Universidad de Oviedo

Programa de Doctorado “Biomedicina y Oncología Molecular”

“Epigenomic links between aging, development and cancer”

TESIS DOCTORAL

Raúl Fernández Pérez
30/09/2022

Universidad de Oviedo

Tesis Doctoral

“Epigenomic links between aging, development and cancer”

Raúl Fernández Pérez
Doctorando

Mario Fernández Fraga
Agustín Fernández Fernández
Directores

José María Pérez Freije
Tutor

Universidad de Oviedo

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: Conexiones epigenómicas entre envejecimiento, desarrollo y cáncer	Inglés: Epigenomic links between aging, development and cancer

2.- Autor	
Nombre: Raúl Fernández Pérez	DNI/Pasaporte/NIE:
Programa de Doctorado: Biomedicina y Oncología Molecular	
Órgano responsable: Centro Internacional de Postgrado	

RESUMEN (en español)

Los mecanismos epigenéticos, como la metilación del ADN o la modificación post-traduccional de las histonas, regulan la interacción del genoma con el ambiente y juegan papeles importantes en el control de procesos biológicos, así como en el desarrollo de enfermedades. El fenómeno universal del envejecimiento está asociado a la aparición de muchas enfermedades, como el cáncer y la neurodegeneración. Todos estos procesos manifiestan alteraciones epigenéticas, y debido a que el envejecimiento y patologías derivadas a menudo muestran similitudes en sus modificaciones epigenéticas, se ha hipotetizado que los mecanismos epigenéticos pueden ser nexos causales entre estos procesos. De esta manera, tanto envejecimiento como cáncer presentan alteraciones análogas en la metilación del ADN, pero la naturaleza concreta del contexto epigenómico de estas alteraciones es todavía desconocida, especialmente con respecto a la pérdida de metilación del ADN. Adicionalmente, las marcas epigenéticas pueden también ser utilizadas para la definición de biomarcadores para la prevención o detección de enfermedades asociadas al envejecimiento, así como para descubrir dianas moleculares para su tratamiento.

En esta tesis doctoral, hemos utilizado datos públicos de estudios ómicos a gran escala, así como generado datos epigenómicos a partir de cohortes humanas y modelos murinos, para caracterizar y comparar las características de las alteraciones en la metilación del ADN y su contexto epigenómico en envejecimiento y enfermedades asociadas. En particular, hemos explorado las dinámicas epigenómicas del cáncer, el desarrollo en etapas tempranas de la vida, y el deterioro cognitivo. Nuestros resultados indican que la hipometilación del ADN asociada a envejecimiento y cáncer ocurre en contextos de cromatina distintos, en múltiples tejidos, sugiriendo que la relación entre las alteraciones de metilación del ADN en envejecimiento y cáncer es más compleja que lo anteriormente pensado. Además, observamos que las características epigenéticas específicas de cada enfermedad están conservadas en modelos humanos y murinos de envejecimiento y cáncer cerebral. Adicionalmente, determinamos que los cambios de metilación asociados al desarrollo que ocurren durante los

primeros años de vida son similares a aquellos asociados al envejecimiento, pero de mayor número y magnitud. Con respecto a las conexiones epigenéticas entre envejecimiento y deterioro cognitivo, observamos que las alteraciones de metilación del ADN pueden aparecer en la sangre de sujetos humanos años antes de la manifestación clínica de síntomas patológicos. Finalmente, utilizando modelos murinos de enriquecimiento ambiental, demostramos cómo la estimulación cognitiva puede revertir características moleculares típicas del proceso de envejecimiento.

En conjunto, nuestros resultados revelan cómo el estudio de alteraciones epigenéticas relacionadas con el envejecimiento nos ayuda a entender la etiología de enfermedades asociadas, nos proporciona nuevas vías para el desarrollo de biomarcadores o la detección de dianas moleculares para su tratamiento, y nos muestra cómo intervenciones sobre el estilo de vida pueden ser utilizadas para prevenir deterioro asociado al envejecimiento.

RESUMEN (en Inglés)

Epigenetic mechanisms, such as DNA methylation and histone post-translational modification, regulate the interaction of the genome with the environment and play major roles in the control of biological processes as well as the development of disease. The universal phenomenon of aging is linked to the emergence of many diseases, including cancer and neurodegeneration. All of these processes manifest associated epigenetic alterations, and because aging and aging-related pathologies often display similarities in their epigenetic modifications, it is thought that epigenetic mechanisms could be causal links between these processes. For instance, both aging and cancer present apparently analogous DNA methylation alterations, but the precise nature of the epigenomic context of these changes is to date unclear, especially regarding the characteristics of DNA methylation loss. Moreover, epigenetic marks may also be used to define biomarkers for the prevention or detection of aging-associated diseases, as well as to discover molecular targets for their treatment.

In this doctoral thesis, we have taken advantage of publicly-available large-scale omic studies and also generated epigenetic data from human cohorts and murine models to characterize and compare the features of DNA methylation alterations and their epigenomic context in aging and aging-associated disease, using state-of-the-art computational methods. In particular, we explored the epigenomic dynamics of cancer, early-life developmental processes and cognitive decline. Our results indicate that aging- and cancer-associated DNA hypomethylation occurs at different chromatin contexts across multiple tissues, suggesting that the relationship between DNA methylation alterations in aging and cancer is more complex than previously thought. Moreover, we observed that disease-specific epigenetic features are conserved between human and mouse models of brain aging and cancer. We additionally determined that early-life, development-associated DNA methylation changes are similar to aging-associated alterations but occur in larger number and magnitude. With regards to the epigenetic links between aging and aging-associated cognitive decline, we observed that DNA methylation alterations can

Universidad de Oviedo

appear in the peripheral blood of human subjects years before the manifestation of pathological clinical symptoms. Finally, using murine models of environmental enrichment, we showed how cognitive stimulation can revert specific molecular features of the aging process.

Overall, our findings reveal how the study of aging-associated epigenetic alterations helps us understand the etiology of aging-associated diseases, provides us with new avenues regarding the development of biomarkers or molecular targets for their treatment, and informs us of how lifestyle interventions may be used to prevent aging-associated decline.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN BIOMEDICINA Y ONCOLOGÍA MOLECULAR

Acknowledgements

This doctoral thesis was supported by the Severo Ochoa program (grant number BP17-114) from Fundación Española para la Ciencia y Tecnología (FECYT) and Gobierno del Principado de Asturias. In addition, several institutions contributed via direct funding or institutional support to the work here described:

- Universidad de Oviedo.
- Centro de Investigación en Nanomateriales y Nanotecnología - Consejo Superior de Investigaciones Científicas (CINN-CSIC).
- Instituto de Investigación Sanitaria de Principado de Asturias (ISPA).
- Fundación para la investigación y la innovación biosanitaria del Principado de Asturias (FINBA).
- Instituto Universitario de Oncología del Principado de Asturias (IUOPA, supported by Obra Social Cajastur-Liberbank).
- Centro de Investigación Biomédica en Red de Enfermedades Raras – Instituto de Saludo Carlos III (CIBER-ISCIII).
- Fundación para la Investigación del Hospital Clínico de la Comunidad Valenciana (INCLIVA).
- Asociación Española Contra el Cáncer (AECC).
- Asociación de Familias de Niños con Cáncer del Principado de Asturias "Galbán".
- Fundación Centro de Investigación Enfermedades Neurológicas (FCIEN).
- Instituto Cajal - Consejo Superior de Investigaciones Científicas (IC-CSIC).
- Biobanks: Red Nacional de Biobancos (RNBB), biobanco HUB-ICO-IDIBELL, biobanco del Instituto de Investigación Sanitaria Galicia Sur (IISGS) y biobanco de Aragón en el Instituto Aragonés de Ciencias de la Salud (IACS).

We also convey our gratitude to the administrative staff of the listed institutions.

Finally, we acknowledge the generous contribution of the patients involved in the studies here described, as well as that of the research animals which participated in the experiments.

TABLE OF CONTENTS

FIGURES AND BOXES

LIST OF FIGURES

LIST OF BOXES

ABBREVIATIONS AND SYMBOLS

5caC – 5-carboxylcytosine

5fC – 5-formylcytosine

5hmC – 5-hydroxymethylcytosine

5mC – 5-methylcytosine

6mA – N6-methyladenine

A – adenine

ADP – adenosine diphosphate

ATAC-seq – assay for transposase-accessible chromatin using sequencing

A^{vy} – agouti viable gene

BC – before Christ

bHLH – basic helix-loop-helix

BS-seq – bisulfite sequencing

bp – base pair

C – cytosine

ChIP-seq – chromatin immunoprecipitation and sequencing

CIMP – CpG island methylator phenotype

CpG – 5'-C-phosphate-G-3'

DNA – deoxyribonucleic acid

DNMT – DNA methyltransferase

ESC – embryonic stem cell

ETS – E26 transformation-specific

G – guanine

GBM – glioblastoma

HAT – histone acetyltransferase

HDAC – histone deacetylase

HDM – histone demethylase

HMG – high-mobility group

HMM – hidden Markov model

HMT – histone methyltransferase

IAP – intracisternal A-particle

IHEC – International Human Epigenome Consortium

KAT – lysine acetylase

LGG – lower-grade glioma

MDB – methyl-CpG-binding domain

mRNA – messenger ribonucleic acid

ncRNA – non-coding ribonucleic acid

NDR – nucleosome-depleted region

NFR – nucleosome-free region

nm – nanometer

NR – nuclear receptor

PCA – principal component analysis

RNA – ribonucleic acid

RNA-seq – RNA sequencing

RRBS – reduced-representation bisulfite sequencing

rRNA – ribosomal RNA

SAM – *S*-adenosylmethionine

sdCpG – species-discordant CpG

SRA – SET- and Ring finger-associated domain

srpRNA – signal recognition particle RNA

T – thymine

TCGA – The Cancer Genome Atlas

TDG – thymine-DNA glycosylase

TF – transcription factor

TFBS – transcription factor binding site

tRNA – transfer RNA

TSS – transcription start site

PTM – post-translational modification

WGBS – whole-genome bisulfite sequencing

ZGA – zygotic genome activation

Resumen

Los mecanismos epigenéticos, como la metilación del ADN o la modificación post-traduccional de las histonas, regulan la interacción del genoma con el ambiente y juegan papeles importantes en el control de procesos biológicos, así como en el desarrollo de enfermedades. El fenómeno universal del envejecimiento está asociado a la aparición de muchas enfermedades, como el cáncer y la neurodegeneración. Todos estos procesos manifiestan alteraciones epigenéticas, y debido a que el envejecimiento y patologías derivadas a menudo muestran similitudes en sus modificaciones epigenéticas, se ha hipotetizado que los mecanismos epigenéticos pueden ser nexos causales entre estos procesos. De esta manera, tanto envejecimiento como cáncer presentan alteraciones análogas en la metilación del ADN, pero la naturaleza concreta del contexto epigenómico de estas alteraciones es todavía desconocida, especialmente con respecto a la pérdida de metilación del ADN. Adicionalmente, las marcas epigenéticas pueden también ser utilizadas para la definición de biomarcadores para la prevención o detección de enfermedades asociadas al envejecimiento, así como para descubrir dianas moleculares para su tratamiento.

En esta tesis doctoral, hemos utilizado datos públicos de estudios ómicos a gran escala, así como generado datos epigenómicos a partir de cohortes humanas y modelos murinos, para caracterizar y comparar las características de las alteraciones en la metilación del ADN y su contexto epigenómico en envejecimiento y enfermedades asociadas. En particular, hemos explorado las dinámicas epigenómicas del cáncer, el desarrollo en etapas tempranas de la vida, y el deterioro cognitivo. Nuestros resultados indican que la hipometilación del ADN asociada a envejecimiento y cáncer ocurre en contextos de cromatina distintos, en múltiples tejidos, sugiriendo que la relación entre las alteraciones de metilación del ADN en envejecimiento y cáncer es más compleja que lo anteriormente pensado. Además, observamos que las características epigenéticas específicas de cada enfermedad están conservadas en modelos humanos y murinos de

envejecimiento y cáncer cerebral. Adicionalmente, determinamos que los cambios de metilación asociados al desarrollo que ocurren durante los primeros años de vida son similares a aquellos asociados al envejecimiento, pero de mayor número y magnitud. Con respecto a las conexiones epigenéticas entre envejecimiento y deterioro cognitivo, observamos que las alteraciones de metilación del ADN pueden aparecer en la sangre de sujetos humanos años antes de la manifestación clínica de síntomas patológicos. Finalmente, utilizando modelos murinos de enriquecimiento ambiental, demostramos cómo la estimulación cognitiva puede revertir características moleculares típicas del proceso de envejecimiento.

En conjunto, nuestros resultados revelan cómo el estudio de alteraciones epigenéticas relacionadas con el envejecimiento nos ayuda a entender la etiología de enfermedades asociadas, nos proporciona nuevas vías para el desarrollo de biomarcadores o la detección de dianas moleculares para su tratamiento, y nos muestra cómo intervenciones sobre el estilo de vida pueden ser utilizadas para prevenir deterioro asociado al envejecimiento.

Abstract

Epigenetic mechanisms, such as DNA methylation and histone post-translational modification, regulate the interaction of the genome with the environment and play major roles in the control of biological processes as well as the development of disease. The universal phenomenon of aging is linked to the emergence of many diseases, including cancer and neurodegeneration. All of these processes manifest associated epigenetic alterations, and because aging and aging-related pathologies often display similarities in their epigenetic modifications, it is thought that epigenetic mechanisms could be causal links between these processes. For instance, both aging and cancer present apparently analogous DNA methylation alterations, but the precise nature of the epigenomic context of these changes is to date unclear, especially regarding the characteristics of DNA methylation loss. Moreover, epigenetic marks may also be used to define biomarkers for the prevention or detection of aging-associated diseases, as well as to discover molecular targets for their treatment.

In this doctoral thesis, we have taken advantage of publicly-available large-scale omic studies and also generated epigenetic data from human cohorts and murine models to characterize and compare the features of DNA methylation alterations and their epigenomic context in aging and aging-associated disease, using state-of-the-art computational methods. In particular, we explored the epigenomic dynamics of cancer, early-life developmental processes and cognitive decline. Our results indicate that aging- and cancer-associated DNA hypomethylation occurs at different chromatin contexts across multiple tissues, suggesting that the relationship between DNA methylation alterations in aging and cancer is more complex than previously thought. Moreover, we observed that disease-specific epigenetic features are conserved between human and mouse models of brain aging and cancer. We additionally determined that early-life, development-associated DNA methylation changes are similar to aging-associated alterations but occur in larger number and magnitude. With regards to the

epigenetic links between aging and aging-associated cognitive decline, we observed that DNA methylation alterations can appear in the peripheral blood of human subjects years before the manifestation of pathological clinical symptoms. Finally, using murine models of environmental enrichment, we showed how cognitive stimulation can revert specific molecular features of the aging process.

Overall, our findings reveal how the study of aging-associated epigenetic alterations helps us understand the etiology of aging-associated diseases, provides us with new avenues regarding the development of biomarkers or molecular targets for their treatment, and informs us of how lifestyle interventions may be used to prevent aging-associated decline.

Introduction

1. THE BIRTH OF EPIGENETICS: FROM ONTOGENY TO MOLECULAR BIOLOGY

The modern field of epigenetics has deep scientific and philosophical roots which go back to, at least, the "Hippocratic Corpus" associated with Hippocrates' (460 BC–370 BC) works and Aristotle's (384–322 BC) "On the Generation of Animals". In these early Western opuses, physician and polymath laid the foundations of developmental biology by tackling a fundamental question: how does life originate and develop from preexisting life? The attempt to understand this biological mystery led to the birth of two opposed theories which would battle, evolve, and intermingle up until contemporary times: "epigenesis" and "preformationism".

From the point of view of epigenesis—termed coined by William Harvey (1578–1657) (Harvey 1651)—an organism developed from a relatively homogeneous, unorganized zygote by a process of differentiation, an "addition of parts"—hence the use of the prefix "epi-": above genesis. On the other hand, preformationism held that the individual was more or less already formed and had mainly to grow in size during its development (Nicoglou & Wolfe 2018) (**FIGURE 1**).

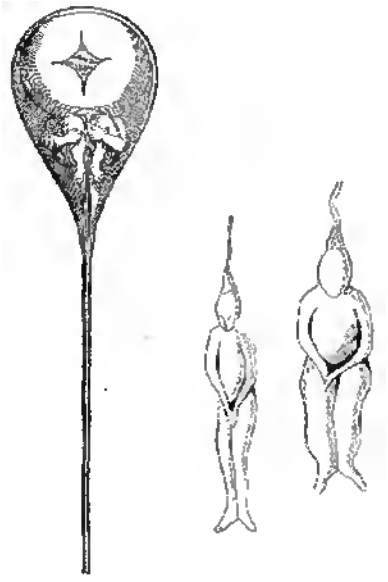


Figure 1. Drawing of a "preformed" individual within human sperm, by Nicolaas Hartsoeker (1656-1725) (public domain).

As science progressed, these concepts evolved to incorporate new knowledge so that by the mid-20th century the question had been narrowed down. The discovery of chromosomes by Walther Flemming (1843–1905) revolutionized genetics (Flemming 1882) and opened the door to understanding that genetic material residing in chromosomes was responsible for carrying biological information (Morgan 1911), in the form of "genes" (Johannsen 1909) with precise physical locations (Sturtevant 1913). It was in this scenario, prior to the identification of deoxyribonucleic acid (DNA) as the molecule comprising genetic material (Avery et al. 1944), that Conrad Hall Waddington (1905–1975) developed the idea of "epigenetics" (Waddington 1940; Waddington 1942). The British biologist sought to build a bridge between developmental biology and genetics by considering the available knowledge at the time (Nicoglou & Wolfe 2018).

Particularly in the contexts of evolution and embryonic development, the concept of epigenetics as established by Waddington referred to those developmental processes which, starting from the genotype, gave rise to the phenotype (Tronick & Hunter 2016) (**FIGURE 2**). Thus, the fundamental "epigenesis question" could now be read as: how can the unicellular zygote, with its single genotype, give rise to multiple cells with different phenotypes?—i.e., there must be a degree of "phenotypic plasticity" evoked from the genotype. In this setting, Waddington already anticipated the importance of the influence of external, environmental stimuli on eliciting different genetic responses, as is demonstrated in his research on "genetic assimilation" (Hall 1992).

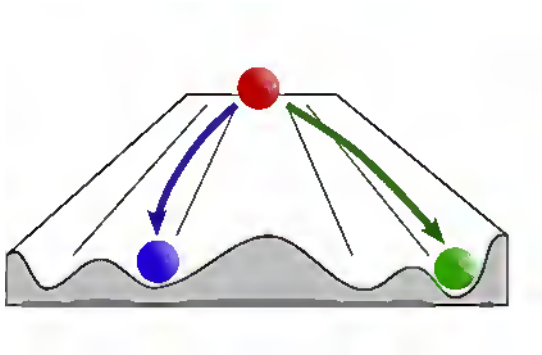


Figure 2. Graphical representation of Waddington's "epigenetic landscape" concept, which illustrates how an initial, individual cell (genotype) may give rise to different cell types (phenotypes) by interaction with internal and external stimuli.

During the second half of the 20th century, various landmark advancements in molecular biology would in turn carry epigenetics from the developmental and evolutionary field to its contemporary standing as a fundamentally molecular science (Felsenfeld 2014): first, the observation in *Drosophila* that changes in the chromosomal position of genetic information could influence its function (Muller 1930; Hannah 1951), thus implying that the physical organization of genes could impact their regulation; second, nuclear transplant experiments performed in amphibians demonstrating that somatic cells carried the same genetic information as embryo cells (Briggs & King 1952; Laskey & Gurdon 1970), suggesting that developmental processes did not occur through mutational, genetic alterations; third, the comprehension that chemical modifications such as the methylation of DNA could play roles in the regulation of gene function (Griffith & Mahler 1969; Riggs 1975; Holliday & Pugh 1975).

We arrive thus at the contemporary notion of epigenetics. Though still a debated concept (Deans & Maggert 2015; Greally 2018), epigenetics is, today, broadly understood to be **the study of molecular changes that occur at the level of chromatin without modifying the DNA sequence**. These changes can be

heritable, especially regarding cellular mitosis or meiosis, and may affect the function of genes.

Classically, the main epigenetic mechanisms have been considered to be those involving direct, covalent modifications of the chromatin molecule, such as DNA or histone modifications. However, other mechanisms not involving covalent alterations like chromatin remodeling and non-coding ribonucleic acids (ncRNA) are often considered to be epigenetic mechanisms, especially when being observed to be stable and/or heritable processes that act on chromatin (Cavalli & Heard 2019). In this sense, perhaps Adrian Bird's definition of epigenetics as "*the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states*" (Bird 2007) provides us with a more concise definition for this complex field.

2. CHROMATIN AND EPIGENETIC MECHANISMS: THE REGULATION OF BIOLOGICAL INFORMATION

In order to delve into the characterization of the different epigenetic mechanisms, the nature of biological information must be first discussed. Deoxyribonucleic acid, DNA, is the molecule responsible for hosting biological information in all forms of life—with the exception of some ribonucleic acid (RNA) viruses. This molecule is a double-chained (bio)polymer composed of covalently-bound nucleotides containing four different types of nitrogenous bases: cytosine (C), guanine (G), adenine (A) and thymine (T). The linear arrangement of these individual bases generates a code which can be interpreted by other molecules to bring about biological function. In eukaryotic organisms, most of the information encoded by DNA is stored in a membrane-bound organelle called the nucleus, usually across a set of various independent molecules termed chromosomes. The biological complexity of eukaryotes has led them to possess genomes of up to billions of base pairs (bp) in size (Blommaert 2020). For these

extremely large molecules to physically fit within cells, they must be intricately packaged and organized in a complex known as chromatin.

Chromatin is a fibrous structure mainly composed of DNA and associated proteins known as histones. The nucleosome represents the basic structure of chromatin and consists of a short span of DNA, measuring around 146 bp, wrapped around a core of eight histones—the "histone octamer"—in what is known as the nucleosome's "core particle". These particles are connected through small segments of "linker" DNA to form a 10-nm "beads-on-a-string" structure, which is the basal form of chromatin (Cutter & Hayes 2015; McGinty & Tan 2015). Through the coupling of an additional histone at the linker DNA, this initial fiber may subsequently be condensed into thicker structures (Hansen et al. 2018; Maeshima et al. 2019) forming the interphase chromosomes. During cell division, these threads are further compacted into the well-known structures observable by light microscopy, by way of mechanisms that have not yet been fully clarified (Antonin & Neumann 2016) (**FIGURE 3**).

The structure of the nucleosome is relatively well-conserved across Eukarya (Koyama & Kurumizaka 2018): its octamer core is formed by two H2A/H2B and two H3/H4 histone heterodimers, while histone H1 associates with the linker DNA. Being the basal level of DNA compaction in the cell, nucleosomes are in fact very dynamic structures which can completely or partially disaggregate, slide across the DNA strand, be subject to epigenetic modifications and incorporate different histone variants within their cores (Lai & Pugh 2017).

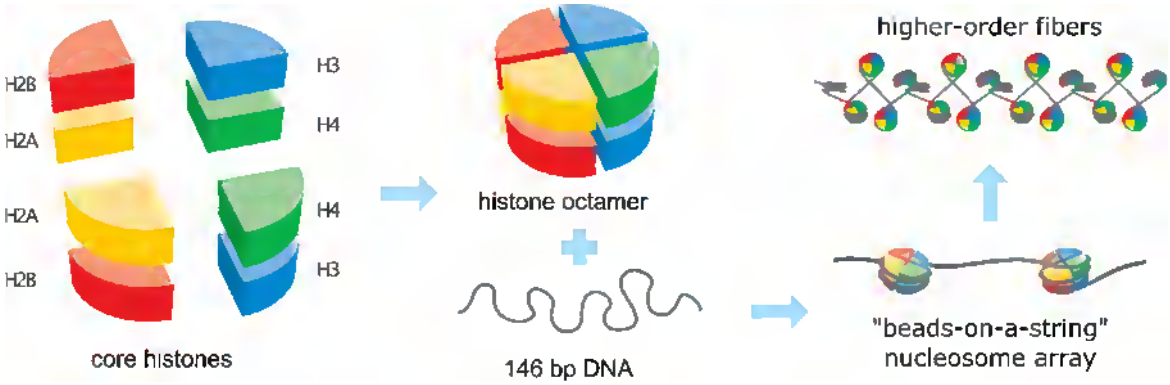


Figure 3. Illustration of the basic components of chromatin, starting with the individual histone components of the nucleosome's core particle leading to the chromatin fiber formed by compaction of beads-on-a-string arrays of nucleosomes (adapted from Morgan 2007).

In this line, it is important to understand that during the last decades, science has gradually moved away from the classic notion of higher-order chromatin organization as a rigid, hierarchical arrangement in favor of a view of chromatin as a markedly dynamic system, in which innumerable agents play their roles in determining its local and global configuration in any cell type, at any time point—including, for example, other non-histone architectural proteins (Cubeñas-Potts & Corces 2015), associated RNA molecules (Li & Fu 2019), interactions with the nuclear lamina (Briand & Collas 2020) and, of course, epigenetic modifications (Bannister & Kouzarides 2011). Not only do these factors directly enact programmed biological functions, but their mere coexistence in the cellular environment leads to the emergence of phase-separated, condensed biophysical states with specific biological purposes (Banani et al. 2017; Hansen et al. 2021).

Together, these tightly regulated processes ultimately control the degree of accessibility to different regions of the genome, thus governing genetic function by determining which information can be read and interpreted by the cell at all times. In the following sections, different epigenetic mechanisms will be discussed, with a focus on covalent histone and DNA modifications.

2.1. HISTONES AND HISTONE MODIFICATIONS

In direct contact with the DNA molecule, the histones H2A, H2B, H3 and H4 form the core of the nucleosome. These proteins are highly-basic, relatively-small (11-15 kDa) molecules in which we can distinguish two main parts (Cutter & Hayes 2015; McGinty & Tan 2015):

- **"Histone fold" motif:** a mostly α-helical domain which comprises the central and C-terminal parts.
- **N-terminal "tail":** a less structured region, rich in positively-charged lysine and arginine residues, which forms the N-terminal end.

The histone fold motif provides an interface through which the proteins can interact to form heterodimers and, subsequently, the full octamer, while the N-terminal tails conform structures which extend outwards from the nucleosome particle, so that they are more accessible to external factors. On the other hand, the linker histone H1 is the least conserved element and possesses a central globular domain with which it interacts with linker DNA, although its various putative roles in regulating chromatin function are much less known than those of its core particle counterparts (Fyodorov et al. 2018).

Proteins—polypeptides—are very diverse polymers, and as such there is a wide array of chemical modifications which can occur at the side chains of their constituent amino acids. Indeed, tens to hundreds of different post-translational histone modifications (PTMs) have been documented in the scientific literature, most of which tend to occur at the more-accessible N-terminal tail domains of the nucleosome octamer (Zhao & Garcia 2015). Within this great variety of epigenetic marks, we can distinguish two main mechanisms through which histone PTMs carry out their biological functions:

- **Modulation of the physicochemical environment of chromatin:** the addition or removal of chemical modifications on histones has a direct impact on the physicochemical interactions between these proteins and the DNA molecule, thus influencing the degree of chromatin compaction or accessibility (Cortini et al. 2016).
- **Constitution of interpretable biological signals:** histone modifications are specific epigenetic signals which can be decoded by chromatin readers in order to execute biological functions in the context of chromatin (Yap & Zhou 2010).

Many histone PTMs are yet to be understood—and some are still to be discovered, even in human (Zhang et al. 2019). These modifications can range from acetylation, methylation, phosphorylation or biotinylation to citrullination, mono-ADP-ribosylation or biotinylation, among others (Zhao & Garcia 2015), and they usually follow the same informal nomenclature (Box 1). Below, the most well-known modifications and their documented functions are summarized.

Box 1. Nomenclature of histone post-translational modifications.

Histone PTMs are usually indicated in the following manner:

[histone] [amino acid] [amino acid position] [modification] [number of modifications]

For example:

- Acetylation of lysine 16 of histone H4: H4K16ac.
- Trimethylation of lysine 4 of histone H3: H3K4me3.
- Phosphorylation of serine 3 of histone H3: H3S10p.

- **Histone acetylation:** histone acetylation occurs through the addition of an acetyl group ($-COCH_3$) to the protonated amino groups ($-NH_3^+$) of lysine amino acid residues. The incorporation of this alteration neutralizes the lysine's positive charge and has an impact on the interaction between the histone and DNA—the negatively-charged DNA phosphates seem to bind more efficiently to positively-charged histones. Although the effect is modification- and histone-dependent (Wang & Hayes 2008), in general, classic *in vitro* studies have shown that acetylation helps to disrupt

chromatin compaction (Cortini et al. 2016) and this histone PTM is usually associated with transcriptional activation in the cell (Barnes et al. 2019). Examples of well-known modifications include H3K9ac and H3K27ac, which are associated with active promoters and enhancers (Pradeepa 2017; Barnes et al. 2019), and H4K16ac, also linked to the aforementioned elements but with additional known roles in DNA damage repair (Dhar et al. 2017).

Histone acetylation is controlled by histone/lysine acetyltransferases (HATs/KATs) and histone deacetylases (HDACs), while the recognition of this mark is mediated by various protein structures, including bromodomains and PHD finger domains (Yap & Zhou 2010).

- **Histone methylation:** histone methylation consists in the transference of methyl groups ($-CH_3$) to the protonated amino group ($-NH_3^+$) of lysine or to the guanidinium group ($-CN_3H_5^+$) of arginine amino acid residues. Importantly, these residues can carry up to three and two methyl groups, respectively, whereby each distinct modification may possess a different biological meaning. Histone H3 is the main known target of histone methylation, and its PTMs have been linked to many different and often opposite functions: across a wide range of organisms, H3K9me2 and H3K9me3 are widely acknowledged repressive marks found at constitutive heterochromatin (Becker et al. 2016), while H3K27me3 is generally related to facultative heterochromatin (Wiles & Selker 2017)[1]. On the other hand, the H3K36me3 modification is associated with the bodies of actively transcribed genes (Wagner & Carpenter 2012) and also plays roles in DNA damage repair (Sun et al. 2020). In addition, the H3K4me3 mark is well-known to be present at the promoters of active genes (Howe et al. 2017), whose adjacent regions are also enriched in the

[1] Heterochromatin regions are compacted DNA domains with reduced accessibility and expression. Two types of regions are generally distinguished: constitutive heterochromatin is always condensed in any cell, such as telomeres or centromeres, while facultative heterochromatin is more dynamic and can be unpacked depending on the cell type, developmental stage or other processes (Trojer & Reinberg 2007).

H3K4me1 modification, which is also characteristic of enhancer locations (Kimura 2013).

A wide variety of histone methyltransferases (HMTs) and histone demethylases (HDMs) exist, and there are several known protein motifs associated with the recognition of methylated histone residues, especially regarding lysine methylation, such as the chromodomain and the Tudor, Royal, PWWP, MBT or PHD finger domains (Yap & Zhou 2010).

- **Histone phosphorylation:** histone phosphorylation involves the addition of a phosphate group ($-PO_4^{2-}$) to the hydroxyl group (-OH) of serine, threonine or tyrosine amino acid residues. Apart from its importance in the regulation of transcription, histone phosphorylation has been classically studied as a key player in cell division and DNA damage response, and often occurs at the core domains of the histones. For example, phosphorylation of histone variant H2A.X at Ser-139 is recognized to play an important role in the signaling and repair of double-strand DNA breaks (Mah et al. 2010), and phosphorylation of H3 at multiple sites, and also of linker histone H1, are linked to chromatin condensation during mitosis (Rossetto et al. 2012).

  In this case, kinases and phosphatases regulate the dynamics of histone phosphorylation, a mark which can be read by protein domains such as those from 14-3-3 proteins, or the BRCT domain (Yap & Zhou 2010).

In addition to the specific biological and physicochemical nature of the different histone PTMs, their dynamics within the cellular environment are also varied so that they can manifest molecular turnover rates ranging from minutes to days, with histone methylation being, in general, a more stable mark than acetylation or phosphorylation (Barth & Imhof 2010). The targeting of chromatin modifiers to specific genomic locations depends on their interaction with multiple factors such as transcription factors (TFs), ncRNAs, other epigenetic marks or the sheer DNA sequence. In this sense, these enzymes typically form part of multiprotein complexes in which some

subunits direct their genomic targeting via chromatin reading or interaction with other molecules while other subunits are responsible for histone modification (Schuettengruber et al. 2017).

As can probably be anticipated, histone PTMs do not exist independently in the cellular context; rather—as is also a general case for all epigenetic marks—they coexist in a certain spatiotemporal circumstance, associated with the DNA molecule, in their almost-infinite possibilities of combinations. Indeed, in the year 2000, Brian D. Strahl and C. David Allis took note of the combinatorial nature of these modifications and proposed the existence of a biologically-meaningful "histone code" (Strahl & Allis 2000), a notion which has been widely adopted in contemporary science (Janssen et al. 2017) and also been extended to epigenetic marks in general (Zhao et al. 2021). Of course, some modifications—and their combinations—do seem to be more prevalent, or conserved across species, than others (Rando 2012; Kimura 2013)[2].This holistic view of epigenetics will be further developed in following sections.

2.1.1. HISTONE VARIANTS

Core histone proteins are relatively-well conserved across evolution, and in higher organisms their genes are grouped in multiple-copy clusters across the genome. The expression of these genes—which are intronless and do not possess the typical mRNA poly(A) tail—is tightly regulated and almost completely associated with cell division, being primarily induced during the S-phase so as to supply histones for the nucleosomes of the newly-synthesized DNA strands (Singh et al. 2018). Aside from relatively-small sequence variation between these "canonical" histone isoforms (H2A, H2B, H3 and H4), there exist other histone genes encoding for so-called histone

[2] Post-translational modifications can also be exclusive: physically exclusive—e.g., a trimethylated lysine residue cannot be acetylated—or biologically exclusive, so that some modifications control impede the deposition of others (Greer & Shi 2012).

"variants" which are considerably different in structure and function. Histone variant genes are not located within canonical clusters and present a more classic configuration in that they produce intron-containing RNAs which can have poly(A) tails and whose expression is not connected to cell cycle dynamics (Talbert & Henikoff 2017). These variants are incorporated and exchanged at nucleosomes, thus adding another layer of epigenetic complexity to chromatin regulation. Prominent examples include the H3 variant CENP-A, which is located at chromosome centromeres and helps define their identity; H3.3, associated with nucleosomal turnover at transcriptionally-active loci; the aforementioned H2A.X, involved in the DNA damage response; H2A.Z, linked to both active and repressive chromatin dynamics; or the testes-specific variants involved in the histone-to-protamine substitution occurring in the germline (Martire & Banaszynski 2020).

2.2. CHROMATIN REMODELING COMPLEXES

As already mentioned, nucleosomes are very dynamic structures which are subject to constant aggregation and disaggregation, displacement or turnover and exchange of their histone components. Indeed, DNA transcription, or its binding by regulatory proteins or RNA, must occur on the bare nucleotide strand, so that genomic locations such as gene promoters or active enhancers typically present a "nucleosome-free" region (NFR)—termed "nucleosome-depleted" (NDR) when it is not constitutively depleted—to which regulatory factors may bind (Lai & Pugh 2017).

These nucleosomal dynamics are controlled by several families of protein complexes known as ATP-dependent chromatin remodelers. These relatively-well conserved complexes all contain a catalytic subunit presenting an ATPase domain and are generally classified into four subfamilies: ISWI, CHD, SWI/SNF and INO80. All remodelers act upon chromatin through a common DNA translocation mechanism, but they carry out different and very specific functions across the genome (Yan &

Chen 2020). In this manner, an "hourglass" model has been proposed to explain the paradigm of chromatin remodeler function (Clapier et al. 2017): in this model, their diverse targeting specificity is mediated by a host of different family subunits under the influence of TFs, ncRNAs and histone PTMs or variants, which then converges on the common ATP-dependent DNA translocation mechanism, to finally enact specific biological functions also dependent on the presence of TFs and histone modifications. These functions are primarily family-specific, and three main categories have been distinguished (Clapier et al. 2017):

- **General nucleosome assembly and organization:** directing the assembly of nucleosomes during DNA replication or transcription to generate the "beads on a string" nucleosome arrays. Primarily directed by the ISWI and CHD subfamilies.
- **Dynamic chromatin access:** mediating chromatin disaggregation to facilitate the binding of DNA by regulatory proteins and RNA. This can involve fully or partially ejecting the components of the nucleosomes. Mostly controlled by the SWI/SNF subfamily.
- **Specific nucleosome editing:** incorporating or exchanging histone variants within the nucleosomes. Mainly regulated by the INO80 subfamily.

To help direct their specific functions, many remodeler proteins possess DNA- and histone PTM-binding domains, such as the aforementioned bromo- or chromodomains (Sundaramoorthy & Owen-Hughes 2020), and they also cooperate with histone modification proteins. Despite not directing the establishment of covalent modifications, chromatin remodeling factors are generally considered to be epigenetic mechanisms because they reconfigure the function of chromatin without altering its sequence. Furthermore, as is the case for other epigenetic mechanisms, the composition—in terms of histone PTMs or variants (Henikoff & Smith 2015)—and even the specific genomic location of nucleosomes has been shown to be inheritable through cell division (Henikoff & Ahmad 2019).

2.3. DNA METHYLATION

DNA methylation is the quintessential epigenetic mark: some of the first hypotheses regarding the possible roles of chromatin modification on the regulation of genomic function in eukaryotes were proposed for DNA methylation (Griffith & Mahler 1969; Holliday & Pugh 1975; Riggs 1975) and this mark is to date, by far, the most well-known and studied of all epigenetic mechanisms. It is present in species across all clades of life, including prokaryotes, and involves the transfer of methyl groups ($-CH_3$) to cytosines or adenines (Yi 2017). In vertebrates, and particularly in mammals, DNA methylation possesses the following general characteristics (Jones 2012; Kulis et al. 2013; Edwards et al. 2017; Greenberg & Bourc'his 2019) (**FIGURE 4**):

- It consists in the addition of a methyl group to the fifth carbon of the cytosine base to form 5-methylcytosine (5mC).
- It appears in regions known as "CpG sites", in which cytosines are followed by guanines.
- CpG sites are symmetrically methylated in both strands of the DNA molecule. When it occurs, the methylation of a single cytosine of the CpG site is called "hemimethylation".
- The distribution of CpG sites across genome is non-random: they appear clustered in high-density regions termed "CpG islands" which are often located at gene promoters, while the rest of the genome, scarce in CpGs, is called "open sea"[3].
- DNA methylation correlates inversely with density in CpG sites, so that CpG islands tend to be unmethylated while the rest of the genome displays high levels of methylation.

[3] Regions adjacent to CpG islands, which may have specific regulatory features, are typically called "shores"—up to 2,000 bp in distance (Irizarry et al. 2009)—and "shelves"—between 2,000 and 4,000 bp in distance (Gao et al. 2020).

There are, of course, exceptions to all of these rules: low levels of N6-methyladenine (6mA) have been reported in the DNA of mammalian and vertebrate species (Koziol et al. 2016; Wu et al. 2016), although the significance of these observations is still disputed (Douvlataniotis et al. 2020); 5mC can be detected—albeit at reduced levels—at non-CpG sites in tissues such as brain or embryonic stem cells (ESCs) in which it can have biological functions (Lister et al. 2009; de Mendoza et al. 2021); CpG sites can be stably- and heritably hemimethylated, for example in human ESCs (Xu & Corces 2018); "orphan" intragenic and intergenic CpG islands exist which may be methylated and display distinct characteristics as compared with promoter-associated islands (Illingworth et al. 2010; Zeng et al. 2014).

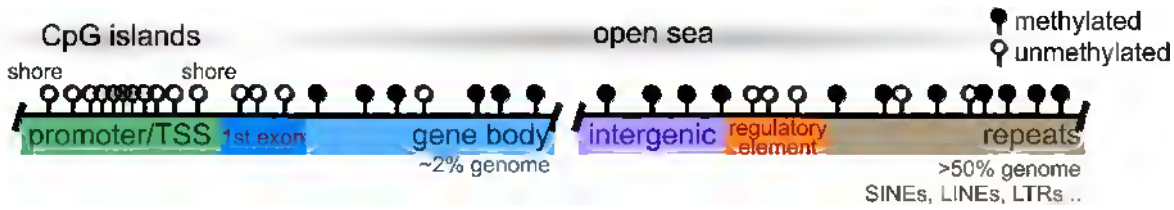


Figure 4. DNA methylation landscape of the human genome. TSS: transcription start site; SINE: short interspersed nuclear element; LINE: long interspersed nuclear element; LTR: long terminal repeat.

Up until the first decade of the 21st century, DNA methylation was primarily considered to be a repressive mark. The classic paradigm of epigenetic regulation held that genes were repressed by gaining methylation—i.e. becoming "hypermethylated"—at the CpG islands associated with their promoters, and, indeed, numerous examples of this phenomenon have been documented, particularly in the context of cancer-related hypermethylation of tumor suppressor genes (Rodríguez-Paredes & Esteller 2011). However, the advent of new technologies allowing for the fine-mapping of DNA methylation across the whole genome has revealed diverse associations of this epigenetic mark with different genomic features (Jones 2012; Kulis et al. 2013; Edwards et al. 2017; Greenberg & Bourc'his 2019) (**BOX 2**).

Box 2. General associations of DNA methylation with genomic features.

CpG islands: these regions often coincide with regulatory elements such as promoters or enhancers and are typically unmethylated, often regardless of the underlying activity of the associated genomic element (Illingworth et al. 2010; Kulis et al. 2013; Bell & Vertino 2017). In physiological conditions, the methylation of CpG islands is linked to the repression of very specific types of genes, which include X-inactivated genes, imprinted genes[4] and germline-specific genes (Greenberg & Bourc'his 2019).

Transcription start sites (TSSs): the methylation of TSSs (Han et al. 2011), and also of the immediately downstream intronic and exonic regions (Brenet et al. 2011; Anastasiadi et al. 2018), is generally associated with repression.

Gene bodies: methylation at gene bodies, particularly at exons, is well conserved across many eukaryotic lineages (Feng et al. 2010), and tends to correlate positively with transcription (Varley et al. 2013). In fact, transcription may also lead to the methylation of CpG islands located within gene bodies (Jeziorska et al. 2017). Gene body methylation is known to influence splicing (Shayevitch et al. 2018), transcriptional elongation (Cholewa-Waclaw et al. 2019), and may play a role in inhibiting spurious intragenic transcription (Neri et al. 2017) thus controlling transcriptional noise (Huh et al. 2013).

Other regulatory elements (enhancers, TF binding sites, insulator binding sites): in general, local regions of accessible chromatin—such as the binding sites of regulatory elements—tend to be unmethylated (Thurman et al. 2012), and it is thought that the occupation by these factors can mediate the loss of methylation at their binding sites (Stadler et al. 2011; Feldmann et al. 2013). Nonetheless, DNA methylation can also influence the interaction of the regulatory factors in the first place (Domcke et al. 2015), having by itself negative, positive or neutral effects on the binding of TFs (Yin et al. 2017). Moreover, the methylation status of particular elements such as enhancers is complex and can be context-specific (Sharifi-Zarchi et al. 2017). Because regulatory elements are often cell type-specific, tissue-specific DNA methylation patterns are typically found at these types of genomic features (Hon et al. 2013; Roadmap Epigenomics Consortium et al. 2015).

Heterochromatin and repetitive DNA: DNA methylation is usually associated with constitutive heterochromatin (Saksouk et al. 2015). These regions are rich in repetitive elements, which are also typically methylated (Edwards et al. 2017) and can be rich or scarce in CpG sites (Rollins et al. 2006). In this setting, DNA methylation is well known to mediate the repression of genomic repeats so as to prevent unwanted transposition and thus ensure chromatin stability (Bourc'his & Bestor 2004; Zamudio et al. 2015; Barau et al. 2016; Pappalardo & Barra 2021).

[4] Imprinted loci are genomic regions which display allele-specific epigenetic patterns inherited from each parent, whereby only one of the alleles—paternal or maternal—is expressed, while the other is methylated and repressed (Elhamamsy 2017).

In addition, DNA methylation can by itself physically influence genome structure and organization (Choy et al. 2010; Lee & Lee 2012; Buitrago et al. 2021), although this effect on chromatin architecture is probably much smaller than those caused by other epigenetic marks such as histone PTMs.

2.3.1. DNA METHYLATION WRITERS AND ERASERS

The modification of cytosine into 5-methylcytosine is catalyzed by a group of proteins known as DNA methyltransferases (DNMTs). All DNMTs use *S*-Adenosylmethionine (SAM) as their methyl donor cosubstrate, and their catalytic domains are relatively well conserved across evolution, with different clades presenting different numbers of orthologs and paralogs (de Mendoza et al. 2019). In human, there are five known DNMTs: DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L. Of these, DNMT2 and DNMT3L are considered to be "non-canonical" members because they do not methylate DNA: DNMT2 is a transfer RNA (tRNA) methyltransferase, while DNMT3L is a catalytically-inactive variant of DNMT3 which acts as a cofactor for the other DNMT3 enzymes. Conversely, DNMT1, DNMT3A and DNMT3B are responsible for the direct methylation of DNA sequences (Lyko 2018).

DNMTs are tightly-regulated enzymes with key roles in numerous biological processes, such as development or cancer (Zhang & Xu 2017; Chen & Zhang 2020). As such, they possess different domains through which they interact with other regulatory factors—including TFs, chromatin readers, or chromatin modifiers—that direct their activity, which can then be sequence-specific or nonspecific (Hervouet et al. 2018). This cooperation with external factors helps understand how different DNA methylation patterns can be associated with different genomic elements such as CpG islands, gene bodies or repetitive DNA (Chen & Zhang 2020). For instance, DNMT3A/B present PWWP and ADD domains that interact with histone PTMs, while DNMT1 possesses, among others, a zinc-finger CXXC domain which binds

unmethylated CpG sites and a RFTS domain which targets replication foci (Lyko 2018). In addition, several families of proteins are able to recognize methylated DNA and interact with DNMTs and other chromatin modifiers, including methyl-CpG-binding domain (MDB) proteins, SET- and Ring finger-associated (SRA) proteins, and the Kaiso family proteins (Du et al. 2015).

In spite of all the involved pleiotropy and interactions, DNMTs have generally been associated with two very different roles: "maintenance" and "de novo" methylation. Maintenance replication, orchestrated by DNMT1, is the process through which DNA methylation is copied to unmethylated newly-synthesized DNA strands during DNA replication. In this mechanism, DNMT1 is targeted to the hemimethylated DNA molecule via interaction with UHRF1 so as to restore the DNA methylation patterns in the new strand and thus prevent the loss of this mark by replication dilution (Petryk et al. 2021). On the other hand, de novo methylation, which is carried out by DNMT3 enzymes, refers to the direct methylation of genomic sequences in order to establish new DNA methylation patterns (Lyko 2018). Nonetheless, DNMTs also have a certain degree of functional promiscuity, so that DNMT1 can participate in the de novo methylation of, for example, transposable elements (Haggerty et al. 2021), while DNMT3A and DNMT3B can be involved in maintenance methylation (Walton et al. 2011; Arand et al. 2012).

2.3.2. DNA METHYLATION IN DEVELOPMENT

Embryonic development is one of the most studied processes from the epigenetic perspective. It is well known that DNA methylation plays a fundamental role during development; indeed, the mutation of DNMT1, DNMT3A and DNMT3B causes lethality at the embryonic or the immediately postnatal stage in mouse (Li et al. 1992; Okano et al. 1999)—as is also the case for many other epigenetic modifiers involving, for example, histone PTMs (Ma & Schultz 2016; Deevy & Bracken 2019).

During development, there occur so-called "waves" of epigenetic reprogramming involving passive and active demethylation, and also de novo methylation, which are thought to be necessary to: 1) erase epigenetic alterations, "epimutations", accumulated in the parental genomes; 2) confer epigenetic plasticity to the embryo cells so that they can correctly develop the organism. In brief, these reprogramming waves consist of (Atlasi & Stunnenberg 2017; Greenberg & Bourc'his 2019):

1. Following fertilization, there is an initial active demethylation of the parental-gamete genomes—especially the paternal—which brings about the zygotic genome activation (ZGA), in which the zygote's genome, as a whole, takes control of the developmental process.
2. Next, both parental genomes are passively demethylated by DNA replication-associated dilution until the blastocyst stage.
3. After implantation of the blastocyst, DNA methylation and other epigenetic marks—and also female X-inactivation—are gradually reestablished as progenitor cells differentiate into the lineages which will give rise to the final somatic cell types.
    - Germline cells, which will generate the reproductive lineage, undergo an additional phase of passive and active demethylation.
    - Subsequently, male gametes are re-methylated prior to birth, while female gametes are re-methylated postnatally prior to ovulation.

2.3.3. DNA HYDROXYMETHYLATION AND DNA METHYLATION ERASERS

In 2009, two back-to-back investigations reported the observation of a new DNA modification, 5-hydroxymethylcytosine (5hmC), in murine brain and ESCs, which in human could be generated from 5mC by a ten-eleven translocation (TET) enzyme,

TET1 (Kriaucionis & Heintz 2009; Tahiliani et al. 2009)[5]. Two years later, another set of studies showed how this modification could be subsequently processed into more oxidized forms (5-formylcytosine, 5fC; 5-carboxylcytosine, 5caC) ultimately leading to is excision from DNA by the DNA repair enzyme thymine-DNA glycosylase (TDG) (Cortázar et al. 2011; Cortellino et al. 2011; He et al. 2011; Ito et al. 2011). Thus, two very important—and somewhat independent—observations had been made: that there existed a new mammalian epigenetic mark and that there existed a molecular pathway enabling active DNA demethylation.

At the present time, 5hmC is considered both to be an epigenetic information carrier with functional relevance and also part of the active DNA demethylation mechanism of the cell. With regards to its functional meaning, this mark is mostly present in the central nervous system, with other tissue types displaying variable or negligible levels of 5hmC (Globisch et al. 2010; Szwagierczak et al. 2010). 5hmC is relatively stable in the cellular environment (Bachman et al. 2014) and tends to be localized within gene bodies, particularly exonic sequences, with tissue specific patterns also being detected at enhancer locations so that, in general, 5hmC is an epigenetic mark associated with regulatory activation and gene expression (Song et al. 2011; Schutsky et al. 2018; Cui et al. 2020; He et al. 2021).

Because 5hmC is generated from 5mC and these modifications cannot coexist within the same nucleotide, these marks have a singular, interdependent relationship (Kochmanski et al. 2019) and it has been observed that alterations in one lead may to changes in the other (Putiri et al. 2014; Vető et al. 2018).

The oxidation of 5mC to 5hmC is catalyzed by the TET family of enzymes, of which there are three members in mammals: TET1, TET2 and TET3. These proteins

[5] The presence of 5hmC had already been reported decades before in animal tissues (Penn et al. 1972) but the significance of these observations has been questioned (Pastor et al. 2013).

are dioxygenases which sequentially oxidize 5mC to 5hmC, 5fC and 5caC by means of a Fe^{2+} and 2-oxoglutarate-dependent mechanism (Pastor et al. 2013). All three TETs possess a common C-terminal catalytic domain and an unmethylated DNA-binding CXXC domain—also carried by the aforementioned DNMT1—which, in the case of TET2, has been separated because of ancestral chromosomal inversion events and now constitutes an independent gene, *IDAX* (An et al. 2017).

The 5mC oxidation process constitutes a bona fide pathway of active DNA demethylation and can in principle occur in two different ways (Kohli & Zhang 2013)[6]:

- **Passive-dilution**: 5mC is modified to 5hmC, which then cannot be copied to new strands during DNA replication, so that the original DNA methylation is lost on the daughter strands.
- **Active restoration**: 5mC is modified to 5hmC and then oxidized to 5fC and 5caC, which TDG can directly excise from DNA generating an abasic site, at which an unmethylated cytosine is restored through DNA repair pathways.

Even though the particularities regarding the importance of the different involved pathways or enzymes in these processes are yet to be clarified, it is indisputable that TET-associated active DNA demethylation occurs and is biologically relevant in mammals. For instance, the previously mentioned developmental reprogramming processes which concern massive demethylation events (see section **2.3.2. DNA METHYLATION IN DEVELOPMENT**) are known to be TET-dependent (Greenberg & Bourc'his 2019), although a part of these phenomena are probably controlled by non-catalytic TET activity (Ross & Bogdanovic 2019).

[6] Other DNA demethylation pathways involving direct deamination of 5mC or 5hmC through AID/APOBEC family enzymes have been proposed but the in vivo evidence is still limited (Bochtler et al. 2017).

3. OMICS DATA AND CHROMATIN STATES: THE EPIGENETIC ORCHESTRA

As already stated, epigenetic marks are not isolated modifications existing independently of one another. The molecular environment enveloping the chromatin molecule is a complex scenario which involves the combination of innumerable factors subject to a constant dynamic regulation. In this sense, no epigenetic mark can be thought to be purely active or repressive: though there can be general associations—as has been successively expounded here—the final activity status of any genomic region will depend, among other things, on the aggregate effects of all the modifications present.

The first two decades of the 21st century have witnessed an explosive revolution in the molecular sciences: spurred by the exponential development of information technology and by the gargantuan scientific and technological enterprises which led to the sequencing of the human genome (Lander et al. 2001; Venter et al. 2001), parallel DNA sequencing technologies, and in general, so-called "high-throughput" technologies have been developed across all fields of molecular biology. These technologies, collectively known as "omics", are capable of analyzing, in parallel, massive quantities of molecules and their goal is to profile the complete set of biomolecules present in biological samples (Hasin et al. 2017) (**BOX 3**).

Box 3. Omics technologies in epigenomics: general characteristics.

Numerous types of omics technologies have been and are being developed continuously. Those based on parallel sequencing usually involve the following general steps:

isolation of biomolecules > modification steps (purification, chemical/enzymatic reactions, fragmentation) > library preparation (adapter ligation, amplification) > sequencing > alignment > quantification

Some of the most widely used technologies, with a focus on the epigenetics field, are commented here:

- **RNA sequencing (RNA-seq, quantifies expression levels of genes and isoforms):** cellular RNA is isolated and typically converted to cDNA, which is then

amplified and sequenced. After aligning the sequencing reads to the genome or transcriptome, counts of reads are obtained for each gene or isoform (Stark et al. 2019).

- **Bisulfite sequencing (BS-seq, quantifies DNA methylation levels at genomic positions):** cellular DNA is isolated and applied a bisulfite treatment which, after amplification, enables the distinction of 5mC from C bases. Then, after sequencing and aligning the reads, the levels of 5mC relative to C can be quantified at each genomic position. The whole genome can be profiled (WGBS), but there are also methods which enrich for a fraction of the genome, e.g. targeting CpG islands, such as reduced-representation bisulfite sequencing (RRBS), or variants which can quantify 5hmC using oxidation, enzymatic or antibody-based protocols (Yong et al. 2016).
- **Chromatin immunoprecipitation and sequencing (ChIP-seq, localizes and quantifies binding sites of proteins of interest):** cellular chromatin bound to a protein of interest is isolated using specific protein-targeting antibodies. Next, the isolated fragments are amplified, and sequencing is performed. Finally, the aligned reads are counted across the genome so as to detect enriched aggregates of reads, called "peaks", which represent the binding sites of the protein of interest (Furey 2012).
- **Assay for transposase-accessible chromatin using sequencing (ATAC-seq, localizes and quantifies regions of open chromatin):** nuclear chromatin is isolated and then open chromatin regions are tagged through the use of a Tn5 transposase. These tagged regions are amplified and sequenced, so that after alignment, "peaks" of accessible chromatin can be detected in a manner similar to that of ChIP-seq (Klemm et al. 2019).
- **Array-based technologies (quantify the levels of sets of molecules of interest):** these technologies do not involve sequencing; rather, they typically rely on highly-sophisticated chips containing oligonucleotides which hybridize with sequences of interest. Thus, they focus on quantifying a set of pre-specified target loci, for example: quantifying the expression of all the genes in the genome (Slonim & Yanai 2009), or measuring the DNA methylation levels of a set of CpG sites in the genome (Wilhelm-Benartzi et al. 2013). Most of these technologies typically quantify by measuring light-based signals produced by the nucleotide hybridization processes.

The large-scale generation, and sharing, of omics data has permitted the development of a myriad of data science-based approaches to tackle old and new scientific questions (Marx 2013). From the epigenetic perspective, this has led to the single-nucleotide resolution mapping of epigenetic modifications across a wide range of tissues and diseases, efforts which are often conducted by international consortia such as the International Human Epigenome Consortium (IHEC) (Stunnenberg et al.

2016), associated with the BLUEPRINT (Martens & Stunnenberg 2013), ENCODE (ENCODE Project Consortium 2012) and NIH Roadmap (Bernstein et al. 2010) projects, or The Cancer Genome Atlas (TCGA) program (Cancer Genome Atlas Research Network 2008). This genome-wide cartography of epigenetic regulation has deepened our understanding of how epigenetic mechanisms come together to regulate chromatin function and has led to the contemporary notion of "chromatin state". Chromatin states are spatial combinations of epigenetic marks and regulatory factors which are repeatedly observed in biological systems and have functional meaning (Baker 2011). These states have been defined through the use of automated statistical methods to integrate large quantities of different omics data (Hon et al. 2008; Ernst & Kellis 2012; Hoffman et al. 2012; Mammana & Chung 2015) and can complement or outperform functional predictions made by genomic sequence analyses (Tsai et al. 2015) (**BOX 4**).

Box 4. An illustration of the concept of chromatin states.

An exemplification of the concept chromatin states can be extracted from the following study (Ernst et al. 2011): in 2011, an ENCODE effort led by Jason Ernst and colleagues integrated epigenomic data from several human cell types measuring the genome-wide occupancy of different histone PTMs (H3K27ac, H3K9ac, H3K4me1/2/3, H4K20me1, H3K27me3, H3K36me3) and also CTCF insulator binding sites. Using hidden Markov models (HMM), they discovered 15 different patterns of combinations of these chromatin marks—15 chromatin states—, which were associated with genomic or functional features. For instance, states linked to active promoters were characterized by the presence of active histone modifications such as H3K4me2/3 or H3K27ac, with inactive promoters also containing H3K27me3, while states located within transcribed gene bodies were mostly associated with H3K36me3.

A very important observation arising from integrative epigenomic studies is that chromatin states appear to be conserved across tissues (Vu & Ernst 2022) and close species like human and mouse (van der Velde et al. 2021), also displaying parallelisms with more distant species like *Drosophila* (Kharchenko et al. 2011) or even plants

(Sequeira-Mendes et al. 2014). Chromatin states do vary in their genomic locations in a tissue- or species-specific manner, but their functional combinations seem to be tightly defined across biological systems, thus arguing in favor of the existence of an "epigenetic code" (de Pretis & Pelizzola 2014).

4. EPIGENETICS, ENVIRONMENT AND DISEASE

The genome carries the basal level of biological information and as such it needs to be a very stable code. In consequence, a host of DNA repair mechanisms have evolved to help maintain this stability and protect it from endogenous or exogenous damage (Chatterjee & Walker 2017). On the other hand, epigenetic mechanisms are not subject to the same constraints: these modifications are dynamic and reversible and thus constitute a flexible system through which the genome can actively respond to changing environments. These dynamic environments can arise from internally-controlled biological processes, such as embryonic development or cellular differentiation, or they can appear because of external stimuli like stress, pathogens, or lifestyle factors. Today, epigenetics is recognized as a critical link between the environment and genomic function (Feil & Fraga 2012; Cavalli & Heard 2019) (**BOX 5**).

Box 5. The agouti mouse: a model for environmental epigenetics.

One of the most studied paradigms of environmental epigenetics is that of the agouti mouse. In this model, the color of the animal's coat is controlled by the expression of the agouti viable gene (A^{vy}), which contains an intracisternal A-particle (IAP)—a retrotransposon-associated sequence—inserted within its promoter (Feil & Fraga 2012). DNA methylation can influence the expression of this region, so that an unmethylated IAP allows gene activation giving rise to a yellow agouti color, while a methylated IAP is repressive and results in the mouse having a brown "pseudo-agouti" color (Michaud et al. 1994). Fascinatingly, the dietary supplementation of gestating mothers with methyl donors such as folic acid or cobalamin, among others, can control the methylation status of the IAP loci in the developing fetus and thus the color of the newborn offspring (Waterland & Jirtle 2003).

Epigenetic alterations are often observed in association with pathological processes such as cancer (Feinberg et al. 2016), aging (Sen et al. 2016), neurodegenerative conditions (Hwang et al. 2017) or metabolic diseases (Ling & Rönn 2019), and can often provide mechanistic explanations for the etiology or pathophysiology of these processes, as will be commented below.

4.1. THE EPIGENETICS OF AGING AND AGING-ASSOCIATED DISEASES

Aging is a near-universal biological process that involves the gradual deterioration of an organism's physiological functions across its lifespan. This time-dependent decline is known to be caused by the accrual of molecular damage at multiple levels, including epigenetic alterations (López-Otín et al. 2013), although it is also thought that this process might be, in part, a programmed biological function (Kowald & Kirkwood 2016). Aging is the major risk factor for many diseases, including cardiovascular pathologies, neurodegenerative disorders or cancer (Niccoli & Partridge 2012), and, in fact, the boundary between aging symptoms and aging-associated disease symptoms is often blurred (Franceschi et al. 2018).

Aging is accompanied by a series of epigenetic alterations which have often been studied through the use of model organisms such as *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* or *Mus musculus* and include general chromatin disruption via the loss of histones and heterochromatin, and the displacement of repressive and active chromatin modifications (Booth & Brunet 2016; Sen et al. 2016). Moreover, genetic models targeting histone modifiers such as histone methylases, or the sirtuin histone deacetylases, have been shown to modulate lifespan in the aforementioned organisms (Molina-Serrano et al. 2019), while laminopathies—pathologies affecting nuclear envelope organization—can generate features of accelerated aging (Shin & Worman 2022). Altogether, the general dysregulation of

"normal" chromatin states, especially at heterochromatic locations, is a typical epigenetic feature of aging (Sen et al. 2016).

With regards to DNA methylation, two hallmarks of aging-associated alterations have been classically described in mammals and particularly human (Huidobro et al. 2013): first, a global loss of DNA methylation was initially supported by chromatography-based methods (Unnikrishnan et al. 2018) and pyrosequencing-based technologies targeting repetitive DNA elements (Bollati et al. 2009), but recent high-throughput sequencing approaches, particularly involving mouse models (Sun et al. 2014; Cole et al. 2017; Hahn et al. 2017; Masser et al. 2017; Hadad et al. 2019; Hernando-Herraez et al. 2019), have often failed to observe global hypomethylation across a range of tissues (Unnikrishnan et al. 2018); secondly, very characteristic local hypermethylation events affecting the promoter CpG islands of a specific subset of genes have been observed across a wide variety of tissues using genome-wide array and sequencing technologies (Rakyan et al. 2010; Teschendorff et al. 2010; Heyn et al. 2012; Day et al. 2013; Fernández et al. 2015). These hypermethylated genes are very particular: they are characterized by presenting a "bivalent chromatin" signature in stem cells consisting of the colocalization of active (H3K4me3) and repressive (H3K27me3) histone modifications (Bernstein et al. 2006; Pan et al. 2007), which are deposited by the Trithorax and Polycomb group proteins, respectively (Schuettengruber et al. 2017). Genes marked by bivalent chromatin usually possess developmental functions and play roles in lineage differentiation; in this sense, it has been posited that they exist in a repressed state which is "poised" for subsequent activation during embryonic development, although the mechanisms involved in this process are still to be clarified (Vastenhouw & Schier 2012). In fact, the aging-associated hypermethylation of these genes could provide an epigenetic explanation

for the typical decline of stem cell function observed during aging (López-Otín et al. 2013).

Aside from the aforementioned specific epigenetic hallmarks, a universal feature of the aging process is an increase in molecular variability or "noise". Indeed, the deterioration of the molecular control mechanisms which occurs during aging leads to a gradual increment in heterogeneity that causes phenotypic variability in the cell, leading to transcriptional noise (Nikopoulou et al. 2019) and the loss of proteostasis (Labbadia & Morimoto 2015). This increase in heterogeneity also affects epigenetic marks, hinting at a mechanism which may help explain the observed alterations in gene and protein expression patterns. It is well known that, with age, there occurs an increase in the inter- and intraindividual variability of global and genome-wide epigenetic patterns (Fraga, Ballestar, Paz, et al. 2005; Hannum et al. 2013; Slieker et al. 2016). Age-associated epigenetic changes have often been referred to as "epigenetic drift" (Teschendorff et al. 2013); nonetheless, this complex phenomenon encompasses several different processes, because the changes can have either biologically-programmed or stochastic origins, can lead to both functional or non-functional effects (Tejedor & Fraga 2017), or can actually reflect alterations in tissue cell type composition (Teschendorff & Zheng 2017). Advances in single cell profiling technologies will no doubt aid in clarifying the sources and mechanisms of molecular heterogeneity associated with the aging process (Hernando-Herraez et al. 2019).

4.1.1. EPIGENETIC CLOCKS

The aforementioned omics data revolution (see section **3. OMICS DATA AND CHROMATIN STATES: THE EPIGENETIC ORCHESTRA**) has led to the fine mapping of aging-associated epigenetic changes in humans across a wide variety of tissues and cohorts. This scenario has allowed researchers to develop algorithms which exploit epigenetic data to predict chronological age, the most successful of which use DNA

methylation biomarkers and have been named “epigenetic clocks” or “DNA methylation clocks”. The first epigenetic clock was developed using saliva samples by Sven Bocklandt and colleagues in 2011 (Bocklandt et al. 2011) and, subsequently, a host of clocks have been derived to predict age both at the single- and multi-tissue levels (Horvath & Raj 2018). These algorithms typically use regression-based multivariate models to combine the DNA methylation levels of tens to hundreds of CpG sites and are able to achieve astonishing levels of precision and correlation with the “real”, chronological age of subjects.

The recent emergence of the epigenetic clocks has had a tremendous impact within the field of aging biology because these methods provide putative biomarkers of aging which can facilitate the verification of clinical anti-aging interventions. Indeed, the degree of negative deviation between the chronological age of patients and their predicted “epigenetic” age, called “age acceleration”, has been shown to be associated with many diseases and biological factors (Oblak et al. 2021), and both embryonic and reprogrammed stem cells show extremely low epigenetic age (Horvath 2013). As such, it is now clear that epigenetic clocks do not strictly measure chronological age or lifespan, but rather a biological age more related to healthspan—the period of life in which one is generally healthy and free of disease—and some algorithms have been developed specifically to predict phenotypic age (Levine et al. 2018).

Nonetheless, the biological significance of the epigenetic clocks remains to be understood, as well as their validity as rejuvenation biomarkers. To date, it is not known whether these CpG sites form part of a programmed chronobiological mechanism or, on the contrary, they are non-functional biomarkers reflecting the influence of other processes (Seale et al. 2022).

4.1.2. AGING AND NEURODEGENERATIVE CONDITIONS

Neurodegenerative disorders are pathologies which affect the nervous system in a mostly gradual and time-dependent manner, thus being tightly associated with aging. Generally, these diseases involve the damage or loss of specific neuronal populations and manifest a proteomic dysregulation which consists in the accumulation of protein aggregates in the brain, often leading to dementia and cognitive disorders (Dugger & Dickson 2017). This cognitive deterioration is intermingled with that occurring during normal aging (Franceschi et al. 2018), and, additionally, lifestyle factors play important roles in the modulation of the development of these diseases, hinting at the relevance of epigenetic mechanisms in their etiology (Popa-Wagner et al. 2020). Epigenetic alterations have been described in association with neurodegenerative disorders at multiple levels, but it is important to study them against the backdrop of aging-associated changes (Barter & Foster 2018; Xia et al. 2018). The characterization of epigenetic marks such as DNA methylation can not only contribute to the understanding of these diseases, but also be used to define biomarkers allowing the early detection of these pathologies (Fransquet et al. 2018), whose timely treatment—often in the form of lifestyle interventions—is crucial for the prevention or amelioration of their symptoms (Kivipelto et al. 2018). In addition, the study of the role of epigenetic mechanisms in normal cognition (Day & Sweatt 2011)—possibly through the use of animal and cellular models (Al Dahhan et al. 2019)—is important in comprehending how neurodegenerative disorders, and aging, alter brain-related functions.

4.1.3. CANCER EPIGENETICS: A LINK WITH AGING?

Cancer is a heterogeneous group of diseases involving the uncontrolled growth of malignant cells that can spread throughout the organism and invade other tissues. Perhaps the most widely studied disease in the field of biology, it involves drastic

molecular alterations—particularly sequence mutations related to multiple types of genes—and also causes epigenetic changes. The contemporary understanding of cancer distinguishes two types of functional roles for mutated genes: "driver" genes which are causally implicated in the tumorigenic process and "passenger" genes displaying mutations which have no apparent function and are not subject to selection (Pon & Marra 2015). In this scenario, epigenetic mechanisms contribute to oncogenesis in different ways so that cancer is characterized by a host of epigenetic alterations and can also involve direct mutations in epigenetic modifiers (Feinberg et al. 2016).

There are well-known chromatin hallmarks of cancer, including the loss of certain histone PTMs (Fraga, Ballestar, Villar-Garea, et al. 2005) and a general dysregulation of heterochromatin domains (Feinberg et al. 2016). With respect to DNA methylation, cancer-associated alterations present two main characteristics which are in fact parallel to those described for the aging process (Rodríguez-Paredes & Esteller 2011): there occurs a global genomic loss of DNA methylation, especially targeting repetitive sequences (Li et al. 2014), which has also been observed using high-throughput sequencing data and is thought to reflect the mitotic activity of tumors (Zhou et al. 2018)—because DNA methylation has to be rewritten after cell division, abnormal proliferation could lead to the incomplete reestablishment of DNA methylation patterns thus causing hypomethylation; in addition, the local hypermethylation of CpG island-associated bivalent chromatin domains during the oncogenic process has also been reported (Ohm et al. 2007; Schlesinger et al. 2007; Widschwendter et al. 2007), suggesting that the epigenetic dysregulation of these developmental genes in cancer can lead to an oncogenic stem-like phenotype which contributes to tumorigenesis (Easwaran et al. 2012).

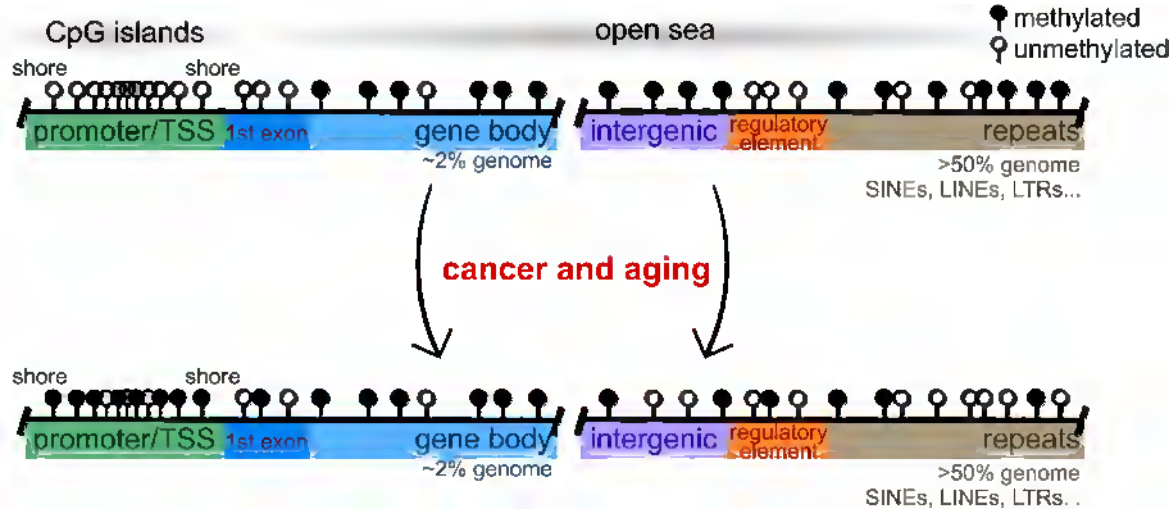


Figure 5. Common alterations in DNA methylation patterns in aging and cancer: local hypermethylation of CpG. TSS: transcription start site; SINE: short interspersed nuclear element; LINE: long interspersed nuclear element; LTR: long terminal repeat.

The striking parallelisms between cancer- and aging-associated DNA methylation alterations—especially regarding the hypermethylation of bivalent, development-associated domains—have led to the hypothesis that epigenetic mechanisms are a causal link between aging and cancer (**FIGURE 5**). Aging is the main risk factor for the development of most tumors, the prevalence of which increases greatly with age, and it is possible that the gradual accumulation of epigenetic damage across lifespan generates a molecular environment that promotes malignant transformation. Nonetheless, while local hypermethylation events have been relatively well characterized in both processes, the similarities involving hypomethylation are not so evident. Thus, there is a need to: 1) integrate the features of the observed cancer and aging DNA methylation alterations within a more general epigenomic framework involving histone modifications and chromatin states; 2) reconcile the findings from studies stemming from human data as opposed to those using animal models such as mouse, which are our main preclinical tools in the study of diseases; 3) characterize

and compare aging-associated alterations to bona fide developmental-associated changes occurring during the first stages of life.

OBJECTIVES AND JUSTIFICATION

On the basis of all of the previously discussed evidence, the work described in this thesis has sought to carry out a systematic study of aging-associated epigenetic alterations—with a focus on DNA methylation—and investigate their relationship with epigenetic changes which take place in aging-associated diseases. Specifically, a first aim has been to describe and compare aging- and cancer-associated epigenetic alterations in order to shed light on the actual similarities and differences between the two processes, adopting an integrative approach that considers all layers of epigenomic regulation and also examines the interspecies evolutionary conservation of these processes across human and mouse. Furthermore, the research also pursued the study of the link between the aforementioned aging alterations, which may have development-associated traits, and actual early-life epigenetic changes. Finally, this work also sought to explore the characteristics of aging-associated epigenetic alterations in the context of cognitive disease so as to gain understanding of how epigenetic changes linked to aging may give rise to different age-associated phenotypes and health related outcomes. To these ends, the following partial objectives were proposed:

1. To describe common and specific features of DNA methylation alterations in aging and cancer in the context of cellular chromatin states.
    a. To compile publicly-available epigenetic data from healthy and tumoral samples across a range of different tissues.
    b. To develop and apply bioinformatic pipelines for the analysis of DNA methylation data and integration with histone PTM and gene expression omics data.
    c. To characterize and compare the chromatin signatures of DNA hyper- and hypomethylation alterations in aging and cancer.
2. To describe the interspecies conservation of aging and cancer DNA methylation alterations between human and mouse in the context of cellular chromatin states.
    a. To generate and analyze epigenetic data from healthy and tumoral brain samples from human and mouse tissues.

   b. To develop and apply bioinformatic pipelines for the analysis of DNA methylation data and integration with histone PTM and gene expression omics data.
   c. To characterize and compare the chromatin signatures of DNA hyper- and hypomethylation alterations in aging and cancer in human and mouse.

3. To describe early-life development-associated DNA methylation alterations and distinguish them from aging-associated changes in the context of cellular chromatin states.
   a. To generate and analyze epigenetic data from the peripheral blood of a longitudinal cohort of early-life subjects.
   b. To develop and apply bioinformatic pipelines for the analysis of DNA methylation data and integration with histone PTM data.
   c. To characterize development-associated early-life DNA methylation alterations.
4. To describe DNA methylation alterations associated with cognitive decline and dementia in elderly subjects.
   a. To generate and analyze epigenetic data from the peripheral blood of a longitudinal cohort of late-life subjects suffering from dementia.
   b. To define DNA methylation biomarkers predictive or explanatory of the development of cognitive decline and dementia.

RESULTS

RESULTS OVERVIEW

The subsequent **RESULTS** section is structured as follows:

- Results related to the study of aging epigenetics and its connection with cancer:
  - **Article 1**: Distinct chromatin signatures of DNA hypomethylation in aging and cancer.
  - **Article 2**: Conservation of Aging and Cancer Epigenetic Signatures across Human and Mouse.
  - **Article 3**: Aging and cancer epigenetics: where do the paths fork?
- Results related to the study of early-life epigenetics:
  - **Article 4**: Longitudinal genome-wide DNA methylation analysis uncovers persistent early-life DNA methylation changes.
- Results related to the study of aging epigenetics and its connection with cognitive function:
  - **Article 5**: Blood DNA methylation patterns in older adults with evolving dementia.

AGING AND CANCER: FIRST ARTICLE

Pérez RF, Tejedor JR, Bayón GF, Fernández AF & Fraga MF (2018) Distinct chromatin signatures of DNA hypomethylation in aging and cancer. Aging Cell 17, e12744.
In this work, we sought to perform a systematic comparison of DNA methylation alterations in aging and cancer, across a wide range of human tissues. We compiled publicly available array-based epigenetic data across seven different tissues from more than 2,000 subjects from various sources, including international consortia (TCGA, NIH Roadmap, ENCODE) and performed bioinformatic analyses to describe aging and cancer DNA methylation alterations. By characterizing these alterations while considering their associated genomic and epigenomic contexts, we first confirmed that DNA hypermethylation was similar in both processes, occurring at CpG island-associated loci which are marked with bivalent, Polycomb-associated chromatin signatures. Strikingly, we observed that DNA hypomethylation affected different regions, being associated to heterochromatin, H3K9me3-marked loci in cancer and active, enhancer-associated, H3K4me1-marked regions in aging. Thus, our results suggest that the epigenomic links between aging and cancer are more complex than previously thought, particularly with regards to the loss of methylation which occurs during these processes.
Personal contribution to the work: I participated in compiling, preprocessing and analyzing all of the data presented in this manuscript, as well as preparing the figure panels and writing the manuscript. The work was carried out in close collaboration with co-first author Dr. Juan Ramón Tejedor, who participated in all of the analyses. The work was co-written and supervised by Dr. Agustín F. Fernández and Dr. Mario F. Fraga.

Accepted: 27 January 2018

DOI: 10.1111/acel.12744

ORIGINAL ARTICLE

WILEY Aging Cell

Distinct chromatin signatures of DNA hypomethylation in aging and cancer

Raúl F. Pérez[1,2] | Juan Ramón Tejedor[2,3] | Gustavo F. Bayón[2] | Agustín F. Fernández[2,3] | Mario F. Fraga[1]

[1]Nanomedicine Group, Nanomaterials and Nanotechnology Research Center (CINN CSIC), Universidad de Oviedo, El Entrego, Asturias, Spain

[2]Cancer Epigenetics Laboratory, Institute of Oncology of Asturias (IUOPA), Hospital Universitario Central de Asturias (HUCA), Universidad de Oviedo, Oviedo, Asturias, Spain

[3]Cáncer Epigenetics Laboratory, Fundación para la Investigación Biosanitaria de Asturias (FINBA), Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, Asturias, Spain

Correspondence
Agustín F. Fernandez, Cancer Epigenetics Laboratory, Institute of Oncology of Asturias (IUOPA), Hospital Universitario Central de Asturias (HUCA), Universidad de Oviedo, Oviedo, Asturias, Spain.
Email: affernandez@hca.es
and
Mario F. Fraga, Nanomedicine Group, Nanomaterials and Nanotechnology Research Center (CINN CSIC), Universidad de Oviedo, El Entrego, Asturias, Spain.
Email: mffraga@cinn.es

Funding information
This work was supported by the Plan Nacional de I+D+I [2013-2016/FEDER PI15/00892 to M.F.F. and A.F.F., 2008-2011/FEDER CP11/00131 to A.F.F.]; the ISCIII · Subdirección General de Evaluación y Fomento de la Investigación [Miguel Servet contract CP11/00131 to A.F.F.]; Instituto Universitario de Oncología del Principado de Asturias (IUOPA) to [G.F.B.]; Fundación Ramón Areces to [M.F.F.]; and the Asturias Regional Government [GRUPIN14-052 to M.F.F.]. R.F.P. is supported by the Retención de Jóvenes Talentos Fellowship from the Obra Social Cajastur-Liberbank. J.R.T. is supported by the Ministry of Economy and

Summary

Cancer is an aging-associated disease, but the underlying molecular links between these processes are still largely unknown. Gene promoters that become hypermethylated in aging and cancer share a common chromatin signature in ES cells. In addition, there is also global DNA hypomethylation in both processes. However, the similarity of the regions where this loss of DNA methylation occurs is currently not well characterized, and it is unknown if such regions also share a common chromatin signature in aging and cancer. To address this issue, we analyzed TCGA DNA methylation data from a total of 2,311 samples, including control and cancer cases from patients with breast, kidney, thyroid, skin, brain, and lung tumors and healthy blood, and integrated the results with histone, chromatin state, and transcription factor binding site data from the NIH Roadmap Epigenomics and ENCODE projects. We identified 98,857 CpG sites differentially methylated in aging and 286,746 in cancer. Hyper- and hypomethylated changes in both processes each had a similar genomic distribution across tissues and displayed tissue-independent alterations. The identified hypermethylated regions in aging and cancer shared a similar bivalent chromatin signature. In contrast, hypomethylated DNA sequences occurred in very different chromatin contexts. DNA hypomethylated sequences were enriched at genomic regions marked with the activating histone posttranslational modification H3K4me1 in aging, while in cancer, loss of DNA methylation was primarily associated with the repressive H3K9me3 mark. Our results suggest that the role of DNA methylation as a molecular link between aging and cancer is more complex than previously thought.

KEYWORDS
aging, cancer, chromatin, DNA methylation, epigenetics, histone modification

Competitiveness Through a Juan de la Cierva Postdoctoral Fellowship [FJCI-2015-26965]. The IUOPA is supported by the Obra Social Cajastur-Liberbank, Spain.

1 | INTRODUCTION

Age is among the most important risk factors for cancer (de Magalhães, 2013; DePinho, 2000). However, the underlying molecular mechanisms governing this relationship are still poorly understood. Recent research has established polycomb-target gene promoter hypermethylation as a common epigenetic characteristic of cancer (Schlesinger et al., 2007; Widschwendter et al., 2007). In this scenario, prior to alteration these promoters display an embryonic stem cell "bivalent chromatin pattern" consisting of the repressive histone mark H3K27me3 and the active mark H3K4me3 (Ohm et al., 2007). Genes affected by this process are associated with developmental regulation (Easwaran et al., 2012), implying a possible stem cell origin of cancer whereby aberrant hypermethylation could promote a continuously self-renewing embryonic-like state in cancer cells (Teschendorff et al., 2010). Interestingly, promoter hypermethylation of polycomb-target genes was later described in aging blood (Rakyan et al., 2010; Teschendorff et al., 2010) and other tissue types such as mesenchymal stem cells (Fernández et al., 2015), ovary (Teschendorff et al., 2010), brain, kidney, and skeletal muscle (Day et al., 2013), findings which were also confirmed using whole-genome bisulfite sequencing (Heyn et al., 2012).

In addition to aberrant locus-specific DNA hypermethylation, tumoral cells are also globally hypomethylated as compared to their healthy counterparts. While this molecular alteration preferentially occurs at gene bodies, intergenic DNA regions, and repeated DNA elements (Ehrlich, 2009) and is proposed to be associated with chromosomal instability, reactivation of transposable elements, and loss of genomic imprinting, its precise functional role in cancer development is still poorly understood (Rodríguez-Paredes & Esteller, 2011). Intriguingly, global loss of genomic DNA methylation has also been reported during the aging and senescence process (Cruickshanks et al., 2013; Fraga & Esteller, 2007). Whole-genome bisulfite sequencing and methylation arrays have confirmed the global loss of DNA methylation in different human tissues including blood (Heyn et al., 2012), mesenchymal stem cells, and brain (Fernández et al., 2015). On the other hand, other important tissues such as skeletal muscle do not seem to become hypomethylated with aging (Zykovich et al., 2014).

Despite the interesting parallelism in aging and cancer recently reported with respect to hypermethylated DNA regions, the relationship between hypomethylated DNA sequences in these two processes has not been sufficiently studied. Moreover, recent analyses mainly performed in mouse tissue have failed to confirm global hypomethylation during the aging process (Cole et al., 2017; Hahn et al., 2017; Masser et al., 2017) and, to date, no study has provided a back-to-back and systematic comparison of the epigenetic changes that occur in aging and cancer. To address this issue, here we have analyzed DNA methylation changes and their associated chromatin patterns in a total of more than 2,300 healthy and tumoral samples obtained from differentially aged individuals, using HumanMethylation450 BeadChip data generated by The Cancer Genome Atlas (TCGA) consortium and other datasets (Bormann et al., 2016; Guintivano, Aryee & Kaminsky, 2013; Hannum et al., 2013). Our results confirmed the relationship between DNA hypermethylation in aging and cancer, but they also revealed important differences in DNA hypomethylation changes in the two processes that might be important to understand the possible role of DNA methylation as a molecular link between decline related to aging and tumor development.

2 | RESULTS

2.1 | DNA methylation profiling in aging and cancer

To identify DNA methylation changes in aging and cancer, we collected DNA methylation data obtained with the HumanMethylation450 BeadChip (Illumina) (see Section 4) and compared the DNA methylation status of a total of 361,698 CpG sites across 1,762 samples corresponding to healthy and tumoral tissues obtained from differentially aged patients with breast, kidney, thyroid, skin, and brain tumors (see Tables 1 and S1 for extended information). Using an empirical Bayes moderated *t* test (see Section 4), we identified a high number of autosomal differentially methylated CpGs (dmCpGs; FDR < 0.05) between normal and tumoral samples, while a lower and more variable number of aging-related dmCpGs between young and old samples was found (Table 1; Figure 1a and Table S2 for additional information). Hierarchical clustering of samples using the dmCpGs enabled us to distinguish between tumoral and control samples with more efficiency than young and old samples (Figure 1b and Figure S1). On the whole, whereas cancer-related DNA methylation changes had no dominance of either hyper- or hypomethylation, aging-related changes tended toward DNA hypermethylation, and showed a much more variable and tissue type-dependent magnitude of change. Globally, methylation changes were found to be more pronounced in cancer than in aging (Figure 1c and Table S3, Wilcoxon tests; all $p < .05$), while comparison of hyper- vs. hypomethylation changes was variable and disease and tissue type dependent. Intriguingly, most tumors obtained from differentially aged patients did not show significant age-associated DNA methylation changes, with the exception of thyroid cancer (Table 1, Figure 1a, bottom panel and Table S2). Furthermore, analyses employing Horvath's

TABLE 1 Description of sample groups and dmCpGs obtained in the analyses

		BRST		KIDN		THYR		SKIN		GLIA	
	Total samples	483		521		363		197		198	
		Normal	Tumor	Normal	Tumor	Normal	Tumor	Normal	Tumor	Normal	Tumor
	Group	98	385	204	317	56	307	108	89	60	138
	Source	TCGA	TCGA	TCGA	TCGA	TCGA	TCGA	E-MTAB 4385	TCGA	GSE 41826	TCGA
Cancer	Total Samples	483		371		363		197		198	
		Normal	Tumor	Normal	Tumor	Normal	Tumor	Normal	Tumor	Normal	Tumor
	Samples group	98	385	154	217	56	307	108	89	60	138
	Age mean	57.6	57.7	62.5	61.3	45.8	47	47.2	64.4	32.6	59.9
	Age range	28–90	26–90	31–90	26–90	15–81	15–89	18–78	24–90	13–79	21–85
	Gender (M,F)	0 98	4 381	106 48	141 76	14 42	74 233	0 108	54 35	29 31	79 59
	DMPs Total	113,314		134,672		38,593		205,134		173,871	
	DMPs hyper	63,196		65,195		16,561		97,362		84,334	
	DMPs hypo	50,118		69,477		22,032		107,772		89,537	
Aging	Total Samples	98		204		56		108		60	
		Young	Old	Young	Old	Young	Old	Young	Old	Young	Old
	Samples group	20	18	23	22	19	18	16	15	20	19
	Age mean	37.7	81.1	43.3	82.0	28.3	66.7	19.9	73.5	16.4	51.8
	Age range	28–44	75–90	31–49	78–90	15–34	55–81	18–22	71–78	13–21	43–79
	Gender (M,F)	0 20	0 18	15 8	13 9	4 15	7 11	0 16	0 15	12 8	13 6
	DMPs Total	2,588		20,019		3,216		66,977		25,849	
	DMPs hyper	2,077		16,227		2,079		40,640		10,533	
	DMPs hypo	511		3,792		1,137		26,337		15,316	
Aging cancer	Total Samples	385		317		307		197		198	
		Young	Old	Young	Old	Young	Old	Young	Old	Young	Old
	Samples group	28	26	29	28	28	27	30	29	28	25
	Age mean	35.1	83.3	40.2	82.1	22.6	75.2	48.4	79.4	41.4	76.3
	Age range	26–40	79–90	26–46	78–90	15–27	70–89	24–58	73–90	21–51	72–85
	Gender (M,F)	0 28	1 25	19 10	18 10	7 21	10 17	20 10	14 15	13 15	12 13
	DMPs Total	–		–		8,480		–		–	
	DMPs hyper	–		–		7,429		–		–	
	DMPs hypo	–		–		1,051		–		–	

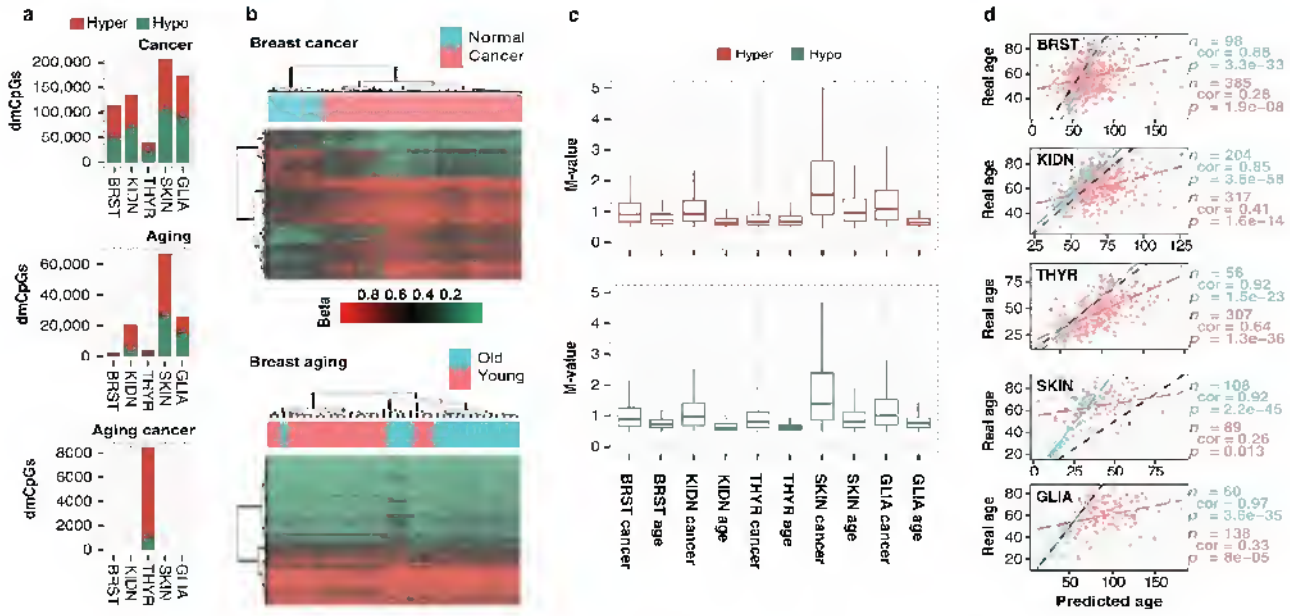


FIGURE 1 DNA methylation changes in aging and cancer. (a) Stacked barplots indicating total number of dmCpGs detected in cancer, aging, and aging cancer tissues. (b) Hierarchical clustering and heatmaps including the 1,000 most significant dmCpGs for breast cancer and aging analyses. Beta-values of DNA methylation are displayed from zero (green) to one (red). (c) Boxplots comparing the magnitude of M-values of methylation changes in cancer and aging. All differences are statistically significant (Wilcoxon tests, all $p < .05$, Table S3). (d) Scatterplots indicating a correlation of chronological age with Horvath's predicted age in normal and cancer samples. Pearson's product-moment correlation coefficient (cor) is indicated, and linear fit lines are added to help with data interpretation

predictor revealed that thyroid cancer had the highest correlation of real-vs.-predicted age across all the cancer types in our dataset (Figure 1d).

2.2 | Genomic distribution of dmCpGs in aging and cancer

The study of the genomic distribution of the dmCpGs revealed that hypomethylated CpG sites followed a similar disease and tissue-independent trend, being preferentially found at low-density CpG DNA regions interrogated by the array in both cancer and aging (average median difference compared to array 49%, Wilcoxon tests; all $p < .001$) (Figure 2a; see also Table S4). Consequently, with respect to the array, these hypomethylated CpG sites were enriched at open sea locations and intronic and intergenic regions (Fisher's tests; all $p < .001$, all odds ratios (ORs) >2.25, >1.34, and >1.21, respectively, except nonsignificant thyroid aging) while impoverished at CpG islands and gene promoters (Fisher's tests; all $p < .001$, all ORs<0.43 and <0.64, respectively) (Figure 2b,c; Figure S2 and Table S5). Density of hypermethylated CpG sites in cancer was variable but comparable to background array density (average median difference <+12%, Wilcoxon tests; all $p < .001$), whereas a noticeably high CpG density was found for breast, kidney, and thyroid in the aging context (average median difference 44%, Wilcoxon tests; all $p < .001$). Consequently, these dmCpGs were enriched at CpG islands and gene promoters (Fisher's tests; all $p < .001$, all ORs>1.89 and >1.08, respectively).

Comparative enrichment analysis confirmed that DNA hypomethylation in aging and cancer mainly occurred at low CpG density DNA regions located at introns, open sea, and intergenic DNA regions, while hypermethylation distribution was more irregular and more similar to array distribution. Nonetheless, a common and strong tendency was found when comparing hyper- to hypomethylation changes in both aging and cancer, whereby hypermethylation changes always occurred in regions with a higher CpG density than did hypomethylation changes (average median difference 51%, Wilcoxon tests; all $p < .001$) (Figure 2a and Table S4), resulting in strong differences in local enrichments at CpG islands and open sea locations, as well as gene promoters and intergenic regions, in most cases (Figure 2b,c; Figure S3 and Table S5), with this effect being even more pronounced for the aging dmCpGs.

2.3 | Tissue type-independent DNA methylation changes in aging and cancer

To determine the effect of tissue type on the DNA methylation changes during aging and cancer, we compared the previously identified dmCpG sites for each of the tissues. In cancer, ~50% of hyper- and hypomethylated CpGs were common to at least two different tumor types (Figure 3a), with 1,962 (1.1%) hyper- and 2,708 (1.5%) hypomethylated CpG sites being common to all five tumor types analyzed (Figure 3b and Table S6). In contrast, the overlap between dmCpGs in aging across different tissues was considerably reduced. Indeed, only 18% of the hyper- and 8% of the hypomethylated CpG

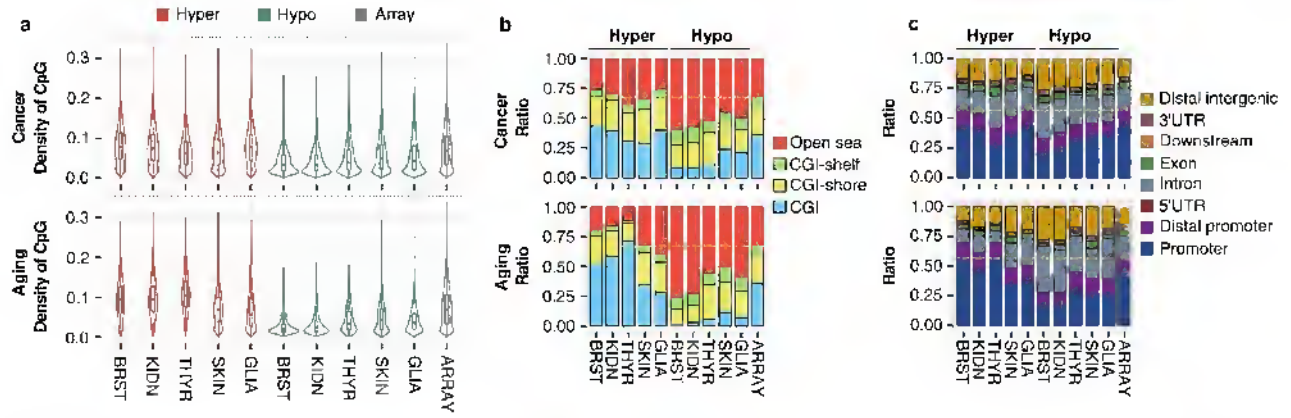


FIGURE 2 Similarities and disparities in the genomic distribution of methylation changes in aging and cancer. (a) Violin plots showing the distribution of CpG density for cancer and aging hyper- and hypomethylated dmCpGs and the background sites in the Infinium HumanMethylation450 microarray. (b) Stacked barplots indicating relative distribution of differentially methylated CpGs according to their CpG island status. A dashed yellow line separates CpG island associated locations from open sea. (c) Stacked barplots indicating relative distribution of differentially methylated CpGs according to their gene location status. A dashed yellow line separates promoter associated locations from the rest

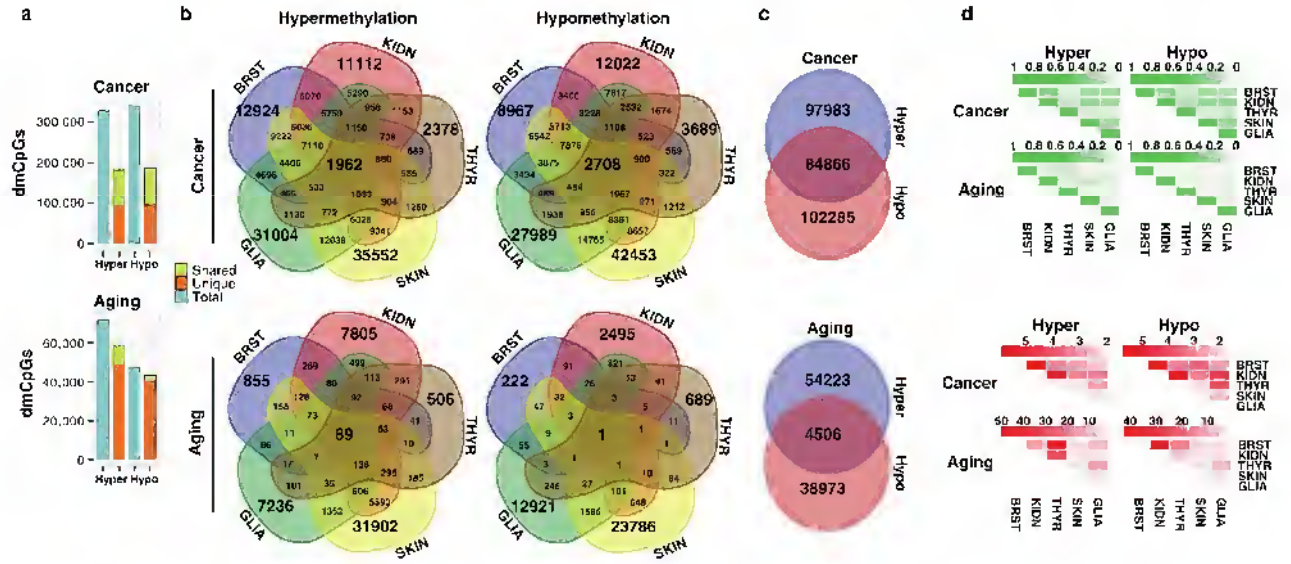


FIGURE 3 DNA methylation signatures of aging and cancer. (a) Stacked barplots indicating, in blue, the total number of hyper- and hypomethylated dmCpGs detected between all of the tissues analyzed in cancer and aging. Of these, the proportion of dmCpGs not shared between any tissues (unique, orange) or those shared by two or more tissues (shared, green) is shown in the adjacent barplot. (b) Venn diagrams depicting the number of differentially hyper- and hypomethylated CpGs in aging and cancer shared by the different tissues. (c) Venn diagrams showing the number and overlap of total nonredundant hyper- and hypomethylated dmCpGs detected in cancer and aging. (d) Heatmaps showing pairwise comparisons between sets of probes: in green, Jaccard Indices; in red, odd ratios (all enrichment Fisher's tests $p < .001$)

sites were common to at least two tissue types and only 89 (0.15%) hyper- and 1 (0.002%) hypomethylated CpG sites were common to all five tissue types analyzed (Figure 3a,b and Table S6). However, statistical analyses of the pairwise overlaps between the different sets of probes showed overall enrichment in every case, especially for aging (Figure 3d, Fisher's tests; all $p < .001$, Table S7). This over-enrichment was also revealed through a simulation of a random sampling of probes from the array (Figure S4a). Taken together, these results suggest that both cancer and aging manifest tissue-independent changes in DNA methylation.

We also identified a subset of dmCpG sites in aging and cancer that could potentially be either hyper- or hypomethylated depending on the tissue type involved (Figure 3c) and showed substantial under-enrichment (Fisher's test $p < .001$ for both,

ORs = 0.65 and 0.56; expected hypergeometric means, EHMs = 94,610 and 7,060; and Jaccard indices, JIs = 0.30 and 0.05, respectively). Interestingly, when examining dmCpGs shared by two or more tissues (Figure S4b), this under-enrichment became more pronounced such that CpGs that were thus affected in more than one tissue were less likely to behave differently in other tissues.

2.4 | Similar chromatin signatures of DNA hypermethylation in aging and cancer

To identify possible chromatin marks associated with hypermethylated CpG sites in aging and cancer, we compared the hypermethylated CpG sites identified in this study with previously published ENCODE and NIH Roadmap Epigenomics ChIP-seq data on the histone modifications H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K27me3 and H3K9me3 across 98 different cell and tissue types (see Section 4). The results confirmed an enrichment of hypermethylated CpG sites in repressive histone modifications H3K27me3 and H3K9me3 and active histone modifications H3K4me1 and H3K4me3 in both in aging and cancer (Figure 4a, upper panel; Table S8), with the H3K27me3 mark being the most consistent enrichment across all of the analyses. Notably, these similarities became more pronounced when examining dmCpGs shared by all five tissue types in cancer, or three of the tissue types in aging (low numbers of common probes, due to tissue-specificity of aging dmCpGs, hindered analysis of dmCpGs shared by more tissues). Interestingly, the embryonic stem cell signature was comparable to other tissue signatures, although, when present, the H3K4me3 mark was more evident in the aging context. Collectively, these results suggest that chromatin signatures of DNA hypermethylation are similar in aging and cancer.

2.5 | Distinct chromatin signatures of DNA hypomethylation in aging and cancer

To determine whether the chromatin signatures of DNA hypomethylation were also similar in aging and cancer, we compared the hypomethylated CpG sites identified in our study with data from the same histone modifications as described in the earlier analyses. Interestingly, the results showed that hypomethylated CpG sites in cancer were enriched in the repressive H3K9me3 histone modification, while in aging, hypomethylated CpGs were more enriched in the activating histone mark H3K4me1 (Figure 4a, lower panel; Table S8). There were though exceptions to this general trend: hypomethylated CpG sites in thyroid tumors were also enriched at H3K4me1, and hypomethylated DNA sequences in aged skin were mainly co-associated with H3K9me3-marked DNA regions. Nevertheless, the ratio H3K4me1/H3K9me3 was always higher in aging than in cancer (Figure 4b). Moreover, when analyzing the dmCpGs shared by all five tissues in cancer or at least three tissues in aging, these distinct chromatin signatures became much more evident. In sum, these results indicate that, in contrast to DNA hypermethylation, chromatin signatures of DNA hypomethylation in aging and cancer differ considerably.

After deriving the chromatin signatures, we performed validation analyses on two additional datasets: the first related to tissue from TCGA control and lung adenocarcinoma and the second to whole blood from the classical Hannum et al. (2013) dataset (Figure S5; see Table S9 for additional information). Interestingly, we were unable to find aging-related methylation changes in normal lung tissue using our pipeline. The magnitude and distribution of the hyper- and hypomethylation changes in lung cancer and whole blood aging followed the same trend as observed for the other datasets (Figure S5a, see Table S2 for a list of dmCpGs). The histone enrichment analyses revealed the same hypermethylation signature previously found for cancer and aging, and very clear and different hypomethylation signatures of H3K9me3 for lung cancer and H3K4me1/3 for whole blood aging (Figure S5b and Table S8).

Finally, we compared the overlap between either hypermethylated or hypomethylated CpGs across tumors and their corresponding age-related tissues (Figure 4c). This approach revealed that the overlap between hypermethylated CpGs (24,442 CpGs) was higher than expected by chance (Fisher's test $p < .001$, OR = 2.61; EHM = 14,689; JI = 0.20) (Figure 4c, upper panel). However, despite the overlap between hypomethylated CpGs (14,521) also being slightly higher than expected (Fisher's test $p < .001$, OR = 1.60; EHM = 11,021; JI = 0.12) (Figure 4c, lower panel), the overall trend observed in this case was weaker than for the hypermethylated CpGs. Furthermore, most of the hypomethylated probes shared by cancer and aging belonged to skin dmCpGs, providing evidence for its similar cancer and aging hypomethylation signatures. Removing skin tissue from the analysis (Figure S6) caused the observed over-enrichment to disappear in the case of DNA hypomethylated probes, although it remained in the hypermethylation scenario (Fisher's tests, both $p < .001$, ORs = 0.8 and 3.0; EHMs = 5,357 and 7,568; JIs = 0.04 and 0.12, respectively), reinforcing the observed overlapping differences between hyper- and hypomethylated probes.

2.6 | Functional characterization of differentially methylated sites in aging and cancer

To determine the possible functional consequences and genomic coincidence of the different histone marks of DNA hyper- and hypomethylation in aging and cancer, we performed an enrichment analysis of NIH Roadmap and ENCODE Hidden Markov Model (HMM) defined "chromatin states" across the same 98 human cell and tissue types used in the previous analyses (see Section 4). In total, 18 states were used for the segmentation of the genome, which were then grouped to highlight predicted functional elements.

As suggested by the earlier chromatin signature analyses, hypermethylated CpGs in both aging and cancer were enriched in states associated with bivalent chromatin domains (i.e., those formed by the combination of repressive histone mark H3K27me3 and activating histone marks H3K4me1/3), polycomb repressive domains, and repeat/ZNF genes. These patterns became more evident when

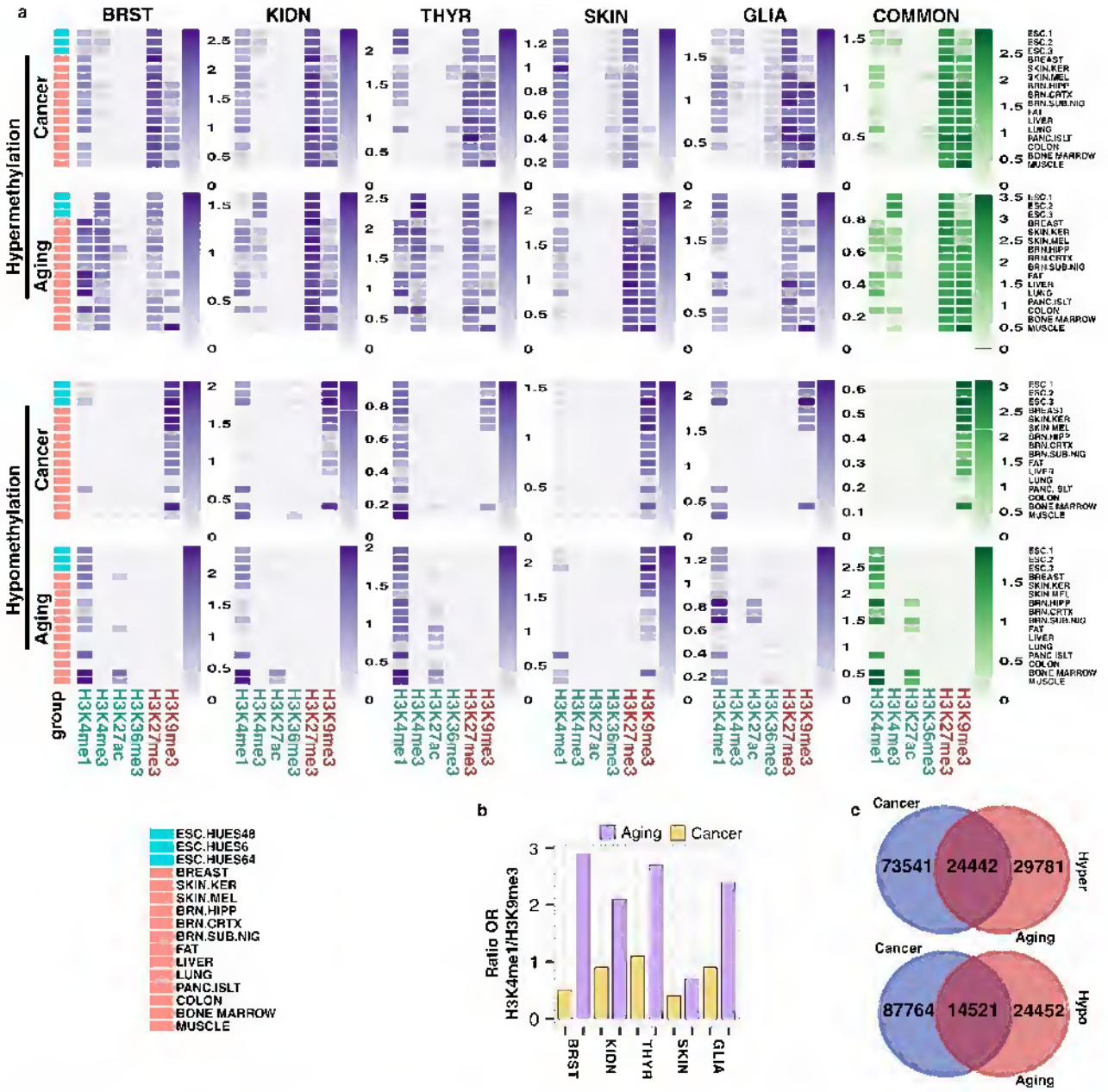


FIGURE 4 Distinct chromatin signatures of hyper- and hypomethylation in aging and cancer. (a) Heatmaps depicting significant ($p < .05$) over-enrichment of hyper- and hypomethylated dmCpG sites with different histone marks in aging and cancer, in a selection of 16 cell and tissue types (see Table S8 for 98 full cell and tissue types). Color code indicates the significant enrichment based on log2 odds ratio (OR). Common signatures are calculated from hyper- and hypomethylated dmCpGs shared between five tissues for cancer (1,962 and 2,708 probes, respectively) or three tissues for aging (904 and 106 probes, respectively) (see Table S6 for CpG lists). (b) Barplots indicating the ratio of OR for the H3K4me1/H3K9me3 marks associated with hypomethylated dmCpGs in aging and cancer. The ratio is calculated by taking the mean of OR of those tracks with significant over-enrichment for each histone mark and dividing the obtained numbers. (c) Venn diagrams showing the number and overlap of total nonredundant hyper- and hypomethylated dmCpGs detected in cancer and aging. dmCpGs that were only hypermethylated or only hypomethylated between all tissues were chosen for the comparison

examining the dmCpGs shared by all five cancer tissues or at least three aging tissues (Figure 5a; see Figure S7 for tissue-specific signatures and Table S10 for full data in all 98 cell and tissue types). Hypomethylated CpG sites in cancer were enriched in chromatin states associated with heterochromatin and repeat/ZNF gene domains and, to a lesser extent, polycomb repressive domains. In contrast, DNA hypomethylation in aging was primarily associated with chromatin states related to DNA enhancers. Again, these marks

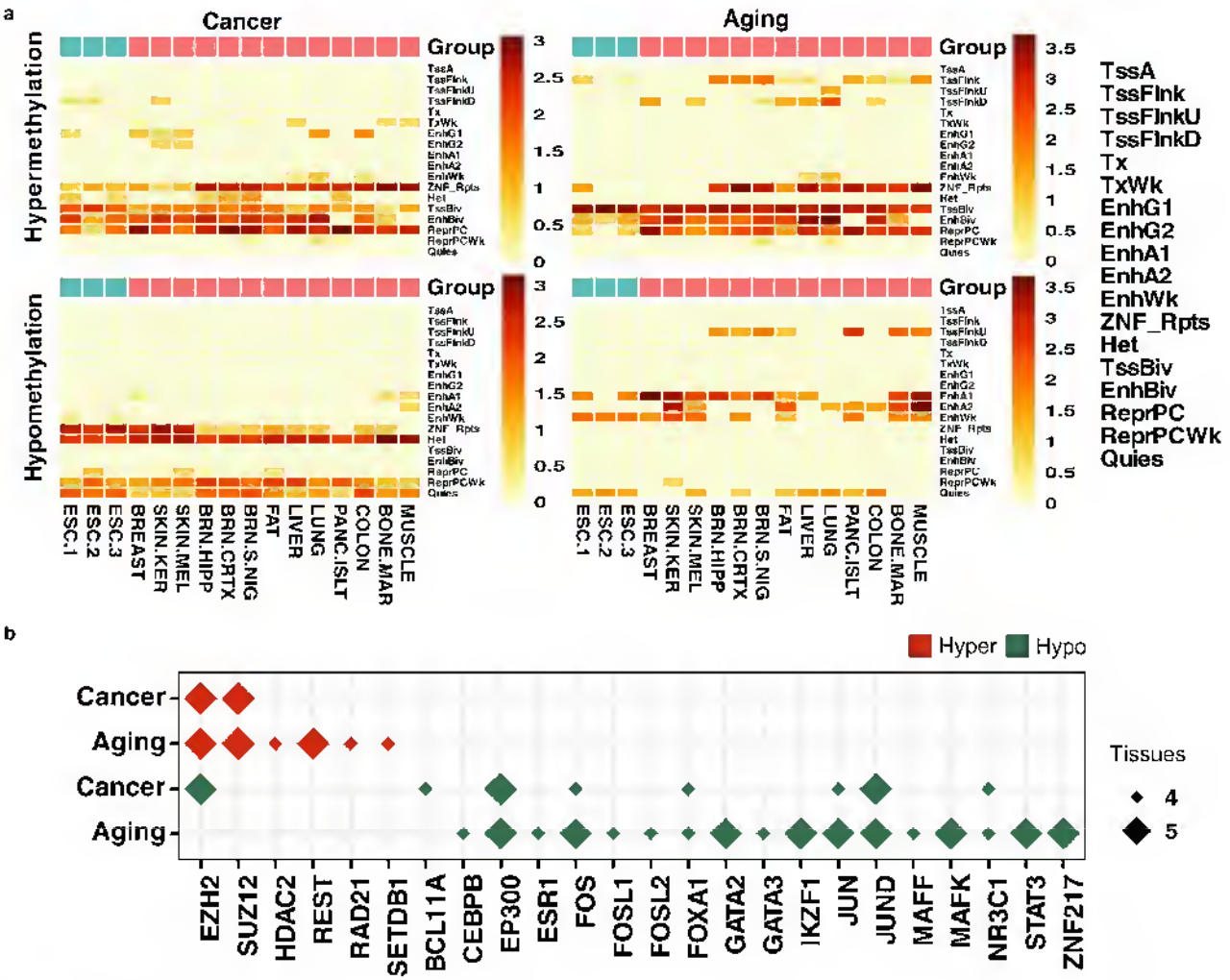


FIGURE 5 Differential chromatin state signatures of hyper- and hypomethylation in aging and cancer. (a) Heatmaps displaying significant ($p < .05$) over-enrichment of hyper- and hypomethylated dmCpG sites with different chromatin states in aging and cancer, in a selection of 16 cell and tissue types (see Figure S7 for tissue-specific signatures, and Table S10 for 98 full cell and tissue types). Color code indicates the significant enrichment based on log2 odds ratio (OR). Common signatures are calculated from hyper- and hypomethylated dmCpGs shared between five tissues for cancer (1,962 and 2,708 probes, respectively) or three tissues for aging (904 and 106 probes, respectively) (see Table S6 for CpG lists). (b) Panel indicating significant ($p < .05$) transcription factor enrichment at hyper- and hypomethylated dmCpG sites in aging and cancer that occurred in at least four or five tissues (see Figure S8 for tissue-specific results and Table S11 for full data). Only the most representative transcription factors (those that appeared as significantly over-enriched in at least three tracks or with an OR > 3 in any track) were selected for data representation

were more pronounced in aging dmCpGs shared by at least three tissues. As occurred with chromatin signatures, hypomethylation chromatin state differences were weaker in skin and thyroid, albeit that the ratio of change of chromatin states always followed the same behavior (data not shown). Collectively, these results support the notion that DNA hypomethylation might have a different functional role in aging as compared to cancer.

To increase our understanding of the functional context of the chromatin signatures characterized, we compared the dmCpG sites identified in this study with publicly available ENCODE ChIP-seq data on transcription factor binding sites in 689 datasets corresponding to 188 transcription factors across 91 different cell types (see Section 4) (Figure 5b; see Figure S8 and Table S11 for full tissue-specific results). As expected, hypermethylated CpGs in aging and cancer were associated in all tissues with the presence of EZH2 and SUZ12, components of the polycomb complex which directly deposits the H3K27me3 mark. Interestingly, aging hypermethylation dmCpGs were specifically associated with other types of transcription factors in various tissues, such as REST, HDAC2, RAD21, and SETDB1. Transcription factor enrichment at hypomethylated dmCpG sites was more heterogeneous than, but different from that of hypermethylated sites. Enrichment of similar factors was found for cancer and aging, for example, EP300, FOS, and JUN, among others. As observed before, specific aging enrichment was found, such as

GATA2/3. In this case, aging hypomethylated dmCpG sites tended to display a more marked enrichment of most of the cancer hypomethylation factors and, additionally, revealed the presence of other family- or function-related proteins, like FOSL1/2, MAFF, MAFK, and STAT3. When examining enrichment at common dmCpG sites shared by different tissues in cancer and aging, the initial observations were further confirmed (Figure S9).

Gene ontology analyses (Figure 6a; Table S12; see Figure S10 for tissue-specific results) revealed that hypermethylated CpGs in both processes belonged to genes that were mainly related to developmental functions. While genes containing hypomethylated CpGs in cancer were associated with extracellular signaling, those for aging were, in general, much less enriched in any gene ontology. In the case of KEGG pathways (Figure 6a; Table S12 and Figure S10), hypermethylated CpGs in both cancer and aging shared enrichment for several ontologies, many related to cell metabolic and signaling pathways. In this respect, hypomethylated CpGs in cancer had some ontologies in common, while others were specific. Once again, aging hypomethylated CpGs exhibited much less enrichment in any function.

To exemplify the similarities and disparities observed for DNA methylation in aging and cancer, we focused on a number of significant dmCpGs from two particular genomic regions, located in chromosomes 11 and 16 (Figure 6b,c). We observed a substantial correlation between bivalent posttranslational histone modifications, especially H3K27me3 and H3K4me1/3, and the presence of hypermethylated probes in aging and cancer. On the other hand, DNA hypomethylated regions were more frequently located near H3K9me3 or H3K4me1 peaks (bottom panel Figure 6b,c) as outlined in our previous histone enrichment analyses. A detailed inspection of the common genes with most abundant dmCpGs in aging and cancer revealed a similar trend toward DNA hypermethylation at the boundaries of the gene *PAX6* (Figure 6b, top panel). Interestingly, a representative set of cancer cell lines, as well as fibroblasts derived from patients with Hutchinson-Gilford progeria, also display higher levels of DNA methylation when

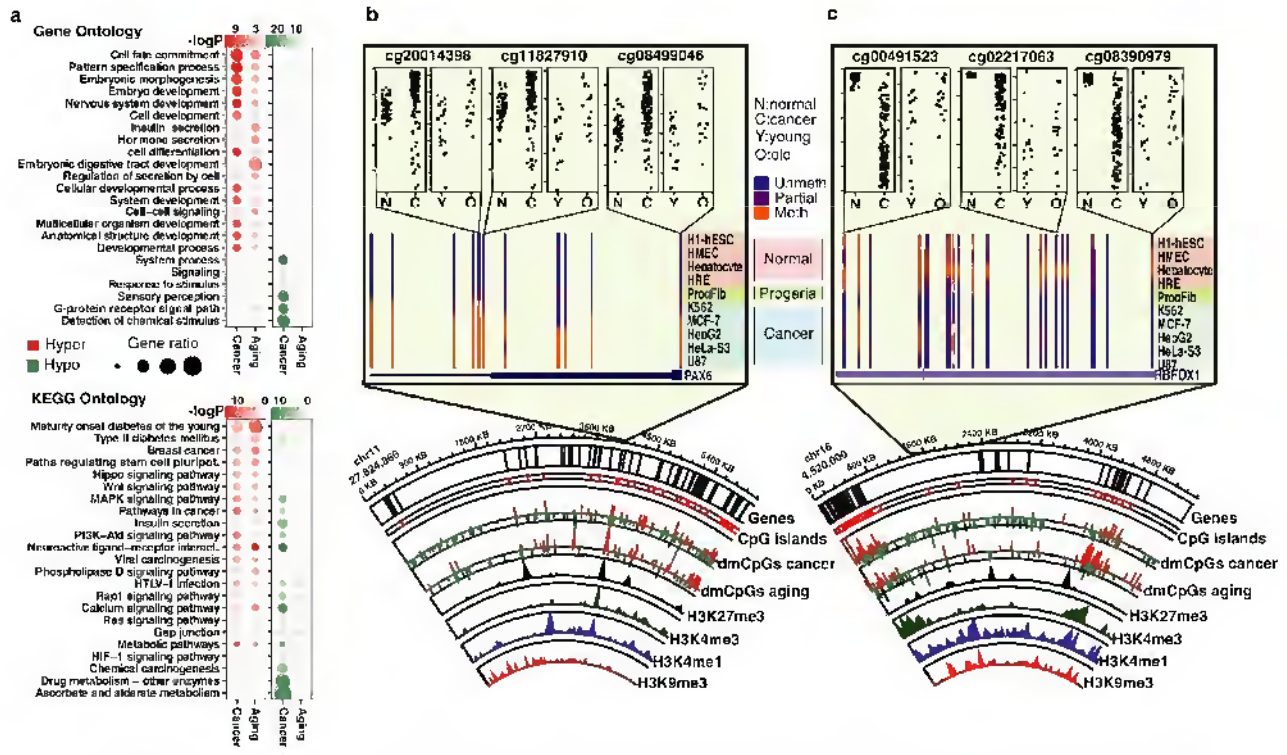
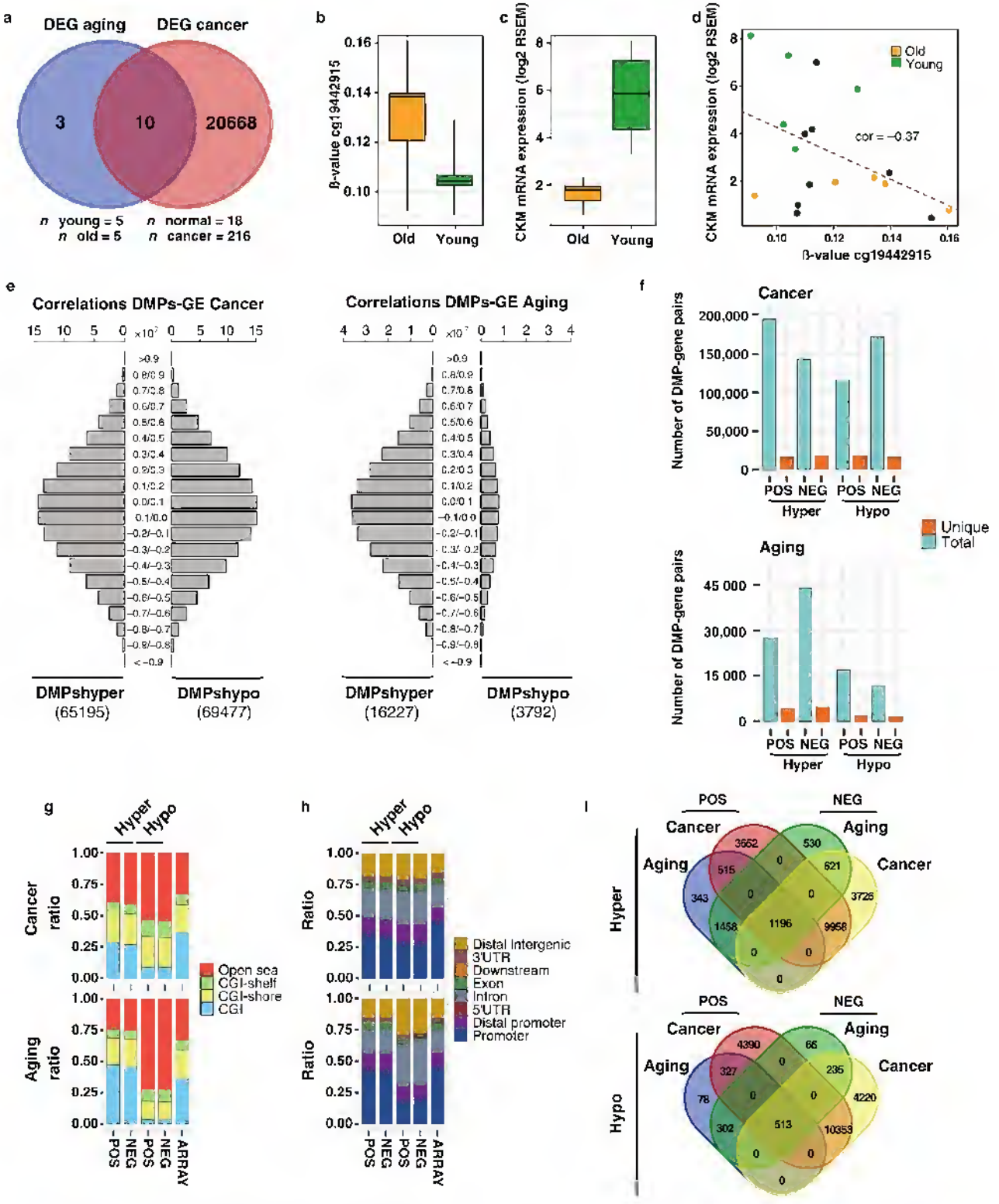


FIGURE 6 Dissimilar functional context of differentially methylated CpGs in aging and cancer. (a) Panels indicating gene and KEGG pathway ontology enrichment for common hyper- and hypomethylated dmCpGs shared between five tissues for cancer (1,962 and 2,708 probes, respectively) or three tissues for aging (904 and 106 probes, respectively) (see Table S6 for CpG lists and Figure S9). Color code indicates the significance of the over-enrichment based on log10 *p*-value. Size of circles indicates gene ratio, calculated as the ratio of a number of identified hits with respect to the total number of components in a given ontology. (b) and (c) Circular representation of illustrative genomic locations indicating hyper- (red) and hypomethylated (green) dmCpG sites in cancer and aging. Inner tracks display chromatin marks (H3K27me3, H3K4me3, H3K4me1, and H3K9me3, respectively), corresponding to ChIP-seq peak data from ESC E014 NIH Roadmap track (circos, lower panel). Two examples of hypo- and hypermethylated genes are highlighted from the circular figures, displaying methylation data obtained from ENCODE/HAIB methyl450 track set from UCSC Genome Browser (hg19) and plots of methylation values extracted from a representative example of our aging and cancer analyses corresponding to glia tissue (upper panel, depicted CpGs also displayed a similar trend in kidney and skin analyses, data not shown)



compared to normal cells in these differentially methylated regions (Figure 6b, middle panel). On the contrary, the abovementioned pattern was mainly reversed in the case of the *RBFOX1* gene, located in a region which was preferentially hypomethylated in cancer (Figure 6c, top and middle panel).

2.7 | Correlations between CpG methylation and gene expression in aging and cancer

Lastly, we looked at the possible impact in the control of gene expression of the methylation changes previously found. To address

FIGURE 7 Relationships between DNA methylation and gene expression in aging and cancer. (a) Venn diagrams illustrating the overlap between DEGs in aging and cancer in the KIRC dataset (see Table S13 for DEG lists). (b) Boxplot depicting the DNA methylation β-values of the CpG cg19442915 in old and young individuals (n = 5) from the KIRC aging condition. (c) Boxplot showing the gene expression values (RSEM) of the CKM gene in old and young individuals (n = 5) from the aging condition of the KIRC dataset. (d) Scatterplot showing the Spearman correlation between DNA methylation (cg19442915) and gene expression (CKM gene) in 18 normal kidney samples. Colored dots indicate old or young individuals used for the aforementioned aging comparisons. (e) Histograms representing the number of pairwise correlations that are contained in a given correlation window (from 1 to −1) obtained as the result of computing the correlation between β-values of dmCpGs identified in cancer or aging and gene expression levels (RSEM) of genes expressed in the KIRC dataset. The number of dmCpGs used for each of the comparisons is indicated at the bottom. (f) Barplots depicting the number of total (blue) and unique (orange) dmCpG-gene pairs identified in the previous analysis which displayed correlation scores above 0.9 (pos) or below −0.9 (neg) in cancer (top) or in aging (bottom) conditions. (g) Stacked barplots indicating relative distribution of unique dmCpGs obtained from the previous correlations according to their CpG island status. (h) Stacked barplots indicating relative distribution of unique dmCpGs obtained from the previous correlations according to their gene location status. (i) Venn diagrams illustrating the overlap between dmCpGs identified in aging and in cancer which displayed strong positive or negative correlations (>0.9 or <−0.9) with genes expressed in the normal kidney dataset

this issue, we focused on kidney tissue (KIRC) as this TCGA dataset displayed a reasonable number of control and cancer patients with paired methylation and gene expression data. We initially performed differential gene expression analyses comparing young vs. old or normal vs. tumoral kidney samples (Figure 7a and Table S13). These results allowed us to identify a total of 13 and 20,678 differentially expressed genes (DEGs) in aging and cancer conditions, respectively. The majority of the aging DEGs were also found in cancer, including, for example, the *CKM* gene, which contained a dmCpG in the proximity of its promoter (Figure 7b), was differentially expressed in both processes (Figure 7c) and displayed a considerable negative correlation between DNA methylation and gene expression in normal kidney (Spearman r = −.37, Figure 7d). To further explore the potential relationships between CpG methylation and gene expression in these processes, and due to the reduced number of DEGs observed in the aging context, we decided to perform all potential pairwise correlations between DNA methylation and gene expression using cancer- or aging-related dmCpGs and genes expressed in a subset of normal kidney tissue samples (n = 18). This approach enabled us to quantify the extent to which CpGs whose methylation status changes in cancer and aging originally influence gene expression in normal tissue.

We computed a total of $2.58e^{09}$ and $3.84e^{08}$ correlations between cancer- and aging-related dmCpGs, respectively, and genes expressed in the normal KIRC dataset (Figure 7e). Despite the considerable difference in a number of dmCpGs between cancer and aging, when compared to the total possible number of correlations, we found similar percentages of strong correlations between DNA methylation and gene expression in both processes (Table S14). Moreover, these proportions were also higher than those observed when sampling random probes from the array and computing their correlations (see Figure S11). These results indicate that both cancer- and aging-related dmCpGs are enriched in CpGs that can influence, to some extent, gene expression in kidney tissue.

A more detailed inspection of the strongest correlations (>= 0.9 or <= −0.9) identified in these datasets revealed that, in cancer, the number of unique dmCpG-gene pairs remained similar (~20,000) regardless of the direction of the observed correlation (Figure 7f, top). Furthermore, while the number of unique dmCpG-gene pairs in the aging context was much reduced (~3,000), the proportion compared to the total number of strong correlations observed in a given dataset remained, to a great extent, similar (Figure 7f, bottom). Interestingly, the differences between the genomic distributions of the unique hyper- and hypomethylated dmCpG-gene pairs identified in aging or in cancer followed the same trend to those observed for the genomic distribution of the hyper- and hypomethylated dmCpGs identified in each of these processes (Figure 2b,c), with hypermethylated dmCpGs being more enriched in CpG islands in aging as compared to the array (Fisher's tests; both p < .001, ORs = 2.0, 1.4, for positively and negatively correlated dmCpGs, respectively), in contrast to the enrichment at open sea locations (Fisher's tests; all p < .001, ORs = 4.6, 4.1, 2.3 and 2.5 for positively and negatively correlated dmCpGs in aging and cancer, respectively) and intronic regions (Fisher's tests p < .002, <.03, <.001 and <.001, ORs = 2.2, 1.9, 1.6, and 1.7 for positively and negatively correlated dmCpGs in aging and cancer, respectively) of the hypomethylated dmCpGs in both processes (Figure 7g,h). It is worth noting that the distribution of the unique hypermethylated dmCpGs which also control gene expression was more enriched in open sea locations as compared to dmCpGs in general in both aging and cancer (Figure 7g as compared to Figure 2b).

Finally, we compared the unique dmCpGs that displayed strong correlations between DNA methylation and gene expression in aging or in cancer (Figure 7i). We observed an extensive overlap between probes that displayed positive or negative correlations in the two processes (Fisher's tests, all p < .001, ORs = 908, 2244, 205, and 171; JIs = 0.57, 0.54, 0.57, and 0.54 for aging and cancer hyper- and hypomethylated CpGs, respectively). This fact might explain their similar genomic distributions (Figure 7g,h), indicating that most of these dmCpGs could play a dual role in the control of their different gene expression targets. We also observed a considerable overlap between dmCpGs associated with gene expression identified in aging and cancer processes (Fisher's tests, both p < .001, ORs = 5.1 and 9.1; JIs = 0.06 and 0.03, for hyper- and hypomethylated CpGs, respectively). Interestingly, regardless of whether the DNA methylation change was toward hyper- or hypomethylation, ~ 60–70% of the aging-related dmCpGs which controlled gene expression were

also present in the group of cancer-related dmCpGs (Figure 7i). These results point toward similarities of cancer- and aging-related dmCpGs in the control of gene expression in normal tissue, despite the fact that the number of cancer-related dmCpGs is clearly larger than their aging counterparts.

3 | DISCUSSION

Although it is widely accepted that cancer is an age-dependent disease, the underlying molecular mechanisms are still poorly characterized. In this work, we have looked at the similarities and differences in epigenetic changes associated with cancer and aging.

In agreement with previously published data (Fernández et al., 2015), we observed that the number of aging- and cancer-associated DNA methylation changes was variable and, in the case of aging, had a marked tissue type-dependent component. In general, cancer displayed strong and bidirectional changes, while, strikingly, hypermethylated CpG sites were predominantly observed for the aging process. These results, which are ostensibly in contrast with the classically described global hypomethylation changes in cancer and aging, might potentially arise from the limitations of our study. As the methylation arrays used in our analyses mainly interrogate genetic elements and do not include repeated DNA, which covers a substantial fraction of the genome and frequently loses DNA methylation in tumors and aged cells, the genome-wide landscape may be different (Ehrlich, 2009). Nonetheless, epigenetic signatures have been successfully derived previously using array technology (Fernández et al., 2015; Rakyan et al., 2010; Teschendorff et al., 2010) and our results are in line with recent studies which report no global decreases in DNA hypomethylation with aging in diverse mouse tissues, such as liver (Cole et al., 2017; Hahn et al., 2017), hippocampus (Masser et al., 2017), or hematopoietic stem cells (Beerman et al., 2013; Sun et al., 2014), thus strengthening the validity of our observations.

The changes in cell type composition that occur with age and cancer are also well-known confounding factors that could affect our datasets (Zheng et al., 2017). However, the application of the SVA method of correction (and Houseman correction for blood) and the use of a pure-cell dataset such as the glia dataset (Guintivano et al., 2013) allowed us to tackle this issue in two different ways. Additionally, the use of the blood validation dataset (Hannum et al., 2013) allowed us to verify the reliability of our workflow, as in terms of whole blood dmCpGs we obtained 89% concordance with previous studies using the same data (Fernández et al., 2015).

When analyzing the genomic distribution of dmCpGs and, in line with previously published reports (Cruickshanks et al., 2013; Kulis et al., 2012; Yuan et al., 2015), we found that hypomethylated CpGs were enriched at open sea DNA regions, principally intronic and intergenic, irrespective of the type of process. The distribution of hypermethylated CpGs was found to be similar to that of the array, which is to a certain extent to be expected because it was designed to interrogate a promoter- and CpG dense-biased portion of the genome. Nonetheless, hypermethylation changes always occurred in far more CpG-dense regions than hypomethylation changes (Day et al., 2013; Yuan et al., 2015), and this observed effect was especially noticeable for aging dmCpGs.

When studying the potential effect of tissue type on DNA methylation changes, we found, in agreement with recently published data (Chen, Breeze, Zhen, Beck & Teschendorff, 2016), that DNA methylation changes in different tumor types were surprisingly similar, regardless of the tendency of the alteration. This observation is conceptually relevant because it has classically been considered that different tumor types are characterized by specific DNA methylation signatures (Ehrlich & Jiang, 2005; Portela & Esteller, 2010). In this sense, our data confirm that, although different tumor types might display specific DNA methylation patterns, there is a significant common nexus between them. The analysis of the DNA methylation changes observed with respect to the aging process also revealed a significant overlap between tissue types, although it is possible that these results are affected by the variability in the sizes of the sets of probes detected in the aging analysis.

Our data revealed that dmCpGs shared by two or more tissues were much less likely to have different behaviors in other tissues, perhaps pointing toward nonstochastic and possibly functional roles for these CpGs.

The systematic DNA methylation analyses described in this study confirm that DNA hypermethylation in aging and cancer is associated with the same set of histone marks, including the repressive H3K27me3 and H3K9me3 marks, and the activating H3K4me1/3 posttranslational modifications. Chromatin state analysis revealed that the hypermethylation-associated H3K27me3 and H3K4me1/3 marks configured bivalent chromatin domains, as has been extensively described in embryonic stem cells (Fernández et al., 2015; Ohm et al., 2007; Rakyan et al., 2010; Schlesinger et al., 2007; Teschendorff et al., 2010; Widschwendter et al., 2007). Moreover, our data reveal that this chromatin signature is not restricted to only embryonic stem cells, but rather this trend should be considered an extended, global tissue-independent chromatin signature of DNA hypermethylation in aging and cancer. Interestingly, Chen and colleagues (Chen et al., 2016) have recently demonstrated that normal tissue signatures are better predictors of DNA hypermethylation changes than ESC signatures. Furthermore, hypermethylation changes were also associated with the repressive histone mark H3K9me3 (Ohm et al., 2007), which was correlated to ZNF genes and DNA repeats in our chromatin state analyses, and which might have a potential relationship with the malignant transformation process (Severson, Tokar, Vrba, Waalkes & Futscher, 2013).

Regarding DNA hypomethylation, our results showed that age-related DNA hypomethylation is associated with the activating histone posttranslational modification H3K4me1, which supports previously published data (Fernández et al., 2015). A slight tendency for the enrichment of H3K27Ac, a histone mark characteristic of active enhancers (Creyghton et al., 2010), was also detected in our analyses. Intriguingly, the chromatin signature of DNA hypomethylation in cancer was substantially different, being primarily enriched in the

posttranslational repressive histone modification H3K9me3, a relationship that has been investigated in colon and breast cancer (Berman et al., 2011; Hon et al., 2012). This observation might be conceptually relevant because DNA methylation has been proposed to be a molecular link between aging and cancer (Fraga, Agrelo & Esteller, 2007; Klutstein, Nejman, Greenfield & Cedar, 2016). However, our results suggest that the role of DNA methylation as a possible link between aging and cancer is more complex than previously proposed. Importantly, even though many of the observed DNA methylation changes in aging were not shared by tissues, we were able to describe a common chromatin signature characteristic of the aging process.

Regarding the analysis of the chromatin states, DNA hypomethylation in cancer was associated with heterochromatin DNA regions, which is in line with previous work (Berman et al., 2011; Kulis et al., 2012). In contrast, chromatin marks of DNA hypomethylation in aging were associated with enhancers, reinforcing previous observations performed with the Infinium HumanMethylation27K Beadchip platform (Day et al., 2013). As DNA methylation changes in enhancers have been shown to play an important role in gene regulation (Aran, Sabato & Hellman, 2013; Blattler et al., 2014; Heyn et al., 2016), our results suggest that DNA hypomethylation during aging might have a different functional role in gene regulation compared to DNA hypomethylation changes in cancer.

With regard to the potential effectors of the distinct chromatin signatures, enrichment analyses of transcription factors revealed the presence of EZH2 and SUZ12 polycomb components at DNA hypermethylated sites, both in cancer and aging. Specific aging hypermethylation-associated factors were also observed in our comparisons, such as REST, which has been reported in previously published data in blood (Yuan et al., 2015), and has also been correlated with longevity (Lu et al., 2014). Concerning DNA hypomethylation, transcription factors such as FOS, JUN, and JUND were detected at both cancer and aging hypomethylated CpG sites, but again aging displayed stronger and more varied enrichment, and included the presence FOSL1/2, other bZIP-domain factors like MAFF and MAFK, and STAT3, which has been associated with recruitment of the H3K4 methyltransferase SET9 at promoters (Yang et al., 2010). Altogether, these observations would imply that hypomethylation in aging displays a more marked functional context than that of cancer, exhibiting an increased enrichment of some factors also detected at cancer hypomethylated sites and other specific factors not found associated with tumoral changes.

Interestingly, our gene ontology analyses revealed similar gene functionalities affected by cancer and aging DNA hypermethylation, mainly related to developmental processes, which is in line with the methylation of bivalent chromatin promoters of developmental regulators in cancer and aging (Easwaran et al., 2012; Rakyan et al., 2010). On the other hand, DNA hypomethylation in cancer was mainly associated with functions identified with cellular signaling, and much lower enrichments in gene functions were found for hypomethylated CpGs in aging. A preponderance of nongenic enhancer hypomethylation in aging could potentially explain this absence of gene function association in our data.

To date, the potential relationships between DNA methylation and gene expression have only been systematically analyzed in a small subset of studies (Gevaert, Tibshirani & Plevritis, 2015; Gutierrez-Arcelus et al., 2013, 2015), and the potential effects of these relationships on aging and cancer are yet to be elucidated. To this end, we explored the establishment of potential correlations between these two processes using the TCGA-KIRC dataset. While most correlative studies focus on CpGs located at particular genomic regions, such as DNA promoters (Moarii, Boeva, Vert & Reyal, 2015) and cis-related correlations with the gene of interest (Gutierrez-Arcelus et al., 2015), we performed a nonbiased approach focusing on all the potential pairwise comparisons that could be identified between any significant dmCpG and the genes expressed in the context of normal kidney tissue. The limitations of these analyses did not allow us to distinguish between direct (i.e., mediated by the effects of the DNA methylation process itself) or indirect regulation of gene expression governed by the subsequent expression of other regulatory factors. Nonetheless, we observed that both aging and cancer dmCpGs influence gene expression to a similar extent, as these processes show the same proportions of strong correlations between DNA methylation and gene expression in normal kidney tissue. Moreover, we observed a similar number of positive and negative correlations between DNA methylation and gene expression, as described in Gutierrez-Arcelus et al., 2013, with most of these positively and negatively correlated dmCpGs overlapping substantially, suggesting that these CpG sites may play a dual role in the control of gene expression, or the involvement of other factors.

Finally, we found that most of the tumor types analyzed in this study did not show age-associated DNA methylation changes, which is in agreement with the reprogramming of the epigenetic clock in cancer cells (Horvath, 2013). As an exception, we identified age-associated dmCpGs in thyroid tumors. Uncommonly, thyroid cancer includes age as a prognostic indicator in most staging systems (Haymart, 2009), implying that these cancers do suffer age-related changes in their behavior. Intriguingly, this tissue displayed the lowest level of DNA methylation changes in cancer and one of the lowest in aging. Although the reasons for the different behavior of DNA methylation changes in thyroid are currently unknown, they could be related to the good prognosis that typically characterizes this type of tumor. In fact, Yang Z. and collaborator's "epiTOC" mitotic clock (Yang et al., 2016) shows thyroid cancer to have the least deviation from the behavior of its normal tissue. In this regard, future research should be conducted to address this issue.

In conclusion, our results indicate that hyper- and hypomethylated changes in aging and cancer each have similar genomic distributions and manifest tissue-independent trends in both processes. We confirm that chromatin signatures of DNA hypermethylation in aging and cancer are similar but, strikingly, we demonstrate that they are different for DNA hypomethylation. Collectively, our data suggest that the possible role of DNA methylation as a molecular link between aging and cancer is more complex than previously thought.

4 | EXPERIMENTAL PROCEDURES

4.1 | Data acquisition

HumanMethylation450 BeadChip (Illumina, California, USA) DNA methylation data (Level 3) corresponding to normal or primary tumors from breast (BRCA), kidney (KIRC), thyroid (THCA), skin (SKCM), and glia (GBM) samples were obtained from TCGA consortium via UCSC Xena Public Data Hub (http://xena.ucsc.edu/). Kidney, skin, and glia tissue datasets were enlarged for control cases using additional samples from KIRP (TCGA), skin (Bormann et al., 2016), and glia (Guintivano et al., 2013), respectively. Tissues were chosen based on disease prevalence, control data availability, and previous literature analyses to include both novel and pre-analyzed tissues. We also performed analyses on two supplementary datasets: lung adenocarcinoma (LUAD) and control TCGA data, and whole blood from a healthy cohort (Hannum et al., 2013). Extended information about the samples for each tissue type is shown in Tables 1, S1 and S9. Data were preprocessed as detailed in supporting information.

4.2 | Differential DNA methylation analyses

Differentially methylated probes (dmCpGs) in aging and cancer were calculated with the R package *limma* (version 3.32.2) (Ritchie et al., 2015). Briefly, a linear model between methylation levels as response variable, the variable of interest (either *age group* or *sample_type*), and surrogate variables (see Supplementary Methods) was fitted for each of the analyses, adjusting *p*-values to control for false discovery rate (FDR < 0.05). For the calculation of age-related dmCpGs, samples were divided into age quantiles in such a way as to obtain groups with sizes of $n = 15–30$, and comparisons were performed between the upper (*OLD*) and the lower (*YOUNG*) quantile. Cancer-related dmCpGs were calculated between normal tissue (*Solid Tissue Normal*) and tumor samples (*Primary Tumor*) as indicated in Tables 1, S1 and S2. Probes with M-value changes of <0.5 were not considered as dmCpGs, as has been suggested elsewhere (Du et al., 2010). Venn diagrams of relationships between dmCpGs were generated with the online resource provided by the UGent/VIB bioinformatics unit (http://bioinformatics.psb.ugent.be/webtools/Venn/). Further enrichment analyses were performed by means of two-sided Fisher's tests ($p < .05$ significance threshold), measuring effect size either by odds ratios (OR), or by the difference between observed counts and expected hypergeometric mean (EHM), employing appropriate backgrounds of interrogated probes for the given context.

Density of CpG (related to Figure 2a), CGI status and genomic region (related to Figure 2b,c), and analysis, DNA methylation age (related to Figure 1d), and gene and KEGG ontology (related to Figure 6a) analyses are further detailed in Supplementary Methods.

4.3 | Region set enrichment analysis

Enrichment analyses were performed with the R package *LOLA* (version 1.4.0) (Sheffield & Bock, 2016), which looks for over-enrichment by conducting one-sided Fisher's tests ($p < .05$ significance threshold), by comparing overlap of probes (10 bp probe-centered windows) with the dataset of interest. Enrichment of histone marks was determined using histone ChIP-seq peak tracks (H3K4me1, H3K4me3, H3K27me3, H3K36me3, H3K9me3, and H3K27ac marks) from 98 epigenomes (primary tissues, cultures, and cell lines) obtained from the NIH Roadmap and ENCODE projects (Bernstein et al., 2010; Consortium 2012) (datasets obtained from http://databio.org/regiondb) (see Table S8). The same method was employed for chromatin-segment analysis using NIH Roadmap's ChromHMM expanded 18-state model tracks for the same 98 epigenomes (see Figure S7 and Table S10, custom database generated with data obtained from http://egg2.wustl.edu/roadmap/). In a similar fashion, ChIP-seq peak tracks from ENCODE for transcription factor binding sites (TFBS) comprising 689 datasets corresponding to 188 TFs analyzed in 91 cell and tissue types were employed for TFBS enrichment analysis (http://databio.org/regiondb, see Table S11).

4.4 | Gene expression analyses

Gene expression data corresponding to RNAseq HTSeq-Counts from the GDC TCGA Kidney Clear Cell Carcinoma (KIRC) cohort were obtained from UCSC Xena Public Data Hub (http://xena.ucsc.edu/, dataset ID: TCGA-KIRC/Xena_Matrices/TCGA-KIRC.htseq_counts.tsv). Samples were filtered to fulfill the criteria of using only those cases with paired DNA methylation and gene expression data. Log2(count+1) data were further transformed to obtain integer count reads per gene condition. Nonvariable and low-expressed genes (sum of expression across all the samples <1,000 counts) were removed to reduce the number of noninformative conditions. Differential expression analyses were performed with the R package DESeq2 (version 1.16.1) (Love, Huber & Anders, 2014), using the standard workflow and parameters, defining differentially expressed genes if they satisfied $p < .05$ after adjustment for multiple testing. For gene expression and DNA methylation correlation analyses, RNAseqV2 log2(RSEM+1) normalized level 3 TCGA gene expression data were obtained for kidney normal tissue (KIRC) via UCSC Xena Public Data Hub (http://xena.ucsc.edu/). Samples were filtered so as to use only those with paired DNA methylation data. Nonvariable and low-expression genes (those with the sum of expression between all the samples of <10) were discarded. After filtering, pairwise Spearman correlations between DNA methylation level and gene expression level were calculated for all the combinations of probes and genes in normal kidney tissue samples, using probes that were previously detected dmCpGs in cancer and aging.

4.5 | Availability

All data generated during this study are included in this published article and its supplementary information files and are also available in the Zenodo public repository, https://doi.org/10.5281/zenodo.1086491.

ACKNOWLEDGMENTS

The authors are grateful to the members from the Cancer Epigenetics laboratory (FINBA, IUOPA) for their positive feedback and to Ronnie Lendrum for manuscript editing.

CONFLICT OF INTEREST

None declared.

AUTHORS' CONTRIBUTION

M.F.F., A.F.F., and G.F.B. conceived, coordinated, and supervised the study. M.F.F, R.F.P., and J.R.T. designed all aspects of the research and contributed equally to this work. R.F.P. and J.R.T. collected the data and performed computational analyses. M.F.F., A.F.F., R.F.P., and J.R.T. wrote the manuscript. All authors revised, read, and approved the final manuscript.

REFERENCES

Aran, D., Sabato, S., & Hellman, A. (2013). DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biology*, *14*, R21. https://doi.org/10.1186/gb-2013-14-3-r21

Beerman, I., Bock, C., Garrison, B. S., Smith, Z. D., Gu, H., Meissner, A., & Rossi, D. J. (2013). Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. *Cell Stem Cell*, *12*, 413–425. https://doi.org/10.1016/j.stem.2013.01.017

Berman, B. P., Weisenberger, D. J., Aman, J. F., Hinoue, T., Ramjan, Z., Liu, Y., … Laird, P. W. (2011). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina–associated domains. *Nature Genetics*, *44*, 40–46.

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., … Thomson, J. A. (2010). The NIH roadmap epigenomics mapping consortium. *Nature Biotechnology*, *28*, 1045–1048. https://doi.org/10.1038/nbt1010-1045

Blattler, A., Yao, L., Witt, H., Guo, Y., Nicolet, C. M., Berman, B. P., & Farnham, P. J. (2014). Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biology*, *15*, 469. https://doi.org/10.1186/s13059-014-0469-0

Bormann, F., Rodríguez-Paredes, M., Hagemann, S., Manchanda, H., Kristof, B., Gutekunst, J., … Lyko, F. (2016). Reduced DNA methylation patterning and transcriptional connectivity define human skin aging. *Aging Cell*, *15*, 563–571. https://doi.org/10.1111/acel.12470

Chen, Y., Breeze, C. E., Zhen, S., Beck, S., & Teschendorff, A. E. (2016). Tissue-independent and tissue-specific patterns of DNA methylation alteration in cancer. *Epigenetics Chromatin*, *9*, 10. Retrieved from http://www.epigeneticsandchromatin.com/content/9/1/10

Cole, J. J., Robertson, N. A., Rather, M. I., Thomson, J. P., McBryan, T., Sproul, D., … Adams, P. D. (2017). Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biology*, *18*, 58. https://doi.org/10.1186/s13059-017-1185-3

Consortium TEP (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*, 57–74.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., … Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 21931–21936. https://doi.org/10.1073/pnas.1016071107

Cruickshanks, H. A., McBryan, T., Nelson, D. M., VanderKraats, N. D., Shah, P. P., van Tuyn, J., … Adams, P. D. (2013). Senescent cells harbour features of the cancer epigenome. *Nature Cell Biology*, *15*, 1495–1506. https://doi.org/10.1038/ncb2879

Day, K., Waite, L. L., Thalacker-Mercer, A., West, A., Bamman, M. M., Brooks, J. D., … Absher, D. (2013). Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biology*, *14*, R102. https://doi.org/10.1186/gb-2013-14-9-r102

de Magalhães, J. P. (2013). How ageing processes influence cancer. *Nature Reviews Cancer*, *13*, 357–365. https://doi.org/10.1038/nrc3497

DePinho, R. A. (2000). The age of cancer. *Nature*, *408*, 248–254. https://doi.org/10.1038/35041694

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, *11*, 587. https://doi.org/10.1186/1471-2105-11-587

Easwaran, H., Johnstone, S. E., Van Neste, L., Ohm, J., Mosbruger, T., Wang, Q., … Baylin, S. B. (2012). A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Research*, *22*, 837–849. https://doi.org/10.1101/gr.131169.111

Ehrlich, M. (2009). DNA hypomethylation in cancer cells. *Epigenomics*, *1*, 239–259. https://doi.org/10.2217/epi.09.33

Ehrlich, M., Jiang, G. (2005). DNA Hypo- vs. Hypermethylation in Cancer. In: *DNA Methylation and Cancer Therapy*. Medical Intelligence Unit. Springer, Boston, MA. pp. 31–41. Retrieved from http://link.springer.com/chapter/10.1007/0-387-27443-X_3

Fernández, A. F., Bayón, G. F., Urdinguio, R. G., Toraño, E. G., García, M. G., Carella, A., … Fraga, M. F. (2015). H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells. *Genome Research*, *25*, 27–40. https://doi.org/10.1101/gr.169011.113

Fraga, M. F., Agrelo, R., & Esteller, M. (2007). Cross-Talk between aging and cancer: The epigenetic language. *Annals of the New York Academy of Sciences*, *1100*, 60–74. https://doi.org/10.1196/annals.1395.005

Fraga, M. F., & Esteller, M. (2007). Epigenetics and aging: The targets and the marks. *Trends in Genetics*, *23*, 413–418. https://doi.org/10.1016/j.tig.2007.05.008

Gevaert, O., Tibshirani, R., & Plevritis, S. K. (2015). Pancancer analysis of DNA methylation-driven genes using MethylMix. *Genome Biology*, *16*, 17. https://doi.org/10.1186/s13059-014-0579-8

Guintivano, J., Aryee, M. J., & Kaminsky, Z. A. (2013). A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics*, *8*, 290–302. https://doi.org/10.4161/epi.23924

Gutierrez-Arcelus, M., Lappalainen, T., Montgomery, S. B., Buil, A., Ongen, H., Yurovsky, A., … Dermitzakis, E. T. (2013). Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife*, *2*, e00523.

Gutierrez-Arcelus, M., Ongen, H., Lappalainen, T., Montgomery, S. B., Buil, A., Yurovsky, A., … Dermitzakis, E. T. (2015). Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLoS Genetics*, *11*, e1004958. https://doi.org/10.1371/journal.pgen.1004958

Hahn, O., Grönke, S., Stubbs, T. M., Ficz, G., Hendrich, O., Krueger, F., … Partridge, L. (2017). Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. *Genome Biology*, *18*, 56. https://doi.org/10.1186/s13059-017-1187-1

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., … Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, *49*, 359–367. https://doi.org/10.1016/j.molcel.2012.10.016

Haymart, M. R. (2009). Understanding the relationship between age and thyroid cancer. *The Oncologist*, *14*, 216–221. https://doi.org/10.1634/theoncologist.2008-0194

Heyn, H., Li, N., Ferreira, H. J., Moran, S., Pisano, D. G., Gomez, A., … Esteller, M. (2012). Distinct DNA methylomes of newborns and centenarians. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 10522–10527. https://doi.org/10.1073/pnas.1120658109

Heyn, H., Vidal, E., Ferreira, H. J., Vizoso, M., Sayols, S., Gomez, A., … Esteller, M. (2016). Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. *Genome Biology*, *17*, 11. https://doi.org/10.1186/s13059-016-0879-2

Hon, G. C., Hawkins, R. D., Caballero, O. L., Lo, C., Lister, R., Pelizzola, M., … Ren, B. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Research*, *22*, 246–258. https://doi.org/10.1101/gr.125872.111

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*, R115. https://doi.org/10.1186/gb-2013-14-10-r115

Klutstein, M., Nejman, D., Greenfield, R., & Cedar, H. (2016). DNA methylation in cancer and aging. *Cancer Research*, *76*, 3446–3450. https://doi.org/10.1158/0008-5472.CAN-15-3278

Kulis, M., Heath, S., Bibikova, M., Queirós, A. C., Navarro, A., Clot, G., … Martín-Subero, J. I. (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nature Genetics*, *44*, 1236–1242. https://doi.org/10.1038/ng.2443

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, T., Aron, L., Zullo, J., Pan, Y., Kim, H., Chen, Y., … Yankner, B. A. (2014). REST and stress resistance in ageing and Alzheimer's disease. *Nature*, *507*, 448–454. https://doi.org/10.1038/nature13163

Masser, D. R., Hadad, N., Porter, H. L., Mangold, C. A., Unnikrishnan, A., Ford, M. M., … Freeman, W. M. (2017). Sexually divergent DNA methylation patterns with hippocampal aging. *Aging Cell*, *16*, 1342–1352. https://doi.org/10.1111/acel.12681

Moarii, M., Boeva, V., Vert, J.-P., & Reyal, F. (2015). Changes in correlation between promoter methylation and gene expression in cancer. *BMC Genomics*, *16*, 873. https://doi.org/10.1186/s12864-015-1994-2

Ohm, J. E., McGarvey, K. M., Yu, X., Cheng, L., Schuebel, K. E., Cope, L., … Baylin, S. B. (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and silencing in adult cancers. *Nature Genetics*, *39*, 237–242. https://doi.org/10.1038/ng1972

Portela, A., & Esteller, M. (2010). Epigenetic modifications and human disease. *Nature Biotechnology*, *28*, 1057–1068. https://doi.org/10.1038/nbt.1685

Rakyan, V. K., Down, T. A., Maslau, S., Andrew, T., Yang, T.-P., Beyan, H., … Spector, T. D. (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Research*, *20*, 434–439. https://doi.org/10.1101/gr.103101.109

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*, e47. https://doi.org/10.1093/nar/gkv007

Rodriguez-Paredes, M., & Esteller, M. (2011). Cancer epigenetics reaches mainstream oncology. *Nature Medicine*, *17*, 330–339. https://doi.org/10.1038/nm.2305

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., … Cedar, H. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nature Genetics*, *39*, 232–236. https://doi.org/10.1038/ng1950

Severson, P. L., Tokar, E. J., Vrba, L., Waalkes, M. P., & Futscher, B. W. (2013). Coordinate H3K9 and DNA methylation silencing of ZNFs in toxicant-induced malignant transformation. *Epigenetics*, *8*, 1080–1088. https://doi.org/10.4161/epi.25926

Sheffield, N. C., & Bock, C. (2016). LOLA: Enrichment analysis for genomic region sets and regulatory elements in R and bioconductor. *Bioinformatics*, *32*, 587–589. https://doi.org/10.1093/bioinformatics/btv612

Sun, D., Luo, M., Jeong, M., Rodriguez, B., Xia, Z., Hannah, R., … Goodell, M. A. (2014). Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell*, *14*, 673–688. https://doi.org/10.1016/j.stem.2014.03.002

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Weisenberger, D. J., Shen, H., … Widschwendter, M. (2010). Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Research*, *20*, 440–446. https://doi.org/10.1101/gr.103606.109

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., … Laird, P. W. (2007). Epigenetic stem cell signature in cancer. *Nature Genetics*, *39*, 157–158. https://doi.org/10.1038/ng1941

Yang, J., Huang, J., Dasgupta, M., Sears, N., Miyagi, M., Wang, B., … Stark, G. R. (2010). Reversible methylation of promoter-bound STAT3 by histone-modifying enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 21499–21504. https://doi.org/10.1073/pnas.1016147107

Yang, Z., Wong, A., Kuh, D., Paul, D. S., Rakyan, V. K., Leslie, R. D., … Teschendorff, A. E. (2016). Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biology*, *17*, 205. https://doi.org/10.1186/s13059-016-1064-3

Yuan, T., Jiao, Y., de Jong, S., Ophoff, R. A., Beck, S., & Teschendorff, A. E. (2015). An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. *PLoS Genetics*, *11*, e1004996. https://doi.org/10.1371/journal.pgen.1004996

Zheng, S. C., Beck, S., Jaffe, A. E., Koestler, D. C., Hansen, K. D., Houseman, A. E., … Teschendorff, A. E. (2017). Correcting for cell-type heterogeneity in epigenome-wide association studies: Revisiting previous analyses. *Nature Methods*, *14*, 216–217. https://doi.org/10.1038/nmeth.4187

Zykovich, A., Hubbard, A., Flynn, J. M., Tarnopolsky, M., Fraga, M. F., Kerksick, C., … Melov, S. (2014). Genome-wide DNA methylation changes with age in disease-free human skeletal muscle. *Aging Cell*, *13*, 360–366. https://doi.org/10.1111/acel.12180

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Pérez RF, Tejedor JR, Bayón GF, Fernández AF, Fraga MF. Distinct chromatin signatures of DNA hypomethylation in aging and cancer. *Aging Cell*. 2018;17:e12744. https://doi.org/10.1111/acel.12744

AGING AND CANCER: SECOND ARTICLE

<u>Pérez RF</u>, Tejedor JR, Santamarina-Ojeda P, Martínez VL, Urdinguio RG, Villamañán L, Candiota AP, Sarró NMV, Barradas M, Fernandez-Marcos PJ, Serrano M, Fernández AF & Fraga MF (2021) Conservation of Aging and Cancer Epigenetic Signatures across Human and Mouse. Mol Biol Evol 38, 3415–3435.

In this work, we expanded upon our previous observations of the similarities and differences in DNA methylation signatures of aging and cancer by studying their possible conservation between human and mouse. Two main reasons motivated this approach: 1) mouse models are one of the most important clinical tools used for the study of diseases; 2) it is important to understand if aging- and cancer-associated epigenetic patterns are conserved across species. To this end, we used reduced representation bisulfite sequencing to generate DNA methylation data for human and mouse brain tissue in the context of aging and cancer and performed a systematic and comparative analysis of the two processes across both species. Our results suggest that the specific DNA methylation characteristics of aging and cancer, and their differences, are conserved between human and mouse across multiple layers of genomic and epigenomic regulation. Moreover, integrative analyses showed how species-specific traits can often be attributed to sequence differences or functional variations. Finally, we developed a database of interspecies DNA methylation alterations which may aid in the identification of biomarkers with a better clinical translation.

Personal contribution to the work: I compiled the human samples from biobanks while other collaborators were responsible for generating the murine samples (labs of Dr. Manuel Serrano and Dr. Ana Paula Candiota). I generated, preprocessed and analyzed all of the data presented in this manuscript, as well as prepared the figure panels and wrote the manuscript. The work was carried out in close collaboration with co-first author Dr. Juan Ramón Tejedor, who was responsible for training the chromatin state models, and supervised by Dr. Agustín F. Fernández and Dr. Mario F. Fraga.

Conservation of Aging and Cancer Epigenetic Signatures across Human and Mouse

Raúl F. Pérez,[†,1,2,3,4,5] Juan Ramón Tejedor,[†,1,2,3,4,5] Pablo Santamarina-Ojeda,[1,2,3,4,5] Virginia López Martínez,[1,2,3,4,5] Rocío G. Urdinguio,[1,2,3,4,5] Lucía Villamañán,[6] Ana Paula Candiota,[6,7] Noemí Vidal Sarró,[8] Marta Barradas,[9] Pablo Jose Fernandez-Marcos,[9] Manuel Serrano,[10,11,12] Agustín F. Fernández,[*,1,2,3,4,5] and Mario F. Fraga[*,1,2,3,4,5]

[1]Cancer Epigenetics and Nanomedicine Laboratory, Nanomaterials and Nanotechnology Research Center (CINN-CSIC), University of Oviedo, Oviedo, Spain

[2]Institute of Oncology of Asturias (IUOPA), University of Oviedo, Oviedo, Spain

[3]Health Research Institute of Asturias (ISPA), University of Oviedo, Oviedo, Spain

[4]Department of Organisms and Systems Biology (B.O.S.), University of Oviedo, Oviedo, Spain

[5]Rare Diseases CIBER (CIBERER) of the Carlos III Health Institute (ISCIII), Madrid, Spain

[6]Unitat de Bioquímica de Biociències, Departament de Bioquímica i Biologia Molecular, Edifici Cs, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

[7]Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain

[8]Servicio Anatomía Patológica, Hospital Universitari de Bellvitge—IDIBELL, Hospitalet de Llobregat, Spain

[9]Metabolic Syndrome Group—BIOPROMET, Madrid Institute for Advanced Studies—IMDEA Food, CEI UAM+CSIC, Madrid, Spain

[10]Tumour Suppression Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

[11]Cellular Plasticity and Disease Group, Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain

[12]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

[†]These authors contributed equally to this work.

[*]**Corresponding authors:** E-mails: agustin.fernandez@cinn.es; mffraga@cinn.es.

Associate editor: Li Liu

Abstract

Aging and cancer are two interrelated processes, with aging being a major risk factor for the development of cancer. Parallel epigenetic alterations have been described for both, although differences, especially within the DNA hypomethylation scenario, have also been recently reported. Although many of these observations arise from the use of mouse models, there is a lack of systematic comparisons of human and mouse epigenetic patterns in the context of disease. However, such comparisons are significant as they allow to establish the extent to which some of the observed similarities or differences arise from pre-existing species-specific epigenetic traits. Here, we have used reduced representation bisulfite sequencing to profile the brain methylomes of young and old, tumoral and nontumoral brain samples from human and mouse. We first characterized the baseline epigenomic patterns of the species and subsequently focused on the DNA methylation alterations associated with cancer and aging. Next, we described the functional genomic and epigenomic context associated with the alterations, and finally, we integrated our data to study interspecies DNA methylation levels at orthologous CpG sites. Globally, we found considerable differences between the characteristics of DNA methylation alterations in cancer and aging in both species. Moreover, we describe robust evidence for the conservation of the specific cancer and aging epigenomic signatures in human and mouse. Our observations point toward the preservation of the functional consequences of these alterations at multiple levels of genomic regulation. Finally, our analyses reveal a role for the genomic context in explaining disease- and species-specific epigenetic traits.

***Key words*: epigenetics, DNA methylation, histone modification, mouse, human, conservation.**

Article

Introduction

Aging is one of the main risk factors associated with the development of cancer. These two processes have been extensively investigated from the epigenetic perspective and consequently analogous epigenetic alterations have been identified for both, providing explanations for the possible molecular links between them (Aunan et al. 2017; Yu et al. 2020). These alterations mainly relate to local gains in DNA

Pérez et al. · doi:10.1093/molbev/msab112 MBE

methylation at CpG-dense regions and global losses at the genomic scale involving repetitive DNA (Baylin and Jones 2016; Sen et al. 2016). Nonetheless, the extent of these parallelisms has not as yet been clearly outlined, particularly within the hypomethylation scenario (Dmitrijeva et al. 2018; Pérez et al. 2018). Moreover, recent whole-genome studies in mouse have also failed to confirm hallmarks such as global hypomethylation with aging in tissues including liver, hippocampus, and stem cells (Sun et al. 2014; Cole et al. 2017; Hahn et al. 2017; Masser et al. 2017; Hadad et al. 2019; Hernando-Herraez et al. 2019). These observations underscore two different, but equally important, issues: first, there is a need to study aging- and cancer-associated epigenetic patterns in a systematic manner to facilitate an integrated comparison between the two processes; second, it is essential to also confirm whether aging- and cancer-associated epigenetic alterations are equivalent in human and mouse models, as the latter are very important experimental proxies of human biology.

Mouse models have been widely used in the characterization of landmark epigenetic mechanisms such as transgenerational inheritance and genomic imprinting (Blewitt and Whitelaw 2013) as well as developmental epigenomic reprogramming (Hanna et al. 2018), and they are often used to characterize alterations in specific disease-associated pathways which sometimes have a translational correspondence in human (Espada and Esteller 2013). To date, reports suggest that DNA methylation patterns preserve global features such as the presence of nonmethylated islands and gene body methylation across mammalian species in general (Long et al. 2013; Schroeder et al. 2015), with other parallelisms extending across more distant vertebrate taxa (Elango and Yi 2008). More directed comparisons involving mouse and human have described a general conservation of their genome-wide epigenetic patterns (Edwards et al. 2010) and have particularly focused on the interspecies conservation of tissue-specific methylation patterns (Kessler et al. 2016; Zhou et al. 2017; Chen et al. 2018) and developmental processes (Lister et al. 2013). Nonetheless, considerable epigenetic differences in proximal species such as primates have been reported (Hernando-Herraez, Heyn, et al. 2015; Mendizabal et al. 2016). So far there have been only limited high-resolution studies that focus on directly assessing specific epigenetic similarities between human and mouse in either aging (Maegawa et al. 2017; Wang et al. 2017) or cancer (Maegawa et al. 2014), and those that exist are often limited by their compilation of data from different studies or the use of different profiling technologies for each species. Some of these issues have arisen from the lack of a large-scale DNA methylation array technology for mouse. Overall, then, there is a lack of back-to-back human–mouse comparisons of DNA methylation at single-base resolution in the context of disease. Even so, this issue is particularly relevant in the study of DNA methylation alterations found independently in both species in order to reliably ascertain the extent to which some of the observed similarities or differences arise from pre-existing species-specific epigenetic traits.

In this study, we sought to perform a systematic, integrated, and intercomparable analysis of the DNA methylation dynamics associated with cancer and aging in both human and mouse. To this end, we profiled the brain methylomes of young and old, tumoral and nontumoral brain samples from human and mouse by reduced representation bisulfite sequencing (RRBS), which provided a robust and interspecies-comparable platform for the analysis at hand (Bock et al. 2010). We first characterized the baseline epigenomic patterns of both species and subsequently focused on the DNA methylation alterations associated with cancer and aging. Next, we described the functional genomic and epigenomic context associated with the alterations, and finally, we integrated our data to study interspecies DNA methylation levels at orthologous CpG sites. Globally, we found considerable differences between the genomic and epigenomic characteristics of DNA methylation alterations in cancer and aging in human and mouse. Furthermore, we describe robust evidence for the conservation of the specific cancer- and aging-associated epigenomic patterns in both species. Our observations point toward the preservation of the functional consequences of these alterations at multiple levels of genomic regulation, including the locations and chromatin context of the alterations, the genetic pathways, and transcription factors (TFs) involved. Additionally, the analysis of orthologous CpG loci allowed us to describe species-common locations of DNA methylation variability and species-common cancer- and aging DNA methylation alterations. Finally, our analyses suggest that the genomic context associated with many CpG sites may help explain the observation of species-discordant DNA methylation as well as cancer and aging-specific DNA methylation alterations.

Results and Discussion

Baseline Epigenomic Patterns Are Conserved between Human and Mouse

To characterize the intricate, conserved epigenetic signatures that may exist in aging and cancer, we studied the genome-wide DNA methylation patterns of young, old, tumoral, and nontumoral brain samples from human and mouse ($n = 24$, 3 per group, fig. 1A). We produced methylation values for an average of 3.94 million and 1.62 million unique CpG sites per sample for human and mouse, respectively, and retained a final total of 1,171,918 and 585,234 CpG sites common to all samples after filtering for coverage (see Materials and Methods and supplementary table 1, Supplementary Material online, for sample and alignment information). We first sought to characterize the global aspects of DNA methylation and its relationship with genomic function in the two species in order to determine the baseline similarities and differences in DNA methylation patterns in our study system. For this initial analysis, we focused on the nontumoral samples because the tumors showed extensive deviation from the rest of the samples using pairwise Pearson correlations (fig. 1B), and our goal was to profile the general species-wise features of DNA methylation in our samples.

Though significant, we did not observe noticeable differences between the species in the distribution of the overall methylation values, with an average methylation of 0.266 for

Aging and Cancer Epigenetic Signatures across Human and Mouse · doi:10.1093/molbev/msab112 MBE

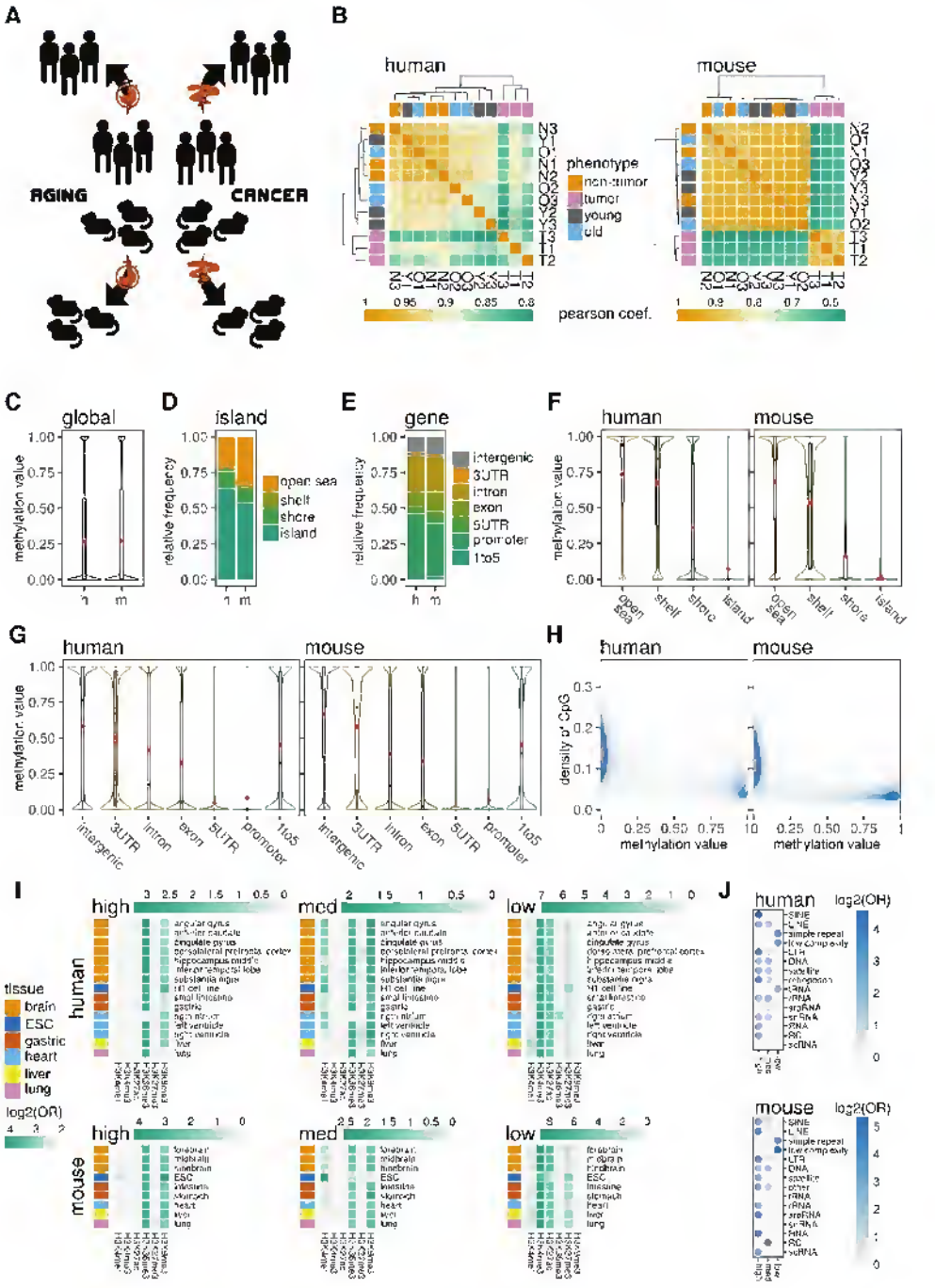


FIG. 1. Baseline epigenomic patterns in human and mouse. (*A*) Schematic of the study design. (*B*) Heatmaps showing the pairwise Pearson correlations between DNA methylation levels of the profiled CpG sites across samples. (*C*) Violin plots of global levels of DNA methylation for the pooled samples (red dots indicate average values; h, human; m, mouse; tumor samples are excluded). (*D* and *E*) Barplots indicating the distribution of the profiled CpG sites in human and mouse across CpG island and gene-region locations. (*F* and *G*) Violin plots showing the DNA methylation levels for the pooled samples with the CpG sites grouped by CpG island locations and gene-region locations (red dots indicate average values). (*H*) Density plots indicating the relationship between DNA methylation values and surrounding density (2-kb bins) in CpG sites for the profiled CpGs. (*I* and *J*) Heatmaps and bubble plots showing the significant overenrichments in log2(odds ratio) of CpG sites classified as high-, medium-, and low-methylation in ChIP-seq peak locations of different histone marks across a panel of tissues or in the locations of repetitive DNA classes across the human and mouse genomes.

Pérez et al. · doi:10.1093/molbev/msab112 MBE

human and 0.272 for mouse (fig. 1C, Wilcoxon rank-sum $P = 0.02$). The low levels of methylation observed are to be expected since the RRBS technology targets CpG-dense regions, such as CpG islands, which tend to be unmethylated. On a genome-wide scale, however, most CpG sites are methylated in mammals (Greenberg and Bourc'his 2019), and technologies such as whole-genome bisulfite sequencing (WGBS), which also targets intergenic regions, would detect higher levels of DNA methylation.

In order to ascertain that the RRBS was evaluating a comparable population of CpGs for each species, we annotated the sites to functional genomic locations in terms of their CpG island membership status and genetic elements. For both species, the technology interrogated a majority of CpG sites located at CpG islands (fig. 1D), with a slight increase in open sea CpGs for mouse (Fisher's $P < 0.001$, OR = 1.89). Regarding genetic elements, the distribution of the quantified CpGs was also very similar for both species (fig. 1E), with the largest proportion of sites being located at promoters and gene bodies. These observations suggest that, for human and mouse, the RRBS technology targets populations of CpG sites which have analogous genomic contexts.

Next, we proceeded to characterize the relationship between DNA methylation and CpG context in both human and mouse. With respect to CpG islands, DNA methylation was highest at open sea locations and lowest at islands for both species, with a transition in methylation across shelves and shores (fig. 1F). For gene elements, the relationship with DNA methylation was also parallel between human and mouse, with promoters and 5′-UTR regions showing the lowest values, followed by gene body exons and introns, whereas intergenic regions were those with the highest DNA methylation (fig. 1G). The presence of DNA methylation is inversely correlated to the density in surrounding CpG sites (Chen et al. 2018), and this may help explain some of the patterns found at genomic locations, especially concerning CpG islands and open sea. We therefore characterized the density in the CpG sites associated with each locus by counting the number of neighboring CpG sites in surrounding 2-kb windows. As expected, in general, low methylation values were found associated with locations with higher density in CpG sites, whereas higher methylation values were found at low-density locations, for both species (fig. 1H). Indeed, the different genomic locations (islands and gene regions) displayed different distributions of density in CpG sites (supplementary fig. S1A, Supplementary Material online) and, moreover, the mean methylation value of each element correlated almost perfectly with its average density in CpG sites (supplementary fig. S1B, Supplementary Material online). Collectively, these results confirm that the relationship between DNA methylation and functional genomic elements such as CpG islands and gene regions is conserved between human and mouse and is mostly explained by the density of CpG sites associated with each element (Edwards et al. 2010).

In the regulation of gene expression, epigenetic marks such as DNA methylation are not independent, rather, they act in combination with interrelated marks such as histone modifications to define functional states (Du et al. 2015). Thus, to better study the relationship of DNA methylation concerning its genomic function in mouse and human, we integrated our data with ENCODE and NIH Roadmap Epigenomics ChIP-seq data profiling the histone modifications H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K27me3, and H3K9me3. We focused on a panel of tissues which was available for the two species, and contained several brain tissue tracks, analyzing a combined total of 144 data sets. We classified CpG sites as high-, medium-, or low-methylation loci (>0.8, [0.2,0.8], and <0.2 average methylation values, respectively) and investigated the colocalization of these sites with the histone marks by overenrichment analyses at FDR <0.05 (see Materials and Methods). With this procedure, we observed practically identical chromatin signatures of methylation status in human and mouse (fig. 1I and supplementary table 2, Supplementary Material online): high methylation sites were associated with active gene-body H3K36me3 and repressive H3K9me3 marks, intermediate methylation also appeared to be related to enhancer/promoter-associated H3Kme1 and low methylation sites were strongly associated with the active promoter and enhancer H3K4me3 and H3K27ac marks. These observations confirm that the relationship between DNA methylation and genomic function is conserved in both species (Roadmap Epigenomics Consortium et al. 2015; Gorkin et al. 2020).

One of the genomic elements to which DNA methylation has been most functionally linked is repetitive DNA (Beisel and Paro 2011). Hence, we mapped the classified CpG sites to RepeatMasker repetitive elements and looked for overenrichment in specific repeat classes. Looking at the most common repeats, we observed, for both human and mouse, that high-methylation loci were associated with SINE, LINE, LTR, and DNA repeats, whereas low-methylation was linked to simple and low-complexity repeats (fig. 1J and supplementary table 3, Supplementary Material online). Expanding the analyses to repetitive element families also revealed parallel patterns in the two species (supplementary fig. S2, Supplementary Material online). These results are also in consonance with existing literature both for mouse and human (Edwards et al. 2010).

Dynamics of DNA Methylation Alterations in Human Aging and Cancer

Once we had confirmed the analogous human and mouse baseline epigenomic patterns in our study system, we proceeded to determine the DNA methylation alterations associated with aging and with cancer for each species independently. Starting with human, we found no general DNA methylation differences associated with cancer or aging by looking at all the sites analyzed by the RRBS, although tumor samples tended to have higher methylation values (fig. 2A). Moreover, PCA clustering suggested differences between the experimental groups as they segregated across the most meaningful dimensions (fig. 2B). We performed a differential methylation analysis with a stringent threshold (FDR <0.05, average change in methylation >0.2) to define differentially methylated CpGs (dmCpGs) in aging and cancer. We identified widespread alterations in both processes, counting

Aging and Cancer Epigenetic Signatures across Human and Mouse · doi:10.1093/molbev/msab112 MBE

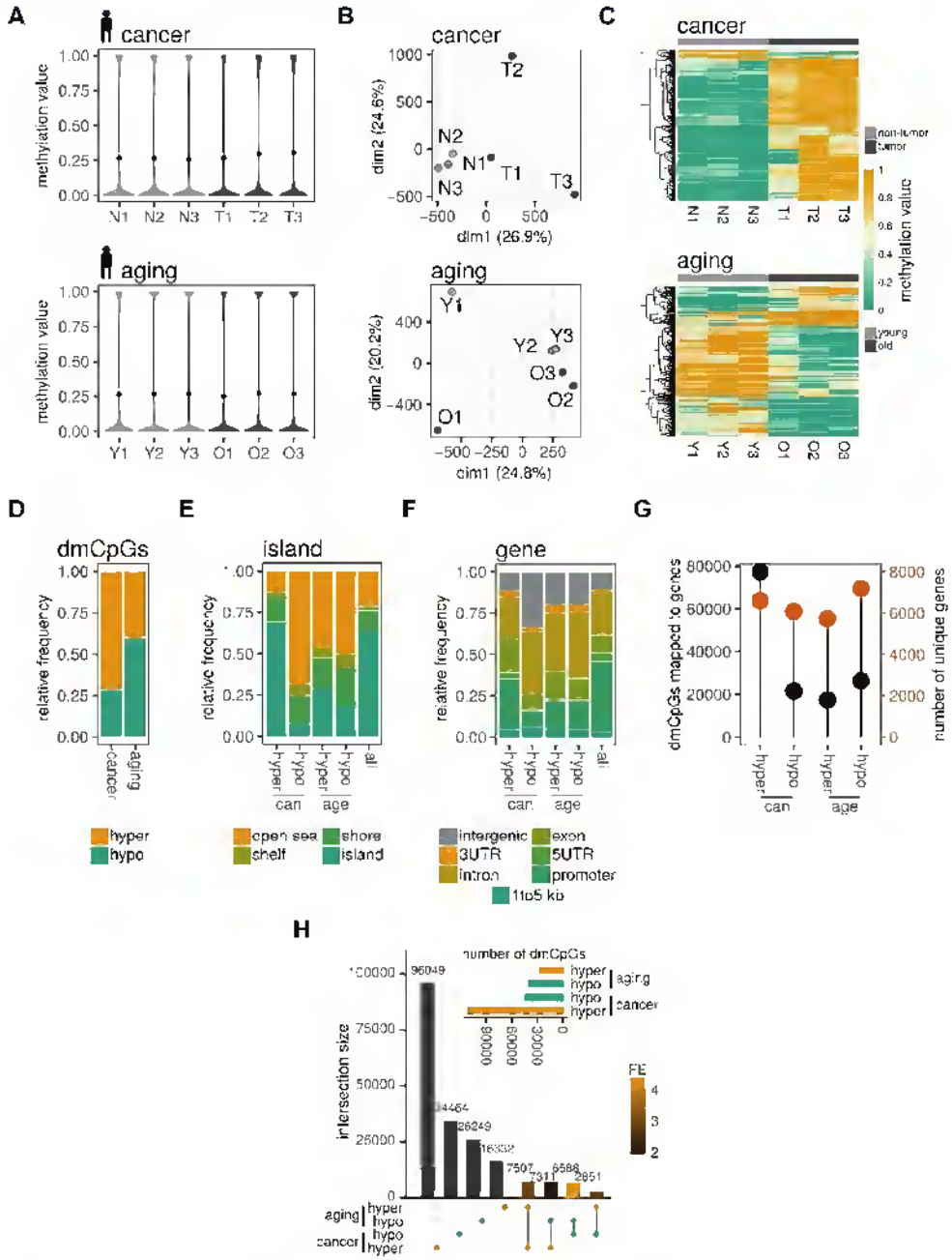


FIG. 2. Cancer and aging DNA methylation alterations in human. (*A*) Violin plots of global levels of DNA methylation of all CpGs across the profiled samples (N, nontumor; T, tumor; Y, young; O, old). (*B*) PCA plots showing the distribution of the samples across the first two principal components, with the percentage of variance each explains. (*C*) Heatmaps showing the methylation values of the top 1,000 dmCpGs found in the aging and cancer comparisons. (*D*) Barplots indicating the relative numbers of hyper- and hypo-dmCpGs (differentially methylated CpGs) in aging and cancer. (*E* and *F*) Barplots indicating the distribution of the dmCpGs across CpG island and gene-region locations, including the distribution of the universe of CpGs profiled by the RRBS ("all"). (*G*) Lollipop plots comparing the number of dmCpGs mapped to genes (left vertical axis, black dots) with the number of different genes that the dmCpGs mapped to (right vertical axis, orange dots). (*H*) UpSet plot describing the dmCpG sets and their intersections, indicating the size of the sets and the fold-enrichment (FE) of the intersections based on their expected overlaps.

154,770 dmCpGs associated with cancer and 66,838 with aging (lists of dmCpGs are available as extended data sets), with hypermethylation dominating the cancer scenario particularly (accounting for 72% of changes, fig. 2*C* and *D*). Furthermore, performing the cancer differential methylation comparison using all nine nontumoral samples (nontumor, young, and old groups) as the control group yielded comparable results and numbers of dmCpGs (supplementary fig. S3A, Supplementary Material online), indicating that the cancer-associated alterations observed are robust and are

Pérez et al. · doi:10.1093/molbev/msab112 MBE

independent of the samples used. When mapping the dmCpGs to CpG island locations (fig. 2E), we observed that hypermethylation occurred for cancer preferentially at CpG islands, with a distribution akin to that of the RRBS background of analyzed sites, whereas hypomethylation and aging-associated changes were, in general, enriched at open sea locations (Fisher's $P < 0.001$, ORs = 8.9, 3.3, and 4.0, respectively). Regarding gene locations, cancer hypomethylation and aging hyper- and hypomethylation were particularly increased at intergenic- and intronic sequences (Fisher's $P < 0.001$, intergenic ORs = 4.5, 2.0, and 2.0, intronic ORs = 1.8, 1.7, and 2.0). In line with this observation, cancer hypermethylation occurred more frequently at CpG-denser regions than the rest of the changes (supplementary fig. S3B, Supplementary Material online) and, additionally, it consisted of the strongest changes (supplementary fig. S3C, Supplementary Material online). These results suggest that there are important differences in the breadth and distribution of hyper- and hypomethylation alterations in cancer, whereas aging-associated alterations appear to be located at similar genomic locations regardless of the direction of change, these locations being comparable to those targeted by cancer-associated hypomethylation.

To expand on the possible functional consequences of the DNA methylation alterations observed, we mapped the dmCpGs to genes and performed pathway enrichment analyses. First, as shown in figure 2F, the largest proportion of dmCpGs mapping to genes was found for cancer hypermethylation. However, of the CpGs that mapped to genes, cancer hypo-dmCpGs and aging dmCpGs, although less numerous, were distributed across similar numbers of genes as cancer hyper-dmCpGs (fig. 2G, χ^2 $P < 0.001$), indicating that the former were more spread-out across different genes. We performed pathway enrichment analyses against several databases which included Gene Ontology, KEGG and Reactome pathways, and the CGP database from MSigDB, which describes empirical pathways mined from biomedical literature. Only cancer hyper-dmCpGs revealed significant enrichment in biological functions, and these were related to Polycomb-target genes, G-protein coupled receptor (GPCR) pathways—with neuropeptide receptors in particular—and cellular development pathways, especially neuronal (supplementary fig. S4, Supplementary Material online, full results are available as extended data sets). Thus, cancer-associated DNA methylation deregulation, particularly hypermethylation, appears to have a more coherent functional impact than aging-associated changes.

In order to investigate common axes of DNA methylation deregulation in both cancer and aging, we intersected the dmCpGs found, observing common CpG sets across all processes (fig. 2H and supplementary table 4, Supplementary Material online). These intersections were all significantly overenriched (SuperExactTest all $P < 0.001$), with common cancer- and aging hypo-dmCpGs displaying the highest fold enrichment, whereas common cancer hyper- and aging hypo-dmCpGs showed the lowest. These findings indicate that, regardless of their direction of change, DNA methylation alterations are readily found at disease-common loci. We performed pathway enrichments on these intersections and only observed enrichments in common CpGs between cancer hypermethylation and aging hyper- and hypomethylation (full results are available as extended data sets). These results suggest that the functional commonalities between DNA methylation deregulation in human cancer and aging mainly pertain to Polycomb-target gene hypermethylation in cancer, whereas these pathways manifest both hyper- and hypomethylation in aging.

Dynamics of DNA Methylation Alterations in Mouse Aging and Cancer

Next, we profiled the DNA methylation alterations associated with aging and cancer in mouse. First, regarding the general levels of DNA methylation, we found, in line with the previous observations in human, an even clearer trend of increased DNA methylation in cancer (fig. 3A). The PCA hinted at some differences between the phenotypes in that a slight separation was observed between the groups when looking at the top two dimensions (fig. 3B). The differential methylation analysis (FDR < 0.05, average change in methylation >0.2) uncovered extensive alterations, with 181,027 dmCpGs detected in cancer and 22,564 in aging (lists of dmCpGs are available as extended data sets). When all of the nine nontumoral samples (nontumor, young, and old groups) were used as the control group in the cancer differential methylation comparison, results and numbers of dmCpGs were similar (supplementary fig. S5A, Supplementary Material online), suggesting that the cancer-associated alterations observed are readily detected irrespective of the samples used. Regarding the direction of the alterations, hypermethylation was the predominant phenomenon in both processes (accounting for 68% and 62% of changes, respectively, fig. 3C and D). The genomic-location distributions of the dmCpGs were analogous to those observed in the human system (fig. 3E): cancer hypermethylation occurred at CpG island-associated sites whereas cancer hypomethylation and aging hyper- and hypomethylation changes were much more enriched at open sea locations (Fisher's $P < 0.001$, ORs = 24.7, 6.7, and 10.3). Similarly, with respect to gene locations, these latter alterations were once more enriched at intergenic and intronic sequences (fig. 3F, Fisher's $P < 0.001$, intergenic ORs = 7.5, 2.4, and 3.0, intronic ORs = 1.6, 2.3, and 2.1, respectively). Again, cancer hypermethylation was directly associated with higher CpG-density (supplementary fig. S5B, Supplementary Material online) and accounted for the largest alterations (supplementary fig. S5C, Supplementary Material online). In general, the patterns of disease-associated DNA methylation changes, and especially the contrast between cancer and aging alterations, followed the same trends as those seen for human tissue.

To examine if the parallelisms observed extended to the functional pathways involved, we mapped the dmCpGs to genes and carried out pathway enrichment analyses. Once more, within the dmCpGs mapping to genes, despite cancer hyper-dmCpGs being by far the most numerous, they were concentrated across fewer genes than expected when compared with hypo-dmCpGs and aging changes (fig. 3G, χ^2

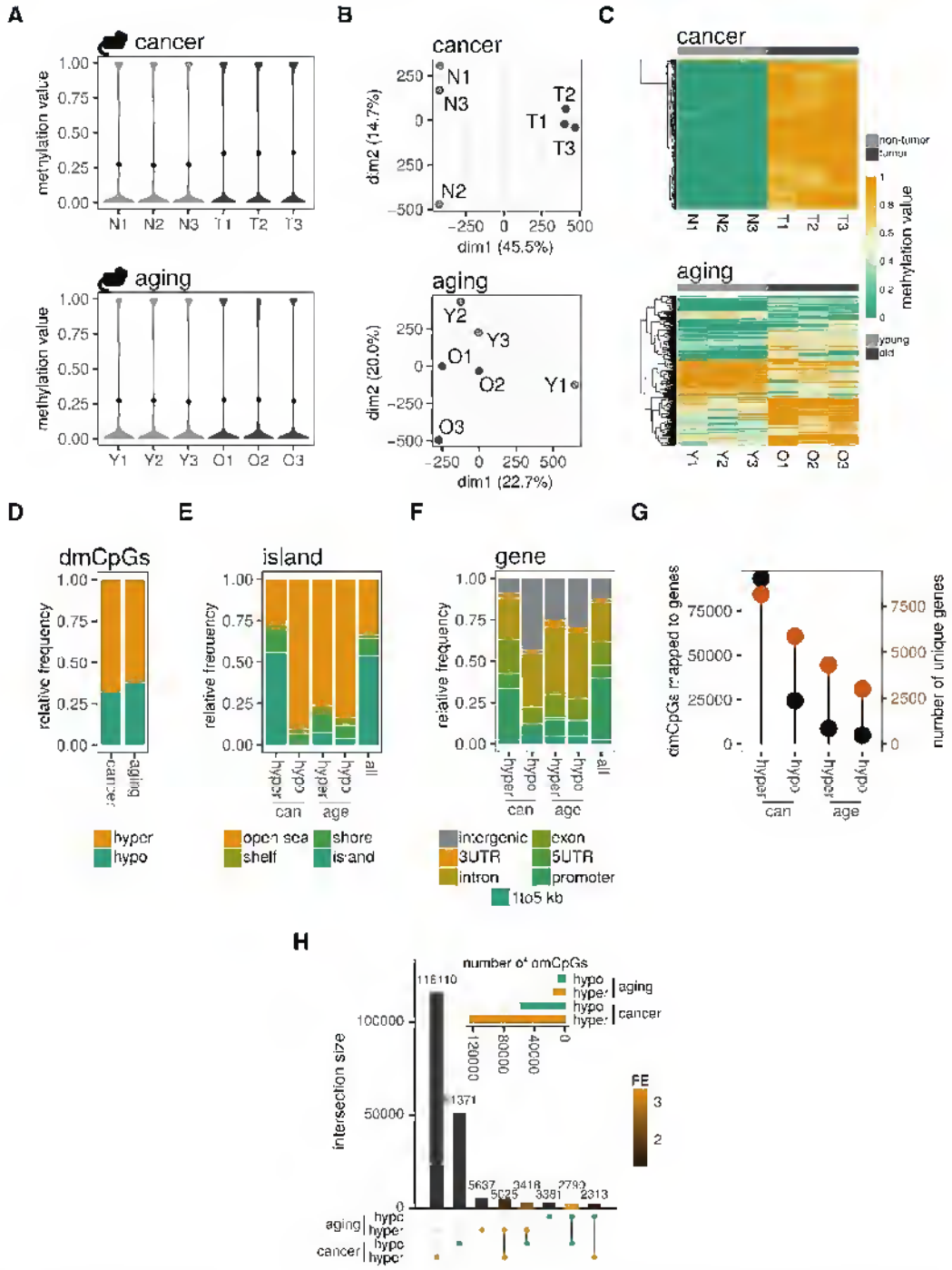


FIG. 3. Cancer and aging DNA methylation alterations in mouse. (*A*) Violin plots of global levels of DNA methylation of all CpGs across the profiled samples (N, nontumor; T, tumor; Y, young; O, old). (*B*) PCA plots showing the distribution of the samples across the first two principal components, with the percentage of variance each explains. (*C*) Heatmaps showing the methylation values of the top 1,000 dmCpGs found in the aging and cancer comparisons. (*D*) Barplots indicating the relative numbers of hyper- and hypo-dmCpGs (differentially methylated CpGs) in aging and cancer. (*E* and *F*) Barplots indicating the distribution of the dmCpGs across CpG island and gene-region locations, including the distribution of the universe of CpGs profiled by the RRBS ("all"). (*G*) Lollipop plots comparing the number of dmCpGs mapped to genes (left axis, black dots) with the number of different genes that the dmCpGs mapped to (right axis, orange dots). (*H*) UpSet plot describing the dmCpG sets and their intersections, indicating the size of the sets and the fold-enrichment (FE) of the intersections based on their expected overlaps.

$P < 0.001$). In this case, the pathway enrichment analyses revealed significant results for all sets of dmCpGs (supplementary fig. S6, Supplementary Material online, full results are available as extended data sets). Cancer hyper-dmCpGs, as with human, were enriched in Polycomb-target genes, GPCRs (Wnt-activated in particular), and cellular development

pathways, including neuronal ones. On the other hand, hypo-dmCpGs were enriched in generic cancer pathways, cell-signaling pathways (including GPCRs), and inflammation terms. Aging hypermethylation was enriched in cell-interaction terms, and hypomethylation in cell-signaling pathways. These results indicate that, in contrast to human processes, aging-associated DNA methylation alterations in mouse may have more directed functional consequences at the gene level.

We integrated the cancer- and aging-associated dmCpGs and found significant overenrichment for all the common sets (fig. 3*H* and supplementary table 4, Supplementary Material online, SuperExactTest all $P < 0.001$). As occurred with human tissue, whereas all the intersections displayed overenrichment, common cancer-, and aging hypo-dmCpGs displayed the highest fold enrichment and cancer hyper- along with aging hypo-dmCpGs the lowest. Next, we performed pathway enrichment analyses of the intersections. This time, we observed no significant enrichments for those containing cancer hyper-dmCpGs, but modest enrichments for inflammation terms did appear, especially for cancer hypo- and aging hyper-dmCpGs (full results are available as extended data sets). Taken together, these results suggest that, in the case of mouse and, in contrast to the human results, the Polycomb-target, developmental gene hypermethylation observed in cancer is not as evident in aging.

The Interspecies Genomic Context of Cancer and Aging DNA Methylation Alterations

In light of the parallelisms observed, we sought to extend the analyses to additional levels of epigenomic regulation by integrating the methylomic data with the previously mentioned histone modification data—involving the histone modifications H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K27me3, and H3K9me3—across a panel of common healthy tissues in human and mouse (see Materials and Methods). We performed overenrichment analyses (FDR < 0.05) to test the association of the cancer- and aging-related dmCpGs discovered in the two species with the different histone modifications, in order to define chromatin signatures linked to the DNA methylation changes. Using this procedure, we were able to find robust and well-defined similarities between the two species (fig. 4*A* and supplementary table 5, Supplementary Material online). As regards cancer-associated changes, we observed an association of DNA hypermethylation with the repressive H3K27me3 modification and the more active enhancer/promoter-associated H3K4me1. In the case of human, cancer hypermethylation was also strongly enriched at H3K9me3 locations, which are linked to heterochromatic and repressive states. On the other hand, cancer hypomethylated dmCpGs were mainly enriched in the H3K9me3 mark for both species. These cancer-associated signatures have been described across multiple tissues for human (Pérez et al. 2018) whereas the literature for mouse is scant.

When looking at aging-related signatures, we again found comparable patterns between human and mouse. Moreover, for both species, these patterns were markedly different to those found for cancer. We observed enrichments at locations marked with gene-body H3K36me3, heterochromatin H3K9me3, and enhancer/promoter H3K4me1. In this case, even though we observed an increase in H3K4m1 association at aging hypomethylated loci when compared with aging hypermethylation—a finding reported for human in several tissues (Fernández et al. 2015; Pérez et al. 2018)—the global patterns of enrichment found for aging dmCpGs differed moderately from those described in the literature, a result which could have been influenced by the fact that the "classical" chromatin signatures have been mostly derived by using array technology, which targets similar but not identical genomic contexts (supplementary fig. S7, Supplementary Material online). In further support of this idea, a recent study comparing aging-associated DNA methylation changes in liver between human and mouse—using array and RRBS data for each species respectively—reported a high enrichment in H3K36me3-associated loci specifically for mouse aging changes as compared with human (Wang et al. 2017).

Because the histone modification results exposed epigenomic similarities between human and mouse for the methylation changes associated with cancer and aging, we extended our analyses to discover and study chromatin states, which are defined by spatial combinations of the histone marks and reflect functional genomic states (Ernst and Kellis 2010). We trained 18-state models using the six core histone modifications for our selected panel of data sets by using multivariate Hidden Markov Models (see Materials and Methods). The nature of the human states defined was similar to that previously described (Roadmap Epigenomics Consortium et al. 2015) in terms of their histone mark composition and co-occurrence with other genomic elements (supplementary fig. S8*A*, Supplementary Material online). The states found for mouse were, in turn, notably similar to human (supplementary fig. S8*B*, Supplementary Material online), with the exception of some specific ones, such as the one assigned as "EnhG2" (genic enhancer 2). These observations suggest that multilevel epigenomic regulation is functionally parallel between the two species.

We evaluated the overenrichment of the dmCpGs in the different chromatin states (fig. 4*B* and supplementary table 6, Supplementary Material online). Cancer hypermethylation was strongly enriched in bivalent transcription start sites and Polycomb-repressed sites (states 14, 16), and somewhat enriched in transcription start sites and enhancer sites in general (states 2–4, 7–11) for both species. Additionally, and as suggested by the stronger enrichment in H3K9me3 for human as compared with mouse (fig. 4*A*), the chromatin-state analysis revealed that cancer hypermethylation also occurred, specifically for human, both at heterochromatic regions and at zinc-finger (ZNF) genes/repeats (states 12, 13). Indeed, ZNF genes, which are characterized by the high presence of H3K9me3 (Blahnik et al. 2011), displayed the highest overenrichment of all the chromatin states in human cancer hypermethylation. Interestingly, this family shows a great evolutionary expansion in human as compared with other species, including mouse, and its members are posited to play roles in novel transcriptional repression activity

Aging and Cancer Epigenetic Signatures across Human and Mouse · doi:10.1093/molbev/msab112 MBE

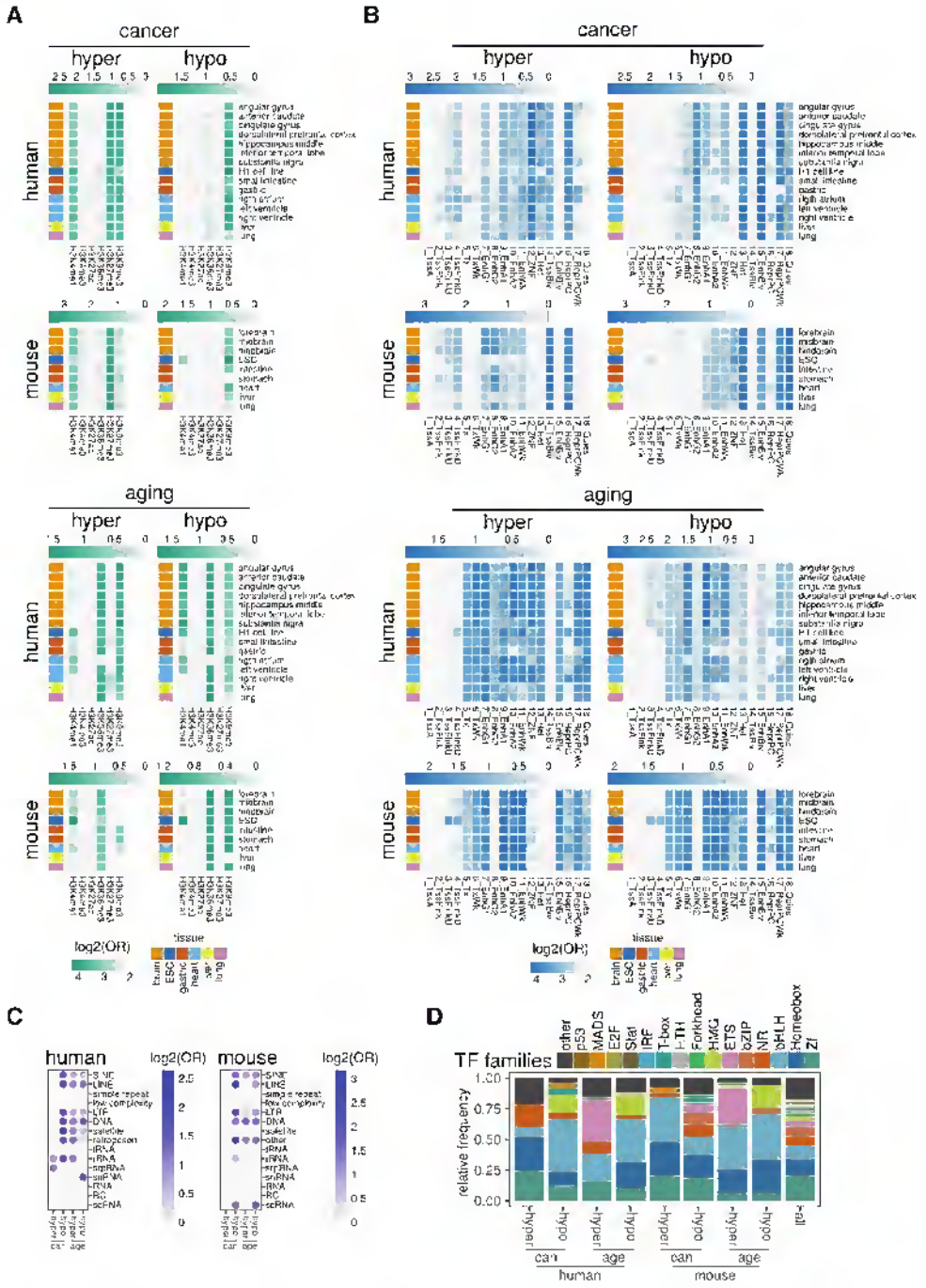


FIG. 4. Chromatin landscape of aging and cancer DNA methylation alterations in human and mouse. (*A* and *B*) Heatmaps showing the significant overenrichments in log2(odds ratio) of hyper- and hypo-dmCpGs in the cancer and aging comparisons, for human and mouse in ChIP-seq peak locations of different histone marks across a panel of tissues or in chromatin states built across the same panel of tissues. (*C*) Bubble plots showing the significant overenrichments in log2(odds ratio) of hyper- and hypo-dmCpGs in the cancer and aging comparisons in the locations of repetitive DNA classes across the genome for human and mouse. (*D*) Barplots indicating the relative numbers of transcription factors, classified by their domains, whose motifs were found to be significantly enriched (HOMER analysis, FDR <0.05) at the genomic locations of the different dmCpG sets, compared with the numbers of all of the transcription factors tested ("all"). Only the top 50 most enriched, significant factors for each set were selected.

(Emerson and Thomas 2009). Thus, the DNA methylation deregulation observed in human cancer could be related to the targeting of human-specific pathways. Regarding cancer hypomethylation, for both species, most of the affected chromatin states (13, 15, 17, 18) involved repressive functions, such as heterochromatin or Polycomb-associated sites, with

Pérez et al. · doi:10.1093/molbev/msab112 MBE

some enrichment also being observed in various enhancer associated states. In the case of aging-associated dmCpGs, and much like with the histone signatures, the enrichments were spread out across similar states for hyper- and hypomethylation, and involved gene-body transcription states (5, 6), enhancer states (7–11), and repressive states (15–18). Especially in the case of human, a particular increase in enrichment in enhancer states was observed for aging hypomethylation. The human and mouse profiles were globally similar, with some differences, such as in EnhG2 (state 8) probably being explained by the pre-existing differences in the chromatin models built (supplementary fig. S8, Supplementary Material online). Collectively, our results show that the chromatin context of DNA methylation changes in cancer and aging is considerably different, and that the specific features of each process are robustly recapitulated in human and mouse.

We also looked at the association between DNA methylation alterations and repetitive DNA elements (fig. 4C and supplementary table 7, Supplementary Material online; see supplementary fig. S9, Supplementary Material online, for the analysis expanded to repetitive DNA classes). First, regarding cancer, repetitive loci were mainly associated with hypomethylation across the same large classes of repeats (SINE, LINE, LTR, DNA), as has been well-described previously (Ross et al. 2010), with the main species difference being that satellite repeats showed no changes in DNA methylation for mouse (fig. 4C). In the case of SINE repeats, both the primate-specific Alu family in human and the related B elements (B1, B2, B4) in mouse were targeted by DNA hypomethylation (supplementary fig. S9, Supplementary Material online). Interestingly, we also observed in human the hypermethylation of ribosome-associated elements (rRNA, srpRNA), a finding which has been reported for some types of cancers (Srivastava et al. 2016), although literature on the subject is scarce. Secondly, in the case of aging, both hyper- and hypomethylation were associated with repetitive DNA, with the patterns observed being similar to those of cancer hypomethylation. Although RRBS is mostly limited to examining single-copy sequences, the bidirectionality of the changes found here for aging may help explain observations in the literature that fail to describe any noticeable trend of loss of methylation during aging in the genome-wide levels of brain methylation both in mouse (Hadad et al. 2019) and human (Lister et al. 2013; McKinney et al. 2019). No associations whatsoever were found with simple or low-complexity repeats, which were the main classes which we had previously linked to low levels of basal methylation in nontumoral tissue (fig. 1J). Taken together, these observations suggest that repetitive DNA elements are subject to similar DNA methylation alterations in cancer and aging for both species, with the exception of satellite repeats, which appear to be more affected by these processes in human as compared with mouse.

We next sought to study the possible effects of the DNA methylation changes found at TF-binding sites, because TFs are downstream genomic regulators directly in contact with the genomic structure. To this end, we performed a HOMER analysis to find enrichment in TF motifs associated with dmCpG sites, selecting the top 50 significant (FDR < 0.05) motifs with the highest enrichments found for each comparison (see supplementary table 8, Supplementary Material online, for extended results). First, to obtain a general overview of the types of TFs affected by the DNA methylation changes, we classified our results into TF domain-families (fig. 4D). With this approach, we found both similarities and differences between the processes as well as between species. Cancer hypermethylation dmCpGs were associated with Homeobox TFs for both species (Fisher's $P < 0.01$, ORs $= 2.7$ and 2.7 for human and mouse, respectively), but also associated with NRs in human (Fisher's $P < 0.05$, OR $= 2.8$) and bHLH-domain TFs (Fisher's $P < 0.001$, OR $= 7.8$) in mouse. Regarding cancer hypomethylation, it was particularly enriched at bHLH- and HMG-domain TF motifs in human (Fisher's $P < 0.001$, ORs $= 11.4$ and 8.8, respectively). On the other hand, the patterns of TF sites targeted by aging hyper- and hypomethylation were similar in both species. Aging hypermethylation particularly affected ETS-domain TF motifs (Fisher's $P < 0.001$, ORs $= 48.5$ and 26.9 for human and mouse, respectively) whereas hypomethylation involved bHLH- and HMG-domain TF sites (Fisher's $P < 0.001$, ORs $= 6.9$, 12.0, 7.8, and 16.6 for human bHLH, HMG and mouse bHLH, HMG, respectively).

Subsequently, we focused on the specific TFs involved in the previous observations, by looking at the main sets of shared or unique TFs affected in each comparison (supplementary fig. S10, Supplementary Material online). As suggested by the previous results, the patterns of species-common alterations at TF motifs were especially noticeable for aging-related changes. Aging-associated hypermethylation of ETS-domain TFs, a family strongly linked to oncogenic processes (Sizemore et al. 2017), was the main common epigenetic phenomenon between mouse and human. Several Sox TFs (HMG domains) were commonly targeted by aging hypomethylation in both species, such as Sox2, whose expression has been shown to be deregulated with aging in both human and mouse brain (Carrasco-Garcia 2018). Interestingly, Sox TFs also appeared affected by cancer hypomethylation in human, suggesting possible links between the regulation of aging and of cancer, as has been shown for Sox4 in other tumors (Foronda et al. 2014). In addition, the deregulation of methylation patterns at bHLH-domain TF sites appeared to be a general trend in both cancer and aging processes, appearing at loci related to specific neuronal development and differentiation, including TFs such as NeuroG2, NeuroD1, Olig2, or Atoh1 (Dennis et al. 2019) and also involving Homeobox TFs such as DLX1/2.

To sum up, the analysis of the links between DNA methylation alterations and TF-binding sites revealed that clear functional parallelisms between human and mouse exist in the aging scenario. This is in contrast to our earlier gene pathway enrichment analyses, which showed that most of the affected genetic pathways common in human and mouse involved cancer hypermethylation (of Polycomb-targets and developmental genes). This discrepancy could be explained by aging-associated methylation changes being more functionally important when occurring at genomic sites not

directly associated with gene locations (such as many TF-binding sites). Moreover, as shown in earlier stages of this study (figs. 2*F* and 3*F*), aging changes concentrate more at nongenic sites than cancer changes. Thus, the identification of functional targets implicated in DNA methylation alterations in aging may be more effective when focusing on TF-binding site alterations.

Interspecies Integration of Cancer and Aging DNA Methylation Alterations

After establishing the parallel features of DNA methylation alterations in human and mouse, we integrated our CpG-specific data, to be able to directly compare methylation values between the two species at orthologous locations, by lifting the mouse coordinates to the human genome (see Materials and Methods). We recovered interspecies measurements for a sizeable set of 59,100 CpG sites, 35% of which had been classified as dmCpGs at any comparison (see supplementary table 9, Supplementary Material online, for the liftover statistics; the full list of orthologous sites is available as an extended data set). These orthologous sites were overenriched in CpG island locations when compared with the original backgrounds (Fisher's $P < 0.001$, ORs = 7.1 and 6.1 for human and mouse, respectively), and largely maintained their relationship to CpG islands in both species, suggesting that they covered similar functional regions in both genomes (fig. 5*A*). Correlation analysis of the DNA methylation values of the interspecies lifted CpGs revealed that, excluding cancer, the samples clustered within their species (fig. 5*B*). Tumors, on the other hand, again showed widespread reconfiguration, clustering independently and, interestingly, mouse tumor CpGs were slightly more correlated to human tumors than to the rest of the mouse samples. Examining only nontumoral samples, methylation was found to be highly correlated between the two species (fig. 5*C*, left panel; Pearson coef. 0.84, $P < 0.001$). Moreover, when we retained only CpG sites which were lowly variable within each species (SD < 0.1 methylation value) most of the species-discordant CpG sites disappeared (fig. 5*C*, right panel; Pearson coef. 0.92, $P < 0.001$). Indeed, the subsets of high-variability sites for human and mouse (CpGs with SD $\geq$0.1 methylation value) showed much higher overlap than expected (fig. 5*D*; Fisher's $P < 0.001$, OR = 16.7), suggesting that locations of methylation variability are conserved between the two species. In order to strictly define species-discordant CpGs (sdCpGs), we performed a differential methylation analysis (FDR < 0.05, average change in methylation >0.2) at the 59,100 loci with interspecies measurements (the list of sites and sdCpGs is available as an extended data set). We identified 1,845 sdCpGs (supplementary fig. S11*A*, Supplementary Material online), evenly split between those presenting higher and those lower methylation in mouse as compared with human (hyper- and hypo-sdCpGs, respectively, supplementary fig. S11*B*, Supplementary Material online). Interestingly, when annotating the sdCpGs either to human or mouse genomic contexts, we observed considerable differences in the distribution of the sites, especially according to their CpG island status (supplementary fig. S11*C* and *D*, Supplementary Material online). Following this observation, we directly assessed the transitions between the genomic contexts of the sdCpGs in the genome of each species (fig. 5*E*), finding that, particularly in the case of hyper-sdCpGs, there was a considerable difference in the genomic association to CpG island status between the two species. In this case, CpGs which in human-occupied CpG islands and in mouse-occupied less-dense regions (shores, open sea) had higher methylation levels in mouse (i.e., were classified as hyper-sdCpGs). This is in line with our previous observations (fig. 1*F*–*H*) of denser genomic loci being associated with lower methylation values in both species. This finding suggests that interspecies differences in DNA methylation levels may be partly explained by local changes in the functional genomic context of their respective genomes, such as CpG island membership (Hernando-Herraez, Garcia-Perez, et al. 2015).

Next, we focused on the fraction of previously described species-specific cancer- or aging-associated dmCpGs for which we had interspecies measurements (see supplementary table 9, Supplementary Material online, and extended data sets). This allowed us to directly compare cancer- and aging changes at specific loci between human and mouse. We found common and statistically significant intersections between the two species for all sets (fig. 5*F* and supplementary table 10, Supplementary Material online; SuperExactTest all $P < 0.001$), with the biggest intersection being found for cancer dmCpGs, although the observed size and fold enrichment of the intersections was probably influenced by the differing sizes of the various sets. Differentiating by direction of change revealed that the biggest intersections were dominated by hypermethylation changes in both species (supplementary fig. S12, Supplementary Material online), especially when involving cancer sets. Aging, on the other hand, was limited to much smaller intersections with directionality not playing such an apparent role. Nonetheless, the observed intersections suggest that, added to the previously described similarities in genomic contexts (fig. 4), when we focus on specific orthologous loci, common sites of DNA methylation alteration across the two diseases and the two species are readily detected.

After the characterization of specific similarities between human and mouse, we turned to examine the influence of sequence conservation in the explanation of the epigenetic patterns found. We integrated our data with UCSC conservation scores for multiple alignments across vertebrate species (see Materials and Methods). First, regarding disease-associated alterations in general (dmCpGs), we observed that, for both species, cancer hypermethylation occurred at more conserved regions than did either cancer hypomethylation or aging changes (fig. 5*G*). This finding can support the notion that cancer hypermethylation may impact conserved axes of gene regulation across species, something which we had observed through the pathway enrichment analyses. However, these trends may also be partly explained by the genomic distribution of the alterations, because hypomethylation and aging alterations occur more often at intergenic and intronic regions (see figs. 2*F* and 3*F*), which we found (using all the profiled CpGs) to be less conserved than other

Pérez et al. · doi:10.1093/molbev/msab112 MBE

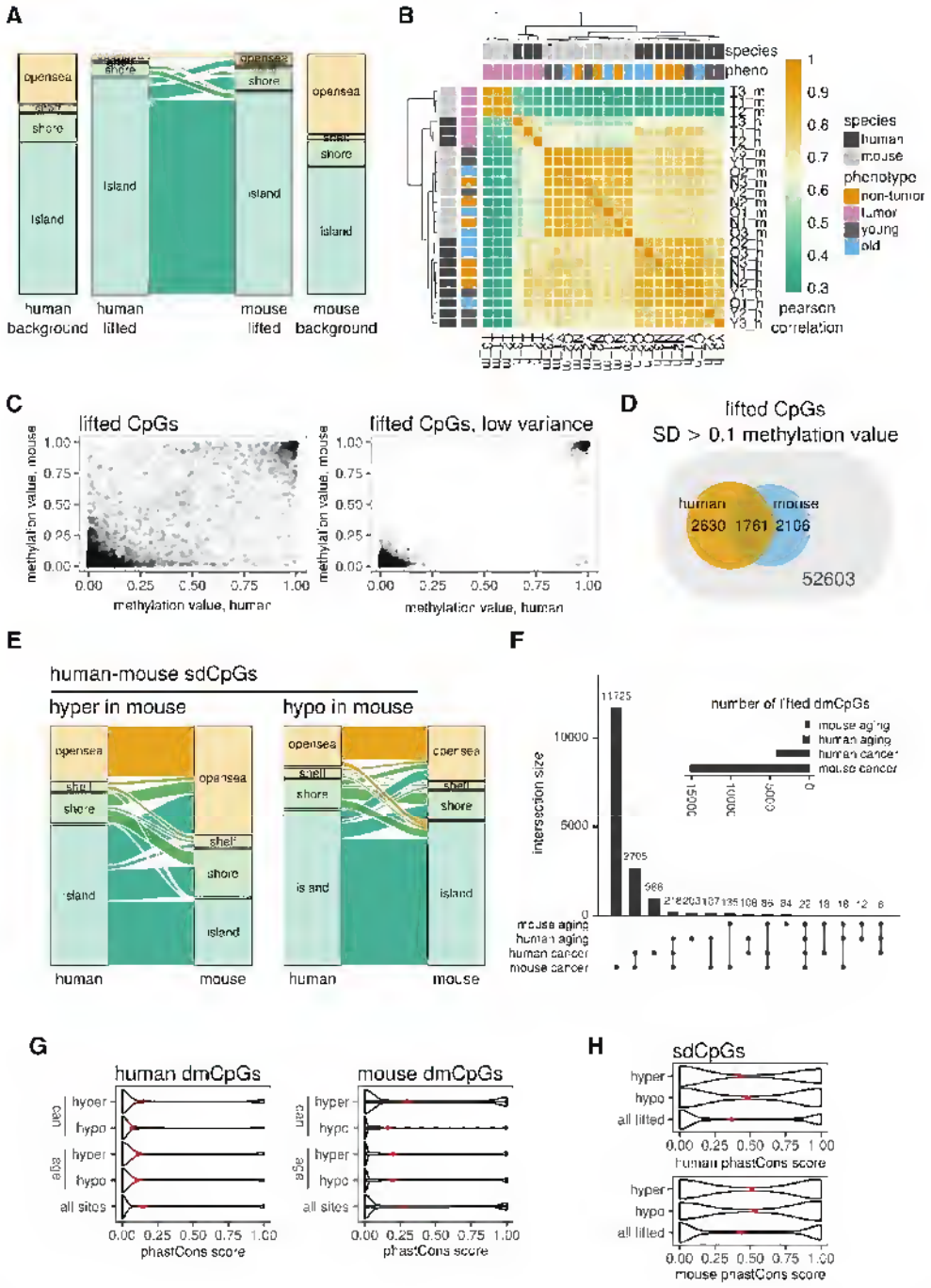


FIG. 5. Interspecies DNA methylation alterations in human and mouse. (*A*) Sankey diagram describing the distribution of changes in the CpG island status of the 59,100 interspecies orthologous CpGs when considering either the human or the mouse genomic annotation. On either side, the original backgrounds of all the profiled CpG sites are shown. (*B*) Heatmap showing the pairwise Pearson correlations of the DNA methylation levels of the interspecies CpGs across all samples. (*C*) Scatter plots showing the correlation between DNA methylation levels at all the profiled interspecies sites (left plot) or excluding intraspecies highly variable CpG sites with SD $\geq$0.1 methylation value (right plot) (tumors excluded). (*D*) Venn diagram showing the intersection of interspecies lifted CpG sites classified as highly variable within each species (SD $\geq$0.1 methylation value) for both mouse and human. (*E*) Sankey diagrams showing the distribution of hyper- and hypo-sdCpGs (species-discordant CpGs) across CpG island locations, taking into account the differences in the genomic context of the human and mouse genome. (*F*) UpSet plot describing the sets of cancer and aging dmCpGs with interspecies measurements, showing their sizes and intersections between human and mouse. (*G*) Violin plots showing the distribution of phastCons conservation scores for the detected aging- and cancer dmCpGs in human and mouse, compared with the profiled RRBS background (red dots indicate average values). (*H*) Violin plots showing the distribution of phastCons scores for the detected interspecies sdCpGs, compared with the 59,100 lifted background, either in human or mouse-centered conservation score (red dots indicate average values).

genetic regions such as exons or promoters, that is, those more impacted by cancer hypermethylation in both species (supplementary fig. S13A and B, Supplementary Material online). Second, within our 59,100 interspecies loci, we found that sdCpGs were enriched at conserved regions (fig. 5H) more frequently than the rest of the interspecies sites. This observation could at first sight appear to challenge the existing literature, which describes the positive correlation between sequence conservation and epigenetic conservation (Zhou et al. 2017), but, again, it may partly be explained by the sdCpGs being enriched at exon locations (see supplementary fig. S11D, Supplementary Material online), which are more conserved (see supplementary fig. S13B, Supplementary Material online). To address these issues, we applied stratified sampling to compare the CpG sets of interest against randomly sampled background sets of CpGs which had matching distributions of gene locations. With this strategy, we found that most of the trends regarding the difference in conservation scores observed could indeed be partly explained by the genomic context of the CpGs involved, both for cancer- and aging dmCpGs (supplementary fig. S14A, Supplementary Material online) and for human- and mouse sdCpGs (supplementary fig. S14B, Supplementary Material online). These results suggest that both disease- and species-specific epigenetic traits are influenced by the genomic context surrounding the specific CpG sites involved.

Interspecies Hotspots of Cancer- and Aging Epigenetic Deregulation

Finally, we used the interspecies CpGs to look for common hotspots of cancer- and aging epigenetic deregulation between human and mouse which could have downstream functional implications. We focused on the 22 CpGs which had been identified as dmCpGs in both species and both processes (described in fig. 5F). Of these, nine CpG sites presented concordant changes in both species (i.e., being deregulated in the same direction, in cancer and aging, for human and mouse) and were associated with a total of six genes (see supplementary table 12, Supplementary Material online). In order to study the possible functional consequences of the DNA methylation alterations, we obtained TCGA expression data for the LGG–GBM glioma cohorts and analyzed the expression of these candidate genes and their association with survival across glioma patients (see Materials and Methods). Notably, we found evidence of the strong deregulation of five out of the six genes in glioma samples as compared with controls (see below), indicating that our robust approach of looking at orthologous and concordant aging-cancer DNA methylation changes leads to the discovery of functionally relevant targets.

We first identified the *AGAP3* gene as being a target of both cancer-associated and aging-associated deregulation in both human and mouse. *AGAP3* is a component of the neural NMDA receptor complex and AGAP proteins regulate receptor trafficking in neurons (Oku and Huganir 2013). In cancer, this gene showed strong hypermethylation of internal CpG sites which, in both species, colocated with CpG islands upstream and downstream of many alternative transcripts (fig. 6A, upper plots), whereas DNA methylation alterations in the same direction, albeit smaller, also appeared with aging in both species (fig. 6A, lower plots). The specific location of the DNA methylation changes found hinted at a nonrandom and perhaps functional role for the alterations. Thus, we analyzed *AGAP3* expression in the TCGA LGG-GBM and found a strong decrease in expression in tumors (fig. 6B, Wilcoxon rank-sum $P < 0.001$). Moreover, expression levels were strongly associated with overall survival in the cohort (fig. 6C, log-rank test $P < 0.001$), with the more lethal GBM subtype having the lowest levels of expression and survival (supplementary fig. S15A and B, Supplementary Material online). Interestingly, as can be seen in figure 6A, the *AGAP3* internal hypermethylation events colocated with the ending and start of several isoforms. We therefore explored the levels of the main *AGAP3* isoforms (supplementary fig. S15C, Supplementary Material online) and their association with survival (supplementary fig. S15D, Supplementary Material online), discovering that a specific isoform (uc003wjf/ENST00000473312) was mostly explanatory of the effect (supplementary fig. S15E, Supplementary Material online, Cox Models log-rank test $P < 0.001$). Moreover, it corresponded to an isoform truncated close to the internal region targeted by the DNA methylation alterations (highlighted in fig. 6A). Indeed, studies have described isoform variants of *AGAP3* with biological functions (Oku and Huganir 2013) some of which may have roles in oncogenic processes through alterations in their mRNA expression (Shimizu et al. 2019). These observations exemplify an interspecies DNA methylation deregulation process which, as shown for human, could have a role in *AGAP3* deregulation in glioma, possibly through an effect on isoform switching, with aging-associated methylation changes preluding the cancer-associated alterations.

In addition to *AGAP3*, five other genes were associated with concordant interspecies dmCpGs (fig. 6D): *GPRIN1*, *AJM1*, *LHX2*, *CCDC177*, and *TPGS1*. Notably, four of the genes identified, much like with *AGAP3*, were clearly downregulated in tumoral samples as compared with healthy tissue (fig. 6E, Wilcoxon rank-sum tests), with the GBM subtype often displaying the strongest downregulation. We also investigated their relationship with survival and found a strong association between gene expression and overall survival for *GPRIN1*, with a more modest effect for *AJM1* and, interestingly, a trend for an association between *CCDC177* expression and survival that was specific to the GBM subtype (supplementary fig. S16, Supplementary Material online). *LHX2* is involved in neural development (Chou and Tole 2019), whereas *GPRIN1* and *TPGS1* play a role in neural cytoskeleton dynamics (Regnard et al. 2003; Nordman and Kabbani 2012) and *AJM1* is related to apical junctions (Köppen et al. 2001). As such, the deregulated genes appear to be predominantly associated with cytoskeletal processes. *LHX2* is known to be oncogenic in several types of cancer (Song et al. 2019) although we have found it to be downregulated in the case of glioma, as has been reported previously (Cheng et al. 2019). Nonetheless, the specific role of *LHX2* and of the rest of the genes in glioma

Pérez et al. · doi:10.1093/molbev/msab112 MBE

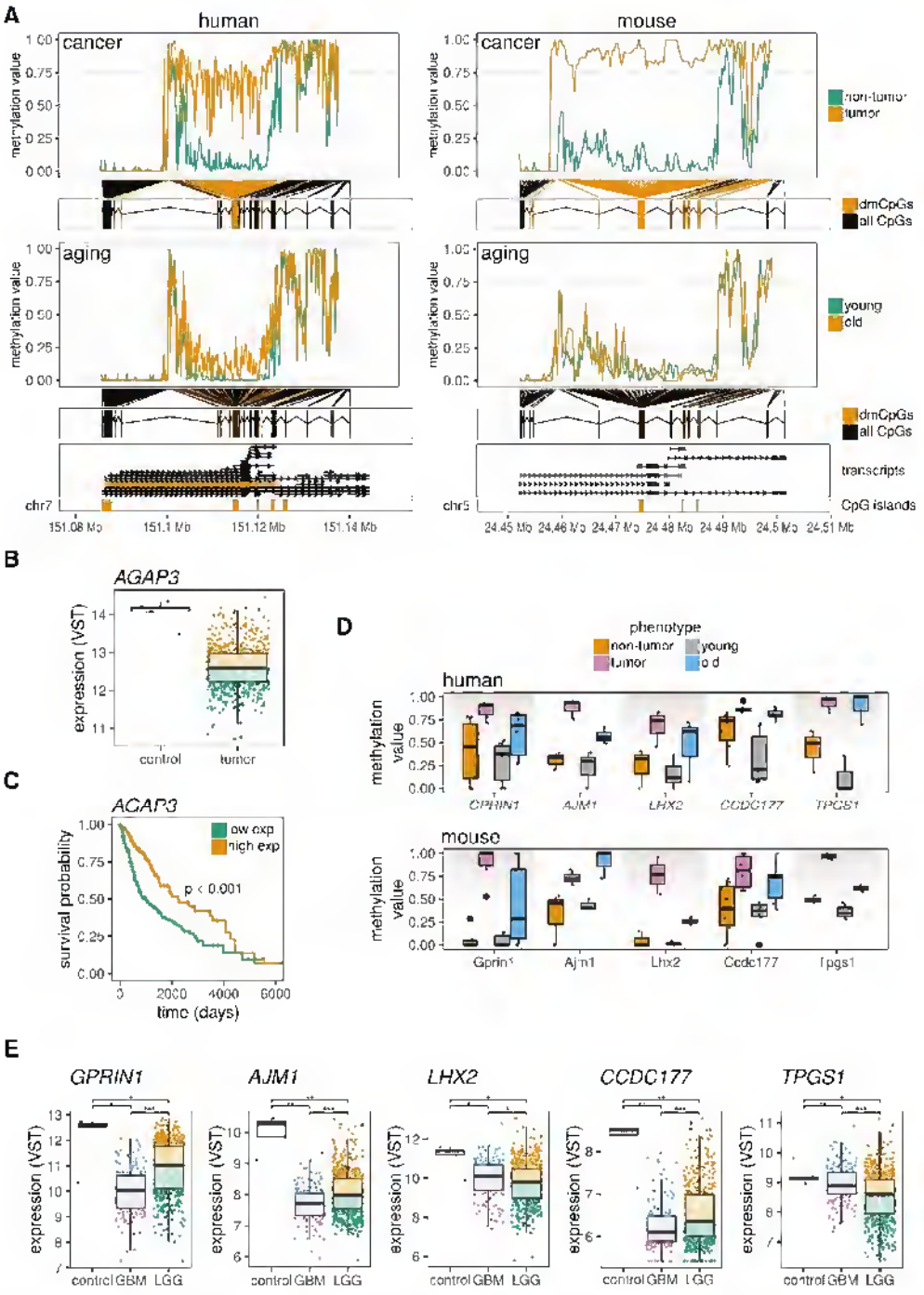


FIG. 6. *AGAP3* is an interspecies cancer- and aging-associated target of DNA methylation alterations. (*A*) Plots describing the average methylation values of tumoral, nontumoral, young, and old samples in human and mouse across the profiled CpG sites mapping to *AGAP3*. Sites marked in yellow represent significant dmCpGs in the aging and cancer comparisons. Below, the transcripts and CpG islands associated with the genomic region described are shown. (*B*) Boxplots showing the expression values of the *AGAP3* gene in control and tumor samples from the TCGA glioma cohort (GBM-LGG). Expression is measured in VST (variance stabilizing transformation) normalized units. High and low-expression samples (above and below the median) are in yellow and green. (*C*) Survival curves showing the association of high and low *AGAP3* expression groups with overall survival across glioma patients. (*D*) Boxplots describing the methylation values of the interspecies dmCpGs that presented concordant changes in cancer and aging across human and mouse and which were associated with specific genes. (*E*) Boxplots showing the expression values (VST normalized units) of the genes associated with the interspecies concordant dmCpGs in control and tumoral samples from the TCGA glioma cohort (GBM-LGG). Glioblastoma (GBM) and lower grade glioma (LGG) samples are shown separately. High- and low-expression samples (above and below the median) are color-coded. Wilcoxon rank-sum tests: $^{*}P<0.05$, $^{**}P<0.001$, $^{***}P<0.0001$, "ns," $P>0.05$.

described here, or in cancer as a whole, remains largely unexplored.

Conclusions

In this work, we have performed a systematic analysis of the DNA methylation dynamics associated with cancer and aging in brain tissue in human and mouse. We employed RRBS as a platform which allowed us to profile comparable genomic locations in both species. With this strategy, we first confirmed, in our study system, the parallelisms in baseline DNA methylation patterns between human and mouse. Subsequently, we explored the alterations associated with cancer and aging in both species. Our analyses revealed that there are considerable differences in the genomic and epigenomic features of DNA methylation alterations in cancer and aging, and that the particular characteristics of the alterations found for each process are globally conserved in human and in mouse (see supplementary fig. S17, Supplementary Material online, for a graphical outline of the main findings). We found specific parallelisms at the level of their genomic locations and the magnitude of the changes, the genetic pathways involved, the chromatin context of the alterations, the families of repetitive DNA elements affected, and the alteration of TF regulation sites, among others. Interestingly, we observed that aging-associated changes in general seem to impact more specific functional gene pathways in mouse than in human. Nonetheless, we found that the classical hypermethylation of developmental gene pathways is more shared between cancer and aging in human than in mouse. Intriguingly, epigenetic alteration of satellite repeats appeared to be a feature of both aging and cancer in human, but not in mouse. Significantly, we found very clear interspecies similarities in the aging-associated alterations occurring at the TF-binding sites of specific families, suggesting that the functional impact of aging methylation changes could indirectly occur through TF deregulation rather than through the direct targeting of specific genes. Finally, we derived a large set of orthologous CpG sites with interspecies measurements, which allowed us to observe that sites of epigenetic variability are conserved between the species, that interspecies differences between human and mouse can be partly explained by changes in the local genomic context associated with the CpG sites and that there are common interspecies loci which are altered during both aging and cancer.

The results presented in this manuscript cover several aspects of the epigenetic dynamics in human and mouse, but we should also draw attention to the limitations of our study. First, we recognize that the sample size used in our study design is limited ($n = 3$ per group), which is principally because of the substantial number of groups being analyzed (four phenotypes in two species). However, the interspecies corroboration of the results does indicate that the observed patterns are robust. Second, we have focused on a single tissue (brain). Although this allowed us to perform a comparative study in greater depth, and it is known that epigenetic alterations that arise in cancer or in aging are similar across tissues (Michalak et al. 2019), it will be of value to explore the interspecies conservation of epigenetic signatures in other tissues (Maegawa et al. 2017; Wang et al. 2017), and also other species such as pig or rhesus. Third, we used RRBS to profile analogous genomic locations in human and mouse, a technology which focuses on CpG-dense locations and is more limited when studying intergenic regions. That said, the use of RRBS allowed us to define a large subset of interspecies measurements (59,100 CpG sites) with adequate coverage across all the samples.

From an evolutionary perspective, systematic multispecies comparisons are of great value in the understanding of the etiology of disease. The uncovering of species-common pathways provides robust molecular explanations for the mechanisms at play, whereas interspecies differences may help explain the phenotypic divergences observed across the species. For instance, our observation of the human-specific epigenetic deregulation of zinc-finger genes in cancer may be related to the importance of this family in primates, which have experienced considerable lineage-specific expansion as compared with rodents (Huntley et al. 2006; Emerson and Thomas 2009). The emphasis on species-specific pathways, whereas relevant to evolutionary biology is, conversely, of less value in clinical biology, which seeks the translation of mechanisms to human. In this respect, systematic, genome-wide, multispecies integrative studies such as the one presented here have the potential to answer questions relating to both ends.

Taking a more clinical viewpoint, we can foresee that our results may spark interest in different areas. The mouse is one of the main preclinical models for human disease and, as such, it is essential to have markers for the prediction of treatment success in human (Day et al. 2015). In this vein, in this work, we found that species-common epigenetic changes across diseases were readily associated with genes showing expression alterations in glioma in human. Thus, focusing on conserved epigenetic alterations such as those described here may help narrow the search for functionally relevant targets that are altered in disease. In this sense, current advances in the development of targeted epigenetic editing tools (Liu et al. 2016), combined with the knowledge of epigenetic conservation in mouse and human, may serve as a powerful tool to explore the overall impact of both epigenetic and pathway alterations in multiple biological scenarios.

On the other hand, the set of epigenetically conserved DNA methylation sites could pave the way for the design and development of novel custom high-throughput platforms, either in the context of cost-effective microarray technology or through the use of more sophisticated targeted bisulfite sequencing approaches. This may be of particular interest within the framework of epigenetic profiling during drug screening in preclinical models as alterations of such conserved disease-associated loci could help predict future treatment success in human. For example, glioma is a cancer type with a high prevalence of isocitrate dehydrogenase mutations which lead to important epigenetic alterations (Han et al. 2020). The main chemotherapy treatment in glioma involves temozolomide (Hirst et al. 2013), and its

effectiveness is also associated with various DNA methylation features beyond the well-known *MGMT* promoter methylation (Cheng et al. 2019). Thus, integrating epigenetic conservation knowledge could aid in the identification of drug-sensitive pathways in animal models, which in turn would improve the current standard-of-care treatments as well as also shed light on the mechanistic insights of novel compounds or those rescued from current drug repositioning approaches.

All in all, this work could constitute a baseline for other future studies focusing on the systematic evaluation of disease-associated epigenetic patterns across multiple tissues and species from a multiomic perspective in order to inform the search for inter- and intraspecies patterns of epigenetic regulation.

Materials and Methods

All statistical analyses were carried out using R statistical software (v3.6.2) unless stated otherwise. Software was managed through Bioconda (Grüning et al. 2018). Graphs were produced using R base functions and the *ggplot2* R package (v3.2.1) (Wickham 2016). For the illustration and testing of multiple-set intersections, the *UpSetR* (v1.4.0) and *SuperExactTest* (v1.0.7) R packages were used (Wang et al. 2015; Conway et al. 2017). For the illustration of gene-level DNA methylation profiles, the *ggbio* R package (v1.34.0) was used (Yin et al. 2012).

Sample Description and DNA Extraction

Regarding the aging cohorts, we collected brain cortex tissue from three young (24, 33, 35 years; female, male, male) and three old (65, 69, 70 years; female, male, female) human subjects and three young (10 weeks; all female) and three old (90 weeks; all female) C57BL/6HsdOla/CBA mice. Humans and mice have different lifespans and in consequence their ages are not equally translatable across their different life stages. The young groups were selected to be at least sexually mature adults, a characteristic which is attained at 10 weeks for mice and 20 years for humans (Dutta and Sengupta 2016). Regarding the old groups and considering an average lifespan of 80 years and 24 months for humans and mice, respectively (Dutta and Sengupta 2016), the 90-week-old mice would correspond to 69-year-old humans. Consequently, the age grouping for both species lies within a similar range.

For the cancer cohorts, we collected glioma and paired cortex tissue from three human subjects (71, 31, 44 years; female, female, male), and glioma and paired cortex tissue from three mouse subjects (12–16 weeks; all female) using an in-house murine model of glioma (see supplementary table 1, Supplementary Material online, for additional information on the samples). The murine glioma model consists in the orthotopic allograft of GL261 glioma cells into the brain of C57BL/6 mice as previously described by us (Ferrer-Font et al. 2017).

DNA was extracted by standard phenol–chloroform procedures. Concentration and quality of the DNA were assessed by Qubit fluorometry (Thermo Fisher Scientific) and Fragment Analyzer capillary electrophoresis (Agilent). Because the human cohorts were composed of both male and female individuals, sex chromosomes were not included in the subsequent analyses for any species (see below) in order to avoid sex bias.

Human samples and data from patients included in this study were provided by the following institutions: the Biobank HUB-ICO-IDIBELL (PT17/0015/0024, integrated in the Spanish Biobank Network), the Biobank of Galicia Sur Health Research Institute (IISGS), and the Aragon Health Sciences Institute in the framework of the Biobank of Aragon. All human and mouse samples were processed following standard operation procedures with the appropriate approval of the corresponding Ethical and Scientific Review Boards: Bellvitge University Hospital Ethics Committee (code 07/19, April 11, 2019) for the human aging samples; Biobank IISGS Ethics Committee (code 2019/238, April 28, 2019) and Aragón Clinical Research Ethics Committee (code 14/2019, July 24, 2019) for the human cancer samples; Health Institute Carlos III Ethics Committee (code CBA PA 13_2013, February 26, 2013) for the mouse aging samples; Autonomous University of Barcelona Ethics Commission (code CEEAH-3665, February 5, 2018) for the mouse cancer samples.

Reduced Representation Bisulfite Sequencing

For each sample, 100 ng of genomic DNA were used for library preparation with the Premium Reduced Representation Bisulfite Sequencing Kit (Diagenode). Subsequently, samples were pooled in groups of 6 or 8. PCR clean-up after the final library amplification was performed using Agencourt AMPure XP (Beckman Coulter). The quality of the pools was assessed by Qubit fluorometry (Thermo Fisher Scientific) and Fragment Analyzer capillary electrophoresis (Agilent). Finally, the library pools were sequenced. Human samples were sequenced on a NovaSeq6000 (Illumina) using 50-bp paired-end read sequencing (PE50). Mouse samples were sequenced on a HiSeq3000 (Illumina) using 50-bp single-read sequencing (SR50).

Reduced Representation Bisulfite Sequencing Data Preprocessing

FASTA quality was assessed with FastQC (v0.11.8; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ last accessed April 29, 2021). Reads were quality- and adapter-trimmed with Trim galore (v0.6.4; https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ last accessed April 29, 2021) with default parameters under *–rrbs* mode and *–2colour 20* in the case of the human data to account for NovaSeq bias toward identifying high-quality G bases from nonsignal basecalls. Next, the reads were aligned to preindexed and bisulfite converted GRCh38/hg38 and GRCm38/mm10 genomes (primary assemblies) with Bismark (v0.22.2) using default parameters (Bowtie 2 alignment; see supplementary table 1, Supplementary Material online, for additional information on the alignment statistics) (Krueger and Andrews 2011). Finally, methylation counts for cytosines belonging to CpG sites were obtained with Bismark's methylation extractor under default parameters, including *–no_overlap* for paired-end data to avoid scoring overlapping methylation calls twice. For each position, methylation values

were computed as the percentage of methylated cytosines with respect to total cytosines on a scale of 0–1.

Differential Methylation Analyses

Differential methylation analyses were performed on CpG sites, not on individual cytosines, using the following process: 1) the counts for cytosines from both strands belonging to the same CpG site were pooled so as to have one methylation measurement per CpG site; 2) CpG sites mapping to sex chromosomes and alternative haplotypes were filtered out; 3) CpG sites were filtered for low coverage (<10 counts) or very high coverage (>99th percentile); 4) CpG sites common to all human or mouse samples were retained for the analyses. Next, the differential methylation analyses were carried out with the *methylKit* R package (v1.14.2) (Akalin et al. 2012) using the *calculateDiffMeth* function with default parameters and multiple testing adjustment by FDR < 0.05. The calculateDiffMeth function uses logistic regression models to define differentially methylated CpGs between groups. Because of the low number of samples ($n = 3$ per group), a strict threshold of >0.2 change in methylation value was used to define significant changes. No other covariates were included in the models because of the limited information available for other variables. Nonetheless, the inclusion of sex or age variables in the human cancer and aging comparisons led to similar results, with 97–99% of the dmCpGs being the same (data not shown).

Annotation of Sites to Genomic Locations

The profiled CpG sites were annotated to hg38 or mm10 CpG islands and gene locations using the *annotatr* R package (v1.12.1) (Cavalcante and Sartor 2017). Overlapping genetic annotations were collapsed with the following priority: Promoter > 5UTR > 3UTR > Exon > Intron > 1–5 > Intergenic. Finally, CpG sites were mapped to genes and their transcripts through the *TxDb.Hsapiens.UCSC.hg38.knownGene* and *TxDb.Mmusculus.UCSC.mm10.knownGene* R packages (v3.1.0). Annotation for the Illumina Infinium Human Methylation 450 K Beadchip was accessed through the *IlluminaHumanMethylation450kanno.ilmn12.hg19* R package (v0.6.0).

To determine the density in surrounding CpGs associated with each site, each locus was expanded to a 2,000-bp window and the number of CpGs present in the surrounding genomic sequence were counted and divided by the number of those possible.

Gene Set Enrichment Analyses

Gene set enrichment analyses were performed with the *goseq* R package (v1.38.0) (2010 P Genome Biol—Young MD). CpGs were mapped to genes, and enrichment in biological functions was carried out with genes that were exclusively hyper- or hypomethylated, using appropriate filtered backgrounds of assayed genes. The bias arising from there being different number of probes per gene was taken into account in the analyses. The GO, KEGG, Reactome, and CGP gene set databases from the Molecular Signature Database (MSigDB v7.0) (Liberzon et al. 2015) were used for the enrichment analyses and were accessed through the *msigdbr* R package (v7.0.1). In order to retain the most important enrichments, only gene sets with FDR < 0.05, OR >2 and containing between 10 and 1,000 genes were represented as statistically significant.

Region Enrichment Analyses

Genomic enrichment analyses were performed with the *LOLA* R package (v1.16.0) (Sheffield and Bock 2016) using custom databases. Sets of CpGs were tested for overenrichment in specific genomic tracks using one-sided Fisher's exact tests (FDR < 0.05) with appropriate filtered backgrounds of the CpGs being used for each case. The genomic regions analyzed included histone marks, chromatin states, and repetitive DNA elements. The histone and chromatin state *LOLA* databases built for the enrichments are available as extended data sets.

Histone ChIP-seq BED tracks for H3K4me1, H3K4me3, H3K36me3, H3K27ac, H3K27me3, and H3K9me3 in 15 healthy human tissues were obtained from LOLA extended universe (databio.org/regiondb) using NIH Roadmap Epigenomics region data corresponding to the hg38 version of the human genome (Roadmap Epigenomics Consortium et al. 2015). Histone BED tracks for the aforementioned marks in nine postnatal mouse tissues mapped to the mm10 version of the mouse genome were obtained from ENCODE (Gorkin et al. 2020).

Chromatin states were built using chromHMM (see below). Consolidated data from the aforementioned six histone tracks (BED tag.align files) and corresponding ChIP-seq input from 15 human tissues were obtained from NIH Roadmap Epigenomics (https://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated; last accessed April 29, 2021). Data corresponding to the aforementioned six histone ChIP-seq tracks (BAM files) and the corresponding ChIP-seq input from nine mouse tissues were obtained from ENCODE (https://www.encodeproject.org/; last accessed April 29, 2021).

Repetitive regions identified by RepeatMasker in the hg38 or mm10 genomes (Smit 1996) were obtained with the UCSC Table Browser (Karolchik 2004) and grouped by family or class. Elements with an uncertain classification, labeled with the "?" symbol, were filtered out.

Chromatin State Model Learning

ChromHMM (v1.18) (Ernst and Kellis 2012) was used to generate an 18-state model to predict chromatin states from 15 human and nine mouse tissues. ChIP-seq data sets corresponding to six histone marks (H3K4me1, H3K4me3, H3K36me3, H3K27ac, H3K27me3, and H3K9me3), along with their respective input controls, were obtained from NIH Roadmap Epigenomics and ENCODE for human and mouse samples respectively (see supplementary table 11, Supplementary Material online, for the description of the files). Human BED files were converted from hg19 to hg38 coordinates via UCSC's liftOver tool (v377) (Hinrichs 2006). Resulting files were binarized using the *binarizeBed* function with default parameters using the hg38 chromosome sizes. A model considering 18 states was learned using the

Pérez et al. · doi:10.1093/molbev/msab112 MBE

LearnModel function with default parameters, as previously described elsewhere (Roadmap Epigenomics Consortium et al. 2015). In the case of mouse data sets (mm10), BAM files corresponding to replicated samples from different mouse tissues were merged using SAMtools (v1.7) (Li et al. 2009), as previous studies have suggested that chromatin state inferences either from single time series or from replicated data are invariably consistent and highly reproducible. Read duplicates were removed using the *MarkDuplicates* function from Picard tools (broadinstitute.github.io/picard/). The resulting BAM files were sorted using SAMtools and binarized using the *binarizeBam* function from chromHMM. A model considering 18 states was learned using the *LearnModel* function using the same conditions as the human chromHMM pipeline. For both human and mouse analyses, the resulting chromatin states were reordered and renamed to facilitate later interpretation purposes. In addition, enrichment analyses of the different chromatin states in external genomic data sets (CpG Islands, RefSeq Exons, RefSeq genes, RefSeq transcription start sites, RefSeq transcript end sites, and regions within 2,000 bp of a RefSeq TSS) obtained from the UCSC genome browser (hg38 and mm10 genomes) were performed using the *OverlapEnrichment* function from chromHMM.

Motif Enrichment Analyses

Transcription factor motif enrichment analyses were performed with Hypergeometric Optimization of Motif EnRichment (HOMER) software (v4.11.1) (Heinz et al. 2010) for cancer- and aging dmCpGs in human and mouse, using the hg38 or mm10 genomes as reference. HOMER was run under default settings and with the *-keepOverlappingBg* parameter, using as background the filtered universe of CpG sites analyzed by the RRBS for human and mouse, respectively. Significant enrichments for known TFs were first filtered by FDR < 0.05, and subsequently only the top 50 results ordered by enrichment with respect to the background were retained for each comparison (see supplementary table 8, Supplementary Material online, for full results).

Liftover of Orthologous CpGs

Liftover of mm10 to hg38 coordinates was performed with UCSC's liftOver tool (v377) (Hinrichs 2006) by using chain files from https://hgdownload.soe.ucsc.edu/downloads.html last accessed April 29, 2021. Lifted CpG sites which did not reciprocally lift back from hg38 to the original mm10 coordinates were discarded (~ 2%).

Conservation Analyses

Base-resolution PhastCons tracks for 100-way (hg38) and 60-way (mm10) alignments (Siepel et al. 2005) were accessed via the *GenomicScores* R package (v1.1.10) (Puigdevall and Castelo 2018).

Gene Expression Analyses

Gene and isoform-level gene expression data in the form of RSEM counts and curated survival data were accessed via the Broad Institute Genomic Data Analysis Center (http://gdac.broadinstitute.org/runs/stddata__2016_01_28/data/; last accessed April 29, 2021) and the UCSC Xena (https://xenabrowser.net/datapages/; last accessed April 29, 2021) for the glioma GBMLGG (glioblastoma—lower grade glioma) cohort from the TCGA (The Cancer Genome Atlas) consortium (Brennan et al. 2013; Ceccarelli et al. 2016) (the list of patients and genes analyzed is available as an extended data set). Prior to analysis, data were normalized by the VST (variance stabilizing transformation) procedure using the *DESeq2* R package (v1.26.0) (Anders and Huber 2010). This transformation stabilizes the variance–mean relationship in the data and corrects for differences in library size between the samples. For the isoform analysis, the four most expressed *AGAP3* isoforms were selected. Cox models were analyzed with the *survival* R package (v3.1.11) (Therneau and Grambsch 2000).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgements

The authors would like to thank Ronnie Lendrum for manuscript editing and the members of the Cancer Epigenetics Laboratory for their positive feedback. Finally, they want to particularly acknowledge the patients and biobanks involved in the study: the Biobank HUB-ICO-IDIBELL (PT17/0015/0024, integrated in the Spanish Biobank Network), the Biobank of Galicia Sur Health Research Institute (IISGS), and the Aragon Health Sciences Institute in the framework of the Biobank of Aragon. Specifically, they thank Laura Arregui Egido (HUB-ICO-IDIBELL), Delia Recalde Frisón (Biobank of Aragon) and Susana Teijeira Bautista, Vanesa Val Varela, and Olga Souto Rodríguez (IISGS) for their contribution in the collection of the samples. This work was supported by the Spanish Association Against Cancer (PROYE18061FERN to M.F.F.), the Asturias Government (PCTI) cofunding 2018-2022/FEDER (IDI/2018/146 to M.F.F.), Fundación General CSIC (0348_CIE_6_E to M.F.F.), and the Health Institute Carlos III (Plan Nacional de I+D+I) cofunding FEDER (PI15/00892 and PI18/01527 to M.F.F and A.F.F.). They also acknowledge support from the Ramón Areces Foundation (CIVP18A3891 to P.J.F.M.), the AECC (SIRTBIO to P.J.F.M.), the MICINN (SAF2017-85766-R to P.J.F.M.), a Ramón y Cajal fellowship (MICINN, RYC-2017-22335 to P.J.F.M.), and the European Commission ATTRACT project (777222 to A.P.C.). J.R.T. is supported by a Juan de la Cierva fellowship from the Spanish Ministry of Science and Innovation (FJCI-2015-26965). R.F.P. and P.S.O. are supported by the Severo Ochoa program (BP17-114 and BP17-165, respectively). R.G.U. is supported by the Centro de Investigación Biomédica en Red de Enfermedades Raras (Health Institute Carlos III). L.V. was supported by the UAB Predoctoral training programme (PIF predoctoral fellowships). They also acknowledge support from the IMDEA Food Institute and IUOPA-ISPA-FINBA (the IUOPA is supported by the Obra Social Cajastur-Liberbank, Spain).

Author Contributions

R.F.P., J.R.T., A.F.F., and M.F.F. conceived, coordinated, and supervised the study. R.F.P. and J.R.T. designed all aspects of the research and contributed equally to this work. R.F.P. and J.R.T. collected the data, performed computational analyses, and wrote the manuscript. A.F.F., P.S.O., V.L.M., and R.G.U. analyzed and interpreted the data. N.V.S. provided clinical samples and biological data for human samples. A.P.C. and L.V. generated the mouse glioma models. M.B., P.J.F.M., and M.S. generated the mouse aging models. All authors revised, read, and approved the final manuscript.

Data Availability

The data underlying this article are available in the article and in its Supplementary Material online. Additionally, extended data sets are available in Zenodo, at https://dx.doi.org/10.5281/zenodo.4573676; last accessed April 29, 2021, principally consisting of scripts used in the analyses, extended results, the matrices of preprocessed methylation values and the lists of orthologous CpG sites, the databases of histones, repetitive DNA elements and chromatin states, information regarding the RNA-seq data analyzed. The sequencing files from the RRBS experiments have been deposited in the European Nucleotide Archive (ENA) under the accession number PRJEB41460 (mouse data) and in the European Genome-Phenome Archive (EGA) under the accession number EGAS00001004851 (human data). Finally, an R Shiny app is available at https://epilabasturias.shinyapps.io/mbe_app/; last accessed April 29, 2021 which describes and facilitates access to the results of the differential methylation analyses.

References

Ai lin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 13(10):R87.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11(10):R106.

Aunan JR, Cho WC, Søreide K. 2017. The biology of aging and cancer: a brief overview of shared and divergent molecular hallmarks. *A&D* 8(5):628.

Baylin SB, Jones PA. 2016. Epigenetic determinants of cancer. *Cold Spring Harb Perspect Biol.* 8(9):a019505.

Beisel C, Paro R. 2011. Silencing chromatin: comparing modes and mechanisms. *Nat Rev Genet.* 12(2):123–135.

Blahnik KR, Dou L, Echipare L, Iyengar S, O'Geen H, Sanchez E, Zhao Y, Marra MA, Hirst M, Costello JF, et al. 2011. Characterization of the contradictory chromatin signatures at the 3′ exons of zinc finger genes. *PLoS One* 6:e17121.

Blewitt M, Whitelaw E. 2013. The use of mouse models to study epigenetics. *Cold Spring Harb Perspect Biol.* 5(11):a017939.

Bock C, Tomazou EM, Brinkman A, Müller F, Simmer F, Gu H, Jäger N, Gnirke A, Stunnenberg HG, Meissner A. 2010. Genome-wide mapping of DNA methylation: a quantitative technology comparison. *Nat Biotechnol.* 28(10):1106–1114.

Brennan CW, Verhaak RGW, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, et al. 2013. The somatic genomic landscape of glioblastoma. *Cell* 155(2):462–477.

Carrasco-García E, Moreno-Cugnon L, García I, Borras C, Revuelta M, Izeta A, Lopez-Lluch G, de Pancorbo MM, Vergara I, Vina J, et al.. 2 9. SOX2 expression diminishes with ageing in several tissues in mice and humans. Mech Ageing Dev. 177:30–36. 10.1016/j.mad.2018.03.008 29574045

Cavalcante RG, Sartor MA. 2017. annotatr: genomic regions in context. *Bioinformatics* 33(15):2381 -2383.

Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, Morozova O, Newton Y, Radenbaugh A, Pagnotta SM, et al. 2016. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 164(3):550–563.

Chen F, Zhang Q, Deng X, Zhang X, Chen C, Lv D, Li Y, Li D, Zhang Y, Li P, et al. 2018. Conflicts of CpG density and DNA methylation are proximally and distally involved in gene regulation in human and mouse tissues. *Epigenetics* 13(7):721–741.

Cheng Q, Huang C, Cao H, Lin J, Gong X, Li J, Chen Y, Tian Z, Fang Z, Huang J. 2019. A novel prognostic signature of transcription factors for the prediction in patients with GBM. *Front Genet.* 10:906.

Chou S-J, Tole S. 2019. Lhx2, an evolutionarily conserved, multifunctional regulator of forebrain development. *Brain Res.* 1705:1–14.

Cole JJ, Robertson NA, Rather MI, Thomson JP, McBryan T, Sproul D, Wang T, Brock C, Clark W, Ideker T, et al. 2017. Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol.* 18(1):58.

Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33(18):2938–2940.

Day C-P, Merlino G, Van Dyke T. 2015. Preclinical mouse cancer models: a maze of opportunities and challenges. *Cell* 163(1):39–53.

Dennis DJ, Han S, Schuurmans C. 2019. bHLH transcription factors in neural development, disease, and reprogramming. *Brain Res.* 1705:48–65.

Dmitrijeva M, Ossowski S, Serrano L, Schaefer MH. 2018. Tissue-specific DNA methylation loss during ageing and carcinogenesis is linked to chromosome structure, replication timing and cell division rates. *Nucleic Acids Res.* 46(14):7022–7039.

Du J, Johnson LM, Jacobsen SE, Patel DJ. 2015. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol.* 16(9):519–532.

Dutta S, Sengupta P. 2016. Men and mice: relating their ages. *Life Sci.* 152:244–248.

Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, Milekic MH, Chanrion B, Fu Y, Su T, Hibshoosh H, et al. 2010. Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Res.* 20(7):972–980.

Elango N, Yi SV. 2008. DNA methylation and structural and functional bimodality of vertebrate promoters. *Mol Biol Evol.* 25(8):1602–1608.

Emerson RO, Thomas JH. 2009. Adaptive evolution in zinc finger transcription factors. *PLoS Genet.* 5(1):e1000325.

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol.* 28(8):817–825.

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

Espada J, Esteller M. 2013. Mouse models in epigenetics: insights in development and disease. *Brief Funct Genomics.* 12(3):279–287.

Fernández AF, Bayón GF, Urdinguio RG, Toraño FG, García MG, Carella A, Petrus-Reurer S, Ferrero C, Martinez-Camblor P, Cubillo I, et al. 2015. H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells. *Genome Res.* 25(1):27–40.

Ferrer-Font L, Villamañan L, Arias-Ramos N, Vilardell J, Plana M, Ruzzene M, Pinna L, Itarte E, Arús C, Candiota A. 2017. Targeting protein kinase CK2: evaluating CX-4945 potential for GL261 glioblastoma therapy in immunocompetent mice. *Pharmaceuticals* 10(4):24.

Foronda M, Martínez P, Schoeftner S, Gómez-López G, Schneider R, Flores JM, Pisano DG, Blasco MA. 2014. Sox4 links tumor suppression

to accelerated aging in mice by modulating stem cell activation. *Cell Rep.* 8(2):487–500.

Gorkin DU, Barozzi I, Zhao Y, Zhang Y, Huang H, Lee AY, Li B, Chiou J, Wildberg A, Ding B, et al. 2020. An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* 583(7818):744–751.

Greenberg MVC, Bourc'his D. 2019. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol.* 20(10):590–607.

Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, Valieris R, Köster J. 2018. Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat Methods.* 15(7):475–476.

Hadad N, Masser DR, Blanco-Berdugo L, Stanford DR, Freeman WM. 2019. Early-life DNA methylation profiles are indicative of age-related transcriptome changes. *Epigenet Chromatin.* 12(1):58.

Hahn O, Grönke S, Stubbs TM, Ficz G, Hendrich O, Krueger F, Andrews S, Zhang Q, Wakelam MJ, Beyer A, et al. 2017. Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. *Genome Biol.* 18(1):56.

Han S, Liu Y, Cai SJ, Qian M, Ding J, Larion M, Gilbert MR, Yang C. 2020. IDH mutation in glioma: molecular mechanisms and potential therapeutic targets. *Br J Cancer.* 122(11):1580–1589.

Hanna CW, Demond H, Kelsey G. 2018. Epigenetic regulation in development: is the mouse a good model for the human? *Hum Reprod Update.* 24(5):556–576.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 38(4):576–589.

Hernando-Herraez I, Evano B, Stubbs T, Commere P-H, Jan Bonder M, Clark S, Andrews S, Tajbakhsh S, Reik W. 2019. Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nat Commun.* 10(1):4361.

Hernando-Herraez I, Garcia-Perez R, Sharp AJ, Marques-Bonet T. 2015. DNA methylation: insights into human evolution. *PLoS Genet.* 11(12):e1005661.

Hernando-Herraez I, Heyn H, Fernandez-Callejo M, Vidal E, Fernandez-Bellon H, Prado-Martinez J, Sharp AJ, Esteller M, Marques-Bonet T. 2015. The interplay between DNA methylation and sequence divergence in recent human evolution. *Nucleic Acids Res.* 43(17):8204–8214.

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J. 2006. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* 34(Database issue):D590–D598.

Hirst TC, Vesterinen HM, Sena ES, Egan KJ, Macleod MR, Whittle IR. 2013. Systematic review and meta-analysis of temozolomide in animal models of glioma: was clinical efficacy predicted? *Br J Cancer.* 108(1):64–71.

Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L. 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res.* 16(5):669–677.

Karolchik D. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 32(90001):493D–496.

Kessler NJ, Van Baak TE, Baker MS, Laritsky E, Coarfa C, Waterland RA. 2016. CpG methylation differences between neurons and glia are highly conserved from mouse to human. *Hum Mol Genet.* 25(2):223–232.

Köppen M, Simske JS, Sims PA, Firestein BL, Hall DH, Radice AD, Rongo C, Hardin JD. 2001. Cooperative regulation of AJM-1 controls junctional integrity in *Caenorhabditis elegans* epithelia. *Nat Cell Biol.* 3(11):983–991.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for bisulfite-Seq applications. *Bioinformatics* 27(11):1571–1572.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1(6):417–425.

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, et al. 2013. Global epigenomic reconfiguration during mammalian brain development. *Science* 341(6146):1237905–1237905.

Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R. 2016. Editing DNA methylation in the mammalian genome. *Cell* 167(1):233–247.e17.

Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, Grützner F, Odom DT, Patient R, Ponting CP, et al. 2013. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *eLife* 2:e00348.

Maegawa S, Gough SM, Watanabe-Okochi N, Lu Y, Zhang N, Castoro RJ, Estecio MRH, Jelinek J, Liang S, Kitamura T, et al. 2014. Age-related epigenetic drift in the pathogenesis of MDS and AML. *Genome Res.* 24(4):580–591.

Maegawa S, Lu Y, Tahara T, Lee JT, Madzo J, Liang S, Jelinek J, Colman RJ, Issa J-PJ. 2017. Caloric restriction delays age-related methylation drift. *Nat Commun.* 8(1):539.

Masser DR, Hadad N, Porter HL, Mangold CA, Unnikrishnan A, Ford MM, Giles CB, Georgescu C, Dozmorov MG, Wren JD, et al. 2017. Sexually divergent DNA methylation patterns with hippocampal aging. *Aging Cell* 16(6):1342–1352.

McKinney BC, Lin C-W, Rahman T, Oh H, Lewis DA, Tseng G, Sibille E. 2019. DNA methylation in the human frontal cortex reveals a putative mechanism for age-by-disease interactions. *Transl Psychiatry.* 9(1):39.

Mendizabal I, Shi L, Keller TE, Konopka G, Preuss TM, Hsieh T-F, Hu E, Zhang Z, Su B, Yi SV. 2016. Comparative methylome analyses identify epigenetic regulatory loci of human brain evolution. *Mol Biol Evol.* 33(11):2947–2959.

Michalak EM, Burr ML, Bannister AJ, Dawson MA. 2019. The roles of DNA, RNA and histone methylation in ageing and cancer. *Nat Rev Mol Cell Biol.* 20(10):573–589.

Nordman JC, Kabbani N. 2012. An interaction between α7 nicotinic receptors and a G-protein pathway complex regulates neurite growth in neural cells. *J Cell Sci.* 125(22):5502–5513.

Oku Y, Huganir RL. 2013. AGAP3 and Arl6 regulate trafficking of AMPA receptors and synaptic plasticity. *J Neurosci.* 33(31):12586–12598.

Pérez RF, Tejedor JR, Bayón GF, Fernández AF, Fraga MF. 2018. Distinct chromatin signatures of DNA hypomethylation in aging and cancer. *Aging Cell* 17(3):e12744.

Puigdevall P, Castelo R. 2018. GenomicScores: seamless access to genomewide position-specific scores from R and Bioconductor. *Bioinformatics* 34(18):3208–3210.

Regnard C, Fesquet D, Janke C, Boucher D, Desbruyéres E, Koulakoff A, Insina C, Travo P, Eddé B. 2003. Characterisation of PGs1, a subunit of a protein complex co-purifying with tubulin polyglutamylase. *J Cell Sci.* 116(Pt 20):4181–4190.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317–330.

Ross JP, Rand KN, Molloy PL. 2010. Hypomethylation of repeated DNA sequences in cancer. *Epigenomics* 2(2):245–269.

Schroeder DI, Jayashankar K, Douglas KC, Thirkill TL, York D, Dickinson PJ, Williams LE, Samollow PB, Ross PJ, Bannasch DL, et al. 2015. Early developmental and evolutionary origins of gene body DNA methylation patterns in mammalian placentas. *PLoS Genet.* 11:e1005442.

Sen P, Shah PP, Nativio R, Berger SL. 2016. Epigenetic mechanisms of longevity and aging. *Cell* 166(4):822–839.

Sheffield NC, Bock C. 2016. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* 32(4):587–589.

Shimizu D, Masuda T, Sato K, Tsuruda Y, Otsu H, Kuroda Y, Eguchi H, Kodera Y, Mimori K. 2019. CRMP5-associated GTPase (CRAG) is a candidate driver gene for colorectal cancer carcinogenesis. *Anticancer Res.* 39(1):99–106.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15(8):1034–1050.

Sizemore GM, Pitarresi JR, Balakrishnan S, Ostrowski MC. 2017. The ETS family of oncogenic transcription factors in solid tumours. *Nat Rev Cancer.* 17(6):337–351.

Smit AF. 1996. The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev.* 6(6):743–748.

Song H, Liu J, Wu X, Zhou Y, Chen X, Chen J, Deng K, Mao C, Huang S, Liu Z. 2019. LHX2 promotes malignancy and inhibits autophagy via mTOR in osteosarcoma and is negatively regulated by miR-129-5p. *Aging* 11(21):9794–9810.

Srivastava R, Srivastava R, Ahn SH. 2016. The epigenetic pathways to ribosomal DNA silencing. *Microbiol Mol Biol Rev.* 80(3):545–563.

Sun D, Luo M, Jeong M, Rodriguez B, Xia Z, Hannah R, Wang H, Le T, Faull KF, Chen R, et al. 2014. Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell* 14(5):673–688.

Therneau TM, Grambsch PM. 2000. Modeling survival data: extending the Cox model. New York: Springer-Verlag.

Wang M, Zhao Y, Zhang B. 2015. Efficient test and visualization of multi-set intersections. *Sci Rep.* 5(1):16923.

Wang T, Tsui B, Kreisberg JF, Robertson NA, Gross AM, Yu MK, Carter H, Brown-Borg HM, Adams PD, Ideker T. 2017. Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol.* 18(1):57.

Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer International Publishing. ISBN 978-0-387-98140-6

Yin T, Cook D, Lawrence M. 2012. ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol.* 13(8):R77.

Yu M, Hazelton WD, Luebeck GE, Grady WM. 2020. Epigenetic aging: more than just a clock when it comes to cancer. *Cancer Res.* 80(3):367–374.

Zhou J, Sears RL, Xing X, Zhang B, Li D, Rockweiler NB, Jang HS, Choudhary MNK, Lee HJ, Lowdon RF, et al. 2017. Tissue-specific DNA methylation is conserved across human, mouse, and rat, and driven by primary sequence conservation. *BMC Genomics* 18(1):724.

AGING AND CANCER: THIRD ARTICLE

Pérez RF, Tejedor JR, Fernández AF & Fraga MF (2022) Aging and cancer epigenetics: where do the paths fork? Aging Cell, in press, e13709.
In this work, we sought to critically re-evaluate the current scientific knowledge of aging and cancer epigenetics while incorporating the new perspectives gained from our previous investigations. To this end, we reviewed the current literature in the field and identified open questions regarding the roles of DNA methylation as a molecular link between aging and cancer. In particular, we highlighted the potential of epigenetic clocks as novel tools in the study of these processes. Moreover, we called attention to a potential knowledge gap regarding epigenetic aging in somatic stem cells, whereby it is not yet clear whether these cells display a "normal" epigenetic aging or maintain a youthful phenotype as is the case of embryonic stem cells. To explore this question, we performed a modest analysis of 13 independent DNA methylation array-based datasets and observed that somatic stem cells seem to display epigenetic aging across multiple tissues. Somatic stem cells play fundamental roles in the development of cancer, and thus their epigenetic deterioration across time may help explain the link between aging and cancer. In the future, single cell-level studies will help shed light on these and other open issues.
Personal contribution to the work: I designed the study, reviewed the literature and preprocessed and analyzed all the data as well as prepared the figure panels presented in this manuscript. The work was co-written and supervised by Dr. Juan Ramón Tejedor, Dr. Agustín F. Fernández and Dr. Mario F. Fraga.

Received: 3 August 2022 | Accepted: 29 August 2022

DOI: 10.1111/acel.13709

Aging Cell WILEY

MINI REVIEW

Aging and cancer epigenetics: Where do the paths fork?

Raúl Fernández Pérez[1,2,3] | Juan Ramón Tejedor[1,2,3] | Agustín Fernández Fernández[1,2,3] | Mario Fernández Fraga[1,2,3]

[1]Cancer Epigenetics and Nanomedicine Laboratory, Nanomaterials and Nanotechnology Research Center (CINN-CSIC), El Entrego, Spain

[2]Health Research Institute of Asturias (ISPA-FINBA), Institute of Oncology of Asturias (IUOPA) and Department of Organisms and Systems Biology (BOS), University of Oviedo, Oviedo, Spain

[3]Rare Diseases CIBER (CIBERER), Carlos III Health Institute (ISCIII), Madrid, Spain

Correspondence
Agustín Fernández Fernández and Mario Fernández Fraga, Cancer Epigenetics and Nanomedicine Laboratory, Nanomaterials and Nanotechnology Research Center (CINN-CSIC).
Emails: agustin.fernandez@cinn.es (A. F. F.) and mffraga@cinn.es (M. F. F.)

Funding information
Asturias Government (PCTI) cofunding 2018-2022/FEDER, Grant/Award Number: IDI/2018/146 and BP17-114; Centro de Investigación Biomédica en Red de Enfermedades Raras CIBERER; Fundación Bancaria Caja de Ahorros de Asturias; Fundación Científica Asociación Española Contra el Cáncer, Grant/Award Number: PROYE18061FERN; Fundación General CSIC, Grant/Award Number: 0848_CIE_6_E; Instituto de Salud Carlos III, Grant/Award Number: PI21/01067; Ministerio de Ciencia e Innovación, Grant/Award Number: IJC2018-36825-I and SGL2021-03-039/40

Abstract

Aging and cancer are clearly associated processes, at both the epidemiological and molecular level. Epigenetic mechanisms are good candidates to explain the molecular links between the two phenomena, but recent reports have also revealed considerable differences, particularly regarding the loss of DNA methylation in the two processes. The large-scale generation and availability of genome-wide epigenetic data now permits systematic studies to be undertaken which may help clarify the similarities and differences between aging and cancer epigenetic alterations. In addition, the development of epigenetic clocks provides a new dimension in which to investigate diseases at the molecular level. Here, we examine current and future questions about the roles of DNA methylation mechanisms as causal factors in the processes of aging and cancer so that we may better understand if and how aging-associated epigenetic alterations lead to tumorigenesis. It seems certain that comprehending the molecular mechanisms underlying epigenetic clocks, especially with regard to somatic stem cell aging, combined with applying single-cell epigenetic-age profiling technologies to aging and cancer cohorts, and the integration of existing and upcoming epigenetic evidence within the genetic damage models of aging will prove to be crucial to improving understanding of these two interrelated phenomena.

KEYWORDS
aging, cancer, DNA methylation, epigenetic clock, stem cells

1 | INTRODUCTION

Aging, a quasi-universal biological phenomenon, is inextricably linked to the development of cancer (de Magalhães, 2013). Understanding the mechanisms underlying the causative role that this gradual, time-dependent accumulation of molecular damage plays in disease etiology is crucial to designing interventions seeking to improve our healthspan—that is, the lifespan free of disease.

Abbreviations: 5mC, 5-methylcytosine; CpG, cytosine-guanine dinucleotide; DNAm, DNA methylation; ESC, embryonic stem cell; HSC, hematopoietic stem cell; MSC, mesenchymal stem cell.

In this scenario, the field of aging research—and aging epigenetics in particular—is experiencing an explosive revolution owing to the development and democratization of high-throughput technologies and computational methods (Hasin et al., 2017), which has led to the large-scale profiling of sizable cohorts using genome-wide tools. Against this background, we can now re-examine the available evidence regarding the molecular parallelisms and differences between aging and cancer in order to shed light on the relationship between the two processes.

Classically, parallel epigenetic alterations have been described for aging and cancer, mainly concerning the local DNA hypermethylation of CpG islands which is typically marked by bivalent chromatin domains in embryonic stem cells (ESCs) (Day et al., 2013; Easwaran et al., 2012; Fernández et al., 2015; Heyn et al., 2012; Ohm et al., 2007; Rakyan et al., 2010; Schlesinger et al., 2007; Teschendorff et al., 2010; Widschwendter et al., 2007) and the global loss of DNA methylation (DNAm), particularly at repetitive sequencies (Bollati et al., 2009; Ehrlich, 2009; Fuke et al., 2004; Li et al., 2014; Wilson et al., 1987).

Nonetheless, these similitudes have recently been challenged, particularly in terms of DNA hypomethylation, by studies using next-generation sequencing technologies to measure 5mC at the genome-wide level (Unnikrishnan et al., 2018). This has been especially the case with research involving mouse models of aging, which have failed to report global DNA hypomethylation in this process, across multiple tissues (Cole et al., 2017; Hadad et al., 2019; Hahn et al., 2017; Hernando-Herraez et al., 2019; Masser et al., 2017; Sun et al., 2014). These observations have prompted systematic studies to be undertaken that seek to accurately characterize the similitudes and differences between aging- and cancer-associated epigenetic alterations, often revealing clear differences between aging, cancer, and senescence (Pérez et al., 2018; Xie et al., 2018), some of which have been detectable across human and mouse (Pérez et al., 2021). There is therefore a need to clarify, within the epigenetic context, the functional relationships between the two processes.

Aside from the canonical hyper- and hypomethylation signatures of aging and cancer, the recent development of epigenetic clocks (Horvath & Raj, 2018) has provided us with a new dimension in which to explore aging-associated epigenetic dysregulation and its link with carcinogenesis. In this work, we explore the current knowledge regarding the DNAm alterations classically described in aging and cancer and consider epigenetic clocks as novel, albeit different, tools which may help clarify the complex relationship between the two processes, paying particular attention to the phenomenon of somatic stem cell epigenetic aging.

2 | THE ROLE OF CLASSIC AGING-ASSOCIATED DNA METHYLATION ALTERATIONS IN CANCER

There appear to exist a number of functional connections between aging and cancer that involve epigenetic mechanisms. Especially with regard to DNA hypermethylation, it seems that the genomic loci which experience aging-associated gains in DNAm in healthy tissue are in turn aberrantly hypermethylated in cancer (Lin & Wagner, 2015; Luebeck et al., 2019), which would seem to indicate that tumors present a "hyperaged" phenotype that could be associated with cancer risk and survival (Klutstein et al., 2017; Lin & Wagner, 2015). These epigenetic alterations can be linked to mitotic activity and, indeed, epigenetic mitotic clocks have been developed by tracking DNAm gains at specific loci (Yang et al., 2016). In this same line, more mechanistic studies have shown how spontaneous epigenetic lesions accumulated over time can facilitate oncogenic transformation, for example, in mouse colorectal organoids through the repression of key genes (Tao et al., 2019). Of course, time-associated alterations may also reflect the accumulation of lifestyle-related aggressions: for instance, long-term exposure to cigarette smoke condensate has been shown to produce epigenetic alterations leading to tumorigenesis in human lung cells (Belinsky et al., 2002; Vaz et al., 2017). With respect to the loss of DNAm, the commonalities between aging and cancer again appear to be related to cell division, and DNA hypomethylation has been observed at late-replicating, lamina-associated domains in both aging and cancer, with cancer once again manifesting stronger alterations (Dmitrijeva et al., 2018; Zhou et al., 2018). However, the functional impact of the gradual loss of methylation at these gene-poor, heterochromatin associated loci remains to be fully clarified: while aging-associated hypermethylation gives rise to alterations which may indeed facilitate the oncogenic process, the canonical global hypomethylation often seen in tumors could actually be a consequence, and not a cause, of the dramatic cellular expansion occurring after malignant transformation.

Thus, it would seem that an important functional epigenetic link between aging and cancer relates to DNAm alterations, which partly reflect the mitotic history of the tissues involved—an "accelerated" history in the case of tumors—and, particularly regarding DNA hypermethylation, these changes may functionally influence the oncogenic process by dysregulating the expression of cancer-related genes. Nonetheless, it is possible that there exist other aging-associated epigenetic alterations which have an impact on cancer development but are otherwise not present, or non-detectable, in tumoral tissues: (1) subtle aging changes could be masked, or reverted, by global epigenetic reconfiguration in tumors (Pérez et al., 2018); (2) aging-related alterations in the tumor microenvironment may have oncogenic roles (Marks et al., 2016); (3) epigenetic changes in the immune system brought about by immunosenescence could also modify the risk of, and response to, cancer in more indirect ways (Pawelec, 2017).

In addition, other aging-associated phenomenon may contribute to concealing the putative functional characteristics of aging-related epigenetic alterations, including the existence of a stochastic component inherent to age-associated DNAm alterations (Seale et al., 2022) and the occurrence of age-associated changes in cellular composition (Campagna et al., 2021). Nonetheless, the phenotypic and epigenetic variability brought about by aging via

mechanisms such as tissue disruption may in itself causally contribute to the oncogenic process, as has been recently proposed (Capp & Thomas, 2021). To tackle these concerns, systematic meta-analysis studies focusing on multi-tissue common alterations in aging and cancer (Chen et al., 2016; Pérez et al., 2018) and also the use of single-cell epigenetic profiling technologies (Cheung et al., 2018; Hernando-Herraez et al., 2019) are clearly needed.

3 | DNA METHYLATION CLOCKS AS NEW TOOLS IN THE STUDY OF DISEASE

The past decade has been witness to a revolution in the field of aging epigenetics: the development of epigenetic clocks. Epigenetic clocks are mathematical algorithms which can, with great precision, predict the chronological age of subjects by combining measured DNAm levels at multiple CpG sites of their genome (Horvath & Raj, 2018). There is now clear evidence of the association between alterations in the predictions made by epigenetic clocks and lifestyle factors, disease—including cancer—or outright mortality (Fransquet et al., 2019; Levine et al., 2018; Lu et al., 2019; Marioni et al., 2015; Oblak et al., 2021), indicating that these algorithms capture a combination of chronological and biological age, the latter being a measure of the "healthiness" of an individual in terms of their risk of developing age-associated adverse outcomes (Jylhävä et al., 2017). For instance, supercentenarian or long-lived subjects, who display reduced incidence, or delayed onset, of diseases (Andersen et al., 2012) also manifest younger epigenetic ages (Armstrong et al., 2017; Horvath et al., 2015). Moreover, there is incipient evidence of epigenetic clock rejuvenation through pharmacological- or lifestyle-based clinical interventions in human (Fahy et al., 2019;Gensous et al., 2020; Noreen et al., 2020; Fiorito et al., 2021; Fitzgerald et al., 2021). Thus, current data support the importance of epigenetic clocks as biomarkers of aging, although studies focusing on the direct modification of clock components will be necessary to demonstrate whether these sets of CpGs can directly modulate the aging process. That said, because these clocks seem to so accurately capture information related to internal or environmental aggression which may be associated with disease, their close examination may help clarify the putative causal relationships between aging and cancer.

3.1 | Functional meaning of the epigenetic clocks and somatic stem cell aging

What—if any—is the functional meaning of epigenetic clocks? Recent reports have indicated that the CpG sites which make up different clocks often share core functional characteristics in terms of their association with genomic elements or with gene expression levels (Jonkman et al., 2022; Liu et al., 2020), and *in vitro* models have been built that partially link cellular passage clocks with organismal aging (Minteer et al., 2022), supporting the notion that these algorithms are not merely abstract constructs but rather reflect underlying biological processes. Indeed, the intrinsic characteristics of DNAm clocks in relation to their differing "ticking rates" during lifespan, which are accelerated during development and subsequently slow down during aging, (Horvath & Raj, 2018) suggest that they may be particularly linked to an underlying biological system. Stem cells display a youthful—pre-natal in fact—epigenetic phenotype (Horvath, 2013) and there are experimental observations indicating that the rejuvenation of the epigenetic clock occurs during embryonic development (Kerepesi et al., 2021) and can be achieved through the transient expression of reprogramming factors in human and mouse models (Chondronasiou et al., 2022; Gill et al., 2022; Sarkar et al., 2020). Thus, the ticking rate of the epigenetic clocks could in fact reflect those developmental and tissue-homeostasis maintenance processes driven by stem cells—and involving alterations in cell composition—that occur throughout life, processes, which can in turn be modified by internal or external factors during aging (Horvath & Raj, 2018; Raj & Horvath, 2020). A recent review summarized three hypotheses regarding the biological processes which could be responsible for the ticking of the clock CpGs whereby the gradual accumulation of DNAm alterations could reflect (Seale et al., 2022): (1) the life-long tissue turnover via asymmetric stem cell division; (2) the decay of circadian mechanisms during aging, which could be linked to metabolic dysregulation; or (3) the genomic re-localization of epigenetic enzymes to DNA damage sites, which appear and accumulate with aging. Nevertheless, these systemic views of epigenetic clocks need to be reconciled with recent evidence demonstrating that epigenetic aging can be measured in individual cells (Trapp et al., 2021). This new clock, termed "scAge", was actually built from bulk-tissue derived aging CpGs, and thus, fits well with models of aging-associated increases in tissue heterogeneity (Rudolph, 2021), but the development of higher-resolution DNAm profiling in single cells will allow for the construction of epigenetic predictors directly from these data. After all, DNAm is a discrete/binary molecular mark, and it remains to be explored whether there is a mechanistic basis to how, or in what order, CpG sites may be altered with age in individual cells.

In addition, the relationship between epigenetic aging and "stemness", or plasticity, must be more clearly defined. Adult stem cells, which reside in tissue-specific niches and are responsible for the maintenance of tissue homeostasis (Cheung & Rando, 2013), show evident signs of aging-associated decline (Oh et al., 2014), including epigenetic alterations (Bork et al., 2010; Fernández et al., 2015). A recent report that studied hematopoietic stem cell (HSC) transplantation has shown that transplanted HSCs give rise to reconstituted blood which manifests a DNAm age similar to that of the HSC donor (Søraas et al., 2019), suggesting that the initial stem cells carry a specific age signature. Thus, the question of whether somatic stem cells display epigenetic clock aging still needs to be clarified. To tackle this issue, we compiled DNAm data from various independent studies profiling different types of somatic stem cells and progenitors (see Methods for full list of datasets) and determined their epigenetic ages using the Horvath clock (Horvath, 2013). Our results clearly show that somatic stem cells display DNAm clock aging across a

wide range of tissues (Figure 1 a-b), as opposed to embryonic stem cells and fetal tissues. In an epigenetic clock model, it makes sense that aging somatic stem cells accumulate a DNAm age signature which is then propagated to tissues during homeostasis and regeneration. Studies that dissect somatic stem cells at the single-cell level in particular are needed to clarify this issue, and there is indeed an example of muscle stem cells showing low epigenetic ages in mouse (Hernando-Herraez et al., 2019), but more data is clearly required.

3.2 | Alterations in epigenetic clocks and the epigenetic age of tumors

With regards to the alterations—acceleration or deceleration—observed in the clocks, various associations have been investigated, particularly the aforementioned links with environmental factors and disease (Fransquet et al., 2019; Levine et al., 2018; Lu et al., 2019; Marioni et al., 2015; Oblak et al., 2021), to ascertain whether the DNAm alterations which occur during these processes could drive the changes observed in the clocks. For instance, a recent study investigating DNAm age acceleration in the non-tumoral breast tissue of breast cancer patients demonstrated that the observed alteration in the epigenetic clock could be explained by a subset of the clock CpGs that suffer from the well-known cancer-associated hypermethylation of Polycomb associated loci (Rozenblit et al., 2022). Indeed, clocks have been constructed which partly reflect the DNAm alterations induced by cancer-related lifestyle factors such as smoking (Lu et al., 2019). There are also genetic factors, some of which also regulate the biology of the aging process, which are associated with the ticking rate of epigenetic clocks (Lu et al., 2018; McCartney et al., 2021). However, a recent report using Mendelian randomization methods to explore the possible causal role of epigenetic clocks in cancer development has demonstrated that there were few causal associations between the ticking rates of these clocks and increased cancer risk (Morales Berstein et al., 2022). Thus, at this stage there is little evidence to suggest that epigenetic clocks are primary drivers of aging or age-related disease as opposed to being biomarkers that reflect the influence of genetic or environmental factors, or other underlying processes

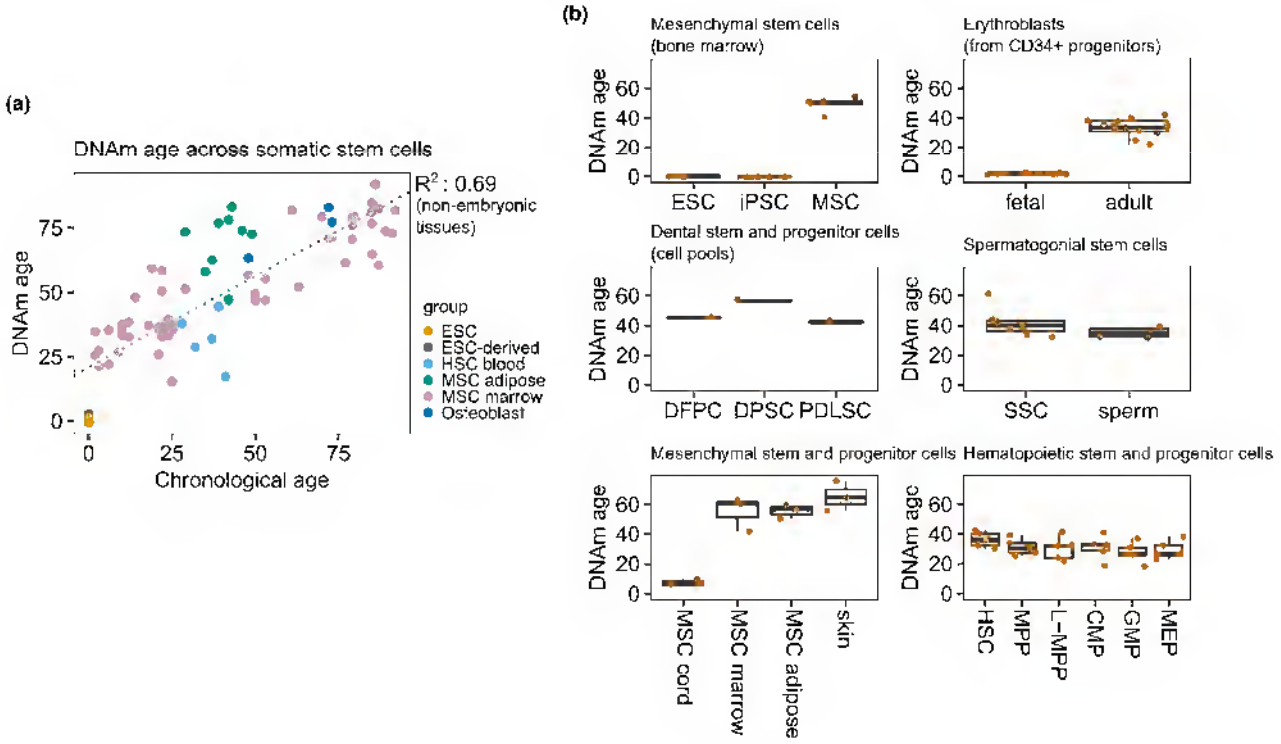


FIGURE 1 Somatic stem cells and progenitors display epigenetic aging. (a) Scatterplot depicting the correlation between chronological and epigenetic age (Horvath clock) across various types of somatic stem cells and progenitors, embryonic stem cells, and derived embryonic tissues. The R-squared coefficient and line of fit are given for the correlation involving non-embryonic tissues (HSC: hematopoietic stem cell, from peripheral blood; MSC: mesenchymal stem cell, from adipose or bone marrow tissue; ESC: embryonic stem cell). (b) Boxplots showing the epigenetic age estimations of different types of somatic stem cells and progenitors for which measurements of chronological age were unavailable (ESC: embryonic stem cell; iPSC: induced pluripotent stem cell; MSC: mesenchymal stem cell; DFPC: dental follicle progenitor cell; DPSC: dental pulp stem cell; PDLSC: periodontal ligament stem cell; SSC: spermatogonial stem cell; HSC: hematopoietic stem cell; MPP: multipotent progenitor; L-MPP: late multipotent progenitor; CMP: common myeloid progenitor; GMP: granulocyte-macrophage progenitor; MEP: megakaryocyte-erythroid progenitor).

(Drew, 2022). Nonetheless, it may also be that some form of cancer-specific clock dysregulation with functional meaning occurs directly during the process of oncogenic transformation, meaning that the tracking of age acceleration in healthy tissues, as done in most studies, would not yield valuable insight.

In addition to the aforementioned findings, we should consider the following observations (Figure 2): (1) the classic aging-associated DNAm alterations described in studies are very different from clock-associated DNAm changes with aging: clock-DNAm changes with age are actually very subtle (Horvath, 2013) and the CpG sites involved could in fact be considered "stable enough" to be able to track time-dependent alterations in spite of internal or external assaults occurring through lifespan, while aging-DNAm changes are much stronger and specific in terms of their biological associations (Day et al., 2013; Fernández et al., 2015; Pérez et al., 2018), and, moreover, there are even aging-DNAm signatures which accelerate with age (Sziráki et al., 2018); (2) as already mentioned, aging DNAm signatures—especially DNA hypermethylation—are typically augmented in cancer (Lin & Wagner, 2015; Luebeck et al., 2019) to the extent that, in this sense, tumors may appear to have an aged phenotype.

On the other hand, we know little regarding the specific reconfiguration of the clock DNAm within tumoral cells: they appear to also have an hyperaged phenotype (Horvath, 2013; Pérez et al., 2018) but this could be related to cancer-associated dramatic hypermethylation events affecting a subset of the clock CpGs. In fact, tumors such as thyroid cancer, for which age is actually a prognosis indicator (Kazaure et al., 2018), show reduced clock-DNAm dysregulation—i.e. they age more "naturally" (Horvath, 2013; Pérez et al., 2018; Yang et al., 2016). It is possible that the age acceleration observed in tumors may partly be caused by epigenetic clocks capturing the mitotic, proliferative history of tissues, which is evidently amplified in cancer. This issue has been tackled by comparing the behavior of a mitotic clock and the Horvath clock in the context of B-cell tumors to show that epigenetic age can sometimes be independent of epigenetic proliferative history (Duran-Ferrer et al., 2020), thus suggesting that there may be mechanisms underlying epigenetic clocks which are not entirely connected with cellular division. Furthermore, if we acknowledge that tumoral cells—and not exclusively cancer stem cells—have an increased quality of stemness, or plasticity, which increases their proliferation and dissemination potential (Batlle & Clevers, 2017), should we expect them to have a higher or lower epigenetic age? There is evidence that some stem-related traits such as telomerase expression do not lead to the rejuvenation of the epigenetic clock (Lu et al., 2018), while, as mentioned previously, the induction of pluripotent reprogramming factors does (Chondronasiou et al., 2022; Gill et al., 2022; Sarkar et al., 2020).

Altogether, it seems that aging-DNAm and clock-DNAm dysregulation present distinct molecular features, and more studies, again particularly those focusing on the single-cell dissection of epigenetic aging in healthy and tumoral tissues, will be needed to clarify these issues.

4 | FUTURE DIRECTIONS: GENETIC AND EPIGENETIC AGING

It is becoming increasingly accepted, in what is known as the "somatic mutation" theory of aging (Kennedy et al., 2012), that the unifying mechanism of aging is the time-dependent accumulation of DNA damage (Schumacher et al., 2021). Indeed, there are recent compelling results suggesting that mutational rates may define lifespan across mammalian species (Cagan et al., 2022), and that this could be partly controlled by differences in the efficiency of DNA repair mechanisms (Gorbunova et al., 2014; Tian et al., 2019). In this same line, recent observations indicate that aging-associated epigenetic dysregulation—including epigenetic clock acceleration—could be a consequence of DNA damage via the re-localization of epigenetic modifiers during DNA repair (Hayano et al., 2019; Kane & Sinclair, 2019). This would suggest that features of epigenetic aging may be brought about by the accumulation of aging-associated DNA damage and, indeed, the stochastic accumulation of genomic alterations is consistent with the epigenetic noise typically observed during the aging process (Tejedor & Fraga, 2017). The recent development of multispecies epigenetic profiling technologies (Arneson et al., 2022) is starting to provide insight into these questions. New studies conducted by the Mammalian Methylation Consortium have provided very interesting results that support the notion that aging mechanisms are evolutionarily conserved: (1) DNA methylation clocks have been successfully built for the prediction of age across mammalian species that may be associated with developmental processes (Lu et al., 2021); (2) epigenetic markers of maximum lifespan have also been developed which are quite different from aging-associated DNAm alterations (Li et al., 2021), although new analytical approaches can be used to define CpG sites that integrate information from both aging and lifespan, which may be particularly useful as biomarkers of anti-aging interventions (Haghani et al., 2022).

Results from epigenetic studies often point towards the existence of development-associated mechanisms underlying the universal features of aging; however, they will have to be reconciled with the existing genetic evidence positing that the accrual of DNA damage is the main driver of this process. Furthermore, with regards to the links between aging mechanisms and the development of cancer, they will have to be examined in the light of recent powerful evidence indicating that, while cancer is also a universal disease across mammals, differences in cancer risk across species are largely independent of longevity—and body mass—and could be related to lifestyle factors such as diet (Vincze et al., 2022). Thus, epigenetic mechanisms may still hold the key to revealing how universal biological processes such as aging or cancer— which may in themselves

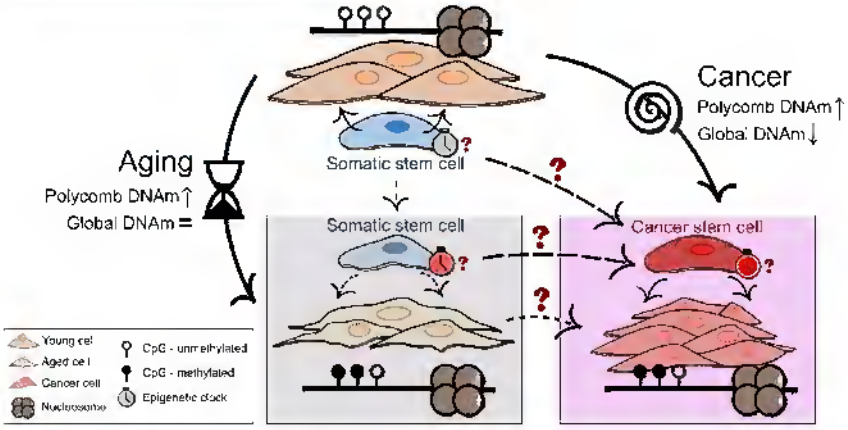


FIGURE 2 Unresolved questions regarding the links between aging and cancer DNAm alterations. There are several loose ends with regards to how DNAm alterations may constitute mechanistic links between aging and cancer. The classic aging-DNAm alterations include Polycomb-associated hypermethylation, which also occurs in cancer, but the similarities with respect to hypomethylation are less clear as tumors display much clearer global hypomethylation, which could be related to their drastic expansion, while aging-associated loss of DNAm is linked to different chromatin signatures. On the other hand, epigenetic clocks seem to be accelerated in cancer in spite of the increased cellular plasticity and activity of cancer cells. It remains to be explained if and how somatic stem cells display epigenetic clock aging, and how this phenomenon may influence cancer development, in particular with regards to cancer stem cells.

be reflected through DNAm biomarkers—are modulated by environmental, non-genetic factors.

5 | METHODS/DATASETS

To explore the dynamics of DNAm aging in somatic stem cells, we retrieved DNAm values in the form of beta values or intensity measurements from 13 publicly available array-based datasets: GSE52114 and GSE17448, containing mesenchymal stem cells (MSCs) from bone marrow (Bork et al., 2010; Fernández et al., 2015); E-MTAB487, containing CD34+hematopoietic stem cells (HSCs) from peripheral blood (Bocker et al., 2011) (reanalyzed samples of MSCs from GSE17448 were removed); GSE202067, containing MSCs and osteoblasts from bone marrow (Giesche et al., 2022); GSE138311, containing MSCs from adipose tissue (Serena et al., 2020) (only healthy subjects were used); GSE26519, containing MSCs from adipose tissue (Schellenberg et al., 2011; 5-passage samples were used while redundant 10passage samples from the same individuals were removed); GSE116754, containing embryonic stem cells (ESCs) and derived embryonic tissue (Colunga et al., 2019); GSE34688, containing MSCs from bone marrow, derived induced pluripotent stem cells (iPSCs) and ESCs (Shao et al., 2013); GSE56491, containing erythroblasts derived from CD34+progenitors from fetal liver and adult bone marrow (Lessard et al., 2015); GSE112933, containing periodontal stem cells (pooled from various patients); GSE72444, containing spermatogonial stem cells and sperm samples (Struijk et al., 2020); GSE41933, containing MSCs from adipose tissue, bone marrow, umbilical cord, and dermal fibroblasts (Reinisch et al., 2015); and GSE63409, containing various hematopoietic progenitors (Jung et al., 2015).

Data were handled with the R statistical software (v4.0.5) and graphs were constructed using the *ggplot2* package (v3.3.3; Wickham, 2016). DNAm ages were estimated using the Horvath clock (Horvath, 2013) via the *ENmix* package (v1.26.10; Xu et al., 2021).

AUTHORS CONTRIBUTION

Raúl Fernández Pérez, Juan Ramón Tejedor, Agustín Fernández Fernández, and Mario Fernández Fraga conceived, coordinated, and supervised the study. Raúl Fernández Pérez designed all aspects of the research, collected the data, performed analyses, and wrote the manuscript. All authors revised, read, and approved the final manuscript.

ACKNOWLEDGEMENTS

We apologize to all authors whose valuable work could not be cited due to length constraints. We thank Ronnie Lendrum for her manuscript editing and the members of the Cancer Epigenetics Laboratory for their positive feedback.

FUNDING INFORMATION

The Spanish Association Against Cancer (grant number PROYE18061FERN to MFF), the Asturias Government (PCTI) co-funding 2018-2022/FEDER (grant number IDI/2018/146 to MFF), Fundación General CSIC (grant number 0348_CIE_6_E to MFF), the Health Institute Carlos III (Plan Nacional de I+D+I) co-funding FEDER (grant number PI21/01067 to MFF and AFF) and the Spanish Ministry of Science and Innovation (grant number SGL2021-03-039/40 to M.F.F.) co-funding NextGenerationEU. JRT is supported by a Juan de la Cierva fellowship from the Spanish Ministry

of Science and Innovation (grant number IJC2018-36825-I). RFP is supported by the Severo Ochoa program (grant number BP17-114). We also acknowledge support from IUOPA-ISPA-FINBA (the IUOPA is supported by the Obra Social Cajastur-Liberbank, Spain) and Consorcio Centro de Investigación Biomédica en Red (CIBERER-ISCIII).

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Raúl Fernández Pérez https://orcid.org/0000-0003-4336-9898
Juan Ramón Tejedor https://orcid.org/0000-0002-4061-9698
Agustín Fernández Fernández https://orcid.org/0000-0002-3792-4085
Mario Fernández Fraga https://orcid.org/0000-0001-8450-2603

REFERENCES

Andersen, S. L., Sebastiani, P., Dworkis, D. A., Feldman, L., & Perls, T. T. (2012). Health span approximates life span among many supercentenarians: compression of morbidity at the approximate limit of life span. *Journals of Gerontology Series A-Biological Sciences and Medical Sciences*, *67*(4), 395–405. https://doi.org/10.1093/gerona/glr223

Armstrong, N. J., Mather, K. A., Thalamuthu, A., Wright, M. J., Trollor, J. N., Ames, D., Brodaty, H., Schofield, P. R., Sachdev, P. S., & Kwok, J. B. (2017). Aging, exceptional longevity and comparisons of the Hannum and Horvath epigenetic clocks. *Epigenomics*, *9*(5), 689–700. https://doi.org/10.2217/epi-2016-0179

Arneson, A., Haghani, A., Thompson, M. J., Pellegrini, M., Kwon, S. B., Vu, H., Maciejewski, E., Yao, M., Li, C. Z., Lu, A. T., Morselli, M., Rubbi, L., Barnes, B., Hansen, K. D., Zhou, W., Breeze, C. E., Ernst, J., & Horvath, S. (2022). A mammalian methylation array for profiling methylation levels at conserved sequences. *Nature Communications*, *13*(1), 783. https://doi.org/10.1038/s41467-022-28355-z

Batlle, E., & Clevers, H. (2017). Cancer stem cells revisited. *Nature Medicine*, *23*(10), 1124–1134. https://doi.org/10.1038/nm.4409

Belinsky, S. A., Palmisano, W. A., Gilliland, F. D., Crooks, L. A., Divine, K. K., Winters, S. A., Grimes, M. J., Harms, H. J., Tellez, C. S., Smith, T. M., Moots, P. P., Lechner, J. F., Stidley, C. A., & Crowell, R. E. (2002). Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer research*, *62*(8), 2370–2377.

Bocker, M. T., Hellwig, I., Breiling, A., Eckstein, V., Ho, A. D., & Lyko, F. (2011). Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging. *Blood*, *117*(19), e182–e189. https://doi.org/10.1182/blood-2011-01-331926

Bollati, V., Schwartz, J., Wright, R., Litonjua, A., Tarantini, L., Suh, H., Sparrow, D., Vokonas, P., & Baccarelli, A. (2009). Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mechanisms of Ageing and Development*, *130*(4), 234–239. https://doi.org/10.1016/j.mad.2008.12.003

Bork, S., Pfister, S., Witt, H., Horn, P., Korn, B., Ho, A. D., & Wagner, W. (2010). DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. *Aging Cell*, *9*, 54–63.

Cagan, A., Baez-Ortega, A., Brzozowska, N., Abascal, F., Coorens, T. H. H., Sanders, M. A., Lawson, A. R. J., Harvey, L. M. R., Bhosle, S., Jones, D., Alcantara, R. E., Butler, T. M., Hooks, Y., Roberts, K., Anderson, E., Lunn, S., Flach, E., Spiro, S., Januszczak, I., … Martincorena, I. (2022). Somatic mutation rates scale with lifespan across mammals. *Nature*, *604*, 517–524.

Campagna, M. P., Xavier, A., Lechner-Scott, J., Maltby, V., Scott, R. J., Butzkueven, H., Jokubaitis, V. G., & Lea, R. A. (2021). Epigenome-wide association studies: current knowledge, strategies and recommendations. *Clinical epigenetics*, *13*, 214.

Capp, J.-P., & Thomas, F. (2021). Tissue-disruption-induced cellular stochasticity and epigenetic drift: Common origins of aging and cancer? *Bioessays*, *43*, e2000140.

McCartney, D. L., Min, J. L., Richmond, R. C., Lu, A. T., Sobczyk, M. K., Davies, G., Broer, L., Guo, X., Jeong, A., Jung, J., Kasela, S., Katrinli, S., Kuo, P.-L., Matias-Garcia, P. R., Mishra, P. P., Nygaard, M., Palviainen, T., Patki, A., Raffield, L. M., … Marioni, R. E. (2021). Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. *Genome biology*, *22*, 194.

Chen, Y., Breeze, C. E., Zhen, S., Beck, S., & Teschendorff, A. E. (2016). Tissue-independent and tissue-specific patterns of DNA methylation alteration in cancer. *Epigenetics Chromatin*, *9*, 10.

Cheung, P., Vallania, F., Warsinske, H. C., Donato, M., Schaffert, S., Chang, S. E., Dvorak, M., Dekker, C. L., Davis, M. M., Utz, P. J., Khatri, P., & Kuo, A. J. (2018). Single-Cell Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging. *Cell*, *173*, 1385–1397.e14.

Cheung, T. H., & Rando, T. A. (2013). Molecular regulation of stem cell quiescence. *Nature Reviews Molecular Cell Biology*, *14*, 329–340.

Chondronasiou, D., Gill, D., Mosteiro, L., Urdinguio, R. G., Berenguer-Llergo, A., Aguilera, M., Durand, S., Aprahamian, F., Nirmalathasan, N., Abad, M., Martin-Herranz, D. E., Stephan-Otto Attolini, C., Prats, N., Kroemer, G., Fraga, M. F., Reik, W., & Serrano, M. (2022). Multi-omic rejuvenation of naturally aged tissues by a single cycle of transient reprogramming. *Aging Cell*, *21*, e13578.

Cole, J. J., Robertson, N. A., Rather, M. I., Thomson, J. P., McBryan, T., Sproul, D., Wang, T., Brock, C., Clark, W., Ideker, T., Meehan, R. R., Miller, R. A., Brown-Borg, H. M., & Adams, P. D. (2017). Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome biology*, *18*, 58.

Colunga, T., Hayworth, M., Kreß, S., Reynolds, D. M., Chen, L., Nazor, K. L., Baur, J., Singh, A. M., Loring, J. F., Metzger, M., & Dalton, S. (2019). Human Pluripotent Stem Cell-Derived Multipotent Vascular Progenitors of the Mesothelium Lineage Have Utility in Tissue Engineering and Repair. *Cell Reports*, *26*, 25662579.e10.

Day, K., Waite, L. L., Thalacker-Mercer, A., West, A., Bamman, M. M., Brooks, J. D., Myers, R. M., & Absher, D. (2013). Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome biology*, *14*, R102.

de Magalhães, J. P. (2013). How ageing processes influence cancer. *Nature Reviews Cancer*, *13*, 357–365.

Dmitrijeva, M., Ossowski, S., Serrano, L., & Schaefer, M. H. (2018). Tissue-specific DNA methylation loss during ageing and carcinogenesis is linked to chromosome structure, replication timing and cell division rates. *Nucleic Acids Research*, *46*, 7022–7039.

Drew, L. (2022). Turning back time with epigenetic clocks. *Nature*, *601*, S20–S22.

Duran-Ferrer, M., Clot, G., Nadeu, F., Beekman, R., Baumann, T., Nordlund, J., Marincevic-Zuniga, Y., Lönnerholm, G., Rivas-Delgado, A., Martín, S., Ordoñez, R., Castellano, G., Kulis, M., Queirós, A. C., Lee, S.-T., Wiemels, J., Royo, R., Puiggrós, M., Lu, J., … MartínSubero, J. I. (2020). The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome. *Nature Cancer*, *1*, 1066–1081.

Easwaran, H., Johnstone, S. E., Van Neste, L., Ohm, J., Mosbruger, T., Wang, Q., Aryee, M. J., Joyce, P., Ahuja, N., Weisenberger, D., Collisson, E., Zhu, J., Yegnasubramanian, S., Matsui, W., & Baylin, S. B. (2012). A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome research*, *22*, 837–849.

Ehrlich, M. (2009). DNA hypomethylation in cancer cells. *Epigenomics*, *1*, 239–259.

Fahy, G. M., Brooke, R. T., Watson, J. P., Good, Z., Vasanawala, S. S., Maecker, H., Leipold, M. D., Lin, D. T. S., Kobor, M. S., & Horvath, S. (2019). Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell*, *18*, e13028.

Fernández, A. F., Bayón, G. F., Urdinguio, R. G., Toraño, E. G., García, M. G., Carella, A., Petrus-Reurer, S., Ferrero, C., Martinez-Camblor, P., Cubillo, I., García-Castro, J., Delgado-Calle, J., Pérez-Campo, F. M., Riancho, J. A., Bueno, C., Menéndez, P., Mentink, A., Mareschi, K., Claire, F., … Fraga, M. F. (2015). H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells. *Genome research*, *25*, 27–40.

Fiorito, G., Caini, S., Palli, D., Bendinelli, B., Saieva, C., Ermini, I., Valentini, V., Assedi, M., Rizzolo, P., Ambrogetti, D., Ottini, L., & Masala, G. (2021). DNA methylation-based biomarkers of aging were slowed down in a two-year diet and physical activity intervention trial: the DAMA study. *Aging Cell*, *20*, e13439.

Fitzgerald, K. N., Hodges, R., Hanes, D., Stack, E., Cheishvili, D., Szyf, M., Henkel, J., Twedt, M. W., Giannopoulou, D., Herdell, J., Logan, S., & Bradley, R. (2021). Potential reversal of epigenetic age using a diet and lifestyle intervention: a pilot randomized clinical trial. *Aging (Albany NY)*, *13*, 9419–9432.

Fransquet, P. D., Wrigglesworth, J., Woods, R. L., Ernst, M. E., & Ryan, J. (2019). The epigenetic clock as a predictor of disease and mortality risk: a systematic review and meta-analysis. *Clinical Epigenetics*, *11*, 62.

Fuke, C., Shimabukuro, M., Petronis, A., Sugimoto, J., Oda, T., Miura, K., Miyazaki, T., Ogura, C., Okazaki, Y., & Jinno, Y. (2004). Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Annals of human genetics*, *68*, 196–204.

Gensous, N., Garagnani, P., Santoro, A., Giuliani, C., Ostan, R., Fabbri, C., Milazzo, M., Gentilini, D., di Blasio, A. M., Pietruszka, B., Madej, D., Bialecka-Debek, A., Brzozowska, A., Franceschi, C., & Bacalini, M. G. (2020). One-year Mediterranean diet promotes epigenetic rejuvenation with country- and sex-specific effects: a pilot study from the NU-AGE project. *Geroscience*, *42*, 687–701.

Giesche, J., Mellert, K., Geißler, S., Arndt, S., Seeling, C., von Baer, A., Schultheiss, M., Marienfeld, R., Möller, P., & Barth, T. F. (2022). Epigenetic lockdown of CDKN1A (p21) and CDKN2A (p16) characterises the neoplastic spindle cell component of giant cell tumours of bone. *The Journal of pathology*, *257*(5), 687–696.

Gill, D., Parry, A., Santos, F., Okkenhaug, H., Todd, C. D., Hernando-Herraez, I., Stubbs, T. M., Milagre, I., & Reik, W. (2022). Multi-omic rejuvenation of human cells by maturation phase transient reprogramming. *Elife*, *11*, e71624.

Gorbunova, V., Seluanov, A., Zhang, Z., Gladyshev, V. N., & Vijg, J. (2014). Comparative genetics of longevity and cancer: insights from long-lived rodents. *Nature Reviews Genetics*, *15*, 531–540.

Hadad, N., Masser, D. R., Blanco-Berdugo, L., Stanford, D. R., & Freeman, W. M. (2019). Early-life DNA methylation profiles are indicative of age-related transcriptome changes. *Epigenetics & Chromatin*, *12*, 58.

Haghani, A., Wang, N., Lu, A. T., Mozhui, K., Consortium, M. M., Raj, K., Yang, X. W., & Horvath, S. (2022). Divergent age-related methylation patterns in long and short-lived mammals. *bioRxiv*, 476530. https://doi.org/10.1101/2022.01.16.476530v1 Accessed June 16, 2022.

Hahn, O., Grönke, S., Stubbs, T. M., Ficz, G., Hendrich, O., Krueger, F., Andrews, S., Zhang, Q., Wakelam, M. J., Beyer, A., Reik, W., & Partridge, L. (2017). Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. *Genome biology*, *18*, 56.

Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology*, *18*, 83.

Hayano, M., Yang, J.-H., Bonkowski, M. S., Amorim, J. A., Ross, J. M., Coppotelli, G., Griffin, P. T., Chew, Y. C., Guo, W., Yang, X., Vera, D. L., Salfati, E. L., Das, A., Thakur, S., Kane, A. E., Mitchell, S. J., Mohri, Y., Nishimura, E. K., Schaevitz, L., … Sinclair, D. A. (2019). DNA Break-Induced Epigenetic Drift as a Cause of Mammalian Aging. *bio Rxiv*, 808659. Available at:. https://doi.org/10.1101/808659v1

Hernando-Herraez, I., Evano, B., Stubbs, T., Commere, P.-H., Jan Bonder, M., Clark, S., Andrews, S., Tajbakhsh, S., & Reik, W. (2019). Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nature Communications*, *10*, 4361.

Heyn, H., Li, N., Ferreira, H. J., Moran, S., Pisano, D. G., Gomez, A., Diez, J., Sanchez-Mut, J. V., Setien, F., Carmona, F. J., Puca, A. A., Sayols, S., Pujana, M. A., Serra-Musach, J., Iglesias-Platas, I., Formiga, F., Fernandez, A. F., Fraga, M. F., Heath, S. C., … Esteller, M. (2012). Distinct DNA methylomes of newborns and centenarians. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 10522–10527.

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

Horvath, S., Pirazzini, C., Bacalini, M. G., Gentilini, D., Di Blasio, A. M., Delledonne, M., Mari, D., Arosio, B., Monti, D., Passarino, G., De Rango, F., D'Aquila, P., Giuliani, C., Marasco, E., Collino, S., Descombes, P., Garagnani, P., & Franceschi, C. (2015). Decreased epigenetic age of PBMCs from Italian semisupercentenarians and their offspring. *Aging (Albany NY)*, *7*, 1159–1170.

Horvath, S., & Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, *19*, 371–384.

Jonkman, T. H., Dekkers, K. F., Slieker, R. C., Grant, C. D., Ikram, M. A., van Greevenbroek, M. M. J., Franke, L., Veldink, J. H., Boomsma, D. I., Slagboom, P. E., BIOS, C., & Heijmans, B. T. (2022). Functional genomics analysis identifies T and NK cell activation as a driver of epigenetic clock progression. *Genome Biology*, *23*, 24.

Jung, N., Dai, B., Gentles, A. J., Majeti, R., & Feinberg, A. P. (2015). An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nature Communications*, *6*, 8489.

Jylhävä, J., Pedersen, N. L., & Hägg, S. (2017). Biological Age Predictors. *EBioMedicine*, *21*, 29–36.

Kane, A. E., & Sinclair, D. A. (2019). Epigenetic changes during aging and their reprogramming potential. *Critical Reviews in Biochemistry and Molecular Biology*, *54*, 61–83.

Kazaure, H. S., Roman, S. A., & Sosa, J. A. (2018). The impact of age on thyroid cancer staging. *Current Opinion in Endocrinology, Diabetes and Obesity*, *25*, 330–334.

Kennedy, S. R., Loeb, L. A., & Herr, A. J. (2012). Somatic mutations in aging, cancer and neurodegeneration. *Mechanisms of Ageing and Development*, *133*, 118–126.

Kerepesi, C., Zhang, B., Lee, S.-G., Trapp, A., & Gladyshev, V. N. (2021). Epigenetic clocks reveal a rejuvenation event during embryogenesis followed by aging. *Science Advances*, *7*, eabg6082.

Klutstein, M., Moss, J., Kaplan, T., & Cedar, H. (2017). Contribution of epigenetic mechanisms to variation in cancer risk among tissues. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 2230–2234.

Lessard, S., Beaudoin, M., Benkirane, K., & Lettre, G. (2015). Comparison of DNA methylation profiles in human fetal and adult red blood cell progenitors. *Genome medicine*, *7*, 1.

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., Hou, L., Baccarelli, A. A., Stewart, J. D., Li, Y., Whitsel, E. A., Wilson, J. G., Reiner, A. P., Aviv, A., Lohman, K., Liu, Y., Ferrucci, L.,

& Horvath, S. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*, *10*, 573–591.

Li, C. Z., Haghani, A., Robeck, T. R., Villar, D., Lu, A. T., Zhang, J., Faulkes, C. G., Vu, H., Ablaeva, J., Adams, D. M., Ardehali, R., Arneson, A., Baker, C. S., Belov, K., Blumstein, D. T., Bors, E. K., Breeze, C. E., Brooke, R. T., Brown, J. L., … Consortium, M. M. (2021). Epigenetic predictors of maximum lifespan and other life history traits in mammals. *bioRxiv*, 444078. https://doi.org/10.1101/2021.05.16.444078v1 Accessed June 16, 2022.

Li, J., Huang, Q., Zeng, F., Li, W., He, Z., Chen, W., Zhu, W., & Zhang, B. (2014). The prognostic value of global DNA hypomethylation in cancer: a meta-analysis. *PLoS One*, *9*, e106290.

Lin, Q., & Wagner, W. (2015). Epigenetic Aging Signatures Are Coherently Modified in Cancer. *PLoS Genetics*, *11*, e1005334.

Liu, Z., Leung, D., Thrush, K., Zhao, W., Ratliff, S., Tanaka, T., Schmitz, L. L., Smith, J. A., Ferrucci, L., & Levine, M. E. (2020). Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging Cell*, *19*, e13229.

Lu, A. T., Fei, Z., Haghani, A., Robeck, T. R., Zoller, J. A., Li, C. Z., Zhang, J., Ablaeva, J., Adams, D. M., Almunia, J., Ardehali, R., Arneson, A., Baker, C. S., Belov, K., Black, P., Blumstein, D. T., Bors, E. K., Breeze, C. E., Brooke, R. T., … Horvath, S. (2021). Universal DNA methylation age across mammalian tissues. *bioRxiv*, 426733. https://doi.org/10.1101/2021.01.18.426733v1 Accessed June 16, 2022.

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., Whitsel, E. A., Assimes, T. L., Ferrucci, L., & Horvath, S. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*, *11*, 303–327.

Lu, A. T., Xue, L., Salfati, E. L., Chen, B. H., Ferrucci, L., Levy, D., Joehanes, R., Murabito, J. M., Kiel, D. P., Tsai, P. C., Yet, I., Bell, J. T., Mangino, M., Tanaka, T., McRae, A. F., Marioni, R. E., Visscher, P. M., Wray, N. R., Deary, I. J., … Horvath, S. (2018). GWAS of epigenetic aging rates in blood reveals a critical role for TERT. *Nature Communications*, *9*, 387.

Luebeck, G. E., Hazelton, W. D., Curtius, K., Maden, S. K., Yu, M., Carter, K. T., Burke, W., Lampe, P. D., Li, C. I., Ulrich, C. M., Newcomb, P. A., Westerhoff, M., Kaz, A. M., Luo, Y., Inadomi, J. M., & Grady, W. M. (2019). Implications of Epigenetic Drift in Colorectal Neoplasia. *Cancer Research*, *79*, 495–504.

Marioni, R. E., Shah, S., McRae, A. F., Chen, B. H., Colicino, E., Harris, S. E., Gibson, J., Henders, A. K., Redmond, P., Cox, S. R., Pattie, A., Corley, J., Murphy, L., Martin, N. G., Montgomery, G. W., Feinberg, A. P., Fallin, M. D., Multhaup, M. L., Jaffe, A. E., … Deary, I. J. (2015). DNA methylation age of blood predicts all-cause mortality in later life. *Genome biology*, *16*, 25.

Marks, D. L., Olson, R. L., & Fernandez-Zapico, M. E. (2016). Epigenetic control of the tumor microenvironment. *Epigenomics*, *8*, 1671–1687.

Masser, D. R., Hadad, N., Porter, H. L., Mangold, C. A., Unnikrishnan, A., Ford, M. M., Giles, C. B., Georgescu, C., Dozmorov, M. G., Wren, J. D., Richardson, A., Stanford, D. R., & Freeman, W. M. (2017). Sexually divergent DNA methylation patterns with hippocampal aging. *Aging Cell*, *16*, 1342–1352.

Minteer, C., Morselli, M., Meer, M., Cao, J., Higgins-Chen, A., Lang, S. M., Pellegrini, M., Yan, Q., & Levine, M. E. (2022). Tick tock, tick tock: Mouse culture and tissue aging captured by an epigenetic clock. *Aging Cell*, *21*, e13553.

Morales Berstein, F., McCartney, D. L., Lu, A. T., Tsilidis, K. K., Bouras, E., Haycock, P., Burrows, K., Phipps, A. I., Buchanan, D. D., Cheng, I., Practical consortium, Martin, R. M., Davey Smith, G., Relton, C. L., Horvath, S., Marioni, R. E., Richardson, T. G., & Richmond, R. C. (2022). Assessing the causal role of epigenetic clocks in the development of multiple cancers: a Mendelian randomization study. *Elife*, *11*, e75374.

Noreen, F., Chaber-Ciopinska, A., Regula, J., Schär, P., & Truninger, K. (2020). Longitudinal analysis of healthy colon establishes aspirin as a suppressor of cancer-related epigenetic aging. *Clinical Epigenetics*, *12*, 164.

Oblak, L., van der Zaag, J., Higgins-Chen, A. T., Levine, M. E., & Boks, M. P. (2021). A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. *Ageing research reviews*, *69*, 101348.

Oh, J., Lee, Y. D., & Wagers, A. J. (2014). Stem cell aging: mechanisms, regulators and therapeutic opportunities. *Nature medicine*, *20*, 870–880.

Ohm, J. E., McGarvey, K. M., Yu, X., Cheng, L., Schuebel, K. E., Cope, L., Mohammad, H. P., Chen, W., Daniel, V. C., Yu, W., Berman, D. M., Jenuwein, T., Pruitt, K., Sharkis, S. J., Watkins, D. N., Herman, J. G., & Baylin, S. B. (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature Genetics*, *39*, 237–242.

Pawelec, G. (2017). Immunosenescence and cancer. *Biogerontology*, *18*, 717–721.

Pérez, R. F., Tejedor, J. R., Bayón, G. F., Fernández, A. F., & Fraga, M. F. (2018). Distinct chromatin signatures of DNA hypomethylation in aging and cancer. *Aging Cell*, *17*, e12744.

Pérez, R. F., Tejedor, J. R., Santamarina-Ojeda, P., Martínez, V. L., Urdinguio, R. G., Villamañán, L., Candiota, A. P., Sarró, N. M. V., Barradas, M., Fernandez-Marcos, P. J., Serrano, M., Fernández, A. F., & Fraga, M. F. (2021). Conservation of Aging and Cancer Epigenetic Signatures across Human and Mouse. *Molecular Biology and Evolution*, *38*, 3415–3435.

Raj, K., & Horvath, S. (2020). Current perspectives on the cellular and molecular features of epigenetic ageing. *Experimental Biology and Medicine (Maywood)*, *245*, 1532–1542.

Rakyan, V. K., Down, T. A., Maslau, S., Andrew, T., Yang, T.-P., Beyan, H., Whittaker, P., McCann, O. T., Finer, S., Valdes, A. M., Leslie, R. D., Deloukas, P., & Spector, T. D. (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Research*, *20*, 434–439.

Reinisch, A., Etchart, N., Thomas, D., Hofmann, N. A., Fruehwirth, M., Sinha, S., Chan, C. K., Senarath-Yapa, K., Seo, E.-Y., Wearda, T., Hartwig, U. F., Beham-Schmid, C., Trajanoski, S., Lin, Q., Wagner, W., Dullin, C., Alves, F., Andreeff, M., Weissman, I. L., … Strunk, D. (2015). Epigenetic and in vivo comparison of diverse MSC sources reveals an endochondral signature for human hematopoietic niche formation. *Blood*, *125*, 249–260.

Rozenblit, M., Hofstatter, E., Liu, Z., O'Meara, T., Storniolo, A. M., Dalela, D., Singh, V., Pusztai, L., & Levine, M. (2022). Evidence of accelerated epigenetic aging of breast tissues in patients with breast cancer is driven by CpGs associated with polycomb-related genes. *Clinical Epigenetics*, *14*, 30.

Rudolph, K. L. (2021). DNA-methylation aging at single-cell level. *Nature Aging*, *1*, 1086–1087.

Sarkar, T. J., Quarta, M., Mukherjee, S., Colville, A., Paine, P., Doan, L., Tran, C. M., Chu, C. R., Horvath, S., Qi, L. S., Bhutani, N., Rando, T. A., & Sebastiano, V. (2020). Transient non-integrative expression of nuclear reprogramming factors promotes multifaceted amelioration of aging in human cells. *Nature Communications*, *11*, 1545.

Schellenberg, A., Lin, Q., Schüler, H., Koch, C. M., Joussen, S., Denecke, B., Walenda, G., Pallua, N., Suschek, C. V., Zenke, M., & Wagner, W. (2011). Replicative senescence of mesenchymal stem cells causes DNA-methylation changes which correlate with repressive histone marks. *Aging (Albany NY)*, *3*, 873–888.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B. E., Bergman, Y., Simon, I., & Cedar, H. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nature genetics*, *39*, 232–236.

Schumacher, B., Pothof, J., Vijg, J., & Hoeijmakers, J. H. J. (2021). The central role of DNA damage in the ageing process. *Nature*, *592*, 695–703.

Seale, K., Horvath, S., Teschendorff, A., Eynon, N., & Voisin, S. (2022). Making sense of the ageing methylome. *Nature Reviews Genetics*, 1–21. https://doi.org/10.1038/s41576-022-00477-6

Serena, C., Millan, M., Ejarque, M., Saera-Vila, A., Maymó-Masip, E., Núñez-Roa, C., Monfort-Ferré, D., Terrón-Puig, M., Bautista, M., Menacho, M., Martí, M., Espin, E., Vendrell, J., & Fernández-Veledo, S. (2020). Adipose stem cells from patients with Crohn's disease show a distinctive DNA methylation pattern. *Clinical Epigenetics*, *12*, 53.

Shao, K., Koch, C., Gupta, M. K., Lin, Q., Lenz, M., Laufs, S., Denecke, B., Schmidt, M., Linke, M., Hemmies, H. C., Hescheler, J., Zenke, M., Zechner, U., Šarić, T., & Wagner, W. (2013). Induced pluripotent mesenchymal stromal cell clones retain donor-derived differences in DNA methylation profiles. *Molecular Therapy*, *21*, 240–250.

Søraas, A., Matsuyama, M., de Lima, M., Wald, D., Buechner, J., Gedde-Dahl, T., Søraas, C. L., Chen, B., Ferrucci, L., Dahl, J. A., Horvath, S., & Matsuyama, S. (2019). Epigenetic age is a cell-intrinsic property in transplanted human hematopoietic cells. *Aging Cell*, *18*, e12897.

Struijk, R. B., Dorssers, L. C. J., Henneman, P., Rijlaarsdam, M. A., Venema, A., Jongejan, A., Mannens, M. M. A. M., Looijenga, L. H. J., Repping, S., & van Pelt, A. M. M. (2020). Comparing genome-scale DNA methylation and CNV marks between adult human cultured ITGA6+ testicular cells and seminomas to assess in vitro genomic stability. *PLoS One*, *15*, e0230253.

Sun, D., Luo, M., Jeong, M., Rodriguez, B., Xia, Z., Hannah, R., Wang, H., Le, T., Faull, K. F., Chen, R., Gu, H., Bock, C., Meissner, A., Göttgens, B., Darlington, G. J., Li, W., & Goodell, M. A. (2014). Epigenomic Profiling of Young and Aged HSCs Reveals Concerted Changes during Aging that Reinforce Self-Renewal. *Cell Stem Cell*, *14*, 673–688.

Sziráki, A., Tyshkovskiy, A., & Gladyshev, V. N. (2018). Global remodeling of the mouse DNA methylome during aging and in response to calorie restriction. *Aging Cell*, *17*, e12738.

Tao, Y., Kang, B., Petkovich, D. A., Bhandari, Y. R., In, J., Stein-O'Brien, G., Kong, X., Xie, W., Zachos, N., Maegawa, S., Vaidya, H., Brown, S., Chiu Yen, R.-W., Shao, X., Thakor, J., Lu, Z., Cai, Y., Zhang, Y., Mallona, I., … Easwaran, H. (2019). Aging-like Spontaneous Epigenetic Silencing Facilitates Wnt Activation, Stemness, and BrafV600EInduced Tumorigenesis. *Cancer Cell*, *35*, 315–328.e6.

Tejedor, J. R., & Fraga, M. F. (2017). Interindividual epigenetic variability: Sound or noise? *BioEssays*, *39*, 1700055.

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Weisenberger, D. J., Shen, H., Campan, M., Noushmehr, H., Bell, C. G., Maxwell, A. P., Savage, D. A., Mueller-Holzner, E., Marth, C., Kocjan, G., Gayther, S. A., Jones, A., Beck, S., Wagner, W., Laird, P. W., … Widschwendter, M. (2010). Agedependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Research*, *20*, 440–446.

Tian, X., Firsanov, D., Zhang, Z., Cheng, Y., Luo, L., Tombline, G., Tan, R., Simon, M., Henderson, S., Steffan, J., Goldfarb, A., Tam, J., Zheng, K., Cornwell, A., Johnson, A., Yang, J.-N., Mao, Z., Manta, B., Dang, W., … Gorbunova, V. (2019). SIRT6 Is Responsible for More Efficient DNA Double-Strand Break Repair in Long-Lived Species. *Cell*, *177*, 622–638.e22.

Trapp, A., Kerepesi, C., & Gladyshev, V. N. (2021). Profiling epigenetic age in single cells. *Nature Aging*, *1*, 1189–1201.

Unnikrishnan, A., Hadad, N., Masser, D. R., Jackson, J., Freeman, W. M., & Richardson, A. (2018). Revisiting the genomic hypomethylation hypothesis of aging. *Annals of the New York Academy of Sciences*, *1418*, 69–79.

Vaz, M., Hwang, S. Y., Kagiampakis, I., Phallen, J., Patil, A., O'Hagan, H. M., Murphy, L., Zahnow, C. A., Gabrielson, E., Velculescu, V. E., Easwaran, H. P., & Baylin, S. B. (2017). Chronic Cigarette Smoke-Induced Epigenomic Changes Precede Sensitization of Bronchial Epithelial Cells to Single-Step Transformation by KRAS Mutations. *Cancer Cell*, *32*, 360–376.

Vincze, O., Colchero, F., Lemaître, J.-F., Conde, D. A., Pavard, S., Bieuville, M., Urrutia, A. O., Ujvari, B., Boddy, A. M., Maley, C. C., Thomas, F., & Giraudeau, M. (2022). Cancer risk across mammals. *Nature*, *601*, 263–267.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer International Publishing.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D. J., Campan, M., Young, J., Jacobs, I., & Laird, P. W. (2007). Epigenetic stem cell signature in cancer. *Nature Genetics*, *39*, 157–158.

Wilson, V. L., Smith, R. A., Ma, S., & Cutler, R. G. (1987). Genomic 5-methyldeoxycytidine decreases with age. *Journal of biological chemistry*, *262*, 9948–9951.

Xie, W., Kagiampakis, I., Pan, L., Zhang, Y. W., Murphy, L., Tao, Y., Kong, X., Kang, B., Xia, L., Carvalho, F. L. F., Sen, S., Chiu Yen, R. W., Zahnow, C. A., Ahuja, N., Baylin, S. B., & Easwaran, H. (2018). DNA Methylation Patterns Separate Senescence from Transformation Potential and Indicate Cancer Risk. *Cancer Cell*, *33*, 309–321.e5.

Xu, Z., Niu, L., & Taylor, J. A. (2021). The ENmix DNA methylation analysis pipeline for Illumina BeadChip and comparisons with seven other preprocessing pipelines. *Clinical Epigenetics*, *13*, 216.

Yang, Z., Wong, A., Kuh, D., Paul, D. S., Rakyan, V. K., Leslie, R. D., Zheng, S. C., Widschwendter, M., Beck, S., & Teschendorff, A. E. (2016). Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biology*, *17*, 205.

Zhou, W., Dinh, H. Q., Ramjan, Z., Weisenberger, D. J., Nicolet, C. M., Shen, H., Laird, P. W., & Berman, B. P. (2018). 764 DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nature genetics*, *50*, 591–602.

How to cite this article: Pérez, R. F., Tejedor, J. R., Fernández, A. F., & Fraga, M. F. (2022). Aging and cancer epigenetics: Where do the paths fork? *Aging Cell*, *00*, e13709. https://doi.org/10.1111/acel.13709

AGING AND DEVELOPMENT: FOURTH ARTICLE

Pérez RF, Santamarina P, Tejedor JR, Urdinguio RG, Álvarez-Pitti J, Redon P, Fernández AF, Fraga MF & Lurbe E (2019) Longitudinal genome-wide DNA methylation analysis uncovers persistent early-life DNA methylation changes. J Transl Med 17, 15.

In this work, we focused on better characterizing aging-associated DNA methylation changes occurring during the first years of life. Developmental and aging processes may be different in nature, and the characterization of early-life DNA methylation alterations is of help to distinguish the two processes. Thus, we compiled blood samples from a longitudinal pediatric cohort with measurements at birth, 5 and 10 years of age and profiled DNA methylation levels at more than 700,000 CpG sites using Infinium MethylationEPIC arrays. Our results reveal that the methylome is drastically remodeled during the first 5 years of life while subsequent epigenomic alterations are much smaller. In addition, development-associated alterations mimic aging-associated alterations but occur in a much bigger scale. These observations shed light on the relationships between developmental- and aging-related DNA methylation changes.

Personal contribution to the work: The samples and clinical information were compiled by the lab of Dr. Empar Lurbe. I generated, preprocessed and analyzed all of the data presented in this manuscript, as well as prepared the figure panels and wrote the manuscript. I performed the pyrosequencing experiments with the assistance of Pablo Santamarina. The work was supervised by Dr. Mario F. Fraga and Dr. Empar Lurbe.

Pérez et al. J Transl Med (2019) 17:15
https://doi.org/10.1186/s12967-018-1751-9

Journal of Translational Medicine

RESEARCH Open Access

Longitudinal genome-wide DNA methylation analysis uncovers persistent early-life DNA methylation changes

Raúl F. Pérez[1,2], Pablo Santamarina[1,2], Juan Ramón Tejedor[1], Rocío G. Urdinguio[1], Julio Álvarez Pitti[3,4], Pau Redon[3,4], Agustín F. Fernández[1], Mario F. Fraga[2*] and Empar Lurbe[3,4*]

Abstract

Background: Early life is a period of drastic epigenetic remodeling in which the epigenome is especially sensitive to extrinsic and intrinsic influence. However, the epigenome-wide dynamics of the DNA methylation changes that occur during this period have not been sufficiently characterized in longitudinal studies.

Methods: To this end, we studied the DNA methylation status of more than 750,000 CpG sites using Illumina MethylationEPIC arrays on 33 paired blood samples from 11 subjects at birth and at 5 and 10 years of age, then characterized the chromatin context associated with these loci by integrating our data with histone, chromatin-state and enhancer-element external datasets, and, finally, validated our results through bisulfite pyrosequencing in two independent longitudinal cohorts of 18 additional subjects.

Results: We found abundant DNA methylation changes (110,726 CpG sites) during the first lustrum of life, while far fewer alterations were observed in the subsequent 5 years (460 CpG sites). However, our analysis revealed persistent DNA methylation changes at 240 CpG sites, indicating that there are genomic locations of considerable epigenetic change beyond immediate birth. The chromatin context of hypermethylation changes was associated with repressive genomic locations and genes with developmental and cell signaling functions, while hypomethylation changes were linked to enhancer regions and genes with immunological and mRNA and protein metabolism functions. Significantly, our results show that genes that suffer simultaneous hyper- and hypomethylation are functionally distinct from exclusively hyper- or hypomethylated genes, and that enhancer-associated methylation is different in hyper- and hypomethylation scenarios, with hypomethylation being more associated to epigenetic changes at blood tissue-specific enhancer elements.

Conclusions: These data show that epigenetic remodeling is dramatically reduced after the first 5 years of life. However, there are certain *loci* which continue to manifest DNA methylation changes, pointing towards a possible functionality beyond early development. Furthermore, our results deepen the understanding of the genomic context associated to hyper- or hypomethylation alterations during time, suggesting that hypomethylation of blood tissue-specific enhancer elements could be of importance in the establishment of functional states in blood tissue during early-life.

Keywords: Epigenetics, DNA methylation, Histone modification, Aging, Newborn, Longitudinal

*Correspondence: mffraga@cinn.es; Empar.Lurbe@uv.es
[2] Nanomedicine Group, Nanomaterials and Nanotechnology Research Center (CINN-CSIC), Universidad de Oviedo, 33940 Oviedo, Asturias, Spain
[3] Servicio de Pediatría, Consorcio Hospital General Universitario de Valencia, 46014 Valencia, Spain
Full list of author information is available at the end of the article

Background

Epigenetic modifications such as DNA methylation are known to influence gene expression and thus biological function [1] and alterations in epigenetic marks are found in processes ranging from physiological development and cellular differentiation [2] to pathological scenarios, as well as aging [3, 4]. Reflecting the developmental changes of the individual, most epigenetic changes occur during embryogenesis, where waves of extensive epigenetic reprogramming take place [5]. After birth, epigenetic modification continues to take place throughout the human lifespan, both at the DNA methylation and chromatin levels [6] and precise DNA methylation markers of age have recently been developed [7].

It seems, however, that the first years of life constitute the post-natal period with the most substantial epigenetic changes [8, 9] and, moreover, prenatal and early life epigenetic aggression has been linked to countless health-related consequences in a wide variety of settings [10]. Thus, the characterization of childhood DNA methylation changes could help uncover genomic locations of functional relevance.

While age-related changes have been functionally linked to many biological processes, other alterations are known to arise in a stochastic fashion [11, 12]. As such, the use of longitudinal experimental designs can allow for the identification of DNA methylation alterations in a more controlled environment [9, 13–16], and this approach has also been employed to study the influence of clinical parameters on the epigenome, especially during early childhood [17]. However, longitudinal DNA methylation studies which analyze data from more than two time points remain scarce [17–19] and most of the previous literature stands on now surpassed methylation screening technologies such as the Infinium Human Methylation 450K BeadChip which mainly interrogates CpG-dense genomic regions, even though the functional association between DNA methylation and gene expression is increasingly being described for CpG-sparse locations such as enhancers and gene bodies [20].

This study, therefore, presents, to the best of our knowledge, the first 3-point longitudinal genome-wide DNA methylation analysis of blood tissue employing the Illumina Infinium MethylationEPIC BeadChip, which allowed us to characterize two distinct phases of early-life epigenetic changes during the first 10 years of life at the same time as focusing on various types of genomic regions, using a clinically well-characterized cohort. Furthermore, we integrated our generated data with external chromatin and enhancer datasets in order to obtain a functional view of the observed age-related DNA methylation changes. Finally, we applied bisulfite pyrosequencing in order to technically and biologically validate the results obtained in the discovery and independent cohorts, respectively.

Methods

Selection of participants

Newborn children born at term (gestational age ≥ 37 weeks) after uncomplicated pregnancies and in the absence of perinatal illness in the General Hospital, University of Valencia, Spain, were randomly selected to enroll in the study. The characteristics of each gestation and delivery were obtained from routine obstetrical records, and the gestational age at birth was measured using the method of Ballard et al. [21]. Groups were established on the basis of birth weight: <10th percentile for their sex (SGA); between 10th and 90th percentile (AGA); and >90th percentile (LGA) [22]. The subjects were followed-up at 5 and 10 years of age.

Anthropometric parameters and biochemical analyses

At both 5 and 10 years, height was measured to the nearest 0.5 cm using a standardized wall-mounted height board, and body weight to the nearest 0.1 kg using a standard beam balance scale with the barefoot subjects wearing light clothing. Body mass index (BMI) was calculated as the weight in kilograms divided by the square of the height in meters (see Table 1 and Additional file 1: Table S9 for this and other clinical information).

Sample collection, DNA extraction, and quantification

Blood samples were collected from a total of 29 subjects at different testing times (n=51). Cord blood samples were taken at birth and peripheral venous blood samples were taken from each child during their fifth and tenth year of life. Genomic DNA was extracted with the RealPure kit (RealPure, REAL, Durviz) and quantified with the Nanodrop-2000C Spectrophotometer. A DNA quality check was performed with Quant-iT PicoGreen dsDNA reagent.

Genome-wide DNA methylation analysis

Microarray-based DNA methylation profiling was performed with the Illumina Infinium MethylationEPIC BeadChip on 33 paired blood samples from 11 subjects collected at birth, 5 and 10 years of age. Bisulfite conversion of DNA was performed using the EZ DNA methylation kit (Zymo Research) following the manufacturer's recommendations, but with the modifications described in the Infinium assay methylation protocol guide. Processed DNA samples were then hybridized to the BeadChip following the Illumina Infinium HD methylation protocol at the Centro Nacional de Genotipado (CEGEN-ISCIII, Spain, http://www.cegen.org).

Table 1 Clinical characteristics of the 11 subjects enrolled in the study

Subject At 0 years	Sex	Birth	Lactation	Gestational age (weeks)	Cephalic perimeter (cm)	Birth weight (g)	Birth weight (group)	Birth length (cm)
1	Male	VD	FF	40	35.0	3580	AGA	51.0
2	Male	CS	FF	40	34.0	3800	AGA	50.0
3	Male	VD	FF	38	31.0	2155	SGA	46.0
4	Male	VD	FF	37	32.5	2720	AGA	47.0
5	Female	VD	BF	37	33.0	2540	AGA	47.0
6	Female	VD	BF	38	32.0	2490	AGA	48.5
7	Male	CS	FF	41	35.0	3550	AGA	52.5
8	Female	CS	FF	38	34.5	3540	AGA	49.5
9	Female	VD	BF	39	33.0	3940	LGA	51.0
10	Male	VD	FF	39	35.0	3850	LGA	50.0
11	Female	CS	BF	39	31.0	2340	SGA	45.5
Subject At 5 years	**Creatinine**	**Insulin**	**HDL**	**LDL**	**Triglycerides**	**Weight (g)**	**Body mass index**	**Height (cm)**
1	0.3	1.8	47	101	60	18,700	14.6	113.0
2	0.4	2.0	60	123	93	22,700	21.4	103.0
3	0.6	2.9	52	118	55	17,000	13.8	111.0
4	0.3	5.8	61	120	69	15,700	14.9	102.5
5	0.3	4.7	58	103	79	22,500	16.7	116.0
6	0.3	2.7	56	109	56	16,600	14.4	107.5
7	0.3	2.1	44	109	45	16,800	14.1	109.0
8	0.4	7.0	59	98	69	26,100	18.6	118.5
9	0.3	4.5	47	118	81	22,000	15.2	120.5
10	0.4	4.2	83	128	58	22,500	17.0	115.0
11	0.4	7.6	54	66	34	23,100	14.6	113.0
Subject At 10 years	**Creatinine**	**Insulin**	**HDL**	**LDL**	**Triglycerides**	**Weight (g)**	**Body mass index**	**Height (cm)**
1	0.5	3.3	47	110	52	27,300	15.6	132.5
2	0.5	3.7	53	117	69	37,600	22.9	128.0
3	0.6	5.2	59	122	46	31,500	14.6	147.0
4	0.5	10.9	64	134	59	34,100	17.2	141.0
5	0.5	12.5	55	147	77	42,900	21.3	142.0
6	0.4	9.1	48	97	66	30,000	15.8	138.0
7	0.4	7.1	40	92	64	27,800	16.5	129.8
8	0.5	9.7	61	111	52	46,000	21.5	146.2
9	0.4	3.2	47	88	88	30,100	17.7	130.5
10	0.6	4.5	99	107	41	38,200	18.3	144.5
11	0.6	6.3	52	65	33	37,000	17.6	145.0

FF formula feeding, *BF* breast feeding, *VD* vaginal delivery, *CS* cesarean section, *AGA* appropriate for gestational age, *SGA* small for gestational age, *LGA* large for gestational age, *HDL* high-density lipoprotein, *LDL* low-density lipoprotein

DNA pyrosequencing

The DNA methylation status of representative CpG sites (cg07547765, cg11047325, cg03830443) was evaluated by bisulfite pyrosequencing using the primers described in Additional file 2: Table S8. Technical validations were performed for all 11 subject samples at the three time points (n = 33) and biological validations were performed on two independent cohorts: (1) peripheral blood from 9 children at birth and 5 years of age (n = 18), and (2) peripheral blood from 9 different children at 5 and 10 years (n = 18). Genomic DNA was isolated following standard phenol–chloroform extraction protocols. Bisulfite conversion of isolated DNA was performed in accordance with the EZ DNA methylation-gold kit

(Zymo Research) following the manufacturer's instructions. After PCR amplification, pyrosequencing was performed using PyroMark Q24 reagents and a vacuum prep workstation, equipment and software (Qiagen).

MethylationEPIC BeadChip data preprocessing

All analyses were carried out using the statistical software R (version 3.4.2). IDAT files from the methylationEPIC BeadChip were preprocessed following a pipeline based on the R/Bioconductor package *minfi* (version 1.22.1) [23]. Probes where at least one sample had a detection p-value > 0.01, probes located in chromosomes X and Y, probes overlapping genetic variants [24] and crossreactive and multimapping probes [25] were discarded for downstream analyses. β-values were normalized using the Noob method [26] implemented in *minfi*, followed by a BMIQ normalization [27] implemented in the R/Bioconductor package *ChAMP* (version 2.8.9) [28]. M-values were calculated from the normalized β-values through a logit transformation (R/Bioconductor package *lumi*, version 2.28.0) [29] and employed for downstream analyses. Multidimensional scaling (MDS) and principal component regression analyses were used to identify potential confounding variables (R/Bioconductor packages *Enmix*, version 1.12.4) [30]. When specified, blood cell-type composition was calculated by the Houseman method [31]. However, to correct for batch effects and unwanted sources of variation in the data we performed a surrogate variable analysis [32] with the R/Bioconductor package *sva* (version 3.23.4) [33] because this procedure seeks to adjust for any confounders, including cell-type heterogeneity [34]. The number of latent factors was estimated using the "*leek*" method, but it is of note that no surrogate variables were found after the normalization procedures for our generated datasets.

Differential DNA methylation analyses

Differentially methylated probes (dmCpGs) were calculated with the R/Bioconductor package *limma* (version 3.32.10) [35]. A linear model was fitted between methylation levels as response variable, the variable of interest (*age* group), the aforementioned surrogate variables and the information of the paired samples (*patient* group) for each of the analyses: 0 vs. 5 and 5 vs. 10 years old. The set of p-values obtained from the tests was adjusted for multiple comparisons using the Benjamini–Hochberg method to control for false discovery rate (FDR < 0.05). An additional threshold of shift size was applied, filtering out significant probes with M-value changes of less than 0.5, as has been suggested elsewhere [36]. Venn diagram representations of the relationships between statistically significant dmCpGs were generated with the online resource provided by the UGent/VIB bioinformatics unit (available at http://bioinformatics.psb.ugent.be/webtools/Venn/). Further enrichment analyses were performed by means of two-sided Fisher's tests ($p < 0.05$ significance threshold), measuring effect size either by odds ratio (OR), or by the difference between observed counts and expected hypergeometric means, employing appropriate backgrounds for the interrogated probes in each given context.

Density of CpG analysis

For each of the probes in the methylation arrays, density of CpG was measured as the number of genomic CpGs present divided by the number of those possible in a 2 kbp window centered on the CpG location under study. Wilcoxon non-parametric tests were used to determine whether there were significant differences between the density distributions of the CpGs belonging to each subset of interest and the densities of the array probes in the background. A significance level of 0.05 was employed for all tests. Shift size was measured using median differences and Cliff's Delta (CD).

CGI status and genomic region analysis

CGI (CpG island) membership was assigned to each probe using the Illumina EPIC annotation with the R/Bioconductor package and *IlluminaHumanMethylationEPICanno.ilm10b2.hg19* (version 0.6.0). Genomic region position was assigned using the R/Bioconductor packages *TxDb.Hsapiens.UCSC.hg19.knownGene* (version 3.2.2) and *ChIPseeker* (version 1.12.1) [37]. Statistical significance with respect to concrete CGI status or genomic region was determined by two-sided Fisher's tests (significance level $p < 0.05$), and ORs were used as a measure of the association effect with respect to a particular feature. Appropriate backgrounds which included all the probes interrogated by the EPIC array in each of the comparisons were used for statistical purposes.

Region set enrichment analysis

Enrichment analyses were performed with the R/Bioconductor package *LOLA* (version 1.6.0) [38], which looks for over-enrichment by conducting one-sided Fisher's tests (p-value < 0.05 significance threshold), by comparing overlap of probes (10 bp probe-centered windows) with the dataset of interest. Enrichment of histone marks was determined using histone ChIP-seq peak tracks (H3K4me1, H3K4me3, H3K27me3, H3K36me3, H3K9me3 and H3K27ac marks) from 98 epigenomes (primary tissues, cultures and cell lines) obtained from the NIH Roadmap and ENCODE projects [39, 40] and integrated into the *LOLA* extended software (datasets obtained from http://databio.org/regiondb). The same method was employed for

chromatin-segment analysis using the expanded NIH Roadmap ChromHMM 18-state model tracks for the same 98 epigenomes, generated from the previous histone marks (custom database generated with data obtained from http://egg2.wustl.edu/roadmap/). In a similar fashion, enhancer enrichment analysis was performed using the enhancer tracks defined in the EnhancerAtlas database [41] (custom database generated with data obtained from http://enhanceratlas.org/download.php) for 65 genomes (primary tissues, cultures and cell lines) defined by the consensus combination of different independent experimental datasets such as histone marks, DNAse hypersensitive sites (DHS) and transcription factor binding sites (TFBS). This last dataset was also used to create the CpG-gene network shown in Fig. 4c: CpGs were first mapped to enhancers in at least one of the 22 EnhancerAtlas tracks corresponding to blood tissue and cell type, and afterwards plotted using the R/CRAN package *igraph*, where network nodes represent CpG sites or genes, while network edges reflect interactions between the enhancer element containing the CpG and the corresponding associated gene.

Gene ontology analyses

Gene ontology enrichments were calculated using the R/Bioconductor package *missMethyl* (version 1.10.0 *gometh* function) [42], which performs one-sided hypergeometric tests taking into account and correcting for any bias derived from the use of differing numbers of probes per gene interrogated by the array. The annotation database that was interrogated is contained within the R/Bioconductor package *GO.db* (version 3.4.1). Appropriate backgrounds of total probes for each given context were employed in the corresponding analyses. Additionally, because a great number of genes contained both hyper- and hypomethylated probes, certain gene ontology enrichments were performed only on those genes that were exclusively hyper- or hypomethylated. Further analyses were also performed by grouping the annotated CpGs into gene regions, using the annotations provided in Additional file 3: Table S1 (column "annotation"): "promoter" is formed by collapsing "Distal promoter" and "Promoter (≤ 1 kb)", "gene body" is formed by collapsing "3′ UTR", "5′ UTR", "Intron", "Exon" and "Downstream"). epigenomes Additional file 4: Table S6 contains the full information on the gene ontology analyses. To help with the visualization of the results, the ontologies were graphically summarized using the online tool REViGO (available at http://revigo.irb.hr/) [43], which performs semantic similarity analyses in order to reduce the number of redundant functional terms obtained.

Results

Genome-wide profiling of DNA methylation changes during the first 10 years of life

To characterize DNA methylation changes that occur during the first years of life, we analyzed the methylation status of 783,659 CpG sites in cord, 5-year-old and 10-year-old paired blood samples from 11 subjects (see Table 1 and Fig. 1a) using the Illumina Infinium MethylationEPIC BeadChip. After preprocessing our data (see "Methods"), we applied an empirical Bayes moderated t test to identify CpGs differentially methylated between the different age groups (dmCpGs; FDR < 0.05) (see "Methods"). We found 110,726 CpGs whose methylation level significantly changed from birth to 5 years ($0 \rightarrow 5$), but only 460 from 5 to 10 years ($5 \rightarrow 10$). The $0 \rightarrow 5$ widespread methylation changes consisted of, respectively, 59,749 and 50,977 CpGs hyper- or hypomethylated over time, while 130 and 330 CpGs, respectively, were found to be hyper- or hypomethylated over the $5 \rightarrow 10$ years period (Fig. 1b, Additional file 3: Table S1). Notably, hypomethylation changes were found to be more pronounced than hypermethylation changes between 0 and 5 years, while the magnitude of change was more balanced in both directions between 5 and 10 years (Additional file 5: Figure S1).

Next, we determined the genomic distribution of the dmCpGs identified, distinguishing between CpG island-related regions and gene-related regions (Fig. 1c). As compared to the distribution of the array background, hypermethylated CpGs were found to be enriched at CpG islands for both $0 \rightarrow 5$ and $5 \rightarrow 10$ changes (Fisher's test; both $p < 0.001$, odds ratios (ORs) = 1.2 and 7.4, respectively) and impoverished at open sea locations (Fisher's test; both $p < 0.001$, ORs = 0.6 and 0.2 respectively), while the inverse was the case for hypomethylated CpGs, that is they were very enriched at open sea regions (Fisher's test; both $p < 0.001$, ORs = 2.4 and 6.9, for $0 \rightarrow 5$ and $5 \rightarrow 10$, respectively) and very impoverished at CpG islands (Fisher's test; both $p < 0.001$, ORs = 0.1 and 0.03 for $0 \rightarrow 5$ and $5 \rightarrow 10$, respectively). As regards gene-related locations, hypermethylated CpGs were enriched at promoters and distal promoters (Fisher's test; both $p < 0.001$, ORs = 1.04 and 1.5 for $0 \rightarrow 5$ and both $p < 0.001$, ORs = 2.2 and 1.4 for $5 \rightarrow 10$) and impoverished at intronic regions for $0 \rightarrow 5$ changes (Fisher's test; $p < 0.001$, OR = 0.8) and at both intronic and intergenic regions for $5 \rightarrow 10$ changes (Fisher's test; both $p < 0.001$, ORs = 0.4 and 0.4), while in contrast, hypomethylated CpGs were preferentially found at intronic and intergenic regions for both age bands, (Fisher's test; $p < 0.001$, ORs = 1.6 and 1.3, in these regions respectively, for $0 \rightarrow 5$ and $p < 0.01$ and < 0.001, ORs = 1.4 and 2.7, respectively, for $5 \rightarrow 10$).

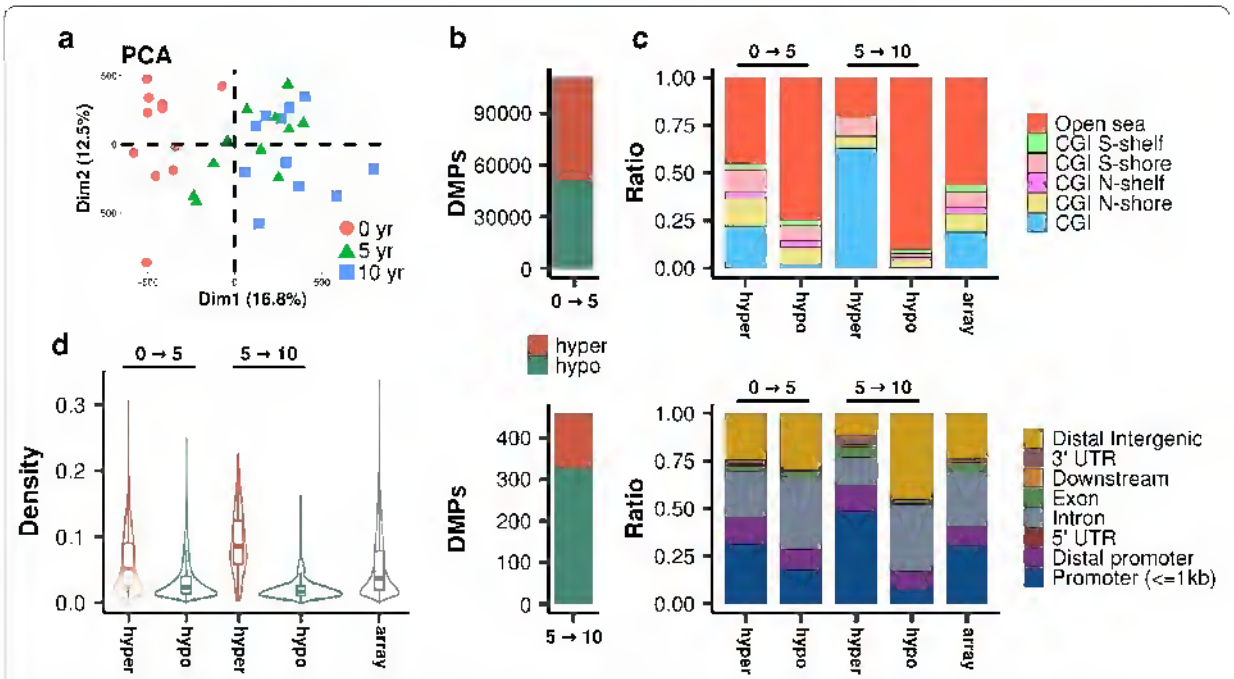


Fig. 1 DNA methylation changes during the first years of life. **a** Principal Component Analysis (PCA) for 783659 CpG sites across all samples in our study. The two PCs represented explain approximately 30% of the variance in methylation levels. **b** Stacked barplots indicating the total number of dmCpGs detected as either 0 → 5 or 5 → 10 changes (see Additional file 3: Table S1 for the detailed names and annotations of the dmCpGs). **c** Stacked barplots depicting the relative distribution of dmCpGs and of the array background according to their CpG island- related or gene location-related status. **d** Violin plots showing the distribution of genomic CpG density for the detected dmCpGs and array background

Lastly, we analyzed the genomic density of the dmCpGs found (see "Methods"), which confirmed that hypermethylation occurred at genomic regions dense in CpG dinucleotides, while hypomethylation appeared at CpG-poor regions (Fig. 1d, Wilcoxon test; all $p < 0.001$, median differences (MD) and Cliff's deltas (CD) compared to array background 0.014/0.15 and −0.014/−0.29 for 0 → 5 and 0.05/0.5 and −0.02/−0.5 for 5 → 10 hyper- and hypomethylated CpGs, respectively).

On the whole, these analyses reveal strong differences between DNA methylation changes that occur during the first years of life, where extensive methylation changes are observed between 0 and 5 years, while only moderate changes occur between 5 and 10 years. Strikingly, while the genomic distribution of these changes appeared to be similar, we found that the changes occurring in the first lustrum of life were of a greater magnitude that those occurring in the second lustrum (Fig. 2a, Wilcoxon test; $p < 0.001$, MD = 0.02, CD = 0.2). Moreover, these differences, which were predicted by the PCA analysis (Fig. 1a), did not seem to be related to any major changes in the composition of blood cell types, as shown by the hierarchical clustering analysis of their predicted cell types (see "Methods" and Additional file 6: Table S2), where the samples did not show extensive age-associated clustering (Fig. 2b).

Chromatin and functional context of early-years DNA methylation changes

In order to study the functional context in which the DNA methylation changes identified occurred, we integrated our results with previously published data from NIH Roadmap Epigenomics and ENCODE relating to ChIP-seq experiments describing the genome-wide localization of the histone modifications H3K4me1, H3K4me3, H3K27ac, H3K36me3, H3K27me3 and H3K9me3 across different blood cell types and cell lines (see "Methods"). Due to the low number of probes, most of the enrichment analyses were unsuccessful for 5 → 10 changes (results available in the Additional file 7: Table S3, Additional file 8: Table S4, Additional file 9: Table S5).

By using this approach, we found that CpGs that were hypermethylated during the first 5 years of life were strongly associated with the repressive histone modification H3K27me3, while hypomethylated CpGs were associated with the active histone modification H3K4me1 (Fig. 3a, see Additional file 7: Table S3 for the full data).

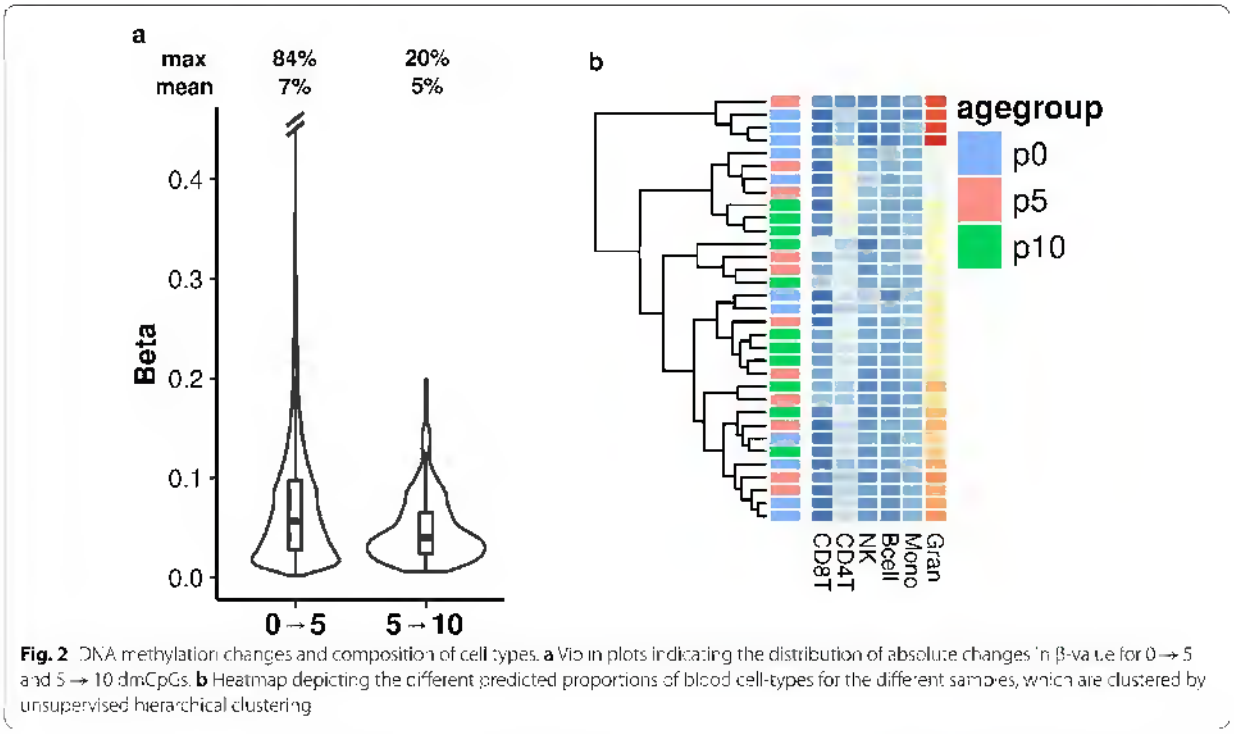


Fig. 2 DNA methylation changes and composition of cell types. **a** Vio in plots indicating the distribution of absolute changes in β-va ue for 0 → 5 and 5 → 10 dmCpGs. **b** Heatmap depicting the different predicted proportions of blood cell-types for the different samples, which are clustered by unsupervised hierarchical clustering

Subsequently we analyzed the enrichment of these dmCpGs for 18 chromatin states, as defined in a Hidden Markov Model (HMM) built from the previous histone modifications (see "Methods"). The results evidenced that hypermethylation occurred at enhancer states associated with bivalent chromatin (i.e. formed by the simultaneous presence of repressive and active marks) and polycomb repressive domains, while hypomethylation occurred at states associated with enhancers, especially of the intragenic type (Fig. 3b, see Additional file 8: Table S4 for the full data). To gain a deeper understanding of enhancer-associated methylation changes, we integrated our data with the EnhancerAtlas database of predicted enhancers across 65 tissues (see "Methods"). This approach enabled us to characterize hypermethylated CpGs as being weakly associated with enhancers present in tissue and embryonic tissue whereas hypomethylated CpGs appeared more strongly associated with enhancers across all datasets, but particularly those specifically related to blood tissue (blood cell lines and primary cells) (Fig. 3c, see Additional file 9: Table S5 for the full data). Taken together, these results highlight the different functionality of the regions where DNA methylation is gained or lost during the first 5 years of life and suggest that hypomethylation could be more relevant than hypermethylation in the establishment of epigenetic patterns at enhancers controlling tissue-specific functions.

Lastly, we performed gene ontology analyses on the genes containing the dmCpGs. In order to avoid bias due to genes with both hyper- and hypomethylated dmCpGs, we conducted the analyses on those genes that were either exclusively hypermethylated or exclusively hypomethylated, also controlling for differences in the number of CpGs analyzed for each gene in the array (see "Methods"). This led to the finding that hypermethylated genes were mainly linked to developmental and cell-to-cell signaling processes, albeit with other notable ontologies also involved, such as response to retinoic acid. Hypomethylated genes, on the other hand, were enriched in metabolic terms related to mRNA and protein metabolism, immune response and mitosis (Fig. 3d, see Additional file 4: Table S6 for the full data). Curiously, when looking at the ontologies of dmCpGs belonging to genes that experienced both hyper- and hypomethylation, the main roles of the genes affected were with respect to cell surface signaling and intracellular component transport (Additional file 10: Figure S2A, Additional file 4: Table S6). This suggests that genes undergoing both hyper- and hypomethylation may well have different cellular functions than those that are differentially methylated in one direction only. This prompted us to examine the ontologies for all of the hyper- or hypomethylated

(See figure on next page.)

Fig. 3 Chromatin and functional context of DNA methylation changes. **a** Heatmaps depicting significant ($p < 0.05$) enrichment of hyper- and hypomethylated 0 → 5 dmCpG sites with different histone marks, in a selection of 21 different blood tissue and cell types (see Additional file 7: Table S3 for 98 full cell and tissue types, and also 5 → 10 dmCpG enrichments). The color code indicates the significant enrichment based on the odds ratio (OR). **b** Heatmaps displaying significant ($p < 0.05$) enrichment of hyper- and hypomethylated 0 → 5 dmCpG sites with different chromatin states, in a selection of 21 different blood tissue and cell types (see Additional file 8: Table S4 for 98 full cell and tissue types, and also 5 → 10 dmCpG enrichments). The color code indicates the significant enrichment based on the OR. **c** Heatmaps showing significant ($p < 0.05$) enrichment of hyper- and hypomethylated 0 → 5 dmCpG sites with predicted enhancers in 65 different cell and tissue types (see Additional file 9: Table S5 for full data, and also 5 → 10 dmCpG enrichments). The color code indicates the significant enrichment based on the OR. **d** Treemap plots indicating the results of REViGO semantic analyses of significantly enriched (FDR < 0.05) gene ontology biological process terms for genes containing 0 → 5 dmCpGs. In total, 402 and 345 significant terms were found for hyper- and hypomethylated dmCpGs, respectively (see Additional file 4: Table S6 for full results, including molecular function and cellular component terms, and also 5 → 10 dmCpG enrichments)

probes (i.e. without taking into account the fact that they might be associated with genes that also contained probes with change in the opposite direction), which were also different (Additional file 10: Figure S2B, Additional file 4: Table S6): while the 0 → 5 hypermethylation ontologies did not substantially change, the hypomethylation terms appeared related more to cellular function, localization and activation, and membrane function, with immune terms being were maintained (Additional file 10: Figure S2, Additional file 4: Table S6). This last observation implies that not taking into account the existence of simultaneously hyper- and hypomethylated genes could alter the results obtained by ontology-enrichment analyses.

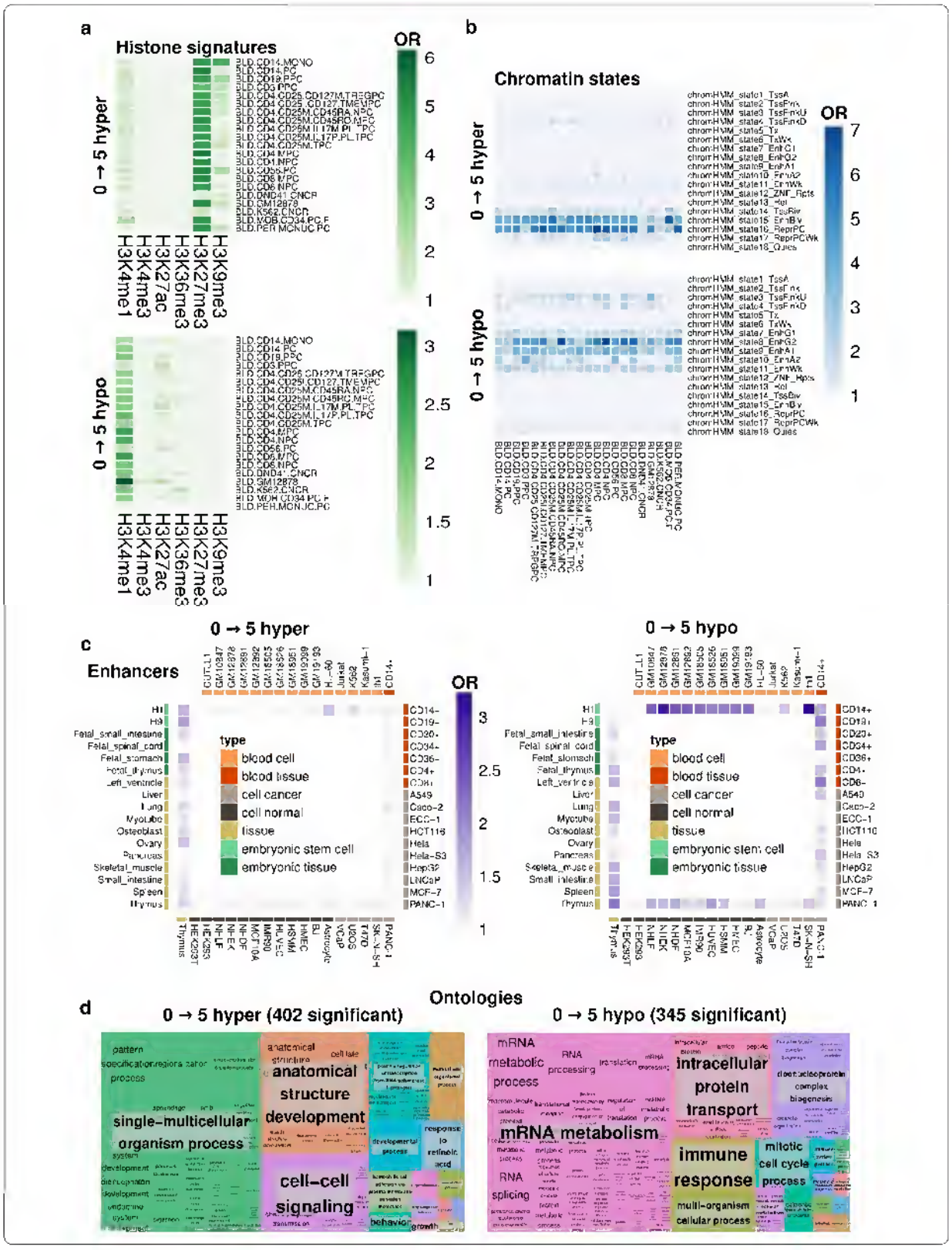
Taking into account these last findings, we extended our analyses to characterize gene ontologies depending on the gene region of the associated dmCpGs. We divided the 0 → 5 dmCpGs into promoter-, exon-, intron- or gene body-associated, hyper- or hypomethylated dmCpGs (8 groups in total), in order to identify differences, for example, between the ontologies of genes that suffered DNA methylation changes in their promoters versus genes with changes in their gene bodies. The results (Additional file 11: Figure S3, Additional file 4: Table S6), when compared to those previously found without segregating dmCpGs by context (shown in Fig. 3d) showed that genes associated to different functions suffered changes at different locations. For hypermethylated CpGs (Additional file 11: Figure S3A), the main functions related to development and signaling seen in Fig. 3d were found on genes with promoter- and exon-associated DNA methylation gains, while the functions associated to genes containing exon- or gene body-(the majority of which is exons) methylation increases were considerably different, manifesting functions such as GTPase signaling, ion transport or cell projection. For hypomethylated CpGs (Additional file 11: Figure S3B), the initially observed functions related to mRNA metabolism, protein transport or immune response were also found in genes with promoter-associated DNA methylation loses (which were, though, much more enriched in immunological functions), while, again, the functions observed for genes containing exon- or gene body-associated methylation reduction changed, being more related to signal transduction and movement of subcellular components. On the whole, these observations reinforce the idea that DNA methylation changes occur at different gene regions depending on the functionality of the affected gene.

Persistent DNA methylation changes during the first 10 years of life

After conducting the analyses of DNA methylation, we made use of the longitudinal design of our study to compare the changes occurring in the 0 → 5 compared to the 5 → 10 period. Interestingly, we discovered an over-enriched subset of 240 dmCpGs that were altered in the first 5 years of life which continued to change in the following 5 years [Fig. 4a, Fisher's test; $p < 0.001$, OR = 6.6, expected hypergeometric mean (EHM) = 65]. No significant gene ontologies were found for this set after correcting for multiple comparisons, although the top hypermethylation terms were related to developmental processes, and those for hypomethylation with metabolic and immunological activities (data not shown).

We subsequently filtered out CpGs with changes of less than 10% in β-value in order to retain those with more plausible biological roles, leaving 36 CpGs (Table 2). These remaining CpGs manifested two kinds of behavior: either strong first-lustrum changes with weaker second-lustrum changes, or a trend of constant (moderate) change (Fig. 4b, see Additional file 12: Figure S4 for all 36 CpG plots). A total of 16 different genes related to these select 0 → 5 → 10 dmCpGs, and, notably, two of the 36 CpGs were mapped to the HOXB7 gene. Since a proportion of these CpG sites were not associated with any gene and many of them were located in open sea regions (see Table 2), we made use of the EnhancerAtlas database to map them to enhancer elements in the 22 different types of blood tissue and cells shown in Fig. 3c (see "Methods"), and constructed a network with the linked to the

Pérez et al. J Transl Med (2019) 17:15



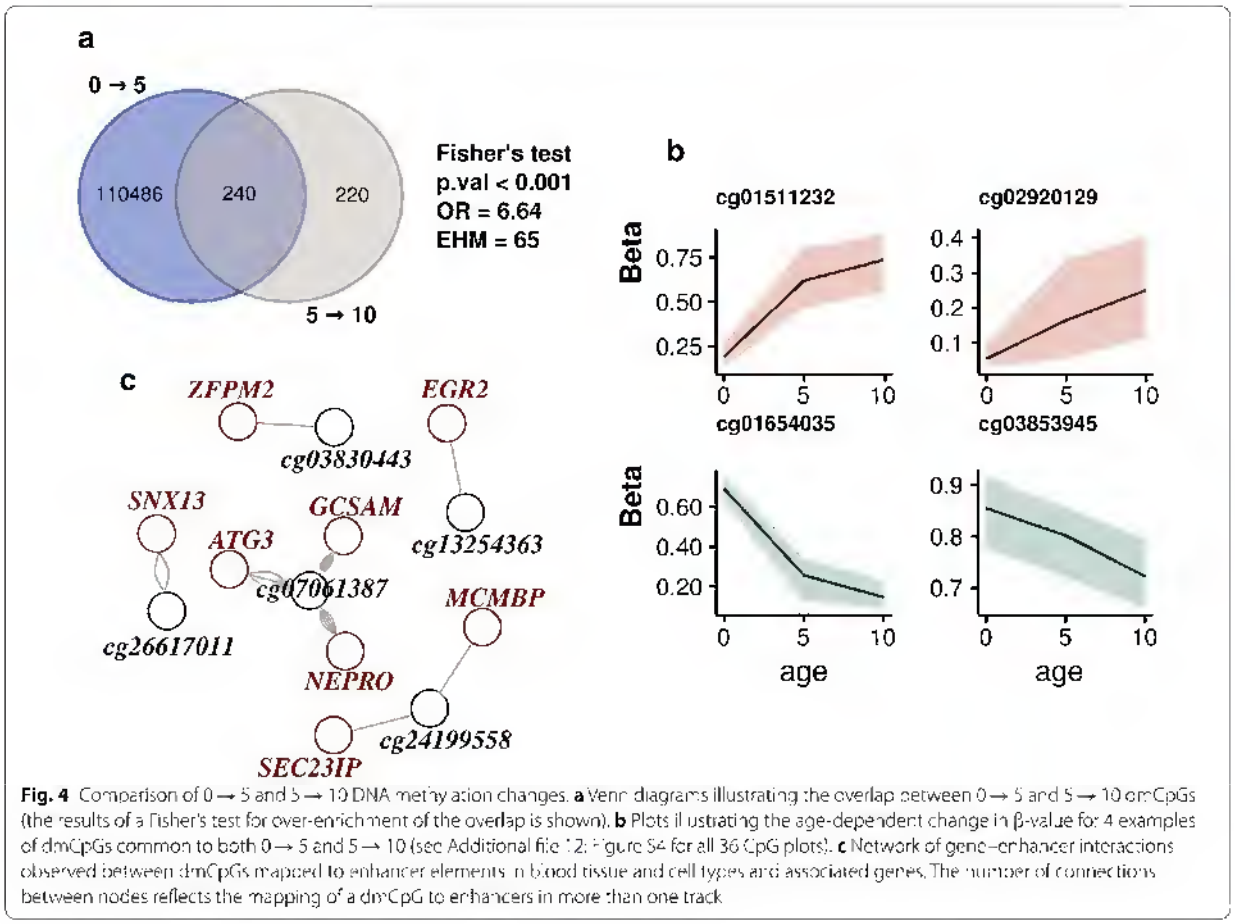


Fig. 4 Comparison of 0 → 5 and 5 → 10 DNA methylation changes. **a** Venn diagrams illustrating the overlap between 0 → 5 and 5 → 10 dmCpGs (the results of a Fisher's test for over-enrichment of the overlap is shown). **b** Plots illustrating the age-dependent change in β-value for 4 examples of dmCpGs common to both 0 → 5 and 5 → 10 (see Additional file 12: Figure S4 for all 36 CpG plots). **c** Network of gene–enhancer interactions observed between dmCpGs mapped to enhancer elements in blood tissue and cell types and associated genes. The number of connections between nodes reflects the mapping of a dmCpG to enhancers in more than one track

Table 2 CpG sites that change substantially during the first 10 years of life

Name	Location	UCSC RefGene	Name	Location	UCSC RefGene	Name	Location	UCSC RefGene
cg13949829	Island	MIR1225;PKD1	cg12024906	Island	HKR1	cg07690222	Open Sea	TMTC2
cg01323777	Island	KCNAB3	cg23201812	Open Sea		cg05498680	Open Sea	
cg00002033	Island	LRFN1	cg01654035	Open Sea		cg11047325	Island	SOCS3
cg23669081	Island	HOXB7	cg03853945	Open Sea		cg01331772	Open Sea	
cg17238334	Open Sea	LOC102477328	cg18311495	Open Sea		cg16022195	S Shore	
cg01511232	Island		cg26617011	Open Sea		cg03465600	Open Sea	PARD3B
cg01577707	Open Sea	MIR100HG	cg03830443	S-Shore	ZFPM2	cg22398226	N-Shore	
cg07547765	Island	HOXB7	cg24199558	Open Sea		cg02920129	Open Sea	ZNF385D
cg07061387	Open Sea	CD200R1	cg12628061	Open Sea		cg13254363	Open Sea	
cg24348981	Open Sea		cg13420364	Open Sea		cg05346619	Open Sea	
cg13329407	Open Sea		cg18647570	Open Sea	DHX8	cg10097598	Open Sea	
cg19567415	Open Sea	TANC2	cg19509778	Island	FOXI2	cg19076536	Open Sea	

enhancer elements containing the dmCpGs (Fig. 4c). In this way we thereby revealed additional relationships with other genes, with some sites, such as cg07061387, showing relationships with up to 3 different genes. Taken together these results indicate that the majority of the later-years epigenetic changes seem to be a weaker

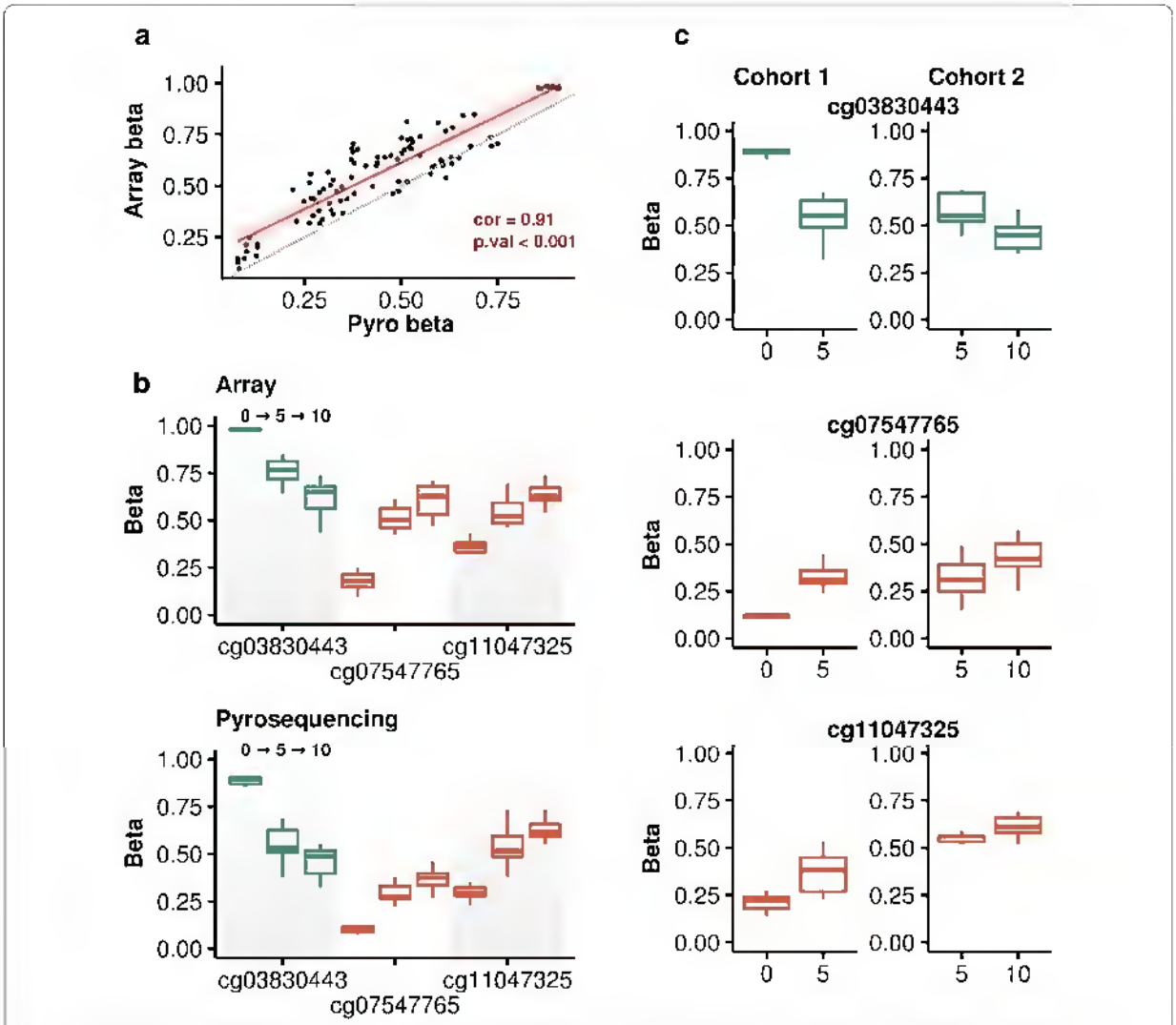


Fig. 5 Technical and experimental validations. **a** Scatter plot showing the correlation between DNA methylation measurements performed by the Infinium MethylationEPIC BeadChip and by bisulfite pyrosequencing for 3 CpGs across the 11 subjects at the 3 time points (red, linear correlation estimation; grey, 1:1 line). **b** Box plots depicting the DNA methylation values of 3 CpGs measured by the Infinium MethylationEPIC BeadChip (top) and by bisulfite pyrosequencing (bottom). **c** Box plots indicating the DNA methylation values of the same 3 CpGs measured by bisulfite pyrosequencing in two independent cohorts, each comprising 9 children: measured (1) at 0 and 5 years of age and (2) at 5 and 10 years of age

continuation of those that occur during the first 5 years of life, although there also appear to be certain CpG sites that show a more linear, and constant, trend of change along the first 10 years of life.

Finally, we performed both technical and experimental bisulfite pyrosequencing validations of three of the CpGs found at the *HOXB7* (cg07547765), *SOCS3* (cg11047325) and *ZFPM2* (cg03830443) genes (Fig. 5, see Additional file 13: Table S7 for full data). Firstly, we analyzed the methylation value of the 11 subjects at three time points (n = 33), which demonstrated a very strong correlation between the Infinium MethylationEPIC BeadChip and pyrosequencing technology (Fig. 5a, b, Pearson correlation coefficient = 0.91, $p < 0.001$). Secondly, we evaluated the methylation status of the same CpGs in the peripheral blood of an independent longitudinal cohort of 9 children measured at birth and 5 years of age (n = 18), along with another independent longitudinal cohort of 9 different children measured at 5 and 10 years of age (n = 18) (Fig. 5c, see Additional file 1: Table S9 for the clinical information relating to these cohorts). This biological validation showed that the DNA methylation

alterations identified can be reproduced in independent cohorts, thus underscoring the significance and robustness of the methylation changes found in this study.

Discussion

In the present study we examined the genome-wide methylation profile of 33 longitudinal blood samples from 11 children at 0 (newborn), 5 and 10 years of age. Firstly, we found that extensive DNA methylation changes occur during the first 5 years of life, while much less epigenetic remodeling takes place in the following 5 years. As many as 110,726 CpG sites were found to have altered DNA methylation values when comparing newborn and 5 year-old samples, with a slight tendency toward an increase in methylation with age (54% of dmCpGs), while only 460 CpGs exhibited significant methylation changes between 5 and 10 years of age, 72% of which lost methylation. These findings are in line with the current literature which describes early-years epigenetic changes as being the most important of the postnatal period [8]. What is more, our three time point study design allowed us to directly compare the alterations occurring during two different early-years intervals, demonstrating that after 5 years, epigenetic reshaping is dramatically reduced. In the same vein, Acevedo and colleagues in their reporting of longitudinal DNA methylation changes during the first 5 years of life noted that in the third year these already seem to be weakening [19]. Aside from being more numerous, we found DNA methylation changes to be far more pronounced in the first lustrum of life than the subsequent one, which may have importance in terms of the functional influence of these methylation alterations on gene expression. Moreover, this observation did not seem to be related to considerable changes in cell-type composition occurring in the first 5 years as compared to the following 5.

We next examined the genomic distribution of the dmCpGs found, which revealed that hyper- and hypomethylation changes occur at very different locations: loss of methylation being observed at CpG-poor regions such as open sea locations, including intronic and intergenic regions, for both $0 \rightarrow 5$ and $5 \rightarrow 10$ changes, while methylation gain occurred at CpG-denser regions like CpG islands and gene promoters for both of the age intervals examined (although $0 \rightarrow 5$ changes were more similar to array distribution). These findings, with exceptions [19] are in line with most studies describing age-related DNA methylation changes, both in the early years [8, 16] and later years of aging [44, 45]. The fact that the great majority of hypomethylation changes were found at open sea regions underscores the value of screening technologies that examine CpG-sparse regions to accurately characterize DNA hypomethylation scenarios.

After identifying the genomic distribution of the age-related methylation changes, we sought to characterize the functional contexts associated with these loci. To achieve this, we integrated our DNA methylation data with external datasets describing histone modifications, chromatin states and enhancers for different tissue and cell types. These analyses primarily considered $0 \rightarrow 5$ dmCpGs because of the low number of $5 \rightarrow 10$ dmCpGs detected, which hindered the enrichments. The results evidenced a link between DNA hypermethylation and the repressive histone mark H3K27me3, and with chromatin states related to polycomb repressive domains and bivalent chromatin enhancers, while DNA hypomethylation occurred mainly at H3K4me1 regions and intragenic enhancers, as has been previously shown [46, 47]. Furthermore, when mapping the methylation changes to an enhancer library of 65 different tissues, we found that hypermethylation displayed a much lower overall enrichment at enhancer sites than hypomethylation, and while the former occurred mainly at enhancers in normal and embryonic tissue, the latter was particularly enriched in tissues and cell types related to blood. These observations suggest differences in the functionality of the enhancer-associated methylation changes that occur during the first 5 years of life depending on whether there is a gain or loss of methylation. Moreover, as studies are increasingly describing enhancer-associated DNA methylation as the main expression-correlated methylation phenomenon [20, 48, 49], our results point towards the enhancer-enriched DNA hypomethylation observed perhaps being of more importance in the control of gene expression than DNA hypermethylation. Indeed, enhancer methylation is in general negatively correlated to their activity [50], and thus it would make sense that the epigenetic changes that occur during the final stages of development are accompanied by the inactivation of general developmental enhancer elements, while tissue-specific (blood) enhancers become activated. In this setting, our data suggest that the early-life DNA methylation loss enriched at tissue-specific enhancer regions observed in our data could reflect the establishment of epigenetic active patterns defining tissue function during this period of life, while enhancers associated to development are already epigenetically repressed and suffer less important (hypermethylation) changes.

Another approach to studying the functional importance of DNA methylation changes is through the mapping of dmCpG sites to genes and then analyzing the related ontologies. Given that DNA methylation is irregularly distributed throughout the different parts of a gene [1], and because a great proportion of our mapped genes

contained both hyper- and hypomethylated dmCpGs, we only performed the ontology analyses, firstly, on genes that exclusively either gained or lost methylation. We found that 0 → 5 DNA hypermethylation changes were principally related to developmental functions and cell-to-cell signaling, while hypomethylation was primarily linked to mRNA and protein metabolism, immune response and mitosis. It is worth mentioning that hypermethylation was also associated with terms such as growth, reproduction and response to retinoic acid, a molecule involved in organogenesis [51] as well as hematopoiesis [52]. Significantly, when we considered the ontologies of genes containing both hyper- and hypomethylated probes, we found considerable differences, implying that these genes could play different functional roles to those that are exclusively hyper- or hypomethylated. Moreover, when we examined the ontologies without making these distinctions, we found that, while hypermethylation ontologies did not considerably change, hypomethylation terms shifted from mRNA and protein metabolism to cellular localization and activation or GTPase activity (maintaining immunological terms). These latter findings are more in line with those reported in previous works which did not describe separating exclusively hyper- and hypomethylated genes [8, 14, 16, 19]. However, the fact that we found a change in function suggests that not taking into consideration the presence of concurrently hyper- and hypomethylated genes could affect the conclusions drawn from gene ontology analyses. On the whole, these results suggest that it is not only differentially hyper- or hypomethylated enhancers that have differing roles, but also differentially hyper- or hypomethylated genes, with the former being more related to general developmental functions, and the latter being involved in, among other things, immunological functions, thus reflecting the functionality differences observed for enhancer elements. In a final analysis, we further segregated the 0 → 5 dmCpGs into groups depending on their gene location (promoter, gene body, intron, exon) and also found changes between the gene ontologies of genes containing promoter- or exon-associated dmCpGs versus intronic- or gene body-associated dmCpGs, both for hyper- and hypomethylated CpGs. These observations imply that different genes, associated to different functions, suffer DNA methylation changes in different regions, which could help explain why DNA methylation changes in different contexts can have different consequences, and also suggests that the response of the gene elements to epigenetic changes during early life is gene region-specific.

Finally, we compared the DNA methylation changes that occur during the first two lustra of life, finding that the majority of the 5 → 10 changes are in fact a continuation of the 0 → 5 changes. By looking at the dmCpGs with the most substantial change in both age groups we defined 36 CpG sites with consistent DNA methylation changes during the first 10 years of life. These locations followed one of two trends: (1) strong 0 → 5 changes followed by weak 5 → 10 changes or (2) moderate overall 0 → 5 → 10 changes. Many of these CpGs were located at genes with important functions, such as development-associated *HOXB7* [53], which contained 2 CpGs, and the GATA-interacting *ZFPM2* [54]. Interestingly, although both these genes are related to developmental functions, the first was found to be hypermethylated while the second was hypomethylated, and, what is more, the dmCpG associated with *ZFPM2* was also mapped to an enhancer element (Fig. 4c). Another zinc finger-family gene which we found to be altered, *ZNF385D*, has been associated to language impairment and reading disability in children [55]. Although this observation could be related to its function during prenatal brain development, the fact that the gene-associated dmCpG (cg02920129) shows a similar trend of methylation gain during both the 0 → 5 and 5 → 10 periods of life (Additional file 12: Figure S4) points towards a possible functional relevance throughout early life. It would thus be of interest to further study the link between the methylation status of *ZNF385D* and language impairment or reading disability. The *CD200R1* gene, which encodes for a myeloid- and T-cell distinctive transmembrane receptor [56] was found to be linked to an hypomethylated dmCpG (cg07061387), which was subsequently found to be associated with enhancer elements related to up to 3 different genes (*ATG3*, *NEPRO*, *GCSAM*) when the 0 → 5 → 10 dmCpGs were mapped to enhancers in different blood tissues, indicating that at least some of the DNA methylation changes found could exert an influence on genes other than the those on which the dmCpGs are located.

Lastly, we performed technical and biological validations of the methylation status of 3 dmCpGs located on the *HOXB7*, the *SOCS3* and the *ZFPM2* gene using bisulfite pyrosequencing. As expected, we observed a high correlation between the results of the Infinium MethylationEPIC BeadChip analysis and the pyrosequencing, albeit the DNA methylation values detected by the array were slightly higher, a fact which has been noted before [57]. Subsequently we examined the methylation of these CpGs in two independent longitudinal cohorts spanning the first and second lustra of life, and found that the DNA methylation changes found in the previous experiments were robust and reproducible in independent subjects.

Conclusions

Collectively our results show that major epigenetic remodeling takes place during the first 5 years of life. However, despite a dramatic reduction, some genomic *loci* continue to experience considerable DNA methylation changes during the second lustrum of life. This observation illustrates that there are CpG sites whose change in methylation status could have functional roles beyond early postnatal development. In addition, we show that genes that experience simultaneous hyper- and hypomethylation are functionally distinct from those that are exclusively hyper- or hypomethylated. Finally, we have found that enhancer-associated methylation is different in the hyper- and hypomethylation scenario, the latter being more relevant and related to blood-specific enhancer elements, while developmental enhancer elements could already be epigenetically defined at birth, suffering changes of lesser importance during early life.

Additional files

Additional file 1: Table S9. Clinical characteristics of the 18 subjects from the independent cohorts.

Additional file 2: Table S8. Primer sequences for bisulfite sequencing of the validated CpGs.

Additional file 3: Table S1. Lists of annotated 0→5 and 5→10 dmCpGs.

Additional file 4: Table S6. Gene ontology enrichment analysis of 0→5 and 5→10 dmCpGs. Enrichments were calculated from the difference between the dmCpGs obtained in each of the analyses and the GO ontology database by the R/Bioconductor package missMethyl. Ontologies for genes with 1) both hyper- and hypomethylated probes 2) exclusively either hyper- or hypomethylated and 3) from mapped hyper- or hypomethylated dmCpGs which did not take into account CpGs differentially methylated in the opposite direction in the same genes are included. Also included are the ontologies found when analyzing 0→5 dmCpGs grouped by gene region (promoter, exon, intron and gene body). All of the analyses include "molecular function", "cellular component" and "biological process" terms.

Additional file 5: Figure S1. Boxplots indicating the distribution of absolute beta values of the DNA methylation changes for 0→5 and 5→10 hyper- and hypomethylated dmCpGs. Effect size is measured as median difference and Cliff's delta.

Additional file 6: Table S2. Blood cell-type compositions as predicted by the Houseman algorithm for the 33 samples.

Additional file 7: Table S3. Histone mark enrichment analysis of 0→5 and 5→10 dmCpGs. Enrichments were calculated based on differences between the dmCpGs obtained in each of the analyses and the full collection of Roadmap epigenomics hg19 regions integrated in the LOLA extended software. Corresponding array backgrounds were used for the different comparisons.

Additional file 8: Table S4. Chromatin state enrichment analysis of 0→5 and 5→10 dmCpGs. Enrichments were calculated based on differences between the dmCpGs obtained in each of the analyses and the hg19 chromatin segmentation regions (ChromHMM, 15 states) obtained from Roadmap and ENCODE consortia. A custom LOLA database including information related to the chromatin states in the different tissues/cell lines and corresponding array backgrounds were used for the correct enrichment calculation.

Additional file 9: Table S5. Enhancer element enrichment analysis of 0→5 and 5→10 dmCpGs. Enrichments were calculated from the difference between the dmCpGs obtained in each of the analyses and the enhancer elements obtained from the EnhancerAtlas database. A customized LOLA database which included information related to the enhancers in the different tissues/cell lines and corresponding array backgrounds were used for the appropriate enrichment calculation.

Additional file 10: Figure S2. a) Treemap plots indicating the results of REVIGO semantic analyses of significantly enriched (FDR < 0.05) gene ontology biological process terms for genes that simultaneously contained 0→5 hyper- and hypomethylated dmCpGs. In total, 460 significant terms were found (see Table S6 for full results, including Molecular Function and Cellular Component terms, and also 5→10 dmCpG enrichments). b) Equivalent plots for genes containing, respectively, hyper- or hypomethylated dmCpGs, irrespective of those same genes also containing dmCpGs that changed in the opposite direction.

Additional file 11: Figure S3. Treemap plots indicating the results of REVIGO semantic analyses of significantly enriched (FDR < 0.05) gene ontology biological process terms for genes containing (a) hyper- and (b) hypomethylated 0→5 dmCpGs. The dmCpGs are grouped by annotated genomic location (see Table S1, column "annotation", for the annotations, "promoter" is formed by collapsing "Distal promoter" and "Promoter (<= 1kb)", "gene body" is formed by collapsing "3'UTR", "5'UTR", "Intron", "Exon" and "Downstream"). See Table S6 for full results, including Molecular Function and Cellular Component terms.

Additional file 12: Figure S4. Boxplots showing the DNA methylation beta values of the 36 common 0→5 → 10 dmCpGs described in Table 2.

Additional file 13: Table S7. DNA methylation values obtained by bisulfite sequencing for 3 CpGs across the original 33 subjects (technical validation) and two independent longitudinal cohorts (biological validation).

Abbreviations

ATG3: autophagy related 3; BF: breast feeding; BMIQ: beta-mixture quantile normalization; CD: Cliff's delta; CS: cesarean section; dmCpGs: differentially methylated CpG; EHM: expected hypergeometric mean; ENCODE: encyclopedia of DNA elements; FDR: false discovery rate; FF: formula feeding; GCSAM: germinal center-associated signaling and motility protein; HDL: high-density lipoprotein; HMM: hidden Markov model; HOXB7: homeobox protein Hox-B7; IDAT: intensity data file; LDL: low-density lipoprotein; MD: median difference; MDS: multidimensional scaling; NEPRO: nucleolus and neural progenitor protein; NIH: National Institutes of Health; OR: odds ratio; PCA: principal component analysis; PCR: polymerase chain reaction; SGA/AGA/LGA: small/appropriate/large for gestational age; SOCS3: suppressor of cytokine signaling 3; VD: vaginal delivery; ZFPM2: zinc finger protein, FOG family member 2; ZNF385D: zinc finger protein 385D

Authors' contributions

RFP, PS, JRT, RGU, AFF and MFF carried out the methylation studies and participated in drafting the manuscript. RFP and JRT performed the statistical analyses. PS and RFP performed the pyrosequencing analyses. EL and MFF conceived of the study and participated in its design and coordination of the manuscript. PR performed critical revision of the manuscript. JA participated in collecting and qualifying mothers and children. All authors read and approved the final manuscript.

Author details

[1] Cancer Epigenetics Laboratory, Institute of Oncology of Asturias (IUOPA)-Instituto de Investigación Sanitaria del Principado de Asturias (ISPA)-Hospital Universitario Central de Asturias (HUCA), 33011 Oviedo, Asturias, Spain. [2] Nanomedicine Group, Nanomaterials and Nanotechnology Research Center (CINN-CSIC), Universidad de Oviedo, 33940 Oviedo, Asturias, Spain. [3] Servicio de Pediatría, Consorcio Hospital General Universitario de Valencia, 46014 Valencia, Spain. [4] Centros de Investigación Biomédica en Red de

Fisiopatología Obesidad y Nutrición (CB06/03), Instituto de Salud Carlos III, Madrid, Spain.

Acknowledgements
The authors are grateful to the members from the Cancer Epigenetics laboratory (FINBA, ISPA, IUOPA) for their positive feedback, to Ronnie Lendrum for manuscript editing and to Feliciano Jesús Ramos Fuentes.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
All data generated during this study are included in this published article and its supplementary information files. The raw IDAT and preprocessed data are also available in the ArrayExpress public repository under accession E-MTAB-7069.

Consent to publish
All parents gave informed consent for their children to participate in the study.

Ethics approval and consent to participate
The Committee for the Protection of Human Subjects of the Hospital General approved the study according to the Declaration of Helsinki.

Funding
This work has been financially supported by: Ministerio de Ciencia, Innovación y Universidades (Grant Numbers PI14/01781 and PI17/00144), Instituto de Salud Carlos III and cofinanced by the European Regional Development Fund (to RL); the Plan Nacional de I+D+I 2013–2016/FEDER (PI15/00893 to MFF and AFF); IUOPA-ISPA-FINBA (to RGU, RFP and PS); AF Fernández is supported by a Miguel Servet II fellowship (contract CPII16/00007); JR Tejedor is supported by the Ministry of Economy and Competitiveness through a Juan de la Cierva postdoctoral fellowship (IJCI-2015-26965); the IUOPA is supported by the Obra Social Cajastur Liberbank, Spain.

Publisher's Note

Received: 14 December 2018 Accepted: 17 December 2018
Published online: 09 January 2019

References

1. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13:484–92.
2. Tang WWC, Kobayashi T, Irie N, Dietmann S, Surani MA. Specification and epigenetic programming of the human germ line. Nat Rev Genet. 2016;17:585–600.
3. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010;28:1057–68.
4. Huidobro C, Fernández AF, Fraga MF. Aging epigenetics: causes and consequences. Mol Aspects Med. 2013;34:765–81.
5. Atsuta Y, Stunnenberg HG. The interplay of epigenetic marks during stem cell differentiation and development. Nat Rev Genet. 2017;18:643–58.
6. Benayoun BA, Pollina EA, Brunet A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. Nat Rev Mol Cell Biol. 2015;16:593–610.
7. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018;19:371–84.
8. Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, et al. Age-associated DNA methylation in pediatric populations. Genome Res. 2012;22:623–32.
9. Martino D, Loke YJ, Gordon L, Ollikainen M, Cruickshank MN, Saffery R, et al. Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. Genome Biol. 2013;14:R42.
10. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet. 2012;13:97–109.
11. Issa J. Aging and epigenetic drift: a vicious cycle. J Clin Invest. 2014;124:24–9.
12. Tejedor JR, Fraga MF. Interindividual epigenetic variability: Sound or noise? BioEssays News Rev Mol Cell Dev Biol. 2017;39:1700055.
13. Wang D, Liu X, Zhou Y, Xie H, Hong X, Tsai H-J, et al. Individual variation and longitudinal pattern of genome-wide DNA methylation from birth to the first two years of life. Epigenetics. 2012;7:594–605.
14. Cruickshank MN, Oshlack A, Theda C, Davis PG, Martino D, Sheehan P, et al. Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. Genome Med. 2013;5:96.
15. Florath I, Butterbach K, Müller H, Bewerunge-Hudler M, Brenner H. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. Hum Mol Genet. 2014;23:1186–201.
16. Urdinguio RG, Torró MI, Bayón GF, Álvarez-Pitti J, Fernández AF, Redon P, et al. Longitudinal study of DNA methylation during the first 5 years of life. J Transl Med. 2016;14:160.
17. Simpkin AJ, Suderman M, Gaunt TR, Lyttleton O, McArdle WL, Ring SM, et al. Longitudinal analysis of DNA methylation associated with birth weight and gestational age. Hum Mol Genet. 2015;24:3752–63.
18. Martino DJ, Tulic MK, Gordon L, Hodder M, Richman TR, Metcalfe J, et al. Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. Epigenetics. 2011;6:1085–94.
19. Acevedo N, Reinius LE, Vitezic M, Fortino V, Söderhäll C, Honkanen H, et al. Age-associated DNA methylation changes in immune genes, histone modifiers and chromatin remodeling factors within 5 years after birth in human blood leukocytes. Clin Epigenetics. 2015;7:34.
20. Gutierrez-Arcelus M, Ongen H, Lappalainen T, Montgomery SB, Buil A, Yurovsky A, et al. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. PLoS Genet. 2015;11:e1004958.
21. Ballard JL, Novak KK, Driver M. A simplified score for assessment of fetal maturation of newly born infants. J Pediatr. 1979;95:769–74.
22. Battaglia FC, Lubchenco LO. A practical classification of newborn infants by weight and gestational age. J Pediatr. 1967;71:159–63.
23. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.
24. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17:208.
25. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8:203–9.
26. Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA methylation BeadArrays. Nucleic Acids Res. 2013;41:e90.
27. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29:189–96.
28. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k chip analysis methylation pipeline. Bioinformatics. 2014;30:428–30.
29. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008;24:1547–8.
30. Xu Z, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. Nucleic Acids Res. 2016;44:e20.
31. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinform. 2012;13:86.
32. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007;3:1724–35.
33. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28:882–3.

34. Teschendorff AE, Zheng SC. Cell type deconvolution in epigenome wide association studies: a review and recommendations. Epigenomics. 2017;9:757–68.
35. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.
36. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinform. 2010;11:587.
37. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinforma Oxf Engl. 2015;31:2382–3.
38. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinforma Oxf Engl. 2016;32:587–9.
39. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28:1045–8.
40. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.
41. Gao T, He B, Liu S, Zhu H, Tan K, Qian J. EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. Bioinforma Oxf Engl. 2016;32:3543–51.
42. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinforma Oxf Engl. 2016;32:286–8.
43. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE. 2011;6:e21800.
44. Day K, Waite LL, Thalacker-Mercer A, West A, Bamman MM, Brooks JD, et al. Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. Genome Biol. 2013;14:R102.
45. Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, Teschendorff AE. An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. PLoS Genet. 2015;11:e1004996.
46. Fernández AF, Bayón GF, Urdinguio RG, Toraño EG, García MG, Carella A, et al. H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells. Genome Res. 2015;25:27–40.
47. Pérez RF, Tejedor JR, Bayón GF, Fernández AF, Fraga MF. Distinct chromatin signatures of DNA hypomethylation in aging and cancer. Aging Cell. 2018;17:e12744.
48. Aran D, Hellman A. DNA methylation of transcriptional enhancers and cancer predisposition. Cell. 2013;154:11–3.
49. Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol. 2013;14:R21.
50. Calo E, Wysocka J. Modification of enhancer chromatin: what, how and why? Mol Cell. 2013;49:825–37.
51. Duester G. Retinoic acid synthesis and signaling during early organogenesis. Cell. 2008;134:921–31.
52. Collins SJ. The role of retinoids and retinoic acid receptors in normal hematopoiesis. Leukemia. 2002;16:1896–905.
53. Cantile M, Schiavo G, Terracciano L, Cillo C. Homeobox genes in normal and abnormal vasculogenesis. Nutr Metab Cardiovasc Dis NMCD. 2008;18:651–6.
54. Cantor AB, Orkin SH. Coregulation of GATA factors by the Friend of GATA (FOG) family of multitype zinc finger proteins. Semin Cell Dev Biol. 2005;16:117–28.
55. Eicher JD, Powers NR, Miller LL, Akshoomoff N, Amaral DG, Bloss CS, et al. Genome-wide association study of shared components of reading disability and language impairment. Genes Brain Behav. 2013;12:792–801.
56. Vaine CA, Soberman R. The CD200-CD200R1 inhibitory signaling pathway: immune regulation and host-pathogen interactions. Adv Immunol. 2014;121:191–211.
57. BLUEPRINT consortium. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. Nat Biotechnol. 2016;34:726–37.

AGING AND OTHER DISEASES: FIFTH ARTICLE

<u>Pérez RF</u>, Alba-Linares JJ, Tejedor JR, Fernández AF, Calero M, Román-Domínguez A, Borrás C, Viña J, Ávila J, Medina M & Fraga MF (2022) Blood DNA methylation patterns in older adults with evolving dementia. J Gerontol A Biol Sci Med Sci, 77(9): 1743–1749.

In this work, our goal was to characterize the epigenetic features of cognitive deterioration in a human model of aging. To this end, we compiled blood samples from a longitudinal cohort of elderly individuals (>70 years) some of whom developed dementia in a span of 4 years. We profiled DNA methylation levels at more than 700,000 CpG sites using Infinium MethylationEPIC arrays. We discovered DNA methylation biomarkers associated with cognitive deterioration, and, importantly, we observed that most of these alterations were detectable in healthy individuals prior to the development of overt clinical symptoms. Several biomarkers were related to dementia-associated genes, and we were able to validate various region-level alterations by integrating our data with external cohorts. Our results show that systemic tissues such as blood incorporate epigenetic information predictive of the appearance of dementia and cognitive decline, and thus DNA methylation may be useful for the development of clinical biomarkers for the prediction and early treatment of these conditions, as well as for understanding the link between aging and aging-associated health deterioration.

Personal contribution to the work: The samples and clinical information were compiled by different labs associated with the Vallecas Project (Drs. Miguel Medina, Jesús Ávila, José Viña and Miguel Calero). I generated, preprocessed and analyzed all of the data presented in this manuscript, including the pyrosequencing experiments, as well as prepared the figure panels and wrote the manuscript. The work was supervised by Dr. Mario F. Fraga and by the aforementioned coauthors.

THE GERONTOLOGICAL SOCIETY OF AMERICA®

Journals of Gerontology: Biological Sciences
cite as: *J Gerontol A Biol Sci Med Sci*, 2022, Vol. 77, No. 9, 1743–1749
https://doi.org/10.1093/gerona/glac068
Advance Access publication March 17, 2022

OXFORD

Research Report

Blood DNA Methylation Patterns in Older Adults With Evolving Dementia

Raúl Fernández Pérez, MSc,[1,2,3,4] Juan José Alba-Linares, MSc,[1,2,3,4] Juan Ramón Tejedor, PhD,[1,2,3,4] Agustín Fernández Fernández, PhD,[1,2,3,4] Miguel Calero, PhD,[5,6,7] Aurora Román-Domínguez, MSc,[8] Consuelo Borrás, PhD,[8,©] José Viña, PhD,[8] Jesús Ávila, PhD,[5,9] Miguel Medina, PhD,[5,7,*,†] and Mario Fernández Fraga, PhD[1,2,3,4,*,©]

[1]Cancer Epigenetics and Nanomedicine Laboratory, Nanomaterials and Nanotechnology Research Center (CINN-CSIC), El Entrego, Spain. [2]Health Research Institute of Asturias (ISPA-FINBA), University of Oviedo, Oviedo, Spain. [3]Institute of Oncology of Asturias (IUOPA) and Department of Organisms and Systems Biology (B.O.S.), University of Oviedo, Oviedo, Spain. [4]Rare Diseases CIBER (CIBERER) of the Carlos III Health Institute (ISCIII), Madrid, Spain. [5]Network Center for Biomedical Research in Neurodegenerative Diseases (CIBERNED), Madrid, Spain. [6]Chronic Disease Programme (UFIEC), Instituto de Salud Carlos III, Madrid, Spain. [7]CIEN Foundation, Queen Sofia Foundation Alzheimer Center, Madrid, Spain. [8]Freshage Research Group, Department of Physiology, Faculty of Medicine, University of Valencia and CIBERFES-ISCIII, Fundación Investigación Hospital Clínico Universitario/INCLIVA, Valencia, Spain. [9]Centro de Biología Molecular Severo Ochoa (CBMSO) CSIC-UAM, Madrid, Spain.

*Address correspondence to: Mario Fernández Fraga, PhD, Cancer Epigenetics and Nanomedicine Laboratory, Nanomaterials and Nanotechnology Research Center (CINN-CSIC), Av. la Vega, 4, 6, 33940 El Entrego, Spain. E-mail: mffraga@cinn.es; or Miguel Medina, PhD, CIEN Foundation, Queen Sofia Foundation Alzheimer Center, C. de Valderrebollo, 5, 28031 Madrid, Spain. E-mail: mmedina@ciberned.es.

†Co-corresponding author.

Received: August 20, 2021; Editorial Decision Date: March 12, 2022

Decision Editor: David Le Couteur, MBBS, FRACP, PhD

Abstract

Dementia and cognitive disorders are major aging-associated pathologies. The prevalence and severity of these conditions are influenced by both genetic and environmental factors. Reflecting this, epigenetic alterations have been associated with each of these processes, especially at the level of DNA methylation, and such changes may help explain the observed interindividual variability in the development of the 2 pathologies. However, the importance of epigenetic alterations in explaining their etiology is unclear because little is known about the timing of when they appear. Here, using Illumina MethylationEPIC arrays, we have longitudinally analyzed the peripheral blood methylomes of cognitively healthy older adults (>70 year), some of whom went on to develop dementia while others stayed healthy. We have characterized 34 individuals at the prediagnosis stage and at a 4-year follow-up in the postdiagnosis stage (total $n = 68$). Our results show multiple DNA methylation alterations linked to dementia status, particularly at the level of differentially methylated regions. These loci are associated with several dementia-related genes, including *PON1*, *AP2A2*, *MAGI2*, *POT1*, *ITGAX*, *PACSIN1*, *SLC2A8*, and *EIF4E*. We also provide validation of the previously reported epigenetic alteration of *HOXB6* and *PM20D1*. Importantly, we show that most of these regions are already altered in the prediagnosis stage of individuals who go on to develop dementia. In conclusion, our observations suggest that dementia-associated epigenetic patterns that have specific biological features are already present before diagnosis, and thus may be important in the design of epigenetic biomarkers for disease detection based on peripheral tissues.

Keywords: Cognitive decline, Dementia, DNA methylation, Epigenetics, Epigenetic age

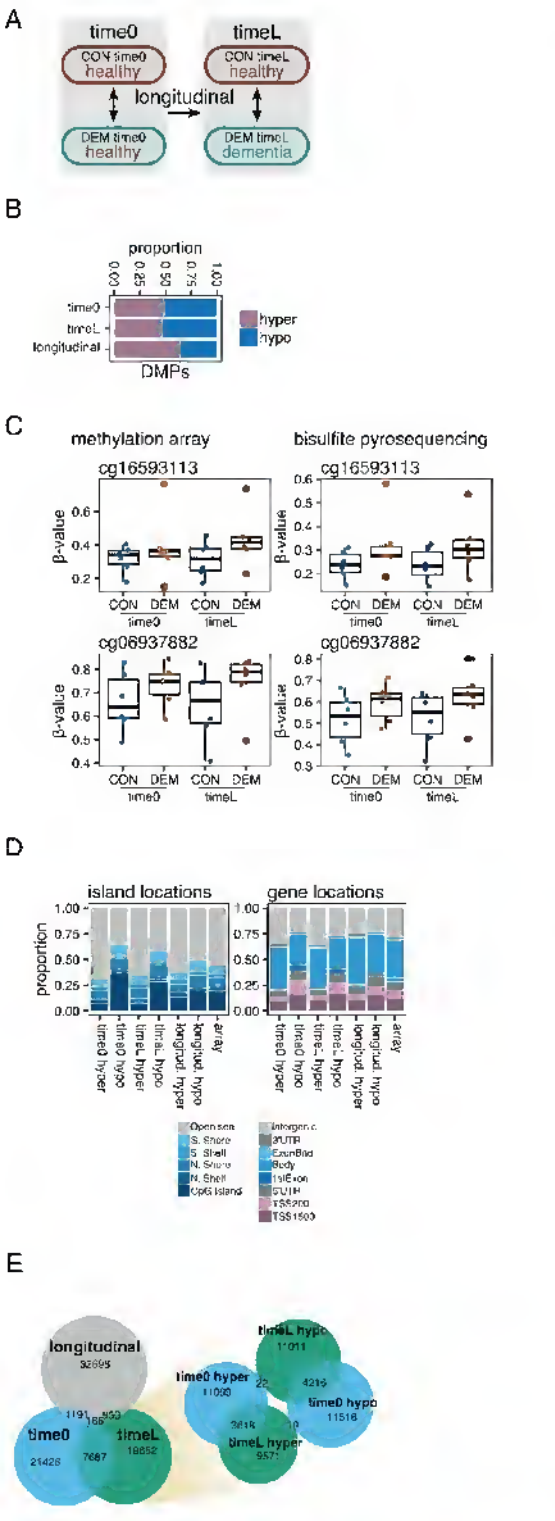


Figure 1. DNA methylation alterations at single-CpG sites in dementia. (A) Schematic of the study design. (B) Barplots depicting the proportion of hyper- and hypo-methylated DMPs (unadjusted $p < .05$) found in the time0, timeL, and longitudinal comparisons. (C) Boxplots comparing the DNA methylation measurements performed by the Infinium MethylationEPIC BeadChip and by bisulfite pyrosequencing for 2 CpGs (cg16593113, cg06937882) on a subset of 24 samples segregated by experimental group. (D) Barplots showing the relative distribution of hyper- and hypo-methylated DMPs in the time0, timeL, and longitudinal comparisons according to their CpG island location status (top) and gene location status (bottom). The rightmost bars reflect the background distribution of all the analyzed array probes. (E) On the left, the Venn diagram describes the numbers and intersections of the DMPs found for the time0, timeL, and longitudinal comparisons. On the right, the Venn diagram shows the specific intersections between hyper- and hypo-methylated DMPs from the time0 to timeL comparisons. CON = control; DEM = dementia; DMPs = differentially methylated probes.

Background

Cognitive decline and dementia are complex diseases in which both genetic and environmental factors play a relevant role (1,2). The well-known preclinical phenotypes of dementia (3) serve to demonstrate that these pathologies are defined by progressive changes whose timely detection is crucial in the management of the disease. Indeed, dementia is strongly associated with aging, although the causal relationships between the overlap in cognitive decline symptoms observed for the 2 processes remain to be clarified (4). Both aging and dementia have been associated with epigenetic alterations, and these molecular mechanisms may contribute to characterizing their relationship (5). During aging, both genetic factors and the accumulation of external stimuli, such as those related to lifestyle, can trigger epigenetic changes which may help explain: (a) the variability in the trajectories of cognitive decline experienced by "healthy" aging individuals (6) and (b) the variability in the appearance of pathological states such as mild cognitive impairment or dementia (5).

Among the known epigenetic changes, DNA methylation alterations have been found to be associated with dementia both in brain and in systemic tissues such as peripheral blood (7). In addition, the recently developed DNA methylation clocks, which are often altered in disease, can serve as proxies that encompass the complex factors (genetic, biological, and environmental) which lead to interindividual differences in phenotype and are thus of great interest in the definition of potential biomarkers of disease (8). Nonetheless, the question still remains as to whether these epigenetic alterations arise prior to or as a consequence of dementia. If the former, they could serve as biological indicators and/or provide novel avenues for interventions to prevent these diseases. Within this scenario, longitudinal studies are of great value in tracing variables that contribute to explaining these phenotypes (9).

Method

Here, we have profiled the peripheral blood mononuclear cell epigenome of 68 samples at more than 770 000 CpG sites by employing Infinium MethylationEPIC BeadChips. We studied a longitudinal cohort of older adults consisting of 17 pairs of age-matched, cognitively healthy individuals where in a 4-year follow-up assessment ($SD = 0.35$ year), one was still cognitively healthy, that is acted as control (CON), while the other had been diagnosed with dementia (DEM, also referred to henceforth as "converter" individuals; Figure 1a, Table 1; Supplementary Table 1 for extended phenotypic data). Subjects are volunteer participants in an ongoing single-center longitudinal study known as "The Vallecas Project" where they annually undergo extensive neurological and neuropsychological assessment (10). We characterized the methylomes of these individuals at an initial, prediagnosis time point, when all were cognitively healthy (time0; CON_time0, and DEM_time0 groups, $n = 34$) and at a longitudinal, postdiagnosis time point at which some had converted to dementia (timeL; CON_timeL, and DEM_timeL groups,

Table 1. Summary of Clinical Information Related to the Subjects. Subjects at Time0 Are all Cognitively Healthy and Grouped Into Stable Controls (CON_time0) or Future Converters to Dementia (DEM_time0)

Time Point	time0, $n = 34$			timeL, $n = 34$		
Group	Stable Control (CON_time0)	Dementia Converter (DEM_time0)	p Value	Stable Control (CON_timeL)	Dementia Converter (DEM_timeL)	p Value
Number of subjects	17	17		17	17	
Sex (M/F)	8/9	3/14	n.s.	8/9	3/14	n.s.
Age, mean yr (*SD*)	76.1 (2.8)	76.6 (4.1)	n.s.	80.1 (2.8)	80.6 (4.2)	n.s.
MMSE, mean (*SD*)	28.6 (1.5)	27.2 (2.4)	n.s.	28.5 (2.0)	21.2 (4.5)	***
FAQ, mean (*SD*)	0.4 (0.5)	0.9 (1.1)	n.s.	0.5 (0.8)	13.5 (8.5)	***
GDS, mean (*SD*)	1.2 (1.4)	1.8 (1.5)	n.s.	1.4 (2.0)	2.9 (2.0)	*
CDR, mean (*SD*)	0.0 (0.0)	0.0 (0.1)	n.s.	0.0 (0.1)	1.1 (0.3)	***

Notes: CDR = Clinical Dementia Rating; FAQ = Functional Activities Questionnaire; GDS = Geriatric Depression Scale; MMSE = Mini-Mental State Examination; *SD* = Standard Deviation. The same subjects are evaluated at timeL, when stable controls remain cognitively healthy (CON_timeL) while converters manifest the disease (DEM_timeL).

n.s., $p \geq .05$, *$p < .05$, ***$p < .001$ for Wilcoxon rank sum or chi-squared tests.

$n = 34$). This allowed us to examine: (a) DNA methylation alterations predictive of the appearance of cognitive pathology (time0 comparison), (b) DNA methylation alterations directly associated with cognitive pathology (timeL comparison), and (c) longitudinal DNA methylation alterations (the full methodology is detailed in Supplementary Methods).

Results

We used empirical Bayes modified t tests in a linear model framework (11) to define differentially methylated probes of CpG sites (DMPs; false discovery rate [FDR] < 0.05) across the different comparisons. The models were adjusted to account for experimental processing batch, blood cell-type composition, sex, and subject-specific effects, with cell-type composition being predicted from the DNA methylation data using the Houseman algorithm (12) (Supplementary Methods). We employed variance decomposition methods to determine the potential effects of experimental or technical variables in our data. A surrogate variable analysis (13) confirmed that the 2 main variables driving latent variation in our data were batch and cell-type composition (principally CD8+T cells; Supplementary Figure 1). After carrying out the differential analyses, we found no statistically significant DMPs between control and converter individuals at either time0 or at timeL, while we did detect 14 DMPs in the longitudinal comparison (described in Supplementary Table 2). However, because the longitudinal comparison involves all subjects (34 at time0 vs 34 at timeL) while the dementia comparisons only involve half of the cohort (17 vs 17), the observed differences in detected DMPs could be due to an increase in statistical power. We performed subsampling of the cohort to retain only 17 individuals and repeated the longitudinal comparisons, finding no statistically significant DMPs across 5 iterations. These initial results suggest that, at the level of individual CpG sites, there are no marked DNA methylation alterations that are predictive of the development of dementia, or directly associated with this disease in the blood of older adults. We also found no evidence of an increase in DNA methylation differences between individuals after onset of symptoms in the DEM group as compared to differences at the prediagnosis stage.

To expand the biological exploration of our data, we next focused on the top probes for each comparison (unadjusted $p < .05$). These corresponded to 30 492 loci, 28 457 loci, and 35 007 loci, respectively, for the time0, timeL, and longitudinal comparisons (Figure 1b). These borderline CpG sites may collectively carry biological insight and, moreover, we validated 2 sites with moderate ($p < .001$, cg16593113) and marginal ($p = .06$, cg06937882) significance in the time0 and timeL comparisons by using bisulfite pyrosequencing in a subset of the samples (Figure 1c; Supplementary Table 3 for primer information), indicating that the array produced robust measurements. Indeed, the array and pyrosequencing measurements were highly concordant across all observations (Pearson correlation coefficient = 0.98, Supplementary Figure 2).

The time0-, timeL-, and longitudinal-DMPs were each associated with specific distributions across CpG island and gene locations (Figure 1d), with parallel hyper- or hypomethylation-specific trends being observed for all 3 comparisons. We analyzed the intersections between the sets of DMPs (Figure 1e, left plot) and found a strong enrichment in shared time0- and timeL-DMPs (Fisher's test $p < .001$, odds ratio [OR] = 12). Moreover, the direction of the dementia-associated alterations was maintained at both time points (Figure 1e, right plot). When looking specifically at the DMPs common to both time0 and timeL that had a concordant direction of change (7 834 out of 7 875), we found no evidence of an increase in the magnitude of change at postdiagnosis (timeL) with respect to prediagnosis (time0; Wilcoxon rank sum test $p = .257$; Supplementary Figure 3). These results suggest that blood dementia-associated DNA methylation patterns are very similar at the prediagnosis stage and after the onset of dementia symptoms, and that these loci are different from those associated with longitudinal drift. We also performed Gene Ontology enrichment analyses on the sets of DMPs (Supplementary Table 4 for full results). Looking at the specific pathways detected for each comparison, we observed common trends for the hypermethylation of neural development pathways associated with time0, timeL, and also longitudinal DMPs (Supplementary Figure 4), indicating that the discernible DNA methylation alterations occurring in dementia at the prediagnosis or diagnosis stage may be linked to specific, and similar, functional pathways.

Recent studies using larger cohorts (14,15) have failed to detect single CpG biomarkers at an adequate significance level, while nonetheless being able to define differentially methylated regions (DMRs). Indeed, it is probable that the subtle DNA methylation alterations associated with dementia and cognitive decline in peripheral blood are better detected when looking at coordinated, region-level changes. Working along these lines, we performed

a regional analysis to look for DMRs using the comb-p method (16) (Supplementary Methods). Interestingly, we detected 61 and 65 significant DMRs (Sidak-corrected $p < .05$; Supplementary Table 5 for lists of DMRs) between DEM and CON individuals at time0 and timeL, respectively, while detecting no significant regions for the longitudinal comparison, in spite of the latter comparison involving more subjects. These significant DMRs were dominated by hypomethylation changes (Figure 2a), and the CpGs involved were notably enriched at CpG islands and transcription start sites (TSS; Figure 2b; Fisher's tests all $p < .001$, ORs = 2.4–3.6 for island and 1.5–4.5 for TSS enrichments, except for timeL hyper-DMRs which had a p value of .09 for TSS enrichment), indicating that they might have more defined roles as regards biological regulation. The majority of these regions (42) overlapped in time0 and timeL (Figure 2c), and all DMRs were altered in the same direction, indicating that the dementia-associated DMRs were already present in time0 individuals prior to the detection of overt cognitive decline symptoms. Indeed, the methylation status of the regions perfectly distinguished CON subjects from DEM subjects at both time points (Figure 2d). We did not, however, observe an increase in the magnitude of the alterations at these regions at timeL as opposed to time0 (Supplementary Figure 5).

In addition, because there is a gradient of cognitive scores within the dementia timeL subjects (Table 1; Supplementary Figure 6), the methylation levels at these regions could be subtly associated with the degree of cognitive decline. To explore this, we correlated neurological scores with mean DNA methylation values at the 42 overlapping DMRs by using linear models within the DEM timeL subgroup (Supplementary Methods), but no relationships were found to be significant after multiple-testing adjustment (FDR < 0.05).

A considerable proportion of the DMRs detected were mapped to genes functionally linked to Alzheimer's or related pathologies via different mechanisms (Figure 2e; Supplementary Table 5) such as: (a) genes associated with polymorphisms related to Alzheimer's disease risk—*PON1* (17), *AP2A2* (18), or *SH3PXD2A* (19) (although the latter is cohort-dependent (20))—or those associated with polymorphisms linked to Aβ related neurodegeneration *MAGI2* (21) or polymorphisms related to cerebrospinal fluid tau phosphorylation levels—*POT1* (22); (b) genes with functional roles in dementia disease models—*ALOX5AP* (23), *PLK2* (24), or *ITGAX* (25); (c) tau protein-interacting genes—*PACSIN1* (26); (d) genes with plasma protein levels associated with Alzheimer's in ApoE4 carriers—*CDH6* (27)—or upregulated in the peripheral blood of fast-progression subjects with early Alzheimer's—*SLC2A8* (28); (e) genes associated with more general brain-pathology pathways—*CBR1* (29).

We next looked for overlaps between our study DMRs and those reported using external cohorts. First, we examined the regions with blood DNA methylation alterations in pre and postdiagnosis Alzheimer's subjects described by Fransquet et al. (14) and found 13 intersections with our DMRs (overlapping or <1 000 bp in distance), 10 of which were altered in the same direction (Supplementary Table 6), including regions mapping to aforementioned genes such as *ALOX5AP*. We also found up to 18 intersections with the Alzheimer's and MCI-associated blood altered regions reported by Wang et al. (30), 6 having the same direction of change (Supplementary Table 6), including, for example, a timeL DMR mapped to the *EIF4E* gene, a gene which has been recently reported as specifically detected in the lacrimal fluid of Alzheimer's patients (31). Lastly, despite the fact that most of our regions were hypomethylated, and although it did not reach statistical significance in the DMR calling, we also confirmed in our cohort the recently reported hypermethylation of the *HOXB6* gene in the blood of Alzheimer's patients (15). Interestingly, as is the case for most of our regions, we observed that the DNA methylation alterations were already present in our prediagnosis time0 samples. Taken together, these results highlight the importance of describing cohort-independent DNA methylation alterations. To further pursue this, we made use of the raw data shared by Roubroeks et al. (15) and performed a DMR analysis by integrating their data set with our own measurements (Supplementary Methods). With this strategy, we discovered 8 cohort-independent DMRs (Supplementary Table 7; Figure 2f), which included genes such as *HOXB6*, mentioned earlier, and also *PM20D1*, which has been recently described as hypomethylated in the peripheral blood of early Alzheimer's (30) and is a quantitative trait locus in this disease (32).

Finally, we screened various epigenetic clocks in order to look for more general epigenomic alterations. We estimated DNAm ages using the "Hannum" blood DNAm clock (33), the "Horvath" universal DNAm clock (34), the "PhenoAge" DNAm clock (35), the "GrimAge" DNAm clock (36), and the "Telomere" DNAm clock (37). We computed DNAm age acceleration values by extracting the residuals from the regression of DNAm age on chronological age, with GrimAge also being adjusted for sex after we observed a significant association with this variable (Supplementary Methods). We found no significant differences in DNAm age acceleration or DNAm telomere length acceleration across the groups (Supplementary Figure 7). DNAm age acceleration has been repeatedly associated with Alzheimer phenotypes in brain tissue (35,38,39), while mixed results have been obtained in blood, with DNAm age acceleration having been linked to cognitive fitness, but not to its longitudinal decline (40), associated with longitudinal cognitive decline (41) or not associated with any differences whatsoever (42). In the case of our cohort, there did not appear to be noticeable differences in epigenetic age acceleration between the groups studied. To take into account the distribution of cognitive scores within the dementia timeL subjects (Table 1; Supplementary Figure 6), we also correlated the acceleration with neurological score values by using linear models across all subjects and also within the dementia timeL subgroup. We again found little evidence of a robust increase in epigenetic age acceleration linked to cognitive decline across the 5 different epigenetic clocks, and no significant association was observed after adjustment for multiple testing (Supplementary Figures 8 and 9).

Discussion

In summary, our work describes DNA methylation alterations in the peripheral blood mononuclear cells of cognitively healthy older adults who in the medium-term (at 4-year follow-up) either develop dementia or remain cognitively healthy. Importantly, most of the observed alterations are present at both the prediagnosis and the postdiagnosis stage, suggesting that DNA methylation alterations associated with dementia have already accumulated in peripheral tissues such as blood prior to clinical symptoms being observed, thus indicating its value for the development of epigenetic biomarkers of disease. Even so, these observations could perhaps also be explained by the presence of preexisting individual genetic traits. While it is true that DNA methylation alterations are better detected when looking at coordinated, regional changes, the exploration of the DNA methylation patterns at the single-CpG level also reveals distinctive signatures associated with biological features.

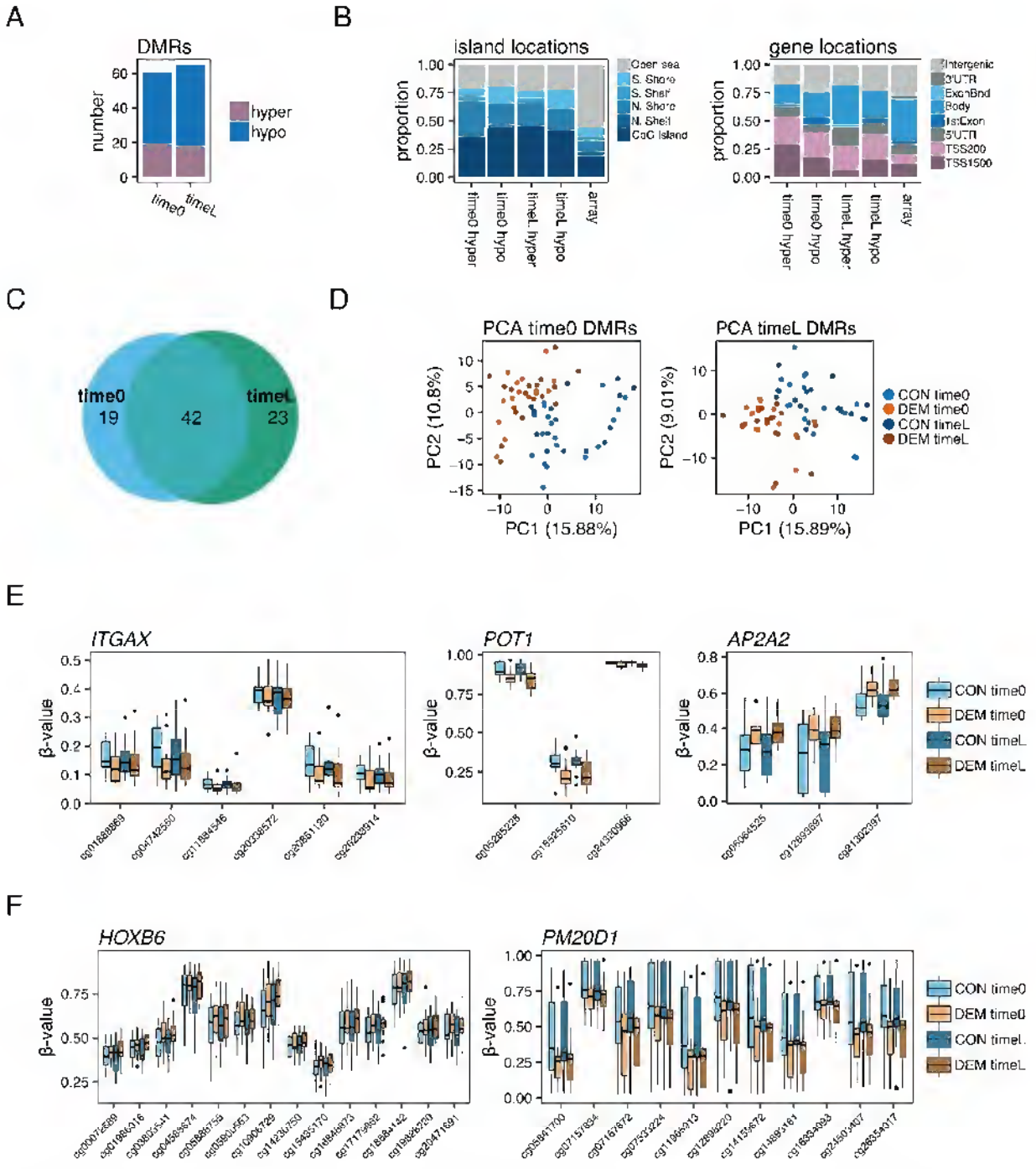


Figure 2. Regional DNA methylation alterations in dementia. (A) Barplots showing the numbers of hyper- and hypo-methylated dementia associated DMRs (Sidak-adjusted $p < .05$) found in the time0 and timeL comparisons. (B) Barplots indicating the relative distribution of CpGs belonging to hyper- and hypo-methylated DMRs in the time0 and timeL comparisons, according to their CpG island location status (left) and gene location status (right). (C) Venn diagram showing the number of DMRs overlapping between the time0 and timeL comparisons. (D) Scatter plots describing the principal component analysis (PCA) of the study subjects according to their mean methylation values for the time0 or timeL DMRs. (E) Boxplots showing the measured DNA methylation values of the individuals, according to their experimental group, at the CpG sites belonging to DMRs associated with the *ITGAX*, *POT1*, and *AP2A2* genes. (F) Boxplots showing the measured DNA methylation values of the individuals, according to their experimental group, at the CpG sites belonging to 2 DMRs discovered in the integrative analysis associated with the *HOXB6* and *PM20D1* genes. CON = control; DEM = dementia; DMRs = differentially methylated regions.

Our study provides valuable epigenetic profiling, using Illumina MethylationEPIC arrays, of a well-characterized longitudinal cohort with comprehensive cognitive measurements separated by a 4-year span. This design allowed for comparisons to be made at 2-time points and also longitudinally. On the other hand, the main limitations of the investigation were: (a) the sample size, which likely

limited the power to detect more subtle alterations; (b) the use of a general diagnosis of dementia without differentiating specific subtypes, so that the alterations described here could be related to other pathologies such as vascular dementia; (c) the short time span in terms of detecting robust aging alterations; and (d) the lack of analyses of lifestyle variables which could reveal additional insights on the biological processes involved.

The DMRs described in this study are related to many Alzheimer's-associated genes, and also overlap with regions reported in other studies, indicating that epigenetic changes reflect the underlying biological processes at play in the development of this disease. Nonetheless, the numbers and extent of dementia-associated DNA methylation alterations are limited, and as such, there is a need for high-powered studies which facilitate the detection of more subtle, poly-epigenetic traits. In this sense, the public availability of epigenetic profiling data sets is of great value for the integration and validation of future studies.

Supplementary Material

Supplementary data are available at *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* online.

Ethics Approval

This study had approval from the Carlos III Institute of Health Research Ethics Committee (HIP CI CEI PI 46_2011-v2015) and all subjects gave written informed consent.

Funding

This work was supported by: the Spanish Association Against Cancer (grant number PROYE18061FERN to M.F.F.), the Asturias Government (PCTI) cofunding 2018-2022/FEDER (grant number IDI/2018/146 to M.F.F.), the Fundación General CSIC (grant number 0348_CIE_6_E to M.F.F.) and the Health Institute Carlos III (Plan Nacional de I + D + I) cofunding FEDER (grant numbers PI15/00892, PI18/01527 to M.F.F. and A.F.F.). J.R.T is supported by a Juan de la Cierva fellowship from the Spanish Ministry of Science and Innovation (grant number FJCI-2015-26965). R.F.P. is supported by the Severo Ochoa program (grant number BP17-114). We also acknowledge support from the Institute of Oncology of Asturias (IUOPA, supported by Obra Social Cajastur Liberbank, Spain), the Health Research Institute of Asturias (ISPA-FINBA) and Consorcio Centro de Investigación Biomédica en Red (CIBERER-ISCIII).

Conflict of Interest

None declared.

Acknowledgments

We thank Ronnie Lendrum for her manuscript editing and the members of the Cancer Epigenetics Laboratory for their positive feedback. Finally, we acknowledge the generous contribution of the patients involved in the study.

Author Contributions

M.F.F., M.M., and J.J.A.-L. conceived, coordinated, and supervised the study. M.M., J.A., and M.C. were involved in the collection of clinical samples. R.F.P. produced the methylation data and performed computational analyses. J.J.A.-L. and J.R.T. assisted in the computational analyses. J.V., C.B., A.R.-D., M.C., and A.F.F. assisted in data analyses and interpretation. R.F.P., M.F.F., M.M., and J.A. participated in drafting the manuscript. All authors revised, read, and approved the final manuscript.

Data Availability

All data generated during this study are included in this published article and its Supplementary Information Files. The raw IDAT and preprocessed data are also available in the ArrayExpress public repository under accession E-MTAB-10600.

References

1. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nat Rev Neurol*. 2018;14(11):653–666. doi:10.1038/s41582-018-0070-3
2. Lourida I, Hannon E, Littlejohns TJ, et al. Association of lifestyle and genetic risk with incidence of dementia. *JAMA*. 2019;322(5):430–437. doi:10.1001/jama.2019.9879
3. Bäckman L. Memory and cognition in preclinical dementia: what we know and what we do not know. *Can J Psychiatry*. 2008;53(6):354–360. doi:10.1177/070674370805300604
4. Irwin K, Sexton C, Daniel T, Lawlor B, Naci L. Healthy aging and dementia: two roads diverging in midlife? *Front Aging Neurosci*. 2018;10:275. doi:10.3389/fnagi.2018.00275
5. Xia X, Jiang Q, McDermott J, Han J-DJ. Aging and Alzheimer's disease: comparison and associations from molecular to system level. *Aging Cell*. 2018;17(5):e12802. doi:10.1111/acel.12802
6. Barter JD, Foster TC. Aging in the brain: new roles of epigenetics in cognitive decline. *Neuroscientist*. 2018;24(5):516–525. doi:10.1177/1073858418780971
7. Fransquet PD, Lacaze P, Saffery R, McNeil J, Woods R, Ryan J. Blood DNA methylation as a potential biomarker of dementia: a systematic review. *Alzheimers Dement*. 2018;14(1):81–103. doi:10.1016/j.jalz.2017.10.002
8. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*. 2018;19(6):371–384. doi:10.1038/s41576-018-0004-3
9. Pettigrew C, Soldan A. Defining cognitive reserve and implications for cognitive aging. *Curr Neurol Neurosci Rep*. 2019;19(1):1. doi:10.1007/s11910-019-0917-z
10. Olazarán J, Valentí M, Frades B, et al. The Valdecas Project: a cohort to identify early markers and mechanisms of Alzheimer's disease. *Front Aging Neurosci*. 2015;7:181. doi:10.3389/fnagi.2015.00181
11. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47–e47. doi:10.1093/nar/gkv007
12. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinf*. 2012;13:86. doi:10.1186/1471-2105-13-86
13. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet*. 2007;3(9):1724–1735. doi:10.1371/journal.pgen.0030161
14. Fransquet PD, Lacaze P, Saffery R, et al. Blood DNA methylation signatures to detect dementia prior to overt clinical symptoms. *Alzheimers Dement (Amst)*. 2020;12(1):e12056. doi:10.1002/dad2.12056
15. Roubroeks JAY, Smith AR, Smith RG, et al. An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. *Neurobiol Aging*. 2020;95:26–45. doi:10.1016/j.neurobiolaging.2020.06.023
16. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics*. 2012;28(22):2986–2988. doi:10.1093/bioinformatics/bts545
17. Erlich PM, Lunetta KL, Cupples LA, et al. Polymorphisms in the PON gene cluster are associated with Alzheimer disease. *Hum Mol Genet*. 2006;15(1):77–85. doi:10.1093/hmg/ddi428
18. Nelson PT, Fardo DW, Katsumata Y. The MUC6/AP2A2 locus and its relevance to Alzheimer's disease: a review. *J Neuropathol Exp Neurol*. 2020;79(6):568–584. doi:10.1093/jnen/nlaa024
19. Harold D, Jehu L, Turic D, et al. Interaction between the ADAM12 and SH3MD1 genes may confer susceptibility to late-onset Alzheimer's

disease. *Am J Med Genet B Neuropsychiatr Genet*. 2007;144B(4):448–452. doi:10.1002/ajmg.b.30456

20. Laumet G, Petitprez V, Sillaire A, et al. A study of the association between the ADAM12 and SH3PXD2A (SH3MD1) genes and Alzheimer's disease. *Neurosci Lett*. 2010;468(1):1–2. doi:10.1016/j.neulet.2009.10.040
21. Kim H-R, Lee T, Choi JK, Jeong Y; Alzheimer's Disease Neuroimaging Initiative. Polymorphism in the MAGI2 gene modifies the effect of Amyloid β on neurodegeneration. *Alzheimer Dis Assoc Disord*. 2021;35(2):114–120. doi:10.1097/WAD.0000000000000422
22. Hohman TJ, Koran MEI, Thornton-Wells TA; Alzheimer's Disease Neuroimaging Initiative. Genetic modification of the relationship between phosphorylated tau and neurodegeneration. *Alzheimers Dement*. 2014;10(6):637–645.e1. doi:10.1016/j.jalz.2013.12.022
23. Giannopoulos PF, Chu J, Joshi YB, et al. 5-lipoxygenase activating protein reduction ameliorates cognitive deficit, synaptic dysfunction, and neuropathology in a mouse model of Alzheimer's disease. *Biol Psychiatry*. 2013;74(5):348–356. doi:10.1016/j.biopsych.2013.04.009
24. Lee JS, Lee Y, André EA, et al. Inhibition of Polo-like kinase 2 ameliorates pathogenesis in Alzheimer's disease model mice. *PLoS One*. 2019;14(7):e0219691. doi:10.1371/journal.pone.0219691. PMID: 31306446; PMCID: PMC6629081.
25. Kamphuis W, Kooijman L, Schetters S, Orre M, Hol EM. Transcriptional profiling of CD11c-positive microglia accumulating around amyloid plaques in a mouse model for Alzheimer's disease. *Biochim Biophys Acta*. 2016;1862(10):1847–1860. doi:10.1016/j.bbadis.2016.07.007
26. Regan P, Mitchell SJ, Kim S-C, et al. Regulation of synapse weakening through interactions of the microtubule associated protein tau with PACSIN1. *J Neurosci*. 2021;41(34):7162–7170. doi:10.1523/JNEUROSCI.3129-20.2021
27. Ahmad S, Milan MDC, Hansson O, et al. CDH6 and HAGH protein levels in plasma associate with Alzheimer's disease in APOE ε4 carriers. *Sci Rep*. 2020;10(1):8233. doi:10.1038/s41598-020-65038-5. PMID: 32427856; PMCID: PMC7237496.
28. Chong MS, Goh LK, Lim WS, et al. Gene expression profiling of peripheral blood leukocytes shows consistent longitudinal downregulation of TOMM40 and upregulation of KIR2DL5A, PLOD1, and SLC2A8 among fast progressors in early Alzheimer's disease. *J Alzheimers Dis*. 2013;34(2):399–405. doi:10.3233/JAD-121621
29. Rashid MA, Haque M, Akbar M. Detoxification of carbonyl compounds by carbonyl reductase in neurodegeneration. *Adv Neurobiol*. 2016;12:355–365. doi:10.1007/978-3-319-28383-8_19
30. Wang Q, Chen Y, Readhead B, et al. Longitudinal data in peripheral blood confirm that PM20D1 is a quantitative trait locus (QTL) for Alzheimer's disease and implicate its dynamic role in disease progression. *Clin Epigenetics*. 2020;12(1):189. doi:10.1186/s13148-020-00984-5. PMID: 33298155; PMCID: PMC7724832.
31. Kenny A, Jiménez-Mateos EM, Zea-Sevilla MA, et al. Proteins and microRNAs are differentially expressed in tear fluid from patients with Alzheimer's disease. *Sci Rep*. 2019;9(1):15437. doi:10.1038/s41598-019-51837-y. PMID: 31659197; PMCID: PMC6817868.
32. Sanchez-Mut JV, Heyn H, Silva BA, et al. PM20D1 is a quantitative trait locus associated with Alzheimer's disease. *Nat Med*. 2018;24(5):598–603. doi:10.1038/s41591-018-0013-y
33. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49(2):359–367. doi:10.1016/j.molcel.2012.10.016
34. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):3156. doi:10.1186/gb-2013-14-10-r115
35. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*. 2018;10(4):573–591. doi:10.18632/aging.101414
36. Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging*. 2019;11(2):303–327. doi:10.18632/aging.101684
37. Lu AT, Seeboth A, Tsai P-C, et al. DNA methylation-based estimator of telomere length. *Aging (Albany NY)*. 2019;11(16):5895–5923. doi:10.18632/aging.102173
38. Levine ME, Lu AT, Bennett DA, Horvath S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging*. 2015;7(12):1198–1211. doi:10.18632/aging.100864
39. Lu AT, Hannon E, Levine ME, et al. Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. *Nat Commun*. 2017;8(1):15353. doi:10.1038/ncomms15353
40. Marioni RE, Shah S, McRae AF, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol*. 2015;44(4):1388–1396. doi:10.1093/ije/dyu277
41. Degerman S, Josefsson M, Nordin Adolfsson A, et al. Maintained memory in aging is associated with young epigenetic age. *Neurobiol Aging*. 2017;55:167–171. doi:10.1016/j.neurobiolaging.2017.02.009
42. Starnawska A, Tan Q, Lenart A, et al. Blood DNA methylation age is not associated with cognitive functioning in middle aged monozygotic twins. *Neurobiol Aging*. 2017;50:60–63. doi:10.1016/j.neurobiolaging.2016.10.025

Discussion

Aging is a major risk factor for the development of multiple pathologies, such as cancer or neurodegeneration. As commented previously (see section **4. EPIGENETICS, ENVIRONMENT AND DISEASE**), epigenetic mechanisms hold promise as putative molecular players explanatory of the links between the aging process and the etiology of disease. Because epigenetic marks are dynamic and reversible, they can inform us of the mechanisms through which the external environment acts upon our health during our lifespan—leading to interindividual differences in health-related outcomes—and can also constitute targets for therapies aiming to improve our healthspan or prevent the appearance of associated diseases.

The general objective of this thesis has been to characterize the epigenomic alterations which are brought about by aging and explore their relationship with epigenomic alterations occurring in aging-related disease. In particular, we focused on cancer, early-life developmental processes and cognitive decline. Across our research, we have made use of different—albeit complementary—approaches to tackle our scientific questions, including the analysis of publicly-available datasets stemming from international biomedical consortia and the in-house generation of epigenetic data from human cohorts and murine models of disease. As can be appreciated from our results, we undertook largely general approaches seeking to comprehend the underlying layers of molecular regulation involved in the aging process; nonetheless, our results have also led to specific findings which we may discuss here.

1. AGING AND CANCER EPIGENOMIC LINKS

It has been classically observed that there are common epigenetic patterns associated with aging and cancer which may contribute to explain the link between the two phenomena. These alterations mainly pertain to DNA methylation changes, so that hypermethylation of development-associated bivalent chromatin CpG islands has been observed in aging (Rakyan et al. 2010; Teschendorff et al. 2010; Heyn et al.

2012; Day et al. 2013; Fernández et al. 2015) and cancer (Ohm et al. 2007; Schlesinger et al. 2007; Widschwendter et al. 2007; Easwaran et al. 2012). Nonetheless, the parallelisms regarding the loss of methylation are less clearly defined and there is a need to accommodate the current knowledge within an epigenomic framework which integrates histone and chromatin state data, and also to deepen our understanding of the evolutionary interspecies conservation of the epigenomic alterations occurring in these processes between human and preclinical models such as mouse.

1.1. DISTINCT CHROMATIN SIGNATURES OF DNA HYPOMETHYLATION IN AGING AND CANCER

We initially compiled epigenetic data for more than 2,000 subjects across seven different tissues: breast, kidney, thyroid, skin, brain, lung and blood. These data consisted of genome-wide DNA methylation measurements from Infinium Human Methylation 450K BeadChip arrays, which profile around 480,000 CpG sites in the human genome (Bibikova et al. 2011), and included healthy and tumoral samples across a range of ages from the TCGA consortium (Cancer Genome Atlas Research Network 2008; Cancer Genome Atlas Network 2012; Brennan et al. 2013; Cancer Genome Atlas Research Network 2013; Cancer Genome Atlas Research Network 2014a; Cancer Genome Atlas Research Network 2014b; Cancer Genome Atlas Network 2015; Cancer Genome Atlas Research Network 2016), which were is some cases complemented with control data from other studies (Guintivano et al. 2013; Hannum et al. 2013; Bormann et al. 2016).

Using the aforementioned datasets, we first set out to map cancer- and aging-associated DNA methylation alterations and found widespread changes across tissues, with cancer presenting more numerous and stronger alterations. Globally, we did not observe a clear trend towards a dominant hyper- or hypomethylation in cancer, while aging changes, more variable in number between tissues, tended towards

hypermethylation. Because the 450K array targets gene- and promoter-enriched genomic loci (Bibikova et al. 2011), inferences regarding global methylation alterations should be taken with caution; nonetheless, intertissue variability in aging alterations has been reported (Fernández et al. 2015) and genome-wide sequencing experiments have also failed to detect global hypomethylation associated with aging, in both human and mouse tissues (Unnikrishnan et al. 2018).

Interestingly, thyroid tissue stood out as a particular case displaying low numbers of cancer-associated changes and more abundant DNA methylation changes with age. Moreover, using Horvath's epigenetic clock to confirm the general disruption of "normal" epigenetic aging within tumors, we observed that again thyroid tissue displayed more coherent epigenetic aging than the rest (Horvath 2013). Age is a very important prognostic factor in thyroid tumors (Haymart 2009; Kazaure et al. 2018) so that DNA methylation alterations could help explain the complex relationship between thyroid cancer and age, and that the epigenetic likeness of thyroid tumors to normal tissues could be related to the good prognosis which generally characterizes this disease (Carling & Udelsman 2014).

Next, the exploration of the genomic distribution of the DNA methylation alterations across aging and cancer revealed tissue-independent similarities between the two processes, with hypermethylation occurring at CpG-dense locations such as CpG islands and gene promoters and hypomethylation being detected at open sea, intergenic and intronic regions. These findings are in line with previously published reports for aging (Day et al. 2013; Yuan et al. 2015) and cancer (Dmitrijeva et al. 2018). Furthermore, using enrichment and permutation testing, we demonstrated the existence of a tissue-independent core of DNA methylation changes in both aging and cancer, as has been described recently for cancer (Chen et al. 2016), suggesting that there is a common, nonstochastic nexus of DNA methylation alterations within both

the cancer and the aging process, independently of the tissue-specific signatures that are classically described in the cancer field (Portela & Esteller 2010).

Up to this point, our observations indicated that there existed general similarities between cancer and aging at the level of DNA methylation alterations and were thus in line with the classic idea proposing that these two processes share an epigenetic mechanism. To further explore this in the context of more layers of epigenomic regulation, we compiled histone PTM (H3K4me1, H3K4me3, H3K27me3, H3K36me3, H3K9me3, and H3K27ac marks), chromatin state and TF binding site (TFBS) data across different tissues and cell lines from the NIH Roadmap (Roadmap Epigenomics Consortium et al. 2015) and ENCODE (ENCODE Project Consortium 2012) projects. The integration of these data led us to confirm that aging and cancer-associated hypermethylation is enriched in chromatin signatures associated with bivalent active (H3K4me1/3) and Polycomb repressive (H3K27me3) marks, as well as heterochromatin (H3K9me3). Strikingly, however, we found quite distinct chromatin signatures for DNA hypomethylation, with aging alterations occurring mostly associated with the enhancer-specific modification H3K4me1 while cancer hypomethylation was found at heterochromatin-associated H3K9me3 locations, indicating that the loss of methylation in aging and cancer occurs in quite different functional contexts. An enhancer-associated hypomethylation chromatin signature has been described for aging (Day et al. 2013; Fernández et al. 2015), while hypomethylation of heterochromatic regions has been reported in cancer (Berman et al. 2011; Hon et al. 2012). In this scenario, our research provides a systematic analysis comparing the methylomic alterations in aging and cancer across a comprehensive set of tissues to demonstrate that the chromatin signatures associated with DNA hypomethylation are quite different between the two processes.

These observations were validated and enhanced with the chromatin state and TFBS data. We also observed some aging-specific hypermethylation events, such as alterations at REST/NRSF binding sites, which have been reported in blood (Yuan et al. 2015). REST is a transcriptional repressor with relevant roles in physiological aging which have been mostly studied in brain tissue (Lu et al. 2014), and which has been shown to preferentially bind to methylated non-CpG sites in mouse development models (Zhang et al. 2017). Regarding DNA hypomethylation, some TFs showed altered binding sites in both aging and cancer, including FOS, JUN, and JUND, suggesting that the AP-1 pathway was targeted in both processes. Indeed, AP-1 has been recently shown to prepare regulatory elements prior to the induction of senescence (Martínez-Zamudio et al. 2020) so it is possible that the hypomethylation at AP-1 sites detected in our data indicates an increase in the occupancy of this factor with both aging and cancer. Aging, however, displayed more abundant alterations than cancer at the binding sites of other bZIP-domain factors, such as FOSL1/2, MAFF and MAFK, and also STAT3, which has been linked to the recruitment of the H3K4 methyltransferase SET9 at promoters (Yang et al. 2010). These observations suggest that H3K4me1-associated hypomethylation in aging could have greater functional consequences than H3K9me3-associated cancer hypomethylation. Nonetheless, when we performed gene set enrichment analyses, we found low enrichments for biological functions at these aging-hypomethylated loci, a result which could be due to these alterations occurring at intergenic enhancer locations. Finally, we performed correlative analyses between DNA methylation alterations and paired gene expression data within the kidney dataset, to find similar numbers of associations in both aging and cancer which displayed both positive and negative correlations between methylation and expression, as has been reported in other studies (Gutierrez-Arcelus et al. 2013).

To sum up, our results suggest that the functional context of DNA hypomethylation in aging and cancer is markedly different, mainly occurring at enhancer-associated locations in aging and at heterochromatin locations in cancer, thus indicating that the epigenetic relationship between these two processes is more complex than previously thought.

1.2. CONSERVATION OF AGING AND CANCER EPIGENETIC SIGNATURES ACROSS HUMAN AND MOUSE

Recent evidence, including our own results, has revealed that DNA methylation alterations in aging and cancer have clear dissimilarities in the hypomethylation scenario (Dmitrijeva et al. 2018; Pérez et al. 2018). Part of the background knowledge spurring these lines of investigation has been the failure to observe global DNA hypomethylation during the aging process using genome-wide sequencing experiments (Unnikrishnan et al. 2018). Often, these studies have been carried out in mouse, and include tissues such as hippocampus, liver or stem cells (Sun et al. 2014; Cole et al. 2017; Hahn et al. 2017; Masser et al. 2017; Hadad et al. 2019; Hernando-Herraez et al. 2019). Thus, it is necessary to: 1) analyze aging and cancer DNA methylation alterations using systematic approaches which allow for the correct comparison of the two diseases; 2) characterize cancer- and aging-associated epigenetic changes in mouse and compare them with observations in human, especially because mouse models are one of the main preclinical instruments used in research (Espada & Esteller 2013). To this end, we compiled young and old, healthy and tumoral brain samples from human subjects and mouse models and performed a genome-wide DNA methylation profiling using reduced representation bisulfite sequencing.

We first characterized the global aspects of the human and mouse methylome using the nontumoral samples and detected similar global DNA methylation levels between the two species. The RRBS interrogated similar sets of CpGs in the two species in

terms of their functional genomic locations, demonstrating that it is a useful technology for the interspecies profiling of DNA methylation signatures. This observation is of importance because, up until quite recently (Arneson et al. 2022; Garcia-Prieto et al. 2022; Zhou et al. 2022), there has been a lack of methylation array technology applicable to mouse. Indeed, the 585,234 CpG sites which we profiled with high coverage across all mouse samples more than double the 280,754 CpG sites covered by the newly developed Infinium Mouse Methylation BeadChip (Garcia-Prieto et al. 2022). Again, it must be noted that RRBS is an enriched technology which preferentially targets CpG island-associated locations, so that its genome-wide measurements—as is also the case of methylation arrays—are impoverished in intergenic and repetitive DNA elements (Meissner et al. 2005; Bock et al. 2010) and thus large scale technologies such as whole-genome bisulfite sequencing (Gong et al. 2022) are more suited for the analysis of global methylomic features. Nonetheless, we were able to use the RRBS data to confirm that the general relationships of DNA methylation with functional genomic and epigenomic elements—including gene and island locations, CpG density, chromatin signatures and repetitive elements—were conserved between human and mouse, in agreement with extensive literature regarding the conservation of DNA methylation patterns across mammals (Lister et al. 2013; Schroeder et al. 2015; Kessler et al. 2016; Zhou et al. 2017), and we also corroborated that CpG density is the main predictor of DNA methylation status throughout the genome in both species (Edwards et al. 2010; Chen et al. 2018).

Next, we parallelly determined cancer- and aging-associated DNA methylation alterations in both human and mouse. We first observed a tendency for cancer-associated hypermethylation in both species, which could be related to our samples belonging to a CpG island methylator phenotype (CIMP), which are tumors displaying high levels of CpG island methylation (Malta et al. 2018). A trend for

hypermethylation was detected in aging for mouse but not human. By mapping the CpGs to genomic locations and genes, we confirmed the analogous interspecies distribution of the alterations, with gains of methylation for both aging and cancer occurring at loci which were denser in CpGs than those suffering from loss of methylation, as has been particularly described in human aging and cancer by us and others (Day et al. 2013; Yuan et al. 2015; Dmitrijeva et al. 2018; Pérez et al. 2018). In consonance with this, we also observed that cancer hypermethylation was more concentrated across a smaller number of genes than the rest of alterations for both species. Gene set enrichment analyses also revealed interspecies agreement regarding the well-known cancer-associated hypermethylation of developmental, Polycomb-associated regions (Ohm et al. 2007; Schlesinger et al. 2007; Widschwendter et al. 2007; Easwaran et al. 2012). Interestingly, we observed more functional enrichment results for aging alterations in mouse as compared with human—with affected pathways related to cell interaction and signaling—, suggesting that mouse aging alterations could be more functionally directed than those in humans. It is probable, nonetheless, that the reduced variability inherent to mouse models, both because of their syngeneic traits and the environmentally-stable housing conditions in which they are bred, helps discern coherent pathway alterations not totally masked by aging-associated stochastic changes.

The integration of the methylomic alterations with ChIP-seq data of histone PTMs across human and mouse tissues (H3K4me1, H3K4me3, H3K27me3, H3K36me3, H3K9me3, and H3K27ac marks) from NIH Roadmap and the recent mouse ENCODE release (Roadmap Epigenomics Consortium et al. 2015; Gorkin et al. 2020) led to two significant observations: 1) we validated that there exist distinct chromatin signatures of DNA hypomethylation between aging and cancer; 2) we demonstrated that these signatures, and their distinction, are conserved between human and mouse. The

cancer-associated signatures involved hypermethylation of active H3K4me1 plus Polycomb H3K27me3 locations and hypomethylation of heterochromatin H3K9me3 locations for both species, with human also displaying a more marked H3K9me3-associated hypermethylation, in consonance with our previous results across multiple human tissues (Pérez et al. 2018), whereas the literature investigating disease-associated chromatin signatures in mouse is scarce. With regards to aging, we again found similar interspecies chromatin signatures which were nonetheless moderately different from those previously described (Fernández et al. 2015; Pérez et al. 2018), with enrichments detected for gene body-related H3K36me3 as well as the known enhancer-associated H3K4me1 mark. These differences could be due to the use of RRBS instead of array technology, from which the classic chromatin signatures have been derived. In support of this idea, a recent report using human array and mouse RRBS data to derive aging-associated methylation signatures reported H3K36me3 enrichment specifically for the mouse RRBS measurements (Wang et al. 2017).

To further explore and validate our observations, we employed Hidden Markov models (Ernst & Kellis 2012) to build chromatin state tracks for human and mouse using the aforementioned histone PTM data from a parallel selection of tissues. We observed a high degree of similarity between the genomic chromatin states of both species, a finding which as very recently been demonstrated using similar ENCODE data (van der Velde et al. 2021). The enrichment analyses of the cancer and aging DNA methylation alterations in the chromatin states supported the conclusions reached in the previous histone PTM enrichments, with analogous interspecies signatures being found. We also detected human-specific zinc-finger gene hypermethylation. These transcription factors, rich in repetitive elements, are characterized by a high presence of the repressive H3K9me3 mark (Blahnik et al. 2011) so that this result could explain the previously detected increase in H3K9me3-

associated hypermethylation in human as compared with mouse. Interestingly, the zinc-finger family of genes presents a significant evolutionary expansion in human in relation to other species such as mouse, and it is thought that novel zinc-finger genes may play roles in genomic regulation via transcriptional repression (Emerson & Thomas 2009). Thus, our observation of this human-specific cancer-associated epigenetic trait may in fact reflect a bona fide pathway of zinc-finger dysregulation which does not exist in other species.

Because of the mentioned conflicts regarding the observation of global DNA hypomethylation in aging, we explored DNA methylation patterns at repetitive DNA elements—since these locations represent a large fraction of the genome—and found parallel aging- and cancer alterations in human and mouse. Curiously, while repetitive DNA was mostly hypomethylated in cancer, we found evidence of both hyper- and hypomethylation of repetitive elements with aging for both species. Even though RRBS in an enriched technology which mostly interrogates single-copy sequences, these findings are in consonance with other genome-wide studies finding no loss of global methylation during aging in both human and mouse brain tissue (Lister et al. 2013; Hadad et al. 2019; McKinney et al. 2019). We also observed some species-specific alterations, such as the human-specific cancer-associated hypermethylation of ribosome-related elements (rRNA, srpRNA), which has been described in some tumors (Srivastava et al. 2016), and satellite repeats being overall more affected by DNA methylation alterations in human. We also found no associations with simple or low-complexity repeats, which we had initially observed to display low levels of basal methylation in normal human and mouse tissue—contrary to the rest of repetitive DNA—, so our observations indicate that the baseline, physiological levels of DNA methylation will often predict the epigenetic behavior of genomic loci during pathological dysregulation.

We additionally studied the enrichment of cancer and aging DNA methylation alterations in transcription factor binding sites using motif discovery tools (Heinz et al. 2010). Interestingly, while our previous pathway enrichment analyses revealed that cancer hypermethylation presented the most common functionalities between human and mouse—related to the alteration of Polycomb targets and developmental genes—, the TF motif analyses revealed interspecies similarities particularly in the aging scenario, which could possibly indicate that aging-associated alterations are more functionally relevant when occurring at regions not directly located within genes, like many TF binding sites. With respect to general trends, we observed the interspecies cancer hypermethylation of Homeobox TFs, and also hypermethylation of nuclear receptor (NR) TFs in human and basic helix-loop-helix (bHLH) domain TFs in mouse, while cancer hypomethylation was particularly focused to bHLH and high-mobility group (HMG) domain TFs in human. All of these families of TFs have particular links to developmental processes (Jones 2004; Furusawa & Cherukuri 2010; Bürglin & Affolter 2016; Weikum et al. 2018) underscoring the importance of the alteration of cellular differentiation pathways during tumorigenesis. Regarding aging alterations, hypermethylation was clearly linked to E26 transformation-specific (ETS) domain TFs, also related to cell differentiation and development (Sharrocks 2001) and hypomethylation to bHLH and HMG-domain TFs for both species. Several of these families, such as ETS and HMG, have important roles in regulating DNA accessibility through epigenetic mechanisms (Sharrocks 2001; Reeves 2010), and also in oncogenesis (Sizemore et al. 2017). Interestingly, we detected the aging-associated hypomethylation of binding sites of several Sox TFs from the HMG family such as SOX2, which has been shown to be dysregulated with aging in human and mouse brain (Carrasco-Garcia et al. 2019). We also observed hypomethylation of some SOX TFs with cancer, indicating a possible nexus between cancer and aging DNA

methylation alterations, a relationship which has been reported for Sox4 in mouse (Foronda et al. 2014). Another observed link between aging and cancer was the epigenetic dysregulation in both processes of bHLH-domain TF binding sites, including factors such as NeuroG2, NeuroD1, Olig2, or Atoh1, which are related to neuronal development and differentiation (Dennis et al. 2019). Overall, these results indicate that cancer and aging DNA methylation alterations impact common and specific TF-regulated pathways in human and mouse, with aging dysregulation displaying a notable conservation between the two species, and that the alteration of some of these pathways may also provide epigenetic mechanistic links between aging and cancer.

Once we had outlined the general features of aging and cancer DNA methylation alterations in both species, we sought to do a direct species-to-species comparison of DNA methylation patterns between human and mouse. To this end, we recovered orthologous measurements by using genomic liftover tools (Hinrichs 2006) to define a final set of 59,100 CpG sites with interspecies measurements across all samples. The majority of these CpGs displayed a similar genomic context in human and mouse in terms of their CpG island membership, suggesting that they were related to similar functional elements in both species. First, correlation analyses revealed that these locations possessed species-specific epigenetic signatures, with tumor samples displaying widespread dysregulation and, very interestingly, we observed that mouse tumors were more similar to human tumors than to the rest of mouse samples, indicating that the epigenetic mechanisms behind tumorigenesis have a strong conservation between human and mouse. Moreover, we also demonstrated that sites of DNA methylation variability were conserved between both species. In order to focus on species-specific epigenetic traits, we defined a subset of species-discordant CpGs (sdCpGs). Curiously, we observed marked differences in the CpG island status of these sdCpGs between human and mouse. Indeed, by looking at the interspecies transitions

between the genomic contexts of these CpGs we showed that interspecies differences in DNA methylation could be partly explained by changes in the genomic context surrounding these CpG sites, as has been previously discussed (Hernando-Herraez et al. 2015; Zhou et al. 2017). Next, we focused on cancer- and aging-associated CpGs with orthologous measurements and observed a common interspecies nexus of DNA methylation alterations in these processes. We detected that cancer hypermethylation tended to occur at more conserved sequence domains than the rest of alterations, as was also the case of sdCpGs; however, stratified sampling analyses demonstrated that most of these observations could be again explained by differences in the functional genomic context surrounding the involved CpGs, in terms of CpG islands and gene locations (Hernando-Herraez et al. 2015). All in all, these results suggest that methylomic patterns, as well as the sites of cancer and aging dysregulation, are mostly conserved between human and mouse, while species-specific epigenetic traits can be partly explained by changes in the genomic context surrounding the implicated CpG sites.

The results hitherto described indicate a strong interspecies conservation of both physiological and pathological epigenetic mechanisms between human and mouse. Because mouse models are very important tools in the study of diseases, epigenetic interspecies biomarkers may lead to better or more confident biological candidates of disease regulation. To investigate this hypothesis, we screened our data to discover six gene candidates with conserved and concordant human-mouse aging-cancer alterations. Then, we compiled and analyzed gene expression data from the TCGA lower-grade glioma (LGG) and glioblastoma (GBM) human cohorts (Cancer Genome Atlas Research Network 2008; Brennan et al. 2013; Cancer Genome Atlas Research Network et al. 2015) and found strong evidence of cancer-associated dysregulation and association with survival for five of these genes across glioma patients. Moreover, in

the case of the glioma-dysregulated *AGAP3* gene—a neural NMDA receptor (Oku & Huganir 2013)—, we were able to spatially associate cancer and aging DNA methylation alterations with a specific isoform explanatory of survival variation within the cancer patients. The other cancer-altered genes included *GPRIN1*—involved in neural cytoskeletal dynamics (Nordman & Kabbani 2012)—, *AJM1*—related to apical junctions (Köppen et al. 2001)—, *LHX2*—involved in neural development (Chou & Tole 2019) and also reported to be downregulated in glioma (Cheng et al. 2019)—and the myelin-like gene *CCDC177*. The importance of these genes in glioma is to date mainly uninvestigated.

Finally, we sought to compile and make available the interspecies-conserved cancer and aging alterations discovered in our study, so that they may aid in the design of future studies analyzing disease-associated DNA methylation biomarkers in mouse models, which may have a translation in human. Therefore, we developed a web database application (available at https://epilabasturias.shinyapps.io/mbe_app/) based on the R shiny web application framework (Chang et al. 2021), which is extensively described in **ANNEX 1. DEVELOPMENT OF A WEB DATABASE APPLICATION CONTAINING AN INTERACTIVE DATABASE OF AGING AND CANCER DNA METHYLATION ALTERATIONS IN HUMAN AND MOUSE**. The application is an interactive database that can be used to interrogate custom genomic regions of interest in the human or mouse genome and outputs cancer and aging DNA methylation alterations which we have observed, in human and mouse, indicating if they are conserved between the two species.

Taken together, our observations suggest that there are relevant differences between the features of DNA methylation alterations in aging and cancer, across multiple levels of epigenomic regulation. Importantly, the specific methylomic signatures of the aging and cancer processes appear to be robustly conserved between

human and mouse and can be used to derive interspecies disease biomarkers which may have a more significant translation to human biology.

2. AGING AND DEVELOPMENTAL DNA METHYLATION ALTERATIONS

2.1. LONGITUDINAL GENOME-WIDE DNA METHYLATION ANALYSIS UNCOVERS PERSISTENT EARLY-LIFE DNA METHYLATION CHANGES

It is well known that, during the aging process, there occurs a gradual accumulation of epigenetic alterations, of which some present specific and directed biological characteristics while others appear to be stochastic noise (Tejedor & Fraga 2017). Nonetheless, this progressive transformation of the epigenome may not follow a chronologically linear path: indeed, evidence from studies measuring the "ticking rate" of epigenetic clocks (Horvath & Raj 2018) and from DNA methylation analyses of pediatric cohorts (Martino et al. 2011; Martino et al. 2013; Simpkin et al. 2015; Acevedo et al. 2015; Urdinguio et al. 2016) suggest that there occurs a strong epigenetic remodeling during the first years of life. This reshaping is probably associated with developmental processes taking place prior to adulthood. However, it remains to be clarified whether these alterations are similar—albeit stronger—to those occurring during the later stages of lifespan, or if they are functionally different to aging-associated changes. To tackle this question, we collected blood samples from a longitudinal pediatric cohort of healthy subjects with measurements at birth, 5 and 10 years of age and profiled their genome-wide DNA methylation levels using Infinium MethylationEPIC BeadChip arrays, which measure over 850,000 CpG sites in the human genome (Pidsley et al. 2016). These chips are the current state-of-the-art array technology and represent a near twofold increase in coverage as compared to the previously mentioned 450K BeadChips, improving the measurement of non-genic regions such as enhancer locations and other regulatory elements (Pidsley et al. 2016).

First, we looked at the global methylomic patterns of the subjects and confirmed that a drastic epigenomic remodeling took place within the first 5 years of life which was considerably reduced in the next lustrum. The determination of DNA methylation alterations at the individual CpG level confirmed that changes during the first 5 years were much more numerous and stronger. There are few longitudinal studies inspecting blood methylomic changes across more than two time points, but the existing evidence using earlier array technology agrees with our observations (Martino et al. 2011; Acevedo et al. 2015; Simpkin et al. 2015). Nonetheless, it is worth noting that our study, as well as others, collects cord blood at birth and peripheral blood in subsequent time points, so that part of the magnitude of the observed changes could be due to the comparison between these two not totally equivalent tissues. To minimize this issue as much as possible, we used cell deconvolution algorithms (Houseman et al. 2012) to derive cell composition measurements from the methylation data which we adjusted for within the linear models. Reassuringly, we did not detect striking differences in cell composition between the cord and peripheral blood samples.

The genomic distribution of the altered CpG sites mimicked that which we had previously observed for the aging process across multiple tissues (Pérez et al. 2018), with hypermethylation occurring at CpG-island associated locations while hypomethylation was linked to intergenic, open sea loci. In light of this, we integrated the results with the previously mentioned ENCODE and NIH Roadmap histone PTM and chromatin state data (ENCODE Project Consortium 2012; Roadmap Epigenomics Consortium et al. 2015). We observed high enrichments in H3K27me3/H3K4me1 and H3K4me1-associated domains for hyper- and hypomethylation, respectively, during the first 5 years of life, with chromatin state enrichment indicating that the gain of methylation occurred at bivalent and Polycomb-associated loci, while loss of methylation was detected at enhancer-associated locations. These results are again in

consonance with the chromatin signatures linked to aging-associated alterations (Day et al. 2013; Fernández et al. 2015; Pérez et al. 2018), though the associations observed by us in the case of the developmental, early life alterations, were particularly well defined.

To exploit the increased coverage of the EPIC array regarding enhancer locations, we integrated our results with the multi-tissue EnhancerAtlas database (Gao et al. 2016). Interestingly, the enhancer enrichments allowed us to dissect and differentiate the active H3K4me1 association within the bivalent chromatin context of DNA hypermethylation from the active H3K4me1 association observed in the context of DNA hypomethylation: the first involved enhancers typical of fetal and developmental tissues, while the latter—with stronger enrichments—affected enhancers linked to blood cell types.

All in all, our data implies that development-associated DNA methylation changes occurring during the first years of life are especially contained within the first 5 years, and that this epigenetic remodeling shares similar epigenomic features with typical aging-associated alterations, which are nonetheless much more limited in number and magnitude.

3. AGING AND OTHER AGING-RELATED DISEASES: COGNITIVE DECLINE

3.1. BLOOD DNA METHYLATION PATTERNS IN OLDER ADULTS WITH EVOLVING DEMENTIA

Epigenetic alterations arising throughout lifespan can not only contribute to explaining the etiology of disease, but also, in a more translational setting, be used to predict the development of pathologies. Neurodegeneration is tightly linked to aging and both processes manifest DNA methylation alterations (Xia et al. 2018); thus, aged subjects may display alterations associated with cognitive deterioration. Nonetheless, the access to this putative biological information is influenced by our capacity to

detect epigenetic signal in biological tissue. In this sense, peripheral blood may be a good candidate for the detection of disease-associated epigenetic alterations because of its accessibility (Fransquet et al. 2018). To test this hypothesis, we analyzed blood samples from a longitudinal cohort of advanced-age cognitively-healthy individuals, of which a subset of subjects developed cognitive deterioration in a medium period of time (~4 years), again using the aforementioned Infinium MethylationEPIC BeadChip arrays.

We discovered DNA methylation biomarkers—particularly at the level of DNA methylation regions—associated with cognitive decline. Importantly, making use of our longitudinal design, we observed that these altered regions were already impacted in the individuals around 4 years before the manifestation of the clinical symptoms of cognitive deterioration. Thus, we can conclude that systemic tissues such as blood can capture, through epigenetic marks, information regarding biological dysregulation even if the core of the pathology is occurring in a different tissue such as brain (Fransquet et al. 2018). Moreover, subtle DNA methylation alterations can allow us to "predict" the appearance of explicit clinical signs of disease, as has been also recently shown (Fransquet et al. 2020). Interestingly, multiple of these alterations were associated with genes previously linked to dementia-related polymorphisms, such as *AP2A2*, *MAGI2, POT1*, *PON1* or *PM20D1* (Erlich et al. 2006; Hohman et al. 2014; Sanchez-Mut et al. 2018; Nelson et al. 2020; Kim et al. 2021). Indeed, many of the detected DNA methylation differences could in fact be surrogates of pre-existing genetic differences at related genomic locations, such as the *PM20D1* gene, whose genetic variation has been shown to modulate DNA methylation and gene expression linked to Alzheimer's disease (Sanchez-Mut et al. 2018). Nonetheless, while genetic differences can explain how it is possible to detect alterations in non-brain tissues such as blood, the fact that dementia-related blood DNA methylation alterations can be by

themselves quite different from those present in brain, as has been shown in a recent cross-tissue study (Silva et al. 2022), suggests that there may be other mechanisms involved.

On the whole, our results show how DNA methylation alterations linked to cognitive decline which appears in aged subjects can be present in peripheral blood some years before the onset of clinical symptoms.

CONCLUSIONS

The main conclusions obtained in this doctoral thesis are the following:

1. The epigenomic context associated with DNA hypermethylation is similar in aging and cancer, being linked to bivalent chromatin signatures in both processes across multiple tissues.
2. The epigenomic context associated with DNA hypomethylation is different in aging and cancer, being linked to enhancer-associated domains in aging and heterochromatin-associated domains in cancer across multiple tissues.
3. The epigenomic characteristics associated with DNA methylation alterations in brain aging and cancer are conserved between human and mouse.
4. The definition of conserved interspecies DNA methylation biomarkers of disease can aid in the identification of clinically relevant pathways in human.
5. The epigenome is drastically remodeled within the first five years of life and becomes much more stable in later years.
6. The epigenomic characteristics of development-associated DNA methylation changes which occur during the first years of life are similar to, but of larger magnitude, than aging-associated DNA methylation alterations.
7. DNA methylation can be used to derive biomarkers predictive of cognitive decline and dementia in elderly cohorts.

CONCLUSIONES

Las principales conclusiones obtenidas en esta tesis doctoral son las siguientes:

1. El contexto epigenómico asociado a la hipermetilación del ADN es similar en cáncer y envejecimiento, estando relacionado con dominios bivalentes de cromatina para ambos procesos en múltiples tejidos.
2. El contexto epigenómico asociado a la hipometilación del ADN es diferente en cáncer y envejecimiento, estando relacionado con dominios *enhancer* en envejecimiento y dominios de heterocromatina en cáncer, en múltiples tejidos.
3. Las características epigenómicas asociadas a los cambios de metilación del ADN en tejido cerebral durante el envejecimiento y cáncer están conservadas entre humano y ratón.
4. La definición de biomarcadores de enfermedad conservados entre especies basados en la metilación del ADN puede ayudar a la identificación de rutas moleculares clínicamente relevantes en humano.
5. El epigenoma sufre una remodelación drástica contenida en los cinco primeros años de vida y se vuelve mucho más estable en años posteriores.
6. Las características epigenómicas de los cambios de metilación del ADN asociados al desarrollo que ocurren durante los primeros años de vida son similares, aunque de mayor magnitud, que las alteraciones de metilación del ADN asociadas al envejecimiento.
7. La metilación del ADN puede utilizarse para desarrollar biomarcadores capaces de predecir el deterioro cognitivo y la demencia en cohortes de individuos de avanzada edad.

REFERENCES

Acevedo N, Reinius LE, Vitezic M, Fortino V, Söderhäll C, Honkanen H, Veijola R, Simell O, Toppari J, Ilonen J, Knip M, Scheynius A, Hyöty H, Greco D & Kere J (2015) Age-associated DNA methylation changes in immune genes, histone modifiers and chromatin remodeling factors within 5 years after birth in human blood leukocytes. *Clin Epigenetics* 7, 34.

Al Dahhan NZ, De Felice FG & Munoz DP (2019) Potentials and Pitfalls of Cross-Translational Models of Cognitive Impairment. *Front Behav Neurosci* 13, 48.

An J, Rao A & Ko M (2017) TET family dioxygenases and DNA demethylation in stem cells and cancers. *Exp Mol Med* 49, e323.

Anastasiadi D, Esteve-Codina A & Piferrer F (2018) Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* 11, 37.

Antonin W & Neumann H (2016) Chromosome condensation and decondensation during mitosis. *Curr Opin Cell Biol* 40, 15–22.

Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, Jenuwein T, Xu G, Leonhardt H, Wolf V & Walter J (2012) In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet* 8, e1002750.

Arneson A, Haghani A, Thompson MJ, Pellegrini M, Kwon SB, Vu H, Maciejewski E, Yao M, Li CZ, Lu AT, Morselli M, Rubbi L, Barnes B, Hansen KD, Zhou W, Breeze CE, Ernst J & Horvath S (2022) A mammalian methylation array for profiling methylation levels at conserved sequences. *Nat Commun* 13, 783.

Atlasi Y & Stunnenberg HG (2017) The interplay of epigenetic marks during stem cell differentiation and development. *Nat Rev Genet* 18, 643–658.

Avery OT, Macleod CM & McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* 79, 137–158.

Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A & Balasubramanian S (2014) 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem* 6, 1049–1055.

Baker M (2011) Making sense of chromatin states. *Nat Methods* 8, 717–722.

Banani SF, Lee HO, Hyman AA & Rosen MK (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* 18, 285–298.

Bannister AJ & Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21, 381–395.

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Hérault Y, Guillou F & Bourc'his D (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* 354, 909–912.

Barnes CE, English DM & Cowley SM (2019) Acetylation & Co: an expanding repertoire of histone acylations regulates chromatin and transcription. *Essays Biochem* 63, 97–107.

Barter JD & Foster TC (2018) Aging in the Brain: New Roles of Epigenetics in Cognitive Decline. *Neuroscientist* 24, 516–525.

Barth TK & Imhof A (2010) Fast signals and slow marks: the dynamics of histone modifications. *Trends in Biochemical Sciences* 35, 618–626.

Becker JS, Nicetto D & Zaret KS (2016) H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet* 32, 29–41.

Bell JSK & Vertino PM (2017) Orphan CpG islands define a novel class of highly active enhancers. *Epigenetics* 12, 449–464.

Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CPE, van Dijk CM, Tollenaar RAEM, Van Den Berg D & Laird PW (2011) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet* 44, 40–46.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL & Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS & Thomson JA (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 28, 1045–1048.

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan J-B & Shen R (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98, 288–295.

Bird A (2007) Perceptions of epigenetics. *Nature* 447, 396–398.

Blahnik KR, Dou L, Echipare L, Iyengar S, O'Geen H, Sanchez E, Zhao Y, Marra MA, Hirst M, Costello JF, Korf I & Farnham PJ (2011) Characterization of the Contradictory Chromatin Signatures at the 3′ Exons of Zinc Finger Genes A. Wutz, ed. *PLoS ONE* 6, e17121.

Blommaert J (2020) Genome size evolution: towards new model systems for old questions. *Proc Biol Sci* 287, 20201441.

Bochtler M, Kolano A & Xu G-L (2017) DNA demethylation pathways: Additional players and regulators. *Bioessays* 39, 1–13.

Bock C, Tomazou EM, Brinkman AB, Müller F, Simmer F, Gu H, Jäger N, Gnirke A, Stunnenberg HG & Meissner A (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol* 28, 1106–1114.

Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S & Vilain E (2011) Epigenetic predictor of age. *PLoS One* 6, e14821.

Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, Sparrow D, Vokonas P & Baccarelli A (2009) Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev* 130, 234–239.

Booth LN & Brunet A (2016) The Aging Epigenome. *Mol Cell* 62, 728–744.

Bormann F, Rodríguez-Paredes M, Hagemann S, Manchanda H, Kristof B, Gutekunst J, Raddatz G, Haas R, Terstegen L, Wenck H, Kaderali L, Winnefeld M & Lyko F (2016) Reduced DNA methylation patterning and transcriptional connectivity define human skin aging. *Aging Cell* 15, 563–571.

Bourc'his D & Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96–99.

Brenet F, Moh M, Funk P, Feierstein E, Viale AJ, Socci ND & Scandura JM (2011) DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One* 6, e14524.

Brennan CW, Verhaak RGW, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, Beroukhim R, Bernard B, Wu C-J, Genovese G, Shmulevich I, Barnholtz-Sloan J, Zou L, Vegesna R, Shukla SA, Ciriello G, Yung WK, Zhang W, Sougnez C, Mikkelsen T, Aldape K, Bigner DD, Van Meir EG, Prados M, Sloan A, Black KL, Eschbacher J, Finocchiaro G, Friedman W, Andrews DW, Guha A, Iacocca M, O'Neill BP, Foltz G, Myers J, Weisenberger DJ, Penny R, Kucherlapati R, Perou CM, Hayes DN, Gibbs R, Marra M, Mills GB, Lander E, Spellman P, Wilson R, Sander C, Weinstein J, Meyerson M, Gabriel S, Laird PW, Haussler D, Getz G, Chin L, Benz C, Barnholtz-Sloan J, Barrett W, Ostrom Q, Wolinsky Y, Black KL, Bose B, Boulos PT, Boulos M, Brown J, Czerinski C, Eppley M, Iacocca M, Kempista T, Kitko T, Koyfman Y, Rabeno B, Rastogi P, Sugarman M, Swanson P, Yalamanchii K, Otey IP, Liu YS, Xiao Y, Auman JT, Chen P-C, Hadjipanayis A, Lee E, Lee S, Park PJ, Seidman J, Yang L, Kucherlapati R, Kalkanis S, Mikkelsen T, Poisson LM, Raghunathan A, Scarpace L, Bernard B, Bressler R, Eakin A, Iype L, Kreisberg RB, Leinonen K, Reynolds S, Rovira H, Thorsson V, Shmulevich I, Annala MJ, Penny R, Paulauskis J, Curley E, Hatfield M, Mallery D, Morris S, Shelton T, Shelton C, Sherman M, Yena P, Cuppini L, DiMeco F, Eoli M, Finocchiaro G, Maderna E, Pollo B, Saini M, Balu S, Hoadley KA, Li L, Miller CR, Shi Y, Topal MD, Wu J, Dunn G, Giannini C, O'Neill BP, Aksoy BA, Antipin Y, Borsu L, Berman SH, Brennan CW, Cerami E, Chakravarty D, Ciriello G, Gao J, Gross B, Jacobsen A, Ladanyi M, Lash A, Liang Y, Reva B, Sander C, Schultz N, Shen R, Socci ND, Viale A, Ferguson ML, Chen Q-R, Demchok JA, Dillon LAL, Shaw KRM, Sheth M, Tarnuzzer R, Wang Z, Yang L, Davidsen T, Guyer MS, Ozenberger BA, Sofia HJ, Bergsten J, Eckman J, Harr J, Myers J, Smith C, Tucker K, Winemiller C, Zach LA, Ljubimova JY, Eley G, Ayala B, Jensen MA, Kahn A, Pihl TD, Pot DA, Wan Y, Eschbacher J, Foltz G, Hansen N, Hothi P, Lin B, Shah N, Yoon J, Lau C,

Berens M, Ardlie K, Beroukhim R, Carter SL, Cherniack AD, Noble M, Cho J, Cibulskis K, DiCara D, Frazer S, Gabriel SB, Gehlenborg N, Gentry J, Heiman D, Kim J, Jing R, Lander ES, Lawrence M, Lin P, Mallard W, Meyerson M, Onofrio RC, Saksena G, Schumacher S, Sougnez C, Stojanov P, Tabak B, Voet D, Zhang H, Zou L, Getz G, Dees NN, Ding L, Fulton LL, Fulton RS, Kanchi K-L, Mardis ER, Wilson RK, Baylin SB, Andrews DW, Harshyne L, Cohen ML, Devine K, Sloan AE, VandenBerg SR, Berger MS, Prados M, Carlin D, Craft B, Ellrott K, Goldman M, Goldstein T, Grifford M, Haussler D, Ma S, Ng S, Salama SR, Sanborn JZ, Stuart J, Swatloski T, Waltman P, Zhu J, Foss R, Frentzen B, Friedman W, McTiernan R, Yachnis A, Hayes DN, Perou CM, Zheng S, Vegesna R, Mao Y, Akbani R, Aldape K, Bogler O, Fuller GN, Liu W, Liu Y, Lu Y, Mills G, Protopopov A, Ren X, Sun Y, Wu C-J, Yung WKA, Zhang W, Zhang J, Chen K, Weinstein JN, Chin L, Verhaak RGW, Noushmehr H, Weisenberger DJ, Bootwalla MS, Lai PH, Triche TJ, Van Den Berg DJ, Laird PW, Gutmann DH, Lehman NL, VanMeir EG, Brat D, Olson JJ, Mastrogianakis GM, Devi NS, Zhang Z, Bigner D, Lipp E & McLendon R (2013) The Somatic Genomic Landscape of Glioblastoma. *Cell* 155, 462–477.

Briand N & Collas P (2020) Lamina-associated domains: peripheral matters and internal affairs. *Genome Biol* 21, 85.

Briggs R & King TJ (1952) Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proc Natl Acad Sci U S A* 38, 455–463.

Buitrago D, Labrador M, Arcon JP, Lema R, Flores O, Esteve-Codina A, Blanc J, Villegas N, Bellido D, Gut M, Dans PD, Heath SC, Gut IG, Brun Heath I & Orozco M (2021) Impact of DNA methylation on 3D genome structure. *Nat Commun* 12, 3243.

Bürglin TR & Affolter M (2016) Homeodomain proteins: an update. *Chromosoma* 125, 497–521.

Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70.

Cancer Genome Atlas Network (2015) Genomic Classification of Cutaneous Melanoma. *Cell* 161, 1681–1696.

Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061–1068.

Cancer Genome Atlas Research Network (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499, 43–49.

Cancer Genome Atlas Research Network (2016) Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med* 374, 135–145.

Cancer Genome Atlas Research Network (2014a) Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550.

Cancer Genome Atlas Research Network (2014b) Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 159, 676–690.

Cancer Genome Atlas Research Network, Brat DJ, Verhaak RGW, Aldape KD, Yung WKA, Salama SR, Cooper LAD, Rheinbay E, Miller CR, Vitucci M, Morozova O, Robertson AG, Noushmehr H, Laird PW, Cherniack AD, Akbani R, Huse JT, Ciriello G, Poisson LM, Barnholtz-Sloan JS, Berger MS, Brennan C, Colen RR, Colman H, Flanders AE, Giannini C, Grifford M, Iavarone A, Jain R, Joseph I, Kim J, Kasaian K, Mikkelsen T, Murray BA, O'Neill BP, Pachter L, Parsons DW, Sougnez C, Sulman EP, Vandenberg SR, Van Meir EG, von Deimling A, Zhang H, Crain D, Lau K, Mallery D, Morris S, Paulauskis J, Penny R, Shelton T, Sherman M, Yena P, Black A, Bowen J, Dicostanzo K, Gastier-Foster J, Leraas KM, Lichtenberg TM, Pierson CR, Ramirez NC, Taylor C, Weaver S, Wise L, Zmuda E, Davidsen T, Demchok JA, Eley G, Ferguson ML, Hutter CM, Mills Shaw KR, Ozenberger BA, Sheth M, Sofia HJ, Tarnuzzer R, Wang Z, Yang L, Zenklusen JC, Ayala B, Baboud J, Chudamani S, Jensen MA, Liu J, Pihl T, Raman R, Wan Y, Wu Y, Ally A, Auman JT, Balasundaram M, Balu S, Baylin SB, Beroukhim R, Bootwalla MS, Bowlby R, Bristow CA, Brooks D, Butterfield Y, Carlsen R, Carter S, Chin L, Chu A, Chuah E, Cibulskis K, Clarke A, Coetzee SG, Dhalla N, Fennell T, Fisher S, Gabriel S, Getz G, Gibbs R, Guin R, Hadjipanayis A, Hayes DN, Hinoue T, Hoadley K, Holt RA, Hoyle AP, Jefferys SR, Jones S, Jones CD, Kucherlapati R, Lai PH, Lander E, Lee S, Lichtenstein L, Ma Y, Maglinte DT, Mahadeshwar HS, Marra MA, Mayo M, Meng S, Meyerson ML, Mieczkowski PA, Moore RA, Mose LE, Mungall AJ, Pantazi A, Parfenov M, Park PJ, Parker JS, Perou CM, Protopopov A, Ren X, Roach J, Sabedot TS, Schein J, Schumacher SE, Seidman JG, Seth S, Shen H, Simons JV, Sipahimalani P, Soloway MG, Song X, Sun H, Tabak B, Tam A, Tan D, Tang J, Thiessen N, Triche T, Van Den Berg DJ, Veluvolu U, Waring S, Weisenberger DJ, Wilkerson MD, Wong T, Wu J, Xi L, Xu AW, Yang L, Zack TI, Zhang J, Aksoy BA, Arachchi H, Benz C, Bernard B, Carlin D, Cho J, DiCara D, Frazer S, Fuller GN, Gao J, Gehlenborg N, Haussler D, Heiman DI, Iype L, Jacobsen A, Ju Z, Katzman S, Kim H, Knijnenburg T, Kreisberg RB, Lawrence MS, Lee W, Leinonen K, Lin P, Ling S, Liu W, Liu Y, Liu Y, Lu Y, Mills G, Ng S, Noble MS, Paull E, Rao A, Reynolds S, Saksena G, Sanborn Z, Sander C, Schultz N, Senbabaoglu Y, Shen R, Shmulevich I, Sinha R, Stuart J, Sumer SO, Sun Y, Tasman N, Taylor BS, Voet D, Weinhold N, Weinstein JN, Yang D, Yoshihara K, Zheng S, Zhang W, Zou L, Abel T, Sadeghi S, Cohen ML, Eschbacher J, Hattab EM, Raghunathan A, Schniederjan MJ, Aziz D, Barnett G, Barrett W, Bigner DD, Boice L, Brewer C, Calatozzolo C, Campos B, Carlotti CG, Chan TA, Cuppini L, Curley E, Cuzzubbo S, Devine K, DiMeco F, Duell R, Elder JB, Fehrenbach A, Finocchiaro G, Friedman W, Fulop J, Gardner J, Hermes B, Herold-Mende C, Jungk C, Kendler A, Lehman NL, Lipp E, Liu O, Mandt R, McGraw M, Mclendon R, McPherson C, Neder L, Nguyen P, Noss A, Nunziata R, Ostrom QT, Palmer C, Perin A, Pollo B, Potapov A, Potapova O, Rathmell WK, Rotin D, Scarpace L, Schilero C, Senecal K, Shimmel K, Shurkhay V, Sifri S, Singh R, Sloan AE, Smolenski K, Staugaitis SM, Steele R, Thorne L, Tirapelli DPC, Unterberg A, Vallurupalli M, Wang Y, Warnick R, Williams F, Wolinsky Y, Bell S, Rosenberg M, Stewart C, Huang F, Grimsby JL, Radenbaugh AJ & Zhang J (2015) Comprehensive,

Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med* 372, 2481–2498.

Carling T & Udelsman R (2014) Thyroid cancer. *Annu Rev Med* 65, 125–137.

Carrasco-Garcia E, Moreno-Cugnon L, Garcia I, Borras C, Revuelta M, Izeta A, Lopez-Lluch G, de Pancorbo MM, Vergara I, Vina J & Matheu A (2019) SOX2 expression diminishes with ageing in several tissues in mice and humans. *Mech Ageing Dev* 177, 30–36.

Cavalli G & Heard E (2019) Advances in epigenetics link genetics to the environment and disease. *Nature* 571, 489–499.

Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A & Borges B (2021) shiny: Web Application Framework for R. Available at: https://shiny.rstudio.com/.

Chatterjee N & Walker GC (2017) Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen* 58, 235–263.

Chen F, Zhang Q, Deng X, Zhang X, Chen C, Lv D, Li Y, Li D, Zhang Y, Li P, Diao Y, Kang L, Owen GI, Chen J & Li Z (2018) Conflicts of CpG density and DNA methylation are proximally and distally involved in gene regulation in human and mouse tissues. *Epigenetics* 13, 721–741.

Chen Y, Breeze CE, Zhen S, Beck S & Teschendorff AE (2016) Tissue-independent and tissue-specific patterns of DNA methylation alteration in cancer. *Epigenetics Chromatin* 9, 10.

Chen Z & Zhang Y (2020) Role of Mammalian DNA Methyltransferases in Development. *Annu Rev Biochem* 89, 135–158.

Cheng Q, Huang C, Cao H, Lin J, Gong X, Li J, Chen Y, Tian Z, Fang Z & Huang J (2019) A Novel Prognostic Signature of Transcription Factors for the Prediction in Patients With GBM. *Front Genet* 10, 906.

Cholewa-Waclaw J, Shah R, Webb S, Chhatbar K, Ramsahoye B, Pusch O, Yu M, Greulich P, Waclaw B & Bird AP (2019) Quantitative modelling predicts the impact of DNA methylation on RNA polymerase II traffic. *Proc Natl Acad Sci U S A* 116, 14995–15000.

Chou S-J & Tole S (2019) Lhx2, an evolutionarily conserved, multifunctional regulator of forebrain development. *Brain Research* 1705, 1–14.

Choy JS, Wei S, Lee JY, Tan S, Chu S & Lee T-H (2010) DNA methylation increases nucleosome compaction and rigidity. *J Am Chem Soc* 132, 1782–1783.

Clapier CR, Iwasa J, Cairns BR & Peterson CL (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* 18, 407–422.

Cole JJ, Robertson NA, Rather MI, Thomson JP, McBryan T, Sproul D, Wang T, Brock C, Clark W, Ideker T, Meehan RR, Miller RA, Brown-Borg HM & Adams PD (2017) Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol* 18, 58.

Cortázar D, Kunz C, Selfridge J, Lettieri T, Saito Y, MacDougall E, Wirz A, Schuermann D, Jacobs AL, Siegrist F, Steinacher R, Jiricny J, Bird A & Schär P (2011) Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* 470, 419–423.

Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, Abramowitz LK, Bartolomei MS, Rambow F, Bassi MR, Bruno T, Fanciulli M, Renner C, Klein-Szanto AJ, Matsumoto Y, Kobi D, Davidson I, Alberti C, Larue L & Bellacosa A (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146, 67–79.

Cortini R, Barbi M, Caré BR, Lavelle C, Lesne A, Mozziconacci J & Victor J-M (2016) The physics of epigenetics. *Rev. Mod. Phys.* 88, 025002.

Cubeñas-Potts C & Corces VG (2015) Architectural proteins, transcription, and the three-dimensional organization of the genome. *FEBS Lett* 589, 2923–2930.

Cui X-L, Nie J, Ku J, Dougherty U, West-Szymanski DC, Collin F, Ellison CK, Sieh L, Ning Y, Deng Z, Zhao CWT, Bergamaschi A, Pekow J, Wei J, Beadell AV, Zhang Z, Sharma G, Talwar R, Arensdorf P, Karpus J, Goel A, Bissonnette M, Zhang W, Levy S & He C (2020) A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat Commun* 11, 6161.

Cutter AR & Hayes JJ (2015) A brief review of nucleosome structure. *FEBS Lett* 589, 2914–2922.

Day JJ & Sweatt JD (2011) Epigenetic mechanisms in cognition. *Neuron* 70, 813–829.

Day K, Waite LL, Thalacker-Mercer A, West A, Bamman MM, Brooks JD, Myers RM & Absher D (2013) Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. *Genome Biol* 14, R102.

Deans C & Maggert KA (2015) What do you mean, "epigenetic"? *Genetics* 199, 887–896.

Deevy O & Bracken AP (2019) PRC2 functions in development and congenital disorders. *Development* 146, dev181354.

Dennis DJ, Han S & Schuurmans C (2019) bHLH transcription factors in neural development, disease, and reprogramming. *Brain Research* 1705, 48–65.

Dhar S, Gursoy-Yuzugullu O, Parasuram R & Price BD (2017) The tale of a tail: histone H4 acetylation and the repair of DNA breaks. *Philos Trans R Soc Lond B Biol Sci* 372, 20160284.

Dmitrijeva M, Ossowski S, Serrano L & Schaefer MH (2018) Tissue-specific DNA methylation loss during ageing and carcinogenesis is linked to chromosome structure, replication timing and cell division rates. *Nucleic Acids Research* 46, 7022–7039.

Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L & Schübeler D (2015) Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* 528, 575–579.

Douvlataniotis K, Bensberg M, Lentini A, Gylemo B & Nestor CE (2020) No evidence for DNA N 6-methyladenine in mammals. *Sci Adv* 6, eaay3335.

Du Q, Luu P-L, Stirzaker C & Clark SJ (2015) Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* 7, 1051–1073.

Dugger BN & Dickson DW (2017) Pathology of Neurodegenerative Diseases. *Cold Spring Harb Perspect Biol* 9, a028035.

Easwaran H, Johnstone SE, Van Neste L, Ohm J, Mosbruger T, Wang Q, Aryee MJ, Joyce P, Ahuja N, Weisenberger D, Collisson E, Zhu J, Yegnasubramanian S, Matsui W & Baylin SB (2012) A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res* 22, 837–849.

Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, Milekic MH, Chanrion B, Fu Y, Su T, Hibshoosh H, Gingrich JA, Haghighi F, Nutter R & Bestor TH (2010) Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Research* 20, 972–980.

Edwards JR, Yarychkivska O, Boulard M & Bestor TH (2017) DNA methylation and DNA methyltransferases. *Epigenetics Chromatin* 10, 23.

Elhamamsy AR (2017) Role of DNA methylation in imprinting disorders: an updated review. *J Assist Reprod Genet* 34, 549–562.

Emerson RO & Thomas JH (2009) Adaptive Evolution in Zinc Finger Transcription Factors S. Myers, ed. *PLoS Genet* 5, e1000325.

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.

Erlich PM, Lunetta KL, Cupples LA, Huyck M, Green RC, Baldwin CT, Farrer LA, & MIRAGE Study Group (2006) Polymorphisms in the PON gene cluster are associated with Alzheimer disease. *Hum Mol Genet* 15, 77–85.

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

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M & Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.

Espada J & Esteller M (2013) Mouse models in epigenetics: insights in development and disease. *Briefings in Functional Genomics* 12, 279–287.

Feil R & Fraga MF (2012) Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* 13, 97–109.

Feinberg AP, Koldobskiy MA & Göndör A (2016) Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet* 17, 284–299.

Feldmann A, Ivanek R, Murr R, Gaidatzis D, Burger L & Schübeler D (2013) Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet* 9, e1003994.

Felsenfeld G (2014) A brief history of epigenetics. *Cold Spring Harb Perspect Biol* 6, a018200.

Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, Ukomadu C, Sadler KC, Pradhan S, Pellegrini M & Jacobsen SE (2010) Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A* 107, 8689–8694.

Fernández AF, Bayón GF, Urdinguio RG, Toraño EG, García MG, Carella A, Petrus-Reurer S, Ferrero C, Martinez-Camblor P, Cubillo I, García-Castro J, Delgado-Calle J, Pérez-Campo FM, Riancho JA, Bueno C, Menéndez P, Mentink A, Mareschi K, Claire F, Fagnani C, Medda E, Toccaceli V, Brescianini S, Moran S, Esteller M, Stolzing A, de Boer J, Nisticò L, Stazi MA & Fraga MF (2015) H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells. *Genome Res.* 25, 27–40.

Fiorito G, Caini S, Palli D, Bendinelli B, Saieva C, Ermini I, Valentini V, Assedi M, Rizzolo P, Ambrogetti D, Ottini L & Masala G (2021) DNA methylation-based biomarkers of aging were slowed down in a two-year diet and physical activity intervention trial: the DAMA study. *Aging Cell* 20, e13439.

Fitzgerald KN, Hodges R, Hanes D, Stack E, Cheishvili D, Szyf M, Henkel J, Twedt MW, Giannopoulou D, Herdell J, Logan S & Bradley R (2021) Potential reversal of epigenetic age using a diet and lifestyle intervention: a pilot randomized clinical trial. *Aging (Albany NY)* 13, 9419–9432.

Flemming W (1882) *Zellsubstanz, Kern und Zelltheilung*, Leipzig: F.C.W. Vogel.

Foronda M, Martínez P, Schoeftner S, Gómez-López G, Schneider R, Flores JM, Pisano DG & Blasco MA (2014) Sox4 links tumor suppression to accelerated aging in mice by modulating stem cell activation. *Cell Rep* 8, 487–500.

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suñer D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu Y-Z, Plass C & Esteller M (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* 102, 10604–10609.

Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, Iyer NG, Pérez-Rosado A, Calvo E, Lopez JA, Cano A, Calasanz MJ, Colomer D, Piris MÁ, Ahn N, Imhof A, Caldas C, Jenuwein T & Esteller M (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* 37, 391–400.

Franceschi C, Garagnani P, Morsiani C, Conte M, Santoro A, Grignolio A, Monti D, Capri M & Salvioli S (2018) The Continuum of Aging and Age-Related Diseases: Common Mechanisms but Different Rates. *Front Med (Lausanne)* 5, 61.

Fransquet PD, Lacaze P, Saffery R, McNeil J, Woods R & Ryan J (2018) Blood DNA methylation as a potential biomarker of dementia: A systematic review. *Alzheimer's & Dementia* 14, 81–103.

Fransquet PD, Lacaze P, Saffery R, Phung J, Parker E, Shah R, Murray A, Woods RL & Ryan J (2020) Blood DNA methylation signatures to detect dementia prior to overt clinical symptoms. *Alzheimers Dement (Amst)* 12, e12056.

Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* 13, 840–852.

Furusawa T & Cherukuri S (2010) Developmental function of HMGN proteins. *Biochim Biophys Acta* 1799, 69–73.

Fyodorov DV, Zhou B-R, Skoultchi AI & Bai Y (2018) Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol* 19, 192–206.

Gao H, He X, Li Q, Wang Y, Tian Y, Chen X, Wang J, Guo Y, Wang W & Li X (2020) Genome-wide DNA methylome analysis reveals methylation subtypes with different clinical outcomes for acute myeloid leukemia patients. *Cancer Med* 9, 6296–6305.

Gao T, He B, Liu S, Zhu H, Tan K & Qian J (2016) EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics* 32, 3543–3551.

Garcia-Prieto CA, Álvarez-Errico D, Musulen E, Bueno-Costa A, N Vazquez B, Vaquero A & Esteller M (2022) Validation of a DNA methylation microarray for 285,000 CpG sites in the mouse genome. *Epigenetics*, 1–9.

Gensous N, Garagnani P, Santoro A, Giuliani C, Ostan R, Fabbri C, Milazzo M, Gentilini D, di Blasio AM, Pietruszka B, Madej D, Bialecka-Debek A, Brzozowska A, Franceschi C & Bacalini MG (2020) One-year Mediterranean diet promotes epigenetic rejuvenation with country- and sex-specific effects: a pilot study from the NU-AGE project. *Geroscience* 42, 687–701.

Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, Koch S, Brückl T, Biel M & Carell T (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 5, e15367.

Gong T, Borgard H, Zhang Z, Chen S, Gao Z & Deng Y (2022) Analysis and Performance Assessment of the Whole Genome Bisulfite Sequencing Data Workflow: Currently

Available Tools and a Practical Guide to Advance DNA Methylation Studies. *Small Methods* 6, e2101251.

Gorkin DU, Barozzi I, Zhao Y, Zhang Y, Huang H, Lee AY, Li B, Chiou J, Wildberg A, Ding B, Zhang B, Wang M, Strattan JS, Davidson JM, Qiu Y, Afzal V, Akiyama JA, Plajzer-Frick I, Novak CS, Kato M, Garvin TH, Pham QT, Harrington AN, Mannion BJ, Lee EA, Fukuda-Yuzawa Y, He Y, Preissl S, Chee S, Han JY, Williams BA, Trout D, Amrhein H, Yang H, Cherry JM, Wang W, Gaulton K, Ecker JR, Shen Y, Dickel DE, Visel A, Pennacchio LA & Ren B (2020) An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* 583, 744–751.

Greally JM (2018) A user's guide to the ambiguous word "epigenetics." *Nat Rev Mol Cell Biol* 19, 207–208.

Greenberg MVC & Bourc'his D (2019) The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* 20, 590–607.

Greer EL & Shi Y (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* 13, 343–357.

Griffith JS & Mahler HR (1969) DNA ticketing theory of memory. *Nature* 223, 580–582.

Guintivano J, Aryee MJ & Kaminsky ZA (2013) A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* 8, 290–302.

Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, Bryois J, Giger T, Romano L, Planchon A, Falconnet E, Bielser D, Gagnebin M, Padioleau I, Borel C, Letourneau A, Makrythanasis P, Guipponi M, Gehrig C, Antonarakis SE & Dermitzakis ET (2013) Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* 2, e00523.

Hadad N, Masser DR, Blanco-Berdugo L, Stanford DR & Freeman WM (2019) Early-life DNA methylation profiles are indicative of age-related transcriptome changes. *Epigenetics & Chromatin* 12, 58.

Haggerty C, Kretzmer H, Riemenschneider C, Kumar AS, Mattei AL, Bailly N, Gottfreund J, Giesselmann P, Weigert R, Brändl B, Giehr P, Buschow R, Galonska C, von Meyenn F, Pappalardi MB, McCabe MT, Wittler L, Giesecke-Thiel C, Mielke T, Meierhofer D, Timmermann B, Müller F-J, Walter J & Meissner A (2021) Dnmt1 has de novo activity targeted to transposable elements. *Nat Struct Mol Biol* 28, 594–603.

Hahn O, Grönke S, Stubbs TM, Ficz G, Hendrich O, Krueger F, Andrews S, Zhang Q, Wakelam MJ, Beyer A, Reik W & Partridge L (2017) Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. *Genome Biol* 18, 56.

Hall BK (1992) Waddington's Legacy in Development and Evolution. *American Zoologist* 32, 113–122.

Han H, Cortez CC, Yang X, Nichols PW, Jones PA & Liang G (2011) DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Hum Mol Genet* 20, 4299–4310.

Hannah A (1951) Localization and function of heterochromatin in Drosophila melanogaster. *Adv Genet* 4, 87–125.

Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan J-B, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T & Zhang K (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* 49, 359–367.

Hansen JC, Connolly M, McDonald CJ, Pan A, Pryamkova A, Ray K, Seidel E, Tamura S, Rogge R & Maeshima K (2018) The 10-nm chromatin fiber and its relationship to interphase chromosome organization. *Biochem Soc Trans* 46, 67–76.

Hansen JC, Maeshima K & Hendzel MJ (2021) The solid and liquid states of chromatin. *Epigenetics Chromatin* 14, 50.

Harvey W (1651) *Exercitationes de generatione animalium. Quibus accedunt quaedam de partu; de membranis ac humoribus uteri; & de conceptione*, London: Typis Du-Gardianis; Impensis O. Pulleyn.

Hasin Y, Seldin M & Lusis A (2017) Multi-omics approaches to disease. *Genome Biology* 18, 83.

Haymart MR (2009) Understanding the relationship between age and thyroid cancer. *Oncologist* 14, 216–221.

He B, Zhang C, Zhang X, Fan Y, Zeng H, Liu J, Meng H, Bai D, Peng J, Zhang Q, Tao W & Yi C (2021) Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nat Commun* 12, 4249.

He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song C-X, Zhang K, He C & Xu G-L (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H & Glass CK (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* 38, 576–589.

Henikoff S & Ahmad K (2019) Nucleosomes remember where they were. *Proc Natl Acad Sci U S A* 116, 20254–20256.

Henikoff S & Smith MM (2015) Histone variants and epigenetics. *Cold Spring Harb Perspect Biol* 7, a019364.

Hernando-Herraez I, Evano B, Stubbs T, Commere P-H, Jan Bonder M, Clark S, Andrews S, Tajbakhsh S & Reik W (2019) Ageing affects DNA methylation drift and

transcriptional cell-to-cell variability in mouse muscle stem cells. *Nat Commun* 10, 4361.

Hernando-Herraez I, Garcia-Perez R, Sharp AJ & Marques-Bonet T (2015) DNA Methylation: Insights into Human Evolution. *PLoS Genet* 11, e1005661.

Hervouet E, Peixoto P, Delage-Mourroux R, Boyer-Guittaut M & Cartron P-F (2018) Specific or not specific recruitment of DNMTs for DNA methylation, an epigenetic dilemma. *Clin Epigenetics* 10, 17.

Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F, Carmona FJ, Puca AA, Sayols S, Pujana MA, Serra-Musach J, Iglesias-Platas I, Formiga F, Fernandez AF, Fraga MF, Heath SC, Valencia A, Gut IG, Wang J & Esteller M (2012) Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A* 109, 10522–10527.

Hinrichs AS (2006) The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research* 34, D590–D598.

Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA & Noble WS (2012) Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods* 9, 473–476.

Hohman TJ, Koran MEI, Thornton-Wells TA, & Alzheimer's Disease Neuroimaging Initiative (2014) Genetic modification of the relationship between phosphorylated tau and neurodegeneration. *Alzheimers Dement* 10, 637-645.e1.

Holliday R & Pugh JE (1975) DNA modification mechanisms and gene activity during development. *Science* 187, 226–232.

Hon G, Ren B & Wang W (2008) ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome. *PLoS Comput Biol* 4, e1000201.

Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, Valsesia A, Ye Z, Kuan S, Edsall LE, Camargo AA, Stevenson BJ, Ecker JR, Bafna V, Strausberg RL, Simpson AJ & Ren B (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res* 22, 246–258.

Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD & Ren B (2013) Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* 45, 1198–1206.

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

Horvath S & Raj K (2018) DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* 19, 371–384.

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK & Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13, 86.

Howe FS, Fischl H, Murray SC & Mellor J (2017) Is H3K4me3 instructive for transcription activation? *Bioessays* 39, 1–12.

Huh I, Zeng J, Park T & Yi SV (2013) DNA methylation and transcriptional noise. *Epigenetics Chromatin* 6, 9.

Huidobro C, Fernandez AF & Fraga MF (2013) Aging epigenetics: Causes and consequences. *Molecular Aspects of Medicine* 34, 765–781.

Hwang J-Y, Aromolaran KA & Zukin RS (2017) The emerging field of epigenetics in neurodegeneration and neuroprotection. *Nat Rev Neurosci* 18, 347–361.

Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R & Bird AP (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* 6, e1001134.

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash J, Sabunciyan S & Feinberg AP (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41, 178–186.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C & Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303.

Janssen KA, Sidoli S & Garcia BA (2017) Recent Achievements in Characterizing the Histone Code and Approaches to Integrating Epigenomics and Systems Biology. *Methods Enzymol* 586, 359–378.

Jeziorska DM, Murray RJS, De Gobbi M, Gaentzsch R, Garrick D, Ayyub H, Chen T, Li E, Telenius J, Lynch M, Graham B, Smith AJH, Lund JN, Hughes JR, Higgs DR & Tufarelli C (2017) DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. *Proc Natl Acad Sci U S A* 114, E7526–E7535.

Johannsen W (1909) *Elemente der exakten erblichkeitslehre. Deutsche wesentlich erweiterte ausgabe in fünfundzwanzig vorlesungen*, Jena: G. Fischer.

Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13, 484–492.

Jones S (2004) An overview of the basic helix-loop-helix proteins. *Genome Biol* 5, 226.

Kazaure HS, Roman SA & Sosa JA (2018) The impact of age on thyroid cancer staging. *Curr Opin Endocrinol Diabetes Obes* 25, 330–334.

Kempermann G (2019) Environmental enrichment, new neurons and the neurobiology of individuality. *Nat Rev Neurosci* 20, 235–245.

Kessler NJ, Van Baak TE, Baker MS, Laritsky E, Coarfa C & Waterland RA (2016) CpG methylation differences between neurons and glia are highly conserved from mouse to human. *Hum. Mol. Genet.* 25, 223–232.

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, Linder-Basso D, Plachetka A, Shanower G, Tolstorukov MY, Luquette LJ, Xi R, Jung YL, Park RW, Bishop EP, Canfield TK, Sandstrom R, Thurman RE, MacAlpine DM, Stamatoyannopoulos JA, Kellis M, Elgin SCR, Kuroda MI, Pirrotta V, Karpen GH & Park PJ (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* 471, 480–485.

Kim H-R, Lee T, Choi JK, Jeong Y, & Alzheimer's Disease Neuroimaging Initiative (2021) Polymorphism in the MAGI2 Gene Modifies the Effect of Amyloid β on Neurodegeneration. *Alzheimer Dis Assoc Disord* 35, 114–120.

Kimura H (2013) Histone modifications for human epigenome analysis. *J Hum Genet* 58, 439–445.

Kivipelto M, Mangialasche F & Ngandu T (2018) Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nat Rev Neurol* 14, 653–666.

Klemm SL, Shipony Z & Greenleaf WJ (2019) Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* 20, 207–220.

Kochmanski J, Savonen C & Bernstein AI (2019) A Novel Application of Mixed Effects Models for Reconciling Base-Pair Resolution 5-Methylcytosine and 5-Hydroxymethylcytosine Data in Neuroepigenetics. *Front Genet* 10, 801.

Kohli RM & Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502, 472–479.

Köppen M, Simske JS, Sims PA, Firestein BL, Hall DH, Radice AD, Rongo C & Hardin JD (2001) Cooperative regulation of AJM-1 controls junctional integrity in Caenorhabditis elegans epithelia. *Nat Cell Biol* 3, 983–991.

Kowald A & Kirkwood TBL (2016) Can aging be programmed? A critical literature review. *Aging Cell* 15, 986–998.

Koyama M & Kurumizaka H (2018) Structural diversity of the nucleosome. *J Biochem* 163, 85–95.

Koziol MJ, Bradshaw CR, Allen GE, Costa ASH, Frezza C & Gurdon JB (2016) Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat Struct Mol Biol* 23, 24–30.

Kriaucionis S & Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.

Kulis M, Queirós AC, Beekman R & Martín-Subero JI (2013) Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochim Biophys Acta* 1829, 1161–1174.

Labbadia J & Morimoto RI (2015) The biology of proteostasis in aging and disease. *Annu Rev Biochem* 84, 435–464.

Lai WKM & Pugh BF (2017) Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat Rev Mol Cell Biol* 18, 548–562.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J, & International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Laskey RA & Gurdon JB (1970) Genetic content of adult somatic cells tested by nuclear transplantation from cultured cells. *Nature* 228, 1332–1334.

Lee JY & Lee T-H (2012) Effects of DNA methylation on the structure of nucleosomes. *J Am Chem Soc* 134, 173–175.

Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitsel EA, Wilson JG, Reiner AP, Aviv A, Lohman K, Liu Y, Ferrucci L & Horvath S (2018) An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* 10, 573–591.

Li E, Bestor TH & Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915–926.

Li J, Huang Q, Zeng F, Li W, He Z, Chen W, Zhu W & Zhang B (2014) The prognostic value of global DNA hypomethylation in cancer: a meta-analysis. *PLoS One* 9, e106290.

Li X & Fu X-D (2019) Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat Rev Genet* 20, 503–519.

Ling C & Rönn T (2019) Epigenetics in Human Obesity and Type 2 Diabetes. *Cell Metab* 29, 1028–1044.

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, Behrens MM & Ecker JR (2013) Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* 341, 1237905–1237905.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B & Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322.

López-Otín C, Blasco MA, Partridge L, Serrano M & Kroemer G (2013) The Hallmarks of Aging. *Cell* 153, 1194–1217.

Lu T, Aron L, Zullo J, Pan Y, Kim H, Chen Y, Yang T-H, Kim H-M, Drake D, Liu XS, Bennett DA, Colaiácovo MP & Yankner BA (2014) REST and stress resistance in ageing and Alzheimer's disease. *Nature* 507, 448–454.

Lyko F (2018) The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet* 19, 81–92.

Ma P & Schultz RM (2016) HDAC1 and HDAC2 in mouse oocytes and preimplantation embryos: Specificity versus compensation. *Cell Death Differ* 23, 1119–1127.

Maeshima K, Ide S & Babokhov M (2019) Dynamic chromatin organization without the 30-nm fiber. *Curr Opin Cell Biol* 58, 95–104.

Mah L-J, El-Osta A & Karagiannis TC (2010) gammaH2AX: a sensitive molecular marker of DNA damage and repair. *Leukemia* 24, 679–686.

Malta TM, de Souza CF, Sabedot TS, Silva TC, Mosella MS, Kalkanis SN, Snyder J, Castro AVB & Noushmehr H (2018) Glioma CpG island methylator phenotype (G-CIMP): biological and clinical implications. *Neuro Oncol* 20, 608–620.

Mammana A & Chung H-R (2015) Chromatin segmentation based on a probabilistic model for read counts explains a large portion of the epigenome. *Genome Biol* 16, 151.

Martens JHA & Stunnenberg HG (2013) BLUEPRINT: mapping human blood cell epigenomes. *Haematologica* 98, 1487–1489.

Martínez-Zamudio RI, Roux P-F, de Freitas JANLF, Robinson L, Doré G, Sun B, Belenki D, Milanovic M, Herbig U, Schmitt CA, Gil J & Bischof O (2020) AP-1 imprints a reversible transcriptional programme of senescent cells. *Nat Cell Biol* 22, 842–855.

Martino D, Loke YJ, Gordon L, Ollikainen M, Cruickshank MN, Saffery R & Craig JM (2013) Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol* 14, R42.

Martino DJ, Tulic MK, Gordon L, Hodder M, Richman TR, Metcalfe J, Prescott SL & Saffery R (2011) Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics* 6, 1085–1094.

Martire S & Banaszynski LA (2020) The roles of histone variants in fine-tuning chromatin organization and function. *Nat Rev Mol Cell Biol* 21, 522–541.

Marx V (2013) The big challenges of big data. *Nature* 498, 255–260.

Masser DR, Hadad N, Porter HL, Mangold CA, Unnikrishnan A, Ford MM, Giles CB, Georgescu C, Dozmorov MG, Wren JD, Richardson A, Stanford DR & Freeman WM (2017) Sexually divergent DNA methylation patterns with hippocampal aging. *Aging Cell* 16, 1342–1352.

McGinty RK & Tan S (2015) Nucleosome structure and function. *Chem Rev* 115, 2255–2273.

McKinney BC, Lin C-W, Rahman T, Oh H, Lewis DA, Tseng G & Sibille E (2019) DNA methylation in the human frontal cortex reveals a putative mechanism for age-by-disease interactions. *Transl Psychiatry* 9, 39.

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES & Jaenisch R (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 33, 5868–5877.

de Mendoza A, Lister R & Bogdanovic O (2019) Evolution of DNA Methylome Diversity in Eukaryotes. *J Mol Biol*, S0022-2836(19)30659-X.

de Mendoza A, Poppe D, Buckberry S, Pflueger J, Albertin CB, Daish T, Bertrand S, de la Calle-Mustienes E, Gómez-Skarmeta JL, Nery JR, Ecker JR, Baer B, Ragsdale CW, Grützner F, Escriva H, Venkatesh B, Bogdanovic O & Lister R (2021) The emergence of the brain non-CpG methylation system in vertebrates. *Nat Ecol Evol* 5, 369–378.

Michaud EJ, van Vugt MJ, Bultman SJ, Sweet HO, Davisson MT & Woychik RP (1994) Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. *Genes Dev* 8, 1463–1472.

Molina-Serrano D, Kyriakou D & Kirmizis A (2019) Histone Modifications as an Intersection Between Diet and Longevity. *Front Genet* 10, 192.

Morgan DO (2007) *The Cell Cycle: Principles of Control*, London: New Science Press.

Morgan TH (1911) An attempt to analyze the constitution of the chromosomes on the basis of sex-limited inheritance in Drosophila. *Journal of Experimental Zoology* 11, 365–413.

Muller HJ (1930) Types of visible variations induced by X-rays inDrosophila. *Journ. of Gen.* 22, 299–334.

Nelson PT, Fardo DW & Katsumata Y (2020) The MUC6/AP2A2 Locus and Its Relevance to Alzheimer's Disease: A Review. *J Neuropathol Exp Neurol* 79, 568–584.

Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, Maldotti M, Anselmi F & Oliviero S (2017) Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 543, 72–77.

Niccoli T & Partridge L (2012) Ageing as a risk factor for disease. *Curr Biol* 22, R741-752.

Nicoglou A & Wolfe CT (2018) Introduction: sketches of a conceptual history of epigenesis. *Hist Philos Life Sci* 40, 64.

Nikopoulou C, Parekh S & Tessarz P (2019) Ageing and sources of transcriptional heterogeneity. *Biol Chem* 400, 867–878.

Nordman JC & Kabbani N (2012) An interaction between α7 nicotinic receptors and a G-protein pathway complex regulates neurite growth in neural cells. *J Cell Sci* 125, 5502–5513.

Oblak L, van der Zaag J, Higgins-Chen AT, Levine ME & Boks MP (2021) A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. *Ageing Res Rev* 69, 101348.

Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, Berman DM, Jenuwein T, Pruitt K, Sharkis SJ, Watkins DN, Herman JG & Baylin SB (2007) A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* 39, 237–242.

Okano M, Bell DW, Haber DA & Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257.

Oku Y & Huganir RL (2013) AGAP3 and Arf6 Regulate Trafficking of AMPA Receptors and Synaptic Plasticity. *Journal of Neuroscience* 33, 12586–12598.

Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir GA, Stewart R & Thomson JA (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1, 299–312.

Pappalardo XG & Barra V (2021) Losing DNA methylation at repetitive elements and breaking bad. *Epigenetics Chromatin* 14, 25.

Pastor WA, Aravind L & Rao A (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* 14, 341–356.

Penn NW, Suwalski R, O'Riley C, Bojanowski K & Yura R (1972) The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem J* 126, 781–790.

Pérez RF, Tejedor JR, Bayón GF, Fernández AF & Fraga MF (2018) Distinct chromatin signatures of DNA hypomethylation in aging and cancer. *Aging Cell* 17, e12744.

Petryk N, Bultmann S, Bartke T & Defossez P-A (2021) Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. *Nucleic Acids Res* 49, 3020–3032.

Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Djik S, Muhlhausler B, Stirzaker C & Clark SJ (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* 17, 208.

Pon JR & Marra MA (2015) Driver and passenger mutations in cancer. *Annu Rev Pathol* 10, 25–50.

Popa-Wagner A, Dumitrascu DI, Capitanescu B, Petcu EB, Surugiu R, Fang W-H & Dumbrava D-A (2020) Dietary habits, lifestyle factors and neurodegenerative diseases. *Neural Regen Res* 15, 394–400.

Portela A & Esteller M (2010) Epigenetic modifications and human disease. *Nat Biotechnol* 28, 1057–1068.

Pradeepa MM (2017) Causal role of histone acetylations in enhancer function. *Transcription* 8, 40–47.

de Pretis S & Pelizzola M (2014) Computational and experimental methods to decipher the epigenetic code. *Front Genet* 5, 335.

Putiri EL, Tiedemann RL, Thompson JJ, Liu C, Ho T, Choi J-H & Robertson KD (2014) Distinct and overlapping control of 5-methylcytosine and 5-hydroxymethylcytosine by the TET proteins in human cancer cells. *Genome Biol* 15, R81.

Rakyan VK, Down TA, Maslau S, Andrew T, Yang T-P, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P & Spector TD (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20, 434–439.

Rando OJ (2012) Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Curr Opin Genet Dev* 22, 148–155.

Reeves R (2010) Nuclear functions of the HMG proteins. *Biochim Biophys Acta* 1799, 3–14.

Riggs AD (1975) X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* 14, 9–25.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu Y-C, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh K-H, Feizi S, Karlic R, Kim A-R, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJM, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai L-H, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T & Kellis M (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330.

Rodríguez-Paredes M & Esteller M (2011) Cancer epigenetics reaches mainstream oncology. *Nat Med* 17, 330–339.

Rollins RA, Haghighi F, Edwards JR, Das R, Zhang MQ, Ju J & Bestor TH (2006) Large-scale structure of genomic methylation patterns. *Genome Res* 16, 157–163.

Ross SE & Bogdanovic O (2019) TET enzymes, DNA demethylation and pluripotency. *Biochem Soc Trans* 47, 875–885.

Rossetto D, Avvakumov N & Côté J (2012) Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics* 7, 1098–1108.

Saksouk N, Simboeck E & Déjardin J (2015) Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* 8, 3.

Sanchez-Mut JV, Heyn H, Silva BA, Dixsaut L, Garcia-Esparcia P, Vidal E, Sayols S, Glauser L, Monteagudo-Sánchez A, Perez-Tur J, Ferrer I, Monk D, Schneider B, Esteller M & Gräff J (2018) PM20D1 is a quantitative trait locus associated with Alzheimer's disease. *Nat Med* 24, 598–603.

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I & Cedar H (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* 39, 232–236.

Schroeder DI, Jayashankar K, Douglas KC, Thirkill TL, York D, Dickinson PJ, Williams LE, Samollow PB, Ross PJ, Bannasch DL, Douglas GC & LaSalle JM (2015) Early Developmental and Evolutionary Origins of Gene Body DNA Methylation Patterns in Mammalian Placentas G. Kelsey, ed. *PLoS Genet* 11, e1005442.

Schuettengruber B, Bourbon H-M, Di Croce L & Cavalli G (2017) Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* 171, 34–57.

Schutsky EK, DeNizio JE, Hu P, Liu MY, Nabel CS, Fabyanic EB, Hwang Y, Bushman FD, Wu H & Kohli RM (2018) Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. *Nat Biotechnol*.

Seale K, Horvath S, Teschendorff A, Eynon N & Voisin S (2022) Making sense of the ageing methylome. *Nat Rev Genet*, 1–21.

Sen P, Shah PP, Nativio R & Berger SL (2016) Epigenetic Mechanisms of Longevity and Aging. *Cell* 166, 822–839.

Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U & Gutierrez C (2014) The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. *Plant Cell* 26, 2351–2366.

Sharifi-Zarchi A, Gerovska D, Adachi K, Totonchi M, Pezeshk H, Taft RJ, Schöler HR, Chitsaz H, Sadeghi M, Baharvand H & Araúzo-Bravo MJ (2017) DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *BMC Genomics* 18, 964.

Sharrocks AD (2001) The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol* 2, 827–837.

Shayevitch R, Askayo D, Keydar I & Ast G (2018) The importance of DNA methylation of exons on alternative splicing. *RNA* 24, 1351–1362.

Shin J-Y & Worman HJ (2022) Molecular Pathology of Laminopathies. *Annu Rev Pathol* 17, 159–180.

Silva TC, Young JI, Zhang L, Gomez L, Schmidt MA, Varma A, Chen XS, Martin ER & Wang L (2022) Cross-tissue analysis of blood and brain epigenome-wide association studies in Alzheimer's disease. *Nat Commun* 13, 4852.

Simpkin AJ, Suderman M, Gaunt TR, Lyttleton O, McArdle WL, Ring SM, Tilling K, Davey Smith G & Relton CL (2015) Longitudinal analysis of DNA methylation associated with birth weight and gestational age. *Hum Mol Genet* 24, 3752–3763.

Singh R, Bassett E, Chakravarti A & Parthun MR (2018) Replication-dependent histone isoforms: a new source of complexity in chromatin structure and function. *Nucleic Acids Res* 46, 8665–8678.

Sizemore GM, Pitarresi JR, Balakrishnan S & Ostrowski MC (2017) The ETS family of oncogenic transcription factors in solid tumours. *Nat Rev Cancer* 17, 337–351.

Slieker RC, van Iterson M, Luijk R, Beekman M, Zhernakova DV, Moed MH, Mei H, van Galen M, Deelen P, Bonder MJ, Zhernakova A, Uitterlinden AG, Tigchelaar EF, Stehouwer CDA, Schalkwijk CG, van der Kallen CJH, Hofman A, van Heemst D, de Geus EJ, van Dongen J, Deelen J, van den Berg LH, van Meurs J, Jansen R, 't Hoen PAC, Franke L, Wijmenga C, Veldink JH, Swertz MA, van Greevenbroek MMJ, van Duijn CM, Boomsma DI, BIOS consortium, Slagboom PE & Heijmans BT (2016) Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. *Genome Biol* 17, 191.

Slonim DK & Yanai I (2009) Getting started in gene expression microarray analysis. *PLoS Comput Biol* 5, e1000543.

Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen C-H, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P & He C (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* 29, 68–72.

Srivastava R, Srivastava R & Ahn SH (2016) The Epigenetic Pathways to Ribosomal DNA Silencing. *Microbiol. Mol. Biol. Rev.* 80, 545–563.

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK & Schübeler D (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495.

Stark R, Grzelak M & Hadfield J (2019) RNA sequencing: the teenage years. *Nat Rev Genet* 20, 631–656.

Strahl BD & Allis CD (2000) The language of covalent histone modifications. *Nature* 403, 41–45.

Stunnenberg HG, International Human Epigenome Consortium & Hirst M (2016) The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* 167, 1145–1149.

Sturtevant AH (1913) The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *Journal of Experimental Zoology* 14, 43–59.

Sun D, Luo M, Jeong M, Rodriguez B, Xia Z, Hannah R, Wang H, Le T, Faull KF, Chen R, Gu H, Bock C, Meissner A, Göttgens B, Darlington GJ, Li W & Goodell MA (2014) Epigenomic Profiling of Young and Aged HSCs Reveals Concerted Changes during Aging that Reinforce Self-Renewal. *Cell Stem Cell* 14, 673–688.

Sun Z, Zhang Y, Jia J, Fang Y, Tang Y, Wu H & Fang D (2020) H3K36me3, message from chromatin to DNA damage repair. *Cell Biosci* 10, 9.

Sundaramoorthy R & Owen-Hughes T (2020) Chromatin remodelling comes into focus. *F1000Res* 9, F1000 Faculty Rev-1011.

Szwagierczak A, Bultmann S, Schmidt CS, Spada F & Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* 38, e181.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L & Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.

Talbert PB & Henikoff S (2017) Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol* 18, 115–126.

Tejedor JR & Fraga MF (2017) Interindividual epigenetic variability: Sound or noise? *BioEssays* 39, 1700055.

Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP, Savage DA, Mueller-Holzner E, Marth C, Kocjan G, Gayther SA, Jones A, Beck S, Wagner W, Laird PW, Jacobs IJ & Widschwendter M (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* 20, 440–446.

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

Teschendorff AE & Zheng SC (2017) Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics* 9, 757–768.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutyavin T, Lajoie B, Lee B-K, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE & Stamatoyannopoulos JA (2012) The accessible chromatin landscape of the human genome. *Nature* 489, 75–82.

Trojer P & Reinberg D (2007) Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* 28, 1–13.

Tronick E & Hunter RG (2016) Waddington, Dynamic Systems, and Epigenetics. *Front Behav Neurosci* 10, 107.

Tsai ZT-Y, Shiu S-H & Tsai H-K (2015) Contribution of Sequence Motif, Chromatin State, and DNA Structure Features to Predictive Models of Transcription Factor Binding in Yeast. *PLoS Comput Biol* 11, e1004418.

Unnikrishnan A, Hadad N, Masser DR, Jackson J, Freeman WM & Richardson A (2018) Revisiting the genomic hypomethylation hypothesis of aging. *Ann N Y Acad Sci* 1418, 69–79.

Urdinguio RG, Torró MI, Bayón GF, Álvarez-Pitti J, Fernández AF, Redon P, Fraga MF & Lurbe E (2016) Longitudinal study of DNA methylation during the first 5 years of life. *J Transl Med* 14, 160.

Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams BA, Stamatoyannopoulos JA, Crawford GE, Absher DM, Wold BJ & Myers RM (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* 23, 555–567.

Vastenhouw NL & Schier AF (2012) Bivalent histone modifications in early embryogenesis. *Curr Opin Cell Biol* 24, 374–386.

van der Velde A, Fan K, Tsuji J, Moore JE, Purcaro MJ, Pratt HE & Weng Z (2021) Annotation of chromatin states in 66 complete mouse epigenomes during development. *Commun Biol* 4, 239.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K,

Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A & Zhu X (2001) The sequence of the human genome. *Science* 291, 1304–1351.

Vető B, Szabó P, Bacquet C, Apró A, Hathy E, Kiss J, Réthelyi JM, Szeri F, Szüts D & Arányi T (2018) Inhibition of DNA methyltransferase leads to increased genomic 5-hydroxymethylcytosine levels in hematopoietic cells. *FEBS Open Bio* 8, 584–592.

Vu H & Ernst J (2022) Universal annotation of the human genome through integration of over a thousand epigenomic datasets. *Genome Biol* 23, 9.

Waddington CH (1940) *Organisers and genes*, London: Cambridge University Press.

Waddington CH (1942) The epigenotype. *Endeavour* 1, 18–20.

Wagner EJ & Carpenter PB (2012) Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol* 13, 115–126.

Walton EL, Francastel C & Velasco G (2011) Maintenance of DNA methylation: Dnmt3b joins the dance. *Epigenetics* 6, 1373–1377.

Wang T, Tsui B, Kreisberg JF, Robertson NA, Gross AM, Yu MK, Carter H, Brown-Borg HM, Adams PD & Ideker T (2017) Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol* 18, 57.

Wang X & Hayes JJ (2008) Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure. *Mol Cell Biol* 28, 227–236.

Waterland RA & Jirtle RL (2003) Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* 23, 5293–5300.

Weikum ER, Liu X & Ortlund EA (2018) The nuclear receptor superfamily: A structural perspective. *Protein Sci* 27, 1876–1892.

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I & Laird PW (2007) Epigenetic stem cell signature in cancer. *Nat Genet* 39, 157–158.

Wiles ET & Selker EU (2017) H3K27 methylation: a promiscuous repressive chromatin mark. *Curr Opin Genet Dev* 43, 31–37.

Wilhelm-Benartzi CS, Koestler DC, Karagas MR, Flanagan JM, Christensen BC, Kelsey KT, Marsit CJ, Houseman EA & Brown R (2013) Review of processing and analysis methods for DNA methylation array data. *Br J Cancer* 109, 1394–1402.

Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K, Liu Y, Byrum SD, Mackintosh SG, Zhong M, Tackett A, Wang G, Hon LS, Fang G, Swenberg JA & Xiao AZ (2016) DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* 532, 329–333.

Xia X, Jiang Q, McDermott J & Han J-DJ (2018) Aging and Alzheimer's disease: Comparison and associations from molecular to system level. *Aging Cell* 17, e12802.

Xu C & Corces VG (2018) Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. *Science* 359, 1166–1170.

Yan L & Chen Z (2020) A Unifying Mechanism of DNA Translocation Underlying Chromatin Remodeling. *Trends Biochem Sci* 45, 217–227.

Yang J, Huang J, Dasgupta M, Sears N, Miyagi M, Wang B, Chance MR, Chen X, Du Y, Wang Y, An L, Wang Q, Lu T, Zhang X, Wang Z & Stark GR (2010) Reversible methylation of promoter-bound STAT3 by histone-modifying enzymes. *Proc Natl Acad Sci U S A* 107, 21499–21504.

Yap KL & Zhou M-M (2010) Keeping it in the family: diverse histone recognition by conserved structural folds. *Crit Rev Biochem Mol Biol* 45, 488–505.

Yi SV (2017) Insights into Epigenome Evolution from Animal and Plant Methylomes. *Genome Biol Evol* 9, 3189–3201.

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schübeler D, Vinson C & Taipale J (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 356, eaaj2239.

Yong W-S, Hsu F-M & Chen P-Y (2016) Profiling genome-wide DNA methylation. *Epigenetics Chromatin* 9, 26.

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

Zamudio N, Barau J, Teissandier A, Walter M, Borsos M, Servant N & Bourc'his D (2015) DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev* 29, 1256–1270.

Zeng J, Nagrajan HK & Yi SV (2014) Fundamental diversity of human CpG islands at multiple biological levels. *Epigenetics* 9, 483–491.

Zhang D, Tang Z, Huang H, Zhou G, Cui C, Weng Y, Liu W, Kim S, Lee S, Perez-Neut M, Ding J, Czyz D, Hu R, Ye Z, He M, Zheng YG, Shuman HA, Dai L, Ren B, Roeder

RG, Becker L & Zhao Y (2019) Metabolic regulation of gene expression by histone lactylation. *Nature* 574, 575–580.

Zhang D, Wu B, Wang P, Wang Y, Lu P, Nechiporuk T, Floss T, Greally JM, Zheng D & Zhou B (2017) Non-CpG methylation by DNMT3B facilitates REST binding and gene silencing in developing mouse hearts. *Nucleic Acids Res* 45, 3102–3115.

Zhang W & Xu J (2017) DNA methyltransferases and their roles in tumorigenesis. *Biomark Res* 5, 1.

Zhao S, Allis CD & Wang GG (2021) The language of chromatin modification in human cancers. *Nat Rev Cancer* 21, 413–430.

Zhao Y & Garcia BA (2015) Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol* 7, a025064.

Zhou J, Sears RL, Xing X, Zhang B, Li D, Rockweiler NB, Jang HS, Choudhary MNK, Lee HJ, Lowdon RF, Arand J, Tabers B, Gu CC, Cicero TJ & Wang T (2017) Tissue-specific DNA methylation is conserved across human, mouse, and rat, and driven by primary sequence conservation. *BMC Genomics* 18, 724.

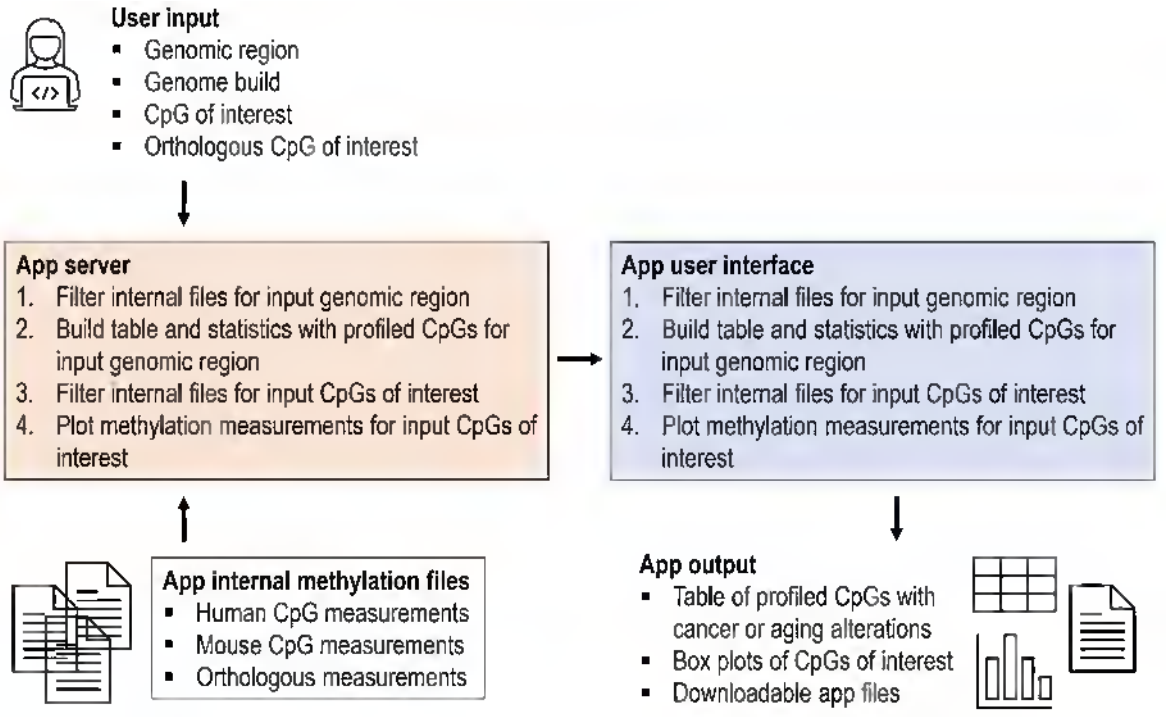
Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, Laird PW & Berman BP (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* 50, 591–602.

Zhou W, Hinoue T, Barnes B, Mitchell O, Iqbal W, Lee SM, Foy KK, Lee K-H, Moyer EJ, VanderArk A, Koeman JM, Ding W, Kalkat M, Spix NJ, Eagleson B, Pospisilik JA, Szabó PE, Bartolomei MS, Vander Schaaf NA, Kang L, Wiseman AK, Jones PA, Krawczyk CM, Adams M, Porecha R, Chen BH, Shen H & Laird PW (2022) DNA methylation dynamics and dysregulation delineated by high-throughput profiling in the mouse. *Cell Genom* 2, 100144.

Annex

ANNEX 1. DEVELOPMENT OF A WEB DATABASE APPLICATION CONTAINING AN INTERACTIVE DATABASE OF AGING AND CANCER DNA METHYLATION ALTERATIONS IN HUMAN AND MOUSE

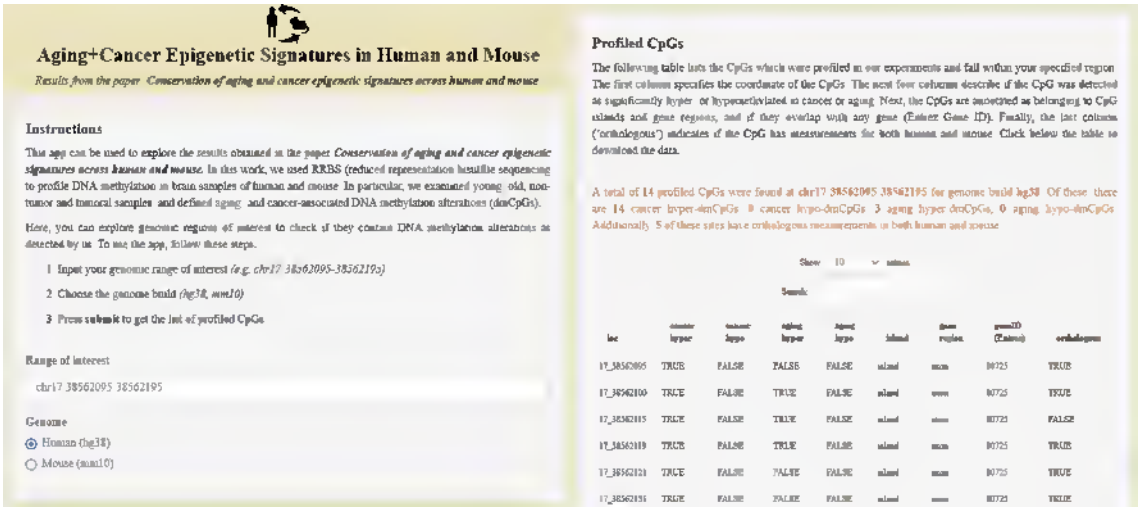
Part of the results detailed in the thesis publication "**Conservation of Aging and Cancer Epigenetic Signatures across Human and Mouse**" involved the development of a web database application using the R shiny application framework (Chang et al. 2021). This application is an interactive database which can be accessed from the following web address: https://epilabasturias.shinyapps.io/mbe_app/. The goal of the application is to output information on genomic locations for which aging- and cancer-associated DNA methylation alterations were detected in human or mouse in the original study. The internal structure of the application is indicated in the following **ANNEX FIGURE 1**:



Annex figure 1. Representation of the structure of the web application.

The use of the application is as follows:

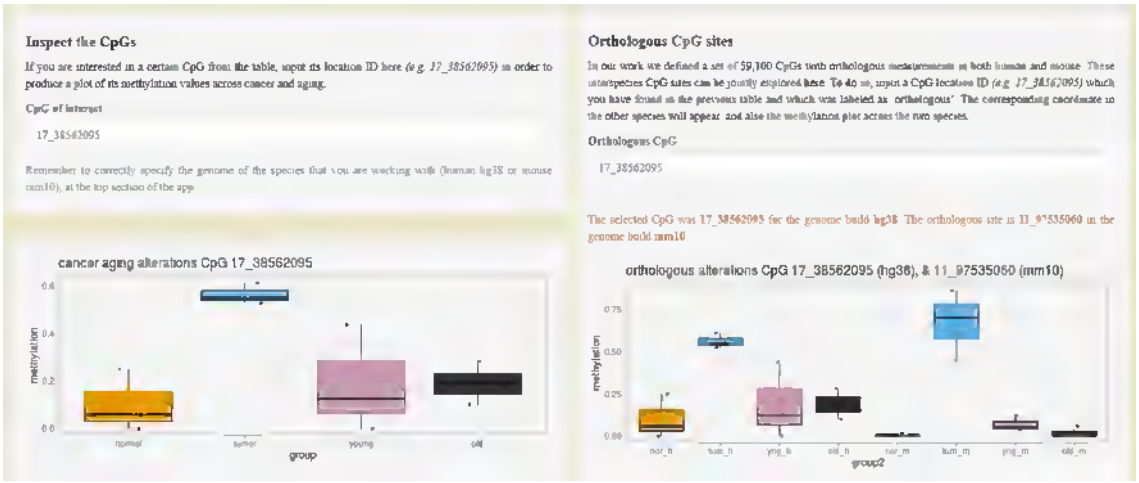
1. The user first inputs a genomic region of interest and a genome build in the first panel (**ANNEX FIGURE 2**, left panel).



Annex figure 2. Screenshots showing the input panel of the application (left) and the output table returned (right).

2. A table describing the CpGs profiled in the study for the genomic region is output (**ANNEX FIGURE 2**, right panel). This table can be downloaded from the application, as well as the full internal methylation tables used by the application.

3. Finally, the user can input CpGs of interest using the internal application identifiers to plot their DNA methylation values across the study groups (**ANNEX FIGURE 3**).



Annex figure 3. Screenshots showing the graphical output panels of the application.

ANNEX 2. SCIENTIFIC CONTRIBUTIONS

This section lists the scientific publications contributed to by the author during the doctoral thesis period (the author's name is **highlighted**):

First or co-first publications:

---

1. **Pérez RF**, Tejedor JR, Fernández AF & Fraga MF (2022) Aging and cancer epigenetics: where do the paths fork? Aging Cell, *in press*, e13709.
2. **Pérez RF**, Alba-Linares JJ, Tejedor JR, Fernández AF, Calero M, Román-Domínguez A, Borrás C, Viña J, Ávila J, Medina M & Fraga MF (2022) Blood DNA methylation patterns in older adults with evolving dementia. J Gerontol A Biol Sci Med Sci, 77(9): 1743–1749.
3. **Pérez RF**, Tejedor JR, Santamarina-Ojeda P, Martínez VL, Urdinguio RG, Villamañán L, Candiota AP, Sarró NMV, Barradas M, Fernandez-Marcos PJ, Serrano M, Fernández AF & Fraga MF (2021) Conservation of Aging and Cancer Epigenetic Signatures across Human and Mouse. Mol Biol Evol 38, 3415–3435.
4. Martínez-García GG, **Pérez RF**, Fernández ÁF, Durand S, Kroemer G & Mariño G (2021) Autophagy Deficiency by Atg4B Loss Leads to Metabolomic Alterations in Mice. Metabolites 11, 481.
5. **Pérez RF**, Fernandez-Morera JL, Romano-Garcia J, Menendez-Torre E, Delgado-Alvarez E, Fraga MF & Fernandez AF (2020) DNA Methylomes and Epigenetic Age Acceleration Associations with Poor Metabolic Control in T1D. Biomedicines 9, 13.
6. **Pérez RF**, Soto Fernández AY, Bousquets Muñoz P, Sierra MI, Tejedor JR, Morales-Sánchez P, Valdés AF, Santamaría R, Blanco C, Torrecillas R, Fraga MF & Fernández AF (2020) No genome-wide DNA methylation changes found associated with medium-term reduced graphene oxide exposure in human lung epithelial cells. Epigenetics 15, 283–293.
7. **Pérez RF**, Santamarina P, Tejedor JR, Urdinguio RG, Álvarez-Pitti J, Redon P, Fernández AF, Fraga MF & Lurbe E (2019) Longitudinal genome-wide DNA

methylation analysis uncovers persistent early-life DNA methylation changes. J Transl Med 17, 15.

8. **Pérez RF**, Santamarina P, Fernández AF & Fraga MF (2019) Epigenetics and Lifestyle: The Impact of Stress, Diet, and Social Habits on Tissue Homeostasis. In Epigenetics and Regeneration. Elsevier, pp.461–489.
9. **Pérez RF**, Tejedor JR, Bayón GF, Fernández AF & Fraga MF (2018) Distinct chromatin signatures of DNA hypomethylation in aging and cancer. Aging Cell 17, e12744.

Other publications:

10. Tejedor JR, Martín G, Roberti A, Mangas C, Santamarina-Ojeda P, **Pérez RF**, López V, González Urdinguio R, Alba-Linares JJ, Peñarroya A, Álvarez-Argüelles ME, Boga JA, Fernández Fernández A, Rojo-Alba S & Fernández Fraga M (2022) Enhanced Detection of Viral RNA Species Using FokI-Assisted Digestion of DNA Duplexes and DNA/RNA Hybrids. Anal Chem 94, 6760–6770.
11. Tejedor JR, Bueno C, Vinyoles M, Petazzi P, Agraz-Doblas A, Cobo I, Torres-Ruiz R, Bayón GF, **Pérez RF**, López-Tamargo S, Gutierrez-Agüera F, Santamarina-Ojeda P, Ramírez-Orellana M, Bardini M, Cazzaniga G, Ballerini P, Schneider P, Stam RW, Varela I, Fraga MF, Fernández AF & Menéndez P (2021) Integrative methylome-transcriptome analysis unravels cancer cell vulnerabilities in infant MLL-rearranged B cell acute lymphoblastic leukemia. J Clin Invest 131, 138833.
12. López V, Tejedor JR, Carella A, García MG, Santamarina-Ojeda P, **Pérez RF**, Mangas C, Urdinguio RG, Aranburu A, de la Nava D, Corte-Torres MD, Astudillo A, Mollejo M, Meléndez B, Fernández AF & Fraga MF (2021) Epigenetic Deregulation of the Histone Methyltransferase KMT5B Contributes to Malignant Transformation in Glioblastoma. Front Cell Dev Biol 9, 671838.
13. Urdinguio RG, Tejedor JR, Fernández-Sanjurjo M, **Pérez RF**, Peñarroya A, Ferrero C, Codina-Martínez H, Díez-Planelles C, Pinto-Hernández P, Castilla-Silgado J, Coto-

Vilcapoma A, Díez-Robles S, Blanco-Agudín N, Tomás-Zapico C, Iglesias-Gutiérrez E, Fernández-García B, Fernandez AF & Fraga MF (2021) Physical exercise shapes the mouse brain epigenome. Mol Metab 54, 101398.

14. Fraga MF, Ramos S, Bravo L, Collado MC, **Pérez RF**, Gómara B, Urdinguio RG, Martín MÁ, Herrero A & Ramos L (2021) Epigenomics and life style. In Volume 3: Genome & Epigenetics. Editorial CSIC, pp.138–158.
15. Carella A, Tejedor JR, García MG, Urdinguio RG, Bayón GF, Sierra M, López V, García-Toraño E, Santamarina-Ojeda P, **Pérez RF**, Bigot T, Mangas C, Corte-Torres MD, Sáenz-de-Santa-María I, Mollejo M, Meléndez B, Astudillo A, Chiara MD, Fernández AF & Fraga MF (2020) Epigenetic downregulation of TET3 reduces genome-wide 5hmC levels and promotes glioblastoma tumorigenesis. Int J Cancer 146, 373–387.
16. Morales-Sánchez P, **Pérez RF**, Santamarina P, Rodriguez-Rodero S, Fernandez-Fernandez A & Fraga MF (2019) Epigenetics: At the Crossroads Between Genetic and Environmental Determinants of Disease. In Bone Health: A Reflection of the Social Mosaic. Singapore: Springer, pp.105–128.
17. Urdinguio RG, Lopez V, Bayón GF, Diaz de la Guardia R, Sierra MI, García-Toraño E, **Pérez RF**, García MG, Carella A, Pruneda PC, Prieto C, Dmitrijeva M, Santamarina P, Belmonte T, Mangas C, Diaconu E, Ferrero C, Tejedor JR, Fernandez-Morera JL, Bravo C, Bueno C, Sanjuan-Pla A, Rodriguez RM, Suarez-Alvarez B, López-Larrea C, Bernal T, Colado E, Balbín M, García-Suarez O, Chiara MD, Sáenz-de-Santa-María I, Rodríguez F, Pando-Sandoval A, Rodrigo L, Santos L, Salas A, Vallejo-Díaz J, C Carrera A, Rico D, Hernández-López I, Vayá A, Ricart JM, Seto E, Sima-Teruel N, Vaquero A, Valledor L, Cañal MJ, Pisano D, Graña-Castro O, Thomas T, Voss AK, Menéndez P, Villar-Garea A, Deutzmann R, Fernandez AF & Fraga MF (2019) Chromatin regulation by Histone H4 acetylation at Lysine 16 during cell death and differentiation in the myeloid compartment. Nucleic Acids Res 47, 5016–5037.

18. Fernandez AF, Bayón GF, Sierra MI, Urdinguio RG, Toraño EG, García MG, Carella A, López V, Santamarina P, **Pérez RF**, Belmonte T, Tejedor JR, Cobo I, Menendez P, Mangas C, Ferrero C, Rodrigo L, Astudillo A, Ortea I, Cueto Díaz S, Rodríguez-Gonzalez P, García Alonso JI, Mollejo M, Meléndez B, Domínguez G, Bonilla F & Fraga MF (2018) Loss of 5hmC identifies a new type of aberrant DNA hypermethylation in glioma. Hum Mol Genet 27, 3046–3059.

19. García MG, Carella A, Urdinguio RG, Bayón GF, Lopez V, Tejedor JR, Sierra MI, García-Toraño E, Santamarina P, **Pérez RF**, Mangas C, Astudillo A, Corte-Torres MD, Sáenz-de-Santa-María I, Chiara M-D, Fernández AF & Fraga MF (2018) Epigenetic dysregulation of TET2 in human glioblastoma. Oncotarget 9, 25922–25934.

20. Rabal O, San José-Enériz E, Agirre X, Sánchez-Arias JA, Vilas-Zornoza A, Ugarte A, de Miguel I, Miranda E, Garate L, Fraga M, Santamarina P, **Pérez RF**, Ordoñez R, Sáez E, Roa S, García-Barchino MJ, Martínez-Climent JA, Liu Y, Wu W, Xu M, Prosper F & Oyarzabal J (2018) Discovery of Reversible DNA Methyltransferase and Lysine Methyltransferase G9a Inhibitors with Antitumoral in Vivo Efficacy. J Med Chem 61, 6518–6545.

21. Tejedor JR, Bueno C, Cobo I, Bayón GF, Prieto C, Mangas C, **Pérez RF**, Santamarina P, Urdinguio RG, Menéndez P, Fraga MF & Fernández AF (2018) Epigenome-wide analysis reveals specific DNA hypermethylation of T cells during human hematopoietic differentiation. Epigenomics 10, 903–923.

22. Bernardo-Castiñeira C, Valdés N, Sierra MI, Sáenz-de-Santa-María I, Bayón GF, **Perez RF**, Fernández AF, Fraga MF, Astudillo A, Menéndez R, Fernández B, Del Olmo M, Suarez C & Chiara M-D (2018) SDHC Promoter Methylation, a Novel Pathogenic Mechanism in Parasympathetic Paragangliomas. J Clin Endocrinol Metab 103, 295–305.