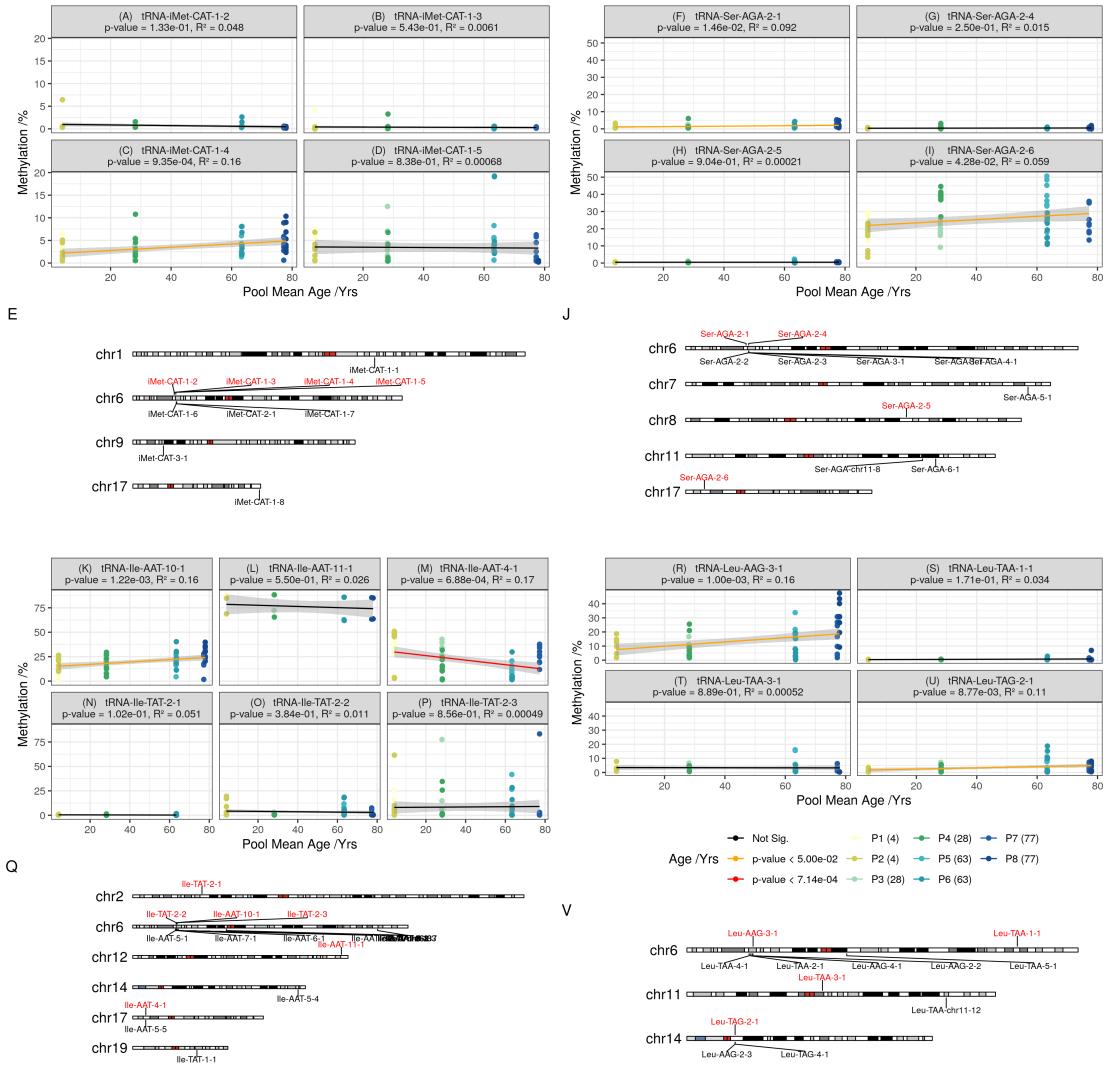


Figure 4.10: (A-Q) Mean methylation across replicates at each CpG in each pool for select tRNAs. Ideograms show the locations of all tRNAs with the same anticodon, with those featured in the scatter plots highlighted in red and placed above the ideogram. (R-V) Except for leucine isoacceptors tRNAs are not all of the same codon and only tRNAs on the same chromosome as those plotted are labelled on the ideograms. (experiment-wide Bonferroni p =  $6.41 \times 10^{-4}$ ).

### 4.3.2.3 DNA methylation 450k Array Data Validates the MeDIP-seq Results

Although DNA methylation array poorly cover the tRNAome, we wished to attempt to see if this BiS-based but differing and well-established technology was supportive at all of our DNA hypermethylation findings. TwinsUK had available 450k array on 587 individuals, and this platform includes 143 probes, covering 103 tRNAs. All the 3 top tRNAs in the MeDIP-seq results were covered by this data set, and 7 of the 16 study-wide significant set. 9 tRNAs show significant ( $p < 4.58 \times 10^{-4}$ ) increases in DNA methylation with age in models corrected for blood cell counts including all 3 of the 3 tRNAs identified in the MeDIP-seq as genome-wide significant and 5 of the 7 study-wide significant set present on the array (Figure 4.11). Although it should be noted that 56 of these 143 probes are within the non-robust set of Zhou et al. [Zhou et al., 2016], including 1 of the genome-wide, and 1 of the study-wide results (covering tRNA-Ile-AAT-4-1 & tRNA-Val-AAC-4-1), respectively.

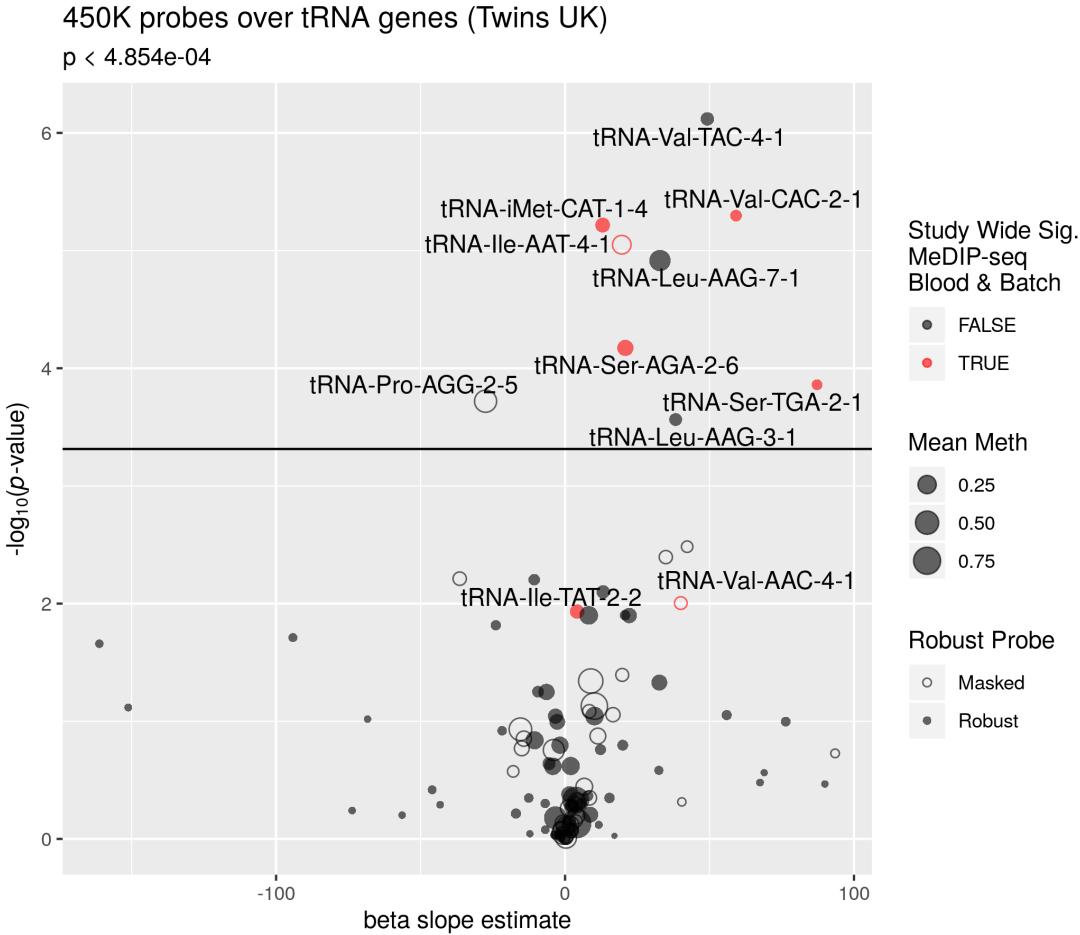


Figure 4.11: Volcano-like plot. tRNAs are labelled if they are significant here or were in the MeDIP-seq data (Red). Model slope: the model coefficient for the methylation values. Unfilled circles indicate those probes in the general mask generated by Zhou et al. [Zhou et al., 2016]. Significance threshold:  $0.05/103 \approx 4.58 \times 10^{-4}$  (the number of tRNA genes examined).

#### 4.3.2.4 Ageing-Related tRNA Loci show increased Enhancer-Related Chromatin Signatures

We further explored the activity of the tRNAome in public Chromatin segmentation data in blood (Epilogos Blood & T-cells set) [Meuleman, 2019]. This shows proportionally more Enhancer-related (Enh, EnhBiv & EnhG) chromatin states at tRNA genes hypermethylating with age than the stronger Promoter-related (TSS) in other tRNAs. (Figure 4.12). Whereas these characteristics are less frequently predominant in the rest of the tRNAs (Figure 4.12). Age-hypermethylating tRNA are enriched for enhancer chromatin states compared to the rest of the tRNAome (Fisher's Exact test  $p = 0.01$ ).

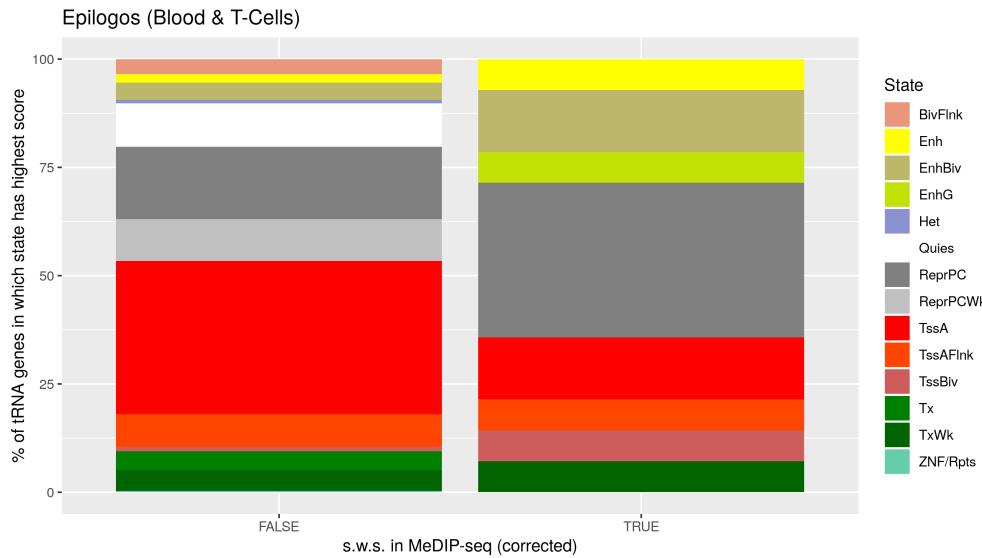


Figure 4.12: Chromatin segmentation data from the Epilogos [Meuleman, 2019] ‘Blood & T-cell’ 15 State model (tRNA genes +/- 200bp). Frequency with which a model state was the predominant state at a given tRNA. Proportions of predominant tRNA state for the 14 study-wide significant age-hypermethylating tRNAs covered compared to other 371 available tRNAs.

#### 4.3.2.5 Age Hypermethylating tRNAs are more methylated in Lymphoid than Myeloid cells

Three tRNA genes remained genome-wide significant and 16 study-wide significant following correction for major cell-type fraction. This is suggestive of either cell-type independent change or, presumably less likely, a very large effect in a minor cell-type fraction. tRNAs have exhibited tissue-specific expression [Schmitt et al., 2014b, Dittmar et al., 2006, Sagi et al., 2016] and blood cell-type populations change with age. Specifically, there is shift to favour the production of cells in the myeloid lineage [Geiger et al., 2013]. These points lead us to examine tRNA gene DNAm in sorted cell populations. We used a publically available 450k array dataset [Reinius et al., 2012]) that has been used in the construction of cell-type specific DNAm references for cell-type fraction prediction using the Houseman algorithm [Houseman et al., 2012] (see Methods). This consists of data from 6 individuals (aged  $38 \pm 13.6/\text{yrs}$ ) from seven isolated cell populations (CD4+ T cells, CD8+ T cells, CD56+ NK cells, CD19+ B cells, CD14+ monocytes, neutrophils, and eosinophils). We found that tRNA gene DNAm could separate myeloid from lymphoid lineages (Figures 4.13 & 4.19).

Of the eight study-wide significant tRNAs with array coverage, we identified that collectively these eight are significantly more methylated in the lymphoid than the myeloid lineage (1.1% difference, Wilcoxon rank sum test  $p = 1.50 \times 10^{-6}$  95% CI 0.7%–∞). Thus, any age related increases in myeloid cell proportion would be expected to dampen rather than exaggerate the age-related hypermethylation signal that we observed. In addition tRNA-Ile-AAT-4-1 and tRNA-Ser-AGA-2-6 have the highest variance in their DNAm of all 129 tRNAs covered in this dataset. This could represent ageing-related changes as these samples range across almost 3 decades. Another possibly may be that these loci as well as hypermethylating are also increasing their variability with age in a similar fashion to those identified by Slieker *et al.* [Slieker et al., 2016]. In that study they identified that those loci accruing methylomic variability were associated with fundamental ageing mechanisms.

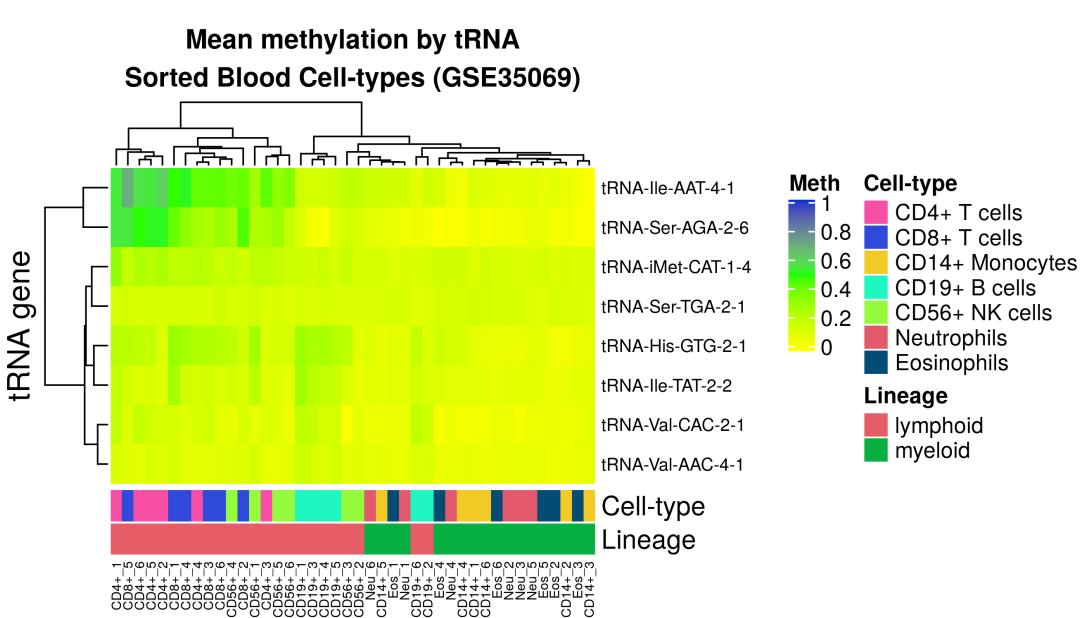


Figure 4.13: Heatmap [Gu et al., 2016] of mean methylation of probes covering each tRNA in 7 cell-type fractions from 6 Male individuals. Data from GSE35069 [Reinius et al., 2012]. Of the 16 study-wide significant hypermethylating tRNAs, 8 are covered by this dataset.

### 4.3.3 tRNA Gene DNA Methylation in Other Tissues

Some tRNA gene expression has been shown to be highly tissue specific [Schmitt et al., 2014b, Dittmar et al., 2006, Sagi et al., 2016]. It follows that our observations of changes in DNAm with age in blood might be specific to that tissue. We used a mix of 450k and 27k array data from ‘solid tissue normal’ samples made available by TCGA (The Cancer Genome Atlas) and data from foetal tissue [Yang et al., 2016, Nazor et al., 2012] downloaded from GEO (see Methods). The samples from TCGA range in age from 15-90 ( $n = 733$ ). Only 43 tRNA genes had adequate data to compare across tissues in this dataset and 115 in the foetal tissue data.

#### 4.3.3.1 tRNA Genes also Hypermethylate with Age in Solid Tissue

Only 2 of the 3 tRNA genes we identified as genome-wide significant and a further 1 of the study-wide significant tRNA genes are present in the set of 45 tRNA genes in the TCGA data, thus limiting our ability to draw conclusions about the tissue specificity of our results. Solid tissue samples have a strong preponderance for low levels of methylation consistent with the active transcription of many tRNA genes and show slight increasing methylation with age but age accounts for very little of the variance (linear regression slope estimate = 1.52;  $R^2 = 0.0002$ ; p-value  $1.34 \times 10^{-3}$  (Figure 4.20d). In a pan-tissue analysis we found that 10 tRNA genes showed changes in DNAm with age, 9 of which were hypermethylation (p-value  $< 1.1 \times 10^{-3}$ ). One of these tRNA genes, tRNA-Ser-TGA-2-1 was also present in study-wide significant set of tRNA genes. Figures 4.21 & 4.22 illustrate minimal tissue specific differences. Interestingly, however, tRNA-iMet-CAT-1-4 and tRNA-Ser-AGA-2-6 appeared more variable in methylation state than many other tRNAs in the TCGA normal tissue samples (Figure 4.21) and indeed have the highest variance in DNA methylation across tissues (Figure 4.20c).

### 4.3.4 Expression of tRNAs in Blood with Age

Having observed specific tRNA gene isodecoders hypermethylating with age we explored the expression of tsRNA in blood cell-types. We devised a bioinformatic approach to attempt to

assay tRNA transcription in order to use standard publicly available small RNA-seq datasets. We created customised MINTmap [Loher et al., 2017] reference designed to include only fragments which unambiguously map to a single tRNA gene locus and which overlap the 5' or 3' end of the genomic tRNA sequence by at least one base with no mismatches. This reference is intended to capture pre-tRNAs prior to processing and CCA addition operating under the assumption that the levels of pre-tRNAs will be informative about the amount of transcription taking place at the tRNA loci (see Methods). Our custom MINTmap reference build yielded 383 fragments mapping to 92 distinct tRNA loci in this data. The lack of coverage of age hypermethylating tRNAs by uniquely attributable RNA-seq reads prevented us from drawing any strong conclusions about the relationship between DNAm changes and changes in tRNA transcription.

Using the original MINTmap reference optimised to detect tRNA fragments derived from mature tRNAs there were 5384 unique fragments derived from as many as 417 tRNA loci. However, the mapping between fragments and loci in this reference is many to many, with each tRNA gene able to give rise to many fragments and each fragment attributable to at least 1 and usually many tRNA genes. We limited our examination of these fragments to those with a length of greater than or equal to 40nt to capture reads more likely to be derived from mature tRNAs rather than tRFs or tRNA halves (Figure 4.23). We identified 48 tsRNAs with nominally significant expression changes ( $p < 0.05$ ), 8 increased and 40 decreased in abundance with age. For example 5 of 6 fragments showing significant age-related expression changes derived from the Gln-CTG family of tRNAs are decreasing with age (Figure 4.14). This is suggestive that expression of some tRNA genes may decline with age but this possibility is in need of additional tRNA expression data before it can be asserted with confidence.

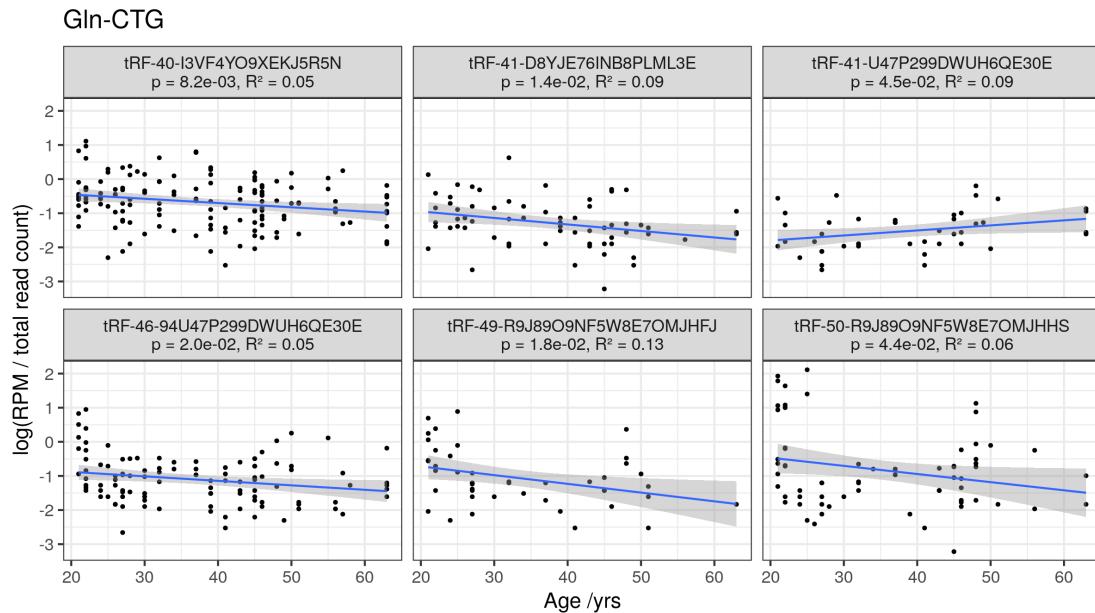


Figure 4.14: tRNA fragments derived from the Gln-CTG family of tRNAs, selected as tRNA-Gln-CTG-7-1 is one of the 16 study-wide significant age-hypermethylating tRNA genes. Panel titles contain the MINTbase Plates, unique identifiers of the tRNA fragments [Pliatsika et al., 2018].

#### 4.3.5 Mice also show age-related tRNA gene DNA hypermethylation

We examined the DNA methylation of the mouse tRNAome in using data from a reduced representation bisulfite sequencing (RRBS) experiment performed by Petkovich et al. [Petkovich et al., 2017]. These data from 152 mice covered 51 tRNA genes and 385 CpGs after QC (see Methods), representing ~11% of the mouse tRNAome. The mice ranged in age from 0.67–35 months.

Three of the 51 tRNAs showed Bonferroni significant DNA methylation changes with age ( $p$ -value  $< 1.08 \times 10^{-4}$ ) and all were in the hypermethylation direction. These three are tRNA-Asp-GTC-1-12, tRNA-Ile-AAT-1-4, tRNA-Glu-TTC-1-3 (Figure 4.15).

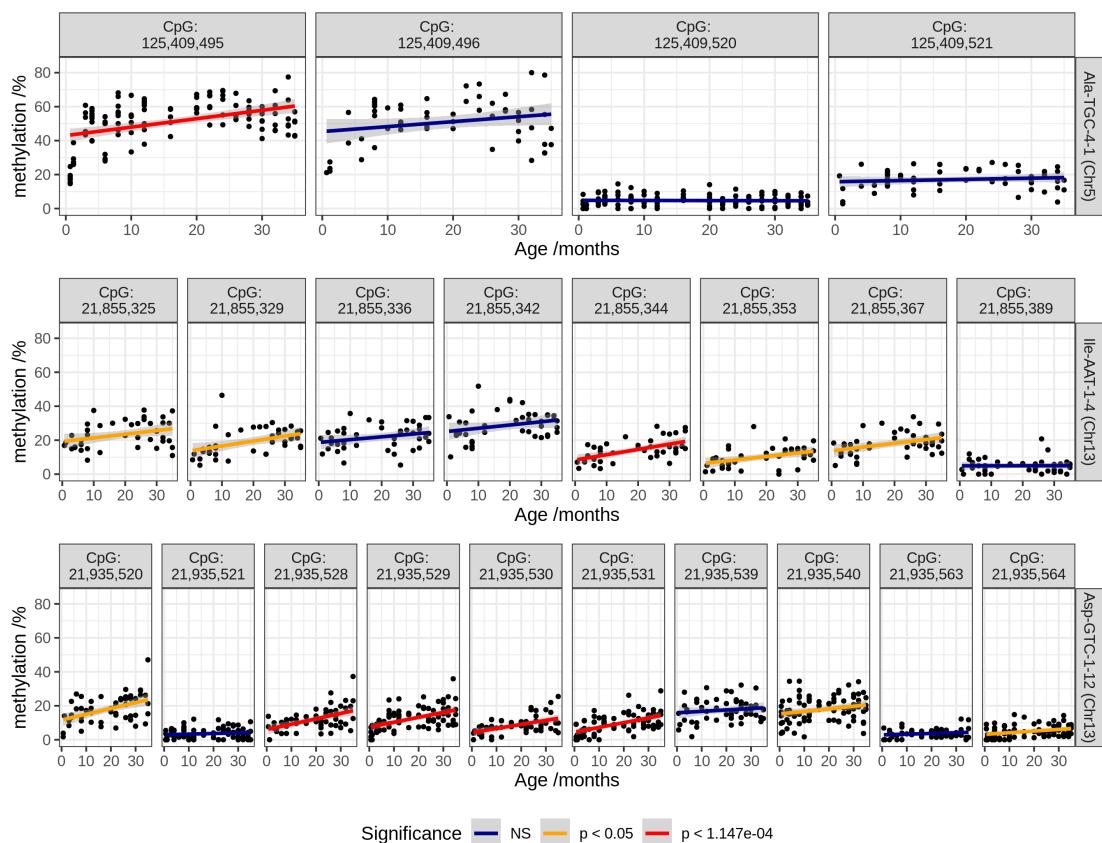


Figure 4.15: DNA methylation of CpGs in 3 tRNA which significantly hypermethylate with age in mice. 6 CpGs reach bonferroni significance and 7 show nominally significant increases.

## 4.4 Discussion

Our work has identified a previously unknown enrichment for age-related epigenetic changes within the tRNA genes of the human genome. This observation was strongly directional with increasing DNA methylation with age [Ehrlich, 2019].

The MeDIP-seq dataset employed brought advantages in exploring this undefined terrain of the tRNAome. Firstly, being genome-wide it provides much increased access, as these regions are poorly covered by current arrays. Secondly, being a fragment-based regional assessment of DNA methylation, the individual but highly similar small tRNA genes can be surrounded by unique sequence.

We determined by genome-wide permutation that this strong hypermethylation signal was specific to the tRNAome, and not merely driven by the underlying CpG density of these loci. A targeted BiS-seq experiment validated the defined nature of the tRNA change in an independent dataset, with a successful pooling approach, which may also be useful for other ageing-related targeted DNA methylome evaluations. Additionally, we gained support for our results from limited DNA methylation array data.

We subsequently explored further what was driving this age-related phenomenon and its possible biological implications. As this result was observed in peripheral blood, we were well aware that we were examining DNA derived from a heterogeneous cell type population [Lappalainen and Greally, 2017]. Moreover, that there are well known age-related proportional changes in peripheral blood cell composition [Geiger et al., 2013]. The TwinsUK MeDIP-seq and 450k array DNA methylation data included measured haematological values. Therefore, we adjusted for major cell type effects, such as a myeloid skew, and distinct tRNAs were still significant. Although, a caveat to our study is that this can not exclude changes in minor specific sub-cell fractions types. However, that these age-related effects were strong enough to be observed in both a regional MEDIP-seq assessment and a pooled sequencing approach, implies that they not extremely subtle. We examined age-related tRNA gene DNA methylation changes in the limited subset of mouse tRNA genes covered in publicly available RRBS data (~13%) and were able to identify tRNAs exhibiting DNA hypermethylation with age in this set. This suggests that age-related tRNA gene hypermethylation may not be unique to humans, but at least observed across mammals.

Due to the high number of hypermethylating tRNA prior to cell-type correction, we were also curious whether the epigenetic state of this small tRNAome fraction of genome could capture and in fact be a defined fingerprint of cell type. We found that tRNA gene DNA methylation could separate myeloid from lymphoid lineages. There also was some suggestion of more fine-grained blood cell-type signatures in tRNA DNAm, such as the separation of CD19+ B cells from CD4/8+ T cells. Ageing is also known to lead to an increase in senescent cells (*e.g.* CD8+ CD28- cells). Whether these epigenetic changes in the tRNAome uniquely represent these cell-types will require technical advances to enable future single cell DNA methylome analysis to accurately assess these regions. If further supported, the epigenetic state of these loci may aid the taxonomy of cell-type definition.

This signal within the tRNA families was observed to occur at specific Isodecoders. After correcting for major cell types, we identified 2 tRNA genes tRNA-iMet-CAT-1-4 and tRNA-Ser-AGA-2-6 which had the most consistent hypermethylation across 3 different assays. Isodecoders expand in number with organismal complexity and the high prevalence in mammals has been suggested due to their additional regulatory functionality [Goodenbour and Pan, 2006, Keam et al., 2014]. They also have distinct translational efficiency [Geslain and Pan, 2010], which can also have consequences in human disease [Kirchner et al., 2017]. Furthermore, there is great complexity to the fragmentation of tRNA [Schimmel, 2018], with physiological processes such as stress shown to induce fragment production [Li et al., 2019]. These resultant tsRNAs can feedback on protein synthesis by regulating ribosome biogenesis [Kim et al., 2017] and others have diverse regulatory functions such as targeting transposable element transcripts [Martinez et al., 2017]. They are also observed to circulate in the blood in a cell-free fashion, and fragment levels can be modulated by ageing and calorie restriction [Dhahbi et al., 2013]. The isodecoder

specific nature of our findings frame a possible hypothesis for regulatory change with age and future work will be required to unravel this potential.

Whilst, the expression of the tRNA genes has long been simplified as ‘constitutive’, some observations have indicated that many tRNA genes are expressed in a tissue-specific fashion in diverse organisms [Dittmar et al., 2006; Sagi et al., 2016]. Although others have found the majority of isodecoders are transcribed in different cell types [Parisien et al., 2013]. Several transcription factors acting via TFIIIB [Gomez-Roman et al., 2003] have a negative (the tumour suppressors p53 [Crighton, 2003] and Rb [Sutcliffe et al., 2000]) or positive (the proto-oncogene c-Myc) influence [Gomez-Roman et al., 2003]. Regulatory sequence in the flanking or the internal regions of tRNA genes do not explain tRNA expression variation [Schmitt et al., 2014a]. Whilst DNAm is able to repress the expression of tRNA genes [Besser et al., 1990], the broader chromatin environment also affects tRNA transcription. Due to the co-ordinated nature of epigenomic modifications, it may also be revealing to evaluate ageing-related histone modification in these tRNA loci.

Changes in the epigenetic state of specific tRNA could be modulating transcription efficiency or even codon availability in the ageing cell. tRNA gene dosage is quite closely matched to amino acid usage frequency in the human exome. However, the transcriptome codon usage frequency and tRNA gene expression have been claimed to vary with the replicative state of cells, separating differentiated from replicating cells [Gingold et al., 2014]. Others have argued that these differences are substantially explained by variation in GC content [Rudolph et al., 2016] and that codon usage frequencies are observed to be mostly invariant in the transcriptomes of a wide range of tissues, as well as across developmental time [Schmitt et al., 2014a]. Although, experimental stress-related states have revealed changes with an over-representation of codons that are translated by rare tRNAs [Gingold et al., 2012].

tRNA sequences themselves are under strong structural (both secondary and tertiary) [Goodenbour and Pan, 2006] as well as functional constraint, which leads to an order of magnitude reduction in variation compared the background genomic mutation rate [Parisien et al., 2013]. However, polymorphic tRNA could be another potential caveat to our work. Although, there is no significant population variation in, for example, tRNA iMet sequences in 1,000 Genomes data. Indeed, there are only 11 new isodecoder sequences with high confidence (tRNAscan scores  $\geq 50$ ) at  $>1\%$  population frequency [Parisien et al., 2013]. There is also some evidence for tRNA copy number variation at specific loci, although this remains under-characterised [Iben and Maraia, 2014; Darrow and Chadwick, 2014]. Another potential cause we considered was whether ageing-related somatic copy number increases could be occurring in these loci. Population or somatic copy number expansions could lead to increased methylated reads in MeDIP-seq without any epigenetic state change. However, this would not be consistent with the targeted and array Bis conversion methodologies, where the proportion of methylated to unmethylated reads would still be constant.

It is worth noting the parallels with known cancer and ageing epigenetic changes, and that tRNAs are also dysregulated in cancer [Huang et al., 2018], with proposed utility as prognostic markers [Krishnan et al., 2016]. Furthermore, the early replicating state of tRNA loci, potentially associated with high expression [Müller and Nieduszynski, 2017], may make them prone to hypermethylate, as is observed in early replicating loci in both cancer [Du et al., 2019] and senescent cells [Cruickshanks et al., 2013]. Interestingly, tRNA gene loci may also play a role in local as well as large scale genome organisation [Hamdani2019; Van Bortle et al., 2017]. tRNA gene clusters act as insulators [Raab et al., 2012] and have extensive long-range chromatin interactions with other tRNA gene loci [Van Bortle et al., 2017]. The coordinated transcription of tRNAs at subnuclear foci and the B-box sequence elements bound by TFIIIC and not PolIII may represent an organising principle for 3D-chromatin by providing spatial constraints [Noma et al., 2006]. Therefore, these tRNA epigenetic changes could contribute to the structural changes that are also observed in ageing [Sun et al., 2018].

In conclusion, due to the unique challenges that make the tRNAome difficult to examine it has remained epigenetically under-characterised despite its critical importance for cell function. We directly interrogated the epigenetic DNA methylation state of the functionally important tRNAome, across the age spectrum in a range of datasets as well as methodologies and identified an enrichment for age-related DNA hypermethylation in the human tRNA genes.

## 4.5 Methods

### 4.5.1 Participants

Participants in the ‘EpiTwins’ study are adult volunteers from the TwinsUK Register. The participants were aged between 16 and 82 years, with a median of ~55 years (cohort profile [Moayyeri et al., 2013]). Ethics for the collection of these data were approved by Guy’s & St Thomas’ NHS Foundation Trust Ethics Committee (EC04/015—15-Mar-04) and written informed consent was obtained from all participants.

Participants for our targeted bisulfite sequencing of select tRNA loci were drawn from two studies. Samples from participants aged 4 and 28 years are from the MAVIDOS [Harvey et al., 2012] study and participants aged 63 and 78 years are from the Hertfordshire cohort study [Syddall et al., 2005]. Due to a limited number of available samples, the two 4 year old pools contained DNA from 20 individuals each, with all other pools having 25 contributing individuals. Pool 1, the first 4 year old pool used DNA from all male samples, with all other pools using all female samples. Thus, the total number of participants was 190 (see Table 4.2). Samples from the 28 year old time point are all from pregnant women at ~11 weeks gestation.

### 4.5.2 tRNA annotation information

Genomic coordinates of the tRNA genes were downloaded from GtRNAdb [Chan and Lowe, 2009]. The 2 tRNAs located in chr1\_gl000192\_random are tRNA-Gly-CCC-8-1 & tRNA-Asn-ATT-1-2 (Supplementary File S1). Stem/loop structure annotations were inferred from output of tRNAscan [Lowe and Chan, 2016] with a custom perl script. The 213 probes overlapping tRNA genes were derived from intersecting the tRNA gene annotation data from gtrRNAdb with the Illumina 450k array manifest annotation for the hg19 genome build using bedtools v2.17.0 [Quinlan and Hall, 2010]. We excluded 107 tRNAs from blacklisted regions of hg19 [Amemiya et al., 2019].

### 4.5.3 DNA methylome data

#### 4.5.3.1 TwinsUK MeDIP-seq methylomes

The Methylated DNA Immunoprecipitation sequencing (MeDIP-seq) data was processed as previously described [Bell et al., 2016, 2018]. These processed data are available from the European Genome-phenome Archive (EGA) (<https://www.ebi.ac.uk/ega>) under study number EGAS00001001910 and dataset EGAD00010000983. The dataset used in this work consists of 4350 whole blood methylomes with age data. 4054 are female and 270 male. 3001 have full blood counts. There are 3652 individuals in this data set. These individuals originate from 1933 unique families. There are 1234 monozygotic (MZ) twin pairs (2468 individuals), and 458 dizygotic (DZ) twin pairs (916 individuals).

MeDIP-seq used a monoclonal anti-5mC antibody to bind denatured fragmented genomic DNA at methylated CpG sites. This antibody-bound fraction of DNA was isolated and sequenced [Down et al., 2008]. MeDIP-seq 50-bp single-end sequencing reads were aligned to the hg19/GRCh37 assembly of the human genome and duplicates were removed. MEDIPS (v1.0) was used for the MeDIP-seq specific analysis [Lienhard et al., 2014]. This produced reads per million base pairs (RPM) values binned into 500bp windows with a 250bp slide in the BED format, resulting in ~12.8 million windows on the genome. MeDIP-seq data from regions of interest was extracted using Bedtools v2.17.0 [Quinlan and Hall, 2010].

#### 4.5.3.2 Analysis of DNA methylome data for Significant Ageing-related changes

All analysis was performed in R/3.5.2. Linear models were fitted to age using the MeDIP-seq DNA methylome data, as quantile normalised RPM scores at each 500bp window. Models were

fitted with: 1. No covariates; 2. Batch information as a fixed effect; 3. Blood cell-type counts for neutrophils, monocytes, eosinophils, and lymphocytes as fixed effects; and 4. Batch and Blood Cell counts as fixed effects. Model 1 & 2 were fitted on the full set of 4350 as batch information was available for all samples but blood cell count data was only available for a subset of 3001 methylomes. Models 1 & 2 fitted in the n=3001 subset were similar to those fitted in the complete set of 4350. Models 3 & 4 were fitted in the n=3001 subset with full covariate information and sets of significant tRNAs identified at study-wide and genome wide levels in model 4 were used in subsequent analyses. Models were also fitted for two unrelated subsets created by selecting one twin from each pair (Monozygotic or Dizygotic), yielding sets with n = 1198 & 1206 DNA methylomes. One additional model was fitted for longitudinal analysis, samples were selected by identifying individuals with a DNA methylome at more than one time point and filtering for only those with a minimum of 5 years between samples. This yielded 658 methylomes from 329 individuals with age differences of 5-16.1 yrs, median 7.6 yrs. Models for this set included participant identifier as a fixed effect in addition to blood cell counts and batch information.

#### 4.5.3.3 Permutation Analysis for Enrichment with Age-related Changes

We performed a permutation analysis to determine whether the CpG distribution of sets of the tRNAome was the principle driver of the ageing-related changes observed. Windows overlapping tRNAs have a higher proportion of windows with a greater CpG density than their surrounding sequences (see supplementary Figure 4.18). CpGs residing within moderate CpG density loci are the most dynamic in the genome [Ziller et al., 2013] and CpG dense CpG island regions include specific ageing-related changes [Teschendorff et al., 2010, Rakyan et al., 2010, Bell et al., 2016]. For comparison we also performed the permutation in the CGI regions from the Polycomb group protein target promoters in Teschendorff *et al.* [Teschendorff et al., 2010] and bivalent loci from ENCODE ChromHMM ‘Poised Promoter’ classification in the GM12878 cell-line [Ernst et al., 2011]. A random set of 500bp windows representing an equivalent CpG density distribution of the feature set in question were selected from the genome-wide data. Above a certain CpG density there are insufficient windows to sample without replacement within a permutation. Furthermore, above  $\sim \geq 18\%$  CpG density CpG Islands become consistently hypomethylated [Bell et al., 2012a]. Therefore, all windows with a CpG density of  $\geq 18\%$  (45 CpGs per 500bp) were grouped and sampled from the same pool. i.e. a window overlapping a tRNA gene which had a 20% density could be represented in permutation by one with any density  $\geq 18\%$ . This permutation was performed 1,000 times to determine an Empirical p value by calculating the number of times the permutation result exceeded the observed number of significant windows in the feature set. *Empirical p – value* =  $\frac{r+1}{N+1}$ , where r is the sum of significant windows in all permutations and N is number of permutations [North et al., 2003].

#### 4.5.3.4 Neonate and Centenarian Whole Genome Bisulfite Sequencing

DNA methylation calls were downloaded from GEO:%5BGSE31263%5D(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31263>) and intersected with tRNA genes using bedtools v2.17.0 [Quinlan and Hall, 2010].

#### 4.5.3.5 Sample pooling and EPIC array

We performed an Illumina Infinium DNA methylation EPIC array ((C) Illumina) and targeted bisulfite sequencing of select tRNA gene loci. Here we used DNA extracted from whole blood and pooled into 8 samples from unrelated individuals at 4 time-points with 2 pools at each time-point. The timepoints were 4, 28, 63, and 78 years. Using the EPIC array we were able to infer the DNAm age using the Horvath DNAm clock [Horvath, 2013] and blood cell-type composition of our samples using the Houseman algorithm [Houseman et al., 2012].

#### 4.5.3.6 Targeted Bisulfite Sequencing

We selected tRNA loci for targeted sequencing in which have had observed changes and DNAm with age and closely related tRNAs in which changes were not observed. Primer design was performed using ‘methPrimer’ [Li and Dahiya, 2002] (Supplementary File S2). A total of 84 tRNA loci were targeted and 79 subsequently generated reliable results post-QC. The targeted tRNAs covered a total of 723 CpGs with a median of 8 CpGs per tRNA (range 1-13), data passing QC was generated for 458 CpGs, median 6 (range 1-9) per tRNA.

Quality was assessed before and after read trimming using `fastqc` [Andrews, 2010] and `multiqc` [Ewels et al., 2016] to visualise the results. Targeting primers were trimmed with `cutadapt` [Martin, 2011] and a custom `perl5` script. Quality trimming was performed with `trim_galore` [Kueger, 2015]. Alignment and methylation calling was performed with `Bismark` (v0.20.0) [Kueger and Andrews, 2011] making use of `bowtie2` [Langmead and Salzberg, 2012]. The alignment was performed against both the whole hg19 genome and just the tRNAome +/- 100bp to assess the possible impact of off-target mapping. Mapping to the whole genome did produce purported methylation calls at a larger number of loci than mapping just to the tRNAome (683,783 vs 45,861 respectively). Introducing a minimum coverage threshold of 25 reads dramatically reduced this and brought the number of sites into line with that in the tRNAome set (36,065 vs 33,664 respectively) suggesting a small number of ambiguously mapping reads. All subsequent analysis was performed using the alignment to just the tRNAome with a minimum coverage of 25 reads.

We performed pairwise differential methylation analysis of the tRNA genes at the different time points using `RnBeads` [Müller et al., 2019] with `limma` [Ritchie et al., 2015] and a minimum coverage of 25 reads. We also performed linear regression predicting age from DNA methylation at the targeted tRNA sites, permitting us to compare rates of increase with age. For the linear regression, we used only CpG sites with more than 25 reads mapped to the regions of the genome targeted for amplification.

#### 4.5.3.7 TwinsUK Illumina 450k array methylomes

Illumina Infinium DNA methylation 450k arrays ((C) Illumina) were also performed on TwinsUK participants, in 770 Blood-derived DNA samples which had matched MeDIP-seq data. These data were preprocessed in the form of methylation ‘beta’ values pre-processed as previously described [Bell et al., 2016, 2018]. Cell-type correction was performed using cell-count data and the following model: `lm(age ~ beta + eosinophils + lymphocytes + monocytes + neutrophils)`.

#### 4.5.4 Chromatin Segmentation Data

Epilogos chromatin segmentation data [Meuleman, 2019] was downloaded for the tRNA gene regions +/- 200bp from [https://explore.altius.org/tabix/epilogos/hg19.15.Blood\\_T-cell.KL.gz](https://explore.altius.org/tabix/epilogos/hg19.15.Blood_T-cell.KL.gz) using the `tabix` utility. The data used was the ‘Blood & T-cell’ 15 State model based on segmentation of 14 cell-types. This data was manipulated and visualised with `R` and `ggplot2`.

#### 4.5.5 Isolated Blood Cell Type Specific Data

Data from 7 cell-type fractions from 6 Male individuals was downloaded from GSE35069 [Reinius et al., 2012] using `GEOquery` [Sean and Meltzer, 2007]. Five of the 6 top age hypermethylating tRNAs are covered by this array dataset.

#### 4.5.6 Cancer and Tissue Specific Methylation Data

Data was downloaded from the TCGA (The Cancer Genome Atlas) via the GDC (genomic data commons) data portal [Grossman et al., 2016] using the `GenomicDataCommons` R package. Data

from foetal tissue [Yang et al., 2016, Nazor et al., 2012] was downloaded from GEO (GSE72867, GSE30654). From the TCGA, we selected samples for which DNAm data was available from both the primary site and normal solid tissue, and for which we could infer an approximate age (within one year). We selected those probes overlapping tRNA genes yielding 73,403 data points across 19 tissues with an age range of 15-90yrs (median 63.4) (Supplementary File S3)

#### 4.5.7 Assaying tRNA expression in blood with MINTmap

We used small RNA-seq data from sorted blood cell fractions [Juzenas et al., 2017] (GSE100467) and the MINTmap [Loher et al., 2017] tRNA fragment alignment tool. This dataset covered 42 individuals aged 21-63. We also created a customised MINTmap reference designed to include only fragments which unambiguously map to a single tRNA gene locus and which overlap the 5' or 3' end of the genomic tRNA sequence by at least one base with no mismatches. This reference is intended to capture pre-tRNAs prior to processing and CCA addition operating under the assumption that the levels of pre-tRNAs will be informative about the amount of transcription taking place at the tRNA loci. This approach provides at most a many to one mapping of tRNA fragment to a tRNA gene.

Assaying the expression of tRNA genes presents numerous difficulties [Torres et al., 2019], and usually requires variants on standard RNA-seq protocols. Our custom MINTmap reference build yielded 383 fragments mapping to 92 distinct tRNA loci in this data. To control quality only fragments with more than 20 total instances in the dataset and present in more than 20 individuals were considered.

The maximum length of a fragment was limited to 50nt, due to the read length of the small RNA-seq data.

#### 4.5.8 Mouse RRBS Analysis

We downloaded methylation calls and coverage information resulting from RRBS performed by Petkovich *et al.* [Petkovich et al., 2017] from GEO using **GEOquery** [Sean and Meltzer, 2007] GSE80672. These data from 152 mice covered 68 tRNA and 436 CpGs after QC requiring >50 reads per CpG and >10 data points per tRNA. We excluded 5 tRNAs from blacklisted regions of mm10 [Amemiya et al., 2019]. After QC there were 58 tRNA genes and 385 CpGs. We performed simple linear modeling to predict age from methylation level at each tRNA and each CpG.

#### 4.5.9 Data availability

The MeDIP-seq data supporting the results of this article are available in the EMBL-EBI European Genome-phenome Archive (EGA) under Data set Accession number EGAD00010000983 (<https://www.ebi.ac.uk/ega/datasets/> EGAD00010000983). The targeted BiS-sequencing data will be available on publication.

#### 4.5.10 Code availability

Available at [https://github.com/RichardJActon/tRNA\\_paper\\_code](https://github.com/RichardJActon/tRNA_paper_code)

#### 4.5.11 Acknowledgements

We gratefully acknowledge the individuals from TwinsUK, Mavidos and the Hertfordshire cohort. TwinsUK received funding from the Wellcome Trust (Ref: 081878/Z/06/Z), European Community's Seventh Framework Programme (FP7/2007-2013), the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. Further funding support for the EpiTwin project was obtained from the European Research Council (project number 250157) and BGI. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR. The authors would like to thank Nikki Graham for her assistance with the identification and pooling of the MAVIDOS and Hertfordshire DNA samples. The authors also acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work. The MRC-LEU is supported by the Medical Research Council (MRC). CGB received support from Diabetes UK (16/0005454). RJA was in receipt of a MRC Doctoral fund (1820097).

#### 4.5.12 Author Contributions

RJA designed experiments and analysed all the processed and experimental data. CGB conceived and designed the experiments. TDS, KW and JW conceived and provided TwinsUK MeDIP-seq data. YX, FG and JW produced raw MeDIP-seq data with WY and JB processing and quality controlling these data. WY contributed an analysis concept. CC, NH, ED, and KL provided MAVIDOS and Hertfordshire sample data. EB, EW, and CAM performed the targeted Bis sequencing experiment. PGH contributed additional data and discussion of results. RJA and CGB wrote the paper. All authors reviewed and approved the final manuscript.

## Supplementary materials

### 4.5.13 Supplementary Figures

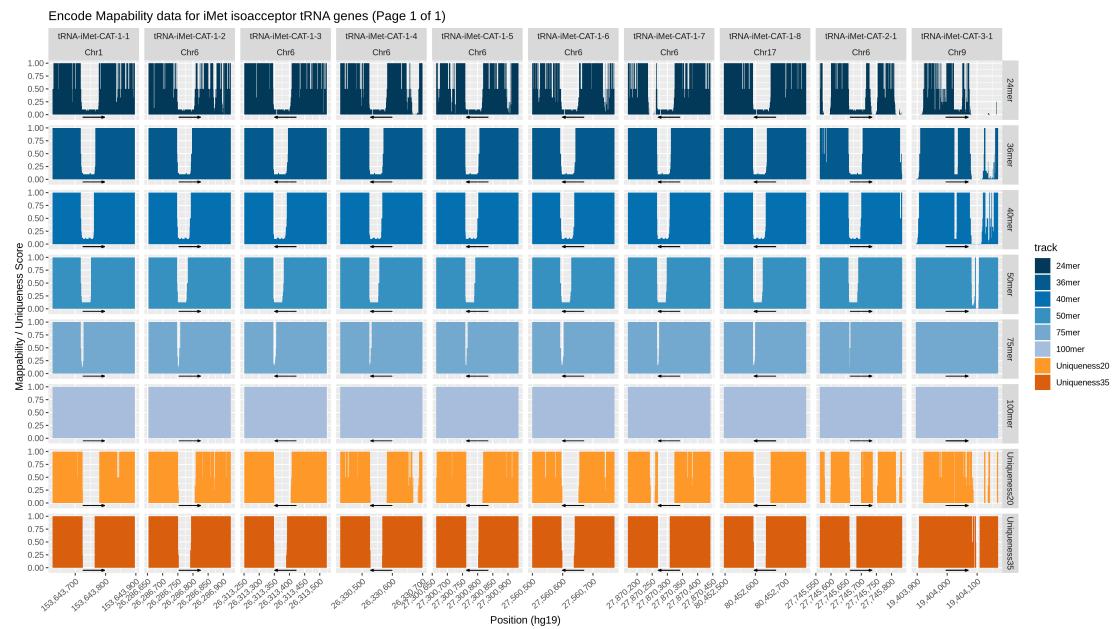


Figure 4.16: Example of mappability data from the encode mappability tracks [Derrien et al., 2012] for the initiator methionine tRNA genes.

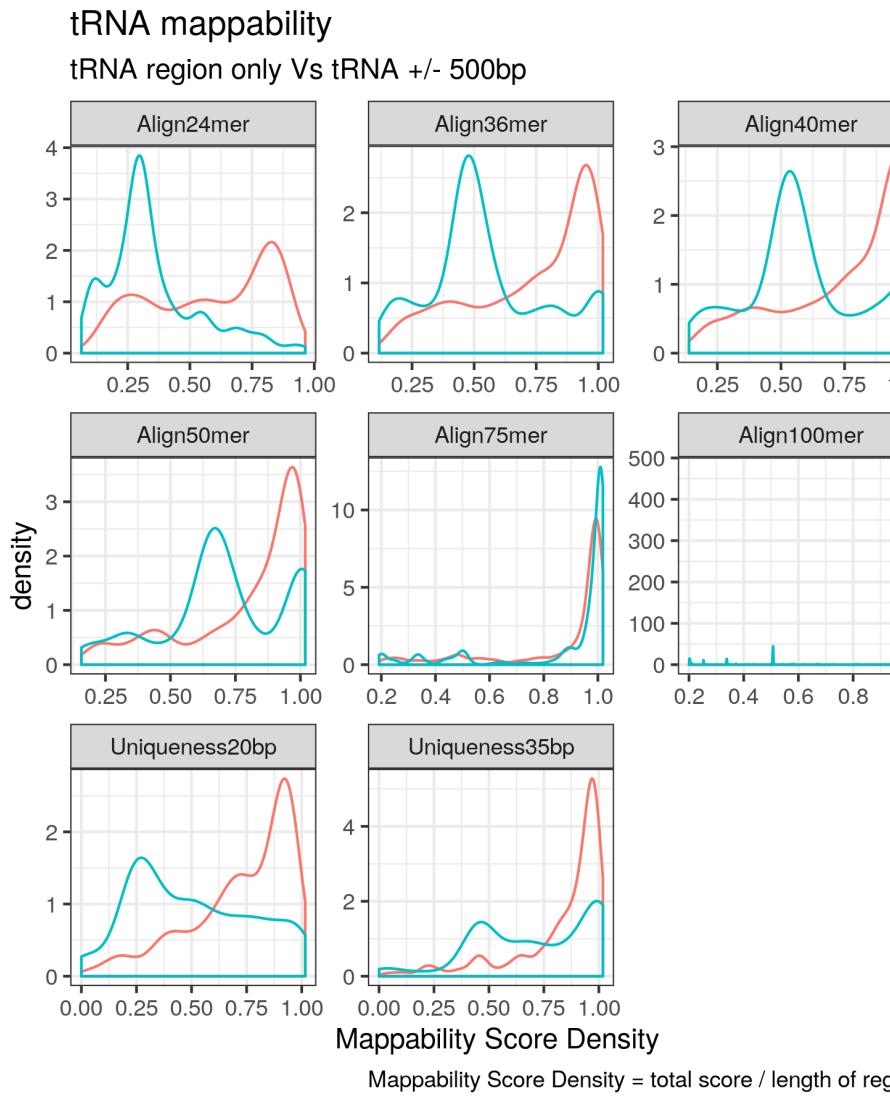


Figure 4.17: Mappability score density of the tRNAome increases with read length and is greater when flanking regions ( $\pm 500\text{bp}$ ) are included. Mappability score density is computed as the area under the encode mappability tracks [Derrien et al., 2012] over the length of the region.

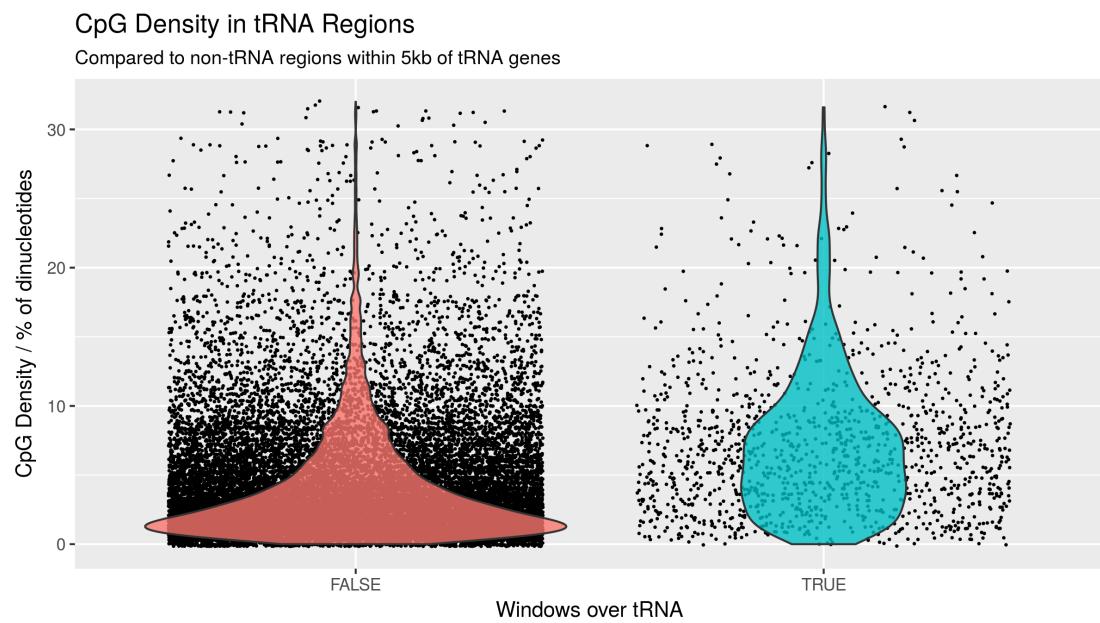


Figure 4.18: CpG Density in windows overlapping tRNA genes compared to that of non-tRNA overlapping windows in flanking sequences (+/-5kb)

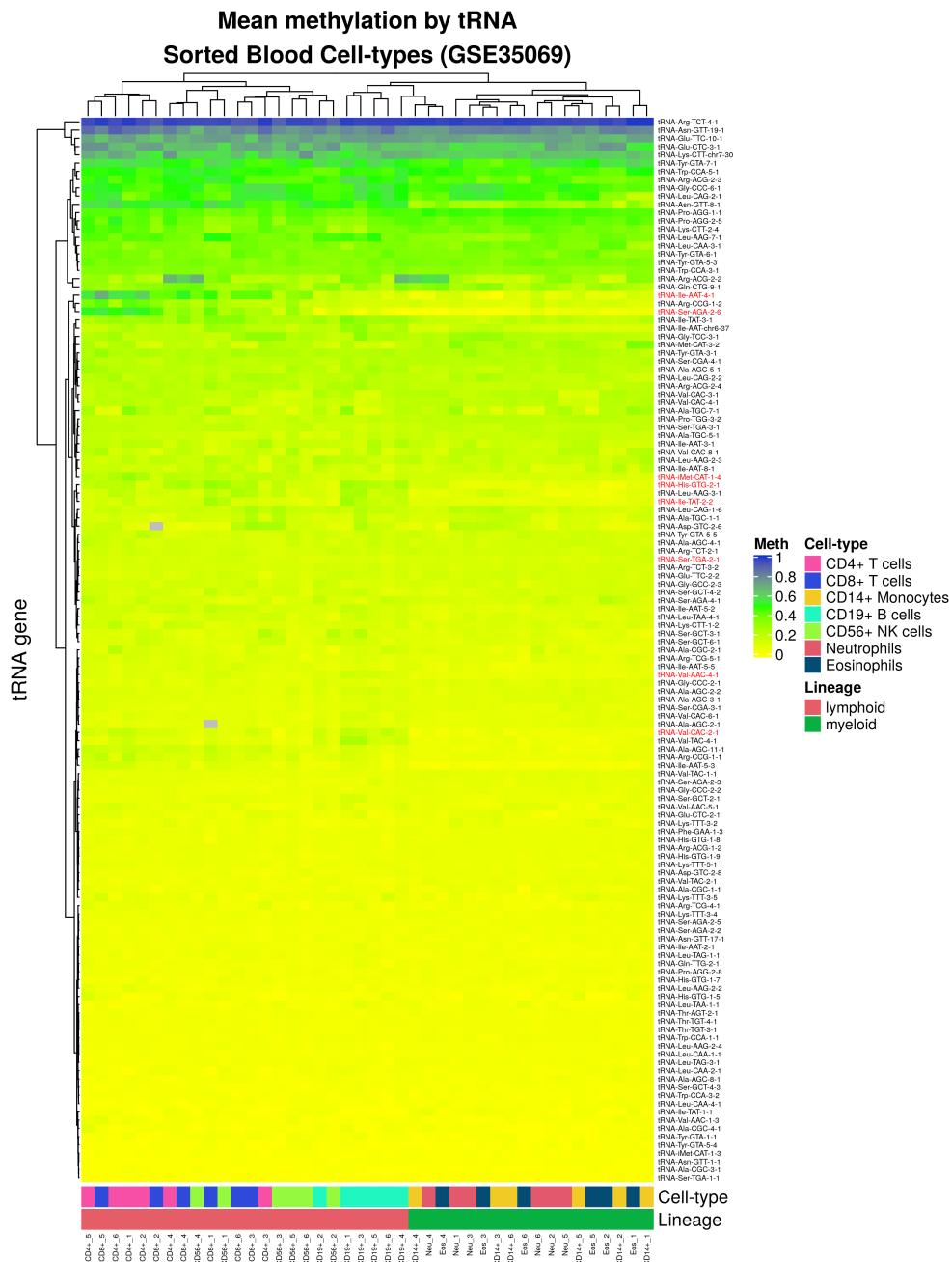


Figure 4.19: Heatmap Mean Methylation of probes covering each tRNA in 7 cell-type fractions from 6 Male individuals. Showing all 150 tRNAs covered by 213 probes on the Illumina 450k array. Data from GSE35069 [Reinius et al., 2012] downloaded using GEOquery [Sean and Meltzer, 2007]. Generated with the ComplexHeatmap R package [Gu et al., 2016].

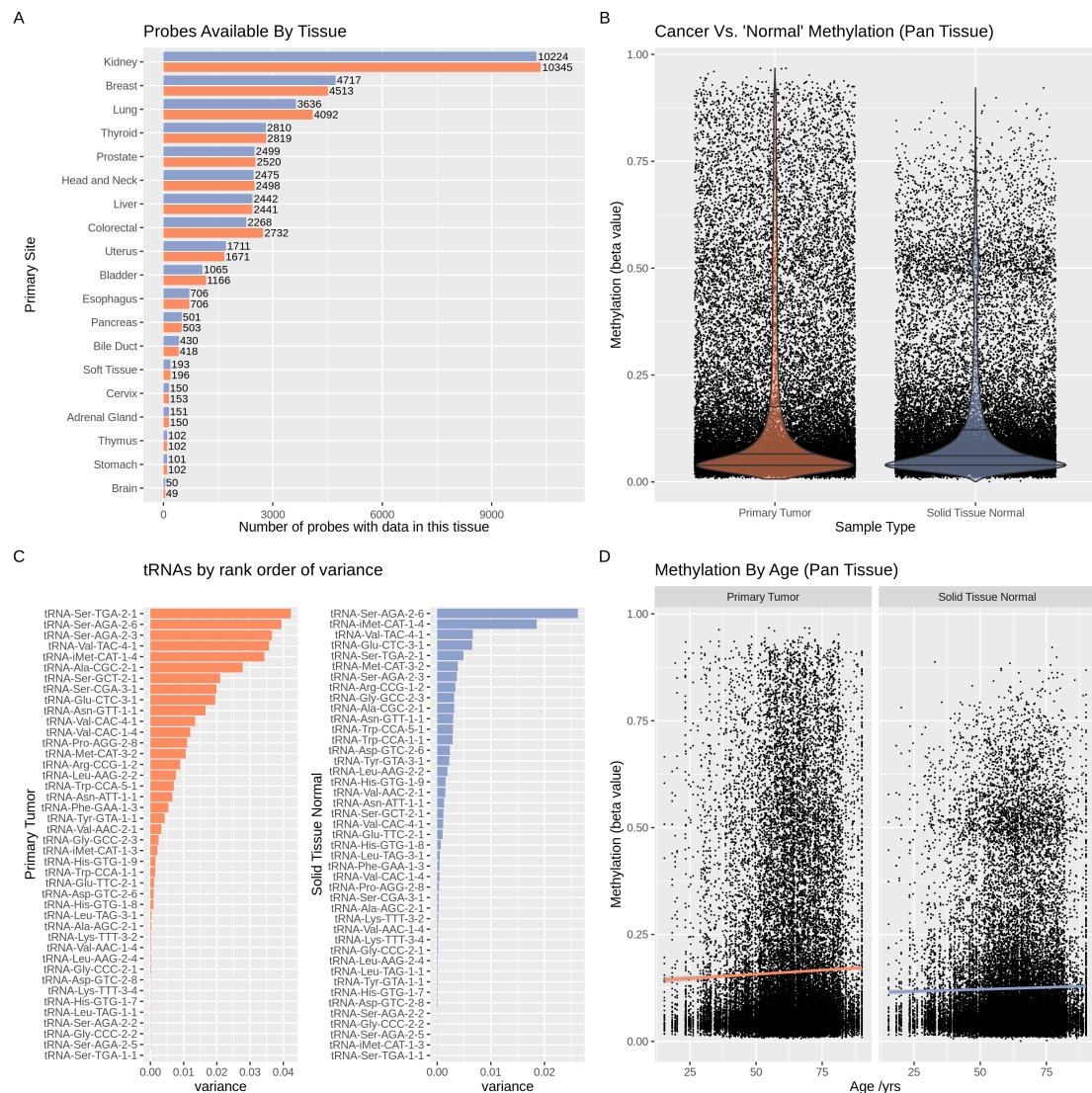


Figure 4.20: Global properties of tRNA methylation data for 45 tRNA genes across 19 tissues with matched normal and tumour samples from 733 cases in TCGA [Yang et al., 2016, Nazor et al., 2012].

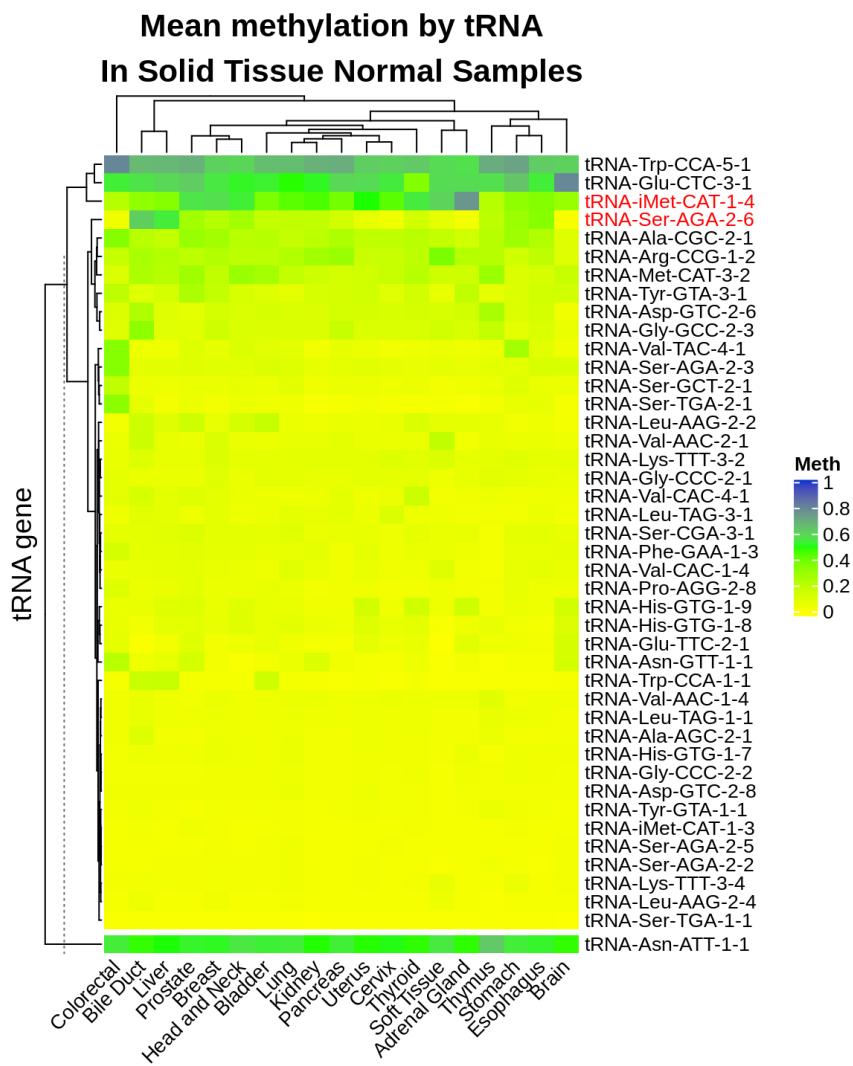


Figure 4.21: Mean Methylation of 43 tRNAs in 19 tissues. Possible pseudogene (tRNA-Asn-ATT-1-1) is shown in a separate cluster beneath the main heatmap [Gu et al., 2016].

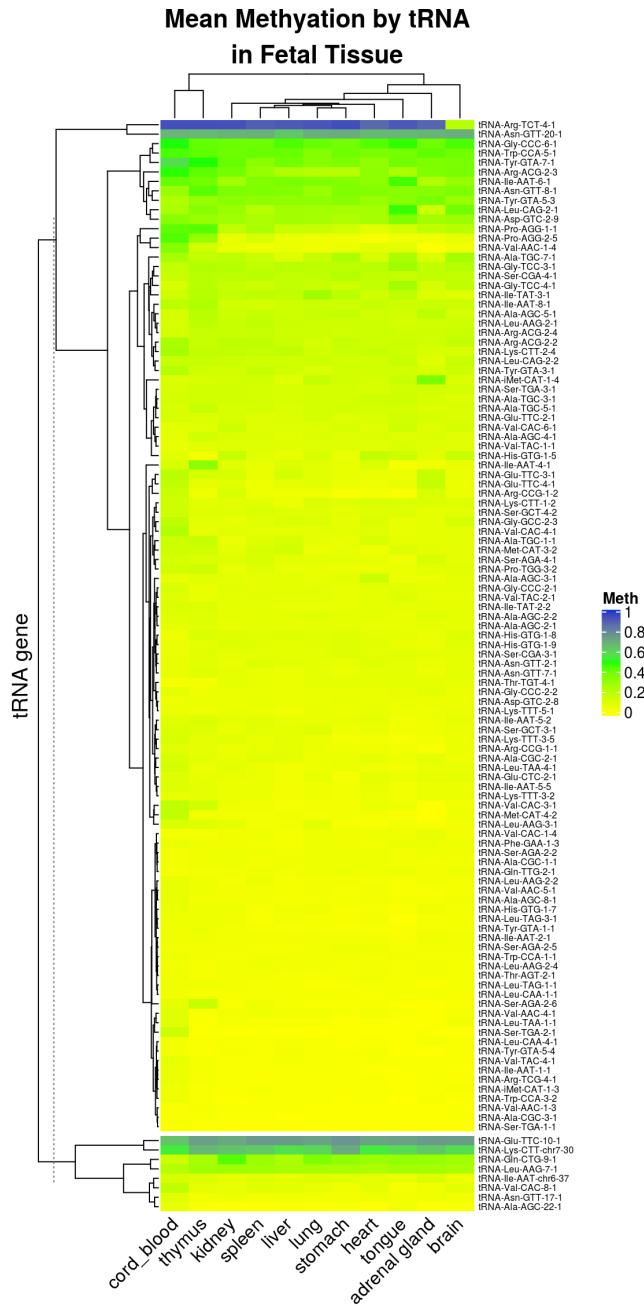


Figure 4.22: Mean Methylation of 115 tRNAs in 11 tissues. Possible pseudogenes are shown in a separate cluster beneath the main heatmap [Gu et al., 2016].

#### 4.5.14 MINTmap reference Fragment distribution

In the original MINTmap reference (Figure 4.23b) there are peaks at around 18nt, 22nt and 32nt. This is consistent with the expected tRNA fragment size distributions with ‘tRNA halves’ at 30-33nt and other tRFs at 18nt and 22nt. In our custom reference (Figure 4.23a) whilst there is still a peak at ~18nt, with suggestions of peaks near 22nt and 32nt the tRNA fragment length distribution is somewhat different from that of the standard MINTmap reference. There are larger peaks at ~28 and ~40nt consistent with the longer fragments expected given that this reference aimed to target fragments derived from pre-tRNAs not tRFs derived from mature tRNAs.

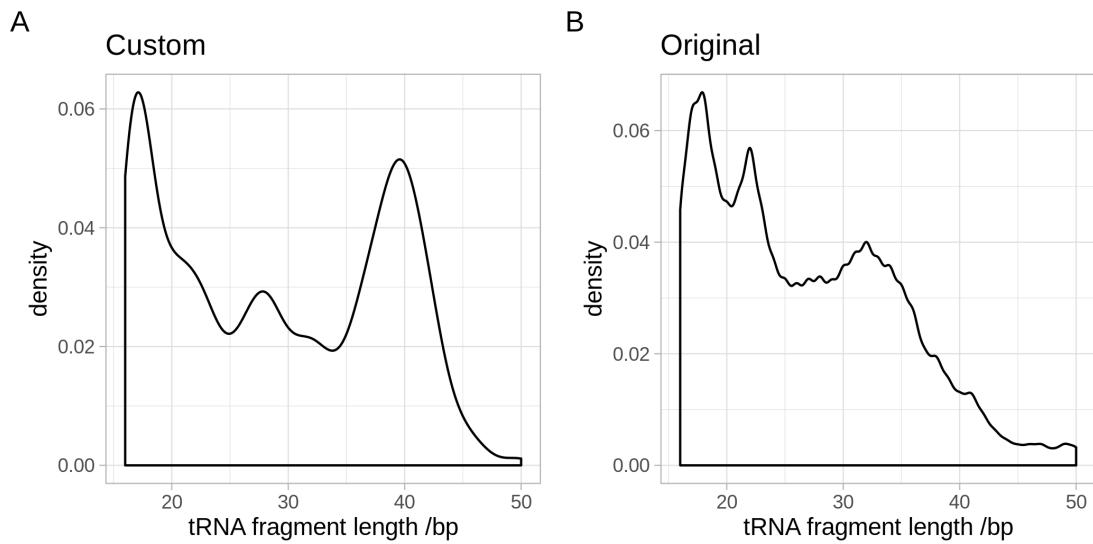


Figure 4.23: Comparison of the fragment size distributions between our custom reference (A) and the original the MINTmap reference (B).

## 4.6 References

## **Part V**

## **Results 3**



## Chapter 5

# Alu Clock Paper Draft Introduction

Age is a significant risk factor across the breadth of chronic diseases, including cardiovascular, obesity-related, ophthalmic, musculoskeletal and neurological disorders [Partridge et al., 2018]. The molecular changes that occur with age include alterations to the epigenome [Booth and Brunet, 2016] – the chemical modifications and packaging of the genome that indicate or facilitate cell-specific activity [Bird, 2007]. This includes age-related changes to DNA methylation, the most common modification of DNA [López-Otín et al., 2013]. This epigenetic mechanism is crucial in repressing the large proportion of the human genome (~45% [Gregory, 2005]) that is comprised of repetitive or transposable elements [Deniz et al., 2019], [Kazazian and Moran, 2017]. It also classically leads to robust repression of gene expression through promoter DNA methylation [Deaton and Bird, 2011], and has more nuanced effects on expression within the gene body [Hellman and Chess, 2007], as well as influencing splicing [Shukla et al., 2011] and transcription factor binding-sites (TFBSs) [Blattler et al., 2014], [Yin et al., 2017].

A significant decrease in DNA methylation with age was first observed with pioneering global measures [Wilson and Jones, 1983], in similar fashion to the reduced methylation seen in cancer tissue [Gama-Sosa et al., 1983]. This DNA methylation loss was seen to be strongly contributed to by the demethylation of repetitive elements [Bollati et al., 2009]. Random drift in the epigenetic state with age was also observed [Bollati et al., 2009], and more recent studies have identified targeted hypermethylation of specific functional regions, such as the promoters of polycomb target genes [Teschendorff et al., 2010], bivalent chromatin domains [Rakyan et al., 2010] and the genomic loci of specific tRNA genes [Acton et al., 2019]

Alu elements are the most successful transposon in terms of numbers within the human genome, totalling ~1.19 million copies and comprising ~10.7% of the genomic sequence [Dewannieux et al., 2003], [Deininger, 2011]. This ~300 bp primate-specific retrotransposon is a member of the short interspersed nuclear elements (SINEs) and is derived from the 7SL RNA gene [Ullu and Tschudi, 1984]. Alu elements arose ~65 million years ago (MYA), with a peak in amplification around ~40 MYA [Ade et al., 2013]. Non-LTR retrotransposons, which include Alus as well as L1 and SVA elements, are active in human. However, Alu elements themselves are non-autonomous and need to commandeer a nearby L1 element's retrotransposition molecular machinery [Cordaux and Batzer, 2009]. Alu insertion mutagenesis as well as Alu element mediated recombination account for 0.4% of all human genetic disorders [Kazazian and Moran, 2017], [Belancio et al., 2008].

DNA methylation represses Alu elements and due to their sequences being rich in CpGs they in fact account for ~25% of all the CpGs in human genome (~7.5 million) [Deininger, 2011]. Alu repeats are predominately methylated across all normal tissues, and even in cancer cells they are the more resistant to hypomethylation than other repetitive elements, implying a strong pressure to maintain their epigenetic state [Jordà et al., 2017]. These elements are an abundant genetic resource of regulatory sequences with numerous TFBSs residing within its canonical

sequence [Polak and Domany, 2006]. They can influence gene expression [Chen and Yang, 2017], [Ferrari et al., 2019] and there is evidence that some may possess enhancer activity [Su et al., 2014]. The deterioration of epigenomic repression can unmask latent regulatory function within repetitive elements [Ward et al., 2013], [Chuong et al., 2017], [Xie et al., 2013]. This has been clearly displayed as a pathogenic mechanism in cancer, with epigenetic reactivation of cryptic cis-regulatory elements (CREs) within transposons acting as oncogenic drivers [Jang et al., 2019], including Alu elements [Rajendiran et al., 2016]. There is the potential for these changes to also influence common non-malignant age-related diseases [Chuong et al., 2017], [Payer et al., 2017].

The identification of specific age-related DNA methylation changes has enabled the construction of ‘DNA methylation clocks’ [Horvath and Raj, 2018] that can predict an individual’s chronological age, but also capture aspects of biological ageing, such as all-cause mortality [Marioni et al., 2015]. Hypermethylated and hypomethylated cytosines approximately equally drive these DNA methylation clocks, although these age-related increases or decreases in methylation are driven by distinct biological mechanisms [Mozhui and Pandey, 2017], [Wu and Zhang, 2017]. Interestingly, hypomethylating CpGs have been identified to be the most indicative of biological ageing-related all-cause mortality effects in blood [Zhang et al., 2017]. However, due to poor coverage of the repetitive space because of technical issues [Bell et al., 2019], much of the genomic hypomethylation is not well captured by array analysis.

In this study we set about to identify and explore age-related DNA methylation changes in Alu elements in a large ( $>3,000$ ) immunoprecipitation-derived sequencing-based dataset (MeDIP-seq) from peripheral blood DNA. We constructed an Alu-specific DNA methylation ageing clock and compared how this contrasted with other clocks’ measures. We also investigated genetic influences on age- acceleration measures from this Alu-clock via GWAS, as has been shown previously with an association in the TERT gene with the Horvath clock [Lu et al., 2017], [Gibson et al., 2019]. These results may reveal mechanisms involved in controlling the epigenetic state of this large portion of the genome. Finally, we interrogated these genetic associations for any potential connection with common age-related diseases through PheWAS in the UKBiobank dataset [Bycroft et al., 2017].

## **Chapter 6**

# **Final Words**

We have finished a nice book.



# Bibliography

- Richard J Acton, Emma Bourne, Jordana Bell, Karen Lillycrop, Jun Wang, Elaine Dennison, Nicholas Harvey, Tim D Spector, Cyrus Cooper, and Christopher G Bell. The Genomic Loci of Specific Human tRNA Genes Exhibit Ageing-Related DNA Hypermethylation. *bioRxiv*, pages 1–48, 2019. doi: 10.1101/870352. URL <https://www.biorxiv.org/content/10.1101/870352v1>.
- Martyna Adamowicz, Klio Maratou, and Timothy J Aitman. Multiplexed DNA Methylation Analysis of Target Regions Using Microfluidics (Fluidigm). In *Methods in molecular biology (Clifton, N.J.)*, volume 1708, pages 349–363. 2018. doi: 10.1007/978-1-4939-7481-8\_18. URL <http://www.ncbi.nlm.nih.gov/pubmed/29224153>.
- Catherine Ade, Astrid M. Roy-Engel, and Prescott L. Deininger. Alu elements: an intrinsic source of human genome instability. *Current Opinion in Virology*, 3(6):639–645, dec 2013. ISSN 18796257. doi: 10.1016/j.coviro.2013.09.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S1879625713001533>.
- Ornella Affinito, Giovanni Scala, Domenico Palumbo, Ermanno Florio, Antonella Monticelli, Gennaro Miele, Vittorio Enrico Avvedimento, Alessandro Usiello, Lorenzo Chiariotti, and Sergio Cocozza. Modeling DNA methylation by analyzing the individual configurations of single molecules. *Epigenetics*, 11(12):881–888, dec 2016. ISSN 1559-2308. doi: 10.1080/15592294.2016.1246108. URL <http://www.ncbi.nlm.nih.gov/pubmed/27748645>.
- Haley M. Amemiya, Anshul Kundaje, and Alan P. Boyle. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports*, 9(1):9354, dec 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-45839-z. URL <http://www.nature.com/articles/s41598-019-45839-z>.
- S Andrews. FastQC, 2010. URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Shan V Andrews, Christine Ladd-Acosta, Andrew P Feinberg, Kasper D Hansen, and M Daniele Fallin. “Gap hunting” to characterize clustered probe signals in Illumina methylation array data. *Epigenetics & Chromatin*, 9(1):56, dec 2016. ISSN 1756-8935. doi: 10.1186/s13072-016-0107-z. URL <http://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/s13072-016-0107-z>.
- Maria Giulia Bacalini, Alessio Boattini, Davide Gentilini, Enrico Giampieri, Chiara Pirazzini, Cristina Giuliani, Elisa Fontanesi, Daniel Remondini, Miriam Capri, Alberto Del Rio, Donata Luiselli, Giovanni Vitale, Daniela Mari, Gastone Castellani, Anna Maria Di Blasio, Stefano Salvioli, Claudio Franceschi, and Paolo Garagnani. A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data. *Aging*, 7(2):97–109, jan 2015. ISSN 1945-4589. doi: 10.18632/aging.100718. URL <http://www.aging-us.com/article/100718>.
- Kelly M. Bakulski, Jason I. Feinberg, Shan V. Andrews, Jack Yang, Shannon Brown, Stephanie L. McKenney, Frank Witter, Jeremy Walston, Andrew P. Feinberg, and M. Daniele Fallin. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics*, 11(5):354–362, may 2016. ISSN 1559-2294. doi: 10.1080/15592294.2016.1161875. URL <http://dx.doi.org/10.1080/15592294.2016.1161875>.

- Andrew J Bannister and Tony Kouzarides. Regulation of chromatin by histone modifications. *Cell Research*, 21(3):381–395, mar 2011. ISSN 1001-0602. doi: 10.1038/cr.2011.22. URL <http://www.ncbi.nlm.nih.gov/pubmed/21321607>.
- Monika Bauden, Theresa Kristl, Roland Andersson, György Marko-Varga, and Daniel Ansari. Characterization of histone-related chemical modifications in formalin-fixed paraffin-embedded and fresh-frozen human pancreatic cancer xenografts using LC-MS/MS. *Laboratory Investigation*, 97(3):279–288, mar 2017. ISSN 0023-6837. doi: 10.1038/labinvest.2016.134. URL <http://www.nature.com/doifinder/10.1038/labinvest.2016.134>.
- Victoria P. Belancio, Dale J. Hedges, and Prescott Deininger. Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. *Genome Research*, 18(3):343–358, feb 2008. ISSN 1088-9051. doi: 10.1101/gr.5558208. URL <http://www.genome.org/cgi/doi/10.1101/gr.5558208>.
- Christopher G. Bell, Gareth A. Wilson, Lee M. Butcher, Christian Roos, Lutz Walter, and Stephan Beck. Human-specific CpG "beacons" identify loci associated with human-specific traits and disease. *Epigenetics*, 7(10):1188–99, oct 2012a. ISSN 1559-2308. doi: 10.4161/epi.22127. URL <http://www.ncbi.nlm.nih.gov/pubmed/22968434>.
- Christopher G. Bell, Yudong Xia, Wei Yuan, Fei Gao, Kirsten Ward, Leonie Roos, Massimo Mangino, Pirro G. Hysi, Jordana Bell, Jun Wang, and Timothy D. Spector. Novel regional age-associated DNA methylation changes within human common disease-associated loci. *Genome Biology*, 17(1):193, dec 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-1051-8. URL <http://www.ncbi.nlm.nih.gov/pubmed/27663977>.
- Christopher G. Bell, Fei Gao, Wei Yuan, Leonie Roos, Richard J. Acton, Yudong Xia, Jordana Bell, Kirsten Ward, Massimo Mangino, Pirro G. Hysi, Jun Wang, and Timothy D. Spector. Obligatory and facilitative allelic variation in the DNA methylome within common disease-associated loci. *Nature Communications*, 9(1):8, dec 2018. ISSN 2041-1723. doi: 10.1038/s41467-017-01586-1. URL <http://www.nature.com/articles/s41467-017-01586-1>.
- Christopher G Bell, Robert Lowe, Peter D Adams, Andrea A Baccarelli, Stephan Beck, Jordana T Bell, Brock C. Christensen, Vadim N. Gladyshev, Bastiaan T. Heijmans, Steve Horvath, Trey Ideker, Jean-Pierre J. Issa, Karl T Kelsey, Riccardo E. Marioni, Wolf Reik, Caroline L Relton, Leonard C Schalkwyk, Andrew E Teschendorff, Wolfgang Wagner, Kang Zhang, and Vardhman K Rakyan. DNA methylation aging clocks: challenges and recommendations. *Genome biology*, 20(1):249, nov 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1824-y. URL <http://www.ncbi.nlm.nih.gov/pubmed/31767039>.
- Jordana T. Bell, Pei-Chien Tsai, Tsun-Po Yang, Ruth Pidsley, James Nisbet, Daniel Glass, Massimo Mangino, Guangju Zhai, Feng Zhang, Ana Valdes, So-Youn Shin, Emma L. Dempster, Robin M. Murray, Elin Grundberg, Asa K. Hedman, Alexandra Nica, Kerrin S. Small, MuTHER Consortium, Emmanouil T. Dermitzakis, Mark I. McCarthy, Jonathan Mill, Tim D. Spector, and Panos Deloukas. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS genetics*, 8(4):e1002629, apr 2012b. ISSN 1553-7404. doi: 10.1371/journal.pgen.1002629. URL <http://www.ncbi.nlm.nih.gov/pubmed/22532803>.
- Joshua S. K. Bell and Paula M. Vertino. Orphan CpG islands define a novel class of highly active enhancers. *Epigenetics*, 12(6):449–464, jun 2017. ISSN 1559-2308. doi: 10.1080/15592294.2017.1297910. URL <http://www.ncbi.nlm.nih.gov/pubmed/28448736>.
- Miles C Benton, Heidi G Sutherland, Donia Macartney-Coxson, Larisa M Haupt, Rodney A. Lea, and Lyn R Griffiths. Methylome-wide association study of whole blood DNA in the Norfolk Island isolate identifies robust loci associated with age. *Aging*, 9(3):753–768, feb 2017. ISSN 1945-4589. doi: 10.18632/aging.101187. URL <http://www.aging-us.com/article/101187/text>.
- G D Berdyshev, G K Korotaev, G V Boiarskikh, and B F Vaniushin. [Nucleotide composition of DNA and RNA from somatic tissues of humpback and its changes during spawning]. *Biokhimija (Moscow, Russia)*, 32(5):988–93, 1967. ISSN 0320-9725. URL <http://www.ncbi.nlm.nih.gov/pubmed/5628601>.

- Daniel Besser, Frank Götz, Kai Schulze-Forster, Herbert Wagner, Hans Kröger, and Dietrich Simon. DNA methylation inhibits transcription by RNA polymerase III of a tRNA gene, but not of a 5S rRNA gene. *FEBS letters*, 269(2):358–62, sep 1990. ISSN 0014-5793. doi: 10.1016/0014-5793(90)81193-R. URL <http://www.ncbi.nlm.nih.gov/pubmed/2401361>.
- Eva Bianconi, Allison Piovesan, Federica Facchin, Alina Beraudi, Raffaella Casadei, Flavia Fabbetti, Lorenza Vitale, Maria Chiara Pelleri, Simone Tassani, Francesco Piva, Soledad Perez-Amodio, Pierluigi Strippoli, and Silvia Canaider. An estimation of the number of cells in the human body. *Annals of human biology*, 40(6):463–71, nov 2013. ISSN 1464-5033. doi: 10.3109/03014460.2013.807878. URL <http://www.ncbi.nlm.nih.gov/pubmed/23829164>.
- Marina Bibikova. High-throughput DNA methylation profiling using universal bead arrays. *Genome Research*, 16(3):383–393, mar 2006. ISSN 1088-9051. doi: 10.1101/gr.4410706. URL <http://www.genome.org/cgi/doi/10.1101/gr.4410706>.
- Marina Bibikova, Jennie Le, Bret Barnes, Shadi Saedinia-Melnyk, Lixin Zhou, Richard Shen, and Kevin L Gunderson. Genome-wide DNA methylation profiling using Infinium ® assay. *Epigenomics*, 1(1):177–200, oct 2009. ISSN 1750-1911. doi: 10.2217/epi.09.14. URL <https://www.futuremedicine.com/doi/10.2217/epi.09.14>.
- Marina Bibikova, Bret Barnes, Chan Tsan, Vincent Ho, Brandy Klotzle, Jennie M. Le, David Delano, Lu Zhang, Gary P. Schroth, Kevin L. Gunderson, Jian-Bing Fan, and Richard Shen. High density DNA methylation array with single CpG site resolution. *Genomics*, 98(4):288–95, oct 2011. ISSN 1089-8646. doi: 10.1016/j.ygeno.2011.07.007. URL <http://www.ncbi.nlm.nih.gov/pubmed/21839163>.
- Adrian Bird. DNA methylation patterns and epigenetic memory. *Genes & development*, 16(1): 6–21, jan 2002. ISSN 0890-9369. doi: 10.1101/gad.947102. URL <http://www.ncbi.nlm.nih.gov/pubmed/11782440>.
- Adrian Bird. Perceptions of epigenetics. *Nature*, 447(7143):396–8, may 2007. ISSN 1476-4687. doi: 10.1038/nature05913. URL <http://www.ncbi.nlm.nih.gov/pubmed/17522671>.
- Hans T. Bjornsson. Intra-individual Change Over Time in DNA Methylation With Familial Clustering. *JAMA*, 299(24):2877, jun 2008. ISSN 0098-7484. doi: 10.1001/jama.299.24.2877. URL <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.299.24.2877>.
- Adam Blattler, Lijing Yao, Heather Witt, Yu Guo, Charles M Nicolet, Benjamin P Berman, and Peggy J Farnham. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biology*, 15(9):469, sep 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0469-0. URL <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0469-0>.
- Christoph Bock, Eleni M Tomazou, Arie B Brinkman, Fabian Müller, Femke Simmer, Hongcang Gu, Natalie Jäger, Andreas Gnirke, Hendrik G Stunnenberg, and Alexander Meissner. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nature Biotechnology*, 28(10):1106–1114, oct 2010. ISSN 1087-0156. doi: 10.1038/nbt.1681. URL <http://www.ncbi.nlm.nih.gov/pubmed/20852634>.
- Sven Bocklandt, Wen Lin, Mary E. Sehl, Francisco J. Sánchez, Janet S. Sinsheimer, Steve Horvath, and Eric Vilain. Epigenetic Predictor of Age. *PLoS ONE*, 6(6):e14821, jun 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0014821. URL <https://dx.plos.org/10.1371/journal.pone.0014821>.
- J. Bohlin, S. E. Håberg, P. Magnus, S. E. Reese, H. K. Gjessing, M. C. Magnus, C. L. Parr, C. M. Page, S. J. London, and W. Nystad. Prediction of gestational age based on genome-wide differentially methylated regions. *Genome Biology*, 17(1):207, dec 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-1063-4. URL <http://dx.doi.org/10.1186/s13059-016-1063-4>.

- Marco P. Boks, Eske M. Derkx, Daniel J. Weisenberger, Erik Strengman, Esther Janson, Iris E. Sommer, René S. Kahn, and Roel A. Ophoff. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *Plos one*, 4(8):e6767, aug 2009. ISSN 1932-6203. doi: 10.1371/journal.pone.0006767. URL <http://www.ncbi.nlm.nih.gov/pubmed/19774229>.
- Valentina Bollati, Joel Schwartz, Robert Wright, Augusto Litonjua, Letizia Tarantini, Helen Suh, David Sparrow, Pantel Vokonas, and Andrea Baccarelli. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mechanisms of Ageing and Development*, 130(4):234–239, apr 2009. ISSN 00476374. doi: 10.1016/j.mad.2008.12.003. URL <https://linkinghub.elsevier.com/retrieve/pii/S0047637408002820>.
- Lauren N. Booth and Anne Brunet. The Aging Epigenome. *Molecular Cell*, 62(5):728–744, jun 2016. ISSN 10972765. doi: 10.1016/j.molcel.2016.05.013. URL <http://dx.doi.org/10.1016/j.molcel.2016.05.013>.
- M. Bostick, J. K. Kim, P.-O. Esteve, A. Clark, S. Pradhan, and S. E. Jacobsen. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science*, 317(5845):1760–1764, sep 2007. ISSN 0036-8075. doi: 10.1126/science.1147939. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1147939>.
- Carrie V. Breton, Carmen J. Marsit, Elaine Faustman, Kari Nadeau, Jaclyn M. Goodrich, Dana C. Dolinoy, Julie Herbster, Nina Holland, Janine M. LaSalle, Rebecca Schmidt, Paul Yousefi, Frederica Perera, Bonnie R. Joubert, Joseph Wiemels, Michele Taylor, Ivana V. Yang, Rui Chen, Kinjal M. Hew, Deborah M. Hussey Freeland, Rachel Miller, and Susan K. Murphy. Small-Magnitude Effect Sizes in Epigenetic End Points are Important in Children’s Environmental Health Studies: The Children’s Environmental Health and Disease Prevention Research Center’s Epigenetics Working Group. *Environmental Health Perspectives*, 125(4):511–526, apr 2017. ISSN 0091-6765. doi: 10.1289/EHP595. URL <https://ehp.niehs.nih.gov/doi/10.1289/EHP595>.
- Adrian W. Briggs, Udo Stenzel, Matthias Meyer, Johannes Krause, Martin Kircher, and Svante Pääbo. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic acids research*, 38(6):e87, apr 2010. ISSN 1362-4962. doi: 10.1093/nar/gkp1163. URL <http://www.ncbi.nlm.nih.gov/pubmed/20028723>.
- Graham C. Burdge, Mark A. Hanson, Jo L. Slater-Jeffries, and Karen A. Lillycrop. Epigenetic regulation of transcription: a mechanism for inducing variations in phenotype (fetal programming) by differences in nutrition during early life? *British Journal of Nutrition*, 97(6):1036–1046, jun 2007a. ISSN 0007-1145. doi: 10.1017/S0007114507682920. URL [https://www.cambridge.org/core/product/identifier/S0007114507682920/type/journal{\\_\\_}article](https://www.cambridge.org/core/product/identifier/S0007114507682920/type/journal{__}article).
- Graham C Burdge, Jo Slater-Jeffries, Christopher Torrens, Emma S Phillips, Mark A. Hanson, and Karen A. Lillycrop. Dietary protein restriction of pregnant rats in the F 0 generation induces altered methylation of hepatic gene promoters in the adult male offspring in the F 1 and F 2 generations. *British Journal of Nutrition*, 97(3):435–439, mar 2007b. ISSN 0007-1145. doi: 10.1017/S0007114507352392. URL [https://www.cambridge.org/core/product/identifier/S0007114507352392/type/journal{\\_\\_}article](https://www.cambridge.org/core/product/identifier/S0007114507352392/type/journal{__}article).
- Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O’Connell, Adrian Cortes, Samantha Welsh, Gil McVean, Stephen Leslie, Peter Donnelly, and Jonathan Marchini. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*, page 166298, 2017. doi: 10.1101/166298. URL <https://www.biorxiv.org/content/10.1101/166298v1>.
- Judith Campisi, Pankaj Kapahi, Gordon J. Lithgow, Simon Melov, John C. Newman, and Eric Verdin. From discoveries in ageing research to therapeutics for healthy ageing. *Nature*, 571(7764):183–192, jul 2019. ISSN 0028-0836. doi: 10.1038/s41586-019-1365-2. URL <http://www.nature.com/articles/s41586-019-1365-2>.

- Donatella Canella, Viviane Praz, Jaime H. Reina, Pascal Cousin, and Nouria Hernandez. Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells. *Genome research*, 20(6):710–21, jun 2010. ISSN 1549-5469. doi: 10.1101/gr.101337.109. URL <http://www.ncbi.nlm.nih.gov/pubmed/20413673>.
- Andres Cardenas, Catherine Allard, Myriam Doyon, E Andres Houseman, Kelly M Bakulski, Patrice Perron, Luigi Bouchard, and Marie-France Hivert. Validation of a DNA methylation reference panel for the estimation of nucleated cells types in cord blood. *Epigenetics*, 11(11):773–779, nov 2016. ISSN 1559-2308. doi: 10.1080/15592294.2016.1233091. URL <http://www.ncbi.nlm.nih.gov/pubmed/27668573>.
- Enrique Carrillo-de Santa-Pau, David Juan, Vera Pancaldi, Felipe Were, Ignacio Martin-Subero, Daniel Rico, and Alfonso Valencia. Automatic identification of informative regions with epigenomic changes associated to hematopoiesis. *Nucleic Acids Research*, 45(16):9244–9259, sep 2017. ISSN 0305-1048. doi: 10.1093/nar/gkx618. URL <https://academic.oup.com/nar/article/45/16/9244/3976483>.
- CellOntology. Cell Ontology, 2017. URL <http://obofoundry.org/ontology/cl.html>.
- CellSystemsVoices. What Is Your Conceptual Definition of “Cell Type” in the Context of a Mature Organism? *Cell Systems*, 4(3):255–259, mar 2017. ISSN 24054712. doi: 10.1016/j.cels.2017.03.006. URL <https://linkinghub.elsevier.com/retrieve/pii/S2405471217300911>.
- Andrea Cerase, Greta Pintacuda, Anna Tattermusch, and Philip Avner. Xist localization and function: new insights from multiple levels. *Genome Biology*, 16(1):166, dec 2015. ISSN 1474-760X. doi: 10.1186/s13059-015-0733-y. URL <http://genomebiology.com/2015/16/1/166>.
- Patricia P. Chan and Todd M. Lowe. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic acids research*, 37(Database issue):D93–7, jan 2009. ISSN 1362-4962. doi: 10.1093/nar/gkn787. URL <http://www.ncbi.nlm.nih.gov/pubmed/18984615>.
- Brian H Chen, Riccardo E Marioni, Elena Colicino, Marjolein J Peters, Cavin K Ward-Caviness, Pei-Chien Tsai, Nicholas S Roetker, Allan C Just, Ellen W Demerath, Weihua Guan, Jan Bressler, Myriam Fornage, Stephanie Studenski, Amy R Vandiver, Ann Zeno-bia Moore, Toshiko Tanaka, Douglas P Kiel, Liming Liang, Pantel Vokonas, Joel Schwartz, Kathryn L Lunetta, Joanne M Murabito, Stefania Bandinelli, Dena G Hernandez, David Melzer, Michael Nalls, Luke C Pilling, Timothy R Price, Andrew B Singleton, Christian Gieger, Rolf Holle, Anja Kretschmer, Florian Kronenberg, Sonja Kunze, Jakob Linseisen, Christine Meisinger, Wolfgang Rathmann, Melanie Waldenberger, Peter M Visscher, Sonia Shah, Naomi R Wray, Allan F McRae, Oscar H Franco, Albert Hofman, André G Uitter-linden, Devin Absher, Themistocles Assimes, Morgan E Levine, Ake T Lu, Philip S Tsao, Lifang Hou, Joann E Manson, Cara L Carty, Andrea Z LaCroix, Alexander P Reiner, Tim D Spector, Andrew P Feinberg, Daniel Levy, Andrea Baccarelli, Joyce van Meurs, Jordana T Bell, Annette Peters, Ian J Deary, James S Pankow, Luigi Ferrucci, and Steve Horvath. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging*, 8(9):1844–1865, sep 2016. ISSN 1945-4589. doi: 10.18632/aging.101020. URL <http://www.ncbi.nlm.nih.gov/pubmed/27690265>.
- Ling-Ling Chen and Li Yang. ALU ternative Regulation for Gene Expression. *Trends in Cell Biology*, 27(7):480–490, jul 2017. ISSN 09628924. doi: 10.1016/j.tcb.2017.01.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S0962892417300028>.
- Brock C. Christensen, E. Andres Houseman, Carmen J. Marsit, Shichun Zheng, Margaret R. Wrensch, Joseph L. Wiemels, Heather H. Nelson, Margaret R. Karagas, James F. Padbury, Raphael Bueno, David J. Sugarbaker, Ru-Fang Yeh, John K. Wiencke, and Karl T. Kelsey. Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context. *PLoS Genetics*, 5(8):e1000602, aug 2009. ISSN 1553-7404. doi: 10.1371/journal.pgen.1000602. URL <https://dx.plos.org/10.1371/journal.pgen.1000602>.

- Lene Christiansen, Adam Lenart, Qihua Tan, James W. Vaupel, Abraham Aviv, Matt McGue, and Kaare Christensen. DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell*, 15(1):149–154, feb 2016. ISSN 14749718. doi: 10.1111/acel.12421. URL <http://doi.wiley.com/10.1111/acel.12421>.
- Edward B. Chuong, Nels C. Elde, and Cédric Feschotte. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics*, 18(2):71–86, feb 2017. ISSN 1471-0056. doi: 10.1038/nrg.2016.139. URL <http://www.nature.com/articles/nrg.2016.139>.
- Annie Vogel Ciernia and Janine LaSalle. The landscape of DNA methylation amid a perfect storm of autism aetiologies. *Nature Reviews Neuroscience*, 17(7):411–423, jul 2016. ISSN 1471-003X. doi: 10.1038/nrn.2016.41. URL <http://dx.doi.org/10.1038/nrn.2016.41>.
- Christine Clark, Priit Palta, Christopher J. Joyce, Carol Scott, Elin Grundberg, Panos Deloukas, Aarno Palotie, and Alison J. Coffey. A Comparison of the Whole Genome Approach of MeDIP-Seq to the Targeted Approach of the Infinium HumanMethylation450 Bead-Chip® for Methylome Profiling. *PLoS ONE*, 7(11):e50233, nov 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0050233. URL <https://dx.plos.org/10.1371/journal.pone.0050233>.
- Jacob Cohen. *Statisticsl Power Analysis for the Behavioural Sciences*. Lawrence Erlbaum Associates, New York, 2 edition, 1988. ISBN 0805802835.
- John Joseph Cole, Neil A. Robertson, Mohammed Iqbal Rather, John P. Thomson, Tony McBryan, Duncan Sproul, Tina Wang, Claire Brock, William Clark, Trey Ideker, Richard R. Meehan, Richard A. Miller, Holly M. Brown-Borg, and Peter D Adams. Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biology*, 18(1):58, dec 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1185-3. URL <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1185-3>.
- Cyrus Cooper, Nicholas C. Harvey, Nicholas J. Bishop, Stephen Kennedy, Aris T. Papageorghiou, Inez Schoenmakers, Robert Fraser, Saurabh V. Gandhi, Andrew Carr, Stefania D’Angelo, Sarah R. Crozier, Rebecca J. Moon, Nigel K. Arden, Elaine M. Dennison, Keith M. Godfrey, Hazel M. Inskip, Ann Prentice, M. Zulf Mughal, Richard Eastell, David M. Reid, and M. Kasim Javaid. Maternal gestational vitamin D supplementation and offspring bone health (MAVIDOS): a multicentre, double-blind, randomised placebo-controlled trial. *The Lancet Diabetes & Endocrinology*, 4(5):393–402, may 2016. ISSN 22138587. doi: 10.1016/S2213-8587(16)00044-9. URL [http://dx.doi.org/10.1016/S2213-8587\(16\)00044-9](http://dx.doi.org/10.1016/S2213-8587(16)00044-9).
- Richard Cordaux and Mark A. Batzer. The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics*, 10(10):691–703, oct 2009. ISSN 1471-0056. doi: 10.1038/nrg2640. URL <http://www.nature.com/articles/nrg2640>.
- Diane Crighton. p53 represses RNA polymerase III transcription by targeting TBP and inhibiting promoter occupancy by TFIIIB. *The EMBO Journal*, 22(11):2810–2820, jun 2003. ISSN 14602075. doi: 10.1093/emboj/cdg265. URL <http://emboj.embopress.org/cgi/doi/10.1093/emboj/cdg265>.
- Hazel A Cruickshanks, Tony McBryan, David M Nelson, Nathan D. VanderKraats, Parisha P Shah, John van Tuyn, Taranjit Singh Rai, Claire Brock, Greg Donahue, Donncha S Duncan, Mark E Drotar, Richard R Meehan, John R Edwards, Shelley L Berger, and Peter D Adams. Senescent cells harbour features of the cancer epigenome. *Nature Cell Biology*, 15(12):1495–1506, dec 2013. ISSN 1465-7392. doi: 10.1038/ncb2879. URL <http://www.ncbi.nlm.nih.gov/pubmed/24270890>.
- Elizabeth M. Curtis, Robert Murray, Philip Titcombe, Eloïse Cook, Rebecca Clarke-Harris, Paula Costello, Emma Garratt, Joanna D. Holbrook, Sheila Barton, Hazel Inskip, Keith M. Godfrey, Christopher G. Bell, Cyrus Cooper, Karen A. Lillycrop, and Nicholas C. Harvey. Perinatal DNA Methylation at CDKN2A Is Associated With Offspring Bone Mass: Findings From the Southampton Women’s Survey. *Journal of Bone and Mineral Research*, 32(10):2030–2040, oct 2017. ISSN 08840431. doi: 10.1002/jbmr.3153. URL <http://doi.wiley.com/10.1002/jbmr.3153>.

- Emily M. Darrow and Brian P. Chadwick. A novel tRNA variable number tandem repeat at human chromosome 1q23.3 is implicated as a boundary element based on conservation of a CTCF motif in mouse. *Nucleic Acids Research*, 42(10):6421–6435, jun 2014. ISSN 1362-4962. doi: 10.1093/nar/gku280. URL <http://academic.oup.com/nar/article/42/10/6421/2435926/A-novel-tRNA-variable-number-tandem-repeat-at>.
- Olivia M. de Goede, Hamid R. Razzaghian, E. Magda Price, Meaghan J. Jones, Michael S. Kobor, Wendy P. Robinson, and Pascal M. Lavoie. Nucleated red blood cells impact DNA methylation and expression analyses of cord blood hematopoietic cells. *Clinical Epigenetics*, 7(1):95, dec 2015. ISSN 1868-7075. doi: 10.1186/s13148-015-0129-6. URL <http://dx.doi.org/10.1186/s13148-015-0129-6>.
- Aimée M. Deaton and Adrian Bird. CpG islands and the regulation of transcription. *Genes & development*, 25(10):1010–22, may 2011. doi: 10.1101/gad.2037511. URL <http://www.ncbi.nlm.nih.gov/pubmed/21576262>.
- Prescott Deininger. Alu elements: know the SINEs. *Genome Biology*, 12(12):236, 2011. ISSN 1465-6906. doi: 10.1186/gb-2011-12-12-236. URL <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-12-236>.
- Özgen Deniz, Jennifer M. Frost, and Miguel R. Branco. Regulation of transposable elements by DNA modifications. *Nature Reviews Genetics*, 20(7):417–431, 2019. ISSN 14710064. doi: 10.1038/s41576-019-0106-6.
- Thomas Derrien, Jordi Estellé, Santiago Marco Sola, David G. Knowles, Emanuele Raineri, Roderic Guigó, and Paolo Ribeca. Fast Computation and Applications of Genome Mappability. *PLoS ONE*, 7(1):e30377, jan 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0030377. URL <https://dx.plos.org/10.1371/journal.pone.0030377>.
- Marie Dewannieux, Cécile Esnault, and Thierry Heidmann. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics*, 35(1):41–48, sep 2003. ISSN 1061-4036. doi: 10.1038/ng1223. URL <http://www.nature.com/articles/ng1223>.
- Joseph M Dhahbi, Stephen R Spindler, Hani Atamna, Amy Yamakawa, Dario Boffelli, Patricia Mote, and David IK Martin. 5' tRNA halves are present as abundant complexes in serum, concentrated in blood cells, and modulated by aging and calorie restriction. *BMC Genomics*, 14(1):298, 2013. ISSN 1471-2164. doi: 10.1186/1471-2164-14-298. URL <http://www.ncbi.nlm.nih.gov/pubmed/23638709>.
- Kimberly A. Dittmar, Jeffrey M. Goodenbour, and Tao Pan. Tissue-specific differences in human transfer RNA expression. *PLoS genetics*, 2(12):e221, dec 2006. ISSN 1553-7404. doi: 10.1371/journal.pgen.0020221. URL <http://www.ncbi.nlm.nih.gov/pubmed/17194224>.
- Meeshanthini V. Dogan, Steven R. H. Beach, and Robert A. Philibert. Genetically contextual effects of smoking on genome wide DNA methylation. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 174(6):595–607, sep 2017. ISSN 15524841. doi: 10.1002/ajmg.b.32565. URL <http://doi.wiley.com/10.1002/ajmg.b.32565>.
- Thomas A Down, Vardhman K Rakyan, Daniel J Turner, Paul Flicek, Heng Li, Eugene Kulesha, Stefan Gräf, Nathan Johnson, Javier Herrero, Eleni M Tomazou, Natalie P Thorne, Liselotte Bäckdahl, Marlis Herberth, Kevin L Howe, David K Jackson, Marcos M Miretti, John C Marioni, Ewan Birney, Tim J P Hubbard, Richard Durbin, Simon Tavaré, and Stephan Beck. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnology*, 26(7):779–785, jul 2008. ISSN 1087-0156. doi: 10.1038/nbt1414. URL <http://www.nature.com/articles/nbt1414>.
- Qian Du, Saul A. Bert, Nicola J. Armstrong, C. Elizabeth Caldon, Jenny Z. Song, Shalima S. Nair, Cathryn M. Gould, Phuc-Loi Luu, Timothy Peters, Amanda Khoury, Wenjia Qu, Elena Zotenko, Clare Stirzaker, and Susan J. Clark. Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nature Communications*, 10(1):416, dec 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-08302-1. URL <http://dx.doi.org/10.1038/s41467-019-08302-1>.

B K Duncan and J H Miller. Mutagenic deamination of cytosine residues in DNA. *Nature*, 287(5782):560–1, oct 1980. ISSN 0028-0836. URL <http://www.ncbi.nlm.nih.gov/pubmed/6999365>.

Florian Eckhardt, Joern Lewin, Rene Cortese, Vardhman K Rakyan, John Attwood, Matthias Burger, John Burton, Tony V Cox, Rob Davies, Thomas A Down, Carolina Haefliger, Roger Horton, Kevin Howe, David K Jackson, Jan Kunde, Christoph Koenig, Jennifer Liddle, David Niblett, Thomas Otto, Roger Pettett, Stefanie Seemann, Christian Thompson, Tony West, Jane Rogers, Alex Olek, Kurt Berlin, and Stephan Beck. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature Genetics*, 38(12):1378–1385, dec 2006. ISSN 1061-4036. doi: 10.1038/ng1909. URL <http://www.nature.com/doifinder/10.1038/ng1909>.

Melanie Ehrlich. DNA hypermethylation in disease: mechanisms and clinical relevance. *Epidemiology*, 14(12):1141–1163, dec 2019. ISSN 1559-2308. doi: 10.1080/15592294.2019.1638701. URL <http://www.ncbi.nlm.nih.gov/pubmed/31284823>.

Melanie Ehrlich, Miguel A. Gama-Sosa, Lan-Hsiang Huang, Rose Marie Midgett, Kenneth C. Kuo, Roy A. McCune, and Charles Gehrke. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Research*, 10(8):2709–2721, 1982. ISSN 0305-1048. doi: 10.1093/nar/10.8.2709. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/10.8.2709>.

Manfred Eigen, B. Lindemann, Manfred Tietze, R Winkler-Oswatitsch, Andreas Dress, and A von Haeseler. How old is the genetic code? Statistical geometry of tRNA provides an answer. *Science*, 244(4905):673–679, may 1989. ISSN 0036-8075. doi: 10.1126/science.2497522. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.2497522>.

Jason Ernst, Pouya Kheradpour, Tarjei S Mikkelsen, Noam Shores, Lucas D Ward, Charles B Epstein, Xiaolan Zhang, Li Wang, Robbyn Issner, Michael Coyne, Manching Ku, Timothy Durham, Manolis Kellis, and Bradley E Bernstein. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49, may 2011. ISSN 0028-0836. doi: 10.1038/nature09906. URL <http://www.nature.com/articles/nature09906>.

M Esteller, J M Silva, G Dominguez, F Bonilla, X Matias-Guiu, E Lerma, E Bussaglia, J Prat, I C Harkes, E A Repasky, E Gabrielson, M Schutte, S B Baylin, and J G Herman. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *Journal of the National Cancer Institute*, 92(7):564–9, apr 2000. ISSN 0027-8874. doi: 10.1093/jnci/92.7.564. URL <http://www.ncbi.nlm.nih.gov/pubmed/10749912>.

Pierre-Olivier Estève, Hang Gyeong Chin, Andrea Smallwood, George R. Feehery, Omkaram Gangisetty, Adam R. Karpf, Michael F. Carey, and Sriharsa Pradhan. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes & development*, 20(22):3089–103, nov 2006. ISSN 0890-9369. doi: 10.1101/gad.1463706. URL <http://www.ncbi.nlm.nih.gov/pubmed/17085482>.

Philip Ewels, Måns Magnusson, Sverker Lundin, and Max Käller. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)*, 32(19):3047–8, 2016. ISSN 1367-4811. doi: 10.1093/bioinformatics/btw354. URL <http://www.ncbi.nlm.nih.gov/pubmed/27312411>.

Andrew P. Feinberg. The Key Role of Epigenetics in Human Disease Prevention and Mitigation. *New England Journal of Medicine*, 378(14):1323–1334, apr 2018. ISSN 0028-4793. doi: 10.1056/NEJMra1402513. URL <http://www.ncbi.nlm.nih.gov/pubmed/29617578>.

Andrew P Feinberg and Benjamin Tycko. The history of cancer epigenetics. *Nature Reviews Cancer*, 4(2):143–153, feb 2004. ISSN 1474-175X. doi: 10.1038/nrc1279. URL <http://www.nature.com/articles/nrc1279>.

Roberto Ferrari, Lara Isabel de Llobet Cucalon, Chiara Di Vona, François Le Dilly, Enrique Vidal, Antonios Lioutas, Javier Quilez Oliete, Laura Jochem, Erin Cutts, Giorgio Dieci, Alessandro Vannini, Martin Teichmann, Susana de la Luna, and Miguel Beato. TFIIIC Binding

- to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation. *Molecular Cell*, pages 1–13, nov 2019. ISSN 10972765. doi: 10.1016/j.molcel.2019.10.020. URL <https://linkinghub.elsevier.com/retrieve/pii/S1097276519307981>.
- Adam E. Field, Neil A. Robertson, Tina Wang, Aaron Havas, Trey Ideker, and Peter D. Adams. DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. *Molecular Cell*, 71(6):882–895, 2018. ISSN 10974164. doi: 10.1016/j.molcel.2018.08.008. URL <https://doi.org/10.1016/j.molcel.2018.08.008>.
- Danny Filer, Maximillian A. Thompson, Vakil Takhayev, Adam J. Dobson, Ilektra Kotronaki, James W. M. Green, Matthias Heinemann, Jennifer M. A. Tullet, and Nazif Alic. RNA polymerase III limits longevity downstream of TORC1. *Nature*, 552(7684):263–267, dec 2017. ISSN 0028-0836. doi: 10.1038/nature25007. URL <http://www.nature.com/articles/nature25007>.
- Ines Florath, Katja Butterbach, H. Muller, M. Bewerunge-Hudler, and Hermann Brenner. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Human Molecular Genetics*, 23(5):1186–1201, mar 2014. ISSN 0964-6906. doi: 10.1093/hmg/ddt531. URL <https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddt531>.
- Jean-Philippe Fortin, Aurélie Labbe, Mathieu Lemire, Brent W. Zanke, Thomas J. Hudson, Elana J. Fertig, Celia M.T. Greenwood, and Kasper D. Hansen. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology*, 15(11):503, nov 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0503-2. URL <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0503-2>.
- Mario F Fraga, E. Ballestar, Maria F Paz, Santiago Ropero, Fernando Setien, Maria L Ballestar, D. Heine-Suner, Juan C Cigudosa, Miguel Urioste, Javier Benitez, Manuel Boix-Chornet, Abel Sanchez-Aguilera, Charlotte Ling, Emma Carlsson, Pernille Poulsen, Allan Vaag, Zarko Stephan, Tim D Spector, Y.-Z. Wu, Christoph Plass, and Manel Esteller. From The Cover: Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences*, 102(30):10604–10609, jul 2005. ISSN 0027-8424. doi: 10.1073/pnas.0500398102. URL <http://www.ncbi.nlm.nih.gov/pubmed/16009939>.
- Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 2010. ISSN 1548-7660. doi: 10.18637/jss.v033.i01. URL <http://www.jstatsoft.org/v33/i01/>.
- Miguel A. Gama-Sosa, Valerie A. Slagel, Ronald W. Trewyn, Ronald Oxenhandler, Kenneth C. Kuo, Charles W. Gehrke, and Melanie Ehrlich. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Research*, 11(19):6883–6894, 1983. ISSN 0305-1048. doi: 10.1093/nar/11.19.6883. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/11.19.6883>.
- Paolo Garagnani, Maria G. Bacalini, Chiara Pirazzini, Davide Gori, Cristina Giuliani, Daniela Mari, Anna M. Di Blasio, Davide Gentilini, Giovanni Vitale, Sebastiano Collino, Serge Rezzi, Gastone Castellani, Miriam Capri, Stefano Salvioli, and Claudio Franceschi. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging cell*, 11(6):1132–4, dec 2012. ISSN 1474-9726. doi: 10.1111/acel.12005. URL <http://www.ncbi.nlm.nih.gov/pubmed/23061750>.
- Hartmut Geiger, Gerald de Haan, and M. Carolina Florian. The ageing haematopoietic stem cell compartment. *Nature Reviews Immunology*, 13(5):376–389, may 2013. ISSN 1474-1733. doi: 10.1038/nri3433. URL <http://dx.doi.org/10.1038/nri3433>.
- Kristina Gervin, Christian Magnus Page, Hans Christian D. Aass, Michelle A. Jansen, Heidi Elisabeth Fjeldstad, Bettina Kulle Andreassen, Liesbeth Duijts, Joyce B. van Meurs, Menno C. van Zelm, Vincent W. Jaddoe, Hedvig Nordeng, Gunn Peggy Knudsen, Per Magnus, Wenche Nystad, Anne Cathrine Staff, Janine F. Felix, and Robert Lyle. Cell type specific DNA methylation in cord blood: A 450K-reference data set and cell count-based validation of estimated cell type composition. *Epigenetics*, 11(9):690–698, sep 2016. ISSN 1559-2294. doi: 10.1080/15592294.2016.1214782. URL <http://dx.doi.org/10.1080/15592294.2016.1214782>.

Renaud Geslain and Tao Pan. Functional Analysis of Human tRNA Isodecoders. *Journal of Molecular Biology*, 396(3):821–831, feb 2010. ISSN 00222836. doi: 10.1016/j.jmb.2009.12.018. URL <https://linkinghub.elsevier.com/retrieve/pii/S002228360901523X>.

Jude Gibson, Tom C Russ, Toni-Kim Clarke, David M Howard, Robert F Hillary, Kathryn L Evans, Rosie M Walker, Mairead L Bermingham, Stewart W Morris, Archie Campbell, Caroline Hayward, Alison D Murray, David J Porteous, Steve Horvath, Ake T Lu, Andrew M McIntosh, Heather C Whalley, and Riccardo E Marioni. A meta-analysis of genome-wide association studies of epigenetic age acceleration. *PLOS Genetics*, 15(11):e1008104, nov 2019. ISSN 1553-7404. doi: 10.1371/journal.pgen.1008104. URL <http://www.ncbi.nlm.nih.gov/pubmed/31738745>.

Hila Gingold, Orna Dahan, and Yitzhak Pilpel. Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome. *Nucleic acids research*, 40(20):10053–63, nov 2012. ISSN 1362-4962. doi: 10.1093/nar/gks772. URL <http://www.ncbi.nlm.nih.gov/pubmed/22941644>.

Hila Gingold, Disa Tehler, Nanna R. Christoffersen, Morten M. Nielsen, Fazila Asmar, Susanne M. Kooistra, Nicolaj S. Christophersen, Lise Lotte Christensen, Michael Borre, Karina D. Sørensen, Lars D. Andersen, Claus L. Andersen, Esther Hulleman, Tom Wurdinger, Elisabeth Ralfkiær, Kristian Helin, Kirsten Grønbæk, Torben Ørntoft, Sebastian M. Waszak, Orna Dahan, Jakob Skou Pedersen, Anders H. Lund, and Yitzhak Pilpel. A Dual Program for Translation Regulation in Cellular Proliferation and Differentiation. *Cell*, 158(6):1281–1292, sep 2014. ISSN 00928674. doi: 10.1016/j.cell.2014.08.011. URL <http://www.sciencedirect.com/science/article/pii/S0092867414010423>.

Natividad Gomez-Roman, Carla Grandori, Robert N. Eisenman, and Robert J. White. Direct activation of RNA polymerase III transcription by c-Myc. *Nature*, 421(6920):290–294, jan 2003. ISSN 0028-0836. doi: 10.1038/nature01327. URL <http://www.nature.com/articles/nature01327>.

Jeffrey M. Goodenbour and Tao Pan. Diversity of tRNA genes in eukaryotes. *Nucleic acids research*, 34(21):6137–46, dec 2006. ISSN 1362-4962. doi: 10.1093/nar/gkl725. URL <http://www.ncbi.nlm.nih.gov/pubmed/17088292>.

T. Ryan Gregory. Synergy between sequence and size in large-scale genomics. *Nature Reviews Genetics*, 6(9):699–708, 2005. ISSN 14710056. doi: 10.1038/nrg1674.

Robert L. Grossman, Allison P. Heath, Vincent Ferretti, Harold E. Varmus, Douglas R. Lowy, Warren A. Kibbe, and Louis M. Staudt. Toward a Shared Vision for Cancer Genomic Data. *New England Journal of Medicine*, 375(12):1109–1112, sep 2016. ISSN 0028-4793. doi: 10.1056/NEJMp1607591. URL <http://www.nejm.org/doi/10.1056/NEJMp1607591>.

Zuguang Gu, Roland Eils, and Matthias Schlesner. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18):2847–2849, sep 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btw313. URL <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw313>.

Junjie U Guo, Yijing Su, Joo Heon Shin, Jaehoon Shin, Hongda Li, Bin Xie, Chun Zhong, Shaohui Hu, Thuc Le, Guoping Fan, Heng Zhu, Qiang Chang, Yuan Gao, Guo-li Ming, and Hongjun Song. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature neuroscience*, 17(2):215–22, feb 2014. ISSN 1546-1726. doi: 10.1038/nn.3607. URL <http://www.ncbi.nlm.nih.gov/pubmed/24362762>.

Jan O. Haerter, Cecilia Lökvist, Ian B. Dodd, and Kim Sneppen. Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic acids research*, 42(4):2235–44, feb 2014. ISSN 1362-4962. doi: 10.1093/nar/gkt1235. URL <http://www.ncbi.nlm.nih.gov/pubmed/24288373> <http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC3936770>.

- Oliver Hahn, Sebastian Grönke, Thomas M. Stubbs, Gabriella Ficz, Oliver Hendrich, Felix Krueger, Simon Andrews, Qifeng Zhang, Michael J. Wakelam, Andreas Beyer, Wolf Reik, and Linda Partridge. Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. *Genome Biology*, 18(1):56, dec 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1187-1. URL <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1187-1>.
- Gregory Hannum, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, Srinivas Sadda, Brandy Klotzle, Marina Bibikova, Jian-Bing Fan, Yuan Gao, Rob Deconde, Menzies Chen, Indika Rajapakse, Stephen Friend, Trey Ideker, and Kang Zhang. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell*, 49(2):359–367, jan 2013. ISSN 1097-4164. doi: 10.1016/j.molcel.2012.10.016. URL <http://www.ncbi.nlm.nih.gov/pubmed/23177740>.
- Malene Hansen, Stefan Taubert, Douglas Crawford, Nataliya Libina, Seung-Jae Lee, and Cynthia Kenyon. Lifespan extension by conditions that inhibit translation in *Caenorhabditis elegans*. *Aging Cell*, 6(1):95–110, feb 2007. ISSN 1474-9718. doi: 10.1111/j.1474-9726.2006.00267.x. URL <http://doi.wiley.com/10.1111/j.1474-9726.2006.00267.x>.
- Alan Harrison and Anne Parle-McDermott. DNA Methylation: A Timeline of Methods and Applications. *Frontiers in Genetics*, 2(3):632–649, 2011. ISSN 1664-8021. doi: 10.3389/fgene.2011.00074. URL <http://journal.frontiersin.org/article/10.3389/fgene.2011.00074/abstract>.
- Nicholas C Harvey, Kassim Javaid, Nicholas Bishop, Stephen Kennedy, Aris T Papageorgiou, Robert Fraser, Saurabh V Gandhi, Inez Schoenmakers, Ann Prentice, and Cyrus Cooper. MAVIDOS Maternal Vitamin D Osteoporosis Study: study protocol for a randomized controlled trial. The MAVIDOS Study Group. *Trials*, 13(1):13, dec 2012. ISSN 1745-6215. doi: 10.1186/1745-6215-13-13. URL <http://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-13-13>.
- Nicholas C. Harvey, Allan Sheppard, Keith M. Godfrey, Cameron McLean, Emma Garratt, Georgia Ntani, Lucy Davies, Robert Murray, Hazel M. Inskip, Peter D. Gluckman, Mark A. Hanson, Karen A. Lillycrop, and Cyrus Cooper. Childhood Bone Mineral Content Is Associated With Methylation Status of the RXRA Promoter at Birth. *Journal of Bone and Mineral Research*, 29(3):600–607, mar 2014. ISSN 08840431. doi: 10.1002/jbmr.2056. URL <http://doi.wiley.com/10.1002/jbmr.2056>.
- Tamar Hashimshony, Jianmin Zhang, Ilana Keshet, Michael Bustin, and Howard Cedar. The role of DNA methylation in setting up chromatin structure during development. *Nature Genetics*, 34(2):187–192, jun 2003. ISSN 1061-4036. doi: 10.1038/ng1158. URL <http://www.nature.com/articles/ng1158>.
- I Hatada, Y Hayashizaki, S Hirotsune, H Komatsubara, and T Mukai. A genomic scanning method for higher organisms using restriction sites as landmarks. *Proceedings of the National Academy of Sciences of the United States of America*, 88(21):9523–7, nov 1991. ISSN 0027-8424. doi: 10.1073/pnas.88.21.9523. URL <http://www.ncbi.nlm.nih.gov/pubmed/1946366>.
- Hikoya Hayatsu, Yusuke Wataya, Kazushige Kai, and Shigeru Iida. Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry*, 9(14):2858–2865, jul 1970. ISSN 0006-2960. doi: 10.1021/bi00816a016. URL <http://pubs.acs.org/doi/abs/10.1021/bi00816a016>.
- Asaf Hellman and Andrew Chess. Gene body-specific methylation on the active X chromosome. *Science*, 315(5815):1141–1143, 2007. ISSN 00368075. doi: 10.1126/science.1136352.
- Holger Heyn, Ning Li, Humberto J Ferreira, Sebastian Moran, David G Pisano, Antonio Gomez, Javier Diez, J. V. Sanchez-Mut, F. Setien, F. J. Carmona, A. A. Puca, S. Sayols, M. A. Pujana, J. Serra-Musach, I. Iglesias-Platas, F. Formiga, A. F. Fernandez, M. F. Fraga, S. C. Heath, A. Valencia, I. G. Gut, J. Wang, and M. Esteller. Distinct DNA methylomes of newborns and centenarians. *Proceedings of the National Academy of Sciences*, 109(26):10522–10527, jun 2012. ISSN 0027-8424. doi: 10.1073/pnas.1203530109. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3408300/>.

2012. ISSN 0027-8424. doi: 10.1073/pnas.1120658109. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1120658109>.
- Michael M Hoffman, Orion J Buske, Jie Wang, Zhiping Weng, Jeff a Bilmes, and William Stafford Noble. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nature methods*, 9(5):473–6, mar 2012. ISSN 1548-7105. doi: 10.1038/nmeth.1937. URL <http://www.ncbi.nlm.nih.gov/pubmed/22426492>.
- Michelle L Holland, Robert Lowe, Paul W Caton, Carolina Gemma, Guillermo Carbajosa, Amy F Danson, Asha A M Carpenter, Elena Loche, Susan E Ozanne, and Vardhman K Rakyan. Early-life nutrition modulates the epigenetic state of specific rDNA genetic variants in mice. *Science*, 353(6298):495–498, jul 2016. ISSN 0036-8075. doi: 10.1126/science.aaf7040. URL <http://www.sciencemag.org/lookup/doi/10.1126/science.aaf7040>.
- Steve Horvath. DNA methylation age of human tissues and cell types. *Genome biology*, 14(10):R115, 2013. ISSN 1474-760X. doi: 10.1186/gb-2013-14-10-r115. URL <http://www.ncbi.nlm.nih.gov/pubmed/24138928>.
- Steve Horvath and Kenneth Raj. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, 19(6):371–384, jun 2018. ISSN 1471-0056. doi: 10.1038/s41576-018-0004-3. URL <http://www.nature.com/articles/s41576-018-0004-3>.
- Eugene Andres Houseman, William P Accomando, Devin C Koestler, Brock C Christensen, Carmen J Marsit, Heather H Nelson, John K Wiencke, and Karl T Kelsey. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, 13(1):86, dec 2012. ISSN 1471-2105. doi: 10.1186/1471-2105-13-86. URL <http://www.biomedcentral.com/1471-2105/13/86>.
- Shi-qiong Huang, Bao Sun, Zong-ping Xiong, Yan Shu, Hong-hao Zhou, Wei Zhang, Jing Xiong, and Qing Li. The dysregulation of tRNAs and tRNA derivatives in cancer. *Journal of Experimental & Clinical Cancer Research*, 37(1):101, dec 2018. ISSN 1756-9966. doi: 10.1186/s13046-018-0745-z. URL <https://jeccr.biomedcentral.com/articles/10.1186/s13046-018-0745-z>.
- Jason T. Huff and Daniel Zilberman. Dnmt1-Independent CG Methylation Contributes to Nucleosome Positioning in Diverse Eukaryotes. *Cell*, 156(6):1286–1297, mar 2014. ISSN 00928674. doi: 10.1016/j.cell.2014.01.029. URL <http://www.ncbi.nlm.nih.gov/pubmed/24630728>.
- HumanCellAtlas. Human Cell Atlas, 2017. URL <https://www.humancellatlas.org/>.
- James R. Iben and Richard J. Maraia. tRNA gene copy number variation in humans. *Gene*, 536(2):376–384, feb 2014. ISSN 03781119. doi: 10.1016/j.gene.2013.11.049. URL <https://linkinghub.elsevier.com/retrieve/pii/S0378111913015758>.
- Robert S. Illingworth, Ulrike Gruenewald-Schneider, Shaun Webb, Alastair R W Kerr, Keith D. James, Daniel J. Turner, Colin Smith, David J. Harrison, Robert Andrews, and Adrian P. Bird. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS genetics*, 6(9):e1001134, sep 2010. ISSN 1553-7404. doi: 10.1371/journal.pgen.1001134. URL <http://www.ncbi.nlm.nih.gov/pubmed/20885785>.
- Rafael A Irizarry, Christine Ladd-Acosta, Bo Wen, Zhijin Wu, Carolina Montano, Patrick Onyango, Hengmi Cui, Kevin Gabo, Michael Rongione, Maree Webster, Hong Ji, James Potash, Sarven Sabunciyan, and Andrew P Feinberg. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature genetics*, 41(2):178–186, feb 2009. ISSN 1546-1718. doi: 10.1038/ng.298. URL <http://www.ncbi.nlm.nih.gov/pubmed/19151715>.
- Andrew E Jaffe and Rafael A Irizarry. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome biology*, 15(2):R31, feb 2014. ISSN 1474-760X. doi: 10.1186/gb-2014-15-2-r31. URL <http://www.ncbi.nlm.nih.gov/pubmed/24495553>.

Siddhartha Jaiswal, Pierre Fontanillas, Jason Flannick, Alisa Manning, Peter V. Grauman, Brenton G. Mar, R. Coleman Lindsley, Craig H. Mermel, Noel Burtt, Alejandro Chavez, John M. Higgins, Vladislav Moltchanov, Frank C. Kuo, Michael J. Kluk, Brian Henderson, Leena Kinnunen, Heikki A. Koistinen, Claes Ladenvall, Gad Getz, Adolfo Correa, Benjamin F. Banahan, Stacey Gabriel, Sekar Kathiresan, Heather M. Stringham, Mark I. McCarthy, Michael Boehnke, Jaakkko Tuomilehto, Christopher Haiman, Leif Groop, Gil Atzmon, James G. Wilson, Donna Neuberg, David Altshuler, and Benjamin L. Ebert. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine*, 371(26):2488–2498, dec 2014. ISSN 0028-4793. doi: 10.1056/NEJMoa1408617. URL <http://www.nejm.org/doi/10.1056/NEJMoa1408617>.

Hyo Sik Jang, Nakul M. Shah, Alan Y. Du, Zea Z. Dailey, Erica C. Pehrsson, Paula M. Godoy, David Zhang, Daofeng Li, Xiaoyun Xing, Sungsu Kim, David O'Donnell, Jeffrey I. Gordon, and Ting Wang. Transposable elements drive widespread expression of oncogenes in human cancers. *Nature Genetics*, 51(4):611–617, apr 2019. ISSN 1061-4036. doi: 10.1038/s41588-019-0373-3. URL <http://dx.doi.org/10.1038/s41588-019-0373-3>.

Garrett Jenkinson, Elisabet Pujadas, John Goutsias, and Andrew P Feinberg. Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nature genetics*, 49(5):719–729, may 2017. ISSN 1546-1718. doi: 10.1038/ng.3811. URL <http://www.ncbi.nlm.nih.gov/pubmed/28346445>.

Garrett Jenkinson, Jordi Abante, Andrew P. Feinberg, and John Goutsias. An information-theoretic approach to the modeling and analysis of whole-genome bisulfite sequencing data. *BMC Bioinformatics*, 19(1):87, dec 2018. ISSN 1471-2105. doi: 10.1186/s12859-018-2086-5. URL <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2086-5>.

T. Jenuwein. Translating the Histone Code. *Science*, 293(5532):1074–1080, aug 2001. ISSN 00368075. doi: 10.1126/science.1063127. URL <http://www.ncbi.nlm.nih.gov/pubmed/20653993>.

Da Jia, Renata Z. Jurkowska, Xing Zhang, Albert Jeltsch, and Xiaodong Cheng. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, 449(7159):248–251, sep 2007. ISSN 0028-0836. doi: 10.1038/nature06146. URL <http://www.nature.com/articles/nature06146>.

Nicholas D Johnson, Howard W Wiener, Alicia K Smith, Shota Nishitani, Devin M Absher, Donna K Arnett, Stella Aslibekyan, and Karen N Conneely. Non-linear patterns in age-related DNA methylation may reflect CD4 + T cell differentiation. *Epigenetics*, 12(6):492–503, jun 2017. ISSN 1559-2294. doi: 10.1080/15592294.2017.1314419. URL <http://www.ncbi.nlm.nih.gov/pubmed/28387568>.

Thomas E. Johnson. Recent results: Biomarkers of aging. *Experimental Gerontology*, 41(12):1243–1246, dec 2006. ISSN 05315565. doi: 10.1016/j.exger.2006.09.006. URL <https://linkinghub.elsevier.com/retrieve/pii/S0531556506002865>.

Meaghan J. Jones, Sarah J. Goodman, and Michael S. Kobor. DNA methylation and healthy human aging. *Aging Cell*, 14(6):924–932, dec 2015. ISSN 14749718. doi: 10.1111/acel.12349. URL <http://doi.wiley.com/10.1111/acel.12349>.

Mireia Jordà, Anna Díez-Villanueva, Izaskun Mallona, Berta Martín, Sergi Lois, Víctor Barrera, Manel Esteller, Tanya Vavouris, and Miguel A. Peinado. The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells. *Genome Research*, 27(1):118–132, jan 2017. ISSN 1088-9051. doi: 10.1101/gr.207522.116. URL <http://www.ncbi.nlm.nih.gov/pubmed/27999094>.

Bonnie R. Joubert, Janine F. Felix, Paul Yousefi, Kelly M. Bakulski, Allan C. Just, Carrie Breton, Sarah E. Reese, Christina A. Markunas, Rebecca C. Richmond, Cheng-Jian Xu, Leanne K. Küpers, Sam S. Oh, Cathrine Hoyo, Olena Gruzieva, Cilla Söderhäll, Lucas A. Salas, Nour Baiz, Hongmei Zhang, Johanna Lepeule, Carlos Ruiz, Symen Ligthart, Tianyuan

- Wang, Jack A. Taylor, Liesbeth Duijts, Gemma C. Sharp, Soesma A. Jankipersadsing, Roy M. Nilsen, Ahmad Vaez, M. Daniele Fallin, Donglei Hu, Augusto A. Litonjua, Bernard F. Fuemmeler, Karen Huen, Juha Kere, Inger Kull, Monica Cheng Munthe-Kaas, Ulrike Gehring, Mariona Bustamante, Marie José Saurel-Coubizolles, Bilal M. Quraishi, Jie Ren, Jörg Tost, Juan R. Gonzalez, Marjolein J. Peters, Siri E. Håberg, Zongli Xu, Joyce B. van Meurs, Tom R. Gaunt, Marjan Kerkhof, Eva Corpeleijn, Andrew P. Feinberg, Celeste Eng, Andrea A. Bacarelli, Sara E. Benjamin Neelon, Asa Bradman, Simon Kebede Merid, Anna Bergström, Zdenko Herceg, Hector Hernandez-Vargas, Bert Brunekreef, Mariona Pinart, Barbara Heude, Susan Ewart, Jin Yao, Nathanaël Lemonnier, Oscar H. Franco, Michael C. Wu, Albert Hofman, Wendy McArdle, Pieter Van der Vlies, Fahimeh Falahi, Matthew W. Gillman, Lisa F. Barcellos, Ashish Kumar, Magnus Wickman, Stefano Guerra, Marie-Aline Charles, John Holloway, Charles Auffray, Henning W. Tiemeier, George Davey Smith, Dirkje Postma, Marie-France Hivert, Brenda Eskenazi, Martine Vrijheid, Hasan Arshad, Josep M. Antó, Abbas Dehghan, Wilfried Karmaus, Isabella Annesi-Maesano, Jordi Sunyer, Akram Ghantous, Göran Pershagen, Nina Holland, Susan K. Murphy, Dawn L. DeMeo, Esteban G. Burchard, Christine Ladd-Acosta, Harold Snieder, Wenche Nystad, Gerard H. Koppelman, Caroline L. Relton, Vincent W.V. Jaddoe, Allen Wilcox, Erik Melén, and Stephanie J. London. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *The American Journal of Human Genetics*, 98(4):680–696, apr 2016. ISSN 00029297. doi: 10.1016/j.ajhg.2016.02.019. URL <http://www.ncbi.nlm.nih.gov/pubmed/27040690>.
- Marc Jung and Gerd P Pfeifer. Aging and DNA methylation. *BMC Biology*, 13(1):7, dec 2015. ISSN 1741-7007. doi: 10.1186/s12915-015-0118-4. URL <http://www.biomedcentral.com/1741-7007/13/7>.
- Simonas Juzenas, Geetha Venkatesh, Matthias Hüenthal, Marc P. Hoeppner, Zhipei Gracie Du, Maren Paulsen, Philip Rosenstiel, Philipp Senger, Martin Hofmann-Apitius, Andreas Keller, Limas Kucinskas, Andre Franke, and Georg Hemmrich-Stanisak. A comprehensive, cell specific microRNA catalogue of human peripheral blood. *Nucleic Acids Research*, 45(16): 9290–9301, sep 2017. ISSN 0305-1048. doi: 10.1093/nar/gkx706. URL <http://academic.oup.com/nar/article/45/16/9290/4080663>.
- Minna U. Kaikkonen, Michael T Y Lam, and Christopher K. Glass. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular research*, 90(3):430–40, jun 2011. ISSN 1755-3245. doi: 10.1093/cvr/cvr097. URL <http://www.ncbi.nlm.nih.gov/pubmed/21558279>.
- Haig H. Kazazian and John V. Moran. Mobile DNA in health and disease. *New England Journal of Medicine*, 377(4):361–370, 2017. ISSN 15334406. doi: 10.1056/NEJMra1510092.
- Simon P. Keam, Paul E. Young, Alexandra L. McCorkindale, Thurston H.Y. Dang, Jennifer L. Clancy, David T. Humphreys, Thomas Preiss, Gyorgy Hutvagner, David I.K. Martin, Jennifer E. Cropley, and Catherine M. Suter. The human Piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells. *Nucleic Acids Research*, 42(14):8984–8995, aug 2014. ISSN 0305-1048. doi: 10.1093/nar/gku620. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku620>.
- Theresa K Kelly, Yaping Liu, Fides D Lay, Gangning Liang, Benjamin P Berman, and Peter a Jones. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Research*, 22(12):2497–2506, dec 2012. ISSN 1088-9051. doi: 10.1101/gr.143008.112. URL <http://www.ncbi.nlm.nih.gov/pubmed/22960375>.
- Brian K. Kennedy and Dudley W. Lamming. The Mechanistic Target of Rapamycin: The Grand ConducTOR of Metabolism and Aging. *Cell Metabolism*, 23(6):990–1003, jun 2016. ISSN 15504131. doi: 10.1016/j.cmet.2016.05.009. URL <http://dx.doi.org/10.1016/j.cmet.2016.05.009>.
- Kristi Kerkel, Alexandra Spadola, Eric Yuan, Jolanta Kosek, Le Jiang, Eldad Hod, Kerry Li, Vundavalli V. Murty, Nicole Schupf, Eric Vilain, Mitzi Morris, Fatemeh Haghghi, and Benjamin Tycko. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent

- allele-specific DNA methylation. *Nature genetics*, 40(7):904–8, jul 2008. ISSN 1546-1718. doi: 10.1038/ng.174. URL <http://www.ncbi.nlm.nih.gov/pubmed/18568024>.
- Hak Kyun Kim, Gabriele Fuchs, Shengchun Wang, Wei Wei, Yue Zhang, Hyesuk Park, Biswajoy Roy-Chaudhuri, Pan Li, Jianpeng Xu, Kirk Chu, Feijie Zhang, Mei-Sze Chua, Samuel So, Qiangfeng Cliff Zhang, Peter Sarnow, and Mark A. Kay. A transfer-RNA-derived small RNA regulates ribosome biogenesis. *Nature*, 552(7683):57–62, dec 2017. ISSN 0028-0836. doi: 10.1038/nature25005. URL <http://www.nature.com/articles/nature25005>.
- Sertac N Kip and Emanuel E Strehler. Vitamin D 3 upregulates plasma membrane Ca 2+ - ATPase expression and potentiates apico-basal Ca 2+ flux in MDCK cells. *American Journal of Physiology-Renal Physiology*, 286(2):F363–F369, feb 2004. ISSN 1931-857X. doi: 10.1152/ajprenal.00076.2003. URL <http://www.ncbi.nlm.nih.gov/pubmed/14583431>.
- Sebastian Kirchner, Zhiwei Cai, Robert Rauscher, Nicolai Kastelic, Melanie Anding, Andreas Czech, Bertrand Kleizen, Lynda S Ostedgaard, Ineke Braakman, David N Sheppard, and Zoya Ignatova. Alteration of protein function by a silent polymorphism linked to tRNA abundance. *PLOS Biology*, 15(5):e2000779, may 2017. ISSN 1545-7885. doi: 10.1371/journal.pbio.2000779. URL <http://dx.plos.org/10.1371/journal.pbio.2000779>.
- Anna K. Knight, Jeffrey M. Craig, Christiane Theda, Marie Bækvad-Hansen, Jonas Bybjerg-Grauholt, Christine S. Hansen, Mads V. Hollegaard, David M. Hougaard, Preben B. Mortensen, Shantel M. Weinsheimer, Thomas M. Werge, Patricia A. Brennan, Joseph F. Cubells, D. Jeffrey Newport, Zachary N. Stowe, Jeanie L. Y. Cheong, Philippa Dalach, Lex W. Doyle, Yuk J. Loke, Andrea A. Baccarelli, Allan C. Just, Robert O. Wright, Mara M. Tellez-Rojo, Katherine Svensson, Letizia Trevisi, Elizabeth M. Kennedy, Elisabeth B. Binder, Stella Iurato, Darina Czamara, Katri Räikkönen, Jari M. T. Lahti, Anu-Katriina Pesonen, Eero Kajantie, Pia M. Villa, Hannele Laivuori, Esa Hämäläinen, Hea Jin Park, Lynn B. Bailey, Sasha E. Parets, Varun Kilaru, Ramkumar Menon, Steve Horvath, Nicole R. Bush, Kaja Z. LeWinn, Frances A. Tylavsky, Karen N. Conneely, and Alicia K. Smith. An epigenetic clock for gestational age at birth based on blood methylation data. *Genome Biology*, 17(1):206, dec 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-1068-z. URL <http://dx.doi.org/10.1186/s13059-016-1068-z>.
- A G Knudson. Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, 68(4):820–3, apr 1971. ISSN 0027-8424. doi: 10.1073/pnas.68.4.820. URL <http://www.ncbi.nlm.nih.gov/pubmed/5279523>.
- Carmen M. Koch and Wolfgang Wagner. Epigenetic-aging-signature to determine age in different tissues. *Aging*, 3(10):1018–27, oct 2011. ISSN 1945-4589. doi: 10.18632/aging.100395. URL <http://www.ncbi.nlm.nih.gov/pubmed/22067257>.
- Sarah E. Kolitz and Jon R. Lorsch. Eukaryotic initiator tRNA: Finely tuned and ready for action. *FEBS Letters*, 584(2):396–404, jan 2010. ISSN 00145793. doi: 10.1016/j.febslet.2009.11.047. URL <http://doi.wiley.com/10.1016/j.febslet.2009.11.047>.
- Preethi Krishnan, Sunita Ghosh, Bo Wang, Mieke Heyns, Dongping Li, John R. Mackey, Olga Kovalchuk, and Sambasivarao Damaraju. Genome-wide profiling of transfer RNAs and their role as novel prognostic markers for breast cancer. *Scientific Reports*, 6(1):32843, dec 2016. ISSN 2045-2322. doi: 10.1038/srep32843. URL <http://dx.doi.org/10.1038/srep32843>.
- Felix Krueger. Trim Galore, 2015. URL [https://www.bioinformatics.babraham.ac.uk/projects/trim{\\_\\_}galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim{__}galore/).
- Felix Krueger and Simon R. Andrews. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27(11):1571–1572, jun 2011. ISSN 1460-2059. doi: 10.1093/bioinformatics/btr167. URL <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr167>.
- Thomas A. Kunkel. DNA Replication Fidelity. *Journal of Biological Chemistry*, 279(17):16895–16898, apr 2004. ISSN 0021-9258. doi: 10.1074/jbc.R400006200. URL <http://www.jbc.org/article-lookup/doi/10.1074/jbc.R400006200>.

- Charles D Laird, Nicole D Pleasant, Aaron D Clark, Jessica L Sneeden, K M Anwarul Hassan, Nathan C Manley, Jay C Vary, Todd Morgan, R Scott Hansen, and Reinhard Stöger. Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 101(1):204–9, jan 2004. ISSN 0027-8424. doi: 10.1073/pnas.2536758100. URL <http://www.ncbi.nlm.nih.gov/pubmed/14673087>.
- E S Lander, L M Linton, B Birren, C Nusbaum, M C Zody, J Baldwin, K Devon, K Dewar, M Doyle, W FitzHugh, R Funke, D Gage, K Harris, A Heaford, J Howland, L Kann, J Lehoczky, R LeVine, P McEwan, K McKernan, J Meldrim, J P Mesirov, C Miranda, W Morris, J Naylor, C Raymond, M Rosetti, R Santos, A Sheridan, C Sougnez, Y Stange-Thomann, N Stojanovic, A Subramanian, D Wyman, J Rogers, J Sulston, R Ainscough, S Beck, D Bentley, J Burton, C Cleo, N Carter, A Coulson, R Deadman, P Deloukas, A Dunham, I Dunham, R Durbin, L French, D Graham, S Gregory, T Hubbard, S Humphray, A Hunt, M Jones, C Lloyd, A McMurray, L Matthews, S Mercer, S Milne, J C Mullikin, A Mungall, R Plumb, M Ross, R Shownkeen, S Sims, R H Waterston, R K Wilson, L W Hillier, J D McPherson, M A Marra, E R Mardis, L A Fulton, A T Chinwalla, K H Pepin, W R Gish, S L Chissoe, M C Wendl, K D Delehaunty, T L Miner, A Delehaunty, J B Kramer, L L Cook, R S Fulton, D L Johnson, P J Minx, S W Clifton, T Hawkins, E Branscomb, P Predki, P Richardson, S Wenning, T Slezak, N Doggett, J F Cheng, A Olsen, S Lucas, C Elkin, E Uberbacher, M Frazier, R A Gibbs, D M Muzny, S E Scherer, J B Bouck, E J Sodergren, K C Worley, C M Rives, J H Gorrell, M L Metzker, S L Naylor, R S Kucherlapati, D L Nelson, G M Weinstock, Y Sakaki, A Fujiyama, M Hattori, T Yada, A Toyoda, T Itoh, C Kawagoe, H Watanabe, Y Totoki, T Taylor, J Weissenbach, R Heilig, W Saurin, F Artiguenave, P Brottier, T Bruls, E Pelletier, C Robert, P Wincker, D R Smith, L Doucette-Stamm, M Rubenfield, K Weinstock, H M Lee, J Dubois, A Rosenthal, M Platzer, G Nyakatura, S Taudien, A Rump, H Yang, J Yu, J Wang, G Huang, J Gu, L Hood, L Rowen, A Madan, S Qin, R W Davis, N A Federspiel, A P Abola, M J Proctor, R M Myers, J Schmutz, M Dickson, J Grimwood, D R Cox, M V Olson, R Kaul, C Raymond, N Shimizu, K Kawasaki, S Minoshima, G A Evans, M Athanasiou, R Schultz, B A Roe, F Chen, H Pan, J Ramser, H Lehrach, R Reinhardt, W R McCombie, M de la Bastide, N Dedhia, H Blöcker, K Hornischer, G Nordsiek, R Agarwala, L Aravind, J A Bailey, A Bateman, S Batzoglou, E Birney, P Bork, D G Brown, C B Burge, L Cerutti, H C Chen, D Church, M Clamp, R R Copley, T Doerks, S R Eddy, E E Eichler, T S Furey, J Galagan, J G Gilbert, C Harmon, Y Hayashizaki, D Haussler, H Hermjakob, K Hokamp, W Jang, L S Johnson, T A Jones, S Kasif, A Kaspryzk, S Kennedy, W J Kent, P Kitts, E V Koonin, I Korf, D Kulp, D Lancet, T M Lowe, A McLysaght, T Mikkelsen, J V Moran, N Mulder, V J Pollara, C P Ponting, G Schuler, J Schultz, G Slater, A F Smit, E Stupka, J Szustakowski, D Thierry-Mieg, J Thierry-Mieg, L Wagner, J Wallis, R Wheeler, A Williams, Y I Wolf, K H Wolfe, S P Yang, R F Yeh, F Collins, M S Guyer, J Peterson, A Felsenfeld, K A Wetterstrand, A Patrinos, M J Morgan, P de Jong, J J Catanese, K Osoegawa, H Shizuya, S Choi, Y J Chen, J Szustakowski, and International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, feb 2001. ISSN 0028-0836. doi: 10.1038/35057062. URL <http://www.ncbi.nlm.nih.gov/pubmed/11237011>.
- Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357–9, mar 2012. ISSN 1548-7105. doi: 10.1038/nmeth.1923. URL <http://www.ncbi.nlm.nih.gov/pubmed/22388286>.
- Tuuli Lappalainen and John M. Greally. Associating cellular epigenetic models with human phenotypes. *Nature Reviews Genetics*, 18(7):441–451, jul 2017. ISSN 1471-0056. doi: 10.1038/nrg.2017.32. URL <http://www.nature.com/articles/nrg.2017.32>.
- Yong Sun Lee, Yoshiyuki Shibata, Ankit Malhotra, and Anindya Dutta. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes & development*, 23(22):2639–49, nov 2009. ISSN 1549-5477. doi: 10.1101/gad.1837609. URL <http://www.ncbi.nlm.nih.gov/pubmed/19933153>.
- Jeffrey T. Leek and John D. Storey. Capturing Heterogeneity in Gene Expression Studies by

- Surrogate Variable Analysis. *PLoS Genetics*, 3(9):e161, 2007. ISSN 1553-7390. doi: 10.1371/journal.pgen.0030161. URL <https://dx.plos.org/10.1371/journal.pgen.0030161>.
- Morgan E Levine, Ake T Lu, Austin Quach, Brian H Chen, Themistocles L Assimes, Stefania Bandinelli, Lifang Hou, Andrea A Baccarelli, James D Stewart, Yun Li, Eric A Whitsel, James G Wilson, Alex P Reiner, Abraham Aviv, Kurt Lohman, Yongmei Liu, Luigi Ferrucci, and Steve Horvath. An epigenetic biomarker of aging for lifespan and healthspan. *Aging*, 10(4):573–591, apr 2018. ISSN 1945-4589. doi: 10.18632/aging.101414. URL <http://www.ncbi.nlm.nih.gov/pubmed/29676998>.
- Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. UpSet: Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1983–1992, dec 2014. ISSN 1077-2626. doi: 10.1109/TVCG.2014.2346248. URL <http://caleydo.org/tools/upset/>.
- Jingqiu Li, Jie Yang, Ping Zhou, Yanping Le, Chengwei Zhou, Shaomin Wang, Dazhi Xu, Hui-Kuan Lin, and Zhaohui Gong. Circular RNAs in cancer: novel insights into origins, properties, functions and implications. *American journal of cancer research*, 5(2):472–80, 2015. ISSN 2156-6976. URL <http://www.ncbi.nlm.nih.gov/pubmed/25973291>.
- Long-Cheng Li and Rajvir Dahiya. MethPrimer: designing primers for methylation PCRs. *Bioinformatics (Oxford, England)*, 18(11):1427–31, nov 2002. ISSN 1367-4803. URL <http://www.ncbi.nlm.nih.gov/pubmed/12424112>.
- Siqi Li, Zhengping Xu, and Jinghao Sheng. tRNA-Derived Small RNA: A Novel Regulatory Small Non-Coding RNA. *Genes*, 9(5):246, may 2018. ISSN 2073-4425. doi: 10.3390/genes9050246. URL <http://www.mdpi.com/2073-4425/9/5/246>.
- Siqi Li, Xiaoliang Shi, Muxiong Chen, Ningqin Xu, Desen Sun, Rongpan Bai, Haiyan Chen, Kefeng Ding, Jinghao Sheng, and Zhengping Xu. Angiogenin promotes colorectal cancer metastasis via tiRNA production. *International Journal of Cancer*, page ijc.32245, mar 2019. ISSN 0020-7136. doi: 10.1002/ijc.32245. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.32245>.
- Emanuele Libertini, Simon C Heath, Rifat A Hamoudi, Marta Gut, Michael J Ziller, Javier Herrero, Agata Czyz, Victor Ruotti, Hendrik G Stunnenberg, Mattia Frontini, Willem H Ouwehand, Alexander Meissner, Ivo G Gut, and Stephan Beck. Saturation analysis for whole-genome bisulfite sequencing data. *Nature biotechnology*, 34(7):691–693, jun 2016. ISSN 1546-1696. doi: 10.1038/nbt.3524. URL <http://www.ncbi.nlm.nih.gov/pubmed/27347755>.
- Florian Lienert, Christiane Wirlbauer, Indrani Som, Ann Dean, Fabio Mohn, and Dirk Schübler. Identification of genetic elements that autonomously determine DNA methylation states. *Nature genetics*, 43(11):1091–7, oct 2011. ISSN 1546-1718. doi: 10.1038/ng.946. URL <http://www.ncbi.nlm.nih.gov/pubmed/21964573>.
- Matthias Lienhard, Christina Grimm, Markus Morkel, Ralf Herwig, and Lukas Chavez. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics*, 30(2):284–286, jan 2014. ISSN 1460-2059. doi: 10.1093/bioinformatics/btt650. URL <http://www.ncbi.nlm.nih.gov/pubmed/24227674>.
- Karen A. Lillycrop, Emma S. Phillips, Christopher Torrens, Mark A. Hanson, Alan A. Jackson, and Graham C. Burdge. Feeding pregnant rats a protein-restricted diet persistently alters the methylation of specific cytosines in the hepatic PPAR $\alpha$  promoter of the offspring. *British Journal of Nutrition*, 100(2):278–282, aug 2008. ISSN 0007-1145. doi: 10.1017/S0007114507894438. URL [https://www.cambridge.org/core/product/identifier/S0007114507894438/type/journal{\\_\\_}article](https://www.cambridge.org/core/product/identifier/S0007114507894438/type/journal{__}article).
- Ryan Lister, Mattia Pelizzola, Robert H Dowen, R David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R Nery, Leonard Lee, Zhen Ye, Que-minh Ngo, Lee Edsall, Jessica Antosiewicz-Bourget, Ron Stewart, Victor Ruotti, A Harvey Millar, James A Thomson, Bing Ren, and Joseph R Ecker. Human DNA methylomes at base resolution show

- widespread epigenomic differences. *Nature*, 462(7271):315–22, nov 2009. ISSN 1476-4687. doi: 10.1038/nature08514. URL <http://www.ncbi.nlm.nih.gov/pubmed/19829295>.
- Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell. *Molecular Cell Biology, 4th edition*. W. H. Freeman, New York, 4th edition, 2000. ISBN 10: 0-7167-3136-3. URL <https://www.ncbi.nlm.nih.gov/books/NBK21475/>.
- Phillipe Loher, Aristeidis G. Telonis, and Isidore Rigoutsos. MINTmap: fast and exhaustive profiling of nuclear and mitochondrial tRNA fragments from short RNA-seq data. *Scientific reports*, 7(1):41184, dec 2017. ISSN 2045-2322. doi: 10.1038/srep41184. URL <http://www.ncbi.nlm.nih.gov/pubmed/28220888>.
- Carlos López-Otín, Maria A. Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer. The hallmarks of aging. *Cell*, 153(6):1194–217, jun 2013. ISSN 1097-4172. doi: 10.1016/j.cell.2013.05.039. URL <http://www.ncbi.nlm.nih.gov/pubmed/23746838>.
- Robert Lowe, Carl Barton, Christopher A. Jenkins, Christina Ernst, Oliver Forman, Denise S. Fernandez-Twinn, Christoph Bock, Stephen J. Rossiter, Chris G. Faulkes, Susan E. Ozanne, Lutz Walter, Duncan T. Odom, Cathryn Mellersh, and Vardhman K. Rakyan. Ageing-associated DNA methylation dynamics are a molecular readout of lifespan variation among mammalian species. *Genome Biology*, 19(1):22, dec 2018. ISSN 1474-760X. doi: 10.1186/s13059-018-1397-1. URL <http://www.ncbi.nlm.nih.gov/pubmed/29452591>.
- Todd M. Lowe and Patricia P. Chan. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, 44(W1):W54–W57, jul 2016. ISSN 0305-1048. doi: 10.1093/nar/gkw413. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw413>.
- Ake T Lu, Eilis Hannon, Morgan E Levine, Eileen M Crimmins, Katie Lunnon, Jonathan Mill, Daniel H Geschwind, and Steve Horvath. Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. *Nature Communications*, 8(May):15353, may 2017. ISSN 2041-1723. doi: 10.1038/ncomms15353. URL <http://dx.doi.org/10.1038/ncomms15353>.
- Yanting Luo, Xuemei Lu, and Hehuang Xie. Dynamic Alu Methylation during Normal Development, Aging, and Tumorigenesis. *BioMed Research International*, 2014:1–12, 2014. ISSN 2314-6133. doi: 10.1155/2014/784706. URL <http://www.hindawi.com/journals/bmri/2014/784706/>.
- Frank Lyko. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, 19(2):81–92, oct 2017. ISSN 1471-0056. doi: 10.1038/nrg.2017.80. URL <http://www.nature.com/doifinder/10.1038/nrg.2017.80>.
- Pamela Mahon, Nicholas Harvey, Sarah Crozier, Hazel Inskip, Sian Robinson, Nigel Arden, Rama Swaminathan, Cyrus Cooper, and Keith Godfrey. Low maternal vitamin D status and fetal bone development: Cohort study. *Journal of Bone and Mineral Research*, 25(1):14–19, jan 2010. ISSN 08840431. doi: 10.1359/jbmr.090701. URL <http://doi.wiley.com/10.1359/jbmr.090701>.
- François Mange, Viviane Praz, Eugenia Migliavacca, Ian M Willis, Frédéric Schütz, and Nouria Hernandez. Diurnal regulation of RNA polymerase III transcription is under the control of both the feeding–fasting response and the circadian clock. *Genome Research*, 27(6):973–984, jun 2017. ISSN 1088-9051. doi: 10.1101/gr.217521.116. URL <http://genome.cshlp.org/lookup/doi/10.1101/gr.217521.116>.
- Riccardo E Marioni, Sonia Shah, Allan F McRae, Brian H Chen, Elena Colicino, Sarah E Harris, Jude Gibson, Anjali K Henders, Paul Redmond, Simon R Cox, Alison Pattie, Janie Corley, Lee Murphy, Nicholas G Martin, Grant W Montgomery, Andrew P Feinberg, M Daniele Fallin, Michael L Multhaup, Andrew E Jaffe, Roby Joehanes, Joel Schwartz, Allan C Just, Kathryn L Lunetta, Joanne M Murabito, John M Starr, Steve Horvath, Andrea A Baccarelli, Daniel Levy, Peter M Visscher, Naomi R Wray, and Ian J Deary. DNA methylation age of blood predicts all-cause mortality in later life. *Genome biology*, 16(1):25, jan 2015. ISSN 1474-760X. doi: 10.1186/s13059-015-0584-6. URL <http://www.ncbi.nlm.nih.gov/pubmed/25633388>.

- Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10, may 2011. ISSN 2226-6089. doi: 10.14806/ej.17.1.200. URL <http://journal.embnet.org/index.php/embnetjournal/article/view/200>.
- German Martinez, Sarah G. Choudury, and R. Keith Slotkin. tRNA-derived small RNAs target transposable element transcripts. *Nucleic Acids Research*, 45(9):5142–5152, may 2017. ISSN 0305-1048. doi: 10.1093/nar/gkx103. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkx103>.
- Alika K Maunakea, Raman P Nagarajan, Mikhail Bilenky, Tracy J Ballinger, Cletus D’Souza, Shaun D Fouse, Brett E Johnson, Chibo Hong, Cydney Nielsen, Yongjun Zhao, Gustavo Turecki, Allen Delaney, Richard Varhol, Nina Thiessen, Ksenya Shchors, Vivi M Heine, David H Rowitch, Xiaoyun Xing, Chris Fiore, Maximiliaan Schillebeeckx, Steven J M Jones, David Haussler, Marco A Marra, Martin Hirst, Ting Wang, and Joseph F Costello. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, 466(7303):253–7, jul 2010. ISSN 1476-4687. doi: 10.1038/nature09165. URL <http://www.ncbi.nlm.nih.gov/pubmed/20613842>.
- Kevin McGregor, Sasha Bernatsky, Ines Colmegna, Marie Hudson, Tomi Pastinen, Aurélie Labbe, and Celia M.T. Greenwood. An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. *Genome Biology*, 17(1):84, dec 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-0935-y. URL <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0935-y>.
- Alexander Meissner, Andreas Gnirke, George W Bell, Bernard Ramsahoye, Eric S Lander, and Rudolf Jaenisch. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic acids research*, 33(18):5868–77, oct 2005. ISSN 1362-4962. doi: 10.1093/nar/gki901. URL <http://www.ncbi.nlm.nih.gov/pubmed/16224102>.
- Wouter Meuleman. epilogos, 2019. URL <https://epilogos.altius.org/> <https://github.com/Altius/epilogos>.
- Josine Min, Gibran Hemani, George Davey Smith, Caroline L Relton, and Matthew Suderman. Meffil: efficient normalisation and analysis of very large DNA methylation samples. *bioRxiv*, 44(0):125963, 2017. doi: 10.1101/125963. URL <https://www.biorxiv.org/content/early/2017/04/27/125963>.
- Alireza Moayyeri, Christopher J. Hammond, Ana M. Valdes, and Timothy D. Spector. Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *International Journal of Epidemiology*, 42(1):76–85, feb 2013. ISSN 0300-5771. doi: 10.1093/ije/dyr207. URL <http://www.ncbi.nlm.nih.gov/pubmed/22253318>.
- Khyobeni Mozhui and Ashutosh K. Pandey. Conserved effect of aging on DNA methylation and association with EZH2 polycomb protein in mice and humans. *Mechanisms of Ageing and Development*, 162:27–37, mar 2017. ISSN 00476374. doi: 10.1016/j.mad.2017.02.006. URL <http://dx.doi.org/10.1016/j.mad.2017.02.006>.
- Carolin A. Müller and Conrad A. Nieduszynski. DNA replication timing influences gene expression level. *The Journal of Cell Biology*, 216(7):1907–1914, jul 2017. ISSN 0021-9525. doi: 10.1083/jcb.201701061. URL <http://www.jcb.org/lookup/doi/10.1083/jcb.201701061>.
- Fabian Müller, Michael Scherer, Yassen Assenov, Pavlo Lutsik, Jörn Walter, Thomas Lengauer, and Christoph Bock. RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biology*, 20(1):55, dec 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1664-9. URL <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1664-9>.
- M. Murawski, B. Szczesniak, T. Zoladek, A. K. Hopper, N. C. Martin, and M. Boguta. maf1 mutation alters the subcellular localization of the Mod5 protein in yeast. *Acta biochimica Polonica*, 41(4):441–8, 1994. ISSN 0001-527X. URL <http://www.ncbi.nlm.nih.gov/pubmed/7732762>.

- Kristopher L. Nazor, Gulsah Altun, Candace Lynch, Ha Tran, Julie V. Harness, Ileana Slavin, Ibon Garitaonandia, Franz-Josef Müller, Yu-Chieh Wang, Francesca S. Boscolo, Eytaylo Fakunle, Biljana Dumevska, Sunray Lee, Hyun Sook Park, Tsaiwei Olee, Darryl D. D'Lima, Ruslan Semechkin, Mana M. Parast, Vasiliy Galat, Andrew L. Laslett, Uli Schmidt, Hans S. Keirstead, Jeanne F. Loring, and Louise C. Laurent. Recurrent Variations in DNA Methylation in Human Pluripotent Stem Cells and Their Differentiated Derivatives. *Cell Stem Cell*, 10(5):620–634, may 2012. ISSN 19345909. doi: 10.1016/j.stem.2012.02.013. URL <http://www.ncbi.nlm.nih.gov/pubmed/10449618>.
- Ken-ichi Noma, Hugh P. Cam, Richard J. Maraia, and Shiv I S Grewal. A role for TFIIC transcription factor complex in genome organization. *Cell*, 125(5):859–72, jun 2006. ISSN 0092-8674. doi: 10.1016/j.cell.2006.04.028. URL <http://www.ncbi.nlm.nih.gov/pubmed/16751097>.
- B.V. North, D. Curtis, and P.C. Sham. A Note on the Calculation of Empirical P Values from Monte Carlo Procedures. *The American Journal of Human Genetics*, 72(2):498–499, feb 2003. ISSN 00029297. doi: 10.1086/346173. URL <https://linkinghub.elsevier.com/retrieve/pii/S0002929707605618>.
- Jamaji C. Nwanaji-Enwerem, Marc G. Weisskopf, and Andrea A. Baccarelli. Multi-tissue DNA methylation age: Molecular relationships and perspectives for advancing biomarker utility. *Ageing Research Reviews*, 45(March):15–23, aug 2018. ISSN 15681637. doi: 10.1016/j.arr.2018.04.005. URL <https://doi.org/10.1016/j.arr.2018.04.005>.
- Cindy Y. Okitsu and Chih-Lin Hsieh. Sensitivity and specificity of immunoprecipitation of DNA containing 5-Methylcytosine. *BMC Research Notes*, 8(1):102, dec 2015. ISSN 1756-0500. doi: 10.1186/s13104-015-1069-0. URL <http://www.biomedcentral.com/1756-0500/8/102>.
- Marc Parisien, Xiaoyun Wang, and Tao Pan. Diversity of human tRNA genes from the 1000-genomes project. *RNA Biology*, 10(12):1853–1867, dec 2013. ISSN 1547-6286. doi: 10.4161/rna.27361. URL <http://www.tandfonline.com/doi/full/10.4161/rna.27361>.
- Linda Partridge, Joris Deelen, and P. Eline Slagboom. Facing up to the global challenges of ageing. *Nature*, 561(7721):45–56, sep 2018. ISSN 0028-0836. doi: 10.1038/s41586-018-0457-8. URL <http://www.nature.com/articles/s41586-018-0457-8>.
- Mariana Pavon-Eternod, Suzana Gomes, Marsha R Rosner, and Tao Pan. Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. *RNA*, 19(4):461–466, apr 2013. ISSN 1355-8382. doi: 10.1261/rna.037507.112. URL <http://www.ncbi.nlm.nih.gov/pubmed/23431330>.
- Lindsay M. Payer, Jared P. Steranka, Wan Rou Yang, Maria Kryatova, Sibyl Medabalimi, Daniel Ardeljan, Chunhong Liu, Jef D. Boeke, Dimitri Avramopoulos, and Kathleen H. Burns. Structural variants caused by Alu insertions are associated with risks for many human diseases. *Proceedings of the National Academy of Sciences*, 114(20):E3984–E3992, may 2017. ISSN 0027-8424. doi: 10.1073/pnas.1704117114. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1704117114>.
- Daniel A. Petkovich, Dmitriy I. Podolskiy, Alexei V. Lobanov, Sang-Goo Lee, Richard A. Miller, and Vadim N. Gladyshev. Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metabolism*, 25(4):954–960.e6, apr 2017. ISSN 15504131. doi: 10.1016/j.cmet.2017.03.016. URL <https://linkinghub.elsevier.com/retrieve/pii/S1550413117301687>.
- G P Pfeifer, S D Steigerwald, R S Hansen, S M Gartler, and A D Riggs. Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proceedings of the National Academy of Sciences*, 87(21):8252–8256, nov 1990. ISSN 0027-8424. doi: 10.1073/pnas.87.21.8252. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.87.21.8252>.

- Gerd P Pfeifer, Swati Kadam, and Seung-Gi Jin. 5-hydroxymethylcytosine and its potential roles in development and cancer. *Epigenetics & chromatin*, 6(1):10, may 2013. ISSN 1756-8935. doi: 10.1186/1756-8935-6-10. URL <http://www.ncbi.nlm.nih.gov/pubmed/23634848>.
- Ruth Pidsley, Elena Zotenko, Timothy J. Peters, Mitchell G. Lawrence, Gail P. Risbridger, Peter Molloy, Susan Van Djik, Beverly Muhlhausler, Clare Stirzaker, and Susan J. Clark. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome biology*, 17(1):208, dec 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-1066-1. URL <http://www.ncbi.nlm.nih.gov/pubmed/27717381>.
- Venetia Pliatsika, Phillip Loher, Rogan Magee, Aristeidis G. Telonis, Eric Londin, Megumi Shigematsu, Yohei Kirino, and Isidore Rigoutsos. MINTbase v2.0: a comprehensive database for tRNA-derived fragments that includes nuclear and mitochondrial fragments from all The Cancer Genome Atlas projects. *Nucleic Acids Research*, 46(D1):D152–D159, jan 2018. ISSN 0305-1048. doi: 10.1093/nar/gkx1075. URL <http://academic.oup.com/nar/article/46/D1/D152/4653530>.
- Krzysztof Pluta, Olivier Lefebvre, Nancy C Martin, Wieslaw J Smagowicz, David R Stanford, Steven R Ellis, Anita K Hopper, Andre Sentenac, and Magdalena Boguta. Maf1p, a Negative Effector of RNA Polymerase III in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 21(15):5031–5040, aug 2001. ISSN 0270-7306. doi: 10.1128/MCB.21.15.5031-5040.2001. URL <http://www.ncbi.nlm.nih.gov/pubmed/11438659>.
- Paz Polak and Eytan Domany. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC genomics*, 7:133, jun 2006. ISSN 1471-2164. doi: 10.1186/1471-2164-7-133. URL <http://www.ncbi.nlm.nih.gov/pubmed/16740159>.
- Aaron R. Quinlan and Ira M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, mar 2010. ISSN 1460-2059. doi: 10.1093/bioinformatics/btq033. URL <http://bedtools.readthedocs.io/en/latest/index.html>.
- Jesse R Raab, Jonathan Chiu, Jingchun Zhu, Sol Katzman, Sreenivasulu Kurukuti, Paul A Wade, David Haussler, and Rohinton T Kamakaka. Human tRNA genes function as chromatin insulators. *The EMBO journal*, 31(2):330–50, jan 2012. ISSN 1460-2075. doi: 10.1038/emboj.2011.406. URL <http://www.ncbi.nlm.nih.gov/pubmed/22085927>.
- Eun-ang Raiber, Guillem Portella, Sergio Martínez Cuesta, Robyn Hardisty, Pierre Murat, Zhe Li, Mario Iurlaro, Wendy Dean, Julia Spindel, Dario Beraldí, Zheng Liu, Mark A. Dawson, Wolf Reik, and Shankar Balasubramanian. 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nature Chemistry*, 10(12):1258–1266, dec 2018. ISSN 1755-4330. doi: 10.1038/s41557-018-0149-x. URL <http://www.nature.com/articles/s41557-018-0149-x>.
- Smrithi Rajendiran, Lee D. Gibbs, Timothy Van Treuren, David L. Klinkebiel, and Jamboor K. Vishwanatha. MIEN1 is tightly regulated by SINE Alu methylation in its promoter. *Oncotarget*, 7(40):65307–65319, oct 2016. ISSN 1949-2553. doi: 10.18632/oncotarget.11675. URL <http://www.oncotarget.com/fulltext/11675>.
- Vardhman K. Rakyan, Thomas A. Down, Siarhei Maslau, Toby Andrew, Tsun Po Yang, Huriya Beyan, Pamela Whittaker, Owen T. McCann, Sarah Finer, Ana M. Valdes, R. David Leslie, Panagiotis Deloukas, and Timothy D. Spector. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Research*, 20(4):434–439, apr 2010. ISSN 1088-9051. doi: 10.1101/gr.103101.109. URL <http://genome.cshlp.org/cgi/doi/10.1101/gr.103101.109>.
- Aviv Regev, Sarah A. Teichmann, Eric S. Lander, Ido Amit, Christophe Benoist, Ewan Birney, Bernd Bodenmiller, Peter Campbell, Piero Carninci, Menna Clatworthy, Hans Clevers, Bart Deplancke, Ian Dunham, James Eberwine, Roland Eils, Wolfgang Enard, Andrew Farmer, Lars Fugger, Berthold Göttgens, Nir Hacohen, Muzlifah Haniffa, Martin Hemberg, Seung

- Kim, Paul Klenerman, Arnold Kriegstein, Ed Lein, Sten Linnarsson, Emma Lundberg, Joakim Lundeberg, Partha Majumder, John C. Marioni, Miriam Merad, Musa Mhlanga, Martijn Nawijn, Mihai Netea, Garry Nolan, Dana Pe'er, Anthony Phillipakis, Chris P. Ponting, Stephen Quake, Wolf Reik, Orit Rozenblatt-Rosen, Joshua Sanes, Rahul Satija, Ton N. Schumacher, Alex Shalek, Ehud Shapiro, Padmanee Sharma, Jay W. Shin, Oliver Stegle, Michael Stratton, Michael J T Stubbington, Fabian J. Theis, Matthias Uhlen, Alexander van Oudenaarden, Alon Wagner, Fiona Watt, Jonathan Weissman, Barbara Wold, Ramnik Xavier, and Nir Yosef. The Human Cell Atlas. *eLife*, 6:1–30, dec 2017. ISSN 2050-084X. doi: 10.7554/eLife.27041. URL <https://elifesciences.org/articles/27041>.
- Lovisa E. Reinius, Nathalie Acevedo, Maaike Joerink, Göran Pershagen, Sven-Erik Dahlén, Dario Greco, Cilla Söderhäll, Annika Scheynius, and Juha Kere. Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. *PLoS ONE*, 7(7):e41361, jul 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0041361. URL <https://dx.plos.org/10.1371/journal.pone.0041361>.
- Anthony Rhoads and Kin Fai Au. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*, 13(5):278–289, oct 2015. ISSN 16720229. doi: 10.1016/j.gpb.2015.08.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S1672022915001345>.
- E. J. Rideout, L. Marshall, and S. S. Grewal. Drosophila RNA polymerase III repressor Maf1 controls body size and developmental timing by modulating tRNAiMet synthesis and systemic insulin signaling. *Proceedings of the National Academy of Sciences*, 109(4):1139–1144, jan 2012. ISSN 0027-8424. doi: 10.1073/pnas.1113311109. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1113311109>.
- Arthur D Riggs and Zhenggang Xiong. Methylation and epigenetic fidelity. *Proceedings of the National Academy of Sciences*, 101(1):4–5, jan 2004. ISSN 0027-8424. doi: 10.1073/pnas.0307781100. URL <http://www.ncbi.nlm.nih.gov/pubmed/14695893>.
- Pauline Rimmelé, Carolina L. Bigarella, Raymond Liang, Brigitte Izac, Rebeca Dieguez-Gonzalez, Gaetan Barbet, Michael Donovan, Carlo Brugnara, Julie M. Blander, David A. Sinclair, and Saghi Ghaffari. Aging-like phenotype and defective lineage specification in SIRT1-deleted hematopoietic stem and progenitor cells. *Stem cell reports*, 3(1):44–59, jul 2014. ISSN 2213-6711. doi: 10.1016/j.stemcr.2014.04.015. URL <http://www.ncbi.nlm.nih.gov/pubmed/25068121>.
- Matthew E Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47–e47, apr 2015. ISSN 1362-4962. doi: 10.1093/nar/gkv007. URL <http://www.ncbi.nlm.nih.gov/pubmed/25605792>.
- G A Romanov and B F Vanyushin. Methylation of reiterated sequences in mammalian DNAs. Effects of the tissue type, age, malignancy and hormonal induction. *Biochimica et biophysica acta*, 653(2):204–18, apr 1981. ISSN 0006-3002. URL <http://www.ncbi.nlm.nih.gov/pubmed/7225396>.
- Nathan R. Rose and Robert J. Klose. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1839(12):1362–1372, dec 2014. ISSN 18749399. doi: 10.1016/j.bbagr.2014.02.007. URL <http://dx.doi.org/10.1016/j.bbagr.2014.02.007>.
- Konrad L M Rudolph, Bianca M. Schmitt, Diego Villar, Robert J. White, John C. Marioni, Claudia Kutter, and Duncan T. Odom. Codon-Driven Translational Efficiency Is Stable across Diverse Mammalian Cell States. *PLoS genetics*, 12(5):e1006024, may 2016. ISSN 1553-7404. doi: 10.1371/journal.pgen.1006024. URL <http://www.ncbi.nlm.nih.gov/pubmed/27166679>.
- Vincenzo Russo, Robert Martienssen, and Arthur D Riggs. *Epigenetic Mechanisms of Gene Regulation*. Cold Spring Harbor Laboratory Press, 1996. ISBN 9780879694906. URL <https://books.google.co.uk/books/about/EpigeneticMechanismsofGeneRegulation.html>.

- html?id=chnwAAAAAMAJhttp://genesdev.cshlp.org/content/11/11/local/front-matter.pdfhttps://cshmonographs.org/index.php/monographs/issue/view/087969490.32.
- Michael C Sachs, Abigail Shoben, Gregory P Levin, Cassy Robinson-Cohen, Andrew N Hoofnagle, Nancy Swords-Jenny, Joachim H Ix, Matthew Budoff, Pamela L Lutsey, David S Siscovick, Bryan Kestenbaum, and Ian H de Boer. Estimating mean annual 25-hydroxyvitamin D concentrations from single measurements: the Multi-Ethnic Study of Atherosclerosis. *The American Journal of Clinical Nutrition*, 97(6):1243–1251, jun 2013. ISSN 0002-9165. doi: 10.3945/ajcn.112.054502. URL <https://academic.oup.com/ajcn/article/97/6/1243/4576828>.
- Dror Sagi, Roni Rak, Hila Gingold, Idan Adir, Gadi Maayan, Orna Dahan, Limor Broday, Yitzhak Pilpel, and Oded Rechavi. Tissue- and Time-Specific Expression of Otherwise Identical tRNA Genes. *PLOS Genetics*, 12(8):e1006264, aug 2016. ISSN 1553-7404. doi: 10.1371/journal.pgen.1006264. URL <http://www.ncbi.nlm.nih.gov/pubmed/27560950>.
- Juan Sandoval, Holger Heyn, Sebastian Moran, Jordi Serra-Musach, Miguel A Pujana, Marina Bibikova, and Manel Esteller. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics*, 6(6):692–702, jun 2011. ISSN 1559-2308. doi: 10.2217/epi.15.114. URL <http://www.ncbi.nlm.nih.gov/pubmed/21593595>.
- Serge Saxonov, Paul Berg, and Douglas L Brutlag. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences*, 103(5):1412–1417, jan 2006. ISSN 0027-8424. doi: 10.1073/pnas.0510310103. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.0510310103>.
- Leonard C. Schalkwyk, Emma L. Meaburn, Rebecca Smith, Emma L. Dempster, Aaron R. Jeffries, Matthew N. Davies, Robert Plomin, and Jonathan Mill. Allelic skewing of DNA methylation is widespread across the genome. *American journal of human genetics*, 86(2):196–212, feb 2010. ISSN 1537-6605. doi: 10.1016/j.ajhg.2010.01.014. URL <http://www.ncbi.nlm.nih.gov/pubmed/20159110>.
- Paul Schimmel. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nature reviews. Molecular cell biology*, 19(1):45–58, jan 2018. ISSN 1471-0080. doi: 10.1038/nrm.2017.77. URL <http://www.ncbi.nlm.nih.gov/pubmed/28875994>.
- Bianca M Schmitt, Konrad L M Rudolph, Panagiota Karagianni, Nuno A Fonseca, Robert J White, Iannis Talianidis, Duncan T Odom, John C Marioni, and Claudia Kutter. High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA-tRNA interface. *Genome research*, 24(11):1797–807, nov 2014a. ISSN 1549-5469. doi: 10.1101/gr.176784.114. URL <http://www.ncbi.nlm.nih.gov/pubmed/25122613>.
- Bianca M. Schmitt, Konrad L.M. Rudolph, Panagiota Karagianni, Nuno A. Fonseca, Robert J. White, Iannis Talianidis, Duncan T. Odom, John C. Marioni, and Claudia Kutter. High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA-tRNA interface. *Genome Research*, 24(11):1797–1807, nov 2014b. ISSN 1088-9051. doi: 10.1101/gr.176784.114. URL <http://www.ncbi.nlm.nih.gov/pubmed/25122613>.
- Laura Schramm. Recruitment of RNA polymerase III to its target promoters. *Genes & Development*, 16(20):2593–2620, oct 2002. ISSN 08909369. doi: 10.1101/gad.1018902. URL <http://www.genesdev.org/cgi/doi/10.1101/gad.1018902>.
- Bernd Schuettengruber, Daniel Chourrout, Michel Vervoort, Benjamin Leblanc, and Giacomo Cavalli. Genome Regulation by Polycomb and Trithorax Proteins. *Cell*, 128(4):735–745, feb 2007. ISSN 00928674. doi: 10.1016/j.cell.2007.02.009. URL <http://www.ncbi.nlm.nih.gov/pubmed/17320510>.
- Matthew D Schultz, Yupeng He, John W Whitaker, Manoj Hariharan, Eran A Mukamel, Danny Leung, Nisha Rajagopal, Joseph R Nery, Mark A Urich, Huaming Chen, Shin Lin, Yiying Lin, Inkyung Jung, Anthony D Schmitt, Siddarth Selvaraj, Bing Ren, Terrence J Sejnowski, Wei Wang, and Joseph R Ecker. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*, 523(7559):212–6, jul 2015. ISSN 1476-4687. doi: 10.1038/nature14465. URL <http://www.ncbi.nlm.nih.gov/pubmed/26030523>.

- Davis Sean and Paul S. Meltzer. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, 23(14):1846–1847, 2007. ISSN 13674803. doi: 10.1093/bioinformatics/btm254.
- David Serre, Byron H. Lee, and Angela H. Ting. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Research*, 38(2):391–399, jan 2010. ISSN 1362-4962. doi: 10.1093/nar/gkp992. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkp992>.
- Robert Shoemaker, Jie Deng, Wei Wang, and Kun Zhang. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome research*, 20(7):883–9, jul 2010. ISSN 1549-5469. doi: 10.1101/gr.104695.109. URL <http://www.ncbi.nlm.nih.gov/pubmed/20418490>.
- Sanjeev Shukla, Ersen Kavak, Melissa Gregory, Masahiko Imashimizu, Bojan Shutinoski, Mikhail Kashlev, Philipp Oberdoerffer, Rickard Sandberg, and Shalini Oberdoerffer. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, 479(7371):74–79, 2011. ISSN 00280836. doi: 10.1038/nature10442. URL <http://dx.doi.org/10.1038/nature10442>.
- Andrew J. Simpkin, Gibran Hemani, Matthew Suderman, Tom R. Gaunt, Oliver Lyttleton, Wendy L. Mcardle, Susan M. Ring, Gemma C. Sharp, Kate Tilling, Steve Horvath, Sonja Kunze, Annette Peters, Melanie Waldenberger, Cavin Ward-Caviness, Ellen A. Nohr, Thorkild I A Sørensen, Caroline L. Relton, and George Davey Smith. Prenatal and early life influences on epigenetic age in children: a study of mother–offspring pairs from two cohort studies. *Human Molecular Genetics*, 25(1):191–201, jan 2016. ISSN 0964-6906. doi: 10.1093/hmg/ddv456. URL <http://www.ncbi.nlm.nih.gov/pubmed/26546615>.
- Andrew J Simpkin, Laura D Howe, Kate Tilling, Tom R Gaunt, Oliver Lyttleton, Wendy L McAr- dle, Susan M Ring, Steve Horvath, George Davey Smith, and Caroline L Relton. The epigenetic clock and physical development during childhood and adolescence: longitudinal analysis from a UK birth cohort. *International Journal of Epidemiology*, 46(2):dyw307, jan 2017a. ISSN 0300-5771. doi: 10.1093/ije/dyw307. URL <http://www.ncbi.nlm.nih.gov/pubmed/28089957>.
- Andrew J. Simpkin, Matthew Suderman, and Laura D. Howe. Epigenetic clocks for gestational age: statistical and study design considerations. *Clinical Epigenetics*, 9(1):100, dec 2017b. ISSN 1868-7075. doi: 10.1186/s13148-017-0402-y. URL <http://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-017-0402-y>.
- Jared T Simpson, Rachael E Workman, P C Zuzarte, Matei David, L J Dursi, and Winston Timp. Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, 14(4):407–410, apr 2017. ISSN 1548-7091. doi: 10.1038/nmeth.4184. URL <http://www.nature.com/articles/nmeth.4184>.
- Ksenia Skvortsova, Elena Zotenko, Phuc-Loi Luu, Cathryn M. Gould, Shalima S. Nair, Susan J. Clark, and Clare Stirzaker. Comprehensive evaluation of genome-wide 5-hydroxymethylcytosine profiling approaches in human DNA. *Epigenetics & Chromatin*, 10(1):16, dec 2017. ISSN 1756-8935. doi: 10.1186/s13072-017-0123-7. URL <http://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/s13072-017-0123-7>.
- Roderick C. Slieker, Maarten van Iterson, René Luijk, Marian Beekman, Daria V Zhernakova, Matthijs H. Moed, Hailiang Mei, Michiel van Galen, Patrick Deelen, Marc Jan Bonder, Alexandra Zhernakova, André G. Uitterlinden, Ettje F. Tigchelaar, Coen D. A. Stehouwer, Casper G. Schalkwijk, Carla J. H. van der Kallen, Albert Hofman, Diana van Heemst, Eco J. de Geus, Jenny van Dongen, Joris Deelen, Leonard H. van den Berg, Joyce van Meurs, Rick Jansen, Peter A C ’t Hoen, Lude Franke, Cisca Wijmenga, Jan H. Veldink, Morris A. Swertz, Marleen M. J. van Greevenbroek, Cornelia M. van Duijn, Dorret I. Boomsma, BIOS consortium, P. Eline Slagboom, and Bastiaan T. Heijmans. Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. *Genome biology*, 17(1):191, sep 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-1053-6. URL <http://www.ncbi.nlm.nih.gov/pubmed/27654999>.

- Alicia K. Smith, Varun Kilaru, Mehmet Kocak, Lynn M. Almli, Kristina B. Mercer, Kerry J. Ressler, Frances A. Tylavsky, and Karen N. Conneely. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics*, 15(1):145, 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-145. URL <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2164-15-145>.
- Rick W A Smith, Cara Monroe, and Deborah A. Bolnick. Detection of Cytosine methylation in ancient DNA from five native american populations using bisulfite sequencing. *PLoS one*, 10(5):e0125344, may 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0125344. URL <http://www.ncbi.nlm.nih.gov/pubmed/26016479>.
- Michael B. Stadler, Rabih Murr, Lukas Burger, Robert Ivanek, Florian Lienert, Anne Schöler, Erik van Nimwegen, Christiane Wirbelauer, Edward J. Oakeley, Dimos Gaidatzis, Vijay K. Tiwari, and Dirk Schübeler. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 480(7378):490–5, dec 2011. ISSN 1476-4687. doi: 10.1038/nature10716. URL <http://www.ncbi.nlm.nih.gov/pubmed/22170606>.
- Thomas M Stubbs, Marc Jan Bonder, Anne-Katrien Stark, Felix Krueger, Ferdinand von Meyenn, Oliver Stegle, and Wolf Reik. Multi-tissue DNA methylation age predictor in mouse. *Genome Biology*, 18(1):68, dec 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1203-5. URL <http://www.ncbi.nlm.nih.gov/pubmed/28399939>.
- Ming Su, Dali Han, Jerome Boyd-Kirkup, Xiaoming Yu, and Jing-Dong J. Han. Evolution of Alu Elements toward Enhancers. *Cell Reports*, 7(2):376–385, apr 2014. ISSN 22111247. doi: 10.1016/j.celrep.2014.03.011. URL <http://dx.doi.org/10.1016/j.celrep.2014.03.011>.
- Luyang Sun, Ruofan Yu, and Weiwei Dang. Chromatin Architectural Changes during Cellular Senescence and Aging. *Genes*, 9(4):211, apr 2018. ISSN 2073-4425. doi: 10.3390/genes9040211. URL <http://www.mdpi.com/2073-4425/9/4/211>.
- Josephine E Sutcliffe, Timothy R P Brown, Simon J Allison, Pamela H Scott, and Robert J White. Retinoblastoma Protein Disrupts Interactions Required for RNA Polymerase III Transcription. *Molecular and Cellular Biology*, 20(24):9192–9202, dec 2000. ISSN 0270-7306. doi: 10.1128/MCB.20.24.9192-9202.2000. URL <http://mcb.asm.org/cgi/doi/10.1128/MCB.20.24.9192-9202.2000>.
- HE Syddall, A Aihie Sayer, EM Dennison, HJ Martin, DJP Barker, and C. Cooper. Cohort Profile: The Hertfordshire Cohort Study. *International Journal of Epidemiology*, 34(6):1234–1242, dec 2005. ISSN 1464-3685. doi: 10.1093/ije/dyi127. URL <http://academic.oup.com/ije/article/34/6/1234/707357/Cohort-Profile-The-Hertfordshire-Cohort-Study>.
- Oluwatosin Taiwo, Gareth A. Wilson, Tiffany Morris, Stefanie Seisenberger, Wolf Reik, Daniel Pearce, Stephan Beck, and Lee M. Butcher. Methylome analysis using MeDIP-seq with low DNA concentrations. *Nature Protocols*, 7(4):617–636, apr 2012. ISSN 1754-2189. doi: 10.1038/nprot.2012.012. URL <http://www.nature.com/articles/nprot.2012.012>.
- Nektarios Tavernarakis. Ageing and the regulation of protein synthesis: a balancing act? *Trends in Cell Biology*, 18(5):228–235, may 2008. ISSN 09628924. doi: 10.1016/j.tcb.2008.02.004. URL <https://linkinghub.elsevier.com/retrieve/pii/S0962892408000792>.
- Andrew E Teschendorff and Shijie C Zheng. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics*, 9(5):757–768, may 2017. ISSN 1750-1911. doi: 10.2217/epi-2016-0153. URL <https://www.futuremedicine.com/doi/10.2217/epi-2016-0153>.
- Andrew E. Teschendorff, Usha Menon, Aleksandra Gentry-Maharaj, Susan J. Ramus, Daniel J. Weisenberger, Hui Shen, Mihaela Campan, Houtan Noushmehr, Christopher G. Bell, A. Peter Maxwell, David A. Savage, Elisabeth Mueller-Holzner, Christian Marth, Gabrijela Kocjan, Simon A. Gayther, Allison Jones, Stephan Beck, Wolfgang Wagner, Peter W. Laird, Ian J. Jacobs, and Martin Widschwendter. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome research*, 20(4):440–6, apr 2010.

- ISSN 1549-5469. doi: 10.1101/gr.103606.109. URL <http://www.ncbi.nlm.nih.gov/pubmed/20219944>.
- Andrew E. Teschendorff, Joanna Zhuang, and Martin Widschwendter. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics*, 27(11):1496–1505, jun 2011. ISSN 1460-2059. doi: 10.1093/bioinformatics/btr171. URL <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr171>.
- Michael J Thompson, Bridgett VonHoldt, Steve Horvath, and Matteo Pellegrini. An epigenetic aging clock for dogs and wolves. *Aging*, 9(3):1055–1068, mar 2017. ISSN 1945-4589. doi: 10.18632/aging.101211. URL <http://www.ncbi.nlm.nih.gov/pubmed/28373601>.
- Adrian Gabriel Torres, Oscar Reina, Camille Stephan-Otto Attolini, and Lluís Ribas de Pouplana. Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. *Proceedings of the National Academy of Sciences*, 116(17):201821120, 2019. ISSN 0027-8424. doi: 10.1073/pnas.1821120116. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1821120116>.
- Hong Tran, Jacob Porter, Ming-an Sun, Hehuang Xie, and Liqing Zhang. Objective and Comprehensive Evaluation of Bisulfite Short Read Mapping Tools. *Advances in Bioinformatics*, 2014:1–11, 2014. ISSN 1687-8027. doi: 10.1155/2014/472045. URL <http://www.hindawi.com/journals/abi/2014/472045/>.
- Elisabetta Ullu and Christian Tschudi. Alu sequences are processed 7SL RNA genes. *Nature*, 312(5990):171–172, nov 1984. ISSN 0028-0836. doi: 10.1038/312171a0. URL <http://www.nature.com/articles/312171a0>.
- Toshikazu Ushijima, Naoko Watanabe, Kimiko Shimizu, Kazuaki Miyamoto, Takashi Sugimura, and Atsushi Kaneda. Decreased fidelity in replicating CpG methylation patterns in cancer cells. *Cancer research*, 65(1):11–7, jan 2005. ISSN 0008-5472. URL <http://www.ncbi.nlm.nih.gov/pubmed/15665274>.
- Kevin Van Bortle, Douglas H. Phanstiel, and Michael P. Snyder. Topological organization and dynamic regulation of human tRNA genes during macrophage differentiation. *Genome Biology*, 18(1):180, dec 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1310-3. URL <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1310-3>.
- Dhaval Varshney, Jana Vavrova-Anderson, Andrew J Oler, Victoria H Cowling, Bradley R Cairns, and Robert J White. SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. *Nature communications*, 6(1):6569, mar 2015. ISSN 2041-1723. doi: 10.1038/ncomms7569. URL <http://www.ncbi.nlm.nih.gov/pubmed/25798578>.
- Paula M. Vertino, Jennifer A. Sekowski, Jennifer M. Coll, Nancy Applegren, Suhua Han, Robert J. Hickey, and Linda H. Malkas. DNMT1 is a component of a multiprotein DNA replication complex. *Cell cycle (Georgetown, Tex.)*, 1(6):416–23, nov 2002. ISSN 1538-4101. doi: 10.4161/cc.1.6.270. URL <http://www.ncbi.nlm.nih.gov/pubmed/12548018>.
- H. T. Viljakainen, E. Saarnio, T. Hytinantti, M. Miettinen, H. Surcel, O. Mäkitie, S. Andersson, K. Laitinen, and C. Lamberg-Allardt. Maternal Vitamin D Status Determines Bone Variables in the Newborn. *The Journal of Clinical Endocrinology & Metabolism*, 95(4):1749–1757, apr 2010. ISSN 0021-972X. doi: 10.1210/jc.2009-1391. URL <https://academic.oup.com/jcem/article-lookup/doi/10.1210/jc.2009-1391>.
- H. T. Viljakainen, T. Korhonen, T. Hytinantti, E. K. A. Laitinen, S. Andersson, O. Mäkitie, and C. Lamberg-Allardt. Maternal vitamin D status affects bone growth in early childhood—a prospective cohort study. *Osteoporosis International*, 22(3):883–891, mar 2011. ISSN 0937-941X. doi: 10.1007/s00198-010-1499-4. URL <http://link.springer.com/10.1007/s00198-010-1499-4>.

- Philip Voigt, Gary LeRoy, William J. Drury, Barry M. Zee, Jinsook Son, David B. Beck, Nicolas L. Young, Benjamin A. Garcia, and Danny Reinberg. Asymmetrically modified nucleosomes. *Cell*, 151(1):181–93, sep 2012. ISSN 1097-4172. doi: 10.1016/j.cell.2012.09.002. URL <http://www.ncbi.nlm.nih.gov/pubmed/23021224>.
- C. H. Waddington. The epigenotype. 1942. *International journal of epidemiology*, 41(1):10–3, feb 2012. ISSN 1464-3685. doi: 10.1093/ije/dyr184. URL <http://www.ncbi.nlm.nih.gov/pubmed/22186258>.
- Simone Wahl, Alexander Drong, Benjamin Lehne, Marie Loh, William R Scott, Sonja Kunze, Pei-Chien Tsai, Janina S Ried, Weihua Zhang, Youwen Yang, Sili Tan, Giovanni Fiorito, Lude Franke, Simonetta Guarnera, Silva Kasela, Jennifer Kriebel, Rebecca C Richmond, Marco Adamo, Uzma Afzal, Mika Ala-Korpela, Benedetta Albetti, Ole Ammerpohl, Jane F Apperley, Marian Beekman, Pier Alberto Bertazzi, S. Lucas Black, Christine Blancher, Marc-Jan Bonder, Mario Brosch, Maren Carstensen-Kirberg, Anton J. M. de Craen, Simon de Lusiganan, Abbas Dehghan, Mohamed Elkalaawy, Krista Fischer, Oscar H Franco, Tom R Gaunt, Jochen Hampe, Majid Hashemi, Aaron Isaacs, Andrew Jenkinson, Sujeet Jha, Norihiro Kato, Vittorio Krogh, Michael Laffan, Christa Meisinger, Thomas Meitinger, Zuan Yu Mok, Valeria Motta, Hong Kiat Ng, Zacharoula Nikolakopoulou, Georgios Ntelopoulos, Salvatore Panico, Natalia Pervjakova, Holger Prokisch, Wolfgang Rathmann, Michael Roden, Federica Rota, Michelle Ann Rozario, Johanna K Sandling, Clemens Schafmayer, Katharina Schramm, Reiner Siebert, P. Eline Slagboom, Pasi Soininen, Lisette Stolk, Konstantin Strauch, E-Shyong Tai, Letizia Tarantini, Barbara Thorand, Ettje F Tigchelaar, Rosario Tumino, Andre G Uitterlinden, Cornelia van Duijn, Joyce B. J. van Meurs, Paolo Vineis, Ananda Rajitha Wickremasinghe, Cisca Wijmenga, Tsun-Po Yang, Wei Yuan, Alexandra Zhernakova, Rachel L Batterham, George Davey Smith, Panos Deloukas, Bastiaan T Heijmans, Christian Herder, Albert Hofman, Cecilia M Lindgren, Lili Milani, Pim van der Harst, Annette Peters, Thomas Illig, Caroline L Relton, Melanie Waldenberger, Marjo-Riitta Järvelin, Valentina Bollati, Richie Soong, Tim D Spector, James Scott, Mark I McCarthy, Paul Elliott, Jordana T Bell, Giuseppe Matullo, Christian Gieger, Jaspal S Kooner, Harald Grallert, and John C Chambers. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature*, 541(7635):81–86, jan 2017. ISSN 1476-4687. doi: 10.1038/nature20784. URL <http://www.ncbi.nlm.nih.gov/pubmed/28002404>.
- Michelle C. Ward, Michael D. Wilson, Nuno L. Barbosa-Moraes, Dominic Schmidt, Rory Stark, Qun Pan, Petra C. Schwalie, Suraj Menon, Margus Lukk, Stephen Watt, David Thybert, Claudia Kutter, Kristina Kirschner, Paul Flückeck, Benjamin J. Blencowe, and Duncan T. Odom. Latent Regulatory Potential of Human-Specific Repetitive Elements. *Molecular Cell*, 49(2): 262–272, jan 2013. ISSN 10972765. doi: 10.1016/j.molcel.2012.11.013. URL <http://dx.doi.org/10.1016/j.molcel.2012.11.013>.
- Christopher M Weber and Steven Henikoff. Histone variants: dynamic punctuation in transcription. *Genes & development*, 28(7):672–82, apr 2014. ISSN 1549-5477. doi: 10.1101/gad.238873.114. URL <http://www.ncbi.nlm.nih.gov/pubmed/24696452>.
- Michael Weber, Jonathan J Davies, David Wittig, Edward J Oakeley, Michael Haase, Wan L Lam, and Dirk Schübeler. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics*, 37(8):853–862, aug 2005. ISSN 1061-4036. doi: 10.1038/ng1598. URL <http://www.ncbi.nlm.nih.gov/pubmed/16007088>.
- Carola Ingrid Weidner, Qiong Lin, Carmen Maike Koch, Lewin Eisele, Fabian Beier, Patrick Ziegler, Dirk Olaf Bauerschlag, Karl-Heinz Jöckel, Raimund Erbel, Thomas Walter Mühleisen, Martin Zenke, Tim Henrik Brümmendorf, and Wolfgang Wagner. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome biology*, 15(2):R24, feb 2014. ISSN 1474-760X. doi: 10.1186/gb-2014-15-2-r24. URL <http://www.ncbi.nlm.nih.gov/pubmed/24490752>.

- V L Wilson and P A Jones. DNA methylation decreases in aging but not in immortal cells. *Science (New York, N.Y.)*, 220(4601):1055–7, jun 1983. ISSN 0036-8075. URL <http://www.ncbi.nlm.nih.gov/pubmed/6844925>.
- V. L. Wilson, R. A. Smith, S. Ma, and R. G. Cutler. Genomic 5-methyldeoxycytidine decreases with age. *The Journal of biological chemistry*, 262(21):9948–51, jul 1987. ISSN 0021-9258. URL <http://www.ncbi.nlm.nih.gov/pubmed/3611071>.
- Tao P. Wu, Tao Wang, Matthew G. Seetin, Yongquan Lai, Shijia Zhu, Kaixuan Lin, Yifei Liu, Stephanie D. Byrum, Samuel G. Mackintosh, Mei Zhong, Alan Tackett, Guilin Wang, Lawrence S. Hon, Gang Fang, James a. Swenberg, and Andrew Z. Xiao. DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature*, 532(7599):329–33, apr 2016. ISSN 1476-4687. doi: 10.1038/nature17640. URL <http://www.ncbi.nlm.nih.gov/pubmed/27027282>.
- Xiaoji Wu and Yi Zhang. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics*, 18(9):517–534, may 2017. ISSN 1471-0056. doi: 10.1038/nrg.2017.33. URL <http://www.nature.com/doifinder/10.1038/nrg.2017.33>.
- Mingchao Xie, Chibo Hong, Bo Zhang, Rebecca F Lowdon, Xiaoyun Xing, Daofeng Li, Xin Zhou, Hyung Joo Lee, Cecile L Maire, Keith L Ligon, Philippe Gascard, Mahvash Sigaroudinia, Thea D Tlsty, Theresa Kadlecak, Arthur Weiss, Henriette O’Geen, Peggy J Farnham, Pamela A F Madden, Andrew J Mungall, Angela Tam, Baljit Kamoh, Stephanie Cho, Richard Moore, Martin Hirst, Marco A Marra, Joseph F Costello, and Ting Wang. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nature Genetics*, 45(7):836–841, jul 2013. ISSN 1061-4036. doi: 10.1038/ng.2649. URL <http://www.ncbi.nlm.nih.gov/pubmed/23708189>.
- Wei-Lin Xu, Ye Yang, Yi-Dan Wang, Liang-Hu Qu, and Ling-Ling Zheng. Computational Approaches to tRNA-Derived Small RNAs. *Non-Coding RNA*, 3(1):2, jan 2017. ISSN 2311-553X. doi: 10.3390/ncrna3010002. URL <http://www.mdpi.com/2311-553X/3/1/2>.
- Zhen Yang, Andrew Wong, Diana Kuh, Dirk S. Paul, Vardhman K. Rakyan, R. David Leslie, Shijie C. Zheng, Martin Widschwendter, Stephan Beck, and Andrew E. Teschendorff. Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biology*, 17(1):205, dec 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-1064-3. URL <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1064-3>.
- Yimeng Yin, Ekaterina Morgunova, Artuu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, Pratyush K. Das, Teemu Kivioja, Kashyap Dave, Fan Zhong, Kazuhiro R. Nitta, Minna Taipale, Alexander Popov, Paul A. Ginno, Silvia Domcke, Jian Yan, Dirk Schübeler, Charles Vinson, and Jussi Taipale. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science (New York, N.Y.)*, 356(6337):eaaj2239, may 2017. ISSN 1095-9203. doi: 10.1126/science.aaaj2239. URL <http://www.ncbi.nlm.nih.gov/pubmed/28473536>.
- Shaza B Zaghloul, Masha Al-Shafai, Wadha A Al Muftah, Pankaj Kumar, Mario Falchi, and Karsten Suhre. Association of DNA methylation with age, gender, and smoking in an Arab population. *Clinical Epigenetics*, 7(1):6, 2015. ISSN 1868-7083. doi: 10.1186/s13148-014-0040-6. URL <http://www.ncbi.nlm.nih.gov/pubmed/25663950>.
- Yan Zhang, Jan Hapala, Hermann Brenner, and Wolfgang Wagner. Individual CpG sites that are associated with age and life expectancy become hypomethylated upon aging. *Clinical Epigenetics*, 9(1):9, dec 2017. ISSN 1868-7075. doi: 10.1186/s13148-017-0315-9. URL <http://clincalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-017-0315-9>.
- Lei Zhao, Ming-an Sun, Zejuan Li, Xue Bai, Miao Yu, Min Wang, Liji Liang, Xiaojian Shao, Stephen Arnovitz, Qianfei Wang, Chuan He, Xuemei Lu, Jianjun Chen, and Hehuang Xie. The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. *Genome Research*, 24(8):1296–1307, aug 2014. ISSN 1088-9051. doi: 10.1101/gr.163147.113. URL <http://genome.cshlp.org/lookup/doi/10.1101/gr.163147.113>.

- Wanding Zhou, Peter W. Laird, and Hui Shen. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Research*, 45(4):gkw967, oct 2016. ISSN 0305-1048. doi: 10.1093/nar/gkw967. URL <http://www.ncbi.nlm.nih.gov/pubmed/27924034>.
- Kun Zhu, Andrew JO Whitehouse, Prue H Hart, Merci Kusel, Jenny Mountain, Stephen Lye, Craig Pennell, and John P Walsh. Maternal Vitamin D Status During Pregnancy and Bone Mass in Offspring at 20 Years of Age: A Prospective Cohort Study. *Journal of Bone and Mineral Research*, 29(5):1088–1095, may 2014. ISSN 08840431. doi: 10.1002/jbmр.2138. URL <http://doi.wiley.com/10.1002/jbmр.2138>.
- Michael J Ziller, Hongcang Gu, Fabian Müller, Julie Donaghey, Linus T-Y Tsai, Oliver Kohlbacher, Philip L De Jager, Evan D Rosen, David A Bennett, Bradley E Bernstein, Andreas Gnirke, and Alexander Meissner. Charting a dynamic DNA methylation landscape of the human genome. *Nature*, 500(7463):477–81, aug 2013. ISSN 1476-4687. doi: 10.1038/nature12433. URL <http://www.ncbi.nlm.nih.gov/pubmed/23925113>.
- Zheng Zuo, Basab Roy, Yiming Kenny Chang, David Granas, and Gary D Stormo. Measuring quantitative effects of methylation on transcription factor–DNA binding affinity. *Science Advances*, 3(11):eaa01799, nov 2017. ISSN 2375-2548. doi: 10.1126/sciadv.aa01799. URL <http://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aa01799>.