Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biodiversità ed Evoluzione

Ciclo XXVII

Settore Concorsuale di afferenza: 05/B1
Settore Scientifico disciplinare: BIO/08

EVOLUTIONARY EPIGENETICS OF MODERN HUMAN POPULATIONS

Presentata da: Cristina Giuliani

Coordinatore Dottorato
prof.ssa Barbara Mantovani

Relatore
prof.ssa Donata Luiselli

Esame finale anno 2015

ABSTRACT

Title: Evolutionary Epigenetics of Modern Human Populations
Running title: Human Evolutionary Epigenetics

Epigenetic variability is a new mechanism for the study of human microevolution, because it creates both phenotypic diversity within an individual and within population. This mechanism constitutes an important reservoir for adaptation in response to new stimuli and recent studies have demonstrated that selective pressures shape not only the genetic code but also DNA methylation profiles.

The aim of this thesis is the study of the role of DNA methylation changes in human adaptive processes, considering the Italian peninsula and macro-geographical areas. A whole-genome analysis of DNA methylation profile across the Italian penisula identified some genes whose methylation levels differ between individuals of different Italian districts (South, Centre and North of Italy). These genes are involved in nitrogen compound metabolism and genes involved in pathogens response. Considering individuals with different macro-geographical origins (individuals of Asians, European and African ancestry) more significant DMRs (differentially methylated regions) were identified and are located in genes involved in glucoronidation, in immune response as well as in cell comunication processes. A "profile" of each ancestry (African, Asian and European) was described. Moreover a deepen analysis of three candidate genes (KRTCAP3, MAD1L and BRSK2) in a cohort of individuals of different countries (Morocco, Nigeria, China and Philippines) living in Bologna, was performed in order to explore genetic and epigenetic diversity.

Moreover this thesis have paved the way for the application of DNA methylation for the study of hystorical remains and in particular for the age-estimation of individuals starting from biological samples (such as teeth or blood). Noteworthy, a mathematical model that considered methylation values of DNA extracted from cementum and pulp of living individuals can estimate chronological age with high accuracy (median absolute difference between age estimated from DNA methylation and chronological age was 1.2 years).

CONTENTS

FIGURES LEGENDS

TABLES LEGENDS

ABBREVIATIONS

DMRs differentially methylated regions

BOP block of probes (refer to 450k array analysis, for detailed see Appendix A)

Amplicon fragment of DNA amplified using PCR

GO gene ontology

CI Centre of Italy

SI South of Italy

NI North of Italy

AFR individuals of African ancestry

EUR individuals of European ancestry

EAS individuals of East Asian ancestry

LCL lymphoblastoid cell line

PBMC peripheral blood mononuclear cell

450k array Infinium HumanMethylation450 BeadChip

DAPC discriminant analysis of principal component

GEO Gene Expression Omnibus (repository)

Chapter 1

EPIGENETICS AND DNA METHYLATION

1.1. EPIGENETICS AND DNA METHYLATION: OVERVIEW AND DEFINITIONS

We are in the middle of an epidemic of epigenetics' related studies In the database of ISI Web of Knowledge, more than 5000 articles published in 2011 contain epigenetic(s) in the title (Figure 1). The term *epigenetics* was originally introduced in 1942 by a British developmental biologist, palaeontologist, geneticist and philosopher, named Conrad Hal Waddington. His writingS were recently published in the article reprinted in this issue (Waddington 2012). There Waddington proposed the term "epigenetics" for studies of the "causal mechanisms" by which "genes of the genotype bring about phenotypic effects".

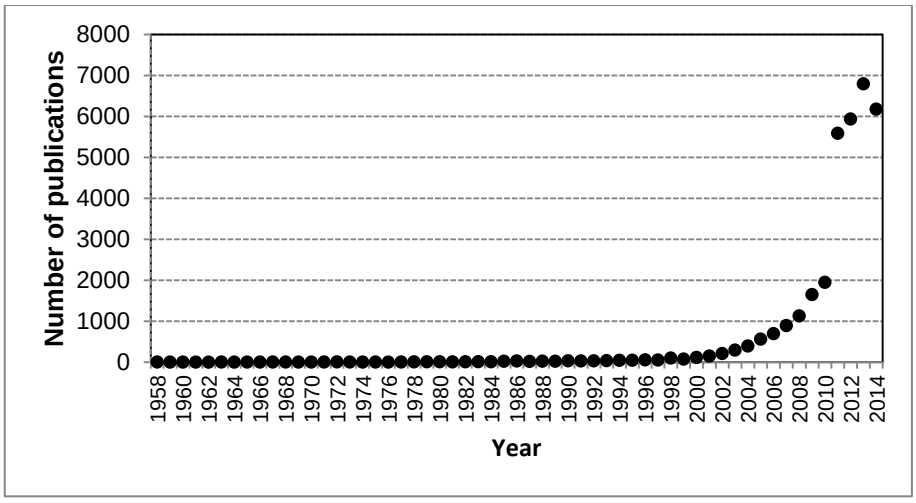


Figure 1. Number of articles with epigenetic or epigenetics in their title. This number is calculated from the ISI Web of Knowledge using the search for term "epigenetic*" in the article titles.

Today the term *epigenetics* refers to modifications of DNA and chromatin (heterochromatin and euchromatin) that affect gene expression without altering the sequences of the DNA itself. These changes are stable, but at the same time can be modulated by many factors such as physiological or pathological circumstances, as well as by the environment itself (Holliday 2006). Among the broad number of epigenetic modifications, in this thesis we will focus on DNA methylation that has been extensively investigated in the last years and represents the most promising (but still unexplored) mechanism for the study of human evolution.

DNA methylation is a molecular mechanism that entails the covalent addition of a methyl group in the fifth carbon of the pyrimidine ring of a cyotsine (m5C) that, in plant and animal genomes, is usually part of a CpG dinucleotide. DNA methylation affects the architecture of genomic DNA within nucleus and it is involved in fundamental cellular processes such as regulation of transposons, transcriptional silencing, genomic imprinting and X chromosome inactivation (Yoder, Walsh, and Bestor 1997; P. A. Jones and Takai 2001; Henderson and Jacobsen 2007).

1.2. DISTRIBUTION OF MCPG AND CPG DINUCLEOTIDES IN GENOMIC DNA

In the human genome, the density of these CpGs is non-random and they also appear to be under-represented with respect to what expected. This happened during evolution due to the presence of spontaneous deaminase activity in nuclei (for review see Peter A. Jones and Baylin 2002). This enzymatic reaction transforms methylated cytosine into thymine, while the unmethylated cytosine is transformed into uracil. Subsequent control and repair mechanisms recognize uracil as an extraneous base on DNA and substitute it, while this substitution does not happen where thymine is involved as it is a common base on DNA. As described in a recent review (Giuliani et al. 2015) this high rate of mutation of

m5C to thymine likely led to a reduction of CpG dinucleotides in mammalian genomes over the course of evolution (Razin and Riggs 1980). Regions with a higher CpGs content than the genomic average (*i.e.* "CpG islands") are supposed to be approximately 29,000 throughout the human genome, although this estimate could vary according to CpG island definition, and 60% of them co-localize with gene promoters. The DNA methylation status of these CpG islands is subjected to dynamic changes during development and cell differentiation and it can affect gene expression levels (Lander et al. 2001; Antequera 2003). In particular, DNA methylation could have different effects according to the position of the involved CpGs. For instance, methylation of CpG islands at transcription start sites (TSSs) is usually associated with inactivation of gene expression by physical interfering with transcription factors or, indirectly, by recruiting chromatin modification enzymes. On the contrary, methylation of CpGs located in gene bodies might even stimulate transcription and influence splicing mechanisms. Many methylated CpGs are also found in repeated sequences, most of which are derived from transposable elements, suggesting that they could play a fundamental role in genome stability, acting as suppressors of transposable elements expression (Peter A Jones 2012). Vertebrate promoters can be divided in two classes - low CpGs (LCG) and high CpGs (HCG) - according to their CpGs content. Most of LCG are methylated and are associated to tissue-specific genes, while HCG are usually hypo-methylated and are related to broadly expressed genes. Some authors suggested that one possible evolutionary reason for such bimodal distribution is that broadly expressed genes have selectively avoided DNA methylation because their consequent inactivation would have been deleterious (Elango and Yi 2008). DNA methylation can occur in different genomic regions and DNA methylation takes place not only in CpG islands, in fact recent studies (Doi et al. 2009; Irizarry et al. 2009) have highlighted the importance of methylation changes in the shores, *i.e.* the regions

nearby the CpGs island. It is known that DNA methylation of these regions is closely associated with transcriptional inactivation.
The papers of Doi et al. and Irizarry et al. anticipate the study of shores and they have demonstrated the importance of these genomic regions. In these studies they defined "shores" all the regions outside the island up to 2000 bp. In particular, the study by Doi et al. demonstrated that the DMRs (differentially methylated regions) that distinguish different types of tissue and tumor tissue from healthy tissue are 13 times more frequently in the shores of the island, despite these regions have a lower density of CpGs.

Chapter2

EVOLUTIONARY RELEVANCE OF DNA METHYLATION

Several theories have been proposed to explain the role that epigenetic variations might have played in human evolutionary history. The fact that DNA methylation constitutes a source of phenotypic variability, together with recent evidences showing that some epigenetic changes can escape transgenerational erasure mechanisms, suggests that these modifications have the potential to influence evolutionary processes.

2.1. PHYLOGENETIC CONSERVATION OF DNA METHYLATION PROCESSES

Comparative studies of DNA methylation suggested the conservation of this mechanism across different evolutionary timescales (for review see Mendizabal et al. 2014). Several studies have investigated highly divergent animal species that span over a billion years of evolution (honey bee, skilworm, sea anemone and sea squirt) and observed a strong conservation of DNA methylation patterns across this species (over 75% of genes are consistently methylated in all species) and that DNA methylation in these species is restricted to gene body. However the data generated in these studies across species usually regarded DNA extracted from whole animal bodies, without any cell sorting and this constitutes a limit that nowadays is still unsolved (tissues and cell specifity of this mechanism will be described in Chapter 3) (Zemach et al. 2010; Feng et al. 2010; Sarda et al. 2012). Recently many studies raised questions about epigenetic divergence by considering humans and closely related primates to investigate molecular mechanisms at the basis of the evolution of the human lineage. A study of Pai and colleagues (2011) considered DNA methylation patterns in livers, hearts, and kidneys from multiple humans and chimpanzees and they argued that some gene expression differences between species might be explained in part by

corresponding differences in methylation levels. Moreover they suggested that inter-tissue methylation patterns are often conserved between humans and chimpanzees, a phenomenon that is also reported by Irizarry and colleagues (2009) that described conserved tissue specific patterns of DNA methylation between humans and mice (diverged around 80 Ma) .

Interestingly this and other studies (Pai et al. 2011; Molaro et al. 2011; Martin et al. 2011; Hodges et al. 2011; Zeng et al. 2012; Wang et al. 2012; Hernando-Herraez et al. 2013) identified an enrichment of DMR in genes associated with neurological functions in the human lineage in different tissues, such as sperm, peripheral blood and prefrontal cortex.

2.2. RECENT THEORIES REGARDING THE ROLE OF DNA METHYLATION IN HUMAN EVOLUTION

Some authors postulated some theories regarding the evolutionary mechanisms that impact on epigenetic variability (reviewed and described in Giuliani et al. 2015). We will briefly describe them as follows:

- Shea et al. (Shea, Pen, and Uller 2011) listed two different epigenetic mechanisms that could have shaped human evolution by producing selection-based and detection-based effects, respectively. The first mechanism is similar to that through which natural selection acts on genes, that is, a specific epigenetic pattern is maintained across generations and is inherited by the offspring from the parents. The second mechanism is based on an "environment detection" process, for which parents experience a particular environment and transmit an epigenetic pattern apt to cope with it. They also considered a third channel of information linked to somatic cells identity and related to the fact that cells belonging to a certain tissue carry information about tissue

differentiation. The authors suggested that in multicellular organisms information conflicts could arise, since a single locus is not able to carry different information (*i.e.* about cell differentiation and about selection- or detection-based effects of specific epigenetic profiles). Obviously the only way to avoid this conflict is to consider the possibility that each locus carries only one type of information, but that different loci could carry different type of information. For example, regarding DNA methylation mechanism, some specific CpG sites are involved in somatic differentiation and they could not be involved also in mechanisms producing selection-based effects at the same time.

- Gluckman et al. (Gluckman, Hanson, and Spencer 2005) described a set of theories, known as "predictive adaptive responses", based on the hypothesis that conditions experienced *in utero* may confer advantages in a later phase of life. It is important to note that it is not known to what extent epigenetic variants are transmitted *per se* or arise via *de novo* mutations at each generation (Rando and Verstrepen 2007).

- Jablonka and Lamb (Jablonka and Lamb 2005) argued that since DNA methylation can alter gene expression, and consequently phenotypes, epialleles can be subjected to natural selection, as through alleles at genetic loci.

- a recent study (Klironomos, Berg, and Collins 2013) hypothesizes that variants beneficial in a population could be of genetic and epigenetic origin and that both of them contribute to increased fitness. However, changes in DNA methylation could be lost more rapidly and need a more strenght of selective pressures to be fixed in the population. The authors claimed that under very strong selective pressures the use of epigenetic

mechanisms to encode new and potential adaptive phenotypes could be advantageous with respect to the use of genetic ones. The authors suggested that once a phenotype has been produced by DNA methylation changes, the related genetic variation is not under selection and viceversa and thus the strength of selective forces on genetic variants could be reduced by epigenetic variations . In this way, they hypothesized that DNA methylation process can constitute a strategy for populations to "buy time" in order to face the new environmental conditions, until a new potentially adaptive geneti mutation arises. In Figure 2 is reported a summary for the methylation hypothesis (Giuliani et al. 2015) in which epigenetic changes could be important to buy time until genetic adaptation arise or when the intensity of the stimulus is not constant over time.

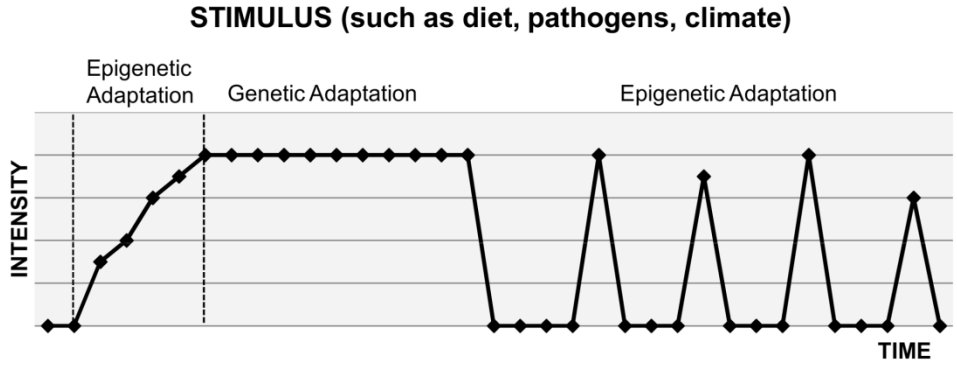


Figure 2. Theory of Klironomos et al to explain the role of epigenetic and genetic adaptation. [Figure published in Annals of Human Biology, Giuliani et al. 2015]

- Feinberg and Irizarry (2009) proposed a fascinating idea for the role that epigenetic mechanisms could have played in human evolution. They suggest that some genetic variants may exert an effect on the variability

of a certain trait, without altering the mean phenotype. According to authors' hypothesis, propensity to stochastically vary is mediated epigenetically and populations carrying genetic variants associated to this higher propensity to variability might have an advantage (or a disadvantage) in changing environments.

In conclusion DNA methylation could contributes not only to create inter individual variability but also to create inter population variability and could constitutes a reversible way to adapt to new stressors typical of a certain environment. In this sense the differences observed among populations could be an important tool in order to deepen the influence of the environment on population. These differences should be important to describe differences that naturally occurs and that could potentially constitute a new strategy for adaptation but also to identify patterns that could contribute to the variability observed in response to drugs or external/internal stimuli. In the Figure 3 (published in Giuliani et al. 2015) different strategies of adaptation to environmental stimuli are reported, according to the time in which they arise. Epigenetics is a molecular mechanism that play a role in human adptations and this reversible mechanism seems to play a role in many situations, flanking the classical way to adapt which is by genetic and cultural changes.

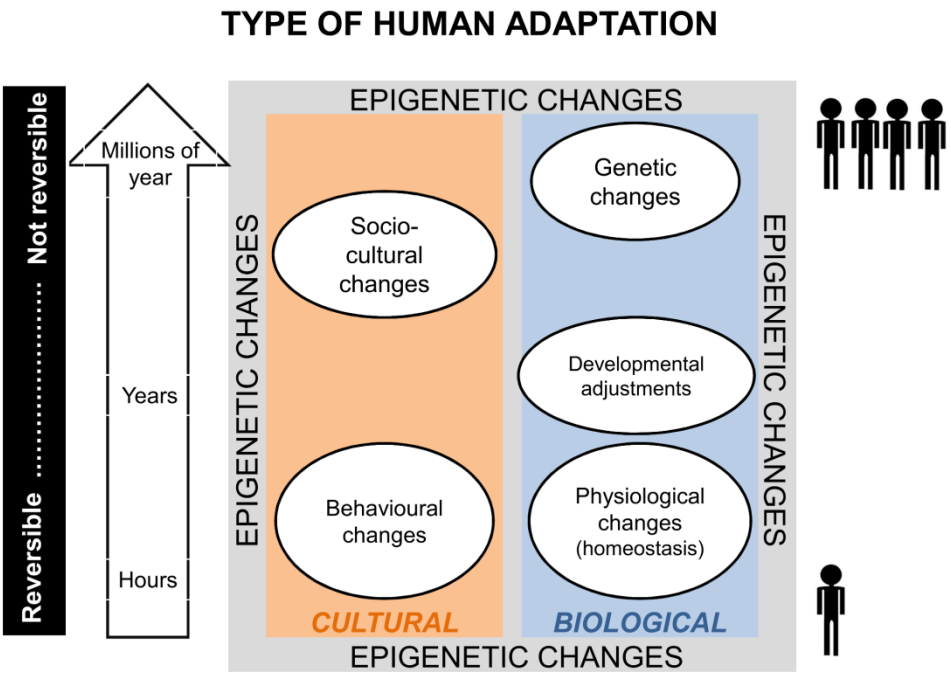


Figure 3. Type of human adaptations. Epigenetics is a reversible way to adapt to new stressors that can play a role in many situations; it is a further level of adaptation to consider in addition to genetic and cultural changes.

Chapter 3

DNA METHYLATION VARIABILITY IN HUMANS

DNA methylation is a dynamic and extremely variable process, this variability can be considered as a source of phenotypic plasticity fundamental for the mechanisms of adaptation.
The variability can be observed at different levels and may involve different pools of CpGs, located in crucial parts of the genomes (promoters, genes body, trascription factor binding sites etc.). As previously mentioned the DNA methylation of certain regions varies depending on the **tissue** and cell type, several studies identified "tissue differently methylated sites" (tDMR) for many tissues, depicting a sort of signature for each tissue (Lister et al. 2009). Differentially methylated loci were also observed when different **populations** living in different countries were considered. Moreover also the **diet** constitutes a factor that can impact on the methylation profiles, both during lifetime (Bacalini et al. 2014) and in evolutionary time. It is also noteworthy that folates play a fundamental role since they constitute a key component in the biochemical pathways that regulated DNA methylation. Also over a lifetime, as a result of the changes in physiological and pathological stimuli, DNA methylation profiles undergo continuous remodelling. Various authors described the phenomenon of the "epigenetic drift" that occurs during **normal aging** (Teschendorff, West, and Beck 2013). Studies of identical twins are an exceptional model and allowed us to describe "pure" epigenetic variability, that is also the one that take place independently of the genetic background during the life of the individual. In the following sections we will briefly discuss physiological processes that could impact on DNA methylation patterns.

3.1. TISSUE SPECIFIC DNA METHYLATION MODIFICATIONS

DNA methylation is an important regulator of tissue differentiation and this means that the levels of DNA methylation of certain CpG sites are closely dependent on the tissue in analysis. This constitutes a level of variability that should be taken into account during the experimental designs. Recent studies showed that tDMR (= tissue differentially methylated regions) are sparse across the genome (Slieker et al. 2013; Lokk et al. 2014). The study by Slieker et al. (2013) showed that tissue-specific DNA methylation was most evident in regions outside CpG islands or CpG island flanking regions. The authors suggested that the role of tDMRs located outside CpG islands might involve the regulation of splicing and alternative transcription. Moreover, the results of Lokk et al (2014) showed a clear correlation between DNA methylation in the gene promoter areas and the gene expression, while this study did not find a positive correlation between gene bodies methylation and mRNA expression as other authors suggested. They also suggested that the percentage of variance explained by tissues is higher than the one explained by individuals.

These characteristics are the basis of one of the major problems in these types of studies: the cellular heterogeneity (Jaffe and Irizarry 2014). In fact, every cell type presumably has its epigenetic profile and the measure of methylation patterns from DNA extracted from a pool of cells inevitably leads to a measure of an average methylation, calculated considering DNA from all the different cells. The "ideal" experiment should include a first step of cell sorting (for example with FACS - Fluorescence Activated Cell Sorting) and then the analysis of methylation. However this is not a cost effective procedure to be introduced in a everyday laboratory workflow and its major limitation is that it could be performed only on fresh tissues/cells. To solve the heterogeneity issue, some authors who worked on blood (the most easy tissue to obtain) discussed the possibility to normalize the data on the basis of cell counts. Other authors,

instead, working on genome wide data, described valid statistical models, based on DNA methylation signature, that can estimate the proportion of cell types in whole blood (NK, T lymphocytes etc..) (Houseman et al. 2012). This is a valid approach that has been utilized also by Guintivano et al. (2013) to estimate neuron and non-neuron components in brain samples.

3.2. AGE RELATED DNA METHYLATION MODIFICATIONS

Although the epigenetic changes in cancer and other diseases have been intensively studied, here we consider only the changes and the variability in the levels of DNA methylation that physiologically accumulates. This can be of great interest both from an adaptive point of view, since humans during their life must continually adapt to stresses that can come from outside and inside of individual body, both from a practical point of view, because age-related epigenetic modifications constitute a promising tool for estimating the age from ancient and historical samples and from forensic findings.

DNA hypomethylation during aging is the most observed phenomenon not only in humans but also in many species such as salmon (Berdyshev et al. 1967) and rats (Vanyushin et al. 1973). Some regions and CGs are hypermethylated during aging as recent literature suggests, in particular at CpG islands (Rodríguez-Rodero et al. 2010; Bell et al. 2012; Christensen et al. 2009). Despite the tissue specificity that could be critic in this type of analysis (Thompson et al. 2010), recent studies identified age-dependent CpG signatures that do not depend from sex, pathology or tissue type. Recently some studies investigated association between age and DNA methylation in saliva and in blood (Bocklandt et al. 2011; Hannum et al. 2013)

Some authors identified CpGs that reveal linear DNA methylation changes during aging and it is likely that they constitutes good candidates for age

prediction (Hannum et al. 2013; Bocklandt et al. 2011; Koch and Wagner 2011)The most detailed study was published in 2013 by Steve Horvath (Horvath 2013), who used a collection of online DNA methylation data sets for defining an age predictor based on DNA methylation data.

3.2.1. DNA-based methods for the age estimation on historical samples

Currently there are many methods to estimate the age of an individual from ancient remains, based on both morphological and molecular data. However it is known that these methods give only a rough estimation of the age of the individual. Moreover there are also some methodological matters, since DNA from historical and ancient samples is a low integrity molecule and moreover few tissues can be collected from post mortem remains (mainly bones and teeth). The issue of what constitutes the standard reference for the estimation of skeletal age-at-death, especially for adults, is the subject of remarkable debate in both physical and forensic anthropology (Franklin 2010). Regarding morphological features it is well established that their variability progressively increases from birth to old age, so the older the sample is, the less accurate the age estimation is (AlQahtani, Hector, and Liversidge 2014; Brooks and Suchey 1990). DNA-based methods for age estimation, such as 4977 bp deletion of mtDNA and telomeres lenght, are usually not suitable if the time period between death and sampling is too long and the age is assigned to a wide age-range, with very limited accuracy (Meissner and Ritz-Timme 2010). The amino acid racemization in dentine represents the best method (±8.7 years with 95% confidence) but it can be used only if postmortem interval is not more than 20 years (Franklin 2010; Meissner and Ritz-Timme 2010).

3.2.2. Age prediction from DNA methylation on forensics samples

Recently DNA methylation has been applied in forensic anthropology. In fact DNA methylation is more stable than RNA molecules and this is a good characteristic for forensic samples. DNA methylation applications are extensively described in a recent review (Vidaki, Daniel, and Court 2013) and it was firstly used in the early 90s for sex typing using the methylation status of a macro-satellite. This approach could be use as complementary analysis in case of sex-reversed individuals (Naito et al. 1993). DNA methylation is also used for the authentication of DNA samples, as it is known that DNA synthesized in vitro is generally unmethylated. DNA methylation profile in this case can be used to validate the material found in crime scene. DNA methylation in forensics is also used to distinguish monozygotic (MZ) twins, that are identical in term of genetic background. At this purpose, it is of particularly interest a recent study that suggests that MZ twins can be distinguished using the methylation status of 377 CpGs sites (Li et al. 2011). Moreover another field of applications is the identification of body fluids, recent studies showed that body fluid can be identified using DNA methylation status of few CpGs. In fact the best biomarker for this purpose is a CpG site whose methylation level is extremely high in the tissue of interests and extremely low in all others tissues. Nowadays with this method it is possible to distinguish with high confidence blood, saliva, semen, and vaginal secretions, while for others tissues genome-wide experiments should be yet validated. Age estimation is a very promising scenario in this context that could help the investigation. At this purpose the model built by Horvath seems to be the more indicated, in fact for this purpose it is important to consider more than one CpGs site and more than one region to increase the accuracy of the age prediction. In fact it is likely that the higher it is the number of CpGs that are considered, the more reliable could be the age estimation based on DNA levels.

3.3. DNA METHYLATION, GEOGRAPHY AND POPULATION HISTORY

Phenotypic plasticity is one of the main conceptual bases of population epigenetic studies, since epigenetic information allows a single genotype to assume multiple epigenotypes, thus enabling the expression of different phenotypes. This is particularly emphasized in plants, because they are obligated to cope with environmental conditions, since they do not have the possibility to migrate away looking for a more apt environment. Different studies, summarized in (Giuliani et al. 2015) (Appendix A) highlight a link between ethnic group identity and DNA methylation variability. Here we will described only the more recent and broad population study published in 2013 by Heyn and colleagues (Heyn et al. 2013) that constitutes the source of a free epigenome wide database to work with. This study analyzed genome-wide methylation (Illumina 450k BeadChip) in three populations with different ethnicity: Americans of African ancestry, Americans of Asian ancestry and Americans of European origins. The study identified 439 CpG sites (pop-CpGs) that distinguish the examined populations, pointing out that this pattern is associated with CpGs that regulate gene expression and that many pop-CpGs are located within or nearby transcription binding sites for NFKB, a transcription factor involved in the immune defence. Accordingly, the authors proposed that local selective pressures, such as a specific pathogens landscape, could shape epigenetic variation in genes involved in immune or xenobiotic responses.

Moreover, this study considered the link between genetic and epigenetic variability and showed that two third of pop-CpGs were associated with the individuals' genetic background, suggesting that part of population-specific DNA methylation variation could be linked to genetic differences between populations and that genetic background drives and modulates DNA methylation levels (Heyn et al. 2013).

DNA methylation differences were also recently investigated at Italian level (Campanella et al. 2014),. This is the first study that, using genome-wide DNA methylation profiles of more than 1000 individuals, looked for DNA methylation changes associated with voluntary south-to-north migration that occurred within Italy. The study identified some diferentially methylated CpG comparing south-to-north migrants with people who lived in north-western Italy. They hypothesized that these changes could indicate an adaptive response to a new environment.

Chapter 4

AIM OF THE THESIS

The study of epigenetic variability is a new and important mechanism for the study of microevolution, because it creates both phenotypic diversity within an individual and within population without genetic variation (Richards 2008). These mechanisms constitute an important reservoir for rapid adaptation in response to internal and external stimuli. The aim of this thesis is the study of how DNA methylation changes may influence human adaptive processes. Moreover this thesis paves the way for the application of DNA methylation for the study of hystorical remains.

This thesis can be theoretically divided in 2 areas of research, linked by the final goal that is the investigation of the role that epigenetic mechanisms play in human biodiversity. (1) The study of the DNA methylation variability across human populations, considering Italian and macro-geographical levels (2) The study of the DNA methylation modifications that naturally occur over time to identify potential biomarkers for age-estimation.

4.1. DNA METHYLATION VARIABILITY ACROSS HUMAN POPULATIONS: ITALIAN AND WORLDWIDE VARIABILITY

4.1.1. Aim

With this thesis we aim at elucidating the role of DNA methylation variability in different populations, in particular we looked for epigenetic structure, trying to elucidate the evolutionary relevance of this mechanism in different populations. We investigated epigenetic variability at macro-geographical and Italian level

4.1.2. Cohorts and Methods

#1 *Italian variability*, we analyzed with a gene candidate approach (IGF2 and H19 genes) DNA methylation profiles of 376 individuals homogeneous for gender (males) and age (age range 30-51), but differing for geographical origin and place of living, *i.e.* four Italian regions (Northern, Central and Southern Italy, and Sardinia). Moreover genome wide analysis (Illumina Infinium HumanMethylation450 BeadChip) of 47 individuals according to geographical origin (North, Centre and South) was performed to identify signature of epigenetic structure in Italy.

#2 *Macro-geographical variability*, starting from the row data available in the public database GEO we used an original pipeline for the analysis of Infinium 450k data of 288 individuals of different ancestry (96 Americans of European ancestry, 96 African-Americans and 96 Han Chinese Americans). This pipeline considers the concomitant variation of a group of adjacent CpG probes (region-centric approach), rather than the methylation level of a single CpG probe. Indeed, changes in DNA methylation usually involve a group of adjacent CpG sites, whose methylation levels vary in a coordinated way and in the same direction, thus potentially affecting chromatin structure. Based on the results from this study, we focused on DNA methylation of three genes (*BRSK2*, *MAD1L1* and KRTCAP2) in blood of individuals with different ancestry and different country of birth (Nigeria, Morocco, Philippines, China and Italy) but who lived in Bologna. Moreover, we performed the analysis of genetic variability in the surrounding regions of the selected candidate genes

4.2. RELATIONSHIP BETWEEN DNA METHYLATION VARIABILITY AND AGE

4.2.1. Aim

DNA methylation changes over time constitute a interesting field of research in anthropology for two main reasons: (1) it allow individuals adaptations during life to be studied, in fact DNA methylation variability, that accumulates during life, can be considered a potential reservoir to cope with internal and external stressors; (2) sites whose DNA methylation level varies according to chronological age, can be used as markers of age in anthropology. The purpose of this thesis is mainly to make use of this knowledge to put the basis for age estimation from remains, starting from modern teeth.

4.2.2. Cohorts and Methods

#1 A genome wide DNA methylation analysis (Illumina Infinium HumanMethylation450 BeadChip) on whole blood of a small cohort of 64 subjects of different ages showed 3 region the CpG islands of ELOVL2, FHL2, and PENK genes, whose methylation level strongly correlates with age. A larger cohort of 501 subjects from 9 to 99 years, including 7 cord blood samples was considered to confirm the DNA methylation correlation between ELOVL2, FHL2, and PENK DNA methylation and age.

#2 Since identifying DNA-based methods to estimate the age of a person is crucial in several fields, such as physical anthropology, biodemography, forensics, evolutionary and ancient DNA studies, in this thesis we analyzed DNA methylation levels of few CpGs located in three genes (*ELOVL2*, *FHL2* and *PENK*) of cementum, dentin and pulp of 21 modern teeth of individuals - from 17 to 77 yrs - and we built a model to predict age.

Chapter 5

MATERIALS AND METHODS

5.1. COHORTS ANALYZED

5.1.1 Italian variability

376 individuals homogeneous for gender (males) and age (age range30-51) were considered in the candidate gene analysis. Individuls were recruited following an accurate sampling strategy on a preliminary surname-based study (Boattini et al. 2012) in order to consider the most representative and informative Italian sample, with an ancient regional ancestry. These individuals were analyzed at candidatte gene levels following the procedures described in the section below.

Moreover PBMC (peripheral blood mononuclear cells) of 47 Italian individuals were analyzed at genome wide level (Illumina 450k array). This cohort is constituted by 17 individuals (6 malea sn 11 females, mean age=70.8±5.5 yr) living in Milan and born in North Italy (NI), 16 individuals (7 males and 8 females, mean age=70.8±7.7 yr) living in Bologna and born in Emilia Romagna region (Centre of Italy, CI) and 14 individuals (9 males and 5 females, mean age=67.6±8.5 yr) living and born in Calabria (South of Italy, SI).

5.1.2. Macro-geographical variability

For the world wide analysis we considered data already published and stored in the public database GEO, Gene Expression Omnibus under accession number GSE36369. DNA methylation data on lymphoblastoid cell lines (LCLs) of 96 Americans of European ancestry (EUR), 96 Americans of African ancestry (AFR) and 96 Han Chinese Americans (EAS) were considered. LCLs considered were included in the Human Variation panel (sample sets HD100AA, HD100CAU, HD100CHI, Coriell Cell Repositories). Mean age was 37.3 ± 16.2

yr for EUR, 29.4 ± 9.9 for AFR and 36.2±15.7 for EAS. Moreover DNA from whole blood of individuals of different ancestry was used to replicate the results. The methylation data regarding blood samples of 10 Americans of European ancestry, 10 Asian American were described in the study of Heyn and colleagues (2013) (GSE36369) while DNA methylation data of 10 Americans of African ancestry have been previously published (Alisch et al. 2012) (GSE36064).
Whole blood of 30 individuals from Morocco (N=16), Nigeria (N=14), Philippines (N=13) and China (N=17) were collected in Bologna before 1997 and stores at -20°C.

5.1.3. Age-dependent variability

32 mother-offspring couples were analyzed at genome-wide level (Illumina 450k array) to identify regions of variability with age. The age range was 42–83 and 9–52 years for mothers and offspring, respectively. The top 5 significant loci mapped within the CpG islands of *ELOVL2*, *FHL2* and *PENK* genes and were analyzed in a large cohort of 501 subject from 9 to 99 years incluing 7 cord blood samples.

Regarding the teeth collection, 21 teeth of individual between 17 and 77 years were used to extract DNA from 3 tissues, cementum, dentin and pulp and DNA was also extracted from the whole teeth of 12 teeth of individuals between 24 and 77 years. The soft tissues were removed with a scalpel and the teeth were washed with distilled water to eliminate blood residues. Cementum was collected scraping the root with a scalpel. The crown was removed using a diamond-cutting disc; pulp and dentine were recovered from the internal cavity of the tooth using a dental burr. DNA was also extracted from the whole tooth of 12 of them by crushing in a mortar.

5.2. EXPERIMENTAL PROCEDURES

5.2.1. DNA extraction and bisulphite treatment of DNA.

Genomic DNA was extracted from 150 μl of whole peripheral blood using the QIAamp 96 DNA Blood Kit (QIAGEN, Hilden, Germany). For Infinium HumanMethylation450 BeadChip, sodium bisulphite conversion of 600 ng of DNA was performed using the EZ DNA Methylation-Gold Kit (Zymo Research Corporation, Orange, CA). For Sequenom EpiTYPER assay, 1000 ng of DNA were bisulphite converted using the EZ-96 DNA Methylation Kit (Zymo Research Corporation, Orange, CA) with the following modifications respect to manufacture's protocol: bisulphite conversion was performed with thermal conditions that repeatedly varied between 55°C for 15 min and 95°C for 30s for 21 cycles; after desulphonation and cleaning steps, bisulphite-treated DNA was eluted in 100 μl of water.

5.2.2. Genome-wide DNA methylation analysis

Whole-genome methylation analysis of the DNA was performed using the Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA) (Bibikova et al. 2011), that allows to assess the methylation level of 485,764 cytosine positions over the entire genome. CpG probes are distributed in CpG islands, CpG shores and CpG shelves and in not-CpG rich intra- and inter-genic regions. Figure 4 modified here but edited by Dedeurwaerder et al (2011), summarized the locations and coverage of the probes included in the 450K array.

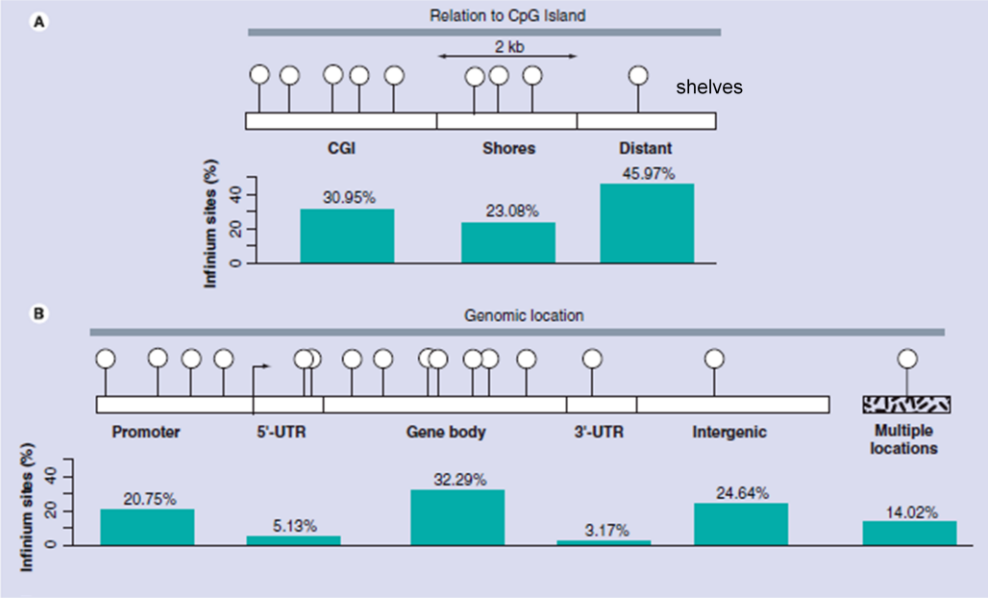


Figure 4. Overview of the coverage of the Infinium Methylation 450K array. Histograms show the percentage of cytosines covered by this technology according to their relation to CpG islands and to their genomic location. Original figure is edited by Dedeurwaerder et al.,2011.

As described in the official Illumina website, the Infinium Methylation Assay uses two different bead types to detect CpG methylation. One bead type (the U bead) matches the unmethylated CpG site while the M bead type matches the methylated site. The Figure 5, reportes in the Illumina official website, the unmethylated CpG target site matches with the U probe, enabling single-base extension and detection is reported. It has a single-base mismatch to the M probe, which inhibits extension. If the CpG locus of interest is methylated, the reverse situation occurs. Illumina 450k array was used to assess DNA methylation in Italian samples (Chapter 6.2.)

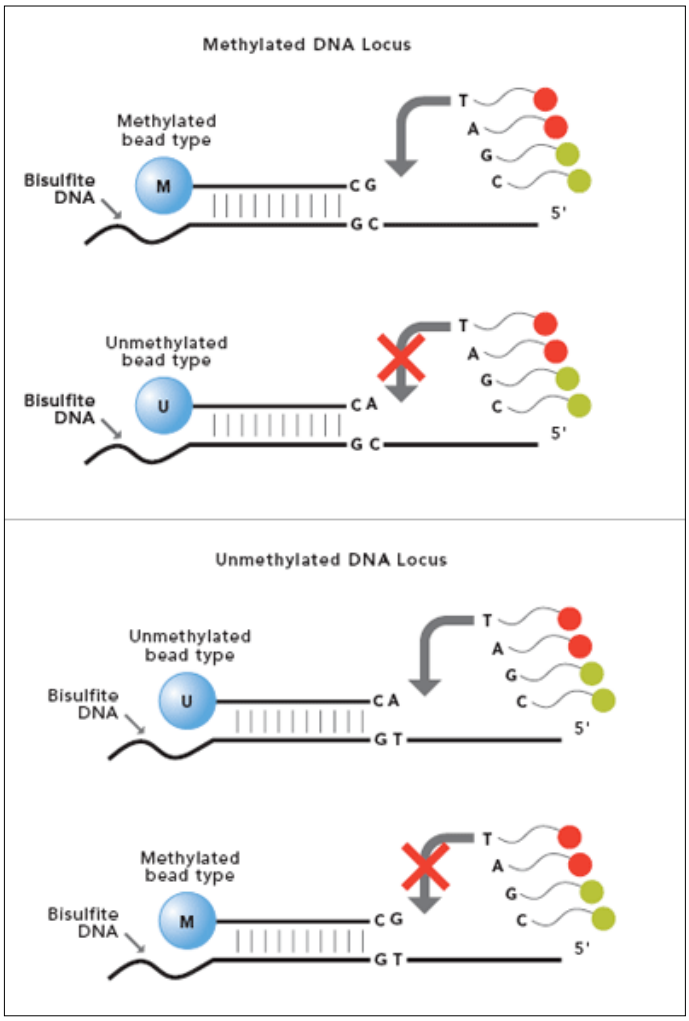


Figure 5. Infinium methylation 450k chemistry (figure of the Illumina official website)

5.2.3. Region-specific DNA methylation analysis

Quantitative analysis of methylation status of CpG sites in candidate genes was performed by the EpiTYPER assay (Sequenom, San Diego, CA), a MALDI-TOF mass spectrometry-based method. 10 ng of bisulphite-treated DNA were

PCR-amplified and then processed following manufacture's instructions. Bisulphite specific primers were reported in Table 1.

Region	Primer Forward	Primer Reverse
H19	AGGAAGAGAGGGGTTTGGGAGAGTTTGTGAGGT	CAGTAATACGACTCACTATAGGGAGAAGGCTATACCTACTACTCCCTACCTACCAAC
IGF2_AS	AGGAAGAGAGTGGATAGGAGATTGAGGAGAAA	CAGTAATACGACTCACTATAGGGAGAAGGCTAAACCCCAACAAAAACCACT
IGF2_island	AGGAAGAGAGTATAGGGGTGGTTTGTTAGGTTAGG	CAGTAATACGACTCACTATAGGGAGAAGGCTAAATCAAAAAAAACCCCAAAAAAAC
IGF2_shore	AGGAAGAGAGGAAGGGGTTGGTTAGTAGGTGTTTGT	CAGTAATACGACTCACTATAGGGAGAAGGCTCCTAAACCCCTTTCCCACTCTCTAA
ELOVL2	AGGAAGAGAGGTAAATTTGTAGGAATAGAGTTATTTTTTT	CAGTAATACGACTCACTATAGGGAGAAGGCTCCCCTCTCCCACAAAAACC
FHL2	AGGAAGAGAGTTTTTTATGGTTATTTGTGGTGTTT	CAGTAATACGACTCACTATAGGGAGAAGGCTCCCTTTATTTACCAAAACTCCTTTC
PENK	AGGAAGAGAGAGGGGTTTTATGATGAAAAGAATTT	CAGTAATACGACTCACTATAGGGAGAAGGCTAAAAAAATCCCAAAAATTTCCAAAC
BRSK2	AGGAAGAGAGTTTGGTATTTGGTGTTAAGTGGTTT	CAGTAATACGACTCACTATAGGGAGAAGGCTAAAAACTAATCTCCACTCTTCATAACA
KRTCAP3	AGGAAGAGAGGGTAAGGGTAGTTTTAGGGTAAGGA	CAGTAATACGACTCACTATAGGGAGAAGGCTAAACCTAAACCTTCTCAACAACC
MAD1L1	AGGAAGAGAGTGAAGATTTATTTTTGGAGTGGGTA	CAGTAATACGACTCACTATAGGGAGAAGGCTTAACACCAACCAAAACACACCTAA

Table 1 Primer sequence of the amplicon analyzed in this thesis using MALDI TOF MS (Sequenom EpiTyper)

5.3. STATISTICAL ANALYSIS

For gene candidate analysis an R package was used to test if bisulphite conversion reaction runs to completion (Thompson et al. 2009). CpG sites with missing values in more than the 20% of the samples were removed, as well as samples with missing values in more than the 20% of CpG sites.

Genome wide DNA methylation values were in part obtained by public database (GSE36369 and GSE36064). Color bias adjustement, background level adjustment and quantile normalization across arrays was performed using the lumi package (Du, Kibbe, and Lin 2008). The pipeline is described in detail in *Appendix A* in the paper published by Bacalini et al. 2015 entitled "*A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data*". Probes on chromosome X and Y, probes containing missing β-values and those with a SNP frequency >1% in the probe sequence were removed (as suggested in Heyn et al. 2013).

For Class A and Class B, BOPs methylation values were compared between groups using the MANOVA function from the R package *car*. BOPs containing one or 2 CpG probes were excluded from the analysis and MANOVA was applied on sliding windows of 3 consecutive CpGs within the same BOP. For each BOP, the lowest p-value among those calculated for the different sliding windows was retained. For Class C and Class D, the methylation values of single CpG site was compared between groups using the ANOVA function from the R package *car*. Benjamini-Hochberg False Discovery Rate correction was performed using the function *mt.rawp2adjp* from the R package *multtest*. Gene ontology annotation of selected DMRs was performed using gene ontology enrichment analysis and visualization tool http://cbl-gorilla.cs.technion.ac.il/ (Eden et al. 2009).

Genetic analysis (Chapter 6.4.4) was performed considering single nucleotide polymorphism (SNPs) located in the regions selected (*KRTCAP3*, *MAD1L1* and *BRSK2* genes) of 1511 individuals belonging to 15 human populations included in the 1000 Genomes project (phase 3). These variants were used to perform population structure analyses. The dataset was filtered using R to retain only biallelic SNPs, then using PLINK the dataset was filtered to prune SNPs according to their linkage disequilibrium (only loci in linkage equilibrium were retained). Loci in LD were filtered out by using a sliding windows approach (Alexander, Novembre, and Lange 2009). Discriminant analysis of principal component was performed using adegenet package to investigate genetic structure (Jombart and Ahmed 2011; Jombart, Devillard, and Balloux 2010). Patterns of differentiation among population was investigated computing Fst index. Fst was calculated using *PopGenReport package* and for differentiation patterns between YRI, CEU and CHB Fst index was calculated by an in-house Perl script on the basis of Weir and Cockerham formula (Cockerham and Weir 1986).

Chapter 6

RESULTS AND DISCUSSION - POPULATION EPIGENETICS

The purpose of this chapter is to address the DNA methylation patterns in different populations. We analyzed epigenetic diversity across Italy, considering individuals of North, Centre and South of Italy at gene candidate level (Section 6.1.) as well as at genome-wide level (Section 6.2.). Moreover we investigated DNA methylation variability across macro-geographic areas (Section 6.3) and in conclusion we deepened the analysis of genetic and epigenetic variability of three candidate genes (*KRTCAP3, MAD1L1* and *BRSK2*) selected on the basis of the comparisons between individuals of Asian, African and European origin (Section 6.4).

6.1. GENE-CANDIDATE APPROACH FOR THE STUDY OF ITALIAN DNA METHYLATION VARIABILITY

We wondered if DNA methylation of a single locus, such as IGF2/H19, could be explained by geographic origin and place of living of the individuals studied. Moreover we investigated the impact of environmental stimuli - different along the Italian peninsula - on methylation patterns of this locus.

To better understand the intricate relationship between spatial (geography/ancestry) and DNA methylation variability, we considered a cohort constituted of 376 individuals homogeneous for gender (males) and age (age range 30-51) recruited in the Genographic project (according to surname criteria (Boattini et al. 2012)), but differing for geographical origin and place of living, i.e four Italian regions (Northern, Central and Southern Italy, and Sardinia) (Pirazzini et al. 2012).

To this aim we measured DNA methylation of four target regions in the imprinted *IGF2/H19* locus, previously analyzed by other groups (Heijmans et al. 2007; Ollikainen et al. 2010). The regions analyzed were the following: *IGF2AS*, 3 kb from a CpG island, in exon 3 of *IGF2AS* transcript, within the *IGF2* DMR 0; *H19,* upstream the transcription start site of *H19* gene, partially overlapping a CpG island (*H19* DMR); *IGF2_island*, within a CpG island in the last exon of *IGF2* gene (DMR 2); *IGF2_shore*, in the shore upstream the island targeted by the *IGF2_island* amplicon. A graphic representation of the genomic localization of the region analyzed is reported in Figure 6. *IGF2AS* and *H19* have been previously analyzed by Heijmans *et al.* and Ollikainen *et al.*, while *IGF2_island* and *IGF2_shore* were analyzed here for the first time. We selected this locuse because it is one of the most studied in literature (Heijmans et al. 2007; Ollikainen et al. 2010). We also observed that in certain regions of this locus DNA methylation diversity accumulates during aging of monozygotic twins. Therefore we concluded that an effect of environment on DNA methylation of this locus could be possible (Pirazzini et al. 2012).

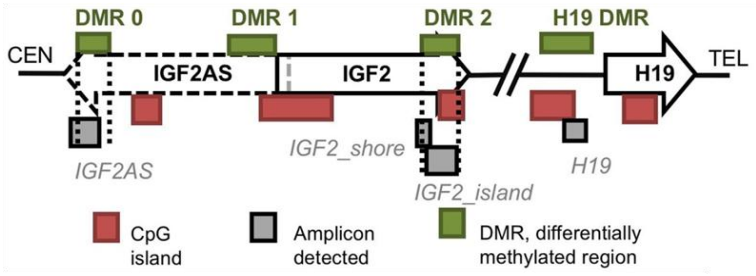


Figure 6. Genomic localization of amplicon analyzed in IGF2/H19 locus, considering, differentially methylated regions known by literature and CpGs islands.

The mean DNA methylation levels observed in *IGF2AS, IGF2_shore*, *IGF2_island* and *H19* were 49%, 60%, 43% and 36% respectively. Notably

IGF2AS and *H19* DNA methylation values were comparable to those previously reported in literature (Heijmans et al. 2007; Ollikainen et al. 2010).

Considerable inter-individual variation in DNA methylation values was observed within each CpG unit. We then evaluated if the observed variability in methylation levels of the *IGF2/H19* locus could be explained by ancestry and place of living of the four populations studied. Mean methylation levels of each CpG unit were calculated for the four geographical areas. The results showed that the methylation patterns were comparable throughout all the samples. No evident differences were detectable according to sample ancestry/geography, except for few CpGs (*i.e.* CpG 20 of *IGF2_island* amplicon). Principal component analysis (PCA) was used to capture the most interesting patterns of variation in each amplicon for the four geographical areas. The percentages of variance explained by principal component 1 (PC1) values are reported in Figure 7. No statistically significant difference was observed between PC1 values, as confirmed by ANOVA.

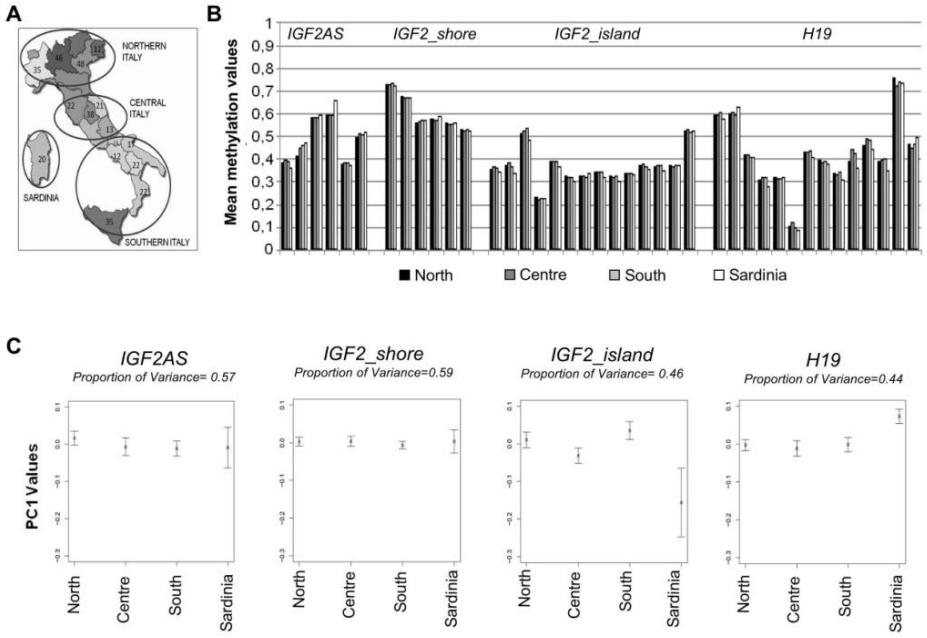


Figure 7. DNA methylation variability of IGF2/H19 locus in Italy. Distribution of the groups considered in the analysis (A). Average of DNA methylation levels according to geographic origin of the individual (B). Principal component analysis of the methylation levels of the 4 amplicons considered (C).

This data showed that methylation variability cannot be explained on the basis of the geographic provenience of subjects. This implies that the genetic structure of the Italian population does not influence the methylation pattern of the analyzed loci. and that there are no environmental cues specifically associated with Italian macro-areas able to generate clear population epigenetic signatures in the *IGF2/H19* locus.

6.2. GENOME-WIDE APPROACH FOR THE STUDY OF ITALIAN DNA METHYLATION VARIABILITY

We did not find any differences in DNA methylation of IGF2/H19 locus according to geographic origin, however since DNA methylation patterns could vary across the genome we investigated genome-wide DNA methylation profiles in individuals coming from North, Centre and South of the Italian peninsula.

To this purpose we measured DNA methylation levels in peripheral blood mononuclear cell (PBMCs) of 47 Italian individuals using Illumina 450k array (genome-wide approach). This cohort is constituted by 17 individuals (6 males and 11 females, mean age=70.8 ± 5.5) living in Milan and born in North of Italy (NI), 16 individuals (7 males and 8 females, mean age=70.8 ± 7.7) living in Bologna and born in Centre of Italy (Emilia Romagna region, CI) and 14 individuals (9 males and 5 females, mean age=67.6 ± 8.5) living and born in Calabria (South of Italy, SI).

6.2.1. Identification of differentially methylated regions between individuals from North, Centre and South of Italy

To identify DMRs between the three populations we used an original pipeline of analysis that we described recently (Bacalini et al. 2015) and that is here briefly summarized. The approach is based on the analysis of groups of adjacent CpGs, in fact DNA methylation patterns of groups of CpGs is more likely to be associated with functional variation than DNA methylation values of a single CpG site. We divided Infinium 450k probes in four Classes: *i)* Class A: probes in CpG islands and CpG islands-surrounding sequences (shores and shelves) that map in regions nearby genes; *ii)* Class B: probes in CpG islands and CpG islands-surrounding regions (shores and shelves) which do not map in genic regions; *iii)*

Class C: probes in genic regions which are not CpG rich; *iv)* Class D: probes in non-genic regions which are not CpG rich. With the term "blocks of probes" (BOPs) we refer to groups of Class A and Class B CpG probes localized in the same island/shore/shelf. Class C and Class D were analyzed by a single CpG approach because CpGs sites are are too distant to support an analysis by groups of adjacent probes.

Here only the results for Class A probes are reported. In Class A, we identified 6857 BOPs able to discriminate individuals from **Centre of Italy from Southern individuals** (q-value < 0.05). Gene ontology enrichment analysis was performed considering all genes with a p-value < 10^{-3} (197 unique BOPs) and showed that there is an enrichment of DMR involved in nitrogen compound transport (GO: 0071705; FDR q-value=ns). This category includes genes involved in the movement of nitrogen-containing compounds into, or out of, a cell or between cells by means of some biological structure or a pore (PLEKHA8, RPSA, SRSF9, SRSF11, RAB3A, OAZ3, SRSF6, TAPBP, RBM8A, SLC44A1, SMPD3, SLC35B4, UPF3A, UBF3A, UBE2Q1). It is interesting beacause nitrogen is contained in many drugs, antibiotics and organic compounds, for example ammonium nitrate has been used as fertilizer and a major source of reactive nitrogen is associated with the growing use of fossil fuels for energy, as described (UNEP, United Nations Environment Programme 2007). In Class A we identified 11208 BOPs that distinguish **Northern individuals from Southern individuals.** Gene ontology enrichment analysis was performed considering all genes with a q-value<10^{-3} (1289 BOPs) and showed that there is an enrichment of DMR involved in cellular nitrogen compound metabolic process (GO:0034461 FDR q-value=$2.7*10^{-2}$), nitrogen compound metabolic process (GO:0006807, FDR q-value=$1.52*10^{-2}$). In Class A, we identified 16161 BOPs able to discriminate **individuals from Centre of Italy from Nothern individuals** (q-value < 0.05). Gene ontology enrichment analysis was performed

considering all genes with a q-value<10^{-3} (4998 unique BOPs) and showed that there is an enrichment of DMR involved in macromolecule metabolic process (GO:0044260, FDR q-value = $2.64*10^{-22}$, GO:0043170 FDR, q-value = $7.71*10^{-20}$), primary metabolic pathway (GO:0044238, FDR q-value = $3.84*10^{-19}$) and gene expression (GO: 0010467, FDR q-value=$3.25*10^{-19}$).

6.2.2. Identification of epigenetic profile of individual of North, Centre and South of Italy

In order to understand the sites that most vary in Italy we considered all the genes that include a DMR in the pairwise comparisons for Class A probes previously reported (here we selected BOPs containing at least 2 adjacent CpG sites for which the DNA methylation difference between the 2 groups analyzed was higher than 0.15 (Du et al. 2010). We then performed a gene ontology enrichment analysis and we demonstrated that DNA methylation differences across Italy mainly involved genes of immunoglobulin mediated immune response (GO: 0002381, FDR q-value=$1.03*10^{-3}$), B cell mediated immunity (GO: 0019724, , FDR q-value=$9.3*10^{-3}$), regulation of interleukin-10 secretion and detection of bacterium. These results, visualized in Figure 8, by a TreeMap obtained by REVIGO analysis (Supek et al. 2011), suggested that microbial distribution across Italy is one of the major forces that shapes epigenetic profiles.

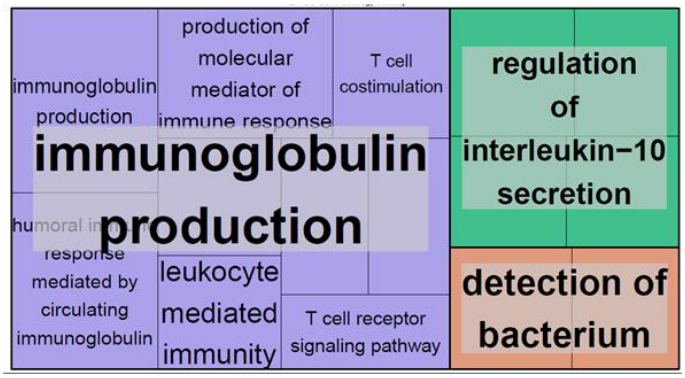
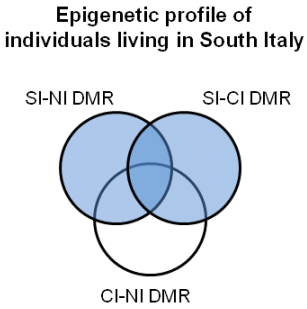


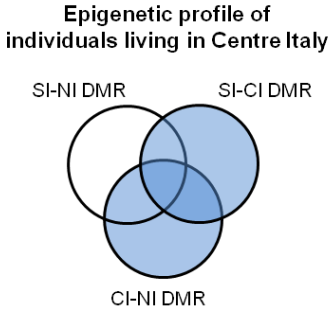
Figure 8. TreeMap obtained by a REVIGO analysis of all genes that contain a DMR considering all comparisons in Italy. Each rectangle is a GO terms that are joined into "superclusters" of related biological process, Size of the rectangle reflecte the p-values of the gene ontology enrichment analysis.

To provide an unambiguous epigenetic profile that include the most informative CpGs for each group of the Italian peninsula (North, Centre, South), from the list of Class A BOPs we selected a short list of DMRs whose DNA methylation status was remarkably different. To this aim, as previously described, we considered only the BOPs containing at least 2 adjacent CpG sites for which the DNA methylation difference between the 2 groups analyzed was higher than 0.15 (Du et al. 2010). Figure 9 reported 32 BOPs whose DNA methylation levels characterized individuals of Southern Italy. Figure 10 reported 30 BOPs whose DNA methylation is characteristic of individuals from Centre of Italy and Figure 11 reported 38 BOPs whose DNA methylation status characterized individual of Northern Italy. DMR are located in genes involved in immune response (such as *HLA-DRB5, HLA-DRB1, HLA-DQB1*), in olfactory receptors (*OR2L13* gene) and in metabolism and energy homeostasis (*LPGAT1*) (Traurig et al. 2013), as well as in a gene involved in cells communication with the microenvironment (*DDR1* gene) (Borza and Pozzi 2014; Phan et al. 2013) and in chronic pain and inflammation (*HCN2 gene*) (Emery, Young, and McNaughton 2012).



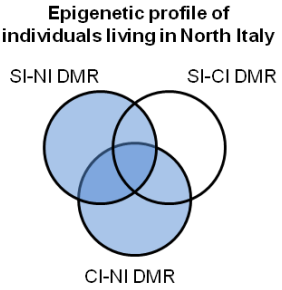
CHR	Genomic Position	Relation to CpG island	GeneName	Best CpG in the BOP
1	248100325-248100726	Island	OR2L13	cg03748376
1	113050813-113052301	Island	WNT2B	cg27181253
1	2266007-2266432	S_Shore	MORN1	cg05392448
1	212003431-212004280	Island	LPGAT1	cg13064658
2	27712554-27712842	S_Shore	IFT172	cg08787988
2	86332659-86333650	Island	POLR1A	cg23655155
2	241989164-241989461	Island	SNED1	cg23491743
5	127418594-127420512	Island	FLJ33630	cg16752652
6	32632158-32633027	Island	HLA-DQB1	cg05341252
6	169977218-169977894	S_Shore	WDR27	cg21696374
6	31148369-31148577	N_Shore	PSORS1C3	cg20073472
6	30852102-30852676	N_Shore	DDR1	cg19178509
6	32489742-32490128	Island	HLA-DRB5	cg26981746
6	32489742-32490128	N_Shore	HLA-DRB5	cg01341801
6	32551851-32552331	N_Shore	HLA-DRB1	cg23905789
6	32632158-32633027	N_Shore	HLA-DQB1	cg09312897
8	6691833-6693135	S_Shore	XKR5	cg27319216
9	130706912-130707273	N_Shore	FAM102A	cg01681863
10	75407413-75407706	S_Shelf	SYNPO2L	cg17411546
10	79396095-79398495	Island	KCNMA1	cg05754819
10	134562566-134563206	S_Shore	INPP5A	cg02201479
10	94526-95301	N_Shore	TUBB8	cg26679884
10	134918755-134918975	S_Shore	GPR123	cg23098789
11	74459858-74460558	N_Shore	RNF169	cg02673002
13	25743998-25746127	S_Shore	FAM123A	cg06779705
13	113424898-113425105	S_Shore	ATP11A	cg08893811
13	113540350-113540558	N_Shore	ATP11A	cg21463262
16	86549069-86550512	N_Shore	FOXF1	cg06834912
17	19617083-19617424	S_Shelf	SLC47A2	cg08102564
17	1492086-1492322	S_Shelf	SLC43A2	cg05291429
19	615691-623505	Island	HCN2	cg04875977
19	56612299-56612743	Island	ZNF787	cg26951705

Figure 9. Epigenetic profile of individuals living in South of Italy. Venn diagram represents with blue color the DMRs selected. In the table chromosome, genomic position (GRCh37/hg19 assembly), the CpG element, the gene name and the CpGs - located in the BOP - that better distinguish between groups, are reported.



CHR	Genomic Position	Relation to CpG island	GeneName	Best CpG in the BOP
1	150206971-150207842	S_Shore	ANP32E	cg21285133
1	113050813-113052301	Island	WNT2B	cg27181253
2	86332659-86333650	Island	POLR1A	cg23655155
2	241989164-241989461	Island	SNED1	cg23491743
3	127541270-127542048	S_Shore	MGLL	cg02116612
3	53289533-53290213	Island	TKT	cg01262337
4	81124468-81124845	Island	PRDM8	cg11983124
5	127418594-127420512	Island	FLJ33630	cg16752652
5	175298515-175300066	Island	CPLX2	cg01469864
6	32632158-32633027	Island	HLA-DQB1	cg05341252
6	31148369-31148577	N_Shore	PSORS1C3	cg20073472
6	30852102-30852676	N_Shore	DDR1	cg19178509
6	32632158-32633027	N_Shore	HLA-DQB1	cg09312897
7	1097405-1097953	Island	GPR146	cg21655171
9	130706912-130707273	N_Shore	FAM102A	cg01681863
10	133109193-133111052	Island	TCERG1L	cg18138206
10	134562566-134563206	S_Shore	INPP5A	cg02201479
10	134918755-134918975	S_Shore	GPR123	cg23098789
11	1247830-1248140	S_Shore	MUC5B	cg26676440
13	113424898-113425105	S_Shore	ATP11A	cg08893811
13	29913886-29914301	Island	MTUS2	cg13506281
16	86549069-86550512	N_Shore	FOXF1	cg06834912
17	80557446-80557686	Island	FOXK2	cg15292356
17	19617083-19617424	S_Shelf	SLC47A2	cg08102564
17	1492086-1492322	S_Shelf	SLC43A2	cg05291429
19	615691-623505	Island	HCN2	cg04875977
19	21265164-21265433	Island	ZNF714	cg06238316
19	3770566-3770942	S_Shore	RAX2	cg27288829
19	21265164-21265433	N_Shore	ZNF714	cg01483656
19	56612299-56612743	Island	ZNF787	cg26951705

Figure 10. Epigenetic profile of individuals living in Centre of Italy. Venn diagram represents with blue color the DMRs selected. In the table chromosome, genomic position (GRCh37/hg19 assembly), the CpG element, the gene name and the CpGs - located in the BOP - that better distinguish between groups, are reported.



CHR	Genomic Position	Relation to CpG island	GeneName	Best CpG in the BOP
1	248100325-248100726	Island	OR2L13	cg03748376
1	113050813-113052301	Island	WNT2B	cg27181253
1	2266007-2266432	S_Shore	MORN1	cg05392448
1	150206971-150207842	S_Shore	ANP32E	cg21285133
1	212003431-212004280	Island	LPGAT1	cg13064658
2	27712554-27712842	S_Shore	IFT172	cg08787988
2	241989164-241989461	Island	SNED1	cg23491743
3	127541270-127542048	S_Shore	MGLL	cg02116612
3	53289533-53290213	Island	TKT	cg01262337
4	81124468-81124845	Island	PRDM8	cg11983124
5	127418594-127420512	Island	FLJ33630	cg16752652
5	175298515-175300066	Island	CPLX2	cg01469864
6	32632158-32633027	Island	HLA-DQB1	cg05341252
6	169977218-169977894	S_Shore	WDR27	cg21696374
6	30852102-30852676	N_Shore	DDR1	cg19178509
6	32489742-32490128	Island	HLA-DRB5	cg26981746
6	32489742-32490128	N_Shore	HLA-DRB5	cg01341801
6	32551851-32552331	N_Shore	HLA-DRB1	cg23905789
6	32632158-32633027	N_Shore	HLA-DQB1	cg09312897
7	1097405-1097953	Island	GPR146	cg21655171
8	6691833-6693135	S_Shore	XKR5	cg27319216
9	130706912-130707273	N_Shore	FAM102A	cg01681863
10	75407413-75407706	S_Shelf	SYNPO2L	cg17411546
10	79396095-79398495	Island	KCNMA1	cg05754819
10	134562566-134563206	S_Shore	INPP5A	cg02201479
10	94526-95301	N_Shore	TUBB8	cg26679884
10	133109193-133111052	Island	TCERG1L	cg18138206
11	74459858-74460558	N_Shore	RNF169	cg02673002
11	1247830-1248140	S_Shore	MUC5B	cg26676440
13	25743998-25746127	S_Shore	FAM123A	cg06779705
13	113540350-113540558	N_Shore	ATP11A	cg21463262
13	29913886-29914301	Island	MTUS2	cg13506281
16	86549069-86550512	N_Shore	FOXF1	cg06834912
17	80557446-80557686	Island	FOXK2	cg15292356
17	19617083-19617424	S_Shelf	SLC47A2	cg08102564
19	21265164-21265433	Island	ZNF714	cg06238316
19	3770566-3770942	S_Shore	RAX2	cg27288829
19	21265164-21265433	N_Shore	ZNF714	cg01483656

Figure 11. Epigenetic profile of individuals living in North of Italy. Venn diagram represents with blue color the DMRs selected. In the table chromosome, genomic position (GRCh37/hg19 assembly), the CpG element, the gene name and the CpGs - located in the BOP - that better distinguish between groups are reported.

6.3. GENOME-WIDE APPROACH FOR THE STUDY OF DNA METHYLATION DIVERSITY ACROSS MACRO-GEOGRAPHIC REGIONS

After the analysis of DNA methylation variability in Italy we questioned if DNA methylation profiles vary across macro-geographic areas and what are the genes and biological patterns whose DNA methylation levels distinguish populations of different ancestry (Africa, Asian and European). This analysis would allow to identify mechanism that likely play a role in recent human adaptation.

We performed a new analysis of the Infinium 450k assay data available online (Heyn et al. 2013) in GEO database. We applied the same pipeline used for Italian samples in order to identify regions with a functional role considering DNA methylation profile of both LCL (lymphoblastoid cell line) and whole blood cells in different populations. We analyzed 288 LCL samples composed by 96 African American (AFR), 96 Han Chinese America (EAS) and 96 American individuals of European ancestry (EUR). After quality check, the exclusion of CpG sites located in chromosomes X and Y and the exclusion of CpGs sites or probes that include SNPs with MAF>1% (The 1000 Genomes Project Consortium 2010) we recovered 440793 out 485577 loci for the analysis. The results obtained were validated in whole blood of a small cohort of individuals of similar origin (Asian, African and European), to reduce the tissue-specificity bias. An overview of the entire pipeline of analysis here performed considering individuals of European, African and Asian ancestry is reported in Figure 12.

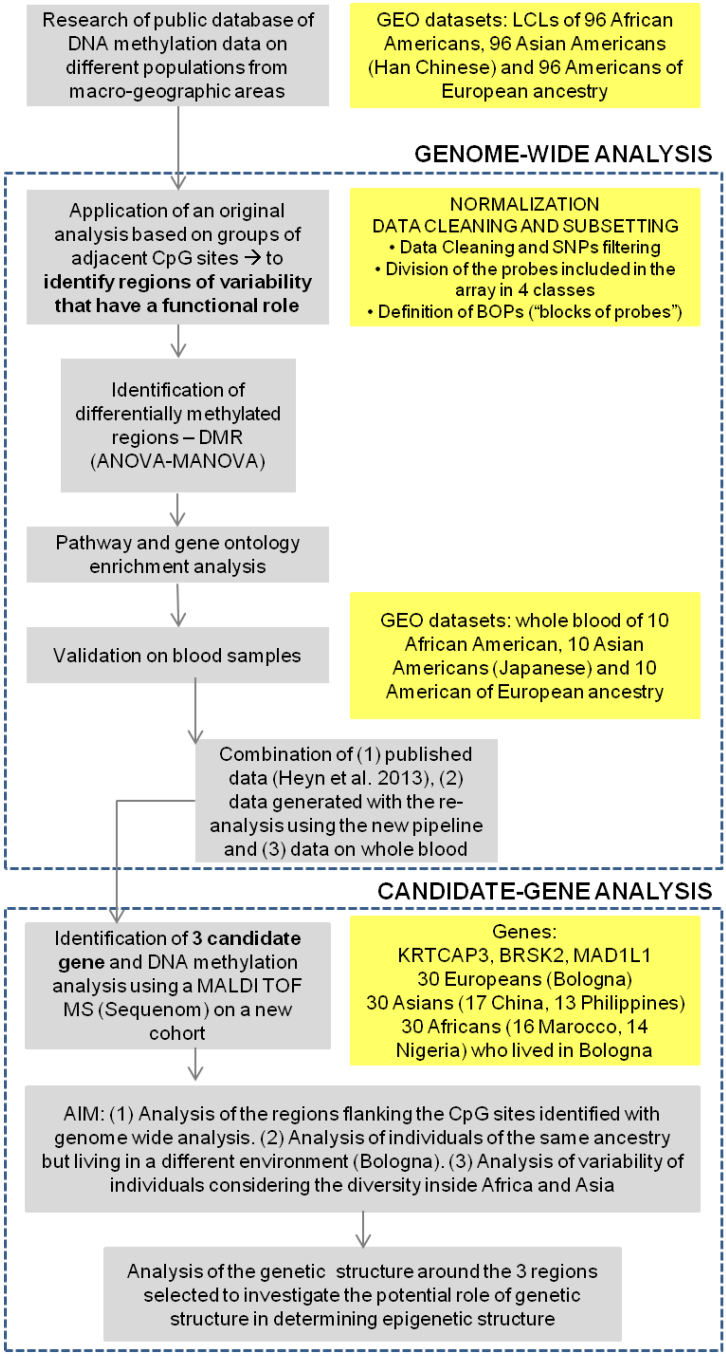


Figure 12. Overview of the pipeline of analyses applied to 3 human populations with different ancestry (European, African and Asian)

6.3.1. Identification of differentially methylated regions between three human populations (AFR, EAS and EUR)

To identify DMRs between the 3 populations we applied the new analysis pipeline described previuously (Bacalini et al. 2015).

In Class A, we identified 77, 217 and 301 BOPs able to discriminate respectively AFR and EU, AFR and EAS, EUR and EAS (q-value $<5*10^{-8}$). In Class B 15, 43 and 122 BOPs resulted differentially methylated in AFR and EU, AFR and EAS, EUR and EAS respectively (q-value$<10^{-8}$).

Methylation values of Class C and Class D probes were compared between groups by ANOVA. 53, 161 and 1815 CpG probes out of 106873 probes mapping in Class C and 62, 107 and 1559 out of 52979 of probes mapping in Class D were differentially methylated between AFR and EU, AFR and EAS, EUR and EAS respectively (q-value $< 10^{-8}$). Number and percentage of diferentially methylated BOPs/CpGs are reported in Table 2.

	Class A		Class B		Class C		Class D	
	N BOP	F BOP (%)	N BOP	F BOP (%)	N CpG	F CpG (%)	N CpG	F CpG (%)
AFR vs EUR	77	0.26	43	0.68	53	0.05	62	0.12
AFR vs EAS	217	0.73	15	0.24	161	0.15	107	0.2
EUR vs EAS	301	1.01	122	1.92	1815	1.70	1559	2.94

Table 2. Number and percentage of differentially nethylated BOPs for Class A and Class B in the pairwise comparisons between populations and number and percentage of diferentially methylathed CpG sites for Class C and Class D in the pairwise comparisons between populations

6.3.2. Gene ontology enrichment analysis

We performed a gene ontology enrichment analysis using the list of Class A BOPs differentially methylated between populations. Regarding the genes that include AFR-EUR DMR (regions differentially methylated between African-Americans and Americans of European ancestry) we did not observe an enrichment in GO terms with a p-value < 10^{-3}. However, DMRs are located in important genes, such as *PACS2* that is involved in the communication between endoplasmatic reticulum (ER) and mitochondria and it is involved also in mechanisms causing cell death, and *MAD1L1* one of the human accelerated regions (HAR3), *i.e.* whose regions that changed very little throughout mammalian evolution, but then experienced many changes in humans since divergence from chimppanzee (Hubisz and Pollard 2014).

AFR-EAS DMR (regions differentially methylated between African-Americans and Han Chinese-Americans) predominantly mapped in genes involved in interferon-gamma mediated signalling pathway such as HLA-C, HCK, IRF7, HLA-DQB2, HLA-E, HLA-DPB1genes (GO: 0060333, p-value nominal=$9.7*10^{-5}$, FDR q-value=ns), moreover the component of MHC protein complex are more represented (GO: 0042611 FDR quvalue=$4.23*10^{-2}$). It is important to note the over representation of the interferon-gamma mediated signalling pathway that is crucial for antigen presentation and immune cells activation. The classic process that activates this pathway is viral infections but interferon gamma is also crucial in inflammatory response and immunoregulation (Schroder et al. 2004). EAS-EUR DMR (regions differentially methylated between Han Chinese-Americans and Americans of European ancestry) predominantly map in genes involved in cell adhesion. The DMR mapped mainly in genes involved in homophilic cell adhesion via plasma membrane adhesion molecules, such as *PCDHA10, CDH4, PCDHA7, PCDHA8, PCDHA6,*

PCDHA5, PCDHA4, PCDHA3, *PCDHA9, PCDHA2, PCDHA1, PCDHB5* (GO: 0007156, FDR q-value = $5.18*10^{-3}$). These proteins regulate process including endocytosis, secretion, phagocytosis, cell adhesion and cell migration.

We then performed a gene ontology enrichment analysis using the list of Class C probes, *i.e.* all the probes that include CpG sites that map near/in a gene but not in islands, shore or shelves. Since the analysis were conducted using a single CpG approach, for the gene ontology enrichment analysis we selected all the genes that included at least 2 CpGs whose levels of DNA methylation differentiate the two populations analyzed. AFR-EAS DMR are located in gene involved in UDP glucuronosyltransferase, this enzyme is involved in the process of glucoronidation (Tukey and Strassburg 2000), i.e. a classic detoxification process characterized by the addiction of glucoronic acid to xenobiotics, process that leads to the formation of water-soluble metabolites. This mechanism is not only the most important pathway for the elimination of vast majority of drugs but also for the removal of toxins, pollutants and fatty acid derivates. The CpGs cg07952421 and cg10632656 are located in correspondence of the binding site of the transcription factor *POLR2A* that is responsible for messenger RNA synthesis located upstream the *UGT2B17*. Regarding the Class C AFR-EUR DMR, analysis of enrichment was no possible because only two genes resulted differentially methylated in more than one CpGs. The genes are *PRDM16* that codes for a protein involved in the differentiation of brown adipose tissue (BAT) and *MLPH* that codes for melanophilin a protein found in melanocytes (pigment producing cells). EAS-EUR comparison revealed an enrichment of DMRs located in genes involved in cellular glucoronidation (as highlighted for AFR-EUR DMR), developmental growth involved in morphogenesis, inorganic anion transport, regulation of membrane potential, flavonoid metabolism, cell adhesion and sensory perception of sound. The last process is interesting because it opens cultural consideration that should be addressed in the future. It is known in fact

the different music perception between Eastern and Western people and the Asian advantage for pitch memory (Schellenberg and Trehub 2008), for example Asians speak tone languages, where different-pitched vowel sounds have different meanings. Some authors suggested that language and also culture influence sound and external perception (for an extensive review see (Jandt 2004; Jandt 2007). What is interesting is the link between culture and molecular mechanisms such as DNA methylation, that here emerged for the first time.

6.3.3. Identification of an epigenetic profile of Americans of African, Asian and European origin

To identify an epigenetic profile characteristic of each descent (African, Asian and Europen) from the list of Class A BOPs we selected a list of DMRs whose DNA methylation status was remarkably different in two comparisons on the basis of p-values and DNA methylation difference, that has to be higher than 0.15 between the 2 groups. In other words epigenetic profile of African origin includes all the DMR that distinguish both AFR-EUR and AFR-EAS (as shows in Figure 13) with a p-value $< 5*10^{-8}$ (after Benjamini Hochberg correction) and with DNA methylation defference higher than 0.15 as previously suggested (Du et al. 2010). Figure 13 summarized the idea behind the identification of epigenetic profile of African origin (Figure 13A), Asian origin (Figure 13B) and European ancestry (Figure 13C) .

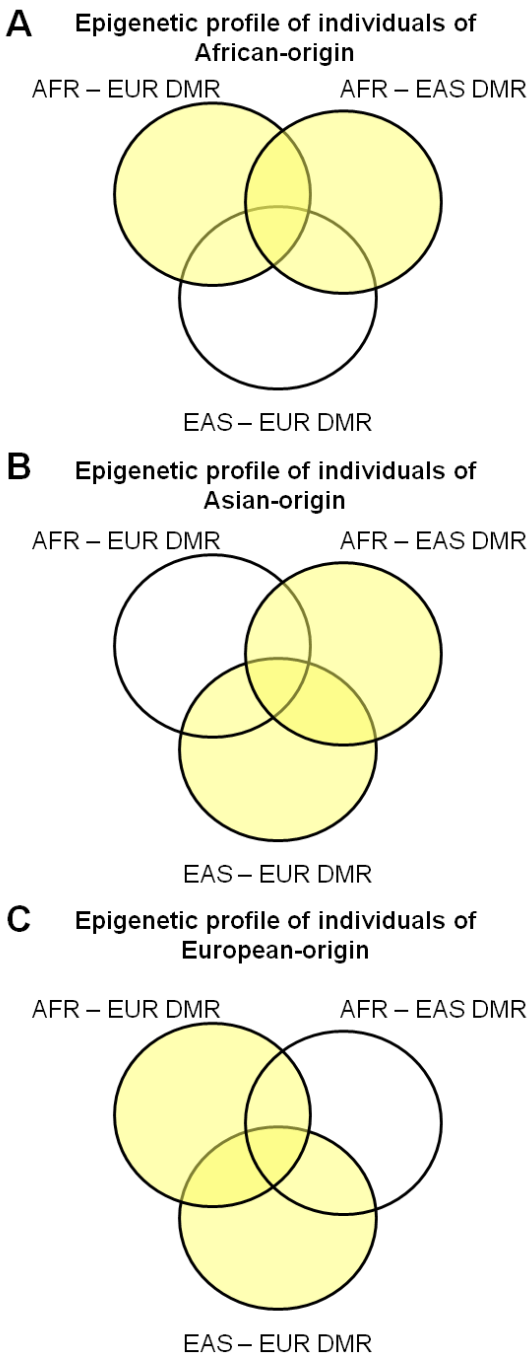


Figure 13. Strategy for the identification of epigenetic profile of individuals of African (A), Asian (B), and European origin (C)

This analysis showed that 66 DMRs differentiate individuals of African origin form from both individual of European and Asian orgin. DMRs identified were reported in Table 3. Gene ontology enrichment analysis was performed and no enriched GO terms were found for these genes.

CHR	Genomic Position	Relation to CpG island	GeneName	Best CpG in the BOP
1	3559952-3561009	Island	WDR8	cg26422465
1	3559952-3561009	N_Shore	WDR8	cg02764188
1	110230238-110230614	Island	GSTM1	cg11680055
1	59042013-59043295	Island	TACSTD2	cg01821018
1	246414748-246415569	Island	SMYD3	cg08887025
1	38157295-38158586	N_Shore	C1orf109	cg06917450
1	5937157-5937392	Island	NPHP4	cg14654471
1	153662203-153662439	Island	NPR1	cg21860360
1	101004471-101005885	N_Shore	GPR88	cg06223162
1	19970255-19971923	Island	NBL1	cg00755546
2	27664939-27665151	Island	*KRTCAP3*	cg21248554
2	99771109-99771637	Island	LIPT1	cg15544633
2	74725039-74727038	Island	LBX2	cg04255230
2	37035698-37035918	Island	VIT	cg06703213
4	3485963-3486381	Island	DOK7	cg15450966
4	1364146-1364349	N_Shore	KIAA1530	cg03698374
5	150051116-150052107	Island	MYOZ3	cg14388237
5	643427-643669	N_Shelf	CEP72	cg15402732
5	79864842-79866447	Island	ANKRD34B	cg24834873
6	30227320-30228255	Island	HLA-L	cg01655658
6	33048416-33048814	Island	HLA-DPB1	cg00570906
7	6570453-6571095	N_Shore	GRID2IP	cg14615128
7	1950279-1950482	N_Shore	*MAD1L1*	cg16178271
7	1885033-1885402	N_Shore	*MAD1L1*	cg16658412
7	75024250-75025212	S_Shelf	TRIM73	cg04681579
7	157937614-157937841	N_Shelf	PTPRN2	cg15935227
7	2054060-2054386	N_Shelf	*MAD1L1*	cg27286614
7	886803-887535	N_Shelf	UNC84A	cg10146442
8	145003589-145004145	Island	PLEC1	cg24891660
8	67976440-67977448	N_Shore	COPS5	cg27247252
9	95526459-95527825	N_Shore	BICD2	cg01028796
9	117160139-117161064	N_Shelf	AKNA	cg13910813
10	1252141-1252505	N_Shore	ADARB2	cg20205188
10	13515185-13515427	Island	BEND7	cg10309386
11	614760-616068	N_Shore	IRF7	cg16486109

11	46389074-46389378	Island	DGKZ	cg12856521
11	124615389-124617006	N_Shore	NRGN	cg26069044
12	124788559-124788884	N_Shore	FAM101A	cg19235645
12	132263617-132263856	Island	SFRS8	cg18997918
14	105830605-105830846	Island	*PACS2*	cg19499452
14	105944421-105945700	Island	CRIP2	cg06285241
14	33402094-33404079	S_Shelf	NPAS3	cg11787167
15	89920793-89922768	Island	LOC254559	cg01317586
16	604801-605348	N_Shelf	SOLH	cg26722972
16	2334108-2334347	Island	ABCA3	cg01491428
16	1493499-1493773	Island	CCDC154	cg06352616
16	85684107-85684491	N_Shore	KIAA0182	cg02817932
16	1358729-1360232	S_Shore	UBE2I	cg06161540
16	85684107-85684491	S_Shore	KIAA0182	cg26657404
17	47091037-47091567	Island	IGF2BP1	cg20966754
17	48585385-48586167	Island	MYCBPAP	cg03168497
17	47091973-47092211	Island	IGF2BP1	cg10950924
17	78039291-78039547	Island	CCDC40	cg07446795
17	113664-113881	N_Shore	RPH3AL	cg16767880
19	14582419-14585098	S_Shore	PTGER1	cg03020379
19	40950025-40950777	Island	SERTAD3	cg14150973
19	49375484-49375928	Island	PPP1R15A	cg19746536
19	56915356-56915856	Island	ZNF583	cg27129744
19	51294876-51295098	N_Shore	ACPT	cg06590444
19	55889005-55889476	N_Shore	TMEM190	cg08806681
20	61877854-61878071	Island	NKAIN4	cg07711085
20	30639908-30640786	Island	HCK	cg09796376
20	4803013-4804146	Island	RASSF2	cg14359894
21	47580492-47582065	Island	C21orf56	cg14789911
21	47601020-47601229	S_Shelf	C21orf56	cg10296238
21	47602431-47602740	S_Shore	C21orf56	cg08742575

Table 3. Genomic regions whose DNA methylation level can disttinguish individuals of African ancestry (profile). Chromosome, genomic position (GRCh37/hg19 Assembly), Relation to CpG island, the name of the gene and the CpG site that better distinguish the groups inside the region considered.

This analysis shows that 102 DMR differentiate individuals of Asian origin. DMRs identified were reported in Table 4. Gene ontology enrichment analysis was performed and BOPs mapped in gene involved in homophilic cell adhesion

via plasma membrane adhesion molecules (GO: 0007156, FDR q-value=$2.47*10^{-5}$) and nervous system development (GO: 0007399, FDR q-value=$3.02*10^{-4}$).

CHR	Genomic Position	Relation to CpG island	GeneName	Best CpG in the BOP
1	110230238-110230614	Island	GSTM1	cg11680055
1	59042013-59043295	Island	TACSTD2	cg01821018
1	246414748-246415569	Island	SMYD3	cg08887025
1	38157295-38158586	N_Shore	C1orf109	cg06917450
1	5937157-5937392	Island	NPHP4	cg14654471
1	153662203-153662439	Island	NPR1	cg21860360
1	101004471-101005885	N_Shore	GPR88	cg06223162
1	3559952-3561009	N_Shore	WDR8	cg02764188
1	19970255-19971923	Island	NBL1	cg00755546
1	26672445-26672650	N_Shore	AIM1L	cg14517454
1	205818898-205819191	S_Shore	PM20D1	cg24503407
1	152487978-152488270	N_Shore	CRCT1	cg21823605
1	155246991-155247748	Island	HCN3	cg14669863
2	27664939-27665151	Island	*KRTCAP3*	cg21248554
2	99771109-99771637	Island	LIPT1	cg15544633
2	74725039-74727038	Island	LBX2	cg04255230
2	37035698-37035918	Island	VIT	cg06703213
2	3321063-3321293	S_Shore	TSSC1	cg07144255
2	109744585-109746833	Island	SH3RF3	cg03846641
4	3485963-3486381	Island	DOK7	cg15450966
4	1364146-1364349	N_Shore	KIAA1530	cg03698374
5	643427-643669	N_Shelf	CEP72	cg15402732
5	79864842-79866447	Island	ANKRD34B	cg24834873
5	176296541-176296933	N_Shelf	UNC5A	cg16635016
5	1226976-1227305	N_Shore	SLC6A18	cg26427908
5	140242073-140243007	Island	PCDHA4	cg01042202
6	30227320-30228255	Island	HLA-L	cg01655658
6	33048416-33048814	Island	HLA-DPB1	cg00570906
6	33048416-33048814	S_Shore	HLA-DPB1	cg20981163
6	33048416-33048814	S_Shelf	HLA-DPB1	cg15138289
7	1885033-1885402	N_Shore	*MAD1L1*	cg16658412
7	75024250-75025212	S_Shelf	TRIM73	cg04681579
7	1950279-1950482	N_Shore	*MAD1L1*	cg16178271
7	157937614-157937841	N_Shelf	PTPRN2	cg15935227
7	2054060-2054386	N_Shelf	*MAD1L1*	cg27286614
7	886803-887535	N_Shelf	UNC84A	cg10146442
7	6570453-6571095	N_Shore	GRID2IP	cg14615128

7	155088705-155090212	S_Shore	INSIG1	cg21341645
7	57247588-57247927	Island	LOC642006	cg02066813
7	157503469-157504144	Island	PTPRN2	cg14168080
7	157524157-157526741	N_Shore	PTPRN2	cg12297440
8	145003589-145004145	Island	PLEC1	cg24891660
8	67976440-67977448	N_Shore	COPS5	cg27247252
8	144659745-144660635	Island	NAPRT1	cg08017634
8	134308328-134310145	N_Shore	NDRG1	cg17129188
9	95526459-95527825	N_Shore	BICD2	cg01028796
9	117160139-117161064	N_Shelf	AKNA	cg13910813
10	1252141-1252505	N_Shore	ADARB2	cg20205188
10	13515185-13515427	Island	BEND7	cg10309386
10	134918755-134918975	S_Shore	GPR123	cg23098789
10	123356616-123358285	N_Shore	FGFR2	cg25052156
10	134911976-134912483	N_Shore	GPR123	cg23713909
10	134925815-134926878	S_Shelf	GPR123	cg04553632
11	614760-616068	N_Shore	IRF7	cg16486109
11	46389074-46389378	Island	DGKZ	cg12856521
11	124615389-124617006	N_Shore	NRGN	cg26069044
11	289771-290010	Island	ATHL1	cg10261205
11	1410119-1412346	S_Shore	*BRSK2*	cg15465743
11	129306659-129306882	Island	BARX2	cg05471602
12	132263617-132263856	Island	SFRS8	cg18997918
12	111806559-111807060	S_Shore	FAM109A	cg10833066
12	124829153-124829490	N_Shore	NCOR2	cg23878260
13	113635654-113635899	N_Shore	MCF2L	cg13108304
13	114783502-114783733	Island	RASA3	cg01758122
14	105944421-105945700	Island	CRIP2	cg06285241
14	105830605-105830846	Island	*PACS2*	cg19499452
14	33402094-33404079	S_Shelf	NPAS3	cg11787167
15	89920793-89922768	Island	LOC254559	cg01317586
15	63892789-63893911	N_Shelf	FBXL22	cg01767116
15	68521418-68522237	S_Shore	CLN6	cg22576950
16	604801-605348	N_Shelf	SOLH	cg26722972
16	1493499-1493773	Island	CCDC154	cg06352616
16	2334108-2334347	Island	ABCA3	cg01491428
16	85684107-85684491	N_Shore	KIAA0182	cg02817932
16	1358729-1360232	S_Shore	UBE2I	cg06161540
16	85684107-85684491	S_Shore	KIAA0182	cg26657404
16	711065-711414	N_Shore	WDR90	cg16808927
17	47091037-47091567	Island	IGF2BP1	cg20966754
17	48585385-48586167	Island	MYCBPAP	cg03168497
17	47091973-47092211	Island	IGF2BP1	cg10950924

17	78039291-78039547	Island	CCDC40	cg07446795
17	113664-113881	N_Shore	RPH3AL	cg16767880
17	43971410-43975040	S_Shelf	MAPT	cg24801230
17	185075-185296	Island	RPH3AL	cg03226844
17	76897078-76897298	S_Shelf	TIMP2	cg20589078
19	40950025-40950777	Island	SERTAD3	cg14150973
19	49375484-49375928	Island	PPP1R15A	cg19746536
19	14582419-14585098	S_Shore	PTGER1	cg03020379
19	56915356-56915856	Island	ZNF583	cg27129744
19	51294876-51295098	N_Shore	ACPT	cg06590444
19	55889005-55889476	N_Shore	TMEM190	cg08806681
19	51128325-51128793	Island	SYT3	cg18118147
19	51051782-51052122	Island	LRRC4B	cg16212298
20	30639908-30640786	Island	HCK	cg09796376
20	4803013-4804146	Island	RASSF2	cg14359894
20	61877854-61878071	Island	NKAIN4	cg07711085
20	62097193-62098254	Island	KCNQ2	cg13782274
20	61150437-61150646	Island	C20orf166	cg15010213
21	47580492-47582065	Island	C21orf56	cg14789911
21	47601020-47601229	S_Shelf	C21orf56	cg10296238
21	47602431-47602740	S_Shore	C21orf56	cg08742575
22	24384134-24384405	S_Shore	GSTT1	cg17005068

Table 4. Genomic regions whose DNA methylation level can disttinguish individuals of Asian ancestry (profile). Chromosome, genomic position (GRCh37/hg19 Assembly), Relation to CpG island, the name of the gene and the CpG site that better distinguish the groups inside the region considered

55 DMRs differentiate individuals of European origin. The identified DMR were reported in Table 5. Gene ontology enrichment analysis was performed and BOPs mapped in gene involved in cell-cell adhesion (GO: 0098742, FDR q-value = $9.56*10^{-7}$), homophilic cell adhesion via plasma membrane adhesion molecules (GO: 0007156, FDR q-value = $8.36*10^{-8}$), nervous system development (GO: 0007399, FDR q-value = $9.6*10^{-5}$) and cranial suture morphogenesis such as FGFR2 and INSIG1 genes (GO: 0060363, p-values = $7.7*10^{-4}$; FDR q-value = ns).

CHR	Genomic Position	Relation to CpG island	GeneName	Best CpG in the BOP
1	3559952-3561009	Island	WDR8	cg26422465
1	3559952-3561009	N_Shore	WDR8	cg02764188
1	246414748-246415569	Island	SMYD3	cg08887025
1	38157295-38158586	N_Shore	C1orf109	cg06917450
1	26672445-26672650	N_Shore	AIM1L	cg14517454
1	205818898-205819191	S_Shore	PM20D1	cg24503407
1	152487978-152488270	N_Shore	CRCT1	cg21823605
1	155246991-155247748	Island	HCN3	cg14669863
2	74725039-74727038	Island	LBX2	cg04255230
2	3321063-3321293	S_Shore	TSSC1	cg07144255
2	109744585-109746833	Island	SH3RF3	cg03846641
4	3485963-3486381	Island	DOK7	cg15450966
5	150051116-150052107	Island	MYOZ3	cg14388237
5	176296541-176296933	N_Shelf	UNC5A	cg16635016
5	1226976-1227305	N_Shore	SLC6A18	cg26427908
5	140242073-140243007	Island	PCDHA4	cg01042202
6	33048416-33048814	S_Shore	HLA-DPB1	cg20981163
6	33048416-33048814	S_Shelf	HLA-DPB1	cg15138289
7	6570453-6571095	N_Shore	GRID2IP	cg14615128
7	1950279-1950482	N_Shore	*MAD1L1*	cg16178271
7	155088705-155090212	S_Shore	INSIG1	cg21341645
7	57247588-57247927	Island	LOC642006	cg02066813
7	157503469-157504144	Island	PTPRN2	cg14168080
7	157524157-157526741	N_Shore	PTPRN2	cg12297440
8	144659745-144660635	Island	NAPRT1	cg08017634
8	134308328-134310145	N_Shore	NDRG1	cg17129188
10	134918755-134918975	S_Shore	GPR123	cg23098789
10	123356616-123358285	N_Shore	FGFR2	cg25052156
10	134911976-134912483	N_Shore	GPR123	cg23713909
10	134925815-134926878	S_Shelf	GPR123	cg04553632
11	289771-290010	Island	ATHL1	cg10261205
11	1410119-1412346	S_Shore	*BRSK2*	cg15465743
11	129306659-129306882	Island	BARX2	cg05471602
12	124788559-124788884	N_Shore	FAM101A	cg19235645
12	111806559-111807060	S_Shore	FAM109A	cg10833066

12	124829153-124829490	N_Shore	NCOR2	cg23878260
12	132263617-132263856	Island	SFRS8	cg18997918
13	113635654-113635899	N_Shore	MCF2L	cg13108304
13	114783502-114783733	Island	RASA3	cg01758122
14	105830605-105830846	Island	*PACS2*	cg19499452
15	63892789-63893911	N_Shelf	FBXL22	cg01767116
15	68521418-68522237	S_Shore	CLN6	cg22576950
16	604801-605348	N_Shelf	SOLH	cg26722972
16	2334108-2334347	Island	ABCA3	cg01491428
16	711065-711414	N_Shore	WDR90	cg16808927
17	43971410-43975040	S_Shelf	MAPT	cg24801230
17	185075-185296	Island	RPH3AL	cg03226844
17	76897078-76897298	S_Shelf	TIMP2	cg20589078
19	14582419-14585098	S_Shore	PTGER1	cg03020379
19	51128325-51128793	Island	SYT3	cg18118147
19	51051782-51052122	Island	LRRC4B	cg16212298
20	61877854-61878071	Island	NKAIN4	cg07711085
20	62097193-62098254	Island	KCNQ2	cg13782274
20	61150437-61150646	Island	C20orf166	cg15010213
22	24384134-24384405	S_Shore	GSTT1	cg17005068

Table 5. Genomic regions whose DNA methylation level can disttinguish individuals of European ancestry (profile). Chromosome, genomic position (GRCh37/hg19 Assembly), Relation to CpG island, the name of the gene and the CpG site that better distinguish the groups inside the region considered

6.4. DEEPENING THE NATURAL EPIGENETIC VARIABILITY OF THREE CANDIDATE GENES KRTCAP3, MAD1L1 AND BRSK2

(1) We wondered if the environment is the major force that shaped the DNA methylation patterns described or if the ancestry is the main force. (2) Moreover we investigated at a candidate gene level the epigenetic variability of individuals coming from different countries belonging to the same continent (e.g. individuals from Morocco and Nigeria for Africa and individuals from China and Philippines for Asia). (3) Since the

Illumina 450k array allows the analysis of DNA methylation of a part of CpGs in the human genome, we investigated the DNA methylation pattern around the most significant CpG sites that distinguish the populations in order to assess if population variability regards only few CpG sites or also the CpG sites that map in the flanking regions.

To this purpose, DNA methylation of 90 individuals of African, Asian and European origins living in Bologna was measured. In particular to investigate the DNA methylation variability observable within macro-geographical areas, such as Europe, Africa and Asia we selected 30 samples of European origin, 17 individuals coming from China, 13 from Philippines, 16 from Morocco and 14 from Nigeria, a Sub-Saharan region. After a meta-analysis conducted between the data published by Heyn and colleagues (2013), the data here obtained and the results achieved using public available data on whole blood of individuals of different populations (*i.e.* African-Americans, Asian-Americans and Americans of European descent), three candidate regions were selected (in *KRTCAP3*, *MAD1L1*, *BRSK2* genes) and DNA methylation analysis was performed. Patterns of population genetic structure at the three genes was also analyzed considering populations sequenced in the 1000Genomes project (1000Genomes Consortium 2010)

6.4.1. Candidate gene selection

Considering BOP of Class A identified previously in the three pairwise comparisons we selected all the CpGs located in this BOPs (11555 CpG sites, q-value $<10^{-8}$ and methylation difference of 0.15 for 2 consecutive CpGs). DNA methylation of these 11555 sites was extracted from a public dataset that considers DNA methylation from whole blood in individuals of comparable origin (GSE36064) . The whole blood dataset here analyzed was composed by 10

individuals of African origin, 10 individuals of Asian origin and 10 individual of African origin. Absolute pairwise difference between blood DNA methylation of the 3 groups were calculated and ANOVA was performed. The most significant CpGs (according to p-value and DNA methylation difference) in LCL and whole blood dataset that are also identify in the study of Heyn and colleagues were selected for further analysis. The CpGs sites are *cg21248554* (GRCh37/hg19 chr2: 27,665,151*), cg16658412 (*GRCh37/hg19 chr 7:1,883,420) and *cg15465743 (*GRCh37/hg19 chr11: 1,413,145) also identified in the populations profiles and mapped in *KRTCAP3*, *MAD1L1* and *BRSK2* genes, respectively. Using a MALDI TOF MS technology (Sequenom Epityper) we measure DNA methylation levels of 16 CpGs units nearby cg21248554 (GRCh37/hg19 chr2: 27,664,896-27,665,359), 13 CpG units nearby cg 16658412 (GRCh37/hg19 chr7: 1,883,322-1,883,759) and 6 CpGs units nearby cg 15465743 (GRCh37/hg19 chr11:1,413,109-1,413,442).

6.4.2. DNA methylation variability between macro-geographic groups

Considering *KRTCAP3* methylation status in individuals of African, European and Asian origin, the DNA methylation of the CpGs sites nearby cg21248554 are concordant and showed a general hypomethylation of individuals of African origins when compared with individuals of European or Asian origin. This hypomethylation of African individuals was also observed in the study of Heyn and colleagues. ANOVA was performed and CpGs 7.8.9.10, 11, 12, 13.14, 34, 35, 36 resulted significantly different between groups with a nominal p-value < 0.05. The *MAD1L1* amplicon showed lower values of DNA methylation for individuals of African origins, the results is confirmed also for CpGs site nearby the cg16658412. This hypomethylation of African individuals was also observed in the study of Heyn and colleagues. ANOVA was performed and CpGs 1,2,4,6,8,9,10,12,13 showed a nominal p-value < 0.05. DNA methylation of the

BRSK2 amplicon showed lower methylation levels for individuals of Asian origin than for European and African individuals. This hypomethylation of Asian individuals was also observed in the study of Heyn and colleagues. ANOVA was performed and all CpG sites showed a significant difference between groups with a nominal p-value < 0.01. Average DNA methylation values for the three regions analyzed is reported in Figure 14.

These results confirmed the data published by Heyn and colleagues (2013). Differences between populations are reported also in the CpGs flanking the CpGs analyzed with a microarray approach (Illumina 450k BeadChip), highlighting the potential possibility of a functional importance of the regions selected. Obviously more data on gene expression and protein activity are needed to draw further conclusions.

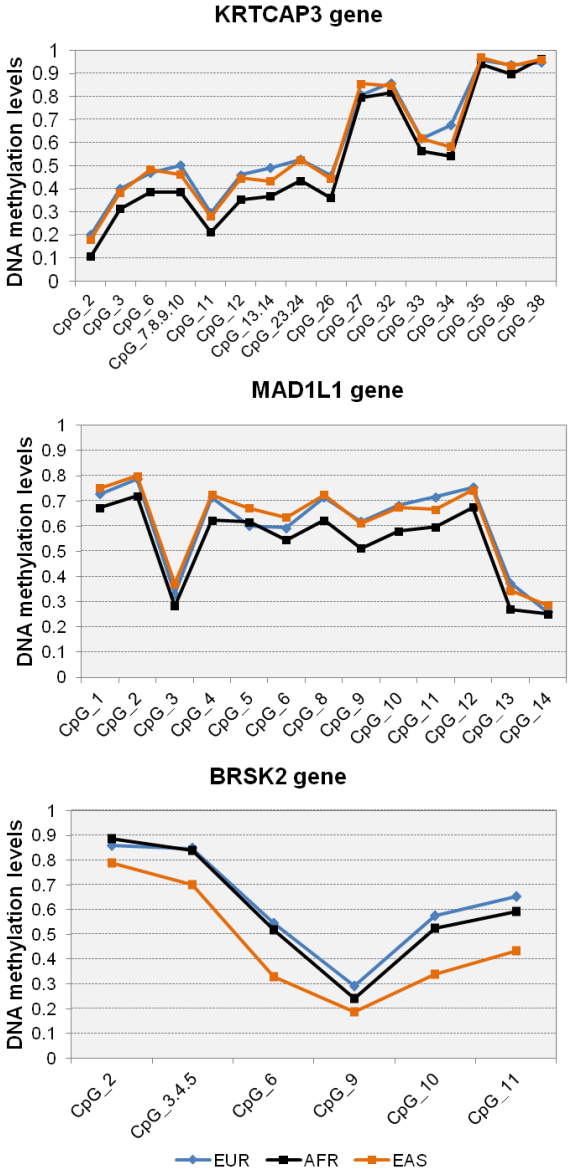


Figure 14. Average DNA methylation levels of *KRTCAP3*, *MAD1L1* and *BRSK2* according to macro-geographic groups (individuals were grouped according to African, European and Asian origin)

6.4.3. DNA methylation variability within macro-geographic areas

Recent studies considered DNA methylation differences across macro-geographic areas (*i.e.* Africa, Asia and Europe) without considering the high genetic, cultural and environmental diversity that is observable within each area. Here we selected samples from two different African countries *(i.e.* Morocco and Nigeria) and from two countries of Asia (*i.e.* China and Philippines). Pairwise t test was performed considering DNA methylation levels of the 3 amplicons (*i.e. KRTCAP3*, *MAD1L1* and *BRSK2*). The *BRSK2* gene showed differences between individuals coming from Morocco and individuals coming from Nigeria (pairwise t-test: *CpG 6* p-value = $5.2*10^{-3}$; *CpG 9* p-value = $2.6*10^{-3}$; *CpG 10* p-value = $3.1*10^{-3}$; *CpG 11* p-value = $2.8*10^{-3}$). This result demonstrates a high level of DNA methylation diversity between individuals belonging to the same cntinent (Africa). Moreover it is noteworthy that despite all individuals lived in Bologna, DNA methylation levels of Maroccans clustered with Italian individuals, whereas DNA methylation values of individuals of Nigerian origin clustered with individuals of Asian origins (*i.e.* China and Philippines) as Figure 15 showed. This indicates that DNA methylation patterns can strongly vary across Africa.

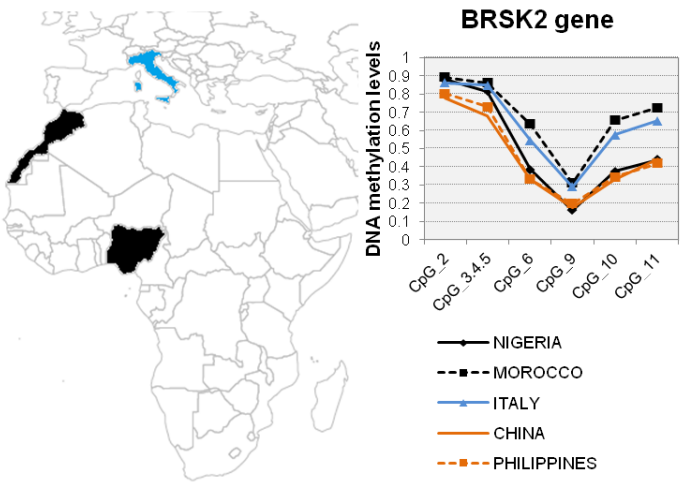


Figure 15. Average DNA methylation for the 6 CpGs analyzed in *BRSK2* gene. Black lines indicate individuals of African orifin (living in Bologna), blue line indicate Italian individuals and ornge lines indicate individuals of Asians origin

Here we suggest two hypothesis to explain this result as regards *BRSK2*, a gene involved in insulin secretion and abundantly expressed in pancreatic islets (Chen et al. 2012). (1) A first hypothesis is that according to their population history each group has evolved a different strategy to cope with a new environment. The European influence (both cultural and in term of genetic admixture) on peolple from Morocco is well known and this could be one possible explanation of the pattern observed. (2) The second hypothesis - more specultive - is based on recent studies that suggest that nutrition transitions are one of the main causes of modern epidemic of diabetes, obesity and metabolic syndrome all over the world. In particular, sub-Saharan Africa migrants are particularly affected by obesity and type 2 diabetes (Agyemang et al. 2009; Hossain, Kawar, and El Nahas 2007), as well as Chinese and Philippine populations (Dahly et al. 2013; Popkin and Gordon-Larsen 2004). The methylation patterns of this gene could be a possible mechanism that contributes to the high susceptibility to these diseases, but more

data are needed to support this hypothesis, since we do not have data on the prevalence of diabetes in our cohort.

6.4.4. Population genetic structure at *BRSK2*, *KRTCAP3*, and *MAD1L1* genes

Heyn and colleagues in their study identified a pool of pop-CpGs whose methylation level is associated to a certain genotype (meQTL), suggesting a genetic influence on DNA methylation levels. Here we investigated patterns of population genetic structure at the 3 examinated genes using genome-wide data from several human groups included in the 1000Genomes database. Elucidating the population genetic structure observable at these genes could be crucial to disentangle the role of genetics in driving the differences observed in DNA methylation patterns. We considered phase 3 genotypes of 15 populations sequenced by 1000Genomes project*, i.e* 5 of East Asian origin (Han Chinese in Bejinh CHB, Japanese in Tokyo - JPT, Southern Han Chinese - CHS, Kinh in Ho Chi Minh City - KHV), 5 of European origin(Utah Residents with Northern and Western European ancestry - CEU, Toscani in Italy - TSI, Finnish in Finland - FIN, British in England and Scotland - GBR, Iberian population in Spain - IBS), 5 of African origin (Yoruba in Ibadan, Nigeria - YRI, Luhya in Webuye, Kenya - LWK, Gambian in Western Divisions in The Gambia - GWD, Mende in Sierra Leone - MSL, Esan in Nigeria - ESN). Data were filtered for allelic state (*i.e.* only biallelic loci were retained), then a pruned subset of SNPs in approximate linkage equilibrium with each other was generated (Alexander, Novembre, and Lange 2009) and monomorphic loci were excluded. The genomic localization and the number of SNPs considered in the analyses are listed in Table 6.

Gene Name	Genomic localization (GRCh 37)	Length (bp)	Number of SNPs after filtering
KRTCAP3	chr2: 27665233-27669348	4848	38
MAD1L1	chr7:1855429 – 2272878	417449	2975
BRSK2	chr11: 1411129-1483919	72790	466

Table 6. Genomic localization and number of SNPs onsidered to assess the genetic population structure of *KRTCAP3*, *MAD1L1* and *BRSK2* genes.

Using this data we investigated genetic diversity of the 15 populations considered by discriminant analysis of principal components (DAPC) that allows a description of homogeneous genetic clusters using few synthetic variables.

Considering the ***KRTCAP3* gene**, DAPC identified 2 main clusters, one including African populations and the second one including both European and Asian groups as reported in the scatterplot (Figure 16). SNPs with the highest Fst value between populations of African and non-African-origin were rs62131871 and rs780112 (Fst = 0.22 and 0.35 respectively, estimated according to Nei 1973). The genetic structure showed in the scatterplot is in agreement with the differences identified via epigenetic analysis, placing individuals of African origin apart from European and East Asian populations. On the basis of this data we cannot exclude that genetic structure could impact on chromatin conformation and therefore on the observed DNA methylation patterns. Moreover we selected only three populations whose individuals' ancestry is more similar to individuals considered in DNA methylation analysis, *i.e.* YRI, individuals of Nigerian origin, CEU, Utah residents with Northern and Western European ancestry and CHB, Han Chinese in Bejing. We calculated Fst values considering SNPs located in a wider genomic region (chr2:27,617,372-27,717,244), to look for those whose

frequency is different between individuals of African ancestry (AFR) and individuals of non-African ancestry (CEU and CHB) as DNA methylation profile suggested. The only SNPs that differentiates these populations are rs1728911, rs780110, rs1260327, rs780105, rs780108 (Fst CEU-YRI were 0.20, 0.21, 0.30, 0.30, 0.31 and Fst YRI-CHB were 0.56, 0.58, 0.70, 0.70, 0.70 respectively). Further experiments of genotyping and methylation are needed to understand the role of these SNPs in determining epigenetic structure and levels.

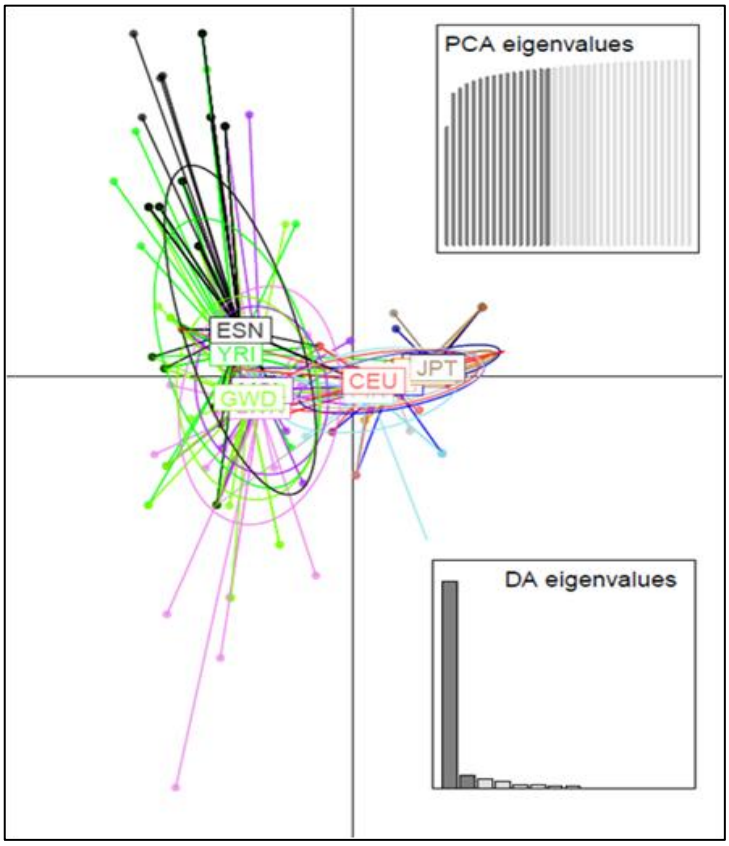


Figure 16. DAPC scatterplot of *KRTCAP3* gene considering 15 populations of African, Asian and European origin. The plot represents individuals as dots and the populations as inertia ellipses. Eigenvalues and principal component of the analysis are displayed.

Considering the ***MAD1L1* gene**, DAPC inferred three main clusters, one including African populations (YRI, LWK,ESN, MSL, GWD), the second one including individuals of European ancestry (GBR, TSI, CEU, FIN, IBS) and the third one including East Asian groups (CHB, CHS, JPT, KHV, CDX) as reported in the scatterplot in Figure 17. The first component of the analysis revealed a clear difference between individuals of African ancestry and individuals of non African ancestry (Europeans and Asians) as DNA methylation patterns suggests. The scenario is similar to *KRTCAP3* in which we cannot exclude that genetic structure could have an impact on DNA methylation profiles. Experiments of DNA sequencing and bisulphite sequencing of candidate regions are needed to disentangle this complicated scenario. Moreover we selected only three populations whose individuals' ancestry is more similar to individuals considered in DNA methylation analysis, *i.e.* YRI, individuals of Nigerian origin, CEU and CHB. We calculated Fst values considering all SNPs in the *MAD1L1* gene (chr7: 1,855,426-2,272,683), to look for SNPs whose frequency is considerably different between individuals of African ancestry (AFR) and individuals of non-African ancestry (CEU and CHB) as DNA methylation profile suggested. The only SNPs that differentiates these populations were rs73288707, rs55934553, rs6968659, rs3839699 (Fst CEU-YRI were 0.33, 0.30, 0.35, 0.49 and Fst YRI-CHB were 0.34, 0.35, 0.35, 0.36 respectively). Further genotyping and methylation experiments are needed to understand the role of these SNPs in determining epigenetic structure.

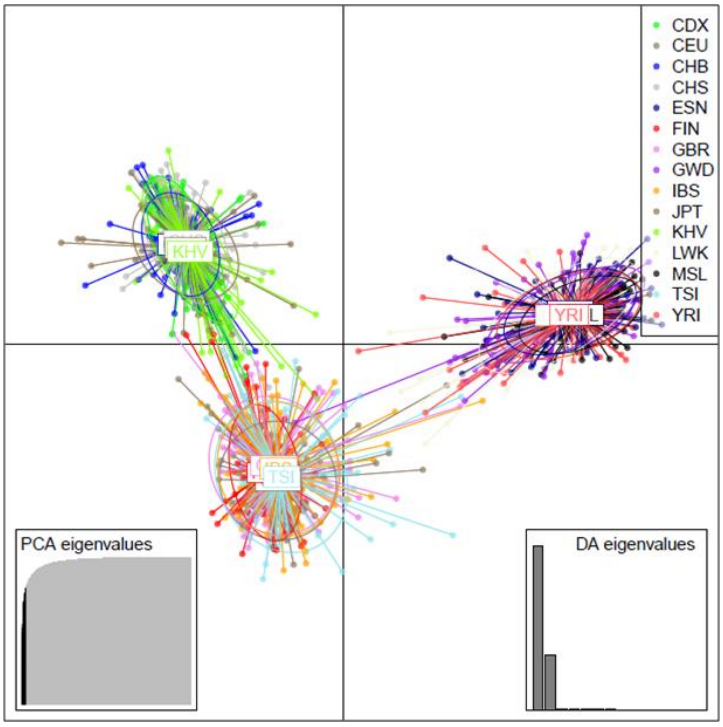


Figure 17. DAPC scatterplot of *MAD1L1* gene considering 15 populations of African, Asian and European origin. The plot represents individuals as dots and the populations as inertia ellipses. Eigenvalues and principal component of the analysis are displayed

Considering the ***BRSK2* gene**, DAPC suggested the presence of three main clusters, one including African populations (YRI, LWK,ESN, MSL, GWD), the second one including individuals of European ancestry (GBR, TSI, CEU, FIN, IBS) and the third one including East Asian groups (CHB, CHS, JPT, KHV, CDX) as reported in the scatterplot (Figure 18). It is interesting to note that DNA methylation profiles of *BRSK2* amplicon cannot be explained with the genetic structure of the region alone. In fact, DNA methylation of individuals of Nigerian origin is similar to DNA methylation of individuals of Asian origin, on the contrary genotypes frequency of these individuals are very diverse considering both PC1 and PC2. We then selected the SNPs with the highest Fst

values between populations of European ancestry and population of non-European ancestry, according to DNA methylation pattenrs identified in Figure 15. rs12419563, rs61869026, rs61868972, rs34805614, rs4881748 and rs7395567 were the SNPs with the highest Fst values in the comparison between the two groups (EUR vs EAS/AFR). Genotyping of this SNPs in the same samples analyzed could be interesting to elucidate the role of genetic variability in determining population specific DNA methylation profiling.

Moreover we selected only three populations (YRI, individuals of Nigerian origin, CEU and CHB) and we calculated Fst values of all SNPs located in *BRSK2* gene (chr11: 1,397,560-1,497,452), in order to identify the SNPs whose frequency is considerably different between individuals of European ancestry (CEU) and individuals of non-European ancestry (YRI and CHB) as DNA methylation profile suggested. The only SNPs that appreacibly differentiate these populations are rs61869028 (Fst CEU-YRI =0.27 and Fst CEU-CHB =0.34 and Fst YRI-CHB = 0.014). An experiment of genotyping and methylation typization should be performed to understand the role of this SNP in determining epigenetic structure.

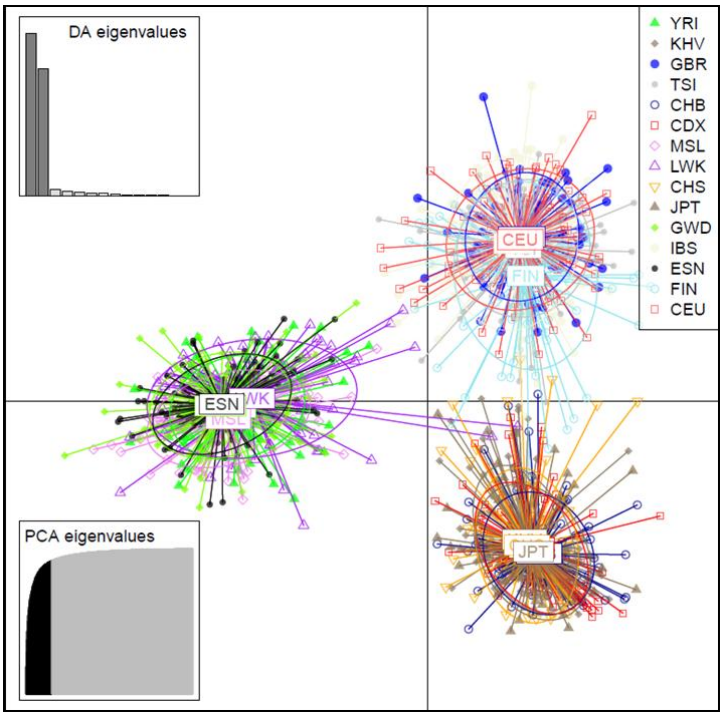


Figure 18. DAPC scatterplot of *BRSK2* gene considering 15 populations of African, Asian and European origin. The plot represents individuals as dots and the populations as inertia ellipses. Eigenvalues and principal component of the analysis are displayed

In conclusion we showed that the epigenetic structure at the *MAD1L1* and *KRTCAP3* genes could be explained by the genetic variation observed between populations. However Heyn and colleagues did not identify meQTL for the CpGs located in *MAD1L1* and *KRTCAP3* even if their data are obtained by SNPs microarray that captures only a part of genetic variation present in the genome (*i.e.* 550K or 660K variants). Noteworthy most of the potential SNPs involved in the genetic diversity here identified are not present in the Illumina 550K array. Our data demonstrated that we cannot exclude that the genetic context nearby the pop-CpGs can influence the chromatin structure and therefore can have an impact on DNA methylation profile. However the main

limit of our approach is that, despite individuals were selected according to their ancestry, DNA methylation profiles and genetic background were analyzed in 2 different cohorts.

Chapter 7

RESULTS AND DISCUSSION - AGE-DEPENDENT EPIGENETIC VARIABILITY

The purpose of this chapter is to elucidate the methylation changes that naturally occur during aging, focusing the analysis on the identification of potential biomarkers useful for the anthropological research. To this aim we firstly performed a genome-wide analysis of DNA methylation status in whole blood considering individuals of different ages (Section 7.1). We identified some regions whose DNA methylation levels correlated with age(ELOVL2, FHL2 and PENK genes) and we analyzed their methylation status in teeth, one of the best source of DNA for anthropological studies (Section 7.2).

7.1.DNA METHYLATION LEVEL IN WHOLE BLOOD AS A BIOMARKERS OF AGE

We wondered what are the changes that occur during physiological aging in whole blood to identify potential biomarkers for age-estimation.

A genome wide DNA methylation analysis (Illumina Infinium HumanMethylation450 BeadChip) on whole blood of a small cohort of 64 (32 mothers and 32 offspring) subjects was performed (Garagnani et al. 2012). The age range was 42-83 for the mothers and 9-52 for offsprings. ANOVA analysis identified 163 CpG whose DNA methylation levels are associated with age (*p-value* < Bonferroni treshold (2.2×10^{-8})). The top 3 significant loci mapped within the CpG islands of ELOVL2, FHL2 and PENK genes. We then replicated these results in a wider cohort constitues by 494 individuals (245 men and 249 women) ranging from 9 to 99 years and 7 cord-blood (3 males and 4 females) (Garagnani et al. 2012) . DNA methylation levels of individual CpG sites in *ELOVL2, FHL2* and *PENK* CpG islands was performed by the EpiTYPER assay

(Sequenom, San Diego, CA), a MALDI-TOF mass spectrometry-based method, starting from DNA extracted from whole blood. The genomic localization of he regions analyzed is reported in Figure 19.

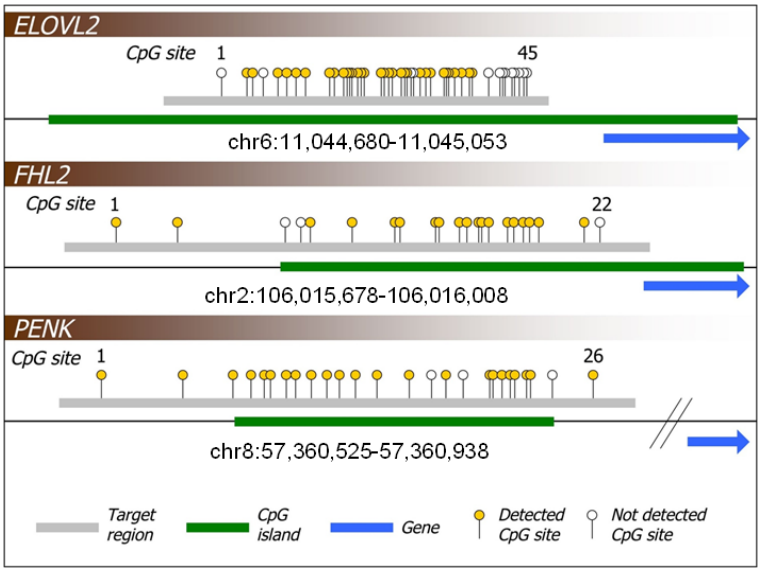


Figure 19. Genomic localization of ELOVL2, FHL2 and PENK genes considering CpG islands and genes (GRCh37/hg19 Assembly)

Spearman's correlation analysis for the 3 genes showed a strong correlation bethween DNA methylation and age. ELOVL2 CpGs unit 11.12.13 showed the highest correlation value of 0.95 (Figure 20A). FHL2 showed a correlation value of 0.80 for CpG 9.10 (Figure 20B) and CpG 19.20, and PENK of 0.63 (CpG 23.24) (Figure 20C). ELOVL2 encodes for a transmembrane protein involved in the synthesis of polyinsaturated fatty acids (PUFA), but its role in blood cells is not completely understood and this is the first study in which ELOVL2 is associated with aging. The role that DNA methylation of ELOVL2 played during aging is yet unknown but considering that PUFA are involved in processes such as modulation of inflammation and maintanance of cell membrane integrity it is

possible that ELOVL2 DNA methylation could paly a role in individual adptation to physiological changing that occurs during aging. Further studies are needed to elucidate the function of DNA hypermethylation during aging, however the outstanding results obtained make these three genes a potential candidate for forensic applications aimed at identifying proband age.

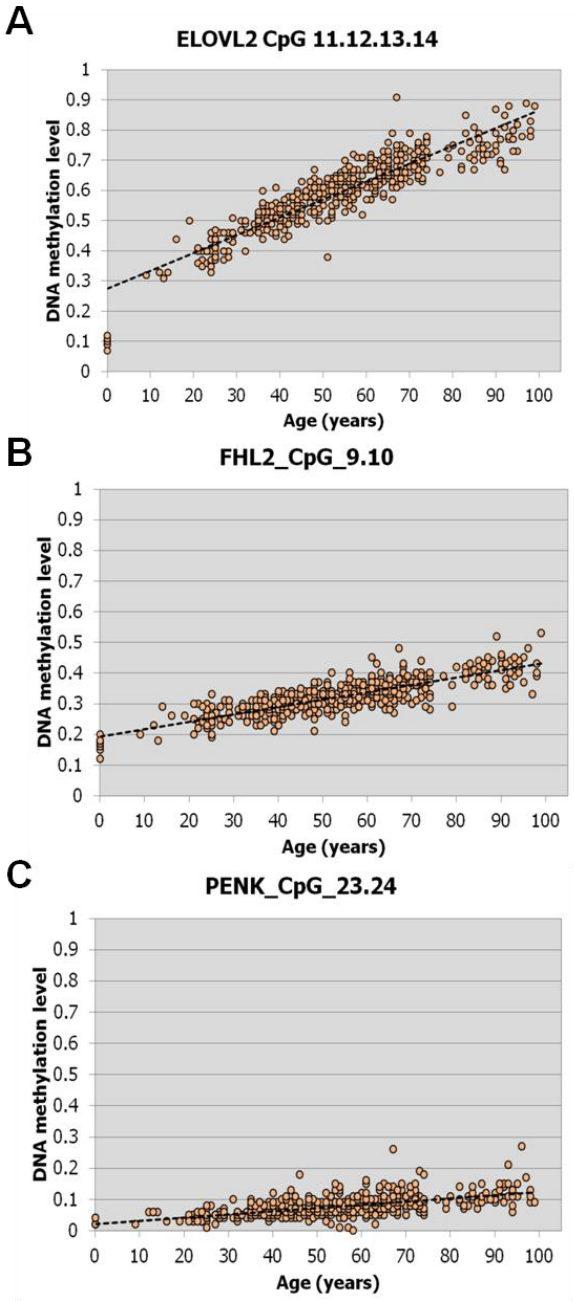


Figure 20. Methylation values of the CpG units that better correlate with age in whole blood. DNA methylation of CpG 11.12.13.14 for ELOVL2 (A), CpG 9.10 for FHL2 (B) and CpG 23.24 for PENK (C)

7.2. DNA METHYLATION LEVEL IN TEETH AS A BIOMARKERS OF AGE

We wondered if age-dependent DNA methylation changes of ELOVL2, FHL2 and PENK identified in whole blood could be a potential candidate for anthropological applications in teeth.

We analyzed DNA methylation levels of few CpGs located in three genes (ELOVL2, FHL2 and PENK) of cementum, dentin and pulp of 21 modern teeth of individuals - from 17 to 77 yrs – and we built a model to predict age. Tooth is one of the best source of DNA in ancient samples due to both its sheltered localization in jawbone and its peculiar structure that provides protection from post-mortem decay. Teeth are also less sensitive to contamination if compared to bones (Adler et al. 2011; Higgins and Austin 2013; Higgins et al. 2013; Pilli et al. 2013). Since DNA methylation is a tissue-specific process (Rakyan et al. 2008), DNA has been extracted frome 3 different layers - pulp, cementum and dentin - of 21 teeth from 17 to 77 years to assess DNA methylation profile in relation to tissue composition and age.

7.2.1. DNA methylation and age: single CpGs analysis

We calculated Pearson's correlation between age and methylation levels for each CpGs unit in the 3 tissues analyzed (cementum, dentin and pulp). Correlation coefficients were also calculated using DNA methylation values of 12 whole teeth. Figure 21 shows, through a gradient of colors, the correlation values for each district for the 3 genes; the red asterisks indicate CpGs sites whose DNA methyation levels significantly correlate with age (nominal p value< 0.01). In all

cases, these sites showed a positive correlation and tended to be hypermethylated with advancing age with the exception of PENK CpG 21.22 and PENK CpG 3.

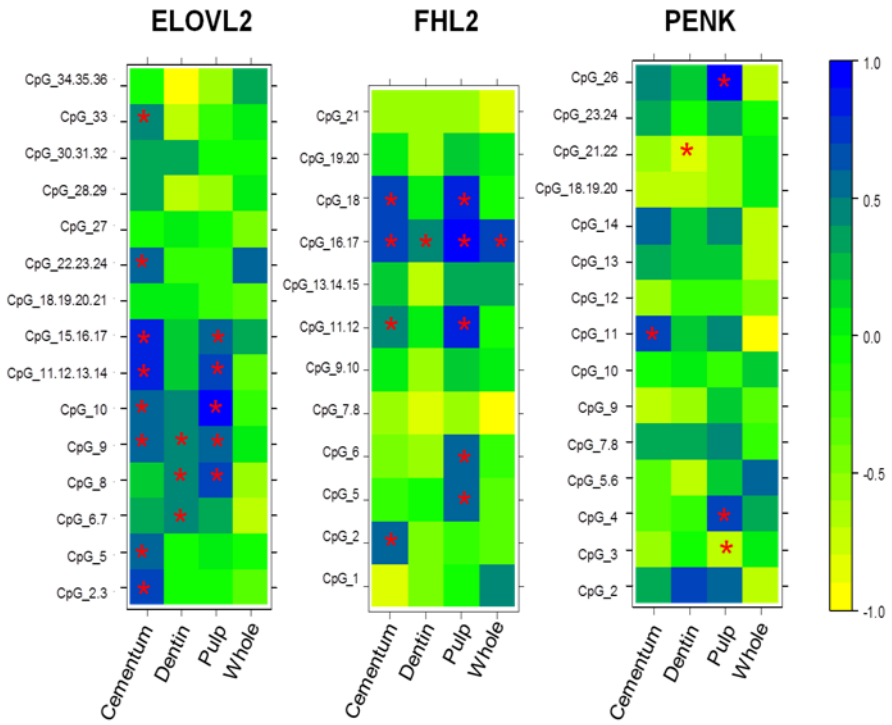


Figure 21. Levelplot for correlation values of DNA methylation and age. Pearson's correlation coefficient were calculated for each CpG unit in ELOVL2, FHL2 and PENK considering the 3 different tissues (cementum, dentin and pulp) and the DNA extracted from the whole tooth. Positive correlations are indicated with blue colors while the inverse correlation is indicated with yellow colors. The red asterisks indicate CpGs that strongly correlate with age (nominal p-value<0.01)

7.2.2. DNA methylation and age: multivariate linear regression model

After a detailed analysis of DNA methylation of the single CpGs, a multiple linear regression was used to predict age from DNA of teeth. A multiple regression model can be used to predict the age by simply plugging the beta values of the selected CpGs into the formula $Age=b_0 + b_1CpG_1 + b_2CpG_2 +... + b_NCpG_N$ as other authors suggested (Hannum et al. 2013; Horvath 2013). For each district (pulp, dentin and cementum) a multiple linear regression model was calculated considering the CpGs units that are strongly correlated with age for the 3 genes. In the pulp a multiple linear regression model was built considering 13 CpGs units and the median of the absolute difference between Age and methAge (age calculated o n the basis of DNA methylation level) was 2.25 (Figura 22A). In the dentin a multiple linear regression model was built considering 5 CpGs units and the median of the absolute difference between Age and methAge was 7.07 (Figura 22B). In the cementum a multiple linear regression model was built considering 13 CpGs units and the median of the absolute difference between Age and methAge was 2.45 (Figura 22C).

However the best way to predict age is to consider DNA methylation values of the best correlating tissues (cementum and pulp). At this purpose DNA methylation values of 8 CpGs units were included in a new model to predict age from cementum and pulp. The median of the absolute difference between Age and methAge was 1.20 for this model. However in this case it is possible to predict age only if it is possible to collect and to separate cementum and pulp from a single tooth.

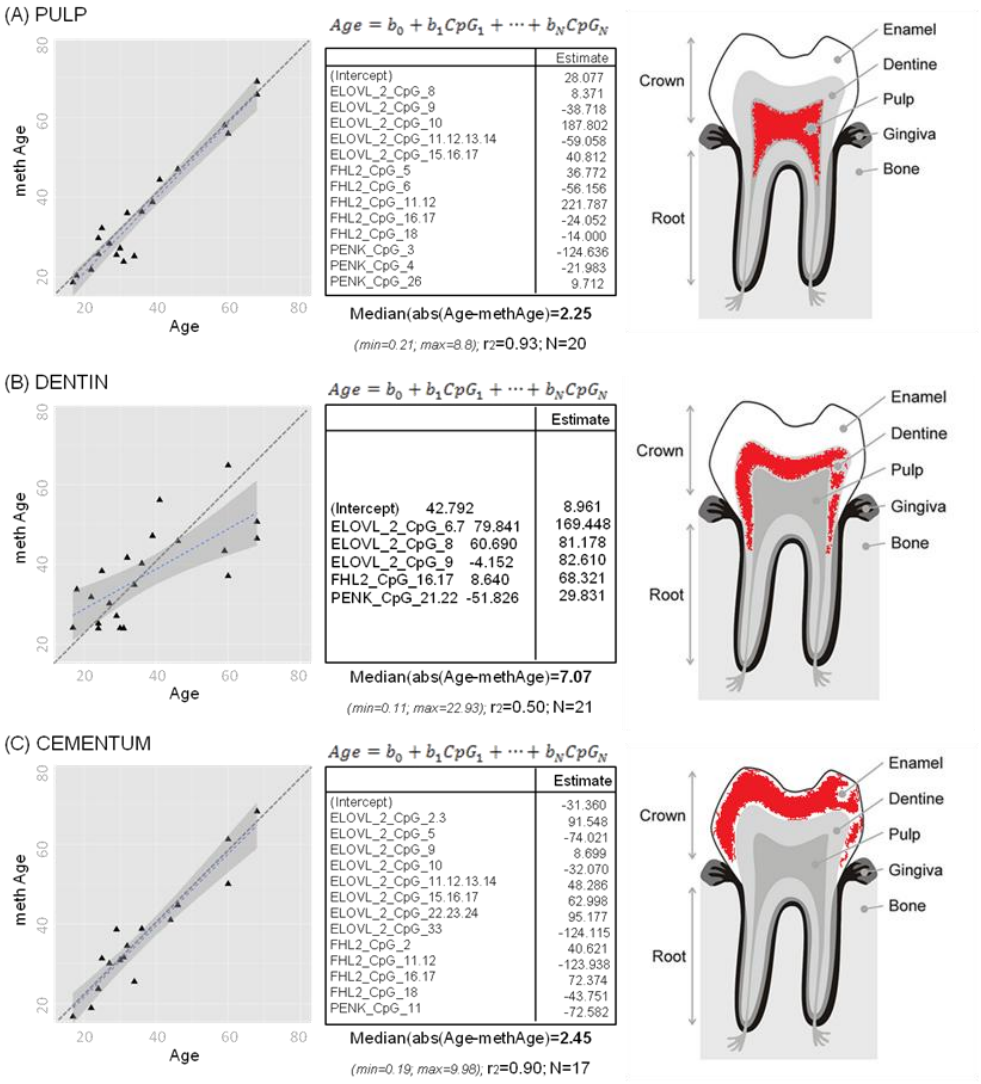


Figura 22. Multiple linear regression model for DNA extracted from dental pulp (A), dentin (B) and cementum (C). The scatter plots report Age against Age calculated according to the multiple linear regression model (methAge) on the basis of the formula reported on the left part of the figure. The gray dashed line indicates the bisector, to indicate the perfect concordance between estimated age and the real age. The regression line is plotted in blue and CI 95% is indicated with a gray shadow. Median, minimum and maximum of the absolute difference between Age and methAge as well as multiple r square and the number of individuals considered in the model are reported.

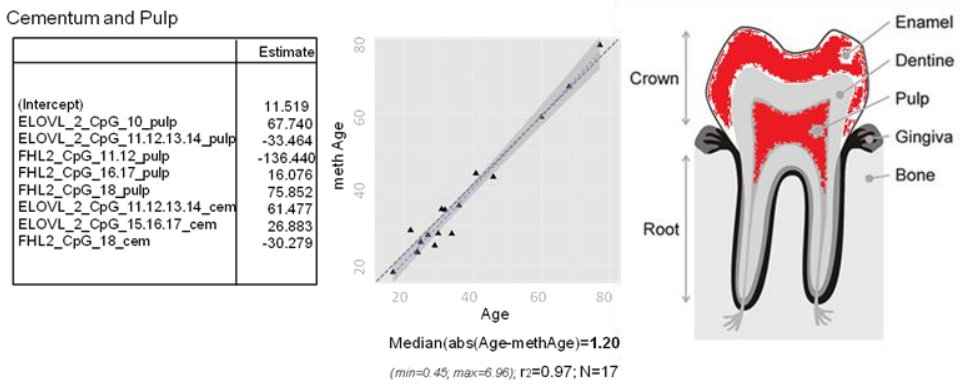


Figure 23. Multiple linear regression model considering DNA methylation of both cementum and pulp. The scatter plots report Age against Age calculated according to the multiple linear regression model (methAge). The gray dashed line indicates the bisector, to indicate the perfect concordance between estimated age and the real age. The regression line is plotted in blue and CI 95% is indicated with a gray shadow. Median, minimum and maximum of the absolute difference between Age and methAge as well as multiple r square and the number of individuals considered in the model are reported.

These data demonstrated that the pulp is the area with the highest correlation with age, but alternatively also the cementum could be used to predict age with high accuracy. The methylation status of DNA extracted from dentin did not show strongly correlation with age. This could be due to the composition of this layer. Recent studies showed that dentin can be subdivided in primary dentin, formed during tooth development, secondary dentin, secreted throughout the life, and tertiary dentin, also called "reparative dentin", that is secreted in response to external damage (Goldberg and Smith 2004; Saygin, Giannobile, and Somerman 2000). The inter-individual variability of the tertiary dentin that changes according to external stimuli could be responsible for the observed low

correlation with age, even if further experiments are needed to test this hypothesis.

Chapter 8

CONCLUSIONS

Epigenetic variability is a new and important mechanism for the study of microevolution, because it creates phenotypic diversity both within an individual and within population without genetic variation (Richards 2008). This mechanism constitutes an important reservoir for rapid adaptation in response to internal and external stimuli. This thesis explored how DNA methylation changes may influence human adaptive processes. Secondly this thesis paves the way for the application of DNA methylation for the study of hystorical remains.

This thesis demonstrated that DNA methylation contributes to natural human variation in Italy and that individuals from Italian peninsula are characterized by different epigenetic pathways. This is the first study that addressed the variability across Italy, to date only a recent study investigated the DNA methylation profile in Italy, focusing on Italian migrations (Campanella et al. 2014). Here we depicted the natural variations that occur in DNA methylation profiles considering individuals who born and lived in different areas of Italy. Italy is a very diverse country in term of culture, traditions, history and genetics (Boattini et al. 2013; Sarno et al. 2014) and for this reason it constitutes a good model to investigate epigenetics variations. We used a new pipeline of analysis in order to highlight the regions with a potential functional role. We reasoned that an oscillation of DNA methylation levels of single CpGs could be less important than the variations of many CpGs in functional elements (such as islands, shore or shelves).

The comparisons between DNA methylation patterns of individuals from North and Centre of Italy with individuals from South of Italy revealed an enrichment of DMRs (differentially methylated region) located in genes involved in nitrogen compund transport and nitrogen compund metabolic process. Reactive nitrogen

has been extensively studied and it is object of research for environmental issues. In fact human intervention has dramatically altered the aquatic and terrestrial ecosystems, as nitrogen oxides (NOx), that are by-products of fossil fuel combustion, have a big impact on them. Moreover gases containing nitrogen can also affect human respiratory health. It is estimated that the major sources of nitrogen compounds are linked to agriculture and to the ever-growing human population that leads to the increasing of sewage and industrial wastes (Galloway et al. 1995; Galloway et al. 2008; UNEP, United Nations Environment Programme 2007). The link between DNA methylation and pollutants is extensively described in literature (see review Breton and Marutani 2014) and even if these results constitute only a preliminar investigation, it is interesting to note that they reflect a different distribution of industries and of agricultural history typical of the Italian peninsula. It is well known that the North of Italy is characterized by a large amount of industries if compared to the south that is more dedicated to agricultural practices.

Considering all genes that included a DMRs, resulting from the pairwise comparison between individuals from North, South and Centre of Italy, we observed an enrichment of genes involved in pathogens response. This is a very interesting point also discussed in recent papers (Bierne, Hamon, and Cossart 2012) as pathogens exert a very strong influence on human genetic/epigenetic variation and some authors suggested that local pathogens diversity constitutes the main selective pressure through human evolution (Fumagalli et al. 2011). We can hypothesize that epigenetic variation could constitute a further level of variability important to counteract pathogens impact. A recent study published by Marr and colleagues (Marr et al. 2014) elucidated the possible role of DNA methylation in pathogens response and they demonstrated that intracellular parasite can alter host cell DNA methylation patterns resulting in altered gene expression possibly to establish diseases. Other studies identified *Helicobacter*

pylori-associated DNA hypermethylation in gastric mucosa at specific genes, such as E-cadherine gene *CDH1* (Chan et al. 2003) and DNA repair genes, *MLH1*(Yao et al. 2006). Moreover it is also known by in-vitro studies that bacteria affect genes involved in cell proliferation. An example is uropathogenic *Escherichia Coli* that, in uroepithelial cells, can induce down-regulation of *CDKN2A* (a cell cycle inhibitor). This is supposed to increase cell proliferation and pathogens persistence (Tolg et al. 2011). However, the exact mechanism by which DNA methylation contributes to pathogens response remains unclear and we do not have enough data to understand whether DNA methylation variability constitutes a way to adapt to pathogens infection (in order to maintain the survival of the organism) or if DNA methylation variability is a way for the pathogens to maintain their survival.

Considering also DNA methylation of individuals of European, Asian and African ancestry we observed an enrichment of DMRs located in genes involved in immune response. The p-values are in general lower if compared to those from the Italian study, this is likely due to the very significant variations in genetic and cultural background of the populations considered. The DMRs comparison between individuals of African ancestry and individuals of Asian ancestry showed an enrichment of genes involved in interferon gamma mediated signalling pathway and in the component of MHC protein. It also showed an increase in genes involved in UDP glucoronosyltransferase, that are enzymes involved in **glucoronidation,** foundamental for detoxification processes. These results confirmed the data of a previous study by Heyn and colleagues (Heyn et al. 2013). They suggested that selection pressure not only shapes the genetic code, but also DNA methylation profiles, as confirmed also by our pipeline.

Moreover, we observed that DMRs are located in genes involved in cell stability, such as genes of the cell adhesion, cell comunication and cell growth. For example, *PACS2* gene (AFR-EUR DMRs) is involved in cell communication between endoplasmatic reticulum and mitochondria and in mechanisms of cell death, *MAD1L1* gene is a component of the mitotic spindle-assembly checkpoint associated with chromosomal stability. It is also a well known gene since it is located in one of the fastest-evolving regions of the human genome (HAR3) and one of the reason for the rapid increase of substitution rate in the human lineage seems to be the big advantage leaded by accumulation of mutations in this gene. In fact it is likely that the mutations of the human lineage lead to the reduction of the error rate during cell division. A dysfunction of this gene is linked to tumorigenesis of various types of human cancer (Tsukasaki et al. 2001) and DNA hypermethylation of the promoter of this gene is linked to cell cycle dysregulation and progression of the ovarian cancer. This data constitutes an important link between genes important for human evolution and genes involved in modern diseases. DNA methylation is the mechanism that could contribute to the evolution of disease vulnerability as other authors suggested (Zeng et al. 2012).

Moreover these pathways remarked the importance of **microenvironmen**t. In fact not only external environment could constitute a selective force that acts on the organism, but also microenvironment could constitute a selective pressures on cells fitness. This theory is particularly important in tumorigenesis (Anderson et al. 2006) but could be applied also in physiological states were some clones of cells, with a common epigenetic profile, could be more apt to cope with the microenvironment (microenvironment here refers to the sum of mediators, molecules, compound, proteins of the extracellular matrix, hormones, chemicals, signalling molecules as well as physical interaction).

Noteworthy, we can observe that gene enrichment analysis considering macro-geographic groups (AFR, EAS and EUR) highlighted some pathways that emerged also in the study of Voight and colleagues (2006). The authors reported p-values for enrichment of GO Categories among genes showing evidence for partial sweeps. Here we noted some pathways also reported to be differentially methylated between populations included in our study, such as MHC-I mediated immunity. This supports the model of Klironomos (Klironomos, Berg, and Collins 2013) in which early adaptation via epigenetic mutations is a reduction of selective constrains on genetic variables while selection pressure acts on the epigenetic variables. Thus, it could be that the selective advantage of the favored allele is weakened because of epigenetic variability can generate a further level of phenotypic variation, reducing the strength of selection on genetic varints (Klironomos, Berg, and Collins 2013; Schmitz et al. 2011).

Regarding the DNA methylation analysis of *KRTCAP3* (keratinocyte associated protein 3), *MAD1L1* gene (component of the mitotic spindle-assembly checkpoint), and *BRSK2* gene (Serine/threonine-protein kinase that plays a key role in polarization of neurons and axonogenesis, cell cycle progress and insulin secretion), we demonstrated that the differences between populations regard the entire amplicon and not only one CpG sites. Moreover, by comparing Americans and Italian individuals of different ancestry, we observed that the DNA methylation patterns are also similar, indicating that ancestry shapes DNA methylation profiles. Noteworthy, DNA methylation patterns of *BRSK2* gene have shown significant differences considering individuals coming from the same macro-geographic area (Maroccans and Nigerians). In particular, the discriminant analysis of principal components shows that individuals from Morocco grouped with Italians while individuals from Nigeria clustered with Asians. This analysis have not been addressed in the study of Heyn and colleagues (2013) because there are no available data regarding the origin of African Americans individuals.

As described in Chapter 6, the more plausible assumption considers the history of Morocco and the European influence exerted on this country. This influence might lead to similar patterns of DNA methylation between individuals from Italy and individuals from Morocco. Analyzing the genetic diversity of these regions (considering 1000Genomes data) we identified a population genetic structure that for *KRTCAP3* and *MAD1L1* could reflect the DNA methylation differences. For *BRSK2* the genetic structure seems to be different. In fact in the Italian cohort, individuals from Nigeria and from Asia share similar values of DNA methylation, while the genetic structure clearly showed three main clusters with Asian population, African individuals and European individuals.

This thesis explored also the relationship between DNA methylation patterns and age both in blood and in human teeth, one of the most used specimens for anthropological studies. In particular, we focused on biomarkers of age to identify genes whose DNA methylation levels could be strongly correlated with individual age. This is particular interesting for forensic anthropology to achive age estimation from blood. ELOVL2, FHL2 and PENK genes, considered in this study, constitute a good model for the use in forensics and for the age estimation on blood traces. This striking correlation with age was also observed in DNA from modern teeth. In particular cementum and dental pulp were the two area where DNA methylation values showed an high correlation with age. DNA methylation of 8 CpGs of two genes (ELOVL2 and FHL2) could be used to predict age with the highest accuracy when both dental pulp and cementum are available. Indeed, median absolute difference between age estimated from DNA methylation and chronological age was 1.20 years and the r2 of the linear model was 0.97. Despite the models were constructed with a modest number of individuals (N=21), these data indicated that DNA methylation from cementum and pulp are the best way to predict age from teeth and could be applied extensively for the estimation of the age. Future studies are needed to test DNA

methylation from historical samples but it seems to be very promising since a recent study showed how DNA methylation could be used to estimate age of a 4000-yr-old Paleo-Eskimo (Pedersen et al. 2014).

In conclusion the data presented constitute a comprehensive view of natural DNA methylation variations across the Italian peninsula and also at macroeographical level, considering continental differences. The genes and the pathways here reported constitute potential mechanisms for human adaptation in different conditions and demonstrated the major forces that are shaping DNA methylation profiles both at a micro- and macro-geographical level.

APPENDIX A

Most relevant publications during the Ph.D course

1: Bacalini MG, Boattini A, Gentilini D, Giampieri E, Pirazzini C, **Giuliani C**, Fontanesi E, Remondini D, Capri M, Del Rio A, Luiselli D, Vitale G, Mari D, Castellani G, Di Blasio AM, Salvioli S, Franceschi C and Garagnani P. *A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data*. Aging (Albany NY). Jan 2015, Advance online publications

2: **Giuliani C**, Bacalini MG, Sazzini M, Pirazzini C, Franceschi C, Garagnani P, Luiselli D. *The epigenetic side of human adaptation: hypotheses, evidences and theories*. Ann Hum Biol. 2015 Jan;42(1):1-9. PubMed PMID: 25413580.

3: Bacalini MG, Friso S, Olivieri F, Pirazzini C, **Giuliani C**, Capri M, Santoro A, Franceschi C, Garagnani P. *Present and future of anti-ageing epigenetic diets*. Mech Ageing Dev. 2014 Mar-Apr;136-137:101-15. Review. PubMed PMID: 24388875.

4: Garagnani P, Bacalini MG, Pirazzini C, Gori D, **Giuliani C**, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S, Rezzi S, Castellani G, Capri M, Salvioli S, Franceschi C. *Methylation of ELOVL2 gene as a new epigenetic marker of age*. Aging Cell. 2012 Dec;11(6):1132-4. PubMed PMID: 23061750.

5: Pirazzini C*, **Giuliani C*,** Bacalini MG, Boattini A, Capri M, Fontanesi E, Marasco E, Mantovani V, Pierini M, Pini E, Luiselli D, Franceschi C, Garagnani P. *Space/population and time/age in DNA methylation variability in humans: a study on IGF2/H19 locus in different Italian populations and in mono- and di-zygotic twins of different age*. Aging (Albany NY). 2012 Jul;4(7):509-20. PubMed PMID: 22879348.

During my PhD programme I have deepened some methodological and theoric issues, such as the study of mitochondrial genomes and in particular of mtDNA heteroplasmy. From a methodological point of view I have deepened the libraries preparation for sequencing a 853 bp segment of mtDNA by ultra-deep sequencing. The aim here was to detect heteroplasmies in blood under a threshold of 2%, reaching an average coverage of 49334-fold per position. As follows I will list publications that better synthetized the mitochondrial questions that I explored in this three years of my PhD programme.

1: **Giuliani C**, Barbieri C, Li M, Bucci L, Monti D, Passarino G, Luiselli D, Franceschi C, Stoneking M, Garagnani P. *Transmission from centenarians to their offspring of mtDNA heteroplasmy revealed by ultra-deep sequencing.* Aging (Albany NY). 2014 Jun;6(6):454-67. PubMed PMID: 25013208

2: Sevini F, **Giuliani C**, Vianello D, Giampieri E, Santoro A, Biondi F, Garagnani P, Passarino G, Luiselli D, Capri M, Franceschi C, Salvioli S. *mtDNA mutations in human aging and longevity: controversies and new perspectives opened by high-throughput technologies*. Exp Gerontol. 2014 Aug;56:234-44. Review. PubMed PMID: 24709341.

Research Paper

A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data

Maria Giulia Bacalini[1,2,3,*], Alessio Boattini[4,*], Davide Gentilini[5], Enrico Giampieri[6], Chiara Pirazzini[1,2], Cristina Giuliani[4], Elisa Fontanesi[1,2], Daniel Remondini[6], Miriam Capri[1,2], Alberto Del Rio[1,7], Donata Luiselli[4], Giovanni Vitale[5,8], Daniela Mari[8,9], Gastone Castellani[6], Anna Maria Di Blasio[5], Stefano Salvioli[1,2], Claudio Franceschi[1,2,10], and Paolo Garagnani[1,2,11]

[1]*Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna 40138, Italy;*
[2]*Interdepartmental Center "L. Galvani", University of Bologna, Bologna 40126, Italy;*
[3]*Personal Genomics S.r.l., Verona 37134, Italy;*
[4]*Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna 40126, Italy;*
[5]*Centro di Ricerche e Tecnologie Biomediche, Istituto Auxologico Italiano IRCCS, Cusano Milanino 20095, Italy;*
[6]*Department of Physics and Astronomy, University of Bologna, Bologna 40126, Italy;*
[7]*Institute of Organic Synthesis and Photoreactivity (ISOF) National Research Council (CNR), Bologna 40126, Italy;*
[8]*Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy;*
[9]*Geriatric Unit, IRCCS Ca' Granda Foundation Maggiore Policlinico Hospital, Milan, Italy;*
[10]*IRCCS Institute of Neurological Sciences, Bologna, Italy;*
[11]*Applied Biomedical Research Center, S. Orsola-Malpighi Polyclinic, Bologna 40138, Italy.*
**Equal contribution*

Key words: *DNA methylation, Infinium HumanMethylation450 BeadChip, Epigenetics, aging*
Received: *12/23/14; Accepted: 01/09/15; Published: 01/11/15*
Correspondence to: *Claudio Franceschi, PhD;* ***E-mail:*** *claudio.franceschi@unibo.it*

Abstract: Aging is characterized by a profound remodeling of the epigenetic architecture in terms of DNA methylation patterns. To date the most effective tool to study genome wide DNA methylation changes is Infinium HumanMethylation450 BeadChip (Infinium 450k). Despite the wealth of tools for Infinium 450k analysis, the identification of the most biologically relevant DNA methylation changes is still challenging. Here we propose an analytical pipeline to select differentially methylated regions (DMRs), tailored on microarray architecture, which is highly effective in highlighting biologically relevant results. The pipeline groups microarray probes on the basis of their localization respect to CpG islands and genic sequences and, depending on probes density, identifies DMRs through a single-probe or a region-centric approach that considers the concomitant variation of multiple adjacent CpG probes. We successfully applied this analytical pipeline on 3 independent Infinium 450k datasets that investigated age-associated changes in blood DNA methylation. We provide a consensus list of genes that systematically vary in DNA methylation levels from 0 to 100 years and that have a potentially relevant role in the aging process.

INTRODUCTION

In the last two years the Infinium HumanMethylation-450 BeadChip (Infinium 450k) [1] has been largely used to investigate age-associated changes in DNA methylation profile of the human genome [2–9]. The Infinium 450k contains 485577 probes, 64% of them mapping to CpG islands and CpG islands surrounding

regions (shores and shelves), while the remaining mapping to dispersed CpG sites in the genome [10]. The array is highly informative, as it covers 96% of the CpG islands of the genome and 99% of RefSeq genes.

Using the Infinium 450k, researchers have identified several CpG sites that either get hypermethylated or hypomethylated during aging in different tissues [11], and a subset of these CpG sites has been successfully combined in predictors of chronological age [12,13].

Although defining a list of CpG sites whose methylation status is age-dependent is an essential step in aging research, the real challenge is to identify biologically relevant DNA methylation changes and their relative contribution to the aging process.

The difficult task of extracting relevant information from microarray data can be made easier if the number of microarray features is reduced on the basis of a biologically meaningful criterion. In this way the top ranking groups of features are more likely to be functionally linked to the phenotype under study than the single features. For expression microarrays this task has been successfully addressed by grouping genes that share common biological functions [14], but this approach is less suitable for methylation microarrays, as the relationship between DNA methylation and biological function is complex. An alternative solution is to adopt a region-centric approach in which the methylation value not of the single CpG probes, but of a group of adjacent CpG probes is considered. This approach is particularly interesting as changes in DNA methylation, especially in the CpG islands, usually involve groups of adjacent CpG sites whose methylation levels are correlated, thus potentially affecting chromatin structure. On the contrary the biological relevance of alterations at individual CpGs, although potentially interesting at specific genomic regions, is less characterized [15].

At present, different region-centric approaches have been proposed. Illumina Methylation Analyzer (IMA) defines for each gene 11 region categories (TSS1500, TSS200, 5' UTR, 1st EXON, GENE BODY, 3' UTR, ILSAND, NSHORE, SSHORE, NSHELF, SSHELF) and calculates their mean or median methylation values, which are then compared between the samples under analysis [16]. As an alternative approach, Numerical Identification of Methylation Biomarker Lists (NIMBL) reports the number of differentially methylated probes within the different annotated regions of each gene [15]. A more sophisticated approach is based on the "bump hunting" method developed for the analysis of CHARM data [17], but its applicability to Infinium 450k data is weakened by the lower density of analysed CpG sites in comparison to CHARM array.

The main point is that the Infinium 450k probes are not evenly distributed across the genome, but they are enriched in specific regions while others are underrepresented. To deal with this issue, a methodology in which differentially methylated regions (DMRs) are defined as regions in which at least two contiguous probes within 1-kb distance have a significant differential statistic was recently proposed [18] and used in a meta-analysis to identify age associated DMRs. A tool for DMRs identification at a region level was implemented also in the RnBeads package, where regions were defined based on the microarray annotation and ranked based on 3 criteria (mean of the differences between average methylation levels of the probes in a region in the two groups under investigation, mean of quotients and a combined p-value calculated from the single p-values of the probes in the region).

Here we implemented an alternative pipeline for the analysis of Infinium 450k data that is based on a careful description of CpG probes distribution within the array. The proposed methodology: 1) classifies CpG probes based on their genomic localization 2) defines groups of adjacent CpG probes based on their density in the region 3) depending on the previous classifications, applies a single-probe or a region centric analysis which considers the concomitant variation of a group of adjacent CpG probes.

As a proof of principle, we used our approach to conduct a meta-analysis on 3 independent datasets in which the Infinium 450k was used to investigate age-associated variations in blood DNA methylation profiles. From this meta-analysis, we extracted a short list of genes that potentially have a biologically relevant role in the aging process.

RESULTS

Grouping Infinium 450k probes in biologically meaningful clusters

The method we propose focuses on grouping CpG probes into clusters, hereafter referred as "blocks of probes" (BOPs). CpG probes were grouped taking into consideration not only their contiguity in DNA sequence, but also their genomic localization, which represents a critical aspect for data interpretation [19]. Using Illumina probe annotation, we first divided the probes included in the array in four classes (Fig. 1A; see Materials and Methods section): *i)* Class A, including probes in CpG islands and CpG islands-surrounding

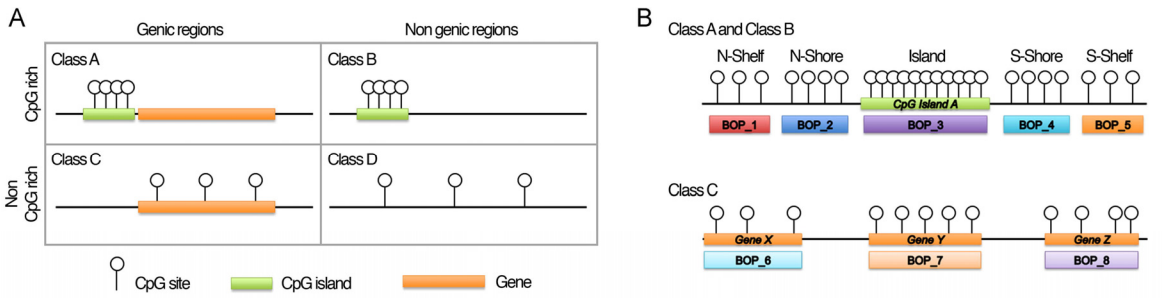


Figure 1. Infinium 450k probes classification and BOPs definition. (**A**) The 485577 probes included in the Illumina HumanMethylation450 BeadChip were divided in 4 classes on the basis of their genomic localization. (**B**) Graphic representation of how probes were grouped in BOPs. Probes mapping in the island and in the surrounding regions of the same CpG island were grouped in 5 functional units: probes in the N-Shelf of the island, probes in the N-Shore of the island, probes in the island, probes in the S-Shore of the island, probes in the S-Shelf of the island. Probes mapping in gene bodies were grouped on the basis of the gene in which they are located.

sequences (shores and shelves) that map in genic regions; *ii)* Class B, including probes in CpG islands and CpG islands-surrounding regions (shores and shelves) which do not map in genic regions; *iii)* Class C, including probes in genic regions which are not CpG rich; *iv)* Class D, including probes in non-genic regions which are not CpG rich. These four classes have different epigenetic functions, as their methylation status can affect gene function and chromatin structure in different ways [20]. Class A, B, C and D included 247394, 62071, 118206 and 57906 CpG probes respectively.

Then, we defined BOPs as follows (Fig. 1B): for probes mapping in CpG islands and in CpG islands surrounding regions (Class A and Class B), we grouped the CpG probes localized in the same island, in the same shore or in the same shelf; for probes mapping in not CpG rich genic regions (Class C), we grouped the CpG probes mapping to the same gene. Class D probes were not grouped because they were highly interspersed across the genome.

Class A, B and C included 77202, 34448 and 20273 BOPs respectively (Fig. 2). Class A BOPs mapped mainly to CpG islands and shores, but CpG islands BOPs were definitely richer in CpG probes (Fig. 2A, left panel). Class B BOPs mapped to shelves with higher frequency than Class A BOPs, but also in this case CpG probes mapped mostly to CpG islands (Fig. 2A, middle panel). As expected Class C BOPs contained a very high number of CpG probes, with a median of 13 and a maximum of 506 (Fig. 2A, right panel).

A region-centric approach is meaningful only if CpG probes are sufficiently close along the DNA sequence under investigation. Indeed, experimental evidences indicate that DNA methylation of nearby CpG sites is correlated within a tract of 250-500 bp [21]. We therefore calculated the mean distance between the probes in each BOP for Class A, Class B and Class C probes (Fig. 2B). Probes in BOPs belonging to Class A and Class B were generally close, with a mean distance of 183.5 and 259.4 bp respectively (Fig. 2B, left and middle panels) and a mode of 73 and 130 bp respectively (Fig. 2C, left and middle panels). In both the cases, as expected, mean distance was lower in islands than in shores and shelves. On the contrary, BOPs belonging to Class C included probes that were scattered across the length of gene sequences and, on average, that were too distant to be analysed together (mean distance 10230 bp, mode 1935 bp; Fig. 2B and 2C, right panels).

DMR identification by multivariate analysis of variance (MANOVA)

Based on the previous observations we propose an analysis pipeline for Infinium 450k "customized" on the characteristics of the different classes of probes (Fig. 3A). BOPs belonging to Class A and Class B are suitable for a region-centric analysis, while for Class C and Class D probes a single-probe analysis is more advisable.

A

	Class A						Class B						Class C
	Total	CpG Islands	N-Shores	S-Shores	N-Shelves	S-Shelves	Total	CpG Islands	N-Shores	S-Shores	N-Shelves	S-Shelves	Total
BOPs	77202	20831	19281	17298	10185	9607	34448	6272	6586	5516	8300	7774	20273
CpG probes	247394	127757	50423	40475	15318	13421	62071	22497	12447	8722	9526	8879	118206
Min	1	1	1	1	1	1	1	1	1	1	1	1	1
1st Quartile	1	5	2	2	1	1	1	3	1	1	1	1	7
Median	2	8	4	3	1	1	1	4	2	2	1	1	13
Mean	3.205	9.203	4.918	4.509	3.341	2.808	1.802	5.117	3.301	2.741	1.591	1.556	28.95
3rd Quartile	4	11	6	6	4	3	2	5	4	4	1	1	30
Max	75	75	44	46	30	32	41	41	38	32	26	23	506

B

	Class A						Class B						Class C
	Total	CpG Islands	N-Shores	S-Shores	N-Shelves	S-Shelves	Total	CpG Islands	N-Shores	S-Shores	N-Shelves	S-Shelves	Total
Min	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.5	3.0	2.0	2.0
1st Quartile	77.0	70.0	104.0	93.0	111.1	110.4	105.8	92.5	157.4	128.5	139.0	112.0	823.5
Median	126.0	106.2	189.7	166.7	199.2	198.2	176.5	144.1	275.7	211.8	243.0	207.5	3644.0
Mean	183.5	127.3	281.1	237.9	297.7	287.7	259.4	179.8	399.8	310.8	375.3	310.9	10230.0
3rd Quartile	212.2	161.3	340.2	290.9	341.0	343.9	294.0	221.0	489.5	376.0	476.0	390.6	10520.0
Max	2972.0	2972.0	1972.0	1967.0	1960.0	1921.0	14780.0	14780.0	1943.0	1951.0	1891.0	1881.0	2494000.0

C

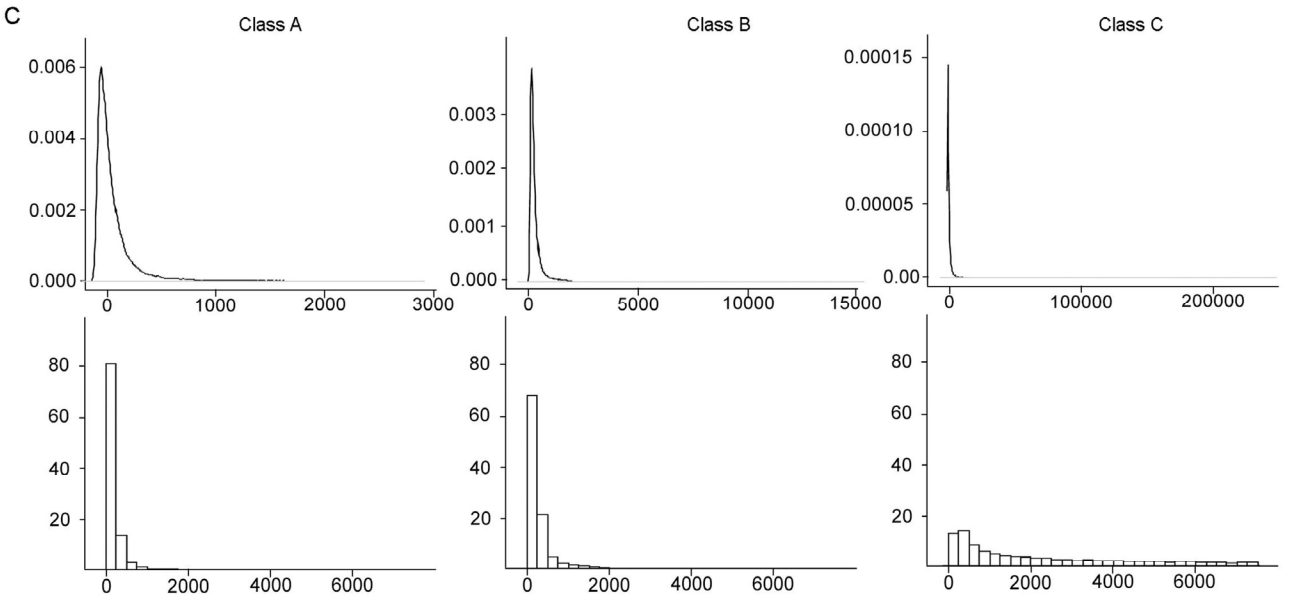


Figure 2. Characteristics of the BOPs belonging to different probe classes. (**A**) Numbers of BOPs and CpG probes in Class A, Class B and Class C. For Class A and Class B, subdivision in CpG islands, N-Shores, S-Shores, N-Shelves and S-Shelves is reported. In the lower part of the tables, descriptive statistics for the distribution of number of probes/BOP in the 3 Classes are reported. (**B**) Descriptive statistics for the distribution of mean bp distance between probes /BOP in the 3 Classes are reported.(**C**) Density distributions (upper panel) and frequency histograms (lower panels) of the mean bp distance between the probes/BOP.

For the region-centric analysis we propose the use of multivariate analysis of variance (MANOVA) to test for general changes in methylation of a genomic region. However, as described above, Class A and Class B BOPs contain a largely variable number of CpGs, spanning from 1 to 75 for Class A BOPs and from 1 to 41 for Class B BOPs. We considered BOPs containing 1 or 2 CpG probes not informative enough for a region-centric approach but more suitable for a single-probe approach. On the contrary, the remaining BOPs (32356 BOPs including 192465 probes in Class A; 7253 BOPs including 31077 probes in Class B) are analysed by applying a sliding-window MANOVA (Fig. 3B). Indeed, the non-homogeneous distribution in BOP probe content could lead to overrepresentation of short BOPs among the top significant BOPs, because when the number of probes in a BOPs is high, it is likely that only a subgroup of them is differentially methylated in the phenotype under study (see the example reported in Fig. 3C). In the sliding window approach we calculated MANOVA for each subgroup of 3 consecutive CpG within the same BOP and we kept the lowest p-value among those calculated, actually normalizing the analysis for the varying number of probes per BOP.

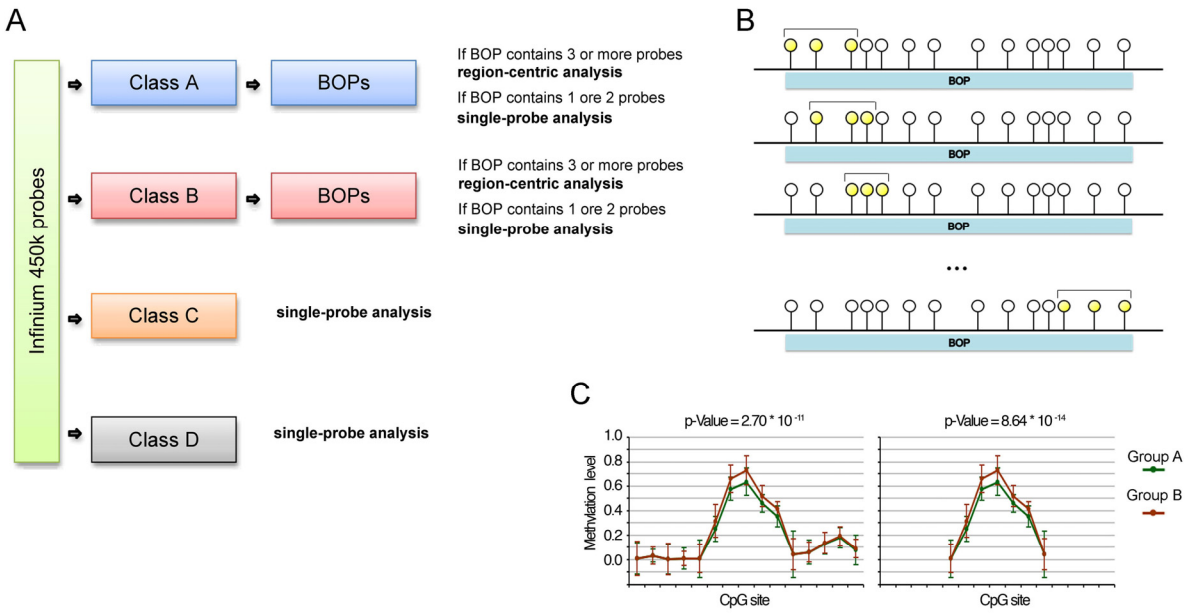


Figure 3. Proposed analytical pipeline for Infinium 450k data. (**A**) Workflow for the use of single-probe or region-centric approaches on Infinium 450k data. (**B**) Graphical representation of the sliding window MANOVA used to normalize for BOPs lengths. CpG probes are represented as circles. The CpG probes considered in each round of MANOVA are highlighted in yellow. (**C**) Example of methylation values of CpG probes within a BOP. The BOP includes 22 CpG probes, 5 of which define a "bubble" of differential methylation between Group A and Group B. The p-value derived from MANOVA on this BOP is 2.70*10e-11. We hypothesized to have a shorter BOP including only the 5 CpG probes differentially methylated between Group A and Group B, plus a probe on both the sites whose methylation level is comparable between the two samples. In this case, although the extent of the bubble of differentially methylation is the same of the longer BOP, the p-value derived from MANOVA is lower, equal to 8.64*10e-14. This simple example shows that if we do not normalize for the length of the BOP, short BOPs tend to rank at higher positions than long BOPs.

We implemented the proposed pipeline in R software environment. Starting from a table containing the beta-values of interest (1 row per CpG site, 1 column per individual), the script annotates the probes and divides them in Class A, Class B, Class B and Class D. Class A and Class B probes are grouped in BOPs and BOPs containing 3 or more probes are analysed by the region-approach (MANOVA on sliding windows of 3 CpG probes within the same BOP). The remaining CpG probes are analysed by ANOVA. Both categorical and continuous covariates can be used. A list of significant CpG probes is outputted from the single-probe analysis. For the region-centric approach, a list of BOPs ranked by nominal or FDR-corrected p-values and, if present, associated genes, is provided as output. Then, authors can choose to select significant BOPs containing at least 2 adjacent CpG sites for which the DNA methylation difference between the considered groups (in the case of a categorical variable) or two selected ranges of values of the continuous variable is higher than a set threshold. Significant BOPs can be ranked according to this criterion allowing authors to identify "bubbles" of different methylation between the conditions under study, which are likely to be biologically meaningful. The pipeline provides as outputs several useful plots, including schematic diagrams in which the chromosomal position of the probes within the selected BOPs is plotted against the methylation level of each probe (see the examples reported in Fig. 6 and in Supplementary Fig. 1). These plots provide a easy-to-interpret visualization of relevant results, that in one shot gives information about probes density within the region and about the changes in DNA methylation between the groups under study.

Identification of age-associated DMR through the region-centric approach

We validated our pipeline on three independent age-related Infinium 450k experiments performed on whole

blood. The first dataset (referred hereafter as D1) includes data from 656 subjects ranging from 19 to 101 years (average 64 ± 15 years) [3]. The second dataset (referred hereafter as D2) includes data from 38 subjects, 19 newborns and 19 nonagenarians [2]. The third dataset (referred hereafter as D3) includes data from 58 subjects ranging from 9 to 83 years (average 44 ± 18 years) [4].

In each dataset, probes containing missing values in at least one sample and probes on X and Y chromosome were removed. Age-associated DMR were identified in D1 and D3 using the age as a continuous variable and in D2 the group (newborns or nonagenarians) as a categorical variable. Ethnicity (Caucasian-European or Hispanic-Mexican) was used as covariate in the D1 analysis. As the relative proportions of the different types of blood cells can vary significantly with age, we inferred blood cell counts from methylation data and use them as covariates in the analysis.

Here we present results only for the region-centric analysis performed on Class A BOPs containing at least 3 CpG probes. MANOVA results were corrected for multiple comparisons through Benjamini-Hochberg False Discovery Rate correction; 0.05 was used as significant threshold for q-values. Based on these criteria 21083, 517 and 2736 BOPs were identified as age-associated DMRs in D1, D2 and D3 respectively.

The considerably lower number of BOPs identified in D2 is ascribable to the effect of blood cell types correction, as cord blood and whole blood significantly differ for this parameter.

To validate our approach, we compared the results of the region-centric approach with the results of a single-probe approach by analysis of variance (ANOVA). As shown in Figure 4, the number of significant CpG probes per BOP tended to be higher in the top ranking BOPs identified by the region-centric analysis. In most the cases, significant CpG probes within the same BOP concordantly moved towards hypermethylation or hypomethylation in younger compared to older subjects. Notably a small number of the selected BOPs in the three datasets included slightly differentially methylated CpG probes that did not reach the significance threshold by themselves, but whose concomitant variation with adjacent probes within the same BOP resulted significant when analysed by a multivariate approach. These results indicate that our approach was successful in identifying chromosomal regions, rather than single CpG sites, whose methylation status is affected by aging.

To confirm that our region-centric approach reduces spurious results and is more likely to identify biologically relevant regions, we compared the results from the three datasets.

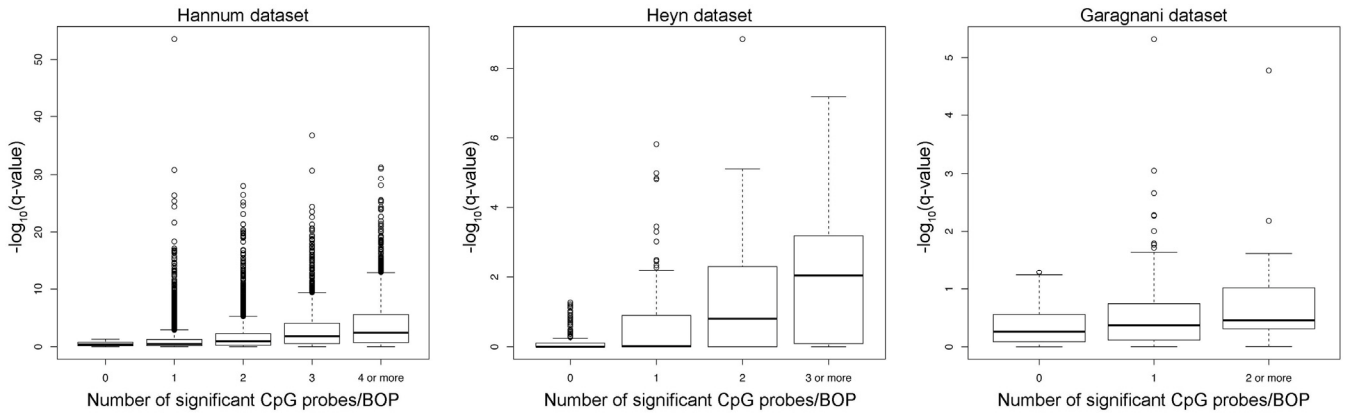


Figure 4. Number of significant CpG probes per significant BOP. For each dataset, the boxplot reports the $-\log_{10}$(q-value) of each significant Class A BOP (MANOVA analysis) against the number of significant CpG probes (q-value < 0.05, ANOVA analysis) included in each BOP.

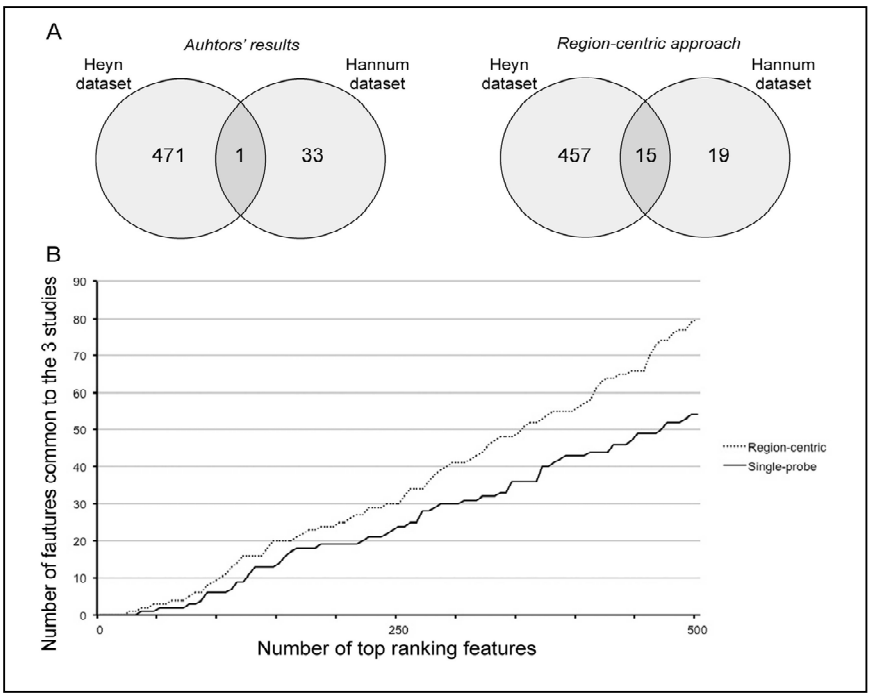


Figure 5. The region-centric approach increases the common findings between the 3 datasets. (**A**) Intersection between the results provided by Hannum et al. and Heyn et al. (left panel) and between the results of the region-centric approach on the two datasets. (**B**) Intersection between a progressively increasing number of top ranking features (BOPs for the region-centric analysis, CpG probes for the single-probe analysis) in the three datasets.

First, we compared the DMRs identified in Hannum et al. and in Heyn et al. with those identified by our approach. Hannum et al. used a multivariate linear model approach based on the Elastic Net algorithm and identified a nucleus of 89 CpG probes whose methylation is associated to age, while Heyn et al selected 3205 age-associated CpG probes that resulted significant after ANOVA test (q-value < 0.01) and that showed a difference in average beta-values between newborns and nonagenarians greater than 0.20. We considered only the CpG probes belonging to Class A probes, that is 45 and 800 probes in D1 and D2 respectively, and we matched them with the corresponding BOPs. In this way, we achieved a list of 34 and 472 BOPs identified by Hannum et al. and by Heyn et al.. Only one BOP was shared by the two lists (Fig 5A, left panel). On the contrary, when we considered the first 34 and the first 472 BOPs identified by our approach respectively in D1 and D2, we observed an overlap of 15 BOPs (Fig 5A, right panel).

Secondly, we considered a progressively increasing number of significant BOPs (region-centric analysis) and CpG probes (single-probe analysis) and we calculated the number of common DMRs between the three datasets. As shown in Figure 5B, the extent of overlap was higher for the region- centric approach compared to the single-probe approach.

Finally, we provided a short list of genomic regions whose methylation levels vary according to age in D1, D2 and D3 (Table 1). We considered the overlap between the first 500 ranked BOPs identified by our region-centric approach in the 3 datasets, resulting in 42 BOPs (44 genes, because some BOPs mapped to multiple genes which share the same CpG island). 11 and 2 of the selected BOPs included at least one CpG probe that was provided also in Hannum's and Heyn's results respectively. Moreover, 4 of the selected BOPs contained at least one CpG probe that was included also in Horvath's epigenetic clock. Two genes (HOXC4 and SST) were included in the GenAge database as related to aging in model systems and/or humans [22]. To have a general view of age-associated changes of the selected BOPs, we joined their beta-values from D1, D2 and D3 and we divided samples in 10 age ranges from 0 to 100 years. The plots reported in Fig 6 and in Supplementary Figure 1 confirm that the 42 genomic regions encounter a systematic hypermethylation (20 BOPs) or hypo-methylation (22 BOPs) with age.

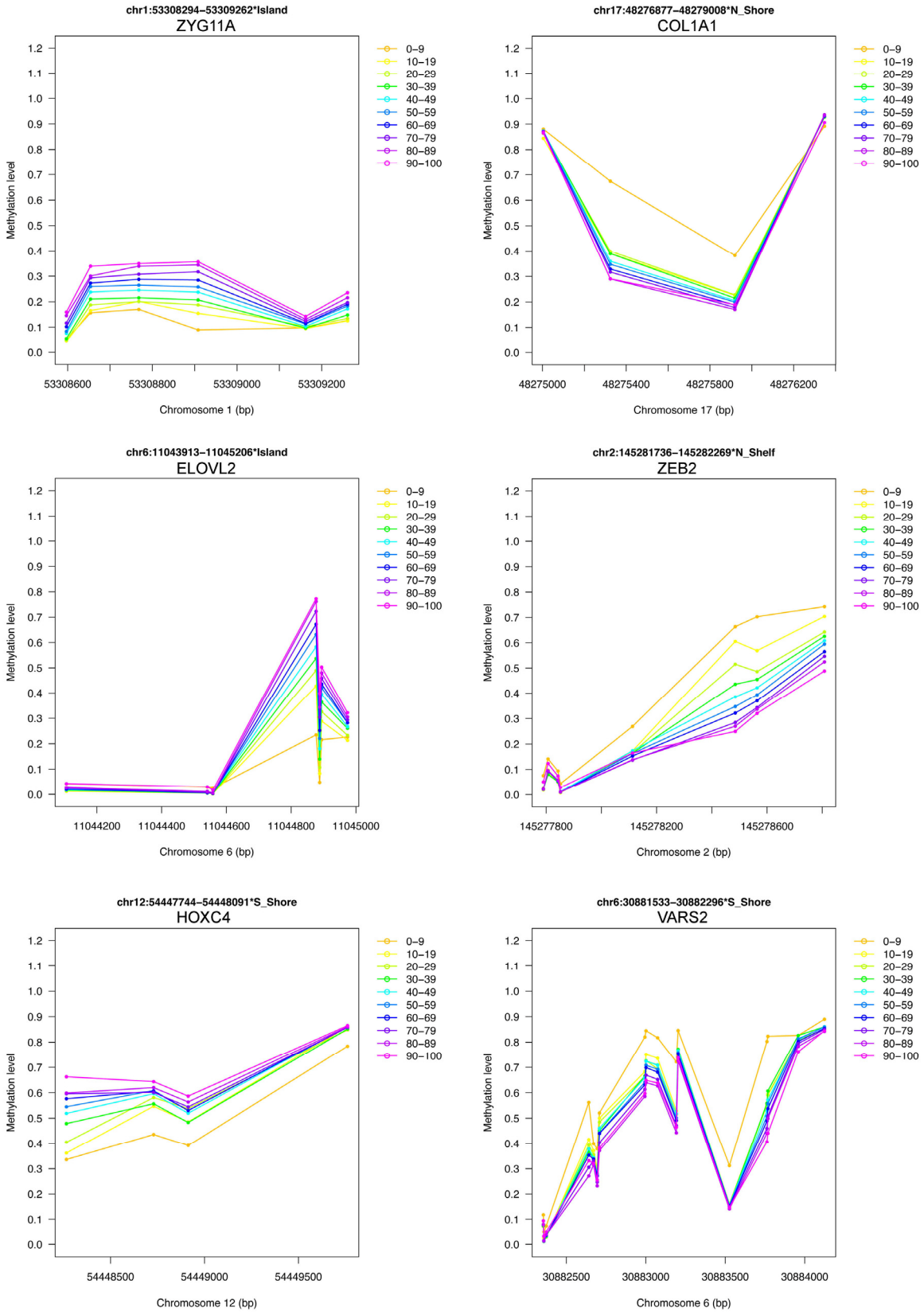


Figure 6. Examples of DNA methylation profiles of selected age-associated BOPs. 6 of the 42 selected BOPs are reported as an example. Mean methylation values in 10 age classes are reported for each CpG probe within the selected BOPs. For each BOP, beta-values from D1, D2 and D3 were joined together.

Table 1. Candidate age-associated genomic regions

Gene Name	Description	BOP	Hannum	Heyn	Horvath
ABCC4	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	chr13:95953337-95954211*N_Shore			
ABHD14A	abhydrolase domain containing 14A	chr3:52008943-52009339*N_Shore	X		X
ABHD14B	abhydrolase domain containing 14B	chr3:52008943-52009339*N_Shore	X		X
AKAP8L	A kinase (PRKA) anchor protein 8-like	chr19:15529290-15529902*S_Shore			
ALDOA	aldolase A, fructose-bisphosphate	chr16:30076310-30077872*N_Shore			
AMER3	APC membrane recruitment protein 3	chr2:131513363-131514183*Island	X		
ATP13A2	ATPase type 13A2	chr1:17337829-17338590*S_Shore	X		
AXL	AXL receptor tyrosine kinase	chr19:41769215-41769417*N_Shore	X		
CACNA1G	calcium channel, voltage-dependent, T type, alpha 1G subunit	chr17:48636103-48639279*Island			
COL1A1	collagen, type I, alpha 1	chr17:48276877-48279008*N_Shore			
CPEB1	cytoplasmic polyadenylation element binding protein 1	chr15:83315116-83317541*Island		X	
CSNK1D	casein kinase 1, delta	chr17:80231019-80231820*S_Shore			X
EDARADD	EDAR-associated death domain	chr1:236558459-236559336*N_Shore	X		X
EIF1	eukaryotic translation initiation factor 1	chr17:39844833-39845950*N_Shore	X		
ELOVL2	ELOVL fatty acid elongase 2	chr6:11043913-11045206*Island	X		
FHL2	four and a half LIM domains 2	chr2:106014878-106015884*Island	X	X	
GIT1	G protein-coupled receptor kinase interacting ArfGAP 1	chr17:27918161-27918398*N_Shore			
GLRA1	glycine receptor, alpha 1	chr5:151304226-151304824*Island			
GPR78	G protein-coupled receptor 78	chr4:8582036-8583364*Island			
GRIN2C	glutamate receptor, ionotropic, N-methyl D-aspartate 2C	chr17:72848166-72848901*Island			
GUSB	glucuronidase, beta	chr7:65446771-65447340*S_Shore			
HNRNPUL1	heterogeneous nuclear ribonucleoprotein U-like 1	chr19:41769215-41769417*N_Shore	X		
HOXC4	homeobox C4	chr12:54447744-54448091*S_Shore	X		
IRX5	iroquois homeobox 5	chr16:54962422-54967805*Island			
LAG3	lymphocyte-activation gene 3	chr12:6882855-6883184*N_Shore			X
MLXIPL	MLX interacting protein-like	chr7:73037528-73038957*Island			
NENF	neudesin neurotrophic factor	chr1:212606105-212606844*N_Shore			
NFIA	nuclear factor I/A	chr1:61548753-61549564*N_Shore			
OTUD7A	OTU deubiquitinase 7A	chr15:31775540-31776988*Island	X		
PI4KB	phosphatidylinositol 4-kinase, catalytic, beta	chr1:151300522-151300724*N_Shore			
PRLHR	prolactin releasing hormone receptor	chr10:120353692-120355821*Island			
PRRT4	proline-rich transmembrane protein 4	chr7:127990926-127992616*Island			
PTGDS	prostaglandin D2 synthase 21kDa (brain)	chr9:139872237-139873143*N_Shore			
PXN	paxillin	chr12:120702976-120703541*S_Shore			
RCSD1	RCSD domain containing 1	chr1:167599464-167599839*N_Shore			
SLC12A5	solute carrier family 12 (potassium/chloride transporter), member 5	chr20:44657463-44659243*Island			
SLC25A22	solute carrier family 25 (mitochondrial carrier: glutamate), member 22	chr11:797640-798544*N_Shore			
SOX1	SRY (sex determining region Y)-box 1	chr13:112720564-112723582*Island			
SST	somatostatin	chr3:187387914-187388176*N_Shore	X		
TFAP2B	transcription factor AP-2 beta (activating enhancer binding protein 2 beta)	chr6:50787286-50788091*Island			
VARS2	valyl-tRNA synthetase 2, mitochondrial	chr6:30881533-30882296*S_Shore			
ZAR1	zygote arrest 1	chr4:48492117-48493589*Island			
ZEB2	zinc finger E-box binding homeobox 2	chr2:145281736-145282269*N_Shelf			
ZYG11A	zyg-11 family member A, cell cycle regulator	chr1:53308294-53309262*Island			

DISCUSSION

In this paper we present an original pipeline for the analysis of Infinium 450k data, in which different genomic regions are analyzed either by a single-probe or a region-centric approach depending on their context and probe content/density.

The Infinium 450k is currently the most used technology for EWAS studies. The reasons for the success of this microarray are to be found in its affordability and simplicity, in addition to a reasonable informativeness on genome-wide DNA methylation profiles. Alternative techniques such as sequencing of immunoprecipitated methylated DNA (MeDIP-seq) or of bisulfite-treated genomic DNA (BS-seq) provide a more comprehensive picture of DNA methylome [23] but, despite the recent introduction of pipelines for their automatization and analysis [24–27], are still more laborious and expensive.

A recent comprehensive review has collected the bioinformatic tools developed to specifically analyze Infinium 450k data [27]. Apart from methods to normalize data and adjust them for technical bias, several algorithms to identify differentially methylated regions (DMRs) between groups of interest have been developed. A growing body of literature suggests to analyze Infinium 450k data by a region-centric approach, with several advantages respect to single-probe analysis [15,18]. First of all, it better resembles the biological basis of the process, as concordant changes in a group of adjacent CpG sites are more likely to affect the phenotype compared to alterations of single CpG sites. Moreover, a region-centric approach simplifies the identification of genomic regions of interest as provides a shorter list of ranked results and is less prone to provide spurious results due for example to the presence of SNPs in the probes included in the array [18].

Here we reasoned that the statistical analysis of the Infinium 450k microarray should take into account the specific architecture of the microarray. Indeed, we clearly showed that different genomic regions present different level of coverage in terms of probes distribution. We grouped probes in blocks (BOPs) based on microarray annotations and verified that only in CpG islands and in the surrounding regions (shores and shelves) the mean distance between probes in a BOP was within 500 bp, a range in which the methylation values of CpG sites are usually correlated [21]. On the contrary, the density of probes mapping in not CpG-rich regions (not CpG-rich promoters, gene bodies, intergenic regions) was strikingly lower. This means that different regions of the array are more suitable to a single-probe or a region-centric analysis. Noteworthy, the proposed probes classification does not have only a mere methodological value, but it has important biological implications. First of all, the region-centric approach is selectively applied to short regions (most CpG islands are within 3000 bp, while shores and shelves are by definition 2000 bp long), in which the methylation level of CpG probes is more likely to be correlated. Moreover, a growing body of evidences indicates that the function of DNA methylation greatly varies with genomic context [28]. This means for example that methylation of CpG islands, of shores or of gene bodies can differently affect gene expression and that it can be differently affected by the condition under study. Also, the methylation status of non-genic regions can have important consequences, for example by influencing chromatin architecture and stability, and it is likely that the effect of methylation at non-genic CpG islands or at open-sea CpG sites can be different. Analyzing separately these regions that have different functional meaning can therefore facilitate the identification of informative variations in methylation profiles in the model under consideration. Moreover, this approach provides shorter ranked lists of results that can be more easily examined by the researcher.

We used multivariate ANOVA (MANOVA) to test for general changes in methylation in the region-centric analysis. This approach has been previously adopted for the analysis of methylation data [29,30] as it allows to explore simultaneously the relationship between several dependent variables (in our case, 3 adjacent CpG probes within a BOP) and the independent variables under study. Notably, most of the other algorithms for DMRs identification is not based on a multivariate approach, but combines the results from univariate analyses on adjacent CpG probes [27]. As shown in the meta-analysis, MANOVA can identify not only regions in which multiple adjacent probes are significantly different between the samples, but also regions in which there are concomitant little variations of adjacent probes, none of which would reach the significance threshold by itself in a univariate test. This could be particularly valuable when small differences exist between the samples under study.

Overall the strength of our approach is that not only microarray features are grouped in biologically meaningful groups, but also that the ranking criterion is based on a multivariate approach. Additionally, significant BOPs can be ranked also on the basis of the "bubble" of differential methylation, defined as at least two adjacent probes whose mean methylation in the groups under investigation differs of at least a minimum value. As a confirm of the validity of the approach we

demonstrated that, when considering the top ranking BOPs, our method increased the number of common genes identified in all three studies compared to a single-probe analysis, indicating that it is likely to provide a more informative overview of biologically relevant results.

The analysis pipeline we propose is implemented in R software environment and is therefore freely available. Researchers just need to define the analysis parameters, such as the covariates to be used in the ANOVA/MANOVA, the FDR correction method, the significance threshold and the minimum difference between mean methylation values between the groups under investigation.

We used our analysis pipeline to perform a meta-analysis on 3 Infinium 450k datasets that investigated age-associated changes in DNA methylation. The 3 datasets are considerably different in terms of both samples number and age range. Nevertheless, we were able to identify a core of genomic regions whose methylation profiles systematically vary with aging, from newborns to nonagenarians. To our knowledge, this is the first report describing a relatively large number of genomic regions with such characteristics. Only a subset of these genes was identified by Hannum *et al.* and by Heyn *et al.* or was included in Horvath's epigenetic clock. Inferring how the methylation of these regions can contribute to the aging process is out of the scope of this paper, although the life-long variations in DNA methylation that we described are suggestive of a profound link between development and aging [31]. Many genes showed marked differences between cord blood and the other age ranges. Although these differences could in principle be ascribed to differences in blood cells composition, it is tempting to suggest that during the first phases of growth a profound epigenetic remodeling occurs. It is interesting to note that 2 genes from our list (SST and HOXC4) are enclosed as age-related genes in the GenAge database [22]. Moreover, a Pubmed search using the query "gene name AND aging [title/abstract]" gave some interesting hints, as the expression of CACNA1G, COL1A1, LAG3, PTGDS and ZEB2 genes resulted modulated by age in several models [32–38]. COL1A1 and PTGDS emerged from the same study [32] in which hippocampal gene expression in senescent female mice was assessed after long-term exercise. The observation that age-dependent expression of the above mentioned genes was detected in tissues other than blood prompts further studies to evaluate general rearrangements in epigenetic landscapes of different cell types. Finally, the methylation status of ELOVL2, FHL2 and EDARADD genomic regions was previously described as associated to aging [4,39].

Collectively the above observations indicate that this short list of genes, selected by means of an analytical pipeline that is tailored on the architecture of the microarray and that is more likely to provide biologically relevant findings, can be used as the basis for deeper investigations to shed light on the molecular basis of the aging process.

METHODS

Datasets. D1 and D2 are publicly available at NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE40279 and GSE30870 respectively. D3 includes 32 mother–offspring couples and is part of a larger datasets submitted to GEO with accession number GSE52588.

Estimation of cell counts. Abundance measures of blood cell types were estimated using the appropriate option of the DNA Methylation Age Calculator, freely available at https://dnamage.genetics.ucla.edu/ [12], which is in part based on a previously published algorithm [40]. As suggested by authors, methylation measures were corrected for "CD8.naive" (Naive CD8 T cells), "CD8pCD28nCD45RAn" (memory and effector T cells), "PlasmaBlast" (plasmablasts), CD4 T cells, monocytes, granulocytes and natural killer cells.

BOPs definition and DMR identification. The analytical pipeline is implemented as an R script freely available at https://immunologyomics.unibo.it/labkey/-450K_pipeline.url. For ease of use, the pipeline is split in 3 step: 1) Definition of Classes and BOPs; 2) DMRs identification 3) FDR correction, selection of significant DMRs/probes, sorting of probes and plots. Step 2 and step 3 are separately provided for Class A BOPs containing 3 or more probes, Class A BOPs containing 1 or 2 probes, Class B BOPs containing 3 or more probes, Class B BOPs containing 1 or 2 probes , Class C probes and Class D probes. Detailed explanations of input and output files are provided. A general description of the 3 analytical steps is provided below:

Step 1: Probes containing missing values and probes with a detection p-value greater than 0.05 in more than 75% samples were removed, together with those localized on sexual chromosomes. Probes that contained SNPs were annotated as previously described [21]. For probes classification and BOPs definition, the RELATION_TO_UCSC_CPG_ISLAND and the UCSC_REFGEN_NAME columns in the Illumina Infinium 450k annotation were used to subset the array probes in four classes and to group probes in BOPs.

Step 2: DMRs identification is based on MANOVA and ANOVA functions from the R package *car*. For sliding windows MANOVA, the function is applied on sliding windows of 3 consecutive CpGs within the same BOP. For each BOP, the lowest p-value among those calculated for the different sliding windows is kept. Both MANOVA and ANOVA analysis support the use of both categorical and continuous covariates. Parallel processing can be used if the computational environment supports it.

Step 3: The analytic pipeline allows to correct MANOVA/ANOVA p-values for multiple testing using the correction methods implemented in the R package *multtest*. BOPs/probes can be selected on the basis of a significance threshold on either nominal or FDR-corrected p-values. In the case of BOP analysis, BOPs can be ranked on the basis of a user-defined minimum mean methylation difference between adjacent CpG probes. MDS plots, heatmap plots and line plots are generated as outputs.

ACKNOWLEDGEMENTS

This work was supported by the European Union's Seventh Framework Programme (grant agreement no. 259679 "IDEAL", grant agreement no. 266486 "NU-AGE", grant agreement no. 305280), by CARISBO foundation and by the Italian Ministry of Health, Progetto Ricerca Finalizzata 2008, convenzione 35: "An integrated approach to identify functional, biochemical and genetic markers for diagnostic and prognostic purposes in the elderly, in the centenarians and in people with dementia, Alzheimer's disease, mild cognitive impairment".

Conflict of interest statement

The authors of this manuscript have no conflict of interests to declare.

REFERENCES

1. 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. High density DNA methylation array with single CpG site resolution. Genomics. 2011; 98:288–295.

2. 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, et al. Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci U S A. 2012; 109:10522–10527.

3. 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, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013; 49:359–367.

4. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S, Rezzi S, Castellani G, Capri M, et al. Methylation of ELOVL2 gene as a new epigenetic marker of age. Aging Cell. 2012; 11:1132–1134.

5. Johansson A, Enroth S, Gyllensten U. Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. PloS One. 2013; 8:e67378.

6. 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. 2013.

7. Raddatz G, Hagemann S, Aran D, Söhle J, Kulkarni PP, Kaderali L, Hellman A, Winnefeld M, Lyko F. Aging is associated with highly defined epigenetic changes in the human epidermis. Epigenetics Chromatin. 2013; 6:36.

8. Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ, Teschendorff AE, Butcher LM, Beck S, Müller M. Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. Age Dordr Neth. 2014; 36:9648.

9. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schönfels W, Ahrens M, Heits N, Bell JT, Tsai P-C, Spector TD, Deloukas P, Siebert R, Sipos B, et al. Obesity accelerates epigenetic aging of human liver. Proc Natl Acad Sci U S A. 2014; 111:15538–15543.

10. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics Off J DNA Methylation Soc. 2011; 6:692–702.

11. Bacalini MG, Friso S, Olivieri F, Pirazzini C, Giuliani C, Capri M, Santoro A, Franceschi C, Garagnani P. Present and future of anti-ageing epigenetic diets. Mech Ageing Dev. 2014; 136-137:101–115.

12. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013; 14:R115.

13. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jöckel K-H, Erbel R, Mühleisen TW, Zenke M, Brümmendorf TH, Wagner W. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol. 2014; 15:R24.

14. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005; 102:15545–15550.

15. Wessely F, Emes RD. Identification of DNA methylation biomarkers from Infinium arrays. Front Genet. 2012; 3:161.

16. Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, Johnson CS, Smiraglia DJ, Liu S. IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinforma Oxf Engl. 2012; 28:729–730.

17. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol. 2012; 41:200–209.

18. Ong M-L, Holbrook JD. Novel region discovery method for Infinium 450K DNA methylation data reveals changes associated with aging in muscle and neuronal pathways. Aging Cell. 2014; 13:142–155.

19. Price ME, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ, Robinson WP, Kobor MS. Additional annotation enhances potential for biologically-relevant analysis of the Illumina

Infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin. 2013; 6:4.
20. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010; 28:1057–1068.
21. Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, Zheng H, Yu J, Wu H, Sun J, Zhang H, Chen Q, Luo R, et al. The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol. 2010; 8:e1000533.
22. De Magalhães JP, Toussaint O. GenAge: a genomic and proteomic network map of human ageing. FEBS Lett. 2004 [cited 2014 November 26]; 571:243–247.
23. Clark C, Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, Palotie A, Coffey AJ. A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip(®) for methylome profiling. PloS One. 2012; 7:e50233.
24. Butcher LM, Beck S. AutoMeDIP-seq: a high-throughput, whole genome, DNA methylation assay. Methods San Diego Calif. 2010; 52:223–231.
25. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinforma Oxf Engl. 2014; 30:284–286.
26. Akman K, Haaf T, Gravina S, Vijg J, Tresch A. Genome-wide quantitative analysis of DNA methylation from bisulfite sequencing data. Bioinforma Oxf Engl. 2014; 30:1933–1934.
27. Morris TJ, Beck S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. Methods San Diego Calif. 2014.
28. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012; 13:484–492.
29. Guerrero-Bosagna CM, Sabat P, Valdovinos FS, Valladares LE, Clark SJ. Epigenetic and phenotypic changes result from a continuous pre and post natal dietary exposure to phytoestrogens in an experimental population of mice. BMC Physiol. 2008; 8:17.
30. Schneider BG, Peng D-F, Camargo MC, Piazuelo MB, Sicinschi LA, Mera R, Romero-Gallo J, Delgado AG, Bravo LE, Wilson KT, Peek RM, Correa P, El-Rifai W. Promoter DNA hypermethylation in gastric biopsies from subjects at high and low risk for gastric cancer. Int J Cancer J Int Cancer. 2010; 127:2588–2597.
31. Blagosklonny MV. Aging is not programmed: genetic pseudo-program is a shadow of developmental growth. Cell Cycle Georget Tex. 2013; 12:3736–3742.
32. Alvarez-López MJ, Castro-Freire M, Cosín-Tomás M, Sanchez-Roige S, Lalanza JF, Del Valle J, Párrizas M, Camins A, Pallás M, Escorihuela RM, Kaliman P. Long-Term Exercise Modulates Hippocampal Gene Expression in Senescent Female Mice. J Alzheimers Dis JAD. 2013; 33:1177–1190.
33. Misra V, Lee H, Singh A, Huang K, Thimmulappa RK, Mitzner W, Biswal S, Tankersley CG. Global expression profiles from C57BL/6J and DBA/2J mouse lungs to determine aging-related genes. Physiol Genomics. 2007; 31:429–440.
34. Nishida N, Nagasaka T, Nishimura T, Ikai I, Boland CR, Goel A. Aberrant methylation of multiple tumor suppressor genes in aging liver, chronic hepatitis, and hepatocellular carcinoma. Hepatol Baltim Md. 2008; 47:908–918.
35. Ohi T, Uehara Y, Takatsu M, Watanabe M, Ono T. Hypermethylation of CpGs in the promoter of the COL1A1 gene in the aged periodontal ligament. J Dent Res. 2006; 85:245–250.
36. Permatasari F, Hu Y, Zhang J, Zhou B, Luo D. Anti-photoaging potential of Botulinum Toxin Type A in UVB-induced premature senescence of human dermal fibroblasts in vitro through decreasing senescence-related proteins. J Photochem Photobiol B. 2014; 133:115–123.
37. Rice RA, Berchtold NC, Cotman CW, Green KN. Age-related downregulation of the CaV3.1 T-type calcium channel as a mediator of amyloid beta production. Neurobiol Aging. 2014; 35:1002–1011.
38. Sataranatarajan K, Feliers D, Mariappan MM, Lee HJ, Lee MJ, Day RT, Yalamanchili HB, Choudhury GG, Barnes JL, Van Remmen H, Richardson A, Kasinath BS. Molecular events in matrix protein metabolism in the aging kidney. Aging Cell. 2012; 11:1065–1073.
39. Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, Vilain E. Epigenetic predictor of age. PloS One. 2011; 6:e14821.
40. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012; 13:86.

SUPPLEMENTARY FIGURES

Please browse the full text version to see the Supplementary Figures of this manuscript.

Annals of Human Biology
Journal of the Society for the Study of Human Biology

http://informahealthcare.com/ahb
ISSN: 0301-4460 (print), 1464-5033 (electronic)

Ann Hum Biol, 2015; 42(1): 1–9
 DOI: 10.3109/03014460.2014.961960

informa
healthcare

REVIEW PAPER

The epigenetic side of human adaptation: hypotheses, evidences and theories

Cristina Giuliani[1], Maria Giulia Bacalini[2,3], Marco Sazzini[1], Chiara Pirazzini[2,3], Claudio Franceschi[2,4,5], Paolo Garagnani[2,3,6], and Donata Luiselli[1]

[1]*Department of Biological, Geological and Environmental Sciences, Laboratory of Molecular Anthropology and Centre for Genome Biology,* [2]*Department of Experimental, Diagnostic and Specialty Medicine (DIMES),* [3]*Interdepartmental Centre 'L. Galvani' for Integrated Studies of Bioinformatics, Biophysics and Biocomplexity (CIG), University of Bologna, Bologna, Italy,* [4]*IRCCS, Institute of Neurological Sciences of Bologna, Bologna, Italy,* [5]*IGM-CNR Institute of Molecular Genetics, Unit of Bologna IOR, Bologna, Italy, and* [6]*Center for Applied Biomedical Research (CRBA), St. Orsola-Malpighi University Hospital, Bologna, Italy*

Abstract

Context: Epigenetics represents a still unexplored research field in the understanding of micro- and macro-evolutionary mechanisms, as epigenetic changes create phenotypic diversity within both individuals and populations.
Objective: The purpose of this review is to dissect the landscape of studies focused on DNA methylation, one of the most described epigenetic mechanisms, emphasizing the aspects that could be relevant in human adaptations.
Methods: Theories and results here considered were collected from the most recent papers published.
Results: The matter of DNA methylation inheritance is here described as well as the recent evolutionary theories regarding the role of DNA methylation—and epigenetics in a broader sense—in human evolution. The complex relation between (1) DNA methylation and genetic variability and (2) DNA methylation and the environmental stimuli crucial in shaping genetic and phenotypic variability through the human lineage—such as diet, climate and pathogens exposure—are described. Papers about population epigenetics are also illustrated due to their high relevance in this context.
Conclusion: Genetic, epigenetic and phenotypic variations of the species, together with cultural ones, are considerably shaped by a vast range of environmental stimuli, thus representing the foundation of all human bio-cultural adaptations.

Keywords

Adaptation, DNA methylation, environment, human evolution

History

Received 10 July 2014
Accepted 2 September 2014
Published online 21 November 2014

Introduction

It has long been accepted that natural selection favours the fittest organism, suggesting that individuals who acquired favourable adaptations to their environment will be positively selected during evolution. In this framework, genetic variation and especially polymorphisms able to influence phenotypic traits become fundamental as potential ''substrates'' on which environmental (and/or, in the case of *H. sapiens*, cultural) selective pressures could act and through which natural selection shapes diversity patterns of populations and species.

However, genetics appears to not be the only domain able to describe processes underlying the adaptive events that occurred in the evolutionary histories of organisms. For instance, epigenetics represents a relatively new and still unexplored research field that promises to contribute to a deeper understanding of micro- and macro-evolutionary mechanisms, according to the capability of epigenetic changes to create phenotypic diversity within both individuals and populations, with or without the occurrence of genetic variation (Richards, 2008).

In particular, this review will focus on the dissection of the landscape of studies focused on DNA methylation, one of the most described epigenetic mechanisms, emphasizing the aspects that could be relevant for a deeper understanding of processes involved in human adaptations. The complex relation between changes in DNA methylation and in DNA sequence, as well as the environmental stimuli able to produce variability in methylation patterns will be, thus, described. Regarding environmental factors, we will especially focus on those that have played a crucial role in shaping genetic and phenotypic variability through the human lineage, such as diet, climate and pathogens exposure.

For this purpose, in the following section we will review the most informative studies that have exploited an epigenetics approach to the study of human phenotypic traits, in the

Correspondence: Dr Paolo Garagnani, DIMES—Department of Experimental, Diagnostic and Specialty Medicine (University of Bologna), Via San Giacomo 12, I-40126 Bologna, Italy. Tel: +39 051 209 4748. Fax: +39 051 209 4747. E-mail: paolo.garagnani2@unibo.it

attempt to shed new light on the recent evolutionary history of our species.

DNA methylation overview

DNA methylation is a molecular mechanism entailing the covalent addition of a methyl group in the fifth carbon of the cytosine pyrimidine ring (m5C) that, in mammals, is usually located within a CpG dinucleotide. In the human genome, the density of these CpGs is non-random and they also appear to be under-represented with respect to what is expected according to such genomic complexity.

This imbalance is due to the high rate of mutation of m5C to thymine (i.e. deamination) that likely led to a reduction of CpG dinucleotides in mammalian genomes over the course of evolution (Razin & Riggs, 1980). Regions with a higher CpGs content than the genomic average (i.e. ''CpG islands'') are supposed to be ~29 000 throughout the human genome, although this estimate could vary according to CpG island definition and 60% of them co-localize with gene promoters. The DNA methylation status of these CpG islands is subjected to dynamic changes during development and cell differentiation and is known to affect gene expression levels (Antequera, 2003; Lander et al., 2001). In particular, it could have different effects according to the position of the involved CpGs. For instance, methylation of CpG islands at transcription start sites (TSSs) is usually associated with inactivation of gene expression by interfering with transcription factors or, indirectly, by recruiting chromatin modification enzymes. On the contrary, CpGs located in gene bodies might even stimulate transcription and influence splicing mechanisms. Many methylated CpGs are also found in repeated sequences, most of which are derived from transposable elements, suggesting that they could play a fundamental role in genome stability, acting as suppressors of transposable elements expression (Jones, 2012).

In general, vertebrate promoters can be divided into two classes—low CpGs (LCG) and high CpGs (HCG)—just according to their CpGs content. Most of LCG are methylated and are associated to tissue-specific genes, while HCG are usually hypo-methylated and are related to broadly expressed genes. One possible evolutionary reason for such bimodal distribution is that broadly expressed genes have selectively avoided DNA methylation because their consequent inactivation would have been deleterious (Elango & Yi, 2008).

Transgenerational inheritance of DNA methylation patterns

The establishment of DNA methylation patterns is a complex and dynamic process, strictly connected with the development of the organism and influenced by genetic and environmental factors.

Soon after fertilization, the genome of the zygote encounters a global DNA de-methylation event, which is necessary to erase the epigenetic patterns characteristic of the gametes and to guarantee the zygote totipotency (Feng et al., 2010; Surani et al., 2007). However, a consistent proportion of the genome is protected from DNA de-methylation and retains its methylation marks. This is the case of imprinted genes (i.e. genes that are expressed only from the maternal or the paternal allele), whose methylation state is established during gametes differentiation. Moreover, during the development, activity of *de novo* DNA methyltransferases 3a and 3b (Dnmt3a/b) is critical to introduce tissue-specific epigenetic marks and to ensure cell differentiation (Chen et al., 2003; Okano et al., 1999). In adult individuals, the methylation status of most CpGs is maintained across cell divisions through the activity of Dnmt1, allowing stable tissue-specific gene expression without changes in DNA sequence. However, the epigenome retains a certain level of plasticity and can be modified by several environmental factors (e.g. diet, climate, stress, exposure to environmental compounds or pathogens) or by physiological (e.g. ageing) or pathological (e.g. cancer) conditions (Bacalini et al., 2014).

For many years, it was believed that phenotypic effects of epigenetic modifications concerned only the individual exposed to related environmental stimuli. However, several recent evidence indicates that some of the changes in DNA methylation induced by environmental stimuli can be also be transmitted to generations derived from the exposed individual (Daxinger & Whitelaw, 2012; Franklin & Mansuy, 2010; Heard & Martienssen, 2014; Skinner et al., 2010). Inheritance of the methylation status of a DNA sequence (defined as epiallele) is well-described in plants (Becker & Weigel, 2012; Schmitz et al., 2011), while its occurrence in mammals is sustained by a minor body of evidences (Dolinoy et al., 2007; Rakyan et al., 2002, 2003). When dealing with this issue, it is important to distinguish between its inter-generational and transgenerational effects (Heard & Martienssen, 2014). Intergenerational effects are observed when an individual inherits a DNA methylation-driven phenotype because it is indirectly exposed to a triggering environmental force, for instance through *in utero* exposure of the offspring and of its developing gametes. On the contrary, transgenerational effects occur when epialleles are maintained in generations that were never exposed to a given triggering environmental stimulus (Morgan et al., 1999; Rakyan et al., 2003). This implies that DNA methylation signatures are inherited through the gametes and escape the erasure that occurs not only in the zygote (see above), but also during germline maturation (Popp et al., 2010). Although the extent and the frequency of this epigenetic inheritance in mammals are still not clear, it has major implications from an evolutionary perspective, as it represents a potentially adaptive mechanism directly modulated by the environment and on which natural selection can act. Moreover, a recent study (Turner, 2011) suggests that epigenetic changes could persist over generations in two ways: (i) through the epigenetic inheritance described above; and (ii) through the persistence of a long-term stimulus that forces a certain epigenetic profile during generations, leading to a consequent genetic change that is able to exert the same phenotypic effect. In this latter case, the acquired phenotype becomes heritable in a sequence-dependent manner.

Evolutionary relevance of DNA methylation

Some studies investigated the role that epigenetic variations might have played in human evolutionary history and, accordingly, several theories have been proposed. Certainly,

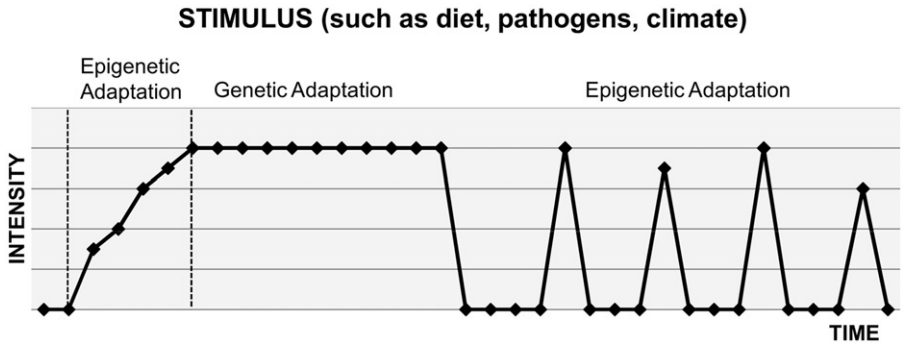


Figure 1. Epigenetic changes as a medium-term way for a population to adapt to environmental stimuli. The type of adaptation can change according to intensity and persistence of a certain environmental stimulus.

the fact that DNA methylation constitutes a source of phenotypic variability, together with recent evidence showing that some epigenetic changes can escape transgenerational erasure mechanisms, suggests that these modifications have the potential to influence micro-evolutionary processes.

In a recent study, Shea et al. (2011) listed two different epigenetic mechanisms that could have shaped human evolution by producing, respectively, selection-based and detection-based effects. The first mechanism is similar to that through which natural selection acts on genes; that is, a specific epigenetic pattern is maintained across generations and is inherited by the offspring from the parents. The second mechanism is instead based on an ''environment detection'' process, for which parents experience a particular environment and transmit an epigenetic pattern apt to cope with it.

According to this view, Gluckman et al. (2005) described a set of theories, known as ''predictive adaptive responses'', based on the hypothesis that conditions experienced *in utero* may confer advantages in a later phase of life.

It is important to note that it is not known to what extent epigenetic variants are transmitted *per se* or arise via *de novo* mutations at each generation (Rando & Verstrepen, 2007). However, Jablonka & Lamb (2005) argued that, since DNA methylation can alter gene expression, and consequently phenotypes, epialleles can be subjected to natural selection, as though alleles at genetic loci. Moreover, Shea et al. (2011) considered a third channel of information linked to somatic cells identity and related to the fact that cells belonging to a given tissue carry information about which specific cell type to become. The authors argued that in multi-cellular organisms information conflicts could arise since a single locus is not able to carry different information (i.e. about cell differentiation and about selection- or detection-based effects of specific epigenetic profiles). The only way to avoid this problem seems to entail the existence of different loci that carry different types of information. For example, regarding the DNA methylation mechanism, some specific CpG sites are involved in somatic differentiation and they could not be involved also in mechanisms producing selection-based effects at the same time. Obviously, these processes depend on how much the epigenetic marks have been erased.

Another study (Klironomos et al., 2013) hypothesizes that variants beneficial in a population could be of genetic and epigenetic origin and that both of them contribute to increased fitness of individuals. However, CpG sites are characterized by a higher mutation rate than single standard nucleotide loci and this means that adaptation achieved through changes in DNA methylation could be lost more rapidly than that achieved via genetic mutations. The authors thus argue that only a high strength of selection could determine the fixation of a given epiallele. In other words, under very strong selective pressures the use of epigenetic mechanisms to encode new and potential adaptive phenotypes could be advantageous with respect to the use of genetic ones. They also argued that, once a phenotype has been produced by DNA methylation changes, the related genetic variation is not under selection and vice versa. In this way, they hypothesized that populations that have to face a new environment could ''buy time'' by using DNA methylation changes to cope with environmental cues, until a new potentially adaptive genetic mutation arises (Figure 1).

Regarding these theories, it is important to cite the study of Feinberg & Irizarry (2010) that proposes a new idea for the role that epigenetic mechanisms could have played in human evolution. In fact, they suggest that some genetic variants may exert an effect on the variability of a certain trait without altering the mean phenotype. According to the authors' hypothesis, propensity to stochastically vary is mediated epigenetically and populations carrying genetic variants associated to this higher propensity to variability might have an advantage (or a disadvantage) in changing environments.

DNA methylation changes affected by genetic variation

In the last years, many studies (Bell et al., 2011; Smith et al., 2014) addressed the association between sequence variants (e.g. single nucleotide polymorphisms, SNPs) and DNA methylation patterns at specific CpGs. In fact, some of them identified DNA methylation quantitative trait loci (meQTLs) as genetic variants associated to DNA methylation at CpG sites (Smith et al., 2014). Moreover, methylation patterns seem to be not only linked to SNPs, but also to DNA composition in terms of local nucleotide sequence content. A recent study (Bock et al., 2006) assigned a degree of ''methylation propensity'' for each CpG island calculated on the basis of the DNA sequence itself. Accordingly, the authors argue that, in normal tissues, DNA composition of CpG islands could be used to predict with high accuracy their methylation status, excluding those differentially methylated in a tissue-specific manner (i.e. 5% of CpG islands in the genome) (Song et al., 2005).

Nevertheless, in order to effectively distinguish the contributions of genetic and environmental factors in the

examined traits, studies on twins continue to constitute a unique model. In general, patterns of DNA methylation in monozygotic twins (MZ) are found to be more concordant than those observed in dizygotic twins or unrelated pairs, indicating a genetic influence on DNA methylation profiles (Fraga et al., 2005). Moreover, this genome-wide study carried out by Fraga et al. (2005) reveals that epigenetic modifications in MZ twins pairs arise also during ageing. The authors suggest that differences in internal and external factors (e.g. smoking habits, lifestyle, diet, etc.), as well as a sort of ''epigenetic drift'', could be the main causes of differences in DNA methylation profiles that occurred within genetic identical backgrounds, as subsequently confirmed by candidate-genes studies (Pirazzini et al., 2012). Data from studies on twins perfectly pointed out that DNA methylation differences between individuals could not be entirely understood considering only their genetic backgrounds. Accordingly, many studies and theories are being focused on the effects of the environment on DNA methylation patterns. Moreover, it should be kept in mind that, with environmental stimuli being equal, DNA methylation patterns can also be affected by physiological processes, such as ageing (Hannum et al., 2013). Methylation profiles of the *ELOVL2*, *FHL2*, and *PENK* genes represent only some examples of how DNA methylation can change during ageing, despite the genetic inter-individual variability (Garagnani et al., 2012). Nevertheless, reviewing all of them is not the purpose of this article, which is instead aimed at describing the relationship between epigenetic modifications, especially DNA methylation ones, and the main environmental stimuli (i.e. diet, climate and pathogens exposure) that have played a crucial role in shaping the genetic variability of our species throughout its evolutionary history. As concerns this issue, the most informative studies are those that have succeeded in exploiting a population epigenetics approach and we have reviewed these in the following section.

DNA methylation changes due to environmental stimuli

Diet in the periconceptional period and in adulthood

As previously mentioned, DNA methylation changes that could be inherited could also be modified by a wide range of environmental stimuli, whose effects directly depend on the duration of the stimulus and on the genetic background of the individual subject to it. It is well known that, from one generation to the next one, methylation status is generally re-set. However, some CpG regions of the genome represent an exception, being not modified and allowing transmission of epigenetic information from one generation to another. Furthermore, it was observed that methylation status in the periconceptional period is very sensitive to external stimuli, such as the diet of the mother and her exposure to toxins (Barnes & Ozanne, 2011; Laker et al., 2013). In fact, recent studies carried out on animal models have shown that the diet of the mother can modulate the epigenome of the zygote, sometimes leading also to increased disease susceptibility in adulthood. The most well known studies on humans in this field take into consideration individuals born from mothers who have experienced famine in 1945 during the Dutch hunger winter (Heijmans et al., 2008; Tobi et al., 2009). The authors found that periconceptional exposure to famine is associated with the lower DNA methylation level observed decades later at *IGF2* and at other candidate genes (imprinted and non-imprinted). Accordingly, they argued that hypomethylation in adulthood could be related to deficiency in methyl donors (e.g. methionine) in early life phases, but they also did not exclude the possible effects of low temperatures and stress to which the mothers were subject to. This study supports the idea that transient environmental stimuli can be recorded as DNA methylation information also in humans. These findings indeed confirm results obtained with rodents that already highlighted changes in nutrition associated with epigenetic instability (mainly due to transposable elements) during early development (Waterland & Jirtle, 2003). However, recent studies showed that not only nutritional habits of the mother can influence DNA methylation patterns of the newborn, but that also folate levels of the father may contribute to healthy pregnancy outcomes (Lambrot et al., 2013). Interestingly, the vast majority of studies focus on *in utero* exposition to stress and environmental stimuli (Langley-Evans, 2014), while investigation of epigenetic plasticity in adulthood is less addressed. In fact, only recently the role of diet in DNA methylation changes of adult individuals, and also during ageing, is acquiring increasing relevance (Bacalini et al., 2014). Moreover, whereas DNA methylation data on murine models are abundant, limited data are still available for humans. For instance, an interesting study by Scoccianti et al. (2011) demonstrated that DNA profiles of some candidate genes change according to 4-week diet intervention. In fact, this study investigated the effects of three dietary regimens on DNA methylation patterns in white blood cells of heavy smokers. One of the tested diets was balanced in fruits and vegetables, another one was enriched in flavonoids and isothiocyanates, particularly cruciferous vegetables, and the third one was supplemented with flavonoids, such as green tea and soy products. The authors demonstrated that an isocaloric diet (balanced in fruits and vegetables) may succeed in stabilizing global epigenetic patterns (estimated by measuring DNA methylation of LINE1 (Yang et al. 2004)), but has no impact at specific genes selected; however, no changes in DNA methylation profiles were observed after a short-term dietary intervention.

A more recent study (Jacobsen et al., 2012) reports for the first time that a short-time high fat overfeeding diet alters DNA methylation profiles at the genome-wide level in human skeletal muscles. The authors also showed that changes were reversed, at least partially, after a washout period of 8 weeks. As mentioned above, the periconceptional period is a critical window in which certain stimuli (e.g. exposure to toxins) can promote epigenetic transgenerational inheritance of complex diseases. For instance, a recent study demonstrates that exposure to the insecticide dichlorodiphenyltrichloroethane (DDT) can affect DNA methylation patterns across generations (Skinner et al., 2010). The authors demonstrated that peculiar methylation profiles are observed not only in F0, F1 and F2, but also in the F3 generation. This is a fundamental point suggesting that F3 generation profiles are likely due to a transgenerational germline-mediated mechanism. Moreover, another study shows that pregnant agouti mothers fed with

bisphenol A (BPA) generate offspring characterized by hypomethylation of the agouti gene that is associated with a more unhealthy outcome (i.e. more incidence of obesity and diabetes) (Dolinoy et al., 2007). This study also demonstrates that a maternal nutrient supplementation of methyl-rich foods can neutralize the negative effects of BPA exposure (Dolinoy et al., 2007).

Furthermore, not only diet, but also physical activity represents another factor that is able to modify DNA methylation status in adulthood. A recent genome-wide study considered DNA methylation profiles of more than 450 000 CpG sites in adipose tissues of healthy individuals, showing that DNA methylation plasticity and modulation is observable, excluding changes in cellular composition, after 6 months of physical activity in 17 975 CpG sites (Rönn et al., 2013). These differentially methylated sites include also obesity- and type 2 diabetes-associated genes, such as *TCF7L2* and *KCNQ1*. Moreover, obtained results indicated that the vast majority of variability is located in non-CpG islands regions, supporting other evidence that demonstrated how CpG islands shores (i.e. regions flanking CpG islands) have higher probability to vary than CpG islands (Irizarry et al., 2009).

Pathogens load

Anatomically modern humans spread out of sub-Saharan Africa ~70 000 years ago, beginning to progressively settle in the other continents (Henn et al., 2012). During this colonization process, different human populations had to face completely new and different environmental challenges, which strongly contributed to shape their genome by triggering several adaptive events. Interestingly, it has been proposed that one of the predominant drivers of these local adaptations was the pathogens load (Fumagalli et al., 2011). In particular, selective pressures imposed on the human genome by helminthes infections seem to be stronger than those related to viral and bacterial ones. The authors also suggest that the genome of helminthes presents an evolutionary rate similar to the human one, making the establishment of effective co-evolutionary processes possible. The same mechanism appears to be not plausible in the case of virus, according to the faster evolving rates with respect to that of our species.

However, what role epigenetic variability has in this scenario is still far from being depicted. What is known is that DNA methylation profiles can vary when cells are infected by viruses or parasites. A study by Jähner & Jaenisch (1985) firstly demonstrated that DNA methylation patterns change after the integration of viral DNA into host cells. This mechanism is associated with the transcriptional silencing that contributes to viral latency. Moreover, some studies linked epigenetic modifications to hepatitis B infection in the liver, stressing a strong interplay between pathogens and the host cell epigenome (Okamoto et al., 2014; Park et al., 2013; Vivekanandan et al., 2010). A recent study indeed considered epigenome-wide data from three human populations with different ancestry (i.e. African-Americans, Asian-Americans and Americans of European origins), succeeding in identifying a core of 439 CpG sites according to which it is possible to distinguish the examined populations (Heyn et al., 2013). It is interesting to note that these data suggest the possibility that pathogens characterizing different environments could leave different and stable signatures in the methylome of the host. In fact, many gene promoters whose DNA methylation profile differs in the three populations turned out to be associated with immune responses and enriched for transcription factors involved in reactions to viral (i.e. hepatitis B) and bacterial (i.e. Shigellosis) infections. Moreover, the authors suggest that DNA methylation could represent a link between the genotype (i.e. risk alleles associated to increased susceptibility to infection) and the phenotype (i.e. the infection incidence). Accordingly, the hypothesis that selective pressures related to the pathogen landscape are able to shape both the human genome and its methylation profiles appears to be highly plausible.

Climatic and environmental conditions

Local adaptations driven by climate-related selective pressures (e.g. UV radiation, humidity, thermal stresses, etc.) have also been proposed as evolutionary events able to strongly influence differential patterns of genetic variation among human populations (Hancock et al., 2011). For instance, as regards adaptive responses to thermal stresses, different strategies and levels of adaptation can be described, ranging from cultural adaptations, through which humans create artificial and regulated microclimates, to plasticity of physiological mechanisms and, finally, to genetic adaptation (Sazzini et al., 2014). All these responses represent complementary strategies that have allowed some human populations to cope with highly challenging climate conditions.

The type of the adopted adaptive strategy and its success strictly depend on the duration of the environmental stimulus. For instance, physiological modifications are optimal short-term adaptive mechanisms, resulting instead in potentially deleterious long-term ones due to the possibility of causing damage to the organism. One clear example is that of adaptation to high altitude (Xing et al., 2008), in which physiological changes, such as increased heart rate, may be decisive for short-term survival, but finally can cause severe tissue and organ damage. Interestingly, epigenetics mechanisms could represent ''medium-term'' strategies to cope with a demanding environmental condition, such as high altitude hypoxia. Unfortunately, epigenetics studies focused on these issues suffer from a lack of human data. A single very elegant study has until now assessed differences in genome-wide levels of DNA methylation between high- and low-altitude populations, by analysing high- and low-altitude blood samples from Amhara (i.e. an Ethiopian ethnic group that is supposed to have lived at 2500 m for at least 5000 years) and high- and low-altitude saliva samples from Oromo (i.e. an Ethiopian ethnic group that have moved to high altitude in 1500 AD (Alkorta-Aranburu et al., 2012; Quirin, 1992). Interestingly, no methylation differences between low- and high-altitude Amhara individuals were observed, while significant differences were found in Oromo for genes involved in defense from pathogens and in response to hypoxia (i.e. *VEGF*, *MT1G*, *APOBEC3G*). The authors suggest that Oromo, moving to high altitude more recently, are still adapting to this new environmental condition by means of

epigenetic variations. This also reflects the fact that more time is needed to adapt genetically to a new environment (as in the case of high-altitude Amhara) and that DNA methylation could represent an effective medium-term response to environmental changes (Figure 1).

Despite these promising findings, further experiments are needed to take into account tissue specificity, since this study focuses on different tissues in the two examined populations. Nevertheless, epigenetics, and in particular DNA methylation, seems to play a key role in hypoxic response and, in general, hypoxia responsive pathways are epigenetically activated in order to promote tissue survival (Watson et al., 2010). Moreover, it has also been demonstrated that, in a cellular model of hypoxia (i.e. prostate cancer), cells are able to adapt to the lack of oxygen by changing their epigenetic profile, especially their DNA methylation and histone acetylation status (Watson et al., 2009).

In the past, similarity between the environment where the human foetus develops and high altitude hypoxia was proposed, suggesting that the foetus is perfectly adapted to a low oxygen pressure and that chemoreceptor sensitivity increases in the neonatal period (Eastman, 1954; Martin et al., 2010). However, mechanisms and involvement of epigenetic processes in the regulation of hypoxic sensing are still not completely understood. Only a recent paper published by Nanduri et al. (2012) focused on the epigenetic regulation of hypoxic sensing. In particular, it reported DNA hypermethylation of the *Sod2* gene and consequent disruption of redox homeostasis by oxidative stress in neonatal rats exposed to intermittent hypoxia. Since hypoxic sensitivity is strongly connected to elevated oxidative stress, the authors described a possible mechanism by which *Sod2* hypermethylation leads to a decreased expression of the gene. However, by treating neonatal rats with an inhibitor of DNA methylation, they observed a reduction of hypersensitivity of the carotid body and of adrenal chromaffin cells to hypoxia. The effects of hypoxia on DNA methylation patterns are also explored by another study that demonstrated how hypoxia alters the epigenetic profile of cultured human placental trophoblasts (Yuen et al., 2013) and how epigenetic processes may contribute to the maintenance of a hypoxia-adapted status by interacting with the hypoxia-induced transcription factor (HIF) (Watson et al., 2010).

Regarding climatic changes, UV radiation is another environmental force that has considerably shaped human genetic diversity (Jablonski & Chaplin, 2000), but so far only a few studies have investigated DNA methylation changes after UV exposure. Accordingly, some hypotheses regarding the pathways linking UV-induced DNA damage and DNA hypomethylation have been proposed. In fact, UVB mutagenesis in skin cells is characterized by C (cytosine) to T (thymidine) transition and 5-methylcytosine is found to have higher energy adsorption in the UVB range than C, so that this propensity to adsorb UVB can lead to a high probability to lose 5 mC rather than unmethylated C (Lee & Pfeifer, 2003). A recent study explored DNA methylation profiles of more than 27 000 CpG sites under chronic sun exposure and compared DNA methylation of the outer forearm (i.e. the sun-exposed area) with the inner arm (i.e. the sun-protected area), also separating dermis and epidermis from the two districts. Interestingly, 14 CpG sites showed a beta value decreased by 0.2 or more in the sun exposed area of the epidermis. In particular, the authors revealed demethylation for the CpGs immediately distal to the one previously detected, concluding that these results demonstrated the demethylation of sun-exposed epidermis samples at the *KRT75* (keratin 75) promoter region (Grönniger et al., 2010). Moreover, the authors suggest that the observed DNA methylation variability is not linked with an altered expression of DNMT genes. After a deep bisulphite sequencing of the *KRT75* gene, the authors do not detect evidence of genetic mutations, thus not confirming previous hypotheses of UV-induced mutations preferentially at 5 mC, at least for the analysed genomic region.

The population epigenetics approach

Phenotypic plasticity is one of the main conceptual bases of population epigenetic studies, since epigenetic information allows a single genotype to assume multiple epigenotypes, thus enabling the expression of different phenotypes. This is particularly emphasized in plants, some of the organisms that are most obligated to do their best with environmental conditions from which they are not able to get away. Different studies, listed below, highlight the link between membership to a specific ethnic group and DNA methylation patterns, but few of them actually take into account individuals' genetic ancestry, environmental characteristics of the place where they live, or the main demographic and historical events that have occurred in the examined geographical areas.

For instance, Terry et al. (2008) observed differences in DNA methylation according to ethnicity among women that live in the same city and were drawn from a birth cohort followed until midlife, with American people of African ancestry appearing more likely to have lower levels of DNA methylation than American individuals of European or Latin American origins (Terry et al., 2008).

Das et al. (2006) instead evaluated methylation patterns in prostate cancer and benign prostatic hyperplasia tissue samples using methylation-specific PCR (MS-PCR), showing that methylation of the *TMS1/ASC* gene was more prevalent in cases than controls in patients of European ancestry. On the contrary, no differences between disease and healthy subjects were observed for people of African origins ($OR = 1.1$, $p = 0.91$). A more recent study further hypothesized that differences in DNA methylation of specific CpG dinucleotides between severe heroin addicts under methadone maintenance treatment and control subjects partially depend on ethnicity. DNA methylation analysis of the μ-opioid receptor gene (*OPRM1*) promoter region was performed on American individuals with African ancestry and Latin American origins and these were compared with a similar cohort of subjects of European ancestry. However, ethnicity was assessed according to the ethnic/cultural background of the subjects and of their parents, grandparents and great-grandparents, without a specific investigation of their actual genetic ancestry components (Nielsen et al., 2010).

Two highly innovative studies, in terms of adopted technologies and approaches, were recently carried out by two different research groups. A first one highlighted

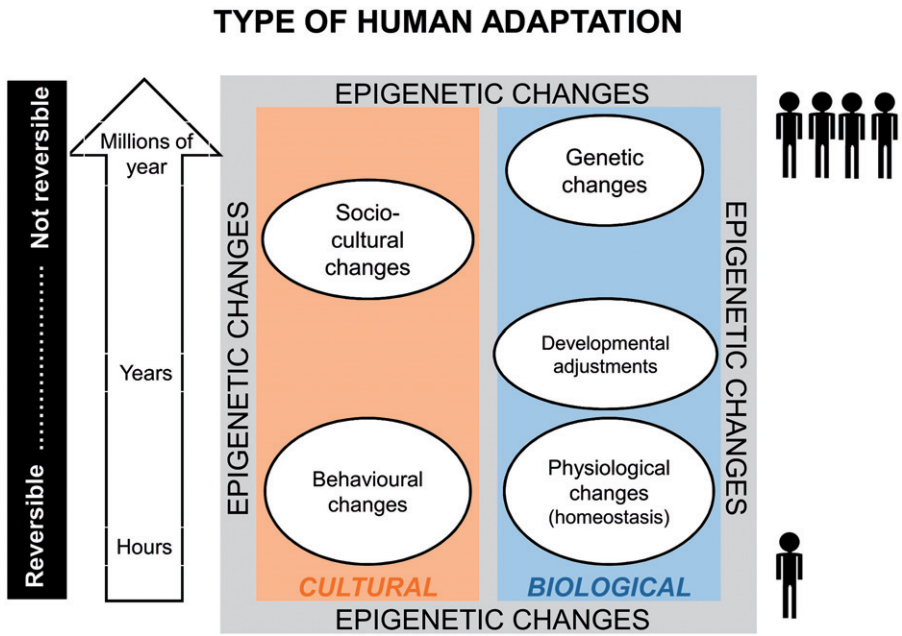


Figure 2. Type of human adaptation. Epigenetics could be considered a reversible way to adapt to new stressors at an individual level as well as at a population level. In the figure different strategies of adaptation to environmental stimuli are listed according to the time in which they arise. Epigenetics, genetics and cultural changes are shaped by a wide range of environmental stimuli, thus representing the foundation of all human bio-cultural adaptations.

significant differences in global genomic DNA methylation of peripheral blood cells by ethnicity, suggesting that this individuals' information should be accurately considered in every methylation study (Zhang et al., 2011). The second one instead tried to assess the impact of population SNPs diversity patterns at the genome-wide level on DNA methylation profiles (Liu et al., 2010). For this purpose, around 200 samples of different self-reported ethnicity were analysed for 1M SNPs and for methylation status at 27 000 CpGs distributed across the whole genome. The authors argued that a significant correlation between genetic and epigenetic population structure exists and they suggest that it is mediated via complex multiple gene interactions.

In line with these findings, Fraser at al. (2012) argue that genetic or environmental interactions likely affect most mSNPs, that is most SNPs identified by evaluating significant correlations between genotypes and methylation levels. To compare DNA methylation profiles between human populations, they utilized lymphoblastoid cell lines (LCLs) from the Hap Map project. In particular, they examined 30 family trios (i.e. mother/father/offspring) of Northern European ancestry (CEU) and 30 trios of Yoruban ancestry (YRI). These samples were characterized for DNA methylation with a 27 k CpGs array, showing that 21.4% of CpG sites differ between the studied populations, even if only 3.9% of these sites actually differ by an average of over 10% of methylation level. Heritability estimates also differ between populations, suggesting extensive divergence in the genetic control of DNA methylation. In the authors' opinion, most of these differences could be attributed to two main factors. One type of divergence that may affect DNA methylation levels is a difference in the CEU/YRI allele frequencies at genetic variants able to influence methylation (e.g. genetic variants can explain over half of the population methylation-specificity observed). On the other hand, a variant that is present in both populations, but affects DNA methylation only in one of them, can be explained uniquely by complex genetic interactions. These interactions could also involve environmental variables, for instance the quantity of available methyl donors, or epistasis with other variants.

A more recent population epigenetics study produced epigenome-wide data for three populations with different ethnicity: Americans of African ancestry, Americans of Asian ancestry and Americans of European origins (Heyn et al., 2013). It succeeded in identifying a core of 439 CpG sites (pop-CpGs), by which it is possible to distinguish the examined populations, pointing out that this pattern is associated with CpGs that regulate gene expression and that many pop-CpGs are located within or nearby transcription binding sites for NFKB, a transcription factor involved in the immune defense. Accordingly, the authors proposed that local selective pressures, such as a specific pathogens landscape, could shape epigenetic variation in genes involved in immune or xenobiotic responses. Moreover, this study demonstrated that two thirds of hypervariable CpG sites were associated with the individuals' genetic background, suggesting that part of population-specific DNA methylation patterns could be linked to genetic differences between populations and that genetic background drives and modulates DNA methylation variability. However, DNA methylation changes could be divided in genotypic-dependent and genotypic-independent ones, with one third of CpG sites that vary according to ethnicity being not linked with genetic variance, thus suggesting that epigenetic variation can also occur independently to genetic background (Heyn et al., 2013).

Conclusions

According to available data, it appears clear that genetic, epigenetic and phenotypic variations of our species, together with cultural ones, are considerably shaped by a vast range of environmental stimuli, thus representing the foundation of all human bio-cultural adaptations (Figure 2).

In this article, we have especially reviewed the most recent studies that provide robust evidence pointing to epigenetic changes, and in particular DNA methylation, as suitable

medium-term evolutionary mechanisms enabling humans to adapt in response to environmental stresses. Although an increasing number of studies are being devoted to explore this issue, the epigenetic contribution to human adaptations is far from being completely understood. For this purpose, we believe that more studies, based on more accurate sampling strategies and deeper investigation of examined individuals' genetic ancestry components according to well-established Anthropological Genetics approaches, are needed to reliably dissect to what extent DNA methylation contributes to human phenotypic variability. It is also essential to note that all these issues are substantially complicated by the fact that DNA methylation is a tissue-specific process and that, therefore, a certain epitype could be associated to a specific phenotype in a tightly tissue-specific manner. Despite these problematic aspects, studies aimed at describing the distribution of epitypes across worldwide human populations actually represent the most suitable approach to characterize and explore the inter-play between genotype and epitype in shaping human phenotypic variation.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

References

Alkorta-Aranburu G, Beall CM, Witonsky DB, Gebremedhin A, Pritchard JK, Di Rienzo A. 2012. The genetic architecture of adaptations to high altitude in Ethiopia. PLoS Genet 8:e1003110.

Antequera F. 2003. Structure, function and evolution of CpG island promoters. CMLS 60:1647–1658.

Bacalini MG, Friso S, Olivieri F, Pirazzini C, Giuliani C, Capri M, Santoro A, et al. 2014. Present and future of anti-ageing epigenetic diets. Mech Ageing Dev 136–137:101–115.

Barnes SK, Ozanne SE. 2011. Pathways linking the early environment to long-term health and lifespan. Prog Biophys Mol Biol 106:323–336.

Becker C, Weigel D. 2012. Epigenetic variation: origin and transgenerational inheritance. Curr Opin Plant Biol 15:562–567.

Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK. 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol 12:R10.

Bock C, Paulsen M, Tierling S, Mikeska T, Lengauer T, Walter J. 2006. CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats, and predicted DNA structure. PLoS Genet 2:243–252, e26.

Chen T, Ueda Y, Dodge JE, Wang Z, Li E. 2003. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol 23:5594–5605.

Das PM, Ramachandran K, Vanwert J, Ferdinand L, Gopisetty G, Reis IM, Singal R. 2006. Methylation mediated silencing of TMS1/ASC gene in prostate cancer. Mol Cancer 5:28.

Daxinger L, Whitelaw E. 2012. Understanding transgenerational epigenetic inheritance via the gametes in mammals. Nat Rev Genet 13:153–162.

Dolinoy DC, Huang D, Jirtle RL. 2007. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc Natl Acad Sci 104:13056–13061.

Eastman NJ. 1954. Mount Everest *in utero*. Am J Obstet Gynecol 67: 701–711.

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

Feinberg AP, Irizarry RA. 2010. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc Natl Acad Sci USA 107(Suppl 1):1757–1764.

Feng S, Jacobsen SE, Reik W. 2010. Epigenetic reprogramming in plant and animal development. Science (N.Y.) 330:622–627.

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suñer D, et al. 2005. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci USA 102: 10604–10609.

Franklin TB, Mansuy IM. 2010. Epigenetic inheritance in mammals: evidence for the impact of adverse environmental effects. Neurobiol Dis 39:61–65.

Fraser HB, Lam LL, Neumann SM, Kobor MS. 2012. Population-specificity of human DNA methylation. Genome Biol 13:R8.

Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Ferrer-Admettla A, Pattini L, Nielsen R. 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet 7:e1002355.

Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, et al. 2012. Methylation of ELOVL2 gene as a new epigenetic marker of age. Aging Cell 11:1132–1134.

Gluckman PD, Hanson MA, Spencer HG. 2005. Predictive adaptive responses and human evolution. Trends Ecol Evol 20: 527–533.

Grönniger E, Weber B, Heil O, Peters N, Stäb F, Wenck H, Korn B, et al. 2010. Aging and chronic sun exposure cause distinct epigenetic changes in human skin. PLoS Genet 6:e1000971.

Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, Utermann G, et al. 2011. Adaptations to climate-mediated selective pressures in humans. PLoS Genet 7: e1001375.

Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, et al. 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 49:359–367.

Heard E, Martienssen RA. 2014. Transgenerational epigenetic inheritance: myths and mechanisms. Cell 157:95–109.

Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, Lumey LH. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci 105:17046–17049.

Henn BM, Cavalli-Sforza LL, Feldman MW. 2012. The great human expansion. Proc Natl Acad Sci USA 109:17758–17764.

Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, Monk D, et al. 2013. DNA methylation contributes to natural human variation. Genome Res 23:1363–1372.

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, et al. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 41:178–186.

Jablonka E, Lamb MJ. 2005. Evolution in four dimensions: genetic, epigenetic, behavioral, and symbolic variation in the history of life. Cambridge, Massachusetts, USA: The MIT Press.

Jablonski NG, Chaplin G. 2000. The evolution of human skin coloration. J Hum Evol 39:57–106.

Jacobsen SC, Brøns C, Bork-Jensen J, Ribel-Madsen R, Yang B, Lara E, Hall E, et al. 2012. Effects of short-term high-fat overfeeding on genome-wide DNA methylation in the skeletal muscle of healthy young men. Diabetologia 55:3341–3349.

Jähner D, Jaenisch R. 1985. Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity. Nature 315: 594–597.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13:484–492.

Klironomos FD, Berg J, Collins S. 2013. How epigenetic mutations can affect genetic evolution: model and mechanism. BioEssays: News Rev Mol Cell Dev Biol 35:571–578.

Laker RC, Wlodek ME, Connelly JJ, Yan Z. 2013. Epigenetic origins of metabolic disease: the impact of the maternal condition to the offspring epigenome and later health consequences. Food Sci Hum Wellness 2:1–11.

Lambrot R, Xu C, Saint-Phar S, Chountalos G, Cohen T, Paquet M, Suderman M. 2013. Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes. Nat Commun 4:2889. doi: 10.1038/ncomms3889.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al, International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Langley-Evans SC. 2014. Nutrition in early life and the programming of adult disease: a review. J Hum Nutr Diet. doi: 10.1111/jhn.12212.

Lee D-H, Pfeifer GP. 2003. Deamination of 5-methylcytosines within cyclobutane pyrimidine dimers is an important component of UVB mutagenesis. J Biol Chem 278:10314–10321.
Liu J, Hutchison K, Perrone-Bizzozero N, Morgan M, Sui J, Calhoun V. 2010. Identification of genetic and epigenetic marks involved in population structure. PLoS One 5:e13209.
Martin DS, Khosravi M, Grocott MP, Mythen MG. 2010. Concepts in hypoxia reborn. Crit Care 14:315.
Morgan HD, Sutherland HG, Martin DI, Whitelaw E. 1999. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 23:314–318.
Nanduri J, Makarenko V, Reddy VD, Yuan G, Pawar A, Wang N, Khan SA, et al. 2012. Epigenetic regulation of hypoxic sensing disrupts cardiorespiratory homeostasis. Proc Natl Acad Sci USA 109: 2515–2520.
Nielsen D, Hamon S, Yuferov V, Jackson C, Ho A, Ott J, Kreek M. 2010. Ethnic diversity of DNA methylation in the OPRM1 promoter region in lymphocytes of heroin addicts. Hum Genet 127:639–649.
Okamoto Y, Shinjo K, Shimizu Y, Sano T, Yamao K, Gao W, Fujii M, et al. 2014. Hepatitis virus infection affects DNA methylation in mice with humanized livers. Gastroenterology 146:562–572.
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.
Park E-S, Park YK, Shin CY, Park SH, Ahn SH, Kim DH, Lim K-H, et al. 2013. Hepatitis B virus inhibits liver regeneration via epigenetic regulation of urokinase-type plasminogen activator. Hepatology (Baltimore, MD) 58:762–776.
Pirazzini C, Giuliani C, Bacalini MG, Boattini A, Capri M, Fontanesi E, Marasco E, et al. 2012. Space/population and time/age in DNA methylation variability in humans: a study on IGF2/H19 locus in different Italian populations and in mono- and di-zygotic twins of different age. Aging 4:509–520.
Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W. 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463:1101–1105.
Quirin J. 1992. Review of The Oromo of Ethiopia: a history, 1570–1860 by Mohammed Hassen. Afr Stud Rev 35:123–125.
Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. 2002. Metastable epialleles in mammals. Trend Genet 18:348–351.
Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KVK, Whitelaw E. 2003. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proc Natl Acad Sci USA 100:2538–2543.
Rando OJ, Verstrepen KJ. 2007. Timescales of genetic and epigenetic inheritance. Cell 128:655–668.
Razin A, Riggs AD. 1980. DNA methylation and gene function. Science 210:604–610.
Richards EJ. 2008. Population epigenetics. Curr Opin Genet Dev 18: 221–226.
Rönn T, Volkov P, Davegårdh C, Dayeh T, Hall E, Olsson AH, Nilsson E, et al. 2013. A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. PLoS Genet 9:e1003572.
Sazzini M, Schiavo G, De Fanti S, Martelli PL, Casadio R, Luiselli D. 2014. Searching for signatures of cold adaptations in modern and archaic humans: hints from the brown adipose tissue genes. Heredity (Edinb) 113:259–267.
Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR. 2011. Transgenerational epigenetic instability is a source of novel methylation variants. Science (New York, N.Y.) 334:369–373.
Scoccianti C, Ricceri F, Ferrari P, Cuenin C, Sacerdote C, Polidoro S, Jenab M, et al. 2011. Methylation patterns in sentinel genes in peripheral blood cells of heavy smokers: influence of cruciferous vegetables in an intervention study. Epigenetics Off J DNA Methylation Soc 6:1114–1119.
Shea N, Pen I, Uller T. 2011. Three epigenetic information channels and their different roles in evolution. J Evol Biol 24:1178–1187.
Skinner MK, Manikkam M, Guerrero-Bosagna C. 2010. Epigenetic transgenerational actions of environmental factors in disease etiology. Trends Endocrinol Metab 21:214–222.
Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, Tylavsky FA, Conneely KN. 2014. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. BMC Genomics 15:145.
Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H, Held WA. 2005. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci USA 102:3336–3341.
Surani MA, Hayashi K, Hajkova P. 2007. Genetic and Epigenetic Regulators of Pluripotency. Cell 128:747–762.
Terry MB, Ferris JS, Pilsner R, Flom JD, Tehranifar P, Santella RM, Gamble MV, Susser E. 2008. Genomic DNA methylation among women in a multiethnic New York City Birth Cohort. Cancer Epidemiol Biomarkers Prevent 17:2306–2310.
Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, Slagboom PE, Heijmans BT. 2009. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Hum Mol Genet 18:4046–4053.
Turner BM. 2011. Environmental sensing by chromatin: an epigenetic contribution to evolutionary change. FEBS Letts 585:2032–2040.
Vivekanandan P, Daniel HD-J, Kannangai R, Martinez-Murillo F, Torbenson M. 2010. Hepatitis B Virus replication induces methylation of both host and viral DNA. J Virol 84:4321–4329.
Waterland RA, Jirtle RL. 2003. transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol 23: 5293–5300.
Watson JA, Watson CJ, McCann A, Baugh J. 2010. Epigenetics, the epicenter of the hypoxic response. Epigenetics Off J DNA Methylation Soc 5:293–296.
Watson JA, Watson CJ, McCrohan A-M, Woodfine K, Tosetto M, McDaid J, Gallagher E, et al. 2009. Generation of an epigenetic signature by chronic hypoxia in prostate cells. Hum Mol Genet 18: 3594–3604.
Xing G, Qualls C, Huicho L, Rivera-Ch M, River-Ch M, Stobdan T, Slessarev M, et al. 2008. Adaptation and mal-adaptation to ambient hypoxia; Andean, Ethiopian and Himalayan patterns. PloS One 3: e2342.
Yang AS, Estecio MRH, Doshi K, Kondo Y, Tajara EH, Issa J-PJ. 2004. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucl Acids Res 32:e38.
Yuen RKC, Chen B, Blair JD, Robinson WP, Nelson DM. 2013. Hypoxia alters the epigenetic profile in cultured human placental trophoblasts. Epigenetics Off J DNA Methylation Soc 8:192–202.
Zhang FF, Cardarelli R, Carroll J, Fulda KG, Kaur M, Gonzalez K, Vishwanatha JK, et al. 2011. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics Off J DNA Methylation Soc 6:623–629.

Mechanisms of Ageing and Development 136-137 (2014) 101–115

Contents lists available at ScienceDirect

Mechanisms of Ageing and Development

journal homepage: www.elsevier.com/locate/mechagedev

Present and future of anti-ageing epigenetic diets

Maria Giulia Bacalini [a,b], Simonetta Friso [c], Fabiola Olivieri [d,e], Chiara Pirazzini [a,b], Cristina Giuliani [f], Miriam Capri [a,b], Aurelia Santoro [a,b], Claudio Franceschi [a,b], Paolo Garagnani [a,b,*]

[a] Department of Experimental, Diagnostic and Specialty Medicine (DIMES), University of Bologna, Via San Giacomo 12, 40126 Bologna, Italy
[b] Interdepartmental Center "L. Galvani" (CIG), University of Bologna, Piazza di Porta San Donato 1, 40126 Bologna, Italy
[c] Department of Medicine, University of Verona School of Medicine, Policlinico "G.B. Rossi," P.le L.A. Scuro 10, 37134 Verona, Italy
[d] Department of Clinical and Molecular Sciences (DISCLIMO), Università Politecnica delle Marche, Via Tronto 10/A, 60020 Ancona, Italy
[e] Center of Clinical Pathology and Innovative Therapy, I.N.R.C.A. National Institute, Via Birarelli n. 8, Ancona, Italy
[f] Department of Biological, Geological and Environmental Sciences (BiGEA), Laboratory of Molecular Anthropology & Centre for Genome Biology, University of Bologna, Via Selmi 3, 40126 Bologna, Italy

ARTICLE INFO

Article history:
Received 24 June 2013
Received in revised form 6 December 2013
Accepted 20 December 2013
Available online 2 January 2014

Keywords:
DNA methylation
microRNAs
Ageing
Nutrition
Inflammaging

ABSTRACT

The rapid technological advancements achieved in the last years have boosted the progressive identification of age-associated epigenetic changes. These studies not only contribute to shed light on the molecular basis of ageing and age-related diseases but, given the plasticity of epigenetic modifications, also provide the basis for anti-ageing interventions to counteract the onset of age-related diseases. In this review we will discuss nutritional interventions as a promising approach that can positively counteract epigenetic changes associated with ageing and promote the health for the elderly. First, we will give an overview of age-associated epigenetic signatures, focusing on DNA methylation. Then, we will report recent evidences regarding the epigenetic changes induced by nutritional interventions in the adulthood (referred as "epigenetic diets"), such as (i) caloric/dietary restriction, (ii) diet supplementation with nutrients involved in one-carbon metabolism and (iii) diet supplementation with bioactive food components. Attention will be drawn on the limits of current studies and the need of proper human models, such as those provided by the ongoing European project NU-AGE. Finally, we will discuss the potential impact of epigenetic diets on inflammaging and age-related diseases, focusing on cardiovascular disease, highlighting the involvement of epigenetic modifications other than DNA methylation, such as microRNA.

1. Introduction

Epigenetic modifications are, by definition, heritable but reversible changes in chromatin structure and gene function. This means that, although epigenetic signatures are transmitted across cellular divisions and – in some cases – also transgenerationally, they can be modified by a wide range of environmental cues, whose effect is in turn dependent on the genetic background of the organism. Epigenetic changes are pivotal in the establishment of developmental patterns in the early life stages, but they appear also to be both triggers and consequences of the phenotypic alterations that characterize ageing. In the past decades a great effort has been placed in order to identify the epigenetic landscapes of ageing and age-related diseases and to interpret them in the framework of ageing theories. From these studies a complex scenario emerges, in which a general relaxation in the mechanisms controlling epigenetic modifications coexists with systemic changes at specific *loci*, suggesting that an epigenetic program originating during development contributes at least in part to the ageing process (de Magalhães, 2012). Given their tissue-specificity, epigenetic modifications could be also able to provide consistent hints regarding the "mosaic of ageing" theory (Cevenini et al., 2008), which states that different tissues/organs age at different rates and differently contribute to the ageing of the whole organism. In addition, the identification of age-related epigenetic

Abbreviations: aDMR, age-associated differentially methylation regions; DNMTs, DNA methyltransferases; ROS, reactive oxygen species; CR, caloric restriction; DR, dietary restriction; ADF, alternate-day fasting; PBMCs, peripheral blood mononuclear cells; AdoMet, S-adenosylmethionine; LINE-1, long interspersed element-1; Treg, regulatory T cells; CVD, cardiovascular disease; T2DM, type 2 diabets mellitus; miRs, microRNAs; DCs, dendritic cells; Hcy, homocysteine; AdoHcy, S-adenosylhomocysteine; CAD, coronary artery disease.

* Corresponding author at: Department of Experimental, Diagnostic and Specialty Medicine (DIMES), University of Bologna, Via San Giacomo 12, 40126 Bologna, Italy. Tel.: +39 0512094748; fax: +39 0512094748.
E-mail address: paolo.garagnani2@unibo.it (P. Garagnani).

0047-6374/$ – see front matter © 2014 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.mad.2013.12.006

Box 1. The challenge of the European project NU-AGE.

As it has been reported in the previous paragraphs, there is a growing amount of data indicating that a series of nutrients, bioactive dietary components and changes in diet habits could modulate the epigenetic pattern impinging upon a series of organs, systems and vital functions that as a whole determine the health status of people at all the stages of life and particularly in elderly. Nevertheless, there is a lack of knowledge regarding the effects of the whole diet on the health status, as the great majority of the past studies have dealt with the effects of a single or few nutrients. New remarkable insights regarding that relationship between nutrition, epigenetic modifications and ageing can be provided by the European project NU-AGE (FP7, n° 266486; www.nu-age.eu). The rationale and the design of the project are deeply described in the paper by Santoro et al. (2014). Briefly, the NU-AGE project will study in a comprehensive and integrated way the effect of a modified/fortified Mediterranean diet (described by Berendsen et al., 2013), specifically designed according to the nutritional needs of people over 65 years of age, on a series of cellular and molecular pathways including epigenetics.
NU-AGE will offer an extraordinary opportunity to elucidate the modulatory role of nutrition on DNA methylation patterns during ageing, considering also the genetic variability (polymorphisms) of some pivotal genes involved in the diet/metabolic response.

changes has two important practical consequences. On one side, epigenetic signatures have emerged as potential biomarkers of biological age. On the other side, the malleability of epigenetic modifications has prompted the search for interventions that, acting on epigenetic regulators, could delay or even revert the aged phenotype (Johnson et al., 2012; Rando and Chang, 2012; Vaiserman and Pasyukova, 2012).

A growing number of studies suggest nutrition as one of most promising anti-ageing interventions (Choi and Friso, 2010; Ford et al., 2011; Ribarič, 2012). Indeed, specific diet components have proved to directly regulate the activity of enzymes that catalyze epigenetic modifications and diet habits can alter the establishment and the maintenance of epigenetic patterns by modifying the intracellular metabolic state or the cellular microenvironment itself.

In this paper we will review the epigenetic changes that occur with ageing and we will provide examples of how nutritional interventions in adulthood, such as that scheduled in the European project NU-AGE (see Box 1), could revert these changes and positively impact the ageing process. Among the broad number of epigenetic modifications, we will specifically focus on DNA methylation, as its role in ageing has been extensively investigated and its responsiveness to nutritional stimuli is the object of several studies. The impact of diet on other age-associated epigenetic modifications, including histone modification, has been deeply reviewed elsewhere (Choi and Friso, 2010; Cosentino and Mostoslavsky, 2013; Huidobro et al., 2013; Li et al., 2011a,b; Martin et al., 2013). Here we will discuss the role of DNA methylation changes during ageing and how DNA methylation patterns can be modified by caloric restriction and by diets enriched in bioactive food components. In this framework we will report current knowledge regarding the contribution of DNA methylation and other epigenetic factors, such as microRNA, to inflammaging, a low-grade systemic inflammation that characterizes the aged phenotype, and their modulation by epigenetic diets. Finally we will discuss how age-related diseases can benefit from nutritional interventions that target epigenetic mechanisms, using cardiovascular disease as example.

2. DNA methylation patterns in human ageing

2.1. DNA methylation

DNA methylation is a covalent modification consisting in the addition of a methyl group to the carbon 5 of a cytosine residue, which usually lies at the 5′ of a guanine residue (CpG dinucleotide). The distribution of CpG dinucleotides in the genome is not random, as they are relatively depleted from the bulk of the genome and enriched at the promoters of around 50% of genes (CpG islands). Recently, the evolutionary dynamics that subtend to the origin and maintenance of CpG islands have been modelled (Cohen et al., 2011). The methylation status of CpG dinucleotides is strictly associated to the regulation of chromatin conformation. CpG islands are usually unmethylated, promoting an open chromatin structure suitable for transcription, and their methylation status is dynamically regulated during development. On the contrary, methylated CpG dinucleotides promote a closed chromatin conformation and are more frequent in the bulk of the genome, in particular in repetitive and intergenic sequences, where they play a central role in maintaining genomic stability and in suppressing spurious transcription.

2.2. Age-associate changes in global DNA methylation

Global DNA methylation level of mammalian genomes decreases with age (Fig. 1, *locus A*). This observation was firstly established more than three decades ago (Vanyushin et al., 1973, 1970) and was then confirmed in multiple species and tissues, including brain, liver, small intestine mucosa, heart, spleen and blood cells (Drinkwater et al., 1989; Fuke et al., 2004; Golbus et al., 1990; Richardson, 2003; Wilson and Jones, 1983; Wilson et al., 1987). Age-dependent global hypomethylation reflects the decrease in methylation of repetitive tandem and interspersed elements, which are highly enriched in CpG dinucleotides (Jintaridth and Mutirangura, 2010; Rodriguez et al., 2008). Longitudinal models have been exploited to assess intra-individual changes in global DNA methylation over time (Bjornsson et al., 2008; Bollati et al., 2009), while the genetic cues that affect the maintenance of epigenetic patterns during ageing have been investigated using family-based cohorts (Bjornsson et al., 2008). A recent study considered the relationship between global DNA methylation and frailty, a common clinical syndrome in the elderly (Bellizzi et al., 2012). Interestingly, the authors showed that global DNA methylation levels were lower in 65–85-year-old frail individuals compared to prefrail and nonfrail subjects, suggesting that the relaxation in the epigenetic control of the genome could be associated with the age-associated functional decline.

2.3. Age-associated changes in DNA methylation at gene promoters

Age-dependent DNA methylation changes have been extensively assessed not only at a global level, but also at specific genomic *loci*. This has allowed the identification of age-associated differentially methylation regions (aDMR) in multiple human tissues. In contrast to global hypomethylation, the trend for gene-associated CpG islands is towards the increase of DNA methylation during ageing (Fig. 1, *locus B*), strictly resembling what happens in cancer (Gilbert, 2009). The list of promoters whose DNA methylation changes upon ageing is continuously increasing (Table 1), prompted by the advancements in technologies to measure sequence-specific DNA methylation (Ndlovu et al., 2011). Currently, candidate gene approaches are generally performed using the Qiagen PyroMark and Sequenom EpiTYPER platforms, that allow high throughput, highly quantitative analysis of 50–500 nt sequences at single base resolution. On the other side, the most

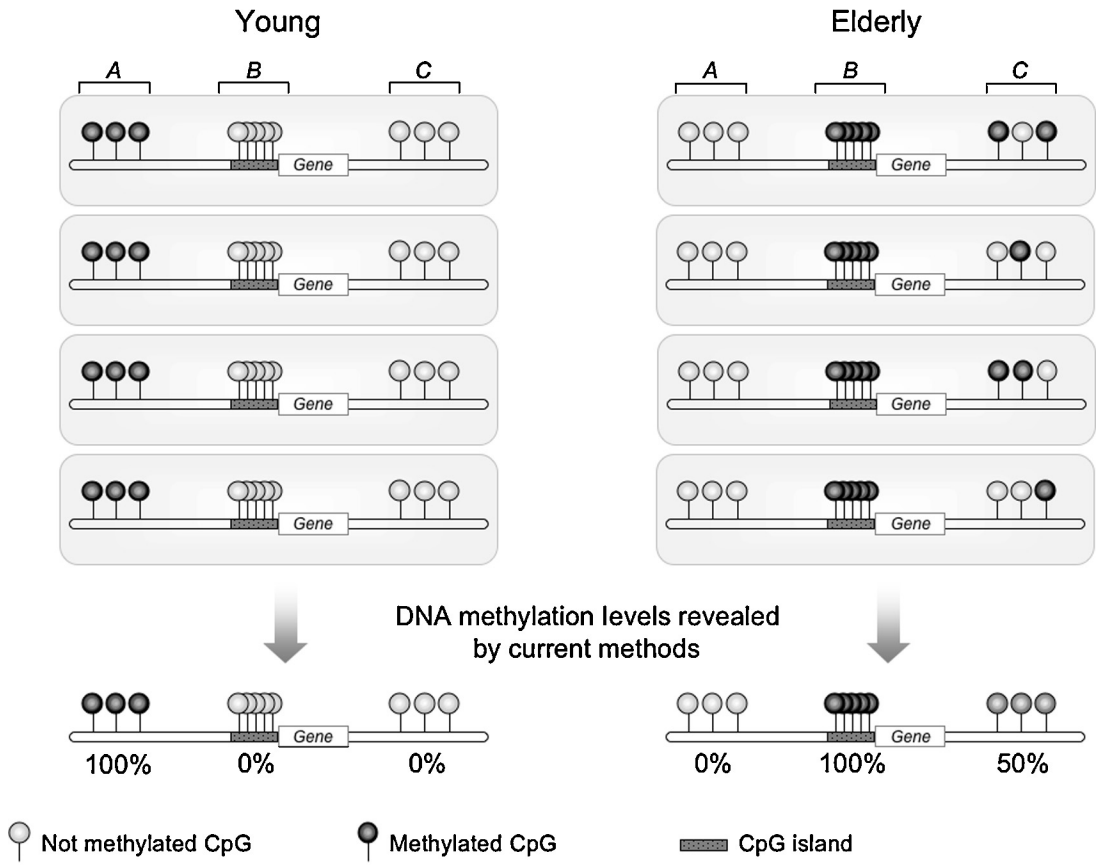


Fig. 1. Age-associated changes in DNA methylation patterns. The figure reports the methylation pattern of 3 genomic *loci*, indicated as *A*, *B* and *C*, in 4 cells from the same tissue in a young (left) and in an old (right) subject. In each DNA molecule, a CpG site can be unmethylated (white circle) or methylated (black circle); for clarity, the methylation profile of only one of the two homologous chromosomes is reported. In the bottom of the figure, the DNA methylation levels measured by current approaches are reported. Note that, as the current methods to analyze DNA methylation do not have single-cell resolution, they provide an estimate of the mean methylation level of each site in the cellular population. Repetitive elements (*locus A*) tend to become unmethylated during ageing, leading to a decrease in the global DNA methylation levels of the genome. The CpG islands at the promoter of some genes (*locus B*) are hypermethylated in elderly respect to young subjects as a consequence of an "ageing program" or of age-associated environmental cues. Finally, the maintenance of the methylation patterns at some genomic *loci* (*locus C*) is impaired in the elderly, leading to increased cell-to-cell heterogeneity in DNA methylation profiles. As a consequence, the methylation levels measured by current approaches tends towards 50%.

Table 1
Gene-specific epigenetic changes associated with ageing.

Gene	Tissue	Methylation change with ageing	References
ER	Human colon	Hypermethylation	Issa et al. (1994)
IGF2	Human colon	Hypermethylation	Issa et al. (1996)
SFRP1	Human colon	Hypermethylation	Wallace et al. (2010)
MYOD1, N33	Human colon	Hypermethylation	Ahuja et al. (1998)
Lox, p16INK4a, RUNX3, TIG1	Gastric mucosa	Hypermethylation	So et al. (2006)
c-fos	Liver	Hypermethylation	Choi et al. (1996)
GCR, iNOS, TLR2	Whole blood	Hypomethylation	Madrigano et al. (2012)
IFNγ, F3, CRAT, OGG	Whole blood	Hypermethylation	Madrigano et al. (2012)
EDARADD, TOM1L1, NPTX2	Saliva	Hypermethylation	Bocklandt et al. (2011) and Hannum et al. (2013)
ELOVL2	Whole blood, breast, kidney, lung	Hypermethylation	Garagnani et al. (2012) and Hannum et al. (2013)

used techniques for genome-wide analysis of DNA methylation include methylated DNA immunoprecipitation coupled with next-generation DNA sequencing (MeDIP-seq) and methylation micro-arrays from Illumina (including GoldenGate assay, the Infinium HumanMethylation27 BeadChip and the higher density Infinium HumanMethylation450 BeadChip), while the use whole-genome bisulfite sequencing is less popular due to computational challenges and high costs. Genome-wide approaches have been adopted in a number of age-focused studies, drastically increasing our knowledge of the epigenetic mechanisms of ageing (Table 2).

The first studies of this kind confirmed that age-associated methylation pattern is dependent on genomic context, with a propensity towards hypermethylation of CpG-island *loci* and towards hypomethylation of non-island CpGs (Christensen et al., 2009). aDMR in whole blood, sorted CD14+ monocytes and sorted CD4+ T-preferentially occur at bivalent chromatin domains (Rakyan et al., 2010), *i.e.* developmental gene promoters that in stem cells show both marks of active and inactive chromatin and that are often hypermethylated in cancer. These observations are consistent with the idea that ageing and cancer share common defects in cell differentiation and development pathways and account for the higher prevalence of cancer in the elder population. Interestingly, bivalent chromatin domains were recently identified as preferential sites of age-associated hypermethylation also in the

Table 2
Genome-wide studies on DNA methylation changes during ageing.

Technique	Samples	References
GoldenGate assay	217 non-pathological human tissues from 10 anatomic sites	Christensen et al. (2009)
Infinium HumanMethylation27 BeadChip	Whole blood from 93 different healthy females (31 twin pairs and 31 singletons) ranging from 49 to 75 years of age; sorted $CD14^+$ monocytes and $CD4^+$ T-cells from 25 subjects (20 females and 5 males) ranging from 20 to 69 years old	Rakyan et al. (2010)
Infinium HumanMethylation27 BeadChip	Skin biopsies from 20 females and 10 males ranging from 18 to 72 years old	Gronniger et al. (2010)
Infinium HumanMethylation27 BeadChip	Whole blood from 261 females ranging from 50 to 85 years old	Teschendorff et al. (2010)
Infinium HumanMethylation27 BeadChip	Saliva from 34 male identical twin pairs ranging from 21 to 55 years old	Bocklandt et al. (2011)
Infinium HumanMethylation27 BeadChip	CD34+ HPCs from umbilical cord blood and whole blood from 35 years old donors.	Bocker et al. (2011)
Infinium HumanMethylation27 BeadChip	Human prefrontal cortex from 108 subjects ranging from 0 to 83 years old	Numata et al. (2012)
Infinium HumanMethylation27 BeadChip	Whole blood from 21 female centenarians, their 21 female offspring, 21 offspring of both non-long-lived parents and 21 young women	Gentilini et al. (2012)
Infinium HumanMethylation27 BeadChip	Normal breast tissue from 38 individuals ranging from 19 to 75 years old	Fackler et al. (2011), Johnson et al. (2013) and Zhuang et al. (2012)
Whole-genome bisulfite sequencing	$CD4^+$ T-cells from 1 male newborn and 1 male centenarian	Heyn et al. (2012)
Infinium HumanMethylation450 BeadChip	$CD4^+$ T-cells from 19 newborns and 19 nonagenarians	Heyn et al. (2012)
Infinium HumanMethylation450 BeadChip	Whole blood from 32 mother–offspring couples ranging from 9 to 83 years old	Garagnani et al. (2012)
Infinium HumanMethylation450 BeadChip	Whole blood from 656 subjects ranging from 19 to 101 years old	Hannum et al. (2013)
Infinium HumanMethylation450 BeadChip	Peripheral blood leukocytes from 421 individuals ranging from 14 to 94 years old	Johansson et al. (2013)
Infinium HumanMethylation450 BeadChip	Whole blood from 965 subjects ranging from 50 to 75 years old	Florath et al. (2013)
Whole-genome bisulfite sequencing	Primary human epidermis samples from 5 young (18–24 years old) and 5 old (70–75 years old) subjects	Raddatz et al. (2013)

human brain (Numata et al., 2012; Watson et al., 2012), suggesting a process common to multiple tissue types.

Two studies compared DNA methylation patterns of centenarians and young subjects (Gentilini et al., 2012; Heyn et al., 2012) and confirmed that methylome encounters drastic changes during ageing. Interestingly, both studies highlighted that age associated DNA hypermethylation preferentially occurs at gene promoters involved in the regulation of developmental patterns. An increasing interest is directed towards the identification of epigenetic biomarkers that could predict the rate of ageing of an individual and possibly discern between chronological and biological age of the subject. Bocklandt and coworkers demonstrated that methylation of the promoters of *EDARADD*, *TOM1L1* and *NPTX2* genes linearly changes with age in saliva and that it can be used to predict age (Bocklandt et al., 2011). Garagnani et al. (2012) analyzed whole blood DNA methylation in 64 subjects from 0 to 99 years using the Infinium HumanMethylation450 BeadChip and identified the promoters of 3 genes, *ELOVL2*, *FHL2* and *PENK*, whose DNA methylation strongly correlates with age. Validation in a larger cohort of 500 subjects aged 0–99 confirmed a very high correlation with age for the 3 *loci* and showed that *ELOVL2* methylation progressively increases from 0% in newborns to almost 100% in nonagenarians. This result is unlikely to be ascribed to the alterations in blood cellular composition that occur with ageing, but it is probably more linked to epigenetic modifications that arise in progenitor stem cells, as previously suggested (Rakyan et al., 2010). *ELOVL2* was included in a quantitative model of ageing based on Infinium HumanMethylation450 BeadChip results on whole blood samples from 656 subjects from 19 to 101 years old (Hannum et al., 2013). This model demonstrated high accuracy in predicting individual's age and accounted for relevant factors in ageing, such as gender and genetic variants. Moreover, the authors reported that the model was successful in predicting chronological age also in tissues other than whole blood, including breast, kidney and lung.

This last observation prompts further considerations regarding the role of epigenetic modifications in tissue homeostasis and ageing. Indeed, ageing is a multi-factorial process in which, as conceptualized by the "mosaic of ageing" theory, different organs and tissues age at different rate and differently contribute to systemic ageing process (Cevenini et al., 2008). In this framework, although the identification of epigenetic changes that are consistent across different tissues (Christensen et al., 2009; Hannum et al., 2013; Koch and Wagner, 2011) can shed light on the general chromatin remodelling that occurs with ageing, equal emphasis should be placed in the discovery of tissue-specific epigenetic alterations. Consistently with these considerations, Hannum and colleagues reported that the entity of variation of the CpG sites included in the model differed among tissues, suggesting that epigenetic variations can reflect different ageing rates/mechanisms of different tissues (Hannum et al., 2013). In addition, when quantitative models of ageing were built independently for each available tissue, most of the markers were different, confirming tissue-specific patterns of epigenetic changes. Notably, *ELOVL2* was included in each tissue-specific model, suggesting that its methylation is uniformly affected by ageing in different tissues and thus highlighting its potential role as biomarker at the organism level.

2.4. Triggers and consequences of DNA methylation changes during ageing

In the previous paragraph we have reported several examples of age-associated changes which are systemic, that is they involve the same genomic region in a proportion of the cellular population large enough to be experimentally detectable (Gravina and Vijg, 2010). Some of the systemic changes arise in response to ageing-associated environmental factors, while others are interpreted as part of an "ageing program" which is part of a development (De Magalhães, 2012). Independently of the causes of these systemic changes, current knowledge suggests that global and *locus*-specific DNA methylation levels are promising markers of human ageing (Johnson, 2006) that can describe the inter-individual variability in functional decline and vulnerability to diseases in the elderly (Bellizzi et al., 2012; Horvath, 2013).

Together with systemic changes, the epigenome is also subjected to stochastic changes, *i.e.* random variations in the epigenetic patterns that have been called epimutations (Holliday, 1987; for an exhaustive review, see Gravina and Vijg, 2010). During ageing a relaxation of the mechanisms that maintain the epigenetic patterns occurs, leading to increased cell-to-cell heterogeneity in both gene expression (Bahar et al., 2006) and epigenetic markers (Fig. 1, *locus C*) (Bennett-Baker et al., 2003; Fu et al., 2008). Epimutations can contribute to the age-associated

epigenetic drift, which consists in the progressive divergence of the epigenetic patterns between individuals with time. In a seminal paper, Fraga and colleagues showed that the methylomes of young twin couples are more similar than those of elderly couples (Fraga et al., 2005). Based on these observations, several studies have analyzed the age-associated divergence in DNA methylation patterns between twins (Heijmans et al., 2007; Pirazzini et al., 2012; Talens et al., 2012) and have shown that (i) some genomic *loci* are epigenetically more stable than others; (ii) the ability to maintain DNA methylation patterns differs among different individuals and it is genetically determined, as suggested also by longitudinal family-based models (Bjornsson et al., 2008). Regarding this epigenetic drift, it is difficult to discern the relative contributes of environmental-driven and stochastic changes in DNA methylation, but recent experiments using laboratory animals, whose environmental conditions can be strictly monitored, have shown that chromatin entropy increases with age, while chromatin homogeneity decreases, pointing for a key role of epimutations in the ageing process (Pantic et al., 2012).

Age-associated changes in DNA methylation are triggered, at least in part, by alterations in the expression/activity of DNA methyltransferases (DNMTs). Indeed, several authors reported that a decrease in Dnmt1 and Dnmt3a expression occurs with ageing, while the levels of Dnmt3b follow the opposite trend (Casillas et al., 2003; Liu et al., 2009; Lopatina et al., 2002). Accordingly, while Dnmt1 knock-down promotes age-related phenotypes (Liu et al., 2011; Ray et al., 2006), the rescue of Dnmt3a2 levels restores cognitive function in aged mice (Oliveira et al., 2012). Beside DNMTs expression, many other mechanisms can affect the maintenance of DNA methylation during ageing. For example, other epigenetic modifications and proteins involved in the DNA damage response can modulate the access of DNMTs to target sequences and affect the methylation status of CpG dinucleotides (Ciccarone et al., 2012; Guastafierro et al., 2008; Jin and Robertson, 2013).

Future researches should address not only the causes of DNA methylation changes during ageing, but also their consequences. While the contribution of DNA methylation changes at promoters should be evaluated on the basis of the function of the specific gene, the global hypomethylation of the genome could exert ageing-effects by promoting genomic instability and activation of transposable elements (Alexeeff et al., 2013; Barbot et al., 2002; De Cecco et al., 2013).

3. Epigenetic mechanisms in anti-ageing nutritional interventions

Epigenetic patterns are set during gametogenesis, fertilization and *in utero* development and they are particularly responsive to environmental factors in these periods. Accordingly, a number of studies both on murine and human models have highlighted the epigenetic consequences of nutritional interventions during preconceptional and *in utero* periods and their transgenerational heritability. These studies have been deeply reviewed in recent manuscripts (Barnes and Ozanne, 2011; Gabory et al., 2011; Lillycrop and Burdge, 2012) and will not be reported here. On the contrary, we will specifically focus on the epigenetic plasticity in the adulthood, reporting available evidences that diet interventions can alter epigenetic patterns later in life and positively affect ageing and lifespan. The example of honeybees is a proof of principle of this deep relationship between after birth diet, epigenetics and ageing, as female larve develop in short-living workers or in long-living queens according to DNA methylation modifications induced by the type of feeding (*i.e.* beebread or royal jelly) (Kucharski et al., 2008). Examples of diet-mediated changes in epigenetic modifications have been reported also in humans.

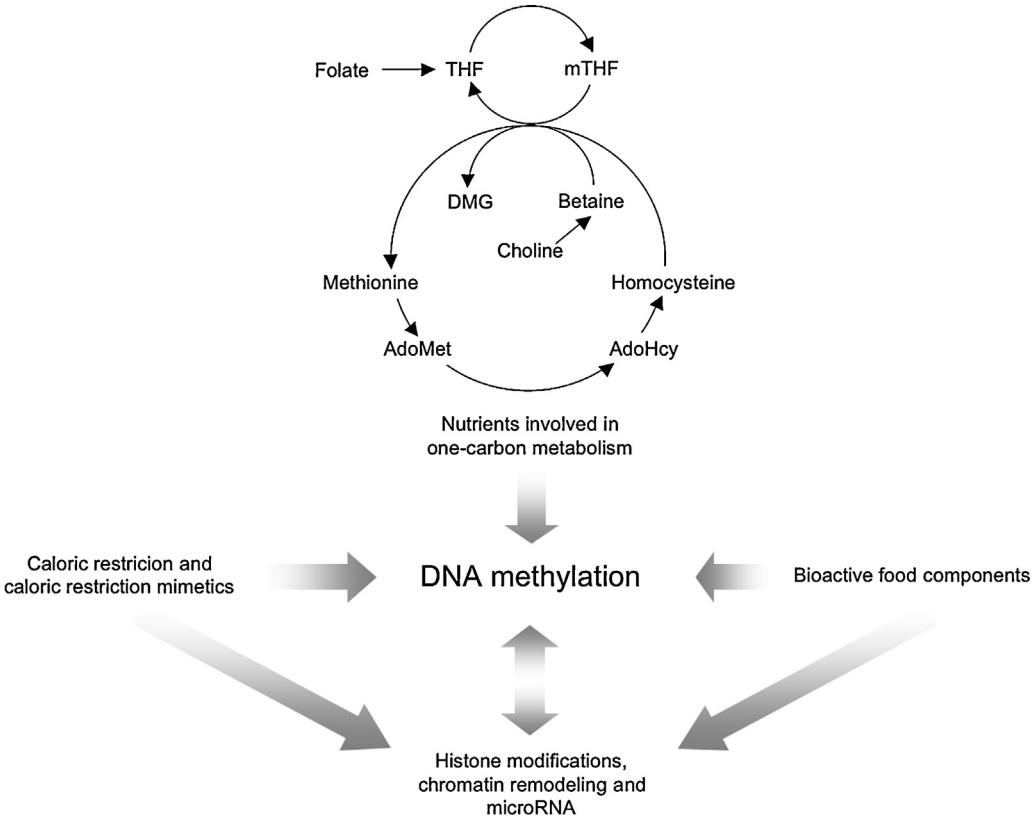


Fig. 2. Nutritional interventions, including caloric/dietary restriction, nutrients involved in one-carbon metabolism and bioactive food components, can modify DNA methylation and/or other epigenetic modifications, that in turn can affect DNA methylation patterns.

The extent of age-related variations in leukocytes global DNA methylation was associated with the daily intake of carbohydrates, lipids, vitamin B6 and magnesium (Gomes et al., 2012). A recent study considered the effect of a diet rich in vegetables and plant oil containing polyunsaturated fatty acids on the methylation of *MLH1*, a gene involved in mismatch repair and in maintenance of genomic integrity (Switzeny et al., 2012). The intervention resulted in significant increase in *MHL1* CpG methylation that was not associated to variations in its expression but that probably reflected a decrease in reactive oxygen species (ROS) and in the demethylating effects of DNA damage related enzymes.

In the following paragraphs we will summarize the effects of dietary restriction and different bioactive compounds on DNA methylation changes during ageing (Fig. 2).

3.1. Caloric and dietary restriction

Although the notion that limited food intake can positively affect health and ageing has deep roots in history, the first scientific demonstration that reducing food intake significantly increases rodents life span dates back to 1935 (McCay et al., 1989). Since then, a number of dietary modifications have been experimented. Caloric restriction (CR) consists in the reduction of caloric intake, typically by the 20–40% respect to the *ad libitum* intake of the species, without leading to malnutrition and deficiency in micronutrients. CR is distinct from dietary restriction (DR), in which one or more macronutrients is reduced or removed without affecting the total caloric intake, and from regular fasting such as alternate-day fasting (ADF) (Trepanowski and Bloomer, 2010).

The effects of these dietary modifications on health and longevity have been investigated in many species, from yeast to primates, leading to exciting but sometimes controversial results (Speakman and Mitchell, 2011). For example CR has proven to extend lifespan in many animal models and to effectively protect against age-related diseases such as diabetes and cancer in murine models. However, some evidences indicate that these outcomes are not universal, and that factors such as genetic background (Liao et al., 2010) and initial basal metabolic rate (Brzek et al., 2012) can affect this phenomenon. The results from two long-term studies on the effects of CR in non-human primates have been recently published (Colman et al., 2009; Mattison et al., 2012). Both studies reported health benefits such as reduction of the incidence of diabetes and of cancer in calorie restricted rhesus monkeys respect to controls, but were discordant on the effects of CR on life span extension. These apparently conflicting results probably arise from the design of the two studies that differ for the diet composition of both caloric restricted and control monkeys. Data on the effects of CR on human models are also available. For example, a randomized clinical study on non-obese, sedentary humans, demonstrated that a 20% reduction in calorie intake for 12 months decreases visceral fat mass, improves insulin sensitivity, increases plasma adiponectin concentration, reduces circulating inflammatory markers, decreases plasma triiodothyronine levels, and reverses some of the age-related deterioration in cardiac diastolic function (Fontana et al., 2010a; Stein et al., 2012; Weiss et al., 2008). Ongoing studies on human subjects who voluntarily follow a caloric restricted diet will contribute to clarify not only the potential anti-ageing effects of CR (Mercken et al., 2013), but also its contribution to the slowdown of age-associated diseases onset (Stein et al., 2012).

The molecular mechanisms by which CR exerts its beneficial effects are poorly understood, but two nutrient-sensing signalling networks, the insulin/IGF (insulin-like growth factor) and the mTOR (mammalian target of rapamycin) pathways, seem to respond to CR in the control of cell growth and ageing pathways (Speakman and Mitchell, 2011). In addition, some authors have suggested the involvement of Sirtuins, a group of NAD-dependent histone acetylase, but their role in linking CR to longevity is, at least in mammals, still controversial (Baur et al., 2010; Fontana et al., 2010b). It is worth to note that both mTOR (Grummt and Voit, 2010; Murayama et al., 2008) and Sirt-1 (Salminen and Kaarniranta, 2009) link the cellular energy condition with the epigenetic status of ribosomal DNA, which in turn plays a role in longevity in model organisms (Larson et al., 2012).

Paradoxically, the possibility that CR could exert anti-ageing effects by reversing epigenetic alterations during ageing and age-related diseases has been relatively poorly investigated (Li et al., 2011b; Vaquero and Reinberg, 2009). In this sense, it is relevant the recent study by Ions et al. (2012) who analyzed available public databases and reported a significant overlap between the genes differentially expressed in response to CR and those whose methylation varies during ageing. Two independent studies reported that CR increases methylation levels of *MYC* in mice liver (Miyamura et al., 1993) and of *HRAS* in pancreatic acinar cells (Hass et al., 1993), reverting the trend towards age-associated hypomethylation of the two oncogenes. More recently, CR was shown to counteract the increase in Dnmt3a expression in the hippocampal dentate gyrus that is observed in aged mice (Chouliaras et al., 2011) and to prevent the age-related changes in DNA methylation levels (Chouliaras et al., 2012). In an *in vitro* model of CR, glucose restriction increased Dnmt1 activity and induced hypermethylation of a E2F-1 binding site in the promoter of the *CDKN2A* gene, down-regulating its expression and extending cell lifespan (Li et al., 2010a,b). As the tumor suppressor p16 accumulates during senescence, its decreased expression could contribute to CR-mediated longevity *in vivo*. The same group reported that p16 reduction upon glucose restriction was ascribable to the combined effect of DNA methylation and Sirt-1 mediated histone deacetylation, supporting a link between these two chromatin modifications (Li et al., 2011b). In line with these evidences, manipulation of Sirt-1 levels by overexpression or silencing of the gene in Caco-2 cells altered the methylation of six genes whose expression is responsive to CR (Ions et al., 2012). Epigenetic data on the effects of CR in humans are limited. In a cohort of overweight/obese postmenopausal women before and after a reduced energy intake intervention over 6 months, epigenomic analysis in subcutaneous adipose tissue showed significant hypermethylation of three *loci* (chromosomes 1p36, 4q21, and 5q13) in responders to the treatment (Bouchard et al., 2010). In a similar study, two CpG sites in the *ATP10A* and in the *WT1* genes were hypermethylated in peripheral blood mononuclear cells (PBMCs) in overweight/obese men after 8 weeks dietary intervention (Milagro et al., 2011). As obesity is a major risk factor for age-related diseases, these studies can shed light on the mechanisms by which CR can potentially exert its anti-ageing effects.

As CR is not a feasible nutritional intervention in the general population, researchers have looked for mimetics of CR that do not imply food restriction (Ribarič, 2012). Protein restriction, and in particular methionine restriction, seems to be an effective alternative to CR in lifespan extension (Miller et al., 2005). Methionine is the precursor of S-adenosylmethionine (AdoMet) in one-carbon metabolism (see below), but the outcomes of its restriction on DNA methylation have been considered only in few studies. For example, 7 weeks 40% methionine restriction induced DNA hypomethylation in rat heart (Sanchez-Roman et al., 2011) but not in rat liver (Sanchez-Roman et al., 2012), suggesting a tissue-specific effect. Two other potential CR mimetics are rapamycin, a bacterial macrolide with antifungal and immunosuppressant activities inhibiting mTOR, and resveratrol, a polyphenol (see below) enriched in the skins of red grapes that can activate Sirt-1 (Barger et al., 2008). However, also in this case the epigenetic mechanisms by which they can exert an anti-ageing

effect have not been assessed yet. These compounds, together with other bioactive food components such as green tea extracts (see below), are being tested by the Interventions Testing Program, a multi-institutional program promoted by the National Institute of Ageing that aims at identify treatments with an anti-ageing effect in mice.

3.2. Nutrients involved in one-carbon metabolism

The one-carbon metabolism is a complex biochemical pathway, also known as the homocysteine cycle (Fig. 1). One of the intermediates of this metabolism is AdoMet, which represents the unique methyl donor in all the DNMTs-mediated DNA methyltransferase reactions in eukaryotes. Various nutrients can regulate the one-carbon metabolism, acting both as sources of methyl donors (for example methionine, choline, betaine and serine) and as co-enzymes (for example vitamins B2, B6, B12 and folate) (Kim et al., 2009). Although the role of these nutrients in physiological and pathological processes has not been completely studied, many evidences support that they can affect DNA methylation changes during ageing.

With the term "folate" we indicate a group of water-soluble B vitamins, also known as B9, that carry a methyl group which can be used for the AdoMet synthesis. As folate and its derivates directly affect reproductive success and survival early in life, natural selection has worked to protect folate levels and preserve the integrity of folate metabolism. Humans cannot synthesize folate, which therefore has to be supplied through the diet. Folate-rich foods include dark green leafy vegetables, asparagus, broccoli, strawberries, liver, organ meats and legumes. In addition, folate can be supplemented with diet and fortified foods in the form of the synthetic molecule folic acid. Many studies have addressed the relationship between DNA methylation and folate intake during pregnancy (Ciappio et al., 2011; Kim et al., 2009) and cancer (Stefanska et al., 2012), while the effects of folate administration on ageing are less well characterized. Age associated decrease of DNA methylation in mouse colon was reduced by folate administration, which induced also *CDKN2A* hypermethylation (Keyes et al., 2007). These results are in agreement with those recently achieved by Li and colleagues (Li et al., 2010a,b). In this study, T lymphocytes from healthy subjects aged 22–81 were cultured in a folate-depleted medium. T cells from 50 years or older donors showed hypomethylation of genes that are activated during ageing, such as *KIR* and *CD70* genes, suggesting a synergic effect between age-associated DNA methylation changes and folate levels. A number of studies confirm that a folate-poor diet is associated to global DNA hypomethylation and to an increased risk for chronic diseases (Axume et al., 2007; Pufulete et al., 2005; Rampersaud et al., 2000; Shelnutt et al., 2004). On the other side, folate supplementation increased global DNA methylation levels in rectal mucosa of a subset of patients with colonic adenomas (Cravo et al., 1998b) and reduced microsatellite instability in the mucosa of ulcerative colitis patients (Cravo et al., 1998a). The effect of folic acid administration on site specific DNA methylation was analyzed in 1000 colorectal mucosa biopsies (Wallace et al., 2010). In this study estrogen receptor gene α (*ESR1*) and *SFRP1* methylation levels, that increase with age and may predispose to colorectal cancer, were positively correlated with folate levels in red blood cells, posing safety concerns regarding folic acid administration in healthy adults.

Beside folate, other intermediates of one-carbon metabolism can affect DNA methylation and ageing. Vitamin B12 (cobalamin) is a water-soluble vitamin that acts as co-enzyme of methionine synthase. Low vitamin B12 concentration reduces methionine synthase activity and prevent folate metabolism, resulting in a decrease in AdoMet concentration and therefore in DNA methylation. Vitamin B-rich foods include eggs, cheese, beef, lamb, seafood and milk. Deficiencies in vitamin B12 can lead to elevated DNA damage and altered DNA methylation (Blount et al., 1997; Lindahl and Wood, 1999) and to an increase of homocysteine level, a risk factor for cardiovascular disease in particular for the elderly (Krishna et al., 2013; Van Dijk et al., 2013).

It should be noted that DNA methylation responses to B vitamin levels may depend on various factors, including genetic background (Axume et al., 2007; Friso et al., 2002; Shelnutt et al., 2004), extent and duration of the depletion/supplementation and the tissues under investigation. For example, Hubner et al. recently reported that one-year vitamin B supplementation did not affect AdoMet plasma levels and long interspersed element-1 (LINE-1) methylation levels in blood cells (Hübner et al., 2013).

3.3. Bioactive food components

Polyphenols constitute a heterogeneous group of natural substances, chemically formed by several phenolic rings, which are particularly known for their positive effects on human health and are enriched in many plant foods. The intake of polyphenols in the human diet varies greatly depending on the type, the quantity, the quality of vegetables consumed and the chosen cooking method that can reduce considerably the polyphenolic content of the food. Polyphenols can exploit a wide range of activities, including antioxidant, antiatherogenic, anticarcinogenic and anti-inflammatory effects. Their antioxidant properties are recognized worldwide: polyphenols may protect cell constituents against oxidative damage and, therefore, limit the risk of degenerative diseases associated to oxidative damage that can be related to both normal cellular metabolism and stressful events, such as radiation, smoke, pollutants, UV rays, emotional and physical stress, chemical additives, viral and bacterial attacks. Numerous studies have shown that, when added to the diet, polyphenols limit the development of cancers (Darvesh and Bishayee, 2013; Gokbulut et al., 2013; Gollucke et al., 2013; Tabrez et al., 2013), cardiovascular diseases (Khurana et al., 2013; Li and Förstermann, 2012), neurodegenerative diseases (Cimini et al., 2013), diabetes (De Bock et al., 2013; Patel et al., 2012), obesity (Siriwardhana et al., 2013; Tian et al., 2013) and osteoporosis (Oka et al., 2012; Sacco et al., 2013).

Regarding the effects of polyphenols on DNA methylation, studies have been focused on different pathologies, including cancer (Vanden Berghe, 2012). Several mechanisms by which polyphenols concur to cancer inhibition have been reported. These mechanisms include the reduction of hypermethylation status that characterizes genes such as *CDKN2A* and *RARB*. Polyphenols may act both directly through physical interaction with Dnmt1 catalytic site and indirectly by the reduction of intracellular concentration of AdoMet, that in turn reduces Dnmt1 activity (Fang et al., 2007; Li and Tollefsbol, 2010).

Lee et al. (2005) evaluated the effects of several tea catechins and bioflavonoids on DNA methylation, finding that each of the tea polyphenols considered [catechin, epicatechin, and (−)-epigallocatechin-3-O-gallate (EGCG)] and bioflavonoids (quercetin, fisetin, and myricetin) inhibited Dnmt1-mediated DNA methylation in a concentration-dependent manner. Many other authors observed the inhibition of DNMTs activity as a consequence of different polyphenols diet supplementation (Fu and Kurzrock, 2010; King-Batoon et al., 2008; Meeran et al., 2010), with tea polyphenols and genistein showing the strongest inhibitory effect.

EGCG is a polyphenolic catechin found in green tea with antioxidant properties that can induce epigenetic changes by inhibiting enzymes involved in DNA methylation and by modifying histone acetylation (Nandakumar et al., 2011; Pandey et al., 2010). Wong et al. reported that the diet EGCG intake could inhibit

DNMTs activity and could induce Foxp3 and IL-10 expression in CD4(+) Jurkat T cells at physiologically relevant concentrations *in vitro*, showing a role in controlling the development and function of regulatory T cells (Treg). Furthermore, authors found that mice treated with EGCG *in vivo* had a significant increase in Treg frequencies, suggesting, on the whole, that EGCG can induce Foxp3 expression and increase Treg frequency through an epigenetic-based mechanism (Wong et al., 2011). Qin et al. conducted a study on 34 healthy premenopausal women consuming isoflavones for one menstrual cycle to test if soy isoflavones have dose-related estrogenic and methylation effects, finding that this kind of polyphenols caused significant changes in the methylation levels of two cancer related genes (*RARB* and *CCND2*) in the breast, according to the circulating levels of genistein (Qin et al., 2009).

Although the data collected until now are encouraging, future studies should specifically consider the effects of polyphenols on DNA methylation in an anti-ageing perspective.

4. MicroRNAs and ageing: the *inflamma-miRs*

During ageing a low-grade systemic inflammation characterized by elevation of circulating acute-phase proteins and proinflammatory cytokines (mainly IL-6 and TNF-α) occurs, a condition that we have previously designated *inflammaging* (Franceschi, 2007; Franceschi et al., 2000). Such inflammatory imbalance is associated to frailty and the development and progression of age-related conditions that include cardiovascular disease (CVD), type 2 diabetes mellitus (T2DM), neurodegenerative diseases, sarcopenia and cancer (Cevenini et al., 2013; Franceschi, 2007; Vasto et al., 2007).

The role of DNA methylation in inflammaging has not been disentangled yet, but some evidences point for an association between age-related global DNA hypomethylation and increased expression of a number of inflammatory markers (Agrawal et al., 2010; Alexeeff et al., 2013; Baccarelli et al., 2010b). Beside DNA methylation, researchers have examined the function of other epigenetic modifications in the establishment of the inflammaged phenotype, focusing in particular on microRNAs (miRs).

MiRs are a broad class of small, non-coding RNAs that exert a powerful gene regulatory role, acting both as repressor as well as activators, mainly at a post-transcriptional level (Breving and Esquela-Kerscher, 2010).

The primary transcripts of miRs (pri-miRs) are transcribed by RNA polymerase II–III in the nucleus. Pri-miRs are processed by the Drosha/DGCR8 enzyme complex into 70 base pair precursor and then transported into the cytoplasm, where RNase III enzymes, Dicer and Loquacious, process them into about 22-nucleotide miR duplexes with guide and passenger strands. The guide strand functions as a mature miR and is incorporated into an RNA-induced silencing complex (RISC), containing an Argonaute (Ago) protein. MiRs guides RISC to recognize target sequences located mainly in the 3′UTR of mRNAs, which lead to the inhibition of translation or degradation of mRNA (Bartel, 2009). Since the specificity of miR targeting is mediated only by 6–11 nucleotides, a single miR can target hundreds of mRNAs (Inukai and Slack, 2013; Park and Kim, 2013). MiRs have been reported to act through autocrine and/or paracrine mechanisms (Kumarswamy et al., 2011; Raitoharju et al., 2011). In addition, circulating miRs can act as hormones, eliciting a systemic response (Wahlgren et al., 2012). Recent studies show that transfer of nucleic acids, including miRs, can be an important means of intercellular communication, occurring both by direct cell-cell contact, for instance *via* gap junctions, or by cell-contact-independent mechanisms, including release of microvesicles into surrounding tissue (Collino et al., 2010; Hosoda et al., 2011) or the blood stream. Plasma exosomes can deliver exogenous short interfering RNA, including miRs to monocytes and lymphocytes (Wahlgren et al., 2012).

Even if a single miR can target hundreds of mRNAs, groups of miRs can induce regulation of specific biological processes by acting in a co-ordinated manner on pathways of functionally related genes (Cloonan et al., 2011). Thus, miRs have recently been indicated as regulators of a number of cellular processes such as apoptosis, differentiation, cell cycle, and immune functions, and their expression can be altered by a wide spectrum of environmental factors including drugs, virus and bacterial pathogens, cigarette smoking, alcohol, sleep, exercise, stress, radiation and nutrition.

It is worth stressing that exceptionally long survival requires dynamic preservation of optimal levels of physiological variables, and that the mean levels of many biomarkers of ageing are not stable, but change in the course of life (Spazzafumo et al., 2013). Thus, a peculiar modulation of miRs expression might contribute to efficient homeostasis in human ageing (Grillari and Grillari-Voglauer, 2010; Inukai and Slack, 2013). Recently we discussed the pleiotropic effects exerting of some miRs on pathways related to inflammation, senescence, and carcinogenesis (Olivieri et al., 2013a). A number of miRs have been reported to play a role in modulating cellular senescence and inflammatory responses (*SA-miRs* and *inflamma-miRs*, respectively). Given the interest elicited by this new area of research, more and more miRs involved in the modulation of inflammation and senescence are expected to be identified in the near future. A clear link between miRs and longevity has been demonstrated in *C. elegans*, suggesting that miRs could have relevant role also in regulation of human lifespan and ageing process (Ibáñez-Ventoso and Driscoll, 2009). However, only few studies have compared the miRs expression profile in tissues and blood of old and young organisms of different species (Table 3). Interestingly, differences in the circulating miRs have been found in a variety of age-associated diseases. The origin of circulating miRs is not completely defined, but senescent cells emerge as a possible source of such secreted miRs, suggesting that these miRs might contribute to the functional decline observed during ageing (Weilner et al., 2012). Overall the results of these studies supported the hypothesis that achievement of extreme longevity probably requires a special gene expression regulation involving inflamma-miRs (ElSharawy et al., 2012; Gombar et al., 2012; Olivieri et al., 2012a; Serna et al., 2012). The prototype of inflamma-miRs are miR-155, miR-21, miR-146a and miR-126 (Olivieri et al., 2013b, 2012b; Quinn and O'Neill, 2011). Under physiological conditions, transcription of the inflamma-miRs is maintained at basal levels; however, as soon as the pro-inflammatory signalling are promoted, their expression is strongly co-induced through a mechanism that is mainly NF-kB-dependent (Boldin and Baltimore, 2012; Olivieri et al., 2013b). Although the importance of inflamma-miRs in the regulation of the innate immune response and the acquisition of senescence-associated secretory phenotype during cellular senescence is clearly demonstrated, the molecular mechanism of their action turned out to be much more complex than initially thought (Olivieri et al., 2012a; Grillari and Grillari-Voglauer, 2010).

5. Epigenetic diets and inflammaging

A growing number of evidences support the role of diet in slow down and at least in part revert inflammaging (Proietti et al., 2009). As diet is known to influence the expression of miRs, nutrients with antioxidant and anti-inflammatory effects were investigated to verify their ability to modulate the expression of specific miRs, among which inflamma-miRs. A potential role of miR-126 in the anti-inflammatory properties of polyphenols from red wine was observed in human colon myofibroblasts cells (CCD-18Co)

Table 3
Main miRs associated with organismal ageing.

MiRs	Organism	Tissue	References
miR-20a, miR-106a, miR-126, miR-155	Humans	Whole blood	ElSharawy et al. (2012)
miR-21	Humans	Plasma	Olivieri et al. (2012b)
miR-126	Humans	Plasma	Our unpublished data
miR-21	Humans	PBMCs	Serna et al. (2012)
miR-21	Humans	B lymphocytes	Gombar et al. (2012)
45 miRs	Male mice	Serum	Dhahbi et al. (2013)
miR-34a	Mice	Plasma, PBMCs, brain	Li et al. (2011a)
93 miRs	Mice	Brains	Inukai et al. (2012)
miR-34a, miR-93, miR-214, miR-669c, miR-709	Mice	Liver	Smith-Vikos and Slack (2012)
miR-22, miR-101a, miR-720, miR-721	Mice	Brain	Smith-Vikos and Slack (2012)
miR-7, miR-468, miR-542, miR-698, miR-124a, miR-181a, miR-221, miR-382, miR-434, miR-455	Mice	Skeletal muscle	Smith-Vikos and Slack (2012)
65 miRs, including miR-21	Mice	Heart	Zhang et al. (2012)
let-7 miR cluster	Mice	Heart	Cao et al. (2012)
miR-34	Drosophila	Brain	Liu et al. (2012)
miR-124	Different species	Different tissues	Dallaire et al. (2012)

(Angel-Morales et al., 2012). MiR-126 targets NF-kB signalling (Olivieri et al., 2013b) and mRNA levels of adhesion molecules, such as Icam-1, Vcam-1, and Pecam-1.

The anti-inflammatory properties of the flavonol quercetin have been intensively investigated using *in vitro* cell systems and are to a great extent reflected by changes in the expression of inflammatory markers (Boesch-Saadatmandi et al., 2012). Quercetin protected CCD-18Co myofibroblasts against ROS in part by increasing activity of antioxidant enzymes and in part by inducing an up-regulation of miR-146a, known as a negative regulator of pro-inflammatory NF-κB activation, thus protecting CCD-18Co from inflammation (Noratto et al., 2011). Moreover, hepatic miR-122 and miR-125b concentrations, previously involved in inflammation (miR-125b) and lipid metabolism (miR-122), were increased by dietary quercetin supplementation, suggesting that miRs modulation can contribute to the gene-regulatory activity of quercetin *in vivo* (Boesch-Saadatmandi et al., 2012). Resveratrol and quercetin in combination are able to decrease the generation of ROS in colon cancer cells decreasing in turn oncogenic miR-27a expression (Del Follo-Martinez et al., 2013).

Interestingly, the stimulation of specific signalling pathways can occur in the cross-talk between probiotic bacteria and gut epithelium cells, which can help to explain the adjuvant properties of probiotic lactobacilli (Vizoso Pinto et al., 2009). Strain specific properties of probiotics in providing supportive health effects in the immune system and the gastrointestinal tract have been widely investigated *in vivo* and *in vitro*, suggesting that specific strains could even modify the immune response at post-transcriptional level by modifying miRs expression in dendritic cells (DCs) (Giahi et al., 2012). Probiotic strains can affect toll like receptor 4 (TLR4) expression in a down-regulatory direction, reducing in turns inflammation level. Inactivated strains of *Lactobacillus rhamnosus GG* (LGG) in human DCs can induce a significant down-regulation of miR-146a expression and a concomitant up-regulation of miR-155 in DCs, which is consistent with the down-regulation of p38MAPK (Giahi et al., 2012). It was reported that these two seemingly co-induced regulatory RNAs, miR-146a and miR-155, dramatically differ in their induction behaviour under different stimuli strengths and act non-redundantly through functional specialization.

In conclusion, inflamma-miRs seem to be modulated by nutrients, highlighting the potential role of probiotic intervention in reducing the pro-inflammatory status associated with human ageing.

6. A notable example: cardiovascular disease

In the developed countries most of the elderly are affected by age-related chronic diseases, which share an inflammatory background. These pathologies share features of accelerated ageing, which from an epigenetic point of view consist in global hypomethylation and site-specific hypermethylation of the genome as discussed above. The ongoing epigenetic characterization of age-related diseases has two important consequences: (1) allows the identification of epigenetic markers of biological age that can be used to early detect the pathologies; (2) provides an interventional target. Indeed, most of the age-related diseases are multi-factorial pathologies in which environmental lifestyle, and in particular nutritional habits, plays a pivotal role. Based on these considerations, it is clear that the anti-ageing dietary interventions provide a powerful tool to counteract or slowdown the age-related diseases onset. In this paragraph, we discuss cardiovascular diseases as a notable example.

Recent studies reported the function of epigenetic mechanisms in CVDs (Baccarelli et al., 2010a; Handy et al., 2011; Turunen et al., 2009; Udali et al., 2013). Epigenetic modification can regulate gene expression of key pathways related to coronary artery disease development in human and animal models (Baccarelli et al., 2010a; Creemers et al., 2012; Hiltunen et al., 2002a,b; Laukkanen et al., 1999) or specifically affect the expression of genes involved in its major thrombotic complication, myocardial infarction (Friso et al., 2012; Handy et al., 2011). Changes in both genome and site-specific DNA methylation have been described in CVD (Friso et al., 2012, 2008) although their specific function is still to be largely investigated (Baccarelli et al., 2010a).

Hiltunen et al. (2002b) described the first report of a global hypomethylation, in advanced atherosclerotic lesions in humans as well as mouse and rabbit models. However, whether global DNA hypomethylation has a causal role or is it a consequence of atherogenesis is not yet known. Global DNA hypomethylation was described also in atherosclerosis-prone *APOE*−/− mice compared to C57BL/6 controls at early disease stages leading to the hypothesis that aberrant DNA methylation may precede the onset of disease even of several years, as described for cancer disease (Ehrlich, 2006; Lund et al., 2004).

Some studies analyzed the levels of global DNA methylation in association with plasma concentrations of homocysteine (Hcy), a biomarker for vascular disease (Handy et al., 2011). Castro et al. (2003) described a significant decrease in DNA global methylation in white blood cells DNA of patients with vascular disease and higher Hcy, compared to a control group. The higher plasma Hcy paralleled with increased concentrations of S-adenosylhomocysteine (AdoHcy) (Castro et al., 2003), the intracellular precursor of Hcy with a well-known function as a strong inhibitor of DNMTs (Yi et al., 2000). Considering that both Hcy (Girelli et al., 1998; Jacques et al., 1996) and DNA methylation are dependent on folate concentrations through a gene-nutrient interaction model

(Friso et al., 2013, 2002) it seems of high interest to explore whether nutritional regulation within one-carbon metabolism may modify the risk of CVD through DNA methylation. A recent observation reported that a nutrient-gene interaction within folate metabolism, not only affect the levels of global DNA methylation but also the risk of cancer disease through this epigenetic phenomenon (Friso et al., 2013). Whether this finding may also apply to CVD is still to be evaluated. Using a surrogate marker for global DNA methylation, LINE-1 methylation was evaluated as a possible marker for cardiovascular risk (Baccarelli et al., 2010c). Different results have been reported, however, when global methylation in peripheral blood leukocytes was evaluated by measuring Alu and juxtacentromeric Satellite 2 repetitive elements (Kim et al., 2010) whose methylation highly correlate with global methylation of DNA (Weisenberger et al., 2005). The precise reason of the inconsistencies observed in those studies may be found in either the different study design but also in the techniques utilized for measuring DNA methylation, namely bisulfite treatment-based methods (Baccarelli et al., 2010c; Kim et al., 2010), restriction enzyme digestion (Castro et al., 2003; Lund et al., 2004; Yi et al., 2000) or HPLC-based methods (Hiltunen et al., 2002b) which may indeed account for the results variability.

Several studies reported the role of methylation at specific gene-sites in CVD. One of the first report refers to the epigenetic regulation through methylation of the estrogen receptors α (*ESR1*) and β (*ESR2*) genes whose functions have been largely described for CVD (Mendelsohn and Karas, 1999). Post and colleagues observed higher methylation levels of *ESR1* promoter region in human coronary atherosclerotic plaques compared to normal vascular tissues (Post et al., 1999). The *ESR2* methylation at promoter region was also described in human atherosclerotic lesions where the increase in DNA methylation correlated with decreased *ESR2* expression (Kim et al., 2007).

Gene-specific DNA methylation in human atherosclerotic lesions was also studied by analyzing the gene coding for 15-lipoxygenase (*ALOX15*), a lipid peroxidating enzyme strongly expressed in atherosclerotic plaque (Hiltunen et al., 2002b). Using a bisulfite-sequencing approach a hypomethylation in CpG promoter region sequences was observed that possibly account for the increased *ALOX15* gene expression.

The role of methylation in CVD was also studied at coagulation factor 7 gene (*F7*) promoter site considering that factor VII plasma concentrations are a strong risk factor for coronary artery disease. *F7* promoter was evaluated in the DNA extracted from PBMCs of patients affected by coronary artery disease (CAD) compared to CAD-free subjects where *F7* promoter hypomethylation correlated with higher plasma concentrations of FVIIa and higher risk for CAD, therefore highlighting a possible role for methylation at *F7* promoter site and CAD risk (Friso et al., 2012).

Regarding the inter-relationships among nutritional factors, DNA methylation and CVD risk, van Straten and colleagues recently showed that a low protein diet influences adult lipid metabolism through DNA methylation by altering, among several differentially methylated genes, liver X-receptor alpha (*NR1H3*) (Van Straten et al., 2010). This gene is involved in the control of cholesterol and fatty acid metabolism and it is hypothesized to contribute to the risk of CVD.

A very interesting model is that proposed by Lillycrop and colleagues who showed, using a rodent model, that a protein restricted diet induces a lower promoter methylation status in glucocorticoid receptor (*NR3C1*) and peroxisomal proliferator-activated receptor α (*PPARA*): both genes are related to cardiovascular disease also in humans by being involved in the regulation of blood pressure and in the lipid and carbohydrate homeostasis, respectively (Lillycrop et al., 2008, 2005).

A growing interest is emerging for the role of miRs and CVD mostly due to the hypothesis of being potentially useful biomarkers of disease development or progression (Bauersachs and Thum, 2011). In this view, miRs have been largely explored in both plasma and serum of stable CAD patients mostly to evaluate whether miRs usually expressed in endothelial cells (miR-126, miR-17, miR-92a), in smooth muscle cells (miR-145) or in inflammatory response-associated cells (miR-155) may also serve as possible CVD markers (Fichtlscherer et al., 2010). However, whether specific miRs plasma levels will be useful markers of disease diagnosis or complication in the near future is not clearly known. Moreover, it is still to be explored whether miRs profiles may be modifiable by environmental or nutritional factors as occurring to other epigenetic features of DNA. The latter issue is of great importance considering that by modulating the environment or nutritional setting it could be possible to modify disease risk or adverse progression.

7. Conclusions

Unravelling the molecular basis of successful ageing is a challenging but exciting goal of current research. Epigenetics is gaining a central position in this scenario, as it provides at the same time mechanistic insights into the ageing process and scientific basis for interventional strategies. A growing number of studies demonstrate that age-associated epigenetic variations can be positively affected by physical exercise, lifestyle habits and diet. However, it should be kept in mind that we are still at the beginning of such studies, and that much has to be done in order to disentangle the complex relationship between epigenetics, nutrition and ageing. For example, the tissue (or even cellular) specificity of age-associated epigenetic changes and their functional contribution to ageing deserve deeper investigation, together with the molecular mechanisms by which dietary habits and food components affect the epigenetic patterns. Among all tissues, adipose tissue appears to be critical for the tight connections between over-nutrition and inflammatory status, being a paradigmatic example of the powerful effects of the underlying epigenetic changes (see the paper by Zamboni et al., 2013, in this special issue). In addition, the stability of the epigenetic modifications induced by nutritional interventions should be evaluated over time. Finally, although the use of murine models is a valuable tool in these studies, the availability of opportune studies on humans, such as those provided by the European Project NU-AGE (see Box 1), appears unavoidable to make in the next future epigenetic diets a valid interventional approach for successful ageing.

Acknowledgements

This study was supported by: the European Union's Seventh Framework Program under grant agreement n° 266486 ('NU-AGE: New dietary strategies addressing the specific needs of the elderly population for healthy aging in Europe'); the European Union's Seventh Framework Program under grant agreement n° 259679 ('IDEAL: Integrated research on DEvelopmental determinants of Aging and Longevity'); MiUR Prin 2009 to CF; Roberto and Cornelia Pallotti Legacy for cancer research to SS and MC.

References

Agrawal, A., Tay, J., Yang, G.-E., Agrawal, S., Gupta, S., 2010. Age-associated epigenetic modifications in human DNA increase its immunogenicity. Aging (Albany, NY) 2, 93–100.

Ahuja, N., Li, Q., Mohan, A.L., Baylin, S.B., Issa, J.P., 1998. Aging and DNA methylation in colorectal mucosa and cancer. Cancer Res. 58, 5489–5494.

Alexeeff, S.E., Baccarelli, A.A., Halonen, J., Coull, B.A., Wright, R.O., Tarantini, L., Bollati, V., Sparrow, D., Vokonas, P., Schwartz, J., 2013. Association between blood pressure and DNA methylation of retrotransposons and pro-inflammatory genes. Int. J. Epidemiol. 42, 270–280.

Angel-Morales, G., Noratto, G., Mertens-Talcott, S., 2012. Red wine polyphenolics reduce the expression of inflammation markers in human colon-derived CCD-18Co myofibroblast cells: potential role of microRNA-126. Food Funct. 3, 745–752.

Axume, J., Smith, S.S., Pogribny, I.P., Moriarty, D.J., Caudill, M.A., 2007. The MTHFR 677TT genotype and folate intake interact to lower global leukocyte DNA methylation in young Mexican American women. Nutr. Res. (New York, NY) 27, 1317–1365.

Baccarelli, A., Rienstra, M., Benjamin, E.J., 2010a. Cardiovascular epigenetics: basic concepts and results from animal and human studies. Circ. Cardiovasc. Genet. 3, 567–573.

Baccarelli, A., Tarantini, L., Wright, R.O., Bollati, V., Litonjua, A.A., Zanobetti, A., Sparrow, D., Vokonas, P., Schwartz, J., 2010b. Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study. Epigenetics Off. J. DNA Methylation Soc. 5 .

Baccarelli, A., Wright, R., Bollati, V., Litonjua, A., Zanobetti, A., Tarantini, L., Sparrow, D., Vokonas, P., Schwartz, J., 2010c. Ischemic heart disease and stroke in relation to blood DNA methylation. Epidemiology (Cambridge, Mass.) 21, 819–828.

Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttil, R.A., Dollé, M.E.T., Calder, R.B., Chisholm, G.B., Pollock, B.H., Klein, C.A., Vijg, J., 2006. Increased cell-to-cell variation in gene expression in ageing mouse heart. Nature 441, 1011–1014.

Barbot, W., Dupressoir, A., Lazar, V., Heidmann, T., 2002. Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. Nucleic Acids Res. 30, 2365–2373.

Barger, J.L., Kayo, T., Pugh, T.D., Prolla, T.A., Weindruch, R., 2008. Short-term consumption of a resveratrol-containing nutraceutical mixture mimics gene expression of long-term caloric restriction in mouse heart. Exp. Gerontol. 43, 859–866.

Barnes, S.K., Ozanne, S.E., 2011. Pathways linking the early environment to long-term health and lifespan. Prog. Biophys. Mol. Biol. 106, 323–336.

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233.

Bauersachs, J., Thum, T., 2011. Biogenesis and regulation of cardiovascular microRNAs. Circ. Res. 109, 334–347.

Baur, J.A., Chen, D., Chini, E.N., Chua, K., Cohen, H.Y., de Cabo, R., Deng, C., Dimmeler, S., Gius, D., Guarente, L.P., Helfand, S.L., Imai, S.-I., Itoh, H., Kadowaki, T., Koya, D., Leeuwenburgh, C., McBurney, M., Nabeshima, Y.-I., Neri, C., Oberdoerffer, P., Pestell, R.G., Rogina, B., Sadoshima, J., Sartorelli, V., Serrano, M., Sinclair, D.A., Steegborn, C., Tatar, M., Tissenbaum, H.A., Tong, Q., Tsubota, K., Vaquero, A., Verdin, E., 2010. Dietary restriction: standing up for sirtuins. Science 329, 1012–1013 (author reply 1013–1014).

Bellizzi, D., D'Aquila, P., Montesanto, A., Corsonello, A., Mari, V., Mazzei, B., Lattanzio, F., Passarino, G., 2012. Global DNA methylation in old subjects is correlated with frailty. Age 34, 169–179.

Bennett-Baker, P.E., Wilkowski, J., Burke, D.T., 2003. Age-associated activation of epigenetically repressed genes in the mouse. Genetics 165, 2055–2062.

Bjornsson, H.T., Sigurdsson, M.I., Fallin, M.D., Irizarry, R.A., Aspelund, T., Cui, H., Yu, W., Rongione, M.A., Ekström, T.J., Harris, T.B., Launer, L.J., Eiriksdottir, G., Leppert, M.F., Sapienza, C., Gudnason, V., Feinberg, A.P., 2008. Intra-individual change over time in DNA methylation with familial clustering. JAMA 299, 2877–2883.

Blount, B.C., Mack, M.M., Wehr, C.M., MacGregor, J.T., Hiatt, R.A., Wang, G., Wickramasinghe, S.N., Everson, R.B., Ames, B.N., 1997. Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. Proc. Natl. Acad. Sci. U. S. A. 94, 3290–3295.

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, e182–e189.

Bocklandt, S., Lin, W., Sehl, M.E., Sánchez, F.J., Sinsheimer, J.S., Horvath, S., Vilain, E., 2011. Epigenetic predictor of age. PLoS One 6, e14821.

Boesch-Saadatmandi, C., Wagner, A.E., Wolffram, S., Rimbach, G., 2012. Effect of quercetin on inflammatory gene expression in mice liver in vivo – role of redox factor 1, miR-122 and miR-125b. Pharmacol. Res. Off. J. Ital. Pharmacol. Soc. 65, 523–530.

Boldin, M.P., Baltimore, D., 2012. MicroRNAs, new effectors and regulators of NF-κB. Immunol. Rev. 246, 205–220.

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.

Bouchard, L., Rabasa-Lhoret, R., Faraj, M., Lavoie, M.-E., Mill, J., Pérusse, L., Vohl, M.-C., 2010. Differential epigenomic and transcriptomic responses in subcutaneous adipose tissue between low and high responders to caloric restriction. Am. J. Clin. Nutr. 91, 309–320.

Breving, K., Esquela-Kerscher, A., 2010. The complexities of microRNA regulation: mirandering around the rules. Int. J. Biochem. Cell Biol. 42, 1316–1329.

Brzek, P., Ksiazek, A., Dobrzyn, A., Konarzewski, M., 2012. Effect of dietary restriction on metabolic, anatomic and molecular traits in mice depends on the initial level of basal metabolic rate. J. Exp. Biol. 215, 3191–3199.

Cao, L., Kong, L.-P., Yu, Z.-B., Han, S.-P., Bai, Y.-F., Zhu, J., Hu, X., Zhu, C., Zhu, S., Guo, X.-R., 2012. microRNA expression profiling of the developing mouse heart. Int. J. Mol. Med. 30, 1095–1104.

Casillas Jr., M.A., Lopatina, N., Andrews, L.G., Tollefsbol, T.O., 2003. Transcriptional control of the DNA methyltransferases is altered in aging and neoplastically-transformed human fibroblasts. Mol. Cell. Biochem. 252, 33–43.

Castro, R., Rivera, I., Struys, E.A., Jansen, E.E.W., Ravasco, P., Camilo, M.E., Blom, H.J., Jakobs, C., Tavares de Almeida, I., 2003. Increased homocysteine and S-adenosylhomocysteine concentrations and DNA hypomethylation in vascular disease. Clin. Chem. 49, 1292–1296.

Cevenini, E., Invidia, L., Lescai, F., Salvioli, S., Tieri, P., Castellani, G., Franceschi, C., 2008. Human models of aging and longevity. Expert Opin. Biol. Ther. 8, 1393–1405.

Cevenini, E., Monti, D., Franceschi, C., 2013. Inflamm-ageing. Curr. Opin. Clin. Nutr. Metab. Care 16, 14–20.

Choi, E.K., Uyeno, S., Nishida, N., Okumoto, T., Fujimura, S., Aoki, Y., Nata, M., Sagisaka, K., Fukuda, Y., Nakao, K., Yoshimoto, T., Kim, Y.S., Ono, T., 1996. Alterations of c-fos gene methylation in the processes of aging and tumorigenesis in human liver. Mutat. Res. 354, 123–128.

Choi, S.-W., Friso, S., 2010. Epigenetics: a new bridge between nutrition and health. Adv. Nutr. 1, 8–16.

Chouliaras, L., van den Hove, D.L.A., Kenis, G., Dela Cruz, J., Lemmens, M.A.M., van Os, J., Steinbusch, H.W.M., Schmitz, C., Rutten, B.P.F., 2011. Caloric restriction attenuates age-related changes of DNA methyltransferase 3a in mouse hippocampus. Brain Behav. Immun. 25, 616–623.

Chouliaras, L., van den Hove, D.L.A., Kenis, G., Keitel, S., Hof, P.R., van Os, J., Steinbusch, H.W.M., Schmitz, C., Rutten, B.P.F., 2012. Prevention of age-related changes in hippocampal levels of 5-methylcytidine by caloric restriction. Neurobiol. Aging 33, 1672–1681.

Christensen, B.C., Houseman, E.A., Marsit, C.J., Zheng, S., Wrensch, M.R., Wiemels, J.L., Nelson, H.H., Karagas, M.R., Padbury, J.F., Bueno, R., Sugarbaker, D.J., Yeh, R.-F., Wiencke, J.K., Kelsey, K.T., 2009. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet. 5, e1000602.

Ciappio, E.D., Mason, J.B., Crott, J.W., 2011. Maternal one-carbon nutrient intake and cancer risk in offspring. Nutr. Rev. 69, 561–571.

Ciccarone, F., Klinger, F.G., Catizone, A., Calabrese, R., Zampieri, M., Bacalini, M.G., De Felici, M., Caiafa, P., 2012. Poly(ADP-ribosyl)ation acts in the DNA demethylation of mouse primordial germ cells also with DNA damage-independent roles. PLoS One 7, e46927.

Cimini, A., Gentile, R., D'Angelo, B., Benedetti, E., Cristiano, L., Avantaggiati, M., Giordano, A., Ferri, C., Desideri, G., 2013. Cocoa powder triggers neuroprotective and preventive effects in a human Alzheimer's Disease model by modulating BDNF signaling pathway. J. Cell. Biochem..

Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., Barbacioru, C., Steptoe, A.L., Martin, H.C., Nourbakhsh, E., Krishnan, K., Gardiner, B., Wang, X., Nones, K., Steen, J.A., Matigian, N.A., Wood, D.L., Kassahn, K.S., Waddell, N., Shepherd, J., Lee, C., Ichikawa, J., McKernan, K., Bramlett, K., Kuersten, S., Grimmond, S.M., 2011. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol. 12, R126.

Cohen, N.M., Kenigsberg, E., Tanay, A., 2011. Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. Cell 145, 773–786.

Collino, F., Deregibus, M.C., Bruno, S., Sterpone, L., Aghemo, G., Viltono, L., Tetta, C., Camussi, G., 2010. Microvesicles derived from adult human bone marrow and tissue specific mesenchymal stem cells shuttle selected pattern of miRs. PLoS One 5, e11803.

Colman, R.J., Anderson, R.M., Johnson, S.C., Kastman, E.K., Kosmatka, K.J., Beasley, T.M., Allison, D.B., Cruzen, C., Simmons, H.A., Kemnitz, J.W., Weindruch, R., 2009. Caloric restriction delays disease onset and mortality in rhesus monkeys. Science 325, 201–204.

Cosentino, C., Mostoslavsky, R., 2013. Metabolism, longevity and epigenetics. Cell. Mol. Life Sci. 70, 1525–1541.

Cravo, M.L., Albuquerque, C.M., Salazar de Sousa, L., Glória, L.M., Chaves, P., Dias Pereira, A., Nobre Leitão, C., Quina, M.G., Costa Mira, F., 1998a. Microsatellite instability in non-neoplastic mucosa of patients with ulcerative colitis: effect of folate supplementation. Am. J. Gastroenterol. 93, 2060–2064.

Cravo, M.L., Pinto, A.G., Chaves, P., Cruz, J.A., Lage, P., Nobre Leitão, C., Costa Mira, F., 1998b. Effect of folate supplementation on DNA methylation of rectal mucosa in patients with colonic adenomas: correlation with nutrient intake. Clin. Nutr. Edinb. Scotl. 17, 45–49.

Creemers, E.E., Tijsen, A.J., Pinto, Y.M., 2012. Circulating microRNAs: novel biomarkers and extracellular communicators in cardiovascular disease? Circ. Res. 110, 483–495.

Dallaire, A., Garand, C., Paquel, E.R., Mitchell, S.J., de Cabo, R., Simard, M.J., Lebel, M., 2012. Down regulation of miR-124 in both Werner syndrome DNA helicase mutant mice and mutant Caenorhabditis elegans wrn-1 reveals the importance of this microRNA in accelerated aging. Aging (Albany, NY) 4, 636–647.

Darvesh, A.S., Bishayee, A., 2013. Chemopreventive and therapeutic potential of tea polyphenols in hepatocellular cancer. Nutr. Cancer 65, 329–344.

De Bock, M., Derraik, J.G.B., Brennan, C.M., Biggs, J.B., Morgan, P.E., Hodgkinson, S.C., Hofman, P.L., Cutfield, W.S., 2013. Olive (*Olea europaea* L.) leaf polyphenols improve insulin sensitivity in middle-aged overweight men: a randomized, placebo-controlled, crossover trial. PLoS One 8, e57622.

De Cecco, M., Criscione, S.W., Peckham, E.J., Hillenmeyer, S., Hamm, E.A., Manivannan, J., Peterson, A.L., Kreiling, J.A., Neretti, N., Sedivy, J.M., 2013. Genomes of

replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. Aging Cell 12, 247–256.
De Magalhães, J.P., 2012. Programmatic features of aging originating in development: aging mechanisms beyond molecular damage? FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 26, 4821–4826.
Del Follo-Martinez, A., Banerjee, N., Li, X., Safe, S., Mertens-Talcott, S., 2013. Resveratrol and quercetin in combination have anticancer activity in colon cancer cells and repress oncogenic microRNA-27a. Nutr. Cancer 65, 494–504.
Dhahbi, J.M., Spindler, S.R., Atamna, H., Yamakawa, A., Guerrero, N., Boffelli, D., Mote, P., Martin, D.I.K., 2013. Deep sequencing identifies circulating mouse miRs that are functionally implicated in manifestations of aging and responsive to calorie restriction. Aging (Albany, NY) 5, 130–141.
Drinkwater, R.D., Blake, T.J., Morley, A.A., Turner, D.R., 1989. Human lymphocytes aged in vivo have reduced levels of methylation in transcriptionally active and inactive DNA. Mutat. Res. 219, 29–37.
Ehrlich, M., 2006. Cancer-linked DNA hypomethylation and its relationship to hypermethylation. Curr. Top. Microbiol. Immunol. 310, 251–274.
ElSharawy, A., Keller, A., Flachsbart, F., Wendschlag, A., Jacobs, G., Kefer, N., Brefort, T., Leidinger, P., Backes, C., Meese, E., Schreiber, S., Rosenstiel, P., Franke, A., Nebel, A., 2012. Genome-wide miR signatures of human longevity. Aging Cell 11, 607–616.
Fackler, M.J., Umbricht, C.B., Williams, D., Argani, P., Cruz, L.-A., Merino, V.F., Teo, W.W., Zhang, Z., Huang, P., Visvananthan, K., Marks, J., Ethier, S., Gray, J.W., Wolff, A.C., Cope, L.M., Sukumar, S., 2011. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. Cancer Res. 71, 6195–6207.
Fang, M., Chen, D., Yang, C.S., 2007. Dietary polyphenols may affect DNA methylation. J. Nutr. 137, 223S–228S.
FichtIscherer, S., De Rosa, S., Fox, H., Schwietz, T., Fischer, A., Liebetrau, C., Weber, M., Hamm, C.W., Röxe, T., Müller-Ardogan, M., Bonauer, A., Zeiher, A.M., Dimmeler, S., 2010. Circulating microRNAs in patients with coronary artery disease. Circ. Res. 107, 677–684.
Florath, I., Butterbach, K., Müller, H., Bewerunge-Hudler, M., Brenner, H., 2013. 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..
Fontana, L., Klein, S., Holloszy, J.O., 2010a. Effects of long-term calorie restriction and endurance exercise on glucose tolerance, insulin action, and adipokine production. Age Dordr. Neth. 32, 97–108.
Fontana, L., Partridge, L., Longo, V.D., 2010b. Extending healthy life span—from yeast to humans. Science 328, 321–326.
Ford, D., Ions, L.J., Alatawi, F., Wakeling, L.A., 2011. The potential role of epigenetic responses to diet in ageing. Proc. Nutr. Soc. 70, 374–384.
Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T.D., 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.
Franceschi, C., 2007. Inflammaging as a major characteristic of old people: can it be prevented or cured? Nutr. Rev. 65, S173–S176.
Franceschi, C., Bonafè, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E., De Benedictis, G., 2000. Inflamm-aging. An evolutionary perspective on immunosenescence. Ann. N. Y. Acad. Sci. 908, 244–254.
Friso, S., Choi, S.-W., Girelli, D., Mason, J.B., Dolnikowski, G.G., Bagley, P.J., Olivieri, O., Jacques, P.F., Rosenberg, I.H., Corrocher, R., Selhub, J., 2002. A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. Proc. Natl. Acad. Sci. U. S. A. 99, 5606–5611.
Friso, S., Lotto, V., Choi, S.-W., Girelli, D., Pinotti, M., Guarini, P., Udali, S., Pattini, P., Pizzolo, F., Martinelli, N., Corrocher, R., Bernardi, F., Olivieri, O., 2012. Promoter methylation in coagulation F7 gene influences plasma FVII concentrations and relates to coronary artery disease. J. Med. Genet. 49, 192–199.
Friso, S., Pizzolo, F., Choi, S.-W., Guarini, P., Castagna, A., Ravagnani, V., Carletto, A., Pattini, P., Corrocher, R., Olivieri, O., 2008. Epigenetic control of 11 beta-hydroxysteroid dehydrogenase 2 gene promoter is related to human hypertension. Atherosclerosis 199, 323–327.
Friso, S., Udali, S., Guarini, P., Pellegrini, C., Pattini, P., Moruzzi, S., Girelli, D., Pizzolo, F., Martinelli, N., Corrocher, R., Olivieri, O., Choi, S.-W., 2013. Global DNA hypomethylation in peripheral blood mononuclear cells as a biomarker of cancer risk. Cancer Epidemiol. Biomarkers Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol. 22, 348–355.
Fu, S., Kurzrock, R., 2010. Development of curcumin as an epigenetic agent. Cancer 116, 4670–4676.
Fu, V.X., Dobosy, J.R., Desotelle, J.A., Almassi, N., Ewald, J.A., Srinivasan, R., Berres, M., Svaren, J., Weindruch, R., Jarrard, D.F., 2008. Aging and cancer-related loss of insulin-like growth factor 2 imprinting in the mouse and human prostate. Cancer Res. 68, 6797–6802.
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. Ann. Hum. Genet. 68, 196–204.
Gabory, A., Attig, L., Junien, C., 2011. Developmental programming and epigenetics. Am. J. Clin. Nutr. 94, 1943S–1952S.
Garagnani, P., Bacalini, M.G., Pirazzini, C., Gori, D., Giuliani, C., Mari, D., Di Blasio, A.M., Gentilini, D., Vitale, G., Collino, S., Rezzi, S., Castellani, G., Capri, M., Salvioli, S., Franceschi, C., 2012. Methylation of ELOVL2 gene as a new epigenetic marker of age. Aging Cell 11, 1132–1134.
Gentilini, D., Mari, D., Castaldi, D., Remondini, D., Ogliari, G., Ostan, R., Bucci, L., Sirchia, S.M., Tabano, S., Cavagnini, F., Monti, D., Franceschi, C., Di Blasio, A.M., Vitale, G., 2012. Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring. Age Dordr. Neth..
Giahi, L., Aumueller, E., Elmadfa, I., Haslberger, A.G., 2012. Regulation of TLR4, p38 MAPkinase, IκB and miRs by inactivated strains of lactobacilli in human dendritic cells. Benef. Microbes 3, 91–98.
Gilbert, S.F., 2009. Ageing and cancer as diseases of epigenesis. J. Biosci. 34, 601–604.
Girelli, D., Friso, S., Trabetti, E., Olivieri, O., Russo, C., Pessotto, R., Faccini, G., Pignatti, P.F., Mazzucco, A., Corrocher, R., 1998. Methylenetetrahydrofolate reductase C677T mutation, plasma homocysteine, and folate in subjects from northern Italy with or without angiographically documented severe coronary atherosclerotic disease: evidence for an important genetic-environmental interaction. Blood 91, 4158–4163.
Gokbulut, A.A., Apohan, E., Baran, Y., 2013. Resveratrol and quercetin-induced apoptosis of human 232B4 chronic lymphocytic leukemia cells by activation of caspase-3 and cell cycle arrest. Hematol. Amst. Neth..
Golbus, J., Palella, T.D., Richardson, B.C., 1990. Quantitative changes in T cell DNA methylation occur during differentiation and ageing. Eur. J. Immunol. 20, 1869–1872.
Gollucke, A.P.B., Aguiar Jr., O., Barbisan, L.F., Ribeiro, D.A., 2013. Use of grape polyphenols against carcinogenesis: putative molecular mechanisms of action using in vitro and in vivo test systems. J. Med. Food 16, 199–205.
Gombar, S., Jung, H.J., Dong, F., Calder, B., Atzmon, G., Barzilai, N., Tian, X.-L., Pothof, J., Hoeijmakers, J.H.J., Campisi, J., Vijg, J., Suh, Y., 2012. Comprehensive microRNA profiling in B-cells of human centenarians by massively parallel sequencing. BMC Genomics 13, 353.
Gomes, M.V.M., Toffoli, L.V., Arruda, D.W., Soldera, L.M., Pelosi, G.G., Neves-Souza, R.D., Freitas, E.R., Castro, D.T., Marquez, A.S., 2012. Age-related changes in the global DNA methylation profile of leukocytes are linked to nutrition but are not associated with the MTHFR C677T genotype or to functional capacities. PLoS One 7, e52570.
Gravina, S., Vijg, J., 2010. Epigenetic factors in aging and longevity. Pflügers Arch. Eur. J. Physiol. 459, 247–258.
Grillari, J., Grillari-Voglauer, R., 2010. Novel modulators of senescence, aging, and longevity: small non-coding RNAs enter the stage. Exp. Gerontol. 45, 302–311.
Gronniger, E., Weber, B., Heil, O., Peters, N., Stab, F., Wenck, H., Korn, B., Winnefeld, M., Lyko, F., 2010. Aging and chronic sun exposure cause distinct epigenetic changes in human skin. PLoS Genet. 6 .
Grummt, I., Voit, R., 2010. Linking rDNA transcription to the cellular energy supply. Cell Cycle Georget. Tex 9, 225–226.
Guastafierro, T., Cecchinelli, B., Zampieri, M., Reale, A., Riggio, G., Sthandier, O., Zupi, G., Calabrese, L., Caiafa, P., 2008. CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. J. Biol. Chem. 283, 21873–21880.
Handy, D.E., Castro, R., Loscalzo, J., 2011. Epigenetic modifications: basic mechanisms and role in cardiovascular disease. Circulation 123, 2145–2156.
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.
Hass, B.S., Hart, R.W., Lu, M.H., Lyn-Cook, B.D., 1993. Effects of caloric restriction in animals on cellular function, oncogene expression, and DNA methylation in vitro. Mutat. Res. 295, 281–289.
Heijmans, B.T., Kremer, D., Tobi, E.W., Boomsma, D.I., Slagboom, P.E., 2007. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. Hum. Mol. Genet. 16, 547–554.
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., Valencia, A., Gut, I.G., Wang, J., Esteller, M., 2012. Distinct DNA methylomes of newborns and centenarians. Proc. Natl. Acad. Sci. U. S. A. 109, 10522–10527.
Hiltunen, M.O., Tuomisto, T.T., Niemi, M., Bräsen, J.H., Rissanen, T.T., Törönen, P., Vajanto, I., Ylä-Herttuala, S., 2002a. Changes in gene expression in atherosclerotic plaques analyzed using DNA array. Atherosclerosis 165, 23–32.
Hiltunen, M.O., Turunen, M.P., Häkkinen, T.P., Rutanen, J., Hedman, M., Mäkinen, K., Turunen, A.-M., Aalto-Setälä, K., Ylä-Herttuala, S., 2002b. DNA hypomethylation and methyltransferase expression in atherosclerotic lesions. Vasc. Med. Lond. Engl. 7, 5–11.
Holliday, R., 1987. The inheritance of epigenetic defects. Science 238, 163–170.
Horvath, S., 2013. DNA methylation age of human tissues and cell types. Genome Biol. 14, R115.
Hosoda, T., Zheng, H., Cabral-da-Silva, M., Sanada, F., Ide-Iwata, N., Ogórek, B., Ferreira-Martins, J., Arranto, C., D'Amario, D., del Monte, F., Urbanek, K., D'Alessandro, D.A., Michler, R.E., Anversa, P., Rota, M., Kajstura, J., Leri, A., 2011. Human cardiac stem cell differentiation is regulated by a mircrine mechanism. Circulation 123, 1287–1296.
Hübner, U., Geisel, J., Kirsch, S.H., Kruse, V., Bodis, M., Klein, C., Herrmann, W., Obeid, R., 2013. Effect of 1 year B and D vitamin supplementation on LINE-1 repetitive element methylation in older subjects. Clin. Chem. Lab. Med. CCLM FESCC 51, 649–655.
Huidobro, C., Fernandez, A.F., Fraga, M.F., 2013. Aging epigenetics: causes and consequences. Mol. Aspects Med. 34, 765–781.

Ibáñez-Ventoso, C., Driscoll, M., 2009. MicroRNAs in *C. elegans* aging: molecular insurance for robustness? Curr. Genomics 10, 144–153.
Inukai, S., de Lencastre, A., Turner, M., Slack, F., 2012. Novel microRNAs differentially expressed during aging in the mouse brain. PLoS One 7, e40028.
Inukai, S., Slack, F., 2013. MicroRNAs and the genetic network in aging. J. Mol. Biol..
Ions, L.J., Wakeling, L.A., Bosomworth, H.J., Hardyman, J.E., Escolme, S.M., Swan, D.C., Valentine, R.A., Mathers, J.C., Ford, D., 2012. Effects of Sirt1 on DNA methylation and expression of genes affected by dietary restriction. Age Dordr. Neth..
Issa, J.P., Ottaviano, Y.L., Celano, P., Hamilton, S.R., Davidson, N.E., Baylin, S.B., 1994. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. Nat. Genet. 7, 536–540.
Issa, J.P., Vertino, P.M., Boehm, C.D., Newsham, I.F., Baylin, S.B., 1996. Switch from monoallelic to biallelic human IGF2 promoter methylation during aging and carcinogenesis. Proc. Natl. Acad. Sci. U. S. A. 93, 11757–11762.
Jacques, P.F., Bostom, A.G., Williams, R.R., Ellison, R.C., Eckfeldt, J.H., Rosenberg, I.H., Selhub, J., Rozen, R., 1996. Relation between folate status, a common mutation in methylenetetrahydrofolate reductase, and plasma homocysteine concentrations. Circulation 93, 7–9.
Jin, B., Robertson, K.D., 2013. DNA methyltransferases, DNA damage repair, and cancer. Adv. Exp. Med. Biol. 754, 3–29.
Jintaridth, P., Mutirangura, A., 2010. Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. Physiol. Genomics 41, 194–200.
Johansson, A., Enroth, S., Gyllensten, U., 2013. Continuous aging of the human DNA methylome throughout the human lifespan. PLoS One 8, e67378.
Johnson, A.A., Akman, K., Calimport, S.R.G., Wuttke, D., Stolzing, A., de Magalhães, J.P., 2012. The role of DNA methylation in aging, rejuvenation, and age-related disease. Rejuvenation Res. 15, 483–494.
Johnson, K.C., Koestler, D.C., Cheng, C., Christensen, B.C., 2013. Age-related DNA methylation in normal breast tissue and its relationship with invasive breast tumor methylation. Epigenetics Off. J. DNA Methylation Soc. 9 .
Johnson, T.E., 2006. Recent results: biomarkers of aging. Exp. Gerontol. 41, 1243–1246.
Keyes, M.K., Jang, H., Mason, J.B., Liu, Z., Crott, J.W., Smith, D.E., Friso, S., Choi, S.-W., 2007. Older age and dietary folate are determinants of genomic and p16-specific DNA methylation in mouse colon. J. Nutr. 137, 1713–1717.
Khurana, S., Piche, M., Hollingsworth, A., Venkataraman, K., Tai, T.C., 2013. Oxidative stress and cardiovascular health: therapeutic potential of polyphenols. Can. J. Physiol. Pharmacol. 91, 198–212.
Kim, J., Kim, J.Y., Song, K.S., Lee, Y.H., Seo, J.S., Jelinek, J., Goldschmidt-Clermont, P.J., Issa, J.-P.J., 2007. Epigenetic changes in estrogen receptor beta gene in atherosclerotic cardiovascular tissues and in-vitro vascular senescence. Biochim. Biophys. Acta 1772, 72–80.
Kim, K., Friso, S., Choi, S.-W., 2009. DNA methylation, an epigenetic mechanism connecting folate to healthy embryonic development and aging. J. Nutr. Biochem. 20, 917–926.
Kim, M., Long, T.I., Arakawa, K., Wang, R., Yu, M.C., Laird, P.W., 2010. DNA methylation as a biomarker for cardiovascular disease risk. PLoS One 5, e9692.
King-Batoon, A., Leszczynska, J.M., Klein, C.B., 2008. Modulation of gene methylation by genistein or lycopene in breast cancer cells. Environ. Mol. Mutagen. 49, 36–45.
Koch, C.M., Wagner, W., 2011. Epigenetic-aging-signature to determine age in different tissues. Aging (Albany, NY) 3, 1018–1027.
Krishna, S.M., Dear, A., Craig, J.M., Norman, P.E., Golledge, J., 2013. The potential role of homocysteine mediated DNA methylation and associated epigenetic changes in abdominal aortic aneurysm formation. Atherosclerosis.
Kucharski, R., Maleszka, J., Foret, S., Maleszka, R., 2008. Nutritional control of reproductive status in honeybees via DNA methylation. Science 319, 1827–1830.
Kumarswamy, R., Volkmann, I., Thum, T., 2011. Regulation and function of miR-21 in health and disease. RNA Biol. 8, 706–713.
Larson, K., Yan, S.-J., Tsurumi, A., Liu, J., Zhou, J., Gaur, K., Guo, D., Eickbush, T.H., Li, W.X., 2012. Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis. PLoS Genet. 8, e1002473.
Laukkanen, M.O., Mannermaa, S., Hiltunen, M.O., Aittomäki, S., Airenne, K., Jänne, J., Ylä-Herttuala, S., 1999. Local hypomethylation in atherosclerosis found in rabbit ec-sod gene. Arterioscler. Thromb. Vasc. Biol. 19, 2171–2178.
Lee, W.J., Shim, J.-Y., Zhu, B.T., 2005. Mechanisms for the inhibition of DNA methyltransferases by tea catechins and bioflavonoids. Mol. Pharmacol. 68, 1018–1030.
Li, H., Förstermann, U., 2012. Red wine and cardiovascular health. Circ. Res. 111, 959–961.
Li, X., Khanna, A., Li, N., Wang, E., 2011a. Circulatory miR34a as an RNAbased, noninvasive biomarker for brain aging. Aging (Albany, NY) 3, 985–1002.
Li, Y., Daniel, M., Tollefsbol, T.O., 2011b. Epigenetic regulation of caloric restriction in aging. BMC Med. 9, 98.
Li, Y., Liu, L., Tollefsbol, T.O., 2010a. Glucose restriction can extend normal cell lifespan and impair precancerous cell growth through epigenetic control of hTERT and p16 expression. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 24, 1442–1453.
Li, Y., Liu, Y., Strickland, F.M., Richardson, B., 2010b. Age-dependent decreases in DNA methyltransferase levels and low transmethylation micronutrient levels synergize to promote overexpression of genes implicated in autoimmunity and acute coronary syndromes. Exp. Gerontol. 45, 312–322.
Li, Y., Tollefsbol, T.O., 2010. Impact on DNA methylation in cancer prevention and therapy by bioactive dietary components. Curr. Med. Chem. 17, 2141–2151.
Liao, C.-Y., Rikke, B.A., Johnson, T.E., Diaz, V., Nelson, J.F., 2010. Genetic variation in the murine lifespan response to dietary restriction: from life extension to life shortening. Aging Cell 9, 92–95.
Lillycrop, K.A., Burdge, G.C., 2012. Epigenetic mechanisms linking early nutrition to long term health. Best Pract. Res. Clin. Endocrinol. Metab. 26, 667–676.
Lillycrop, K.A., Phillips, E.S., Jackson, A.A., Hanson, M.A., Burdge, G.C., 2005. Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. J. Nutr. 135, 1382–1386.
Lillycrop, K.A., Phillips, E.S., Torrens, C., Hanson, M.A., Jackson, A.A., Burdge, G.C., 2008. Feeding pregnant rats a protein-restricted diet persistently alters the methylation of specific cytosines in the hepatic PPAR alpha promoter of the offspring. Br. J. Nutr. 100, 278–282.
Lindahl, T., Wood, R.D., 1999. Quality control by DNA repair. Science 286, 1897–1905.
Liu, L., van Groen, T., Kadish, I., Li, Y., Wang, D., James, S.R., Karpf, A.R., Tollefsbol, T.O., 2011. Insufficient DNA methylation affects healthy aging and promotes age-related health problems. Clin. Epigenetics 2, 349–360.
Liu, N., Landreh, M., Cao, K., Abe, M., Hendriks, G.-J., Kennerdell, J.R., Zhu, Y., Wang, L.-S., Bonini, N.M., 2012. The microRNA miR-34 modulates ageing and neurodegeneration in Drosophila. Nature 482, 519–523.
Liu, Y., Chen, Y., Richardson, B., 2009. Decreased DNA methyltransferase levels contribute to abnormal gene expression in "senescent" CD4(+)CD28(−) T cells. Clin. Immunol. (Orlando, Fla) 132, 257–265.
Lopatina, N., Haskell, J.F., Andrews, L.G., Poole, J.C., Saldanha, S., Tollefsbol, T., 2002. Differential maintenance and de novo methylating activity by three DNA methyltransferases in aging and immortalized fibroblasts. J. Cell. Biochem. 84, 324–334.
Lund, G., Andersson, L., Lauria, M., Lindholm, M., Fraga, M.F., Villar-Garea, A., Ballestar, E., Esteller, M., Zaina, S., 2004. DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E. J. Biol. Chem. 279, 29147–29154.
Madrigano, J., Baccarelli, A., Mittleman, M.A., Sparrow, D., Vokonas, P.S., Tarantini, L., Schwartz, J., 2012. Aging and epigenetics: longitudinal changes in gene-specific DNA methylation. Epigenetics Off. J. DNA Methylation Soc. 7, 63–70.
Martin, S.L., Hardy, T.M., Tollefsbol, T.O., 2013. Medicinal chemistry of the epigenetic diet and caloric restriction. Curr. Med. Chem. 20, 4050–4059.
Mattison, J.A., Roth, G.S., Beasley, T.M., Tilmont, E.M., Handy, A.M., Herbert, R.L., Longo, D.L., Allison, D.B., Young, J.E., Bryant, M., Barnard, D., Ward, W.F., Qi, W., Ingram, D.K., de Cabo, R., 2012. Impact of caloric restriction on health and survival in rhesus monkeys from the NIA study. Nature 489, 318–321.
McCay, C.M., Crowell, M.F., Maynard, L.A., 1989. The effect of retarded growth upon the length of life span and upon the ultimate body size. 1935. Nutrition (Burbank, Los Angeles Cty., Calif.) 5, 155–171 (discussion 172).
Meeran, S.M., Patel, S.N., Tollefsbol, T.O., 2010. Sulforaphane causes epigenetic repression of hTERT expression in human breast cancer cell lines. PLoS One 5, e11457.
Mendelsohn, M.E., Karas, R.H., 1999. The protective effects of estrogen on the cardiovascular system. N. Engl. J. Med. 340, 1801–1811.
Mercken, E.M., Crosby, S.D., Lamming, D.W., Jebailey, L., Krzysik-Walker, S., Villareal, D., Capri, M., Franceschi, C., Zhang, Y., Becker, K., Sabatini, D.M., de Cabo, R., Fontana, L., 2013. Calorie restriction in humans inhibits the PI3K/AKT pathway and induces a younger transcription profile. Aging Cell..
Milagro, F.I., Campión, J., Cordero, P., Goyenechea, E., Gómez-Uriz, A.M., Abete, I., Zulet, M.A., Martínez, J.A., 2011. A dual epigenomic approach for the search of obesity biomarkers: DNA methylation in relation to diet-induced weight loss. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 25, 1378–1389.
Miller, R.A., Buehner, G., Chang, Y., Harper, J.M., Sigler, R., Smith-Wheelock, M., 2005. Methionine-deficient diet extends mouse lifespan, slows immune and lens aging, alters glucose, T4, IGF-I and insulin levels, and increases hepatocyte MIF levels and stress resistance. Aging Cell 4, 119–125.
Miyamura, Y., Tawa, R., Koizumi, A., Uehara, Y., Kurishita, A., Sakurai, H., Kamiyama, S., Ono, T., 1993. Effects of energy restriction on age-associated changes of DNA methylation in mouse liver. Mutat. Res. 295, 63–69.
Murayama, A., Ohmori, K., Fujimura, A., Minami, H., Yasuzawa-Tanaka, K., Kuroda, T., Oie, S., Daitoku, H., Okuwaki, M., Nagata, K., Fukamizu, A., Kimura, K., Shimizu, T., Yanagisawa, J., 2008. Epigenetic control of rDNA loci in response to intracellular energy status. Cell 133, 627–639.
Nandakumar, V., Vaid, M., Katiyar, S.K., 2011. (−)-Epigallocatechin-3-gallate reactivates silenced tumor suppressor genes, Cip1/p21 and p16INK4a, by reducing DNA methylation and increasing histones acetylation in human skin cancer cells. Carcinogenesis 32, 537–544.
Ndlovu, M.N., Denis, H., Fuks, F., 2011. Exposing the DNA methylome iceberg. Trends Biochem. Sci. 36, 381–387.
Noratto, G.D., Kim, Y., Talcott, S.T., Mertens-Talcott, S.U., 2011. Flavonol-rich fractions of yaupon holly leaves (Ilex vomitoria, Aquifoliaceae) induce microRNA-146a and have anti-inflammatory and chemopreventive effects in intestinal myofibroblast CCD-18Co cells. Fitoterapia 82, 557–569.
Numata, S., Ye, T., Hyde, T.M., Guitart-Navarro, X., Tao, R., Wininger, M., Colantuoni, C., Weinberger, D.R., Kleinman, J.E., Lipska, B.K., 2012. DNA methylation signatures in development and aging of the human prefrontal cortex. Am. J. Hum. Genet. 90, 260–272.
Oka, Y., Iwai, S., Amano, H., Irie, Y., Yatomi, K., Ryu, K., Yamada, S., Inagaki, K., Oguchi, K., 2012. Tea polyphenols inhibit rat osteoclast formation and differentiation. J. Pharmacol. Sci. 118, 55–64.

Oliveira, A.M.M., Hemstedt, T.J., Bading, H., 2012. Rescue of aging-associated decline in Dnmt3a2 expression restores cognitive abilities. Nat. Neurosci. 15, 1111–1113.

Olivieri, F., Lazzarini, R., Recchioni, R., Marcheselli, F., Rippo, M.R., Di Nuzzo, S., Albertini, M.C., Graciotti, L., Babini, L., Mariotti, S., Spada, G., Abbatecola, A.M., Antonicelli, R., Franceschi, C., Procopio, A.D., 2012a. MiR-146a as marker of senescence-associated pro-inflammatory status in cells involved in vascular remodelling. Age (Dordr. Neth.).

Olivieri, F., Rippo, M.R., Monsurrò, V., Salvioli, S., Capri, M., Procopio, A.D., Franceschi, C., 2013a. MicroRNAs linking inflamm-aging, cellular senescence and cancer. Ageing Res. Rev..

Olivieri, F., Rippo, M.R., Prattichizzo, F., Babini, L., Graciotti, L., Recchioni, R., Procopio, A.D., 2013b. Toll like receptor signaling in "inflammaging": microRNA as new players. Immun. Ageing 10, 11.

Olivieri, F., Spazzafumo, L., Santini, G., Lazzarini, R., Albertini, M.C., Rippo, M.R., Galeazzi, R., Abbatecola, A.M., Marcheselli, F., Monti, D., Ostan, R., Cevenini, E., Antonicelli, R., Franceschi, C., Procopio, A.D., 2012b. Age-related differences in the expression of circulating microRNAs: miR-21 as a new circulating marker of inflammaging. Mech. Ageing Dev. 133, 675–685.

Pandey, M., Shukla, S., Gupta, S., 2010. Promoter demethylation and chromatin remodeling by green tea polyphenols leads to re-expression of GSTP1 in human prostate cancer cells. Int. J. Cancer 126, 2520–2533.

Pantic, I., Pantic, S., Paunovic, J., 2012. Aging increases nuclear chromatin entropy of erythroid precursor cells in mice spleen hematopoietic tissue. Microsc. Microanal. Off. J. Microsc. Soc. Am. Microbeam Anal. Soc. Microsc. Soc. Can. 18, 1054–1059.

Park, K., Kim, K.-B., 2013. miRTar Hunter: a prediction system for identifying human microRNA target sites. Mol. Cells 35, 195–201.

Patel, D., Prasad, S., Kumar, R., Hemalatha, S., 2012. An overview on antidiabetic medicinal plants having insulin mimetic property. Asian Pac. J. Trop. Biomed. 2, 320–330.

Pirazzini, C., Giuliani, C., Bacalini, M.G., Boattini, A., Capri, M., Fontanesi, E., Marasco, E., Mantovani, V., Pierini, M., Pini, E., Luiselli, D., Franceschi, C., Garagnani, P., 2012. Space/population and time/age in DNA methylation variability in humans: a study on IGF2/H19 locus in different Italian populations and in mono- and di-zygotic twins of different age. Aging (Albany, NY) 4, 509–520.

Post, W.S., Goldschmidt-Clermont, P.J., Wilhide, C.C., Heldman, A.W., Sussman, M.S., Ouyang, P., Milliken, E.E., Issa, J.P., 1999. Methylation of the estrogen receptor gene is associated with aging and atherosclerosis in the cardiovascular system. Cardiovasc. Res. 43, 985–991.

Proietti, A.R., del Balzo, V., Dernini, S., Donini, L.M., Cannella, C., 2009. Mediterranean diet and prevention of non-communicable diseases: scientific evidences. Ann. Ig. Med. Prev. E Comunità 21, 197–210.

Pufulete, M., Al-Ghnaniem, R., Rennie, J.A., Appleby, P., Harris, N., Gout, S., Emery, P.W., Sanders, T.A., 2005. Influence of folate status on genomic DNA methylation in colonic mucosa of subjects without colorectal adenoma or cancer. Br. J. Cancer 92, 838–842.

Qin, W., Zhu, W., Shi, H., Hewett, J.E., Ruhlen, R.L., MacDonald, R.S., Rottinghaus, G.E., Chen, Y.-C., Sauter, E.R., 2009. Soy isoflavones have an antiestrogenic effect and alter mammary promoter hypermethylation in healthy premenopausal women. Nutr. Cancer 61, 238–244.

Quinn, S.R., O'Neill, L.A., 2011. A trio of microRNAs that control Toll-like receptor signalling. Int. Immunol. 23, 421–425.

Raddatz, G., Hagemann, S., Aran, D., Söhle, J., Kulkarni, P.P., Kaderali, L., Hellman, A., Winnefeld, M., Lyko, F., 2013. Aging is associated with highly defined epigenetic changes in the human epidermis. Epigenet. Chromatin 6, 36.

Raitoharju, E., Lyytikäinen, L.-P., Levula, M., Oksala, N., Mennander, A., Tarkka, M., Klopp, N., Illig, T., Kähönen, M., Karhunen, P.J., Laaksonen, R., Lehtimäki, T., 2011. miR-21, miR-210, miR-34a, and miR-146a/b are up-regulated in human atherosclerotic plaques in the Tampere Vascular Study. Atherosclerosis 219, 211–217.

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 Res. 20, 434–439.

Rampersaud, G.C., Kauwell, G.P., Hutson, A.D., Cerda, J.J., Bailey, L.B., 2000. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. Am. J. Clin. Nutr. 72, 998–1003.

Rando, T.A., Chang, H.Y., 2012. Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock. Cell 148, 46–57.

Ray, D., Wu, A., Wilkinson, J.E., Murphy, H.S., Lu, Q., Kluve-Beckerman, B., Liepnieks, J.J., Benson, M., Yung, R., Richardson, B., 2006. Aging in heterozygous Dnmt1-deficient mice: effects on survival, the DNA methylation genes, and the development of amyloidosis. J. Gerontol. A: Biol. Sci. Med. Sci. 61, 115–124.

Ribarič, S., 2012. Diet and aging. Oxid. Med. Cell. Longev. 2012, 741468.

Richardson, B., 2003. Impact of aging on DNA methylation. Ageing Res. Rev. 2, 245–261.

Rodriguez, J., Vives, L., Jordà, M., Morales, C., Muñoz, M., Vendrell, E., Peinado, M.A., 2008. Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells. Nucleic Acids Res. 36, 770–784.

Sacco, S.M., Horcajada, M.-N., Offord, E., 2013. Phytonutrients for bone health during ageing. Br. J. Clin. Pharmacol. 75, 697–707.

Salminen, A., Kaarniranta, K., 2009. SIRT1 regulates the ribosomal DNA locus: epigenetic candles twinkle longevity in the Christmas tree. Biochem. Biophys. Res. Commun. 378, 6–9.

Sanchez-Roman, I., Gomez, A., Gomez, J., Suarez, H., Sanchez, C., Naudi, A., Ayala, V., Portero-Otin, M., Lopez-Torres, M., Pamplona, R., Barja, G., 2011. Forty percent methionine restriction lowers DNA methylation, complex I ROS generation, and oxidative damage to mtDNA and mitochondrial proteins in rat heart. J. Bioenerg. Biomembr. 43, 699–708.

Sanchez-Roman, I., Gómez, A., Pérez, I., Sanchez, C., Suarez, H., Naudí, A., Jové, M., Lopez-Torres, M., Pamplona, R., Barja, G., 2012. Effects of aging and methionine restriction applied at old age on ROS generation and oxidative damage in rat liver mitochondria. Biogerontology 13, 399–411.

Serna, E., Gambini, J., Borras, C., Abdelaziz, K.M., Mohammed, K., Belenguer, A., Sanchis, P., Avellana, J.A., Rodriguez-Mañas, L., Viña, J., 2012. Centenarians, but not octogenarians, up-regulate the expression of microRNAs. Sci. Rep. 2, 961.

Shelnutt, K.P., Kauwell, G.P.A., Gregory 3rd, J.F., Maneval, D.R., Quinlivan, E.P., Theriaque, D.W., Henderson, G.N., Bailey, L.B., 2004. Methylenetetrahydrofolate reductase 677C→ T polymorphism affects DNA methylation in response to controlled folate intake in young women. J. Nutr. Biochem. 15, 554–560.

Siriwardhana, N., Kalupahana, N.S., Cekanova, M., LeMieux, M., Greer, B., Moustaid-Moussa, N., 2013. Modulation of adipose tissue inflammation by bioactive food compounds. J. Nutr. Biochem. 24, 613–623.

Smith-Vikos, T., Slack, F.J., 2012. MicroRNAs and their roles in aging. J. Cell Sci. 125, 7–17.

So, K., Tamura, G., Honda, T., Homma, N., Waki, T., Togawa, N., Nishizuka, S., Motoyama, T., 2006. Multiple tumor suppressor genes are increasingly methylated with age in non-neoplastic gastric epithelia. Cancer Sci. 97, 1155–1158.

Spazzafumo, L., Olivieri, F., Abbatecola, A.M., Castellani, G., Monti, D., Lisa, R., Galeazzi, R., Sirolla, C., Testa, R., Ostan, R., Scurti, M., Caruso, C., Vasto, S., Vescovini, R., Ogliari, G., Mari, D., Lattanzio, F., Franceschi, C., 2013. Remodelling of biological parameters during human ageing: evidence for complex regulation in longevity and in type 2 diabetes. Age Dordr. Neth. 35, 419–429.

Speakman, J.R., Mitchell, S.E., 2011. Caloric restriction. Mol. Aspects Med. 32, 159–221.

Stefanska, B., Karlic, H., Varga, F., Fabianowska-Majewska, K., Haslberger, A., 2012. Epigenetic mechanisms in anti-cancer actions of bioactive food components—the implications in cancer prevention. Br. J. Pharmacol. 167, 279–297.

Stein, P.K., Soare, A., Meyer, T.E., Cangemi, R., Holloszy, J.O., Fontana, L., 2012. Caloric restriction may reverse age-related autonomic decline in humans. Aging Cell 11, 644–650.

Switzeny, O.J., Müllner, E., Wagner, K.-H., Brath, H., Aumüller, E., Haslberger, A.G., 2012. Vitamin and antioxidant rich diet increases MLH1 promoter DNA methylation in DMT2 subjects. Clin. Epigenet. 4, 19.

Tabrez, S., Priyadarshini, M., Urooj, M., Shakil, S., Ashraf, G.M., Khan, M.S., Kamal, M.A., Alam, Q., Jabir, N.R., Abuzenadah, A.M., Chaudhary, A.G.A., Damanhouri, G.A., 2013. Cancer chemoprevention by polyphenols and their potential application as nanomedicine. J. Environ. Sci. Heal. Part C: Environ. Carcinog. Ecotoxicol. Rev. 31, 67–98.

Talens, R.P., Christensen, K., Putter, H., Willemsen, G., Christiansen, L., Kremer, D., Suchiman, H.E.D., Slagboom, P.E., Boomsma, D.I., Heijmans, B.T., 2012. Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. Aging Cell 11, 694–703.

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., Jacobs, I.J., 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.

Tian, C., Ye, X., Zhang, R., Long, J., Ren, W., Ding, S., Liao, D., Jin, X., Wu, H., Xu, S., Ying, C., 2013. Green tea polyphenols reduced fat deposits in high fat-fed rats via erk1/2-PPARγ-adiponectin pathway. PLoS One 8, e53796.

Trepanowski, J.F., Bloomer, R.J., 2010. The impact of religious fasting on human health. Nutr. J. 9, 57.

Turunen, M.P., Aavik, E., Ylä-Herttuala, S., 2009. Epigenetics and atherosclerosis. Biochim. Biophys. Acta 1790, 886–891.

Udali, S., Guarini, P., Moruzzi, S., Choi, S.-W., Friso, S., 2013. Cardiovascular epigenetics: from DNA methylation to microRNAs. Mol. Aspects Med. 34, 883–901.

Vaiserman, A.M., Pasyukova, E.G., 2012. Epigenetic drugs: a novel anti-aging strategy? Front. Genet. 3, 224.

Van Dijk, S.C., Smulders, Y.M., Enneman, A.W., Swart, K.M.A., van Wijngaarden, J.P., Ham, A.C., van Schoor, N.M., Dhonukshe-Rutten, R.A.M., de Groot, L.C.P.G.M., Lips, P., Uitterlinden, A.G., Blom, H.J., Geleijnse, J.M., Feskens, E.J., van den Meiracker, A.H., Raso, F.M., van der Velde, N., 2013. Homocysteine level is associated with aortic stiffness in elderly: cross-sectional results from the B-PROOF study. J. Hypertens. 31, 952–959.

Van Straten, E.M.E., Bloks, V.W., Huijkman, N.C.A., Baller, J.F.W., van Meer, H., Lütjohann, D., Kuipers, F., Plösch, T., 2010. The liver X-receptor gene promoter is hypermethylated in a mouse model of prenatal protein restriction. Am. J. Physiol. Regul. Integr. Comp. Physiol. 298, R275–R282.

Vanden Berghe, W., 2012. Epigenetic impact of dietary polyphenols in cancer chemoprevention: lifelong remodeling of our epigenomes. Pharmacol. Res. Off. J. Ital. Pharmacol. Soc. 65, 565–576.

Vanyushin, B.F., Nemirovsky, L.E., Klimenko, V.V., Vasiliev, V.K., Belozersky, A.N., 1973. The 5-methylcytosine in DNA of rats. Tissue and age specificity and the changes induced by hydrocortisone and other agents. Gerontologia 19, 138–152.

Vanyushin, B.F., Tkacheva, S.G., Belozersky, A.N., 1970. Rare bases in animal DNA. Nature 225, 948–949.

Vaquero, A., Reinberg, D., 2009. Calorie restriction and the exercise of chromatin. Genes Dev. 23, 1849–1869.
Vasto, S., Candore, G., Balistreri, C.R., Caruso, M., Colonna-Romano, G., Grimaldi, M.P., Listi, F., Nuzzo, D., Lio, D., Caruso, C., 2007. Inflammatory networks in ageing, age-related diseases and longevity. Mech. Ageing Dev. 128, 83–91.
Vizoso Pinto, M.G., Rodriguez Gómez, M., Seifert, S., Watzl, B., Holzapfel, W.H., Franz, C.M.A.P., 2009. Lactobacilli stimulate the innate immune response and modulate the TLR expression of HT29 intestinal epithelial cells in vitro. Int. J. Food Microbiol. 133, 86–93.
Wahlgren, J., De, L., Karlson, T., Brisslert, M., Vaziri Sani, F., Telemo, E., Sunnerhagen, P., Valadi, H., 2012. Plasma exosomes can deliver exogenous short interfering RNA to monocytes and lymphocytes. Nucleic Acids Res. 40, e130.
Wallace, K., Grau, M.V., Levine, A.J., Shen, L., Hamdan, R., Chen, X., Gui, J., Haile, R.W., Barry, E.L., Ahnen, D., McKeown-Eyssen, G., Baron, J.A., Issa, J.P.J., 2010. Association between folate levels and CpG Island hypermethylation in normal colorectal mucosa. Cancer Prev. Res. Phila. Pa 3, 1552–1564.
Watson, C.T., Disanto, G., Sandve, G.K., Breden, F., Giovannoni, G., Ramagopalan, S.V., 2012. Age-associated hyper-methylated regions in the human brain overlap with bivalent chromatin domains. PLoS One 7, e43840.
Weilner, S., Schraml, E., Redl, H., Grillari-Voglauer, R., Grillari, J., 2012. Secretion of microvesicular miRs in cellular and organismal aging. Exp. Gerontol..
Weisenberger, D.J., Campan, M., Long, T.I., Kim, M., Woods, C., Fiala, E., Ehrlich, M., Laird, P.W., 2005. Analysis of repetitive element DNA methylation by MethyLight. Nucleic Acids Res. 33, 6823–6836.
Weiss, E.P., Villareal, D.T., Racette, S.B., Steger-May, K., Premachandra, B.N., Klein, S., Fontana, L., 2008. Caloric restriction but not exercise-induced reductions in fat mass decrease plasma triiodothyronine concentrations: a randomized controlled trial. Rejuvenation Res. 11, 605–609.
Wilson, V.L., Jones, P.A., 1983. DNA methylation decreases in aging but not in immortal cells. Science 220, 1055–1057.
Wilson, V.L., Smith, R.A., Ma, S., Cutler, R.G., 1987. Genomic 5-methyldeoxycytidine decreases with age. J. Biol. Chem. 262, 9948–9951.
Wong, C.P., Nguyen, L.P., Noh, S.K., Bray, T.M., Bruno, R.S., Ho, E., 2011. Induction of regulatory T cells by green tea polyphenol EGCG. Immunol. Lett. 139, 7–13.
Yi, P., Melnyk, S., Pogribna, M., Pogribny, I.P., Hine, R.J., James, S.J., 2000. Increase in plasma homocysteine associated with parallel increases in plasma S-adenosylhomocysteine and lymphocyte DNA hypomethylation. J. Biol. Chem. 275, 29318–29323.
Zhang, X., Azhar, G., Wei, J.Y., 2012. The expression of microRNA and microRNA clusters in the aging heart. PLoS One 7, e34688.
Zhuang, J., Jones, A., Lee, S.-H., Ng, E., Fiegl, H., Zikan, M., Cibula, D., Sargent, A., Salvesen, H.B., Jacobs, I.J., Kitchener, H.C., Teschendorff, A.E., Widschwendter, M., 2012. The dynamics and prognostic potential of DNA methylation changes at stem cell gene loci in women's cancer. PLoS Genet. 8, e1002517.

References of papers in this special issue of MAD:

Berendsen, A., Santoro, A., Pini, E., Cevenini, E., Ostan, R., Pietruszka, B., Rolf, K., Cano, N., Caille, A., Lyon-Belgy, N., Fairweather-Tait, S., Feskens, E., Franceschi, C., de Groot, C.P.G.M., 2013. A parallel randomized trial on the effect of a healthful diet on inflammageing and its consequences in European elderly people: design of the NU-AGE dietary intervention study. Mech. Ageing Dev. 134, 523–530.
Santoro, A., Pini, E., Scurti, M., Palmas, G., Berendsen, A., Brzozowska, A., Pietruszka, B., Sczenziska, A., Cano, N., Meunier, N., de Groot, C.P.G.M., Feskens, E., Fairweather-Tait, S., Salvioli, S., Capri, M., Brigidi, P., Franceschi, C., the NU-AGE Consortium, 2014. Combating inflammaging through a Mediterranean whole diet approach: the NU-AGE project's conceptual framework and design. Mech. Ageing Dev. 136–137, 3–13.
Zamboni, M., Rossi, A.P., Fantin, F., Zamboni, G., Chirumbolo, S., Zoico, E., Mazzali, G., 2013. Adipose tissue, diet and aging. Mech. Ageing Dev..

Aging Cell (2012) pp1–3 Doi: 10.1111/acel.12005

SHORT TAKE

Methylation of *ELOVL2* gene as a new epigenetic marker of age

Paolo Garagnani,[1,2] Maria G. Bacalini,[1,2] Chiara Pirazzini,[1,2] Davide Gori,[3] Cristina Giuliani,[4] Daniela Mari,[5,6] Anna M. Di Blasio,[7] Davide Gentilini,[7] Giovanni Vitale,[5,7] Sebastiano Collino,[8] Serge Rezzi,[8] Gastone Castellani,[9] Miriam Capri,[1,2] Stefano Salvioli[1,2] and Claudio Franceschi[1,2]

[1]Department of Experimental Pathology, University of Bologna, Bologna, Italy
[2]Interdept. Centre for Bioinformatics, Biophysics and Biocomplexity L. Galvani, University of Bologna, Bologna, Italy
[3]Department of Medicine and Public Health, University of Bologna, Bologna, Italy
[4]Department of Experimental Evolutionary Biology, University of Bologna, Bologna, Italy
[5]Department of Medical Sciences, University of Milan, Milan, Italy
[6]Geriatric Unit IRCSS Cà Grande Foundation Maggiore Policlinico Hospital, Milan, Italy
[7]Istituto Auxologico Italiano, Milan, Italy
[8]Metabolomics and Biomarkers, Department of BioAnalytical Science, Nestlé Research Center, Nestec Ltd., Lausanne, Switzerland
[9]Physics and Biophysics, Department of Physics, University of Bologna, Bologna, Italy

Summary

The discovery of biomarkers able to predict biological age of individuals is a crucial goal in aging research. Recently, researchers' attention has turn toward epigenetic markers of aging. Using the Illumina *Infinium* HumanMethylation450 BeadChip on whole blood DNA from a small cohort of 64 subjects of different ages, we identified 3 regions, the CpG islands of *ELOVL2*, *FHL2,* and *PENK* genes, whose methylation level strongly correlates with age. These results were confirmed by the Sequenom's EpiTYPER assay on a larger cohort of 501 subjects from 9 to 99 years, including 7 cord blood samples. Among the 3 genes, *ELOVL2* shows a progressive increase in methylation that begins since the very first stage of life (Spearman's correlation coefficient = 0.92) and appears to be a very promising biomarker of aging.

Key words: aging; biomarker; DNA methylation; *ELOVL2*; *FHL2*; *PENK*.

Aging is a complex process characterized by a global decline in physiological functions and is associated with an increased risk for several diseases. A great effort is made to find reliable biomarkers of aging and epigenetic represents one of the most promising fields (Berdasco & Esteller, 2012). Epigenome-wide association studies (EWAS) report that CpG island, mainly placed within genes promoter regions, are hypermethylated in the elderly (Bell *et al.*, 2012). In this study, we analyzed the whole blood DNA methylation profile of 32 mother–offspring couples using Illumina *Infinium* HumanMethylation450 BeadChip. The age range was 42–83 and 9–52 years for mothers and offspring, respectively. ANOVA analysis identified 163 CpG sites differentially methylated between these two groups. The top 5 significant loci mapped within the CpG islands of *ELOVL2*, *FHL2* and *PENK* genes, and other loci within *ELOVL2* and *PENK* had a *P-value* below the Bonferroni threshold (Fig. 1A,B). The CpG islands of *ELOVL2*, *FHL2,* and *PENK* – all located in the respective gene promoter – resulted hypermethylated in mothers compared with offspring (Fig. 1C), with no sex or family-associated bias. Spearman's correlation analysis for the 3 genes showed striking correlation values between methylation levels and age of the subjects.

We replicated these results in a larger sex-balanced cohort using Sequenom's EpiTYPER assay. The analyzed samples included whole blood DNA from 494 individuals (245 men and 249 women) ranging from 9 to 99 years, plus 7 cord-blood DNA samples (3 males and 4 females). Depending on the sequence, the EpiTYPER assay returns the methylation values of single CpGs or small groups of close CpGs (CpG units). Samples were divided in 5 age classes, whose mean methylation values for each CpG unit are reported in Fig. 2A. We then calculated Spearman correlation between age and methylation level for each CpG unit. The highest correlation values obtained were 0.92 (CpG_11.12.13.14), 0.80 (CpG_9.10 and CpG_19.20), and 0.63 (CpG_23.24) for *ELOVL2*, *FHL2,* and *PENK,* respectively (Fig. 2B). In all cases, the considered sites tended to be hypermethylated with advancing age (Fig. 2C, for *FHL2* only the CpG_9.10 is shown). *ELOVL2* displayed the widest methylation range, from 7% to 91% (for FHL2 and PENK, they were 12% to 53% and 1% to 27%, respectively, Fig. 2C).

In each gene, we identified a subset of CpG units, which displayed high coefficients of correlation with age and whose methylation values were closely correlated with each other (Fig. 2D). The subset of highly correlated CpG units was used to perform principal components analysis (PCA). The first principal component (PC1) was calculated, and boxplot distributions of PC1 values in the 10 decades of age (Fig. 2E) showed an increase in methylation level of the considered regions.

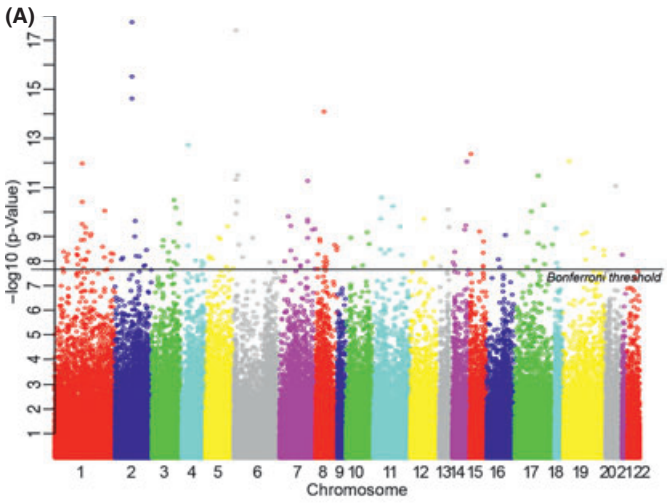
In this study, we identified and validated 3 genes, *ELOVL2*, *FHL2,* and *PENK*, whose CpG islands methylation changes with age. *FHL2* and *PENK* showed very high correlation values, but with a small difference between the different age classes. At variance, *ELOVL2* displayed not only striking correlation levels, but also an almost 'on-off' methylation trend between the two extremes of life, ranging from 7% to 91% of methylation.

The hypermethylation of CpG islands during aging is well described (Bell *et al.*, 2012), and several DNA methylation biomarkers displaying a good correlation with age have been described (Bocklandt *et al.*, 2011). In our study, taking advantage of a more dense DNA methylation microarray technology, we identified and validated more striking and reproducible age biomarkers. To date, the lack of highly reproducible aging biomarkers is explained by the high levels of heterogeneity of aging phenotypes (Cevenini *et al.*, 2008), and only within this framework it is possible to appreciate the extraordinary, progressive hypermethylation of *ELOVL2* that continuously increases from the very first stage of life to nonagenarians. We cannot exclude that this is due to numerical alterations of specific blood cell subpopulations, and further studies are needed to clarify this issue. Nevertheless, it is worth noting

Correspondence
Professor Claudio Franceschi, Department of Experimental Pathology, via S. Giacomo, 12, I-40126 Bologna, Italy. Tel.: +39 051 209 4743; fax: +39 051 209 4747; e-mail: claudio.franceschi@unibo.it

Accepted for publication *03 September 2012*

Aging Cell



(B)

Rank position	Target ID	Gene name	*p*-Value	Correlation with age
1	cg06639320	FHL2	1.82E-18	0.91
2	cg16867657	ELOVL2	3.97E-18	0.91
3	cg22454769	FHL2	3.00E-16	0.92
4	cg24079702	FHL2	2.37E-15	0.86
5	cg16419235	PENK	8.10E-15	0.76
13	cg21572722	ELOVL2	4.92E-12	0.89
26	cg24724428	ELOVL2	1.16E-10	0.85
129	cg16219603	PENK	1.05E-08	0.62
149	cg12877723	PENK	1.65E-08	0.63

Fig. 1 EWAS study. (A) Manhattan plot of genome-wide *P*-values from mother–offspring comparison. (B) *P*-values and Spearman correlation values of the selected CpG sites in *ELOVL2*, *FHL2,* and *PENK* CpG islands. The *P*-value of each probe was ranked from the smallest to the largest, and the position of the selected probes in the list is reported in the first column. (C) Methylation of the selected CpG sites respect to age. Circles: mothers; triangles: offspring.

that in previous studies of gene expression in T cells, correlations with gene expression patterns and changes of T subpopulations with age were not observed (Remondini *et al.*, 2010).

ELOVL2 encodes for a transmembrane protein involved in the synthesis of long (C22 and C24) ω3 and ω6 polyunsaturated fatty acids (PUFA) (Leonard *et al.*, 2002), and it is mainly expressed in the liver, while its expression and role in human blood cells have not been properly addressed. Genome-wide studies identified *ELOVL2* genetic variants associated with serum metabolic profile, especially with the serum concentration of specific n-3 PUFAs (Tanaka *et al.*, 2009; Lemaitre *et al.*, 2011). To date, *ELOVL2* has not been associated with aging. In light of our results, and considering that PUFAs are involved in crucial biological functions including energy production, modulation of inflammation, and maintenance of cell membrane integrity, it is possible that *ELOVL2* methylation plays a role in the aging process through the regulation of different biological pathways.

The outstanding results obtained on *ELOVL2* make it a strong candidate for forensic applications aimed at identifying proband age. To this purposes, other tissues, such as saliva and hair, should be investigated. Secondly, *ELOVL2* age-dependent hypermethylation is also a promising candidate as biomarker for the evaluation of individual fitness in elder people, with the potential for early diagnosis of age-related diseases or for monitoring therapeutic intervention or disease course.

In conclusion, (i) DNA methylation of CpG islands of *ELOVL2*, *FHL2,* and *PENK* shows strong correlation with age and, in particular *ELOVL2* is the most extreme example of age-related hypermethylation, constituting a bridge between the first developmental stages and the aging process. (ii) *ELOVL2* could be used in forensic sciences and in clinical applications. (iii) Finally, *ELOVL2* could be proposed as a sort of rheostat for aging – the more is methylated, the more aged is the subject. Further studies are needed to understand whether *ELOVL2* hypermethylation only represents an indicator of chronological age or rather is functionally correlated with physiological status and specific clinical conditions.

Acknowledgments

This research has received funding from the European Union's Seventh Framework Programme (grant agreement no. 259679, "IDEAL").

References

Bell JT, Tsai P-C, Yang T-P, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin SY, Dempster EL, Murray RM, Grundberg E, Hedman AK, Nica A, Small KS, MuTHER Consortium, Dermitzakis ET, McCarthy MI, Mill J, Spector TD, Deloukas P (2012) Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population. *PLoS Genet.* **8**, e1002629.

Berdasco M, Esteller M (2012) Hot topics in epigenetic mechanisms of aging: 2011. *Aging Cell* **11**, 181–186.

Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, Vilain E (2011) Epigenetic predictor of age. *PLoS ONE* **6**, e14821.

Cevenini E, Invidia L, Lescai F, Salvioli S, Tieri P, Castellani G, Franceschi C (2008) Human models of aging and longevity. *Expert Opin. Biol. Ther.* **8**, 1393–1405.

Lemaitre RN, Tanaka T, Tang W, Manichaikul A, Foy M, Kabagambe EK, Nettleton JA, King IB, Weng L-C, Bhattacharya S, Bandinelli S, Bis JC, Rich SS, Jacobs DR Jr, Cherubini A, McKnight B, Liang S, Gu X, Rice K, Laurie CC, Lumley T, Browning BL, Psaty BM, Chen YD, Friedlander Y, Djousse L, Wu JH, Siscovick DS, Uitterlinden AG, Arnett DK, Ferrucci L, Fornage M, Tsai MY, Mozaffarian D, Steffen LM (2011) Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genet.* **7**, e1002193.

Leonard AE, Kelder B, Bobik EG, Chuang L-T, Lewis CJ, Kopchick JJ, Mukerji P, Huang Y-S (2002) Identification and expression of mammalian long-chain PUFA elongation enzymes. *Lipids* **37**, 733–740.

Remondini D, Salvioli S, Francesconi M, Pierini M, Mazzatti DJ, Powell JR, Zironi I, Bersani F, Castellani G, Franceschi C (2010) Complex patterns of gene expression in human T cells during *in vivo* aging. *Mol. BioSyst.* **6**, 1983–1992.

Tanaka T, Shen J, Abecasis GR, Kisialiou A, Ordovas JM, Guralnik JM, Singleton A, Bandinelli S, Cherubini A, Arnett D, Tsai MY, Ferrucci L (2009) Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. *PLoS Genet.* **5**, e1000338.

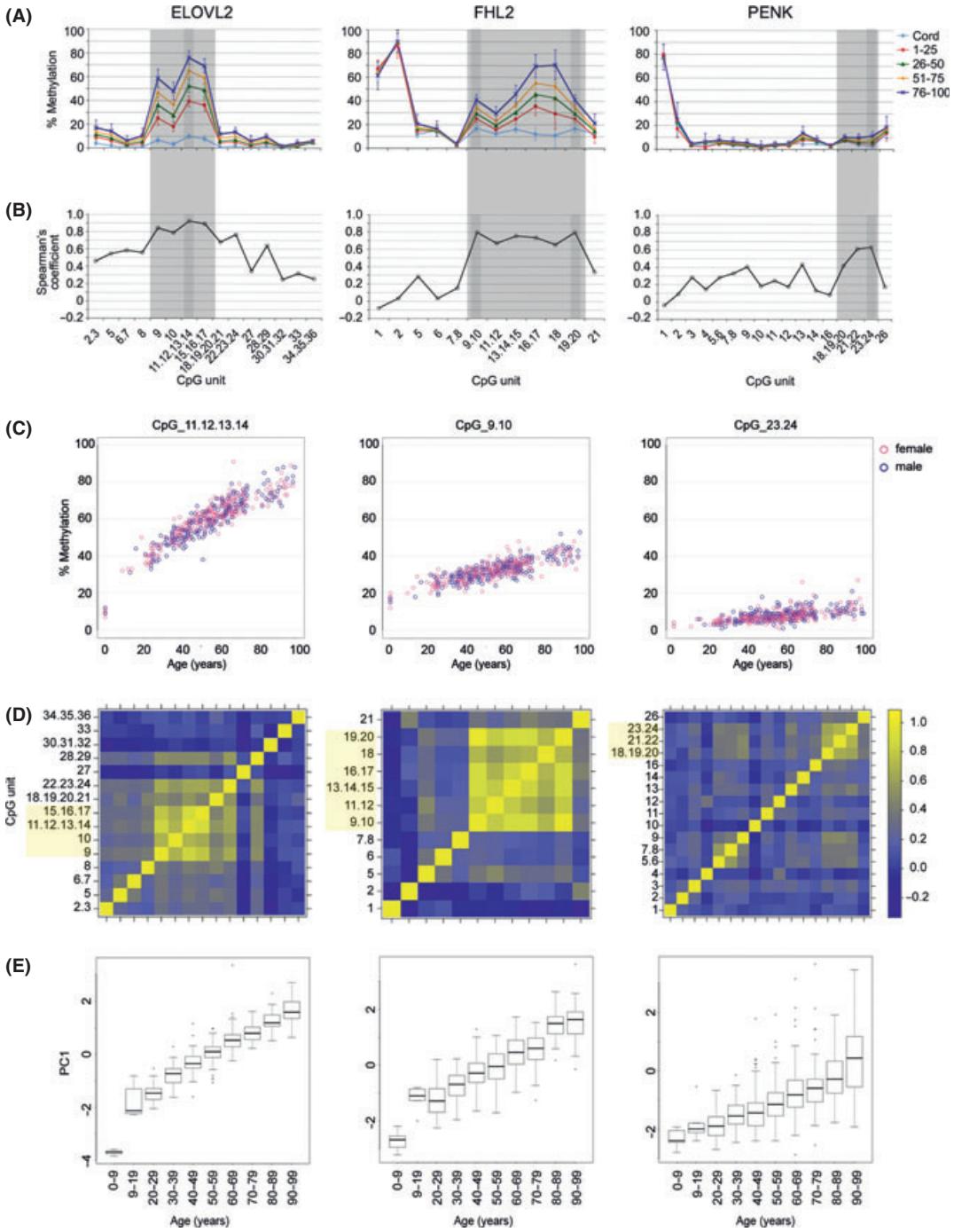


Fig. 2 Replication study. (A) Mean methylation values ± standard deviation in 5 age classes are reported for each CpG unit. (B) Spearman's correlation coefficients for each CpG unit. Highly correlated regions are marked in gray. (C) Methylation values of the CpG unit that better correlates with age in each gene. (D) Correlation between CpG units within each CpG island. The most correlated CpG units are highlighted in yellow. (E) Distribution of PC1 values calculated in 10 age classes.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Experimental procedures.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Research paper

Space/Population and Time/Age in DNA methylation variability in humans: a study on IGF2/H19 locus in different Italian populations and in mono- and di-zygotic twins of different age

Chiara Pirazzini [1,2 *], Cristina Giuliani [2,3 *], Maria Giulia Bacalini [1,2], Alessio Boattini [3], Miriam Capri [1,2], Elisa Fontanesi [2], Elena Marasco [4], Vilma Mantovani [4], Michela Pierini [1,2], Elisa Pini [1,2], Donata Luiselli [3], Claudio Franceschi [1,2], Paolo Garagnani [1,2]

[1] *CIG - Interdipartimental Center L. Galvani, University of Bologna, Bologna, 40126, Italy*
[2] *Department of Experimental Pathology, University of Bologna, Bologna, 40126, Italy*
[3] *Department of Experimental Evolutionary Biology, University of Bologna, Bologna, 40126, Italy*
[4] *CRBA - Applied Biomedical Research Center, S. Orsola-Malpighi Polyclinic, Bologna, 40138 Italy*
** equally contributed*

Key words: *DNA methylation, IGF2/H19, twins, aging, population epigenetics*
Received: *6/21/12; Accepted: 7/27/12; Published: 7/31/12*
Correspondence to: *Maria Giulia Bacalini, PhD;* ***E-mail:*** *mariagiuli.bacalini2@unibo.it*

Abstract: Little is known about the impact of space (geography/ancestry) and time (age of the individuals) on DNA methylation variability in humans. We investigated DNA methylation of the imprinted *IGF2/H19* locus in: i) a cohort of individuals homogeneous for age and gender (males with restricted age range: 30-50 years) belonging to four Italian districts representative of the major genetic clines, informative for the geographical dimension; ii) a cohort of monozygotic (MZ) and dizygotic (DZ) twins of different ages (age-range: 22-97 years), informative for the temporal dimension. DNA methylation of the analyzed regions displayed high levels of inter-individual variability that could not be ascribed to any geographical cline. In MZ twins we identified two *IGF2/H19* regions where the intra-couple variations significantly increased after the age of 60 years. The analysis of twins' individual life histories suggests that the within twin pairs difference is likely the result of the aging process itself, as sharing a common environment for long periods had no effect on DNA methylation divergence. On the whole, the data here reported suggest that: i) aging more than population genetics is responsible for the inter-individual variability in DNA methylation patterns in humans; ii) DNA methylation variability appears to be highly region-specific.

INTRODUCTION

DNA methylation is widespread across the genomes of different organisms and in mammals usually consists in the enzymatic addition of a methyl group to the carbon-5 of cytosine ring of a CpG dinucleotide. Through the recruitment of methyl-binding proteins, this modification induces an inactive chromatin structure that represses transcription. While the bulk of human genome is generally methylated, the promoters of around 40% of genes contain CpG-rich regions, termed CpG islands, whose methylation status is strictly regulated during development and cellular differentiation [1-3]. Deregulation in methylation patterns can lead to disease onset. More recently it has been shown that tissue- and disease-specific differentially methylated regions (DMR) are more frequent in CpG island shores rather than in CpG islands [4, 5].

Although DNA methylation is generally considered a stable modification, it is well established that methylation profiles vary in pathological conditions, and it is also accepted that external factors, such as diet, age, toxins and lifestyle, can induce quantitative hypo- or hyper- DNA methylation [6-13]. Only few studies have focused the effect of such factors on inter-

individual variability of DNA methylation patterns among human populations and subjects [14-17]. Regarding time/age, variations in methylation levels of the *IGF2/H19* locus were observed in a cohort of newborn twins, both between individuals and within twin pairs [17]. In this study, dizygotic twins (DZ) resulted more discordant than monozygotic twins (MZ), suggesting that a heritable component can affects the epigenetic status of this locus. Moreover, Heijmans and colleagues [16] analyzed the variations in DNA methylation of the same *IGF2/H19* locus in a cohort of adolescent and middle-aged twins (13-62 years of age) and demonstrated that a substantial part of the variation observed across individuals was ascribable to heritable factors and single nucleotide polymorphisms (SNPs) *in cis*, rather than to the cumulative effect of environmental and stochastic factors occurring with age. Although this study involved a high number of twins (N=372), it did not include old twin pairs, and therefore could not appreciate the possible epigenetic variability of the locus occurring later in life (after sixties).

Regarding space/population, only few studies addressed the DNA methylation variability at the population level (*population epigenetics*) [18], despite its great potential interest owing to the interaction of both environmental and genetic variables in determining the DNA methylation architecture. To our knowledge, the only two available studies have investigated DNA methylation in multiethnic cohorts of healthy women [19] and of men affected by prostate cancer [20].

To better understand the intricate relationship between spatial (geography/ancestry) and temporal (age of the individuals) dimensions on DNA methylation variability, we took advantage of two *ad hoc* models: i) Cohort 1, constituted of 376 individuals homogeneous for gender (males) and age (age range 30-51), but differing for ancestral geographical origin and place of living, i.e four Italian regions (Northern, Central and Southern Italy, and Sardinia); ii) Cohort 2, constituted of 31 monozygotic (MZ) and 16 dizygotic (DZ) twin couples of different ages, spanning from 22 to 97 years, homogeneous for geographic origin (Northern Italy, Bologna area). Both the models were analyzed for variations in DNA methylation status of 4 target regions in the imprinted *IGF2/H19* locus, previously analyzed by other groups [16, 17].

RESULTS

Characterization of the *IGF2/H19* target regions

Four target regions (amplicons) in the *IGF2/H19* locus were analyzed: *IGF2AS*, 3 kb from a CpG island, in exon 3 of *IGF2AS* transcript, within the *IGF2* DMR 0; *H19*, upstream the transcription start site of *H19* gene, partially overlapping a CpG island (*H19* DMR); *IGF2_island*, within a CpG island in the last exon of *IGF2* gene (DMR 2); *IGF2*_shore, in the shore upstream the island targeted by the *IGF2_island* amplicon (Figure 1A). *IGF2AS* and *H19* have been previously analyzed by Heijmans et al. [16] and Ollikainen et al. [17], while *IGF2_island* and *IGF2_shore* were analyzed here for the first time.

Mean methylation levels and inter-individual variability of *IGF2/H19* locus

Figure 1B reports mean methylation values and their standard deviations for the 4 target regions considering cohort 1 and 2 separately. For cohort 1 the mean DNA methylation levels observed in *IGF2AS, IGF2_shore, IGF2_island* and *H19* were 49%, 60%, 43% and 36% respectively. These values are similar for cohort 2, where the mean methylation values in *IGF2AS, IGF2_shore, IGF2_island* and *H19* were 49%, 62%, 47% and 36%. Notably *IGF2AS* and *H19* DNA methylation values were comparable to those previously reported in literature [16, 17].

Considerable inter-individual variation in DNA methylation values was observed within each CpG unit. Within a single amplicon, DNA methylation values of the CpG units were strongly correlated with each others, while correlation values were lower between CpGs belonging to different amplicons (Figure 1C). Intermediate correlation levels were observed only between *IGF2_island* and *IGF2_shore*, as expected because of their adjacent chromosomal position.

Space/ancestry: geographical dimension

The Italian cohort comprises 376 middle-aged males (age range: 30-51 years) representative of the four main Italian geographical areas: Northern Italy (141 samples), Central Italy (94 samples), Southern Italy (121 samples) and Sardinia (20 samples) (Figure 2A). We evaluated if the observed variability in methylation levels of the *IGF2/H19* locus could be explained by ancestry and place of living of the four populations studied.

Mean methylation levels of each CpG unit were calculated for the 4 geographical areas (Figure 2B). The results showed that the methylation patterns were comparable throughout all the samples. No evident differences were detectable according to sample ancestry/geography, except for few CpGs (*i.e.* CpG 20 of *IGF2_island* amplicon).

Principal component analysis (PCA) was used to capture the most salient patterns of variation in each amplicon for the 4 geographical areas. The percentages of variance explained by principal component 1 (PC1) values are reported in Figure 2C. No statistically significant difference was observed between PC1 values, as confirmed by analysis of variance (ANOVA) (Table 1).

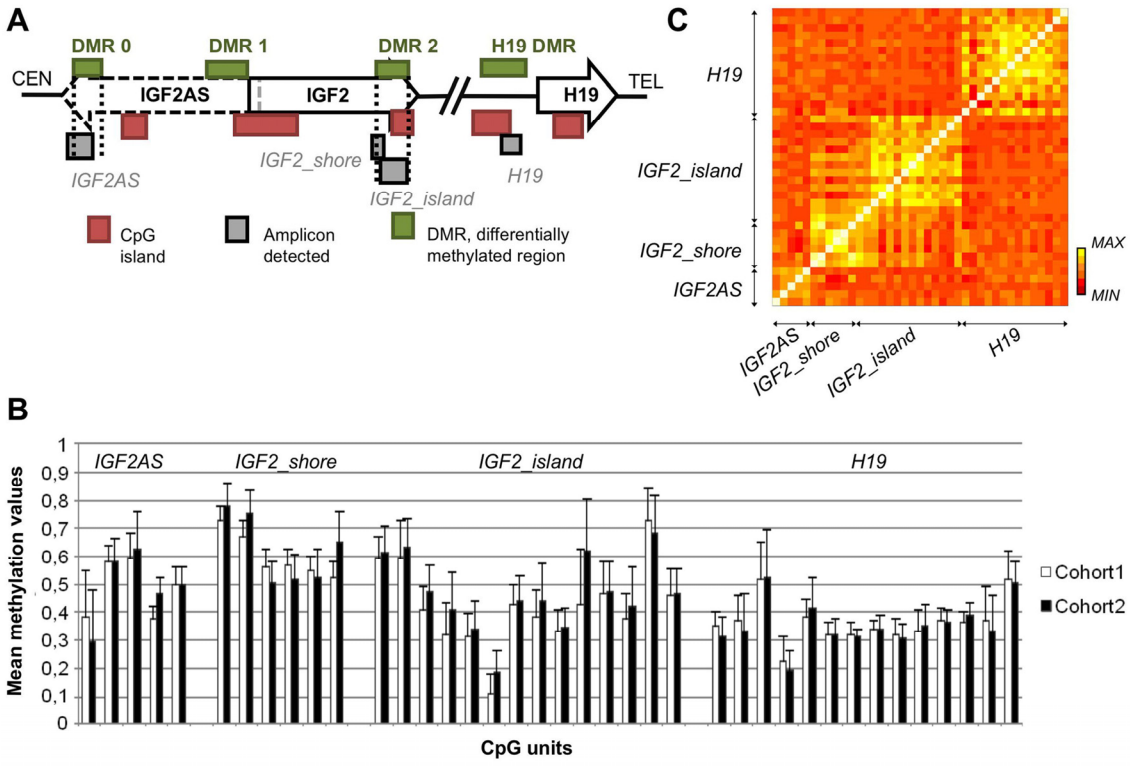


Figure 1. CpGs spatial localization and their methylation patterns. (**A**) Spatial disposition of the amplicon analyzed in relation to functional elements. (**B**) Mean methylation level with the corresponding standard deviation for each CpGs considering the whole dataset. (**C**) Correlation matrix of CpG sites of *IGF2_shore*, I*GF2AS*, *H19* and *IGF2_island* amplicons

Table 1. *p_values* from ANOVA on PC1 values calculated based on DNA methylation levels in samples belonging to different Italian districts

	ANOVA – *p_values*
IGF2_shore	0.93
IGF2_island	0.45
IGF2AS	0.80
H19	0.44

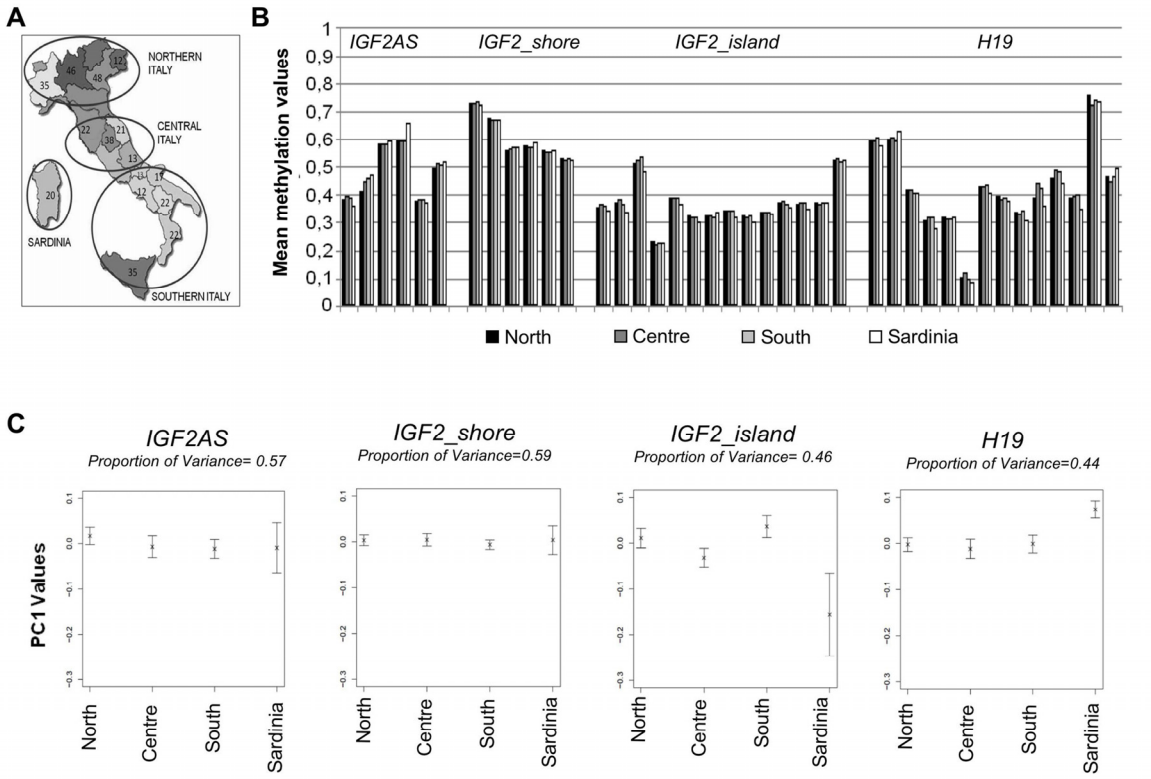


Figure 2. Methylation variation in cohort 1. (**A**) Number of samples from different Italian districts and corresponding groups made for the statistical analysis. (**B**) Mean methylation levels of each CpGs unit in the 4 macroarea. (**C**) PC1 values plotted for each macro-area and the corresponding percentage of variance.

Time/age: temporal dimension

The twins cohort (cohort 2) includes 94 twins (62 MZ and 32 DZ) with age ranging from 22 to 97 years all recruited in the area of Bologna (Emilia Romagna, Italy). Cohort 2 was divided into four age classes (20-45, 46-60, 61-75, 76-97) and for each amplicon, mean methylation value and mean standard deviation were calculated (Figure 3). No significant variations in mean methylation levels were observed. In the first 2 age classes IGF2_shore was characterized by a minor range of variation if compared to the other amplicons. Moreover, while mean standard deviations in *IGF2_island*, IGF2AS and H19 were stable over the time, *IGF2_shore* showed a doubling in their values in the older age classes.

In order to estimate DNA methylation variability within the MZ twins couples, Pearson's correlation and Euclidean distance values for each twins pair were calculated. Pearson's correlation was calculated considering methylation values of each CpG from the same amplicon. The correlation values were then plotted against the age of the subjects (Figure 4). H19 was characterized by a high degree of within-twin pairs correlation (Figure 4A) which was maintained up to old ages (Spearman's correlation with age ρ= -0.094). *IGF2_island* also did not display variations of correlation levels with age (Spearman's correlation with age ρ = -0.105). However, in this case the intra-couple correlation scores were lower, indicating that high levels of variability characterize the CpG island of *IGF2* (Figure 4B). Interestingly, both *IGF2AS* and *IGF2_shore* showed an age dependent trend (Spearman's correlation ρ = -0.613 and ρ = -0.651 respectively), with high correlation levels till the age of 50, followed by a progressive decrease of correlation in the couples of older age (Figure 4C-4D). A possible threshold between the age of 56 and 60 years can be envisaged, and after such age limit the values of intra-couple correlation levels are much more scattered.

Then, the Euclidean distance between the CpG units of each amplicon. was calculated for each MZ twins pair.

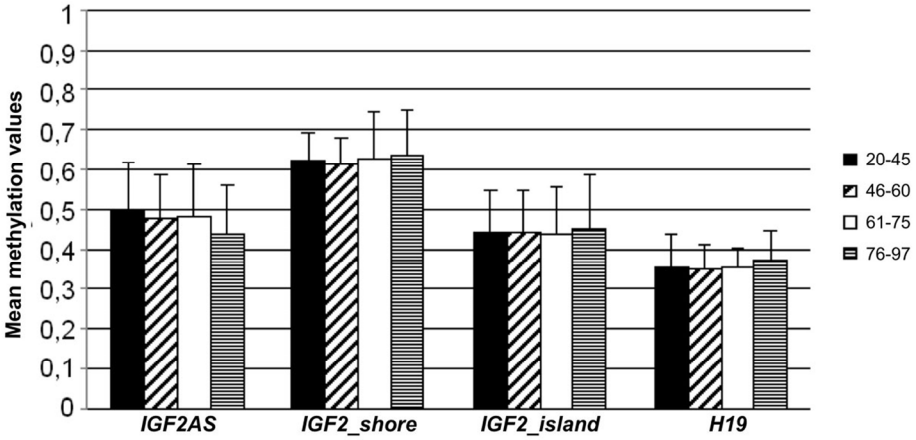


Figure 3 Methylation variation in cohort 2. The twins cohort was divided in 4 age classes (20-45, 46-60, 61-75,76-97). Mean methylation values and the corresponding standard deviation for the 4 analyzed amplicons (*IGF2AS, IGF2_shore, IGF2_island* and *H19*) are reported for each age-class.

Table 2. Student t-test on correlation and Euclidean distance scores using 60 years old as age threshold

Amplicon	Age limit	Values considered	*p_values*
IGF2_shore	60	Correlation	0.00047***
IGF2AS	60	Correlation	0.022*
H19	60	Correlation	0.27
IGF2_island	60	Correlation	0.77
IGF2_shore	60	Euclidean distance	0.0075**
IGF2AS	60	Euclidean distance	0.023*
H19	60	Euclidean distance	0.48
IGF2_island	60	Euclidean distance	0.57

Regarding *H19* and *IGF2_island* the distance values did not show any age-dependent trend (Figure 5A and 5B, Spearman's correlation $\rho = 0.099$ $\rho = 0.216$ respectively). On the contrary, *IGF2AS* and *IGF2_shore* (Figure 5C and 5D) showed an age-dependent increase in the intra-couple Euclidean distances (Spearman's correlation $\rho = 0.462$ and $\rho = 0.553$ respectively), with a scatter of the distance values after the age threshold of 60, as reported for Pearson's correlation analysis.

The data were also subdivided according to sex to exclude an influence on the patterns of correlation and distance values (Figure 4 and Figure 5).

To confirm the previous observations, samples from cohort 2 were divided into 2 groups according to age (below and over 60 years old) and Student t-test on correlation and Euclidean distance values was performed. Significant differences between the 2 groups were found only for *IGF2_shore* and *IGF2AS* amplicons (Table 2).

Evaluation of heritability and environmental influences on DNA methylation

MZ and DZ twins were used to estimate the genetic and environmental influences on DNA methylation of the *IGF2/H19* locus.

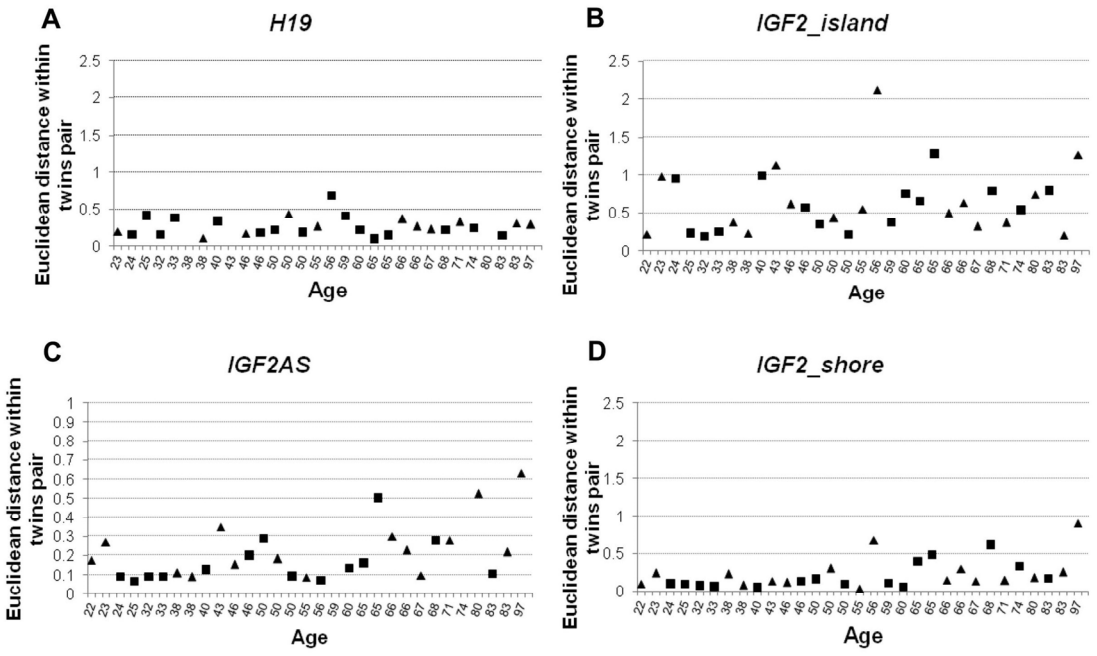


Figure 4. Pearson correlation values calculated within twin pairs in the 4 analyzed amplicons. (**A**) Correlation values within MZ twin couple in *H19* amplicon. (**B**) Correlation values within MZ twin couple in *IGF2_island*. (**C**) Correlation values within MZ twin couple in IGF2AS. (**D**) Correlation values within MZ twin couple in *IGF2_shore* amplicon. In each figure male individuals are indicated with squares and female using a triangle.

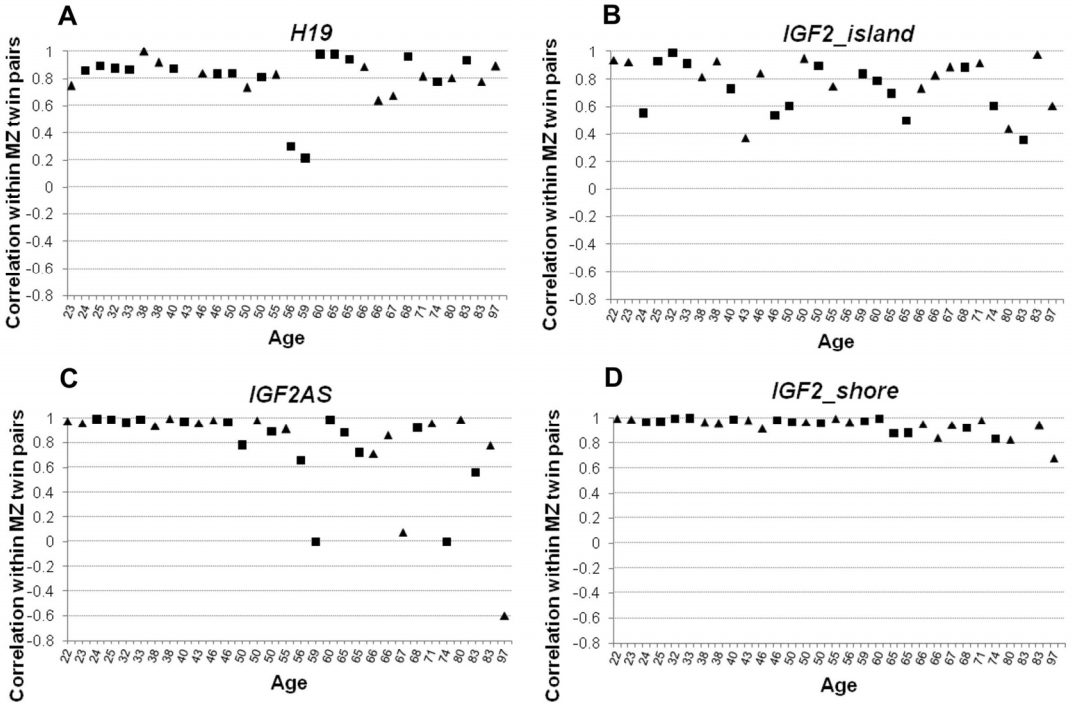


Figure 5. Euclidean distance calculated using DNA methylation values within MZ twin pairs considering the 4 considered amplicons. (**A**) Distance within twin pairs in *H19* amplicon. (**B**) Distance values within twin pairs in *IGF2_island* amplicon. (**C**) Distance values within twin pairs in *IGF2AS* amplicon. (**D**) Distance values within twin pairs in *IGF2_shore* amplicon. In each figure male individuals are indicated with squares and female using a triangle

Firstly for each amplicon intra-couple levels of correlation between MZ and DZ twins were compared. Like MZ, also DZ twins showed high values of correlation in the 4 amplicons (Figure 6), suggesting a strong epigenetic control on this locus. In *IGF2AS, H19* and *IGF2_island*, the intra-couple correlation in MZ was higher compared with that of DZ. Only in *IGF2_shore* median correlation values were similar among MZ (r=0.94) and DZ (r=0.90) twins. Falconer method confirmed this observation, showing that IGF2_shore is more influenced by environmental factors (h2=0.07) when compared with *IGF2AS* (h2=0.37), *H19* (h2=0.34) and *IGF2_island* (h2=0.20). As *IGF2_shore* methylation seems to be more affected by environmental factors, the total lifetime in which the twins did not shared the same environment (i.e. the total lifetime in which they did not live in the same house) was considered. As expected, an increase in percentage of life in which the twins lived separated (S = (years lived in a different house / age)*100) was observed for older couples (Figure 7A).

Low Euclidean distance levels characterized twins with small S values, while a scatter after the threshold of *S*=35, which in our cohort corresponds to the age of 60, was observed (Figure 7B). There were 3 main exceptions, indicated with an arrow in Figure 7B. These subjects (aged 50, 56 and 74) were still living in the same house at the time of recruitment (small S values) but were characterized by high Euclidean distance values for *IGF2_shore*.

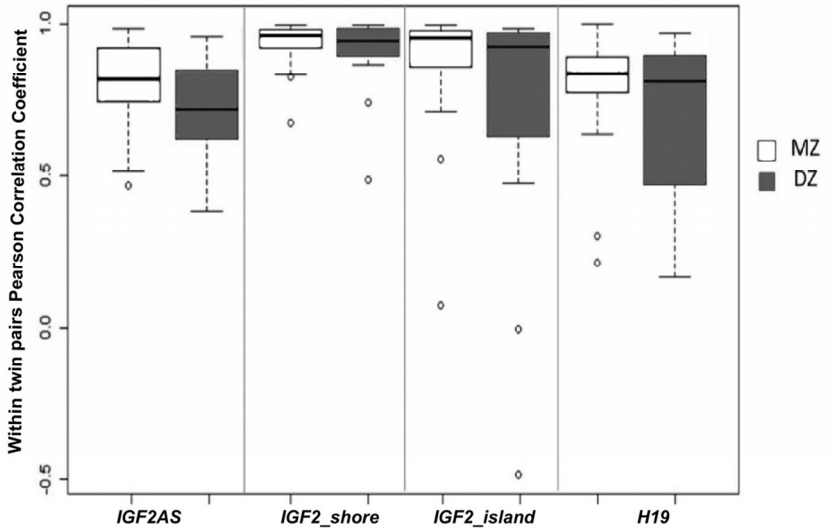


Figure 6. Comparison between intra-couple correlation (ICC) in MZ and DZ twins. Boxplot of ICC for each amplicon dived on the basis of zygosity.

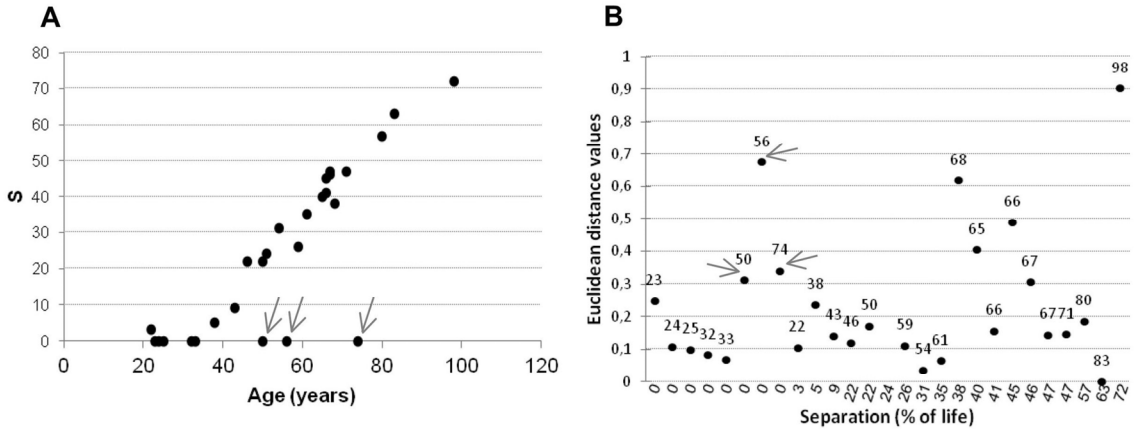


Figure 7. Analysis of twins individual histories. (**A**) Percentage of years lived separately for each twin pairs in relation to their ages. (**B**) Euclidean distance within twin pairs related to years the twins lived separately in *IGF2_shore* amplicon. S = (years lived in a different house / age)*100

DISCUSSION

Our results indicate that the 4 analyzed regions in the *IGF2/H19* locus show high inter-individual variability both in cohort 1 (spatial/population dimension) and in cohort 2 (temporal/aging dimension). *IGF2AS* and *H19* have mean methylation values comparable to previous reports [16, 17], indicating that methylation profiles are consistent among different populations and that the experimental technique used for methylation analysis is highly reproducible between independent laboratories. On the contrary, *IGF2_island* and *IGF2_shore* were analyzed for their methylation values here for the first time.

Previous studies have analyzed methylation variability in multiethnic populations, but the enrolled subjects lived in the same geographical area and were therefore exposed to the same environmental factors [19, 20]. On the contrary, cohort 1 was selected in order to include individuals both belonging to different genetic clines and living in their original geographical area. Our results show that methylation variability in cohort 1 cannot be explained on the basis of the geographic provenience of subjects. This implies that the genetic structure of the Italian population does not influence the methylation pattern of the analyzed loci. At the same time, it means that there are no environmental cues specifically associated with Italian macro-areas able to generate clear population epigenetic signatures in the *IGF2/H19* locus.

Methylation of IGF2 promoter has been shown to increase with aging [21]. On the contrary, mean methylation values of the 4 regions that we considered in *IGF2/H19* locus do not change according to age in cohort 2. This suggests that methylation levels of different regions in the same locus are differently affected by the aging process. At the same time, in young individuals from cohort 2, DNA methylation variations is smaller in *IGF2_shore* respect to the other analyzed amplicons indicating a stricter control of its methylation levels. Accordingly previous studies have highlighted that the range of methylation variation depends on the genomic position [14]. Moreover, only in *IGF2_shore* the range of methylation values increases in the elderly suggesting a clear susceptibility of this region to epimutations that occur with aging.

MZ twins model allowed the investigation of the maintenance of DNA methylation profiles during the lifespan. *IGF2AS* and *IGF2_shore* methylation levels were highly concordant within twin pairs, but only until the 60 years age threshold. Homogeneous methylation values in *IGF2AS* and *IGF2_shore* until the age of 60 years can be in part due to the genetic identity and the sharing of intrauterine and neonatal environments. It is interesting to observe that the threshold of 60 years coincides with the end or with the significant reduction of the reproductive capacity of individuals. A loss in molecular fidelity after the threshold of 60 years has been described [22]. This can affect also the DNA methylation machinery, impairing its ability to maintain the methylation patterns across cellular divisions [23]. The extent and the rate of this process can be influenced by the genetic background of subjects [24], accordingly in our cohort, a subset of couples maintains the concordance in DNA methylation profiles also at old age.

Comparison between MZ and DZ twins confirms a genetic component in the variability observed in *IGF2/H19* locus in our cohort. These analysis highlight also an effect of non-genetic factor, which is stronger for *IGF2_shore*. This observation is supported by the 3 old twins pairs that had always lived together, in the same house, sharing the same daily habits. They show high distance values in *IGF2_shore*, similar to those observed in the old couples that lived separately. This data suggests that age-related epimutations are independent from the shared environment and are more likely related to individual histories or stochasticity. A larger sample of old twins that lived together should be considered to confirm this hypothesis.

Previous data show that different imprinted genes have different ability to maintain their methylation status [25]. Our results not only confirm these observations, but also show that within the same imprinted locus there are regions whose methylation maintenance is differently influenced by aging.

It is interesting to highlight that, between the 4 regions that we have analyzed in *IGF2/H19* locus, *IGF2_shore* is the one in which the scatter in methylation variability during aging is more evident. This result confirms recent evidences that indicate CpG shores as the regions most variable in terms of methylation levels across normal and pathological tissues and between different cell types [4].

The ability of imprinted genes to maintain their methylation profile has been proposed as a marker of epigenetic stability [25]. In fact, although imprinted loci are under a strict epigenetic control, a relaxation in their regulation can be observed both in pathological [26] and normal conditions [27]. Further studies are needed to deepen our knowledge on age-related DNA methylation variability of imprinted genes, with specific attention to the shore elements.

Table 3. Cohort 1 characteristics

	Twins		All
	Monozygotic	**Dizygotic**	**Total samples**
N	62 (31 pairs)	32 (16 pairs)	94 (47 pairs)
Mean age **(SD; range)**	54.0 (19.4; 22-97)	57.8 (12.8;36-79)	55.3 (17.4; 22-97)
Male (%)	32 (51.6%)	20 (62.5%)	52 (55.3%)

In conclusion our data confirm the loss of control over DNA methylation maintenance with age only in specific genomic regions (shore). In particular we observe a sudden increase in intra-couple variability from the age of 60, when the reproductive capacities decline and, as a consequence, the effect of natural selection falls off. On one side this decline could be interpreted as a loss of capacities of transmitting the methylation patterns across cellular divisions. On the other side, as the weaker epigenetic control can result in a reduction of gene expression control, it is possible to speculate that this mechanism could represent a molecular strategy to counteract the physiological fitness decline that occurs during aging by offering a wider range of expression possibilities.

MATERIALS AND METHODS

Ethics statement. Informed consent was obtained from all partecipants and was approved by the Medical Ethics Committee of the S. Orsola-Malpighi Polyclinic (Bologna, Italy).

Samples. *Twins:* The twins cohort was recruited between 2004 and 2005 in Bologna and neighboring districts to reduce possible bias associated to different geographical origin of the samples. The cohort comprises 47 twin pairs (31 MZ an 16 DZ pairs) with a mean age of 55 years (age range: 22-97 years; Table 3). *Italian population:* in order to assemble the most representative and informative Italian sample, with an ancient regional ancestry in the area, and an adequate coverage for both the number of populations and the number of individuals within each population sample, an appropriate and accurate sampling strategy was built based on a preliminary surname-based study (Boattini et al., submitted).

Groups of homogeneous provinces ("sampling areas") were aggregated according to the surname-based clusters. Provinces ("sampling points") were selected within each sampling area according to the complex geographic and historical Italian backgrounds. Individuals ("samples") were chosen, within each province, based on two different strategies, the "grandparents criterion" and "founder surnames analysis", so that were included into the study only those individuals whose four grandparents were born in the same sampling area, taking also into account the presence of founder surnames.

Using this sampling strategy 378 male individuals from 29 provinces scattered in 15 regions of Italy were collected (Figure 2A) .

DNA methylation measurement and associated data cleaning. The level of DNA methylation was measured on genomic DNA using a MALDI-TOF mass spectrometry-based method (Epityper, Sequenom, San Diego, CA) as previously described [28]. DNA was extracted from whole blood (QIAamp 96 DNA Blood Kit, Qiagen, Hilden, Germany), quantified using Picogreen (Quant-iT dsDNA Broad-Range Assay Kit Invitrogen, Carlsbad, CA), and 1 μg was treated with sodium bisulfite using the EZ methylation kit (Zymo-Research, Irvine, CA). The treatment converts non-methylated cytosine into uracil, leaving methylated cytosine unchanged. In this way variations in the sequence are produced depending on DNA methylation status of the original DNA molecule. PCR amplification, addition of SAP solution and Transcription/RNase A cocktails were performed according to the protocol provided by Sequenom and the mass spectra were analyzed by EpiTYPER analyzer (Sequenom, San Diego, CA). *IGF2_shore* amplicon encompassed 276 bp (NCBI build 36, chr11: 2 111 039 - 2 111 314). IGF2AS encompassed 338 bp (NCBI build 36, chr11: 2 126 035 - 2 126 372). H19 encompassed 413 bp (NCBI build 36, chr11: 1 975 948 - 1 976 360). IGF2_island encompassed 453 bp (NCBI build 36, chr11: 2 111 119 - 2 110 666). The primers sequences used to amplify this regions are listed in Table 4.

Table 4. CpG Units considered after data cleaning and primers used to amplify the 4 target regions

Amplicon Name	CpG sites after data cleaning	Total Cpg Units	Primers sequences
IGF2_ shore	CpG 1,2 – CpG 3,4 - CpG 6 - CpG 7 - CpG 8 - CpG 9,10,11	6	Forward: aggaagagagGAAGGGGTTGGTTAGTAGGTGTTTGT Reverse: cagtaatacgactcactatagggagaaggctCCTAAACCCCTTTCCCACTCTCTAA
IGF2_ island	CpG 1,2 - CpG 4 - CpG 11,12 - CpG 13 - CpG 14 - CpG 15 - CpG 16 - CpG 17 - CpG 18,19 - CpG 20 - CpG 21 - CpG 25 - CpG 26 - CpG 27	14	Forward: aggaagagagTATAGGGGTGGTTTGTTAGGTTAGG Reverse: cagtaatacgactcactatagggagaaggcTAAATCAAAAAAAACCCCAAAAAAAC
IGF2AS	CpG 1 - CpG 3 - CpG 4 - CpG 6,7 - CpG 8	5	Forward: aggaagagagTGGATAGGAGATTGAGGAGAAA Reverse: cagtaatacgactcactatagggagaaggctAAACCCCAACAAAAACCACT
H19	CpG 1 - CpG 2 - CpG 6 - CpG 7 - CpG 8 - CpG 9,10 - CpG 12 - CpG 13 - CpG 14,15 - CpG 17 - CpG 18,19 - CpG 20 - CpG 22 - CpG 24	14	Forward: aggaagagagGGGTTTGGGAGAGTTTGTGAGGT Reverse: cagtaatacgactcactatagggagaaggctATACCTACTACTCCCTACCTACCAAC

In this way, a total of 50 CpG units in the 4 gene regions were interrogated for their methylation level. A rigorous data cleaning process was performed to remove unreliable measurements before statistical analysis. Firstly, DNA samples for which methylation level could be established for <60% of the CpG sites were removed. Secondly, CpG sites with more than 30% missing data points within an amplicon were not analyzed. Finally CpG sites that show a bimodal pattern due to the presence of a SNP were removed. Six CpG sites on four fragments could not be measured independently because the fragments had the same molecular weight and were overlapping in the spectrum. After data cleaning, we were able to analyze a total of 39 CpG units in the 4 genomic regions (Table 4).

Statistical analysis. All analysis were performed using R (http://cran.r-project.org/). Prior to statistical analysis, methylation data were checked for complete bisulfite conversion using the MassArray package [29]. h^2 values were calculated according to Falconer's formula [30].

ACKNOWLEDGEMENTS

I thank Wei Gu, Arnold Levine and Bert Vogelstein for critical reading of the manuscript and excellent suggestions. The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement n° 259679 ("IDEAL").

CONFLICT OF INTERESTS STATEMENT

The author of this manuscript has no conflict of interest to declare.

REFERENCES

1. Reik W, Dean W and Walter J. Epigenetic reprogramming in mammalian development. Science (New York, NY). 2001; 293:1089-1093.

2. Morgan HD, Santos Ft, Green K, Dean W and Reik W. Epigenetic reprogramming in mammals. Human Molecular Genetics. 2005; 14 Spec No 1:R47-58-R47-58.

3. Feng S, Jacobsen SE and Reik W. Epigenetic reprogramming in plant and animal development. Science (New York, NY). 2010; 330:622-627.

4. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash JB, Sabunciyan S, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature Genetics. 2009; 41:178-186.

5. Doi A, Park I-H, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nature Genetics. 2009; 41:1350-1353.

6. Aguilera O, Fernandez AF, Munoz A and Fraga MF. Epigenetics and environment: a complex relationship. J Appl Physiol. 2010; 109:243-251.

7. Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, Slagboom PE and Heijmans BT. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Human Molecular Genetics. 2009; 18:4046-4053.

8. Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE and Lumey LH. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:17046-17049.

9. Waterland RA, Lin J-R, Smith CA and Jirtle RL. Post-weaning diet affects genomic imprinting at the insulin-like growth factor 2 (Igf2) locus. Human Molecular Genetics. 2006; 15:705-716.

10. Heijmans BT, Tobi EW, Lumey LH and Slagboom PE. The epigenome: archive of the prenatal environment. Epigenetics: Official Journal of the DNA Methylation Society. 2009; 4:526-531.

11. Champagne FA and Curley JP. Epigenetic mechanisms mediating the long-term effects of maternal care on development. Neuroscience and biobehavioral reviews. 2009; 33:593-600.

12. Weaver ICG, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, Dymov S, Szyf M and Meaney MJ. Epigenetic programming by maternal behavior. Nature neuroscience. 2004; 7:847-854.

13. Weaver ICG, Champagne FA, Brown SE, Dymov S, Sharma S, Meaney MJ and Szyf M. Reversal of maternal programming of stress responses in adult offspring through methyl supplementation: altering epigenetic marking later in life. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2005; 25:11045-11054.

14. Schneider E, Pliushch G, El Hajj N, Galetzka D, Puhl A, Schorsch M, Frauenknecht K, Riepert T, Tresch A, Mvºller AM, Coerdt W, Zechner U and Haaf T. Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns. Nucleic acids research. 2010; 38:3880-3890.

15. Bock C, Walter Jr, Paulsen M and Lengauer T. Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping. Nucleic acids research. 2008; 36:e55-e55.

16. Heijmans BT, Kremer D, Tobi EW, Boomsma DI and Slagboom PE. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. Human Molecular Genetics. 2007; 16:547-554.

17. Ollikainen M, Smith KR, Joo EJ-H, Ng HK, Andronikos R, Novakovic B, Abdul Aziz NK, Carlin JB, Morley R, Saffery R and Craig JM. DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Human Molecular Genetics. 2010; 19:4176-4188.

18. Liu J, Hutchison K, Perrone-Bizzozero N, Morgan M, Sui J and Calhoun V. Identification of genetic and epigenetic marks involved in population structure. PloS One. 2010; 5:e13209-e13209.

19. Terry MB, Ferris JS, Pilsner R, Flom JD, Tehranifar P, Santella RM, Gamble MV and Susser E. Genomic DNA methylation among women in a multiethnic New York City birth cohort. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2008; 17:2306-2310.

20. Das PM, Ramachandran K, Vanwert J, Ferdinand L, Gopisetty G, Reis IM and Singal R. Methylation mediated silencing of TMS1/ASC gene in prostate cancer. Molecular cancer. 2006; 5:28-28.

21. Issa JP, Vertino PM, Boehm CD, Newsham IF and Baylin SB. Switch from monoallelic to biallelic human IGF2 promoter methylation during aging and carcinogenesis. Proceedings of the National Academy of Sciences of the United States of America. 1996; 93:11757-11762.

22. Hayflick L. Entropy explains aging, genetic determinism explains longevity, and undefined terminology explains misunderstanding both. PLoS genetics. 2007; 3:e220-e220.

23. Goyal R, Reinhardt R and Jeltsch A. Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. Nucleic acids research. 2006; 34:1182-1188.
24. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekstrv∂m TJ, Harris TB, Launer LJ, Eiriksdottir G, Leppert MF, et al. Intra-individual change over time in DNA methylation with familial clustering. JAMA: The Journal of the American Medical Association. 2008; 299:2877-2883.
25. Woodfine K, Huddleston JE and Murrell A. Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. Epigenetics & chromatin. 2011; 4:1-1.
26. Cui H. Loss of imprinting of IGF2 as an epigenetic marker for the risk of human cancer. Disease markers. 2007; 23:105-112.
27. Sakatani T, Wei M, Katoh M, Okita C, Wada D, Mitsuya K, Meguro M, Ikeguchi M, Ito H, Tycko B and Oshimura M. Epigenetic heterogeneity at imprinted loci in normal populations. Biochemical and Biophysical Research Communications. 2001; 283:1124-1130.
28. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK and van den Boom D. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:15785-15790.
29. Thompson RF, Suzuki M, Lau KW and Greally JM. A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry. Bioinformatics (Oxford, England). 2009; 25:2164-2170.
30. Falconer DS. (1989). Introduction to quantitative genetics. (Burnt Mill, Harlow, Essex, England, New York: Longman Wiley).

Research Paper

Transmission from centenarians to their offspring of mtDNA heteroplasmy revealed by ultra-deep sequencing

Cristina Giuliani[1] , Chiara Barbieri[1, 2] , Mingkun Li[2] , Laura Bucci[3,4] , Daniela Monti[5] , Giuseppe Passarino[6] , Donata Luiselli[1] , Claudio Franceschi[3,4,7,8,9], Mark Stoneking[2]* , Paolo Garagnani[3,4,10] *

[1] Department of Biological, Geological and Environmental Sciences, Laboratory of Molecular Anthropology and Centre for Genome Biology, University of Bologna, Bologna, Italy
[2] Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[3] Department of Experimental, Diagnostic and Specialty Medicine (DIMES), University of Bologna, Bologna, Italy
[4] Interdepartmental Centre "L. Galvani" for Integrated Studies of Bioinformatics, Biophysics and Biocomplexity (CIG), University of Bologna, Bologna, Italy;
[5] Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence, Italy
[6] Department of Cell Biology, University of Calabria, Rende, Italy
[7] IRCCS, Institute of Neurological Sciences of Bologna, Bologna, Italy
[8] CNR, Institute of Organic Synthesis and Photoreactivity (ISOF), Bologna, Italy
[9] IGM-CNR Institute of Molecular Genetics, Unit of Bologna IOR, Bologna, Italy
[10] Center for Applied Biomedical Research (CRBA), St. Orsola-Malpighi University Hospital, Bologna, Italy
** Denotes co-senior authorship.*

Key words: *mtDNA heteroplasmy, longevity, aging, transmission, centenarians*
Received: *5/02/14; Accepted: 5/09/14; Published: 5/13/14*
Correspondence to: *Claudio Franceschi, PhD; Mark Stoneking, PhD* ***E-mail:*** *claudio.franceschi@unibo.it; stoneking@eva.mpg.de*

Abstract **The role that mtDNA heteroplasmy plays in healthy aging, familial longevity and the heritability patterns of low levels heteroplasmy in the elderly are largely unknown. We analyzed the low levels of mtDNA heteroplasmy in blood in a cohort of centenarians, their offspring and a group of offspring of non long-lived parents, characterized by a less favorable health phenotype. The aims of this study are to: (i) investigate the transmission of low level heteroplasmies in the elderly; (ii) explore the association of heteroplasmy with age and longevity and (iii) investigate heteroplasmy patterns in these three groups. We sequenced a 853 bp mtDNA fragment in 88 individuals to an average coverage of 49334-fold, using quality control filtering and triplicate PCR analysis to reduce any methodological bias, and we detected 119 heteroplasmic positions with a minor allele frequency ≥ 0.2%. The results indicate that low-level heteroplasmies are transmitted and maintained within families until extreme age. We did not find any heteroplasmic variant associated with longevity and healthy aging but we identified an unique heteroplasmy profile for each family, based on total level and positions. This familial profile suggests that heteroplasmy may contribute to familial longevity.**

INTRODUCTION

Many studies have demonstrated that human healthy ageing and longevity are familial traits inherited across generations [1-4] and that centenarian offspring have higher chances for healthy aging[5, 6]. To assess the factors that can contribute to this phenotype, many studies of long-lived families have been carried out from several perspectives, such as genomics [7],

metabolomics [8], epigenetics [9], gut microbiome [8, 10] and mitochondrial DNA (mtDNA) variability [11]. Among these domains it is well established that the physiological decline of mitochondrial machinery plays a critical role in the aging process, given its role in providing energy to cells and tissues and the high production of reactive oxidative species. One factor that can contribute to age-related mitochondrial dysfunction is the accumulation of mtDNA mutations linked to errors in the replication machinery [12-14]. MtDNA heteroplasmy, *i.e.* intra-individual variation in mtDNA sequences, may be of relevance.

Some studies have addressed the role that mtDNA heteroplasmy may play in promoting longevity [15-18]. A possible beneficial effect on healthy aging is that mtDNA heteroplasmy represents a reservoir of genetic variability, that can bring new functions and increase the capacity of cells to cope with environmental and physiological stressors during life [15, 17, 19, 20] but the forces that drive the fate of heteroplasmic mutations are still a matter of debate [21, 22]. Moreover recent studies demonstrated that the heteroplasmy pattern is inherited [23-26], including low-frequency heteroplasmic variants [27]. This new insight identifies heteroplasmy as a possible genetic contributor to familial longevity, but to date there are no data that show that the inherited heteroplasmic profile is conserved until extreme age. Various forces contribute to heteroplasmy, but the extent to which heteroplasmies are inherited or arise via new mutations is still uncertain, making the study of heteroplasmy in the extreme elderly even more pertinent. The heteroplasmy pattern observed in middle-age individuals is a mixture of both inherited and somatic mtDNA mutations that can potentially change according to tissue composition and cell clonality [28]. It is because of the bottleneck that occurs during postnatal folliculogenesis [29] and the somatic events that naturally occur during life [20, 30] that the frequency of heteroplasmies can change from one generation to another [20].

A further challenge in the study of heteroplasmy and aging is represented by technological limits in heteroplasmy detection. Previous technologies (such as DHPLC or pyrosequencing) did not allow a high site-specific resolution of very low-level heteroplasmic variants [31]. Thus, the advent of next-generation sequencing technologies, with the potential for very high coverage, seems ideal for investigating heteroplasmic events in very old people [27].

In the present study we analyzed 30 female centenarians (CENT), 30 offspring of these centenarians (CO) and a control group of 28 offspring of non-long-lived parents (ONLL), *i.e.* offspring of parents who died before reaching the average life expectancies for their cohort. We detected heteroplasmic positions by an ultra-deep re-sequencing method on the Illumina GA platform, focusing on 853 bp of the mtDNA control region (including HV1 and part of HV2). We focused on this region because heteroplasmies occur preferentially at positions with high mutation rates, in particular in the mtDNA control region, and because this region was already analyzed in previous studies of longevity [17, 19, 32].

The aims of this study are three-fold: (i) to address differences and similarities between centenarians (whose extended lifespan could be the result of a better adaptation to environmental stress and stimuli [33]), their offspring, characterized by a better health phenotype [34] and the offspring of non-long-lived parents; (ii) to describe mtDNA heteroplasmy patterns in extreme longevity; and (iii) to address if low levels of mtDNA heteroplasmy are trasmitted, maintained and detectable in very old mother offspring pairs (centenarians and their offspring) and how this mechanism could contribute to human longevity.

RESULTS

Reliability of mtDNA heteroplasmies detected by ultra-deep resequencing

Many studies raise several doubts related to ultra-deep sequencing based on PCR amplicons [31, 35] such as a biased amplification of nuclear mitochondrial pseudogenes (numts), or a very high number of false positives linked to a high sequence error rate or biased PCR amplification. To reduce the possibility of amplifying numts we first tested the pair of primers by in silico PCR to exclude non-specific amplifications, then the PCR products were checked on agarose gels to exclude the presence of non-specific amplicons. Rigorous criteria were used to distinguish true heteroplasmies from sequencing error, such as double strand validation and filtering (Methods), previously implemented in other studies [36]. Most importantly, to evaluate the potential for PCR errors and to obtain an estimate of the number of false positives, 3 different PCR products from 12 samples were sequenced and the concordance of heteroplasmic sites (identified using 0.2% as the lowest heteroplasmy threshold [24, 27]) between replicates checked.

Thus sensitivity and specificity of heteroplasmy detection were calculated. Sensitivity (*i.e.* a measure of the proportion of heteroplasmic sites correctly identified as such) provides the false negative rate while

specificity (*i.e.*, a measure of the proportion of homoplasmic sites correctly identified as such) provides the false positive rate. Because there is no suitable alternative method to detect very low levels of heteroplasmy (*i.e.* 0.2%), in the calculation of sensitivity and specificity levels a site was defined as "true heteroplasmic" if heteroplasmy was supported by at least 2 of the 3 observations. Mean sensitivity and specificity values were calculated as 0.95 ± 0.09 and 1.00 ± 0.0025 respectively. These high values obtained here support the possibility to detect and quantify heteroplasmic point mutations with high accuracy (for detailed analysis see Table 1S).

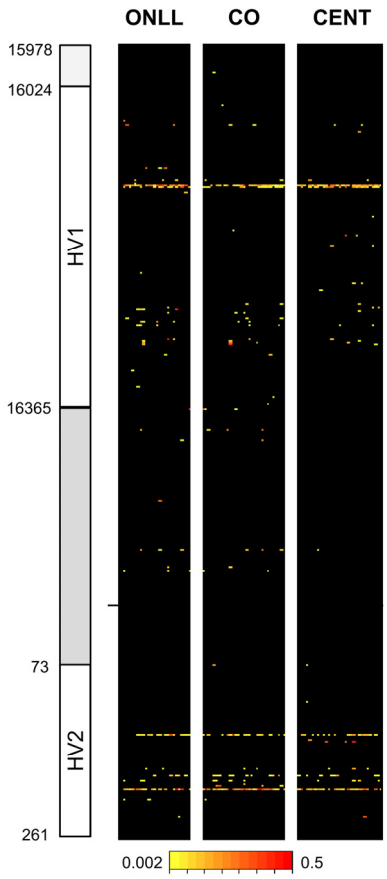


Figure 1. Overview of heteroplasmy pattern in the 853 bp region considered. Heteroplasmic positions are indicated with different colors according to the level of heteroplasmy. Positions are reported on the y-axis while individuals are reported on the x axis.

Heteroplasmy profiling in centenarians, centenarian offspring and offspring of not long-lived parents

After quality control (see Methods) there were results for 30 CENT, 30 CO and 28 ONLL. Heteroplasmy was defined with a minimum threshold of 0.2%, resulting in 119 heteroplasmic positions in the 853 bp fragment analyzed (Figure 1). In particular, we identified 75 heteroplasmic sites in CENT, 101 in CO and 72 in ONLL. We then performed a Fisher-exact test to assess if HV1 (positions: form 15978 to 16364), the intermediate region (positions: from 16365 to 72) and HV2 (positions: from 73 to 261) have similar proportions of heteroplasmic sites. The test showed that HV1 and HV2 are enriched in heteroplasmic sites (OR = 1.25; CI 95% 1.10-1.43 and OR = 1.66; CI 95% 1.43-1.92 respectively) when compared to the intermediate region (OR = 0.24 CI 95% 0.18-0.32). Unsupervised hierarchical clustering was also performed to evaluate the distance between individuals (considering minor allele frequency); CENT, CO and ONLL individuals did not cluster according to group (Figure 1S).

Since heteroplasmies occur preferentially at positions with high mutation rate [31, 32] we explored the correlation between the incidence of heteroplasmy and mutation rates (MR) in our cohort. Mean MR was first calculated considering all positions sequenced in this study (853 bp) and then considering all heteroplasmic sites (Table 1). Similarly we calculated the average MR in HV1, HV2 and the intermediate region considering all sites and then selecting only the heteroplasmic positions for each region.

In order to combine information on both the incidence and the level of heteroplasmy, each position was plotted according to the number of heteroplasmic events and the mean heteroplasmy level in each group (Figure 2). The plot consists of four regions: the upper left quarter shows positions that occur frequently and are characterized by a low level of heteroplasmy; the upper right quarter is empty, indicating that there are no positions characterized by both a high frequency and a high level of heteroplasmy; the lower right quarter includes positions with high heteroplasmy levels but that occur in only a few individuals; and the lower left quarter includes most of the positions, which occur in only a few individuals and with a low heteroplasmy level.

Pearson's correlation between the number of occurrences in CENT, CO and ONLL and the MR of each position was calculated (r CENT = 0.43, p value CENT = $1.012*10^{-6}$; r CO = 0.46, p value CO = $1.51*10^{-7}$; r ONLL = 0.47, p value ONLL = $5.442*10^{-8}$ respectively).

We then compared heteroplasmy levels in CO and ONLL and ANOVA was performed to determine if the level of heteroplasmy at each position differs between

the two groups. Only position 16131 showed a weak significant nominal p values (p value = 0.0062) that is not significant when corrected for multiple test comparisons (Table 2).

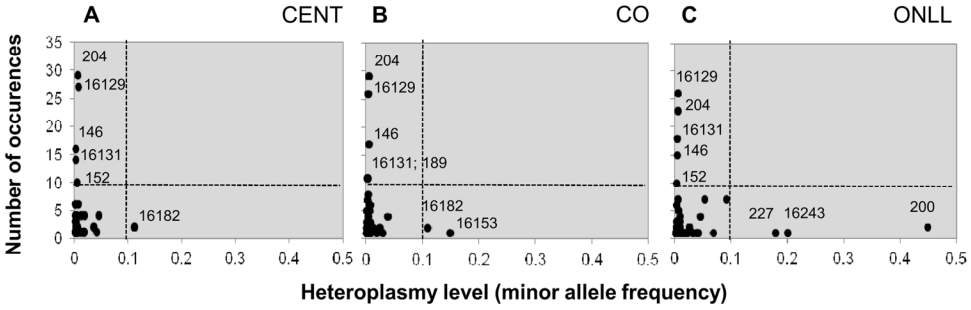


Figure 2. Heteroplasmic positions in the 3 groups. Heteroplasmic positions plotted according to mean heteroplasmy level and number of occurrences in CENT (**A**), CO (**B**) and ONLL (**C**) respectively.

Table 1. Average mutation rate calculated in all sites and only in heteroplasmic sites considering the entire region sequenced, HV1, HV2, and the intermediate region (from position 16365 to position 73).

	Average Mutation Rate	
	All Sites	Heteroplasmic sites
Entire region (853bp)	3.6	15.9 *
HV1	4.7	14.2 *
HV2	6.2	30.9 *
Intermediate region	1.6	6.4
*** p < 0.0001 (Mann Whitney test)**		

Table 2. Significant results of the ANOVA considering the minor allele frequency site by site.

mtDNA position	**p value**	**Significant comparisons**
153	0.0299	CENT-ONLL
204	0.047	CENT-ONLL
16129	0.028	CENT-CO
16131	0.025	CO-ONLL

We next compared each of the offspring groups (CO and ONLL) to centenarians (CENT) to assess whether heteroplasmy levels increase with age. Only positions 153, 204 and 16129 showed weak significant nominal p values (< 0.05 in the ANOVA test) that are not significant when we corrected for multiple test comparisons (Table 2).

To determine if heteroplasmy increases with age we compared the total number of heteroplasmic sites across the three groups by ANOVA; no significant differences among groups were detected (p value = 0.798). We tested for correlations between the age of all individuals (irrespective of group) and the heteroplasmy level for each site. Two positions showed a moderate trend with age: 204 (r = 0.20, p value = 0.06) and 16129 (r = 0.30, p value = 0.0054) (Figure 2S). In addition, the correlation between age and the number of heteroplasmic sites for each individual was not significant (r = 0.17, p value = 0.11).

Heteroplasmy in centenarians and their offspring

Total heteroplasmy analysis. One centenarian and one centenarian's offspring could not be considered in this analysis (for more details see Methods), therefore in the following analyses we considered 29 mother-offspring pairs. Total heteroplasmy was then calculated for each individual (total heteroplasmy = sum of minor allele frequencies at all heteroplasmic sites) to quantify the mutation load of the entire mtDNA genome. These total heteroplasmy values are significantly correlated between centenarians and their offspring (Figure 3A; r = 0.75, p = $2.88*10^{-6}$). To further validate this result, the CENT and CO individuals were reshuffled 2000 times, matching centenarians with unrelated offspring, and the correlation values were calculated. This analysis (Figure 3B) indicates that the observed correlation between CENT and their true offspring in total heteroplasmy is indeed unlikely to arise by chance. In addition, correlation values of 2000 possible combinations of CENT-ONLL and CO-ONLL are centered around zero (Figure 3B), consistent with an absence of any correlation in these comparisons. This analysis suggests that the total heteroplasmy values are maintained in families.

Number of heteroplasmic sites. Centenarians and their offspring were then analyzed to assess shared heteroplasmic positions. The total number of positions shared by CENT and CO over 29 families is 130 (total number of common sites, TNCS); the results for each family are depicted schematically in Figure 4A, and a detailed list is provided in Table 2S. The TNCS was recalculated 2000 times, matching randomly subjects of each the two groups (Figure 4B). This analysis indicates that more TNCS are shared by centenarians and their offspring than between any other combinations of individuals.

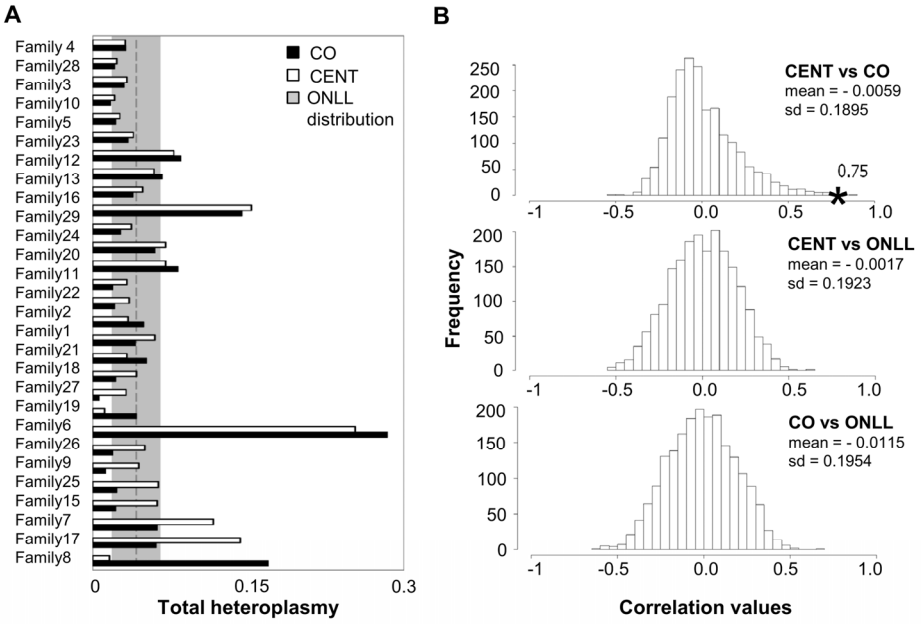


Figure 3. Total heteroplasmy in centenarians and their offspring and reshuffling analysis. (**A**) Total heteroplasmy in centenarians and their offspring. White bars: centenarians; black bars: matched offspring. The dashed line and shading show the mean level and standard deviation of total heteroplasmy in ONLL. (**B**) Frequency distribution of correlation values based on 2000 reshufflings of total heteroplasmy values for CENT and unrelated CO, CENT and ONLL and CO and ONLL. Mean and standard deviation of the distribution is reported. The black star indicates the observed correlation value matching each centenarian with her own offspring.

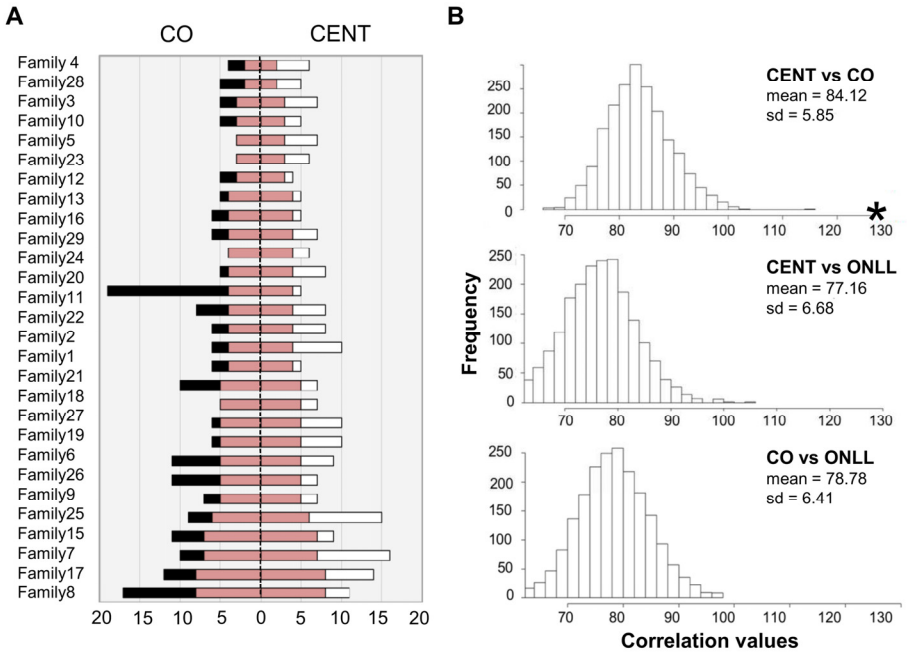


Figure 4. The number of common heteroplasmic sites shared by CENT and CO and frequency distributions of TNCS in all the reshuffled pairs. (**A**) The white bar indicates the number of heteroplasmic position in CENT, the black bar indicates the number of heteroplasmic sites in CO. CENT and CO are here matched according to consanguinity (x-axis). The red bar shows the number of shared heteroplasmic sites in each parent-offspring pair. (**B**) Frequency distributions of TNCS in all the reshuffled pairs considering CENT and unrelated CO, CENT and ONLL, and CO and ONLL. Black star = total number of common heteroplasmic sites matching centenarians with their own offspring.

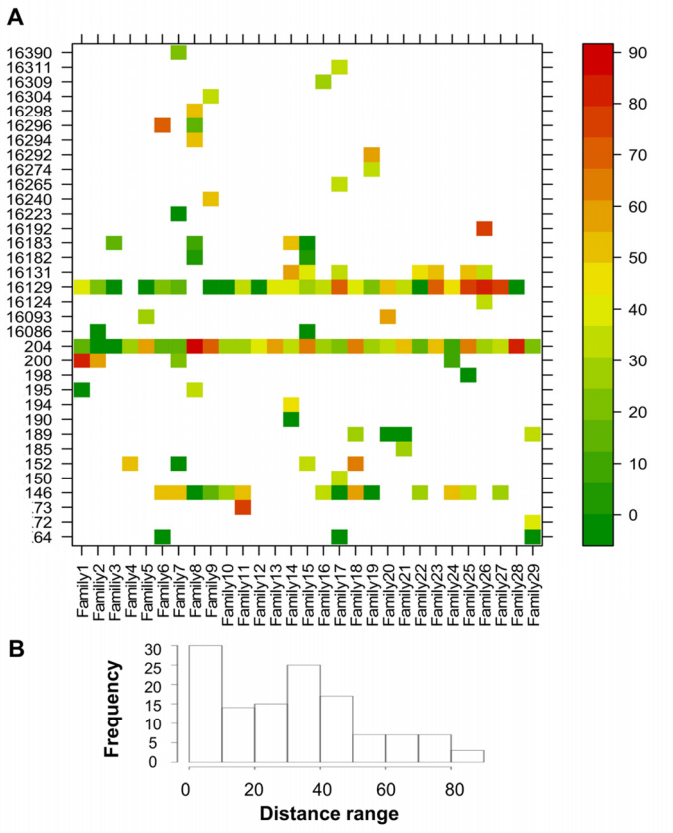


Figure 5. CENT and CO distances of positions that showed heteroplasmy in both individuals. Distance values were calculated as the ratio between the absolute difference between CENT and CO MAF and the higher MAF value of the two (expressed as percentage). (**A**) Heteroplasmy distances plotted for each position and colored according to the scale bar on the right. (**B**) Frequency distribution of the distance values.

Heteroplasmy analysis site by site. To evaluate the similarity in minor allele frequencies (MAF) across sites between each CENT mother and her offspring, a site-by-site distance was calculated (Figure 3S). The ratio between the absolute difference in MAF (abs[CENT-CO]) and the highest MAF of the two was calculated as a measure of the "distance" between each CENT and CO (Figure 5A). Each pair shows *private sites* that are heteroplasmic in CENT and CO with variable values of concordance but different between pairs. Some positions, such as 146, 204 and 16129, are heteroplasmic in most pairs (*public sites*). The frequency distribution of all distance values is reported in Figure 5B; only positions that show heteroplasmy in both individuals were considered in this analysis.

DISCUSSION

To date, several studies have showed that heteroplasmies tend to accumulate during human aging in many tissues, such as brain, heart, skeletal muscles and blood [20, 27, 37-41], but there is still controversy regarding the cause of such age related accumulation of heteroplasmic mutations. The most recent hypothesis suggests that replication errors arise during life and accumulate by clonal expansion and genetic drift, placing the accumulation of ROS damage in the background [14, 42]. This scenario becomes even more intricate when we consider that even low-level heteroplasmies can be inherited [27]. Taking into account the high number of independent mtDNA copies, it is worthwhile to sum up the possible sources of mtDNA variability: (i) heteroplasmy derives from inherited variants and *de novo* mutations [20]; (ii) this variability originates prenatally, from the very first stages of cell divisions and it is modified subsequently during life by segregating events; and (iii) mother and offspring share similars pattern of somatic mtDNA variability.

Given these assumptions we explored whether the inherited heteroplasmy profile is preserved in the elderly or if *de novo* mutations overcome the inherited variants. Moreover we assessed the correlation between heteroplasmy and longevity, and we focused on the identification of specific heteroplasmic profiles for centenarians, their offspring and offspring of non-long lived parents to highlight features linked to familial predisposition to longevity [9, 39, 43, 44]. Centenarian offspring have higher chances for healthy aging [5, 6, 45] even when they share a similar lifestyle with the general population [46]. Moreover clinical [34] and molecular data [8-10] demonstrated that they are characterized by a better health status than offspring of non long lived parents. Applying next-generation sequencing to this cohort, we investigated human mtDNA heteroplasmy transmission and association with longevity. We obtained extremely high coverage (average 49334 fold per position) and replicates analysis demonstrates the possibility to detect very low levels of heteroplasmy (down to 0.2%).

The main results of this paper indicate that low-level heteroplasmies in blood are detectable, transmitted and maintained within families until extreme age. Moreover we did not find any heteroplasmic variant associated with longevity and healthy aging but we identified an unique heteroplasmy profile for each family, resulting from a combination of total heteroplasmy level and specific heteroplasmic positions.

Regarding the technological approach in the first analysis we considered triplicate samples to demonstrate the accuracy and reliability of the method used. We performed double strand validation that considerably reduces the impact of sequencing error [36] and then we used replicates to assess the level of sensitivity and specificity. Sequencing errors, assembly strategies, biased PCR amplification and stochastic effects are unlikely to produce false positives under our criteria for detecting heteroplasmy. The resulting data show that a small fraction of heteroplasmic sites could not be reproduced (possibly because they are false negative), but sites called as *heteroplasmic* are likely to be real mutations. Our results confirmed the possibility to detect with high accuracy a low level of mutation (< 2% of minor allele frequency), as also shown in a recent study [27]. It is worthwhile to note that previous studies, with relatively high coverage, did not show the error rate, sensitivity and specificity and it is likely that these values are directly dependent on the mean coverage. From a technical point of view we can conclude that sequencing depth is very important to identify and to quantify low-level heteroplasmic variants with high accuracy.

We analyzed heteroplasmies focusing on the non-coding region (from position 15978 to position 261). We observed that HV1 and HV2 are enriched in heteroplasmic sites when compared to the intermediate region. We also observed that in our cohort heteroplasmies tend to occur in positions characterized by a high mutation rate. This phenomenon was described previously [31, 32] suggesting that mutation is the major driving force behind heteroplasmies and that subsequent drift to high frequencies in an individual can lead to polymorphisms among individuals [31].

We then analyzed heteroplasmy patterns in our groups (CENT, CO and ONLL). Comparing the minor allele

frequency between groups via ANOVA, the obtained *p values* were not significant, (when corrected for multiple test comparisons), indicating the overall absence of significant heteroplasmy differences between groups. We did not find any differences also in the number of heteroplasmic sites between groups. We also analyzed the heteroplasmic variation across ages and we find only two sites that showed a modest correlation with age, positions 204 and 16129. A previous study that measured heteroplasmies in blood of individuals with different ages (0-60 yrs) found several heteroplasmic sites whose frequency correlates with age [20]. However even if we did not observe a clear pattern of age-related heteroplasmy accumulation, we cannot exclude that heteroplasmies accumulate during life because of the limited age range here considered (average age = 80.7 ± 15.1).

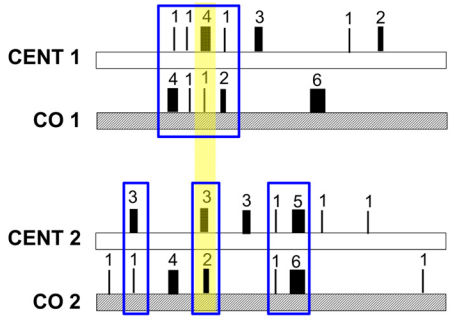


Figure 6. Model of the mechanism of heteroplasmic patterns in blood of long-lived families. The number and thickness of each rectangle represents the different heteroplasmic level of each site, while the sum of all numbers represents the total heteroplasmy value. Total heteroplasmy is more similar within parents-offspring pairs than between pairs (CENT1 = 13 and CO1 = 14 while CENT2 = 17 and CO2 = 16). Each family shares a common core of heteroplasmic sites (here marked in blue). This core is composed of two types of heteroplasmic sites (i) *public heteroplasmic positions* that all pairs share (here marked with a yellow shadow) and (ii) *private heteroplasmic sites* that characterize each family.

Regarding the heteroplasmy pattern in families (mother-offspring pairs), we found a clear signature of heteroplasmy inheritance in consideration of the centenarians-offspring concordance of both total heteroplasmies levels and of the mutated loci (heteroplasmy profile). This concordance was tested and confirmed by reshuffling analyses. Our data suggest that families share a *common core* of heteroplasmic sites that could be divided, according to their frequency in the cohort, into (i) *public heteroplasmic positions* that all parent-offspring couples shared (such as positions 146, 204 and 16129) and (ii) *private heteroplasmic sites* that are common to only one or few families. Figure 6 simplifies the heteroplasmic profiles we observed. We note that public heteroplasmic positions may reflect recurrent mutations as well as inherited hetreroplasmies [47]. These data are consistent with previous studies [17, 19] that found that a high concordance of heteroplasmy can be observed within families whose ages are shifted toward the last decades of life. However, the use of DHPLC (Denaturating High Performance Liquid Chromatography) in these previous studies did not allow the precise identification of heteroplasmic positions, as we did in the present study. Taken together the groups comparison and the familial analysis suggested that the heteroplasmy profile is a partially inherited individual characteristic.

In conclusion many findings showed that exceptional longevity runs in families [1-4], but to date genetic studies on longevity and healthy aging have shown that centenarians carry the same numbers of GWAS-identified common disease variants of ordinary people. In this context rare and private variants seem to have a relevant role in buffering the presence of other variants [48]. Thus we reasoned that the long-term heteroplasmy inheritance, that we demonstrate in the present study, is a potential source of rare variants that could play a role in determining healthy aging and longevity. The issue is to demonstrate the effect of these patterns on longevity, but this limitation is also valid for all the rare variants that can be described in whole genome.

Other studies had already shown that the description of heteroplasmy is complicated by the characteristics of the blood itself such as the origin of the cells, the rapid turnover and the high number of different cell types [27]. We can speculate that at the somatic level rare heteroplasmic mutations can expand or be lost, through segregation events. Thus in a given anatomical region such as muscle or brain, heteroplasmies expansion occurs and remains localized in the same physical area [37]. Similarly this may occur during hematopoietic cell expansion in the bone marrow but the local heteroplasmy profiles are diluted when the new cells are released in the bloodstream. This implies that the heteroplasmic variants that we describe at low frequencies in blood are compatible with the high heteroplasmy level detected in other tissues (such as skeletal muscle and brain). It is likely that low level heteroplasmies have the potential to increase in frequency in critical tissues. Further studies considering different tissues of the same individual are needed to experimentally test this last consideration and to differentiate between acquired or inherited heteroplasmic mutations [24]. Indeed the exact

mechanism of inheritance as well as the origin of these heteroplasmic events are not completely understood, and the forces that shape heteroplasmies are under study [29, 49, 50]. Evidence from mice and humans suggests that nonsynonymous protein-coding heteroplasmies are subject to negative selection during growth and aging [31, 50]. A further critical issue concerns the minimum percentage of detectable mutations compatible with the size of the bottleneck that can be observed in blood and the exact timing of these mutations. Many studies have attempted to estimate the size of the bottleneck [29, 51], but more data are necessary to answer, at least on a theoretical level, this question.

Finally these findings add significantly to our knowledge of mechanisms of heteroplasmy transmission and association with longevity, and indicate that heteroplasmic events are an interesting candidate for further genetic studies on familial longevity. Our results are compatible with recent evidence that shows that the transmission of low levels of mtDNA heteroplasmy seems to influence offspring aging [37, 52], and by focussing on long-lived individuals and their offspring, adds a new perspective on the contribution of low level heteroplasmies to the genetics of longevity.

METHODS

Sample collection. A total of 90 Italians were sequenced, including 31 female centenarians (average age 101.6 ± 2.8), 31 female offspring of these centenarians (average age 69.9 ± 6.6) and 28 female offspring whose parents both died before the average life expectancy (average age 70.7 ± 4.8). The samples here analyzed were selected according to a stringent demographic strategy: centenarians and offspring were recruited with the constraint that also the other parent was long lived (born in the same birth cohort 1900-1908 and died at an age > 77 years). Offspring of non-long lived parents were matched with centenarian offspring, with both parents born between 1900 and 1908 but dying before the average life expectancy calculated at 15 years of age (67 years if male and 72 years if female) by the Italian mortality tables (see http://www.mortality.org/ - Max Planck Institute for Demography, Rostock, Germany). Both groups were comparable with respect to age but very different if we considered the probability of becoming long lived [34]. Indeed recent studies suggest that the offspring of centenarians have a significant survival advantage, a higher probability to become long-lived and a lower risk for age-related diseases (such as cardiovascular diseases) compared to control subjects [4, 5, 45, 53]. Individuals with cancer at the time of the interview or treated with immunosuppressive or anticoagulant therapies were excluded from this study.

Genomic DNA was extracted from whole blood using the Wizard genomic DNA purification kit (PROMEGA, Madison WI, USA) according to the manufacturer's standard protocols.

Ethics statement. The study protocol was approved by the Ethical Committee of the Sant'Orsola - Malpighi University Hospital (Bologna, Italy). After obtaining written informed consent, a standard structured questionnaire was administered to collect information regarding health status. Past and current diseases were checked against the medical documentation.

DNA amplification and sequencing. DNA was first quantified by NanoDrop (NanoDrop Products Thermo Scientific Wilmington, DE) and an 853 bp segment of the mtDNA control region (from position 15978 to 261, including HV1 and part of HV2) was amplified. We focus on this region because heteroplasmies occur preferentially at positions with high mutation rates, like the mtDNA control region, and also because this region was already analyzed in previous studies on longevity. Primers were selected after in silico PCR to avoid amplification of unspecific nuclear DNA (L15997, CACCATTAGCACCCAAAGCT and H242, GCTGTGCAGACATTCAATTG). Amplified fragments were then sheared using the Bioruptor UCD-200 (Diagenode) to approximately 200 – 800 bps. Libraries were prepared following a previously published protocol [54]. Two different indexes were attached to individual samples by performing a PCR amplification using the Herculase Mastermix (NEB). Samples were pooled in equimolar ratios after indexing and subsequently sequenced in one lane of the Illumina GAIIx (Solexa) sequencer, with a paired-end run of 76+7 cycles, which yielded ~45 million paired-end reads.

Assembly and data analysis. Base quality scores were recalibrated by the IBIS software using spiked-in PhiX sequencing data as the training dataset [55]. The adaptor sequence was trimmed and paired-end reads were merged if they were completely overlapped. Reads were mapped to the revised Cambridge reference sequence (GenBank ID:NC_012920, the starting point was changed to 15978 on the original sequence) using BWA [56]; reads with mapping quality score lower than 20 and bases with quality score lower than 20 were excluded from the analysis, resulting in a final average coverage of 49334-fold per position in the target region. Samples that showed low coverage (mean coverage <10000) were excluded from the analysis of

heteroplasmy. Mean coverage distribution is reported in Figure 4S. Initially 31 CENT, 31 CO and 28 ONLL were sequenced but 1 CENT and 1 unrelated CO did not pass the quality control (Figure 4S) and consequently for the centenarians-offspring paired analysis there were 29 CENT-CO pairs. Analyses were performed using R version 3.0.1. We called heteroplasmic sites according to minor allele frequency even if some heteroplasmic positions were then filtered out after double strand validation, as described previously [36].

The heteroplasmy detection threshold was set to 0.2%; the sensitivity and specificity of detected heteroplasmies were determined from 12 samples (one sample was discarded after quality filters) whose levels of heteroplasmy were measured using 3 different PCR products. The concordance was used as a parameter of reliability in the definition of heteroplasmic sites. For this analysis we create a confusion matrix using the *caret packages* and in this calculation a site was defined as "true heteroplasmic" if heteroplasmy was supported by at least 2 of the 3 observations [57-59]. The overview of heteroplasmy pattern (Figure 1) was perfomed using the *lattice* graphics package [60]. The mutation rate was estimated for each site from Phylotree [61]. Unsupervised hierarchical clustering was performed on the basis of the *complete linkage* method, using the function heatmap.2 in the *gplots package*.

ACKNOWLEDGEMENTS

We thank dr Maria Giulia Bacalini and dr Chiara Pirazzini for insightful comments on this manuscript. We thank and mention the Italian Ministry of University and Research (Project PRIN 2006 to DM and Project PRIN 2009 to CF). ML and MS were supported by the Max Planck Society. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived and designed the experiments and data analysis: PG, MS, CF, CG, DL. Performed the experiment: CG, CB. Analyzed the data: CG, ML, CB. Wrote the paper: CG, PG, MS, CF, GP, ML. Recruited patient samples, clinical evaluation of participating subjects; DM, LB. Provided critical edits and discussion of manuscript: CG, CB, ML, LB, DM, GP, DL, CF, MS, PG.

Conflict of interest statement

The authors of this manuscript have no conflict of interest to declare.

REFERENCES

1. Atzmon G, Rincon M, Rabizadeh P and Barzilai N. Biological evidence for inheritance of exceptional longevity. Mech Ageing Dev. 2005; 126:341-345.

2. De Benedictis G and Franceschi C. The unusual genetics of human longevity. Sci Aging Knowledge Environ. 2006; 2006:pe20.

3. Franceschi C, Bezrukov V, Blanché H, Bolund L, Christensen K, de Benedictis G, Deiana L, Gonos E, Hervonen A, Yang H, Jeune B, Kirkwood TBL, Kristensen P, et al. Genetics of healthy aging in Europe: the EU-integrated project GEHA (GEnetics of Healthy Aging). Ann N Y Acad Sci. 2007; 1100:21-45.

4. Perls TT, Bubrick E, Wager CG, Vijg J and Kruglyak L. Siblings of centenarians live longer. Lancet. 1998; 351:9115.

5. Adams ER, Nolan VG, Andersen SL, Perls TT and Terry DF. Centenarian offspring: start healthier and stay healthier. J Am Geriatr Soc. 2008; 56:2089-2092.

6. Balistreri CR, Accardi G, Buffa S, Bulati M, Martorana A, Candore G, Colonna-Romano G, Lio D and Caruso C. Centenarian Offspring: a model for Understanding Longevity. Curr Vasc Pharmacol. 2013. Epub ahead of print

7. Beekman M, Blanché H, Perola M, Hervonen A, Bezrukov V, Sikora E, Flachsbart F, Christiansen L, De Craen AJM, Kirkwood TBL, Rea IM, Poulain M, Robine J-M, et al. Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. Aging cell. 2013; 12:184-193.

8. Collino S, Montoliu I, Martin F-PJ, Scherer M, Mari D, Salvioli S, Bucci L, Ostan R, Monti D, Biagi E, Brigidi P, Franceschi C and Rezzi S. Metabolic signatures of extreme longevity in northern Italian centenarians reveal a complex remodeling of lipids, amino acids, and gut microbiota metabolism. PLoS ONE. 2013; 8.

9. Gentilini D, Mari D, Castaldi D, Remondini D, Ogliari G, Ostan R, Bucci L, Sirchia SM, Tabano S, Cavagnini F, Monti D, Franceschi C, Di Blasio AM, et al. Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring. Age (Dordr). 2013; 35:1961-1973.

10. Rampelli S, Candela M, Turroni S, Biagi E, Collino S, Franceschi C, O'Toole PW and Brigidi P. Functional metagenomic profiling of intestinal microbiome in extreme ageing. Aging (Albany NY). 2013:902-912.

11. Raule N, Sevini F, Li S, Barbieri A, Tallaro F, Lomartire L, Vianello D, Montesanto A, Moilanen JS, Bezrukov V, Blanché H, Hervonen A, Christensen K, et al. The co-occurrence of mtDNA mutations on different oxidative phosphorylation subunits, not detected by haplogroup analysis, affects human longevity and is population specific. Aging cell. 2013.

12. Kirkwood TBL. Understanding the odd science of aging. Cell. 2005; 120:437-447.

13. Trifunovic A, Wredenberg A, Falkenberg M, Spelbrink JN, Rovio AT, Bruder CE, Bohlooly-Y M, Gidlöf S, Oldfors A, Wibom R, Törnell J, Jacobs HT and Larsson N-G. Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature. 2004; 429:417-423.

14. Kennedy SR, Salk JJ, Schmitt MW and Loeb LA. Ultra-Sensitive Sequencing Reveals an Age-Related Increase in Somatic Mitochondrial Mutations That Are Inconsistent with Oxidative Damage. PLoS Genet. 2013; 9: e1003794.

15. Zhang J, Asin-Cayuela J, Fish J, Michikawa Y, Bonafe M, Olivieri F, Passarino G, De Benedictis G, Franceschi C and Attardi

G. Strikingly higher frequency in centenarians and twins of mtDNA mutation causing remodeling of replication origin in leukocytes. Proc Natl Acad Sci USA. 2003; 100:1116-1121.

16. Fish J, Raule N and Attardi G. Discovery of a major D-loop replication origin reveals two modes of human mtDNA synthesis. Science. 2004; 306:2098-2101.

17. Rose G, Passarino G, Scornaienchi V, Romeo G, Dato S, Bellizzi D, Mari V, Feraco E, Maletta R, Bruni A, Franceschi C and De Benedictis G. The mitochondrial DNA control region shows genetically correlated levels of heteroplasmy in leukocytes of centenarians and their offspring. BMC Genomics. 2007; 8: 293.

18. Michikawa Y, Mazzucchelli F, Bresolin N, Scarlato G and Attardi G. Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. Science. 1999; 286:774-779.

19. Rose G, Romeo G, Dato S, Crocco P, Bruni AC, Hervonen A, Majamaa K, Sevini F, Franceschi C, Passarino G and Consortium GEoHAP. Somatic point mutations in mtDNA control region are influenced by genetic background and associated with healthy aging: a GEHA study. PLoS ONE. 2010; 5: e13395

20. Sondheimer N, Glatz CE, Tirone JE, Deardorff MA, Krieger AM and Hakonarson H. Neutral mitochondrial heteroplasmy and the influence of aging. Hum Mol Genet. 2011; 20:1653-1659.

21. Kowald A and Kirkwood TBL. Mitochondrial mutations and aging: random drift is insufficient to explain the accumulation of mitochondrial deletion mutants in short-lived animals. Aging cell. 2013; 12:728-731.

22. Khrapko K and Vijg J. Mitochondrial DNA mutations and aging: devils in the details? Trends in Genetics. 2009; 25:91-98.

23. White DJ, Wolff JN, Pierson M and Gemmell NJ. Revealing the hidden complexities of mtDNA inheritance. Molecular Ecology. 2008; 17:4925-4942.

24. He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA, Jr., Kinzler KW, Vogelstein B and Papadopoulos N. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature. 2010; 464:610-614.

25. Ameur A, Stewart JB, Freyer C, Hagström E, Ingman M, Larsson N-G and Gyllensten U. Ultra-deep sequencing of mouse mitochondrial DNA: mutational patterns and their origins. PLoS Genet. 2011; 7: e1002028.

26. Avital G, Buchshtav M, Zhidkov I, Tuval Feder J, Dadon S, Rubin E, Glass D, Spector TD and Mishmar D. Mitochondrial DNA heteroplasmy in diabetes and normal adults: role of acquired and inherited mutational patterns in twins. Hum Mol Genet. 2012; 21:4214-4224.

27. Payne BAI, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, Taylor RW, Samuels DC, Santibanez-Koref M and Chinnery PF. Universal heteroplasmy of human mitochondrial DNA. Hum Mol Genet. 2013; 22:384-390.

28. Wallace DC. Bioenergetics, the origins of complexity, and the ascent of man. Proceedings of the National Academy of Sciences. 2010; 107:8947-8953.

29. Wai T, Teoli D and Shoubridge EA. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. Nat Genet. 2008; 40:1484-1488.

30. Wallace DC. Mitochondrial DNA mutations in disease and aging. Environmental and Molecular Mutagenesis. 2010; 5:440-450.

31. Li M, Schönberg A, Schaefer M, Schroeder R, Nasidze I and Stoneking M. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. Am J Hum Genet. 2010; 87:237-249.

32. Stoneking M. Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet. 2000; 67:1029-1032.

33. Franceschi C, Capri M, Monti D, Giunta S, Olivieri F, Sevini F, Panourgia MP, Invidia L, Celani L, Scurti M, Cevenini E, Castellani GC and Salvioli S. Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. Mech Ageing Dev. 2007; 128:92-105.

34. Gueresi P, Miglio R, Monti D, Mari D, Sansoni P, Caruso C, Bonafede E, Bucci L, Cevenini E, Ostan R, Palmas MG, Pini E, Scurti M, et al. Does the longevity of one or both parents influence the health status of their offspring? Exp Gerontol. 2013; 48:395-400.

35. Wilber Quispe-Tintaya RRW. Fast mitochondrial DNA isolation from mammalian cells for next-generation sequencing. BioTechniques. 2013; 55:133-136.

36. Li M and Stoneking M. A new approach for detecting low-level mutations in next-generation sequence data. Genome Biology. 2012; 13:R34.

37. Ross JM, Stewart JB, Hagström E, Brené S, Mourier A, Coppotelli G, Freyer C, Lagouge M, Hoffer BJ, Olson L and Larsson N-G. Germline mitochondrial DNA mutations aggravate ageing and can impair brain development. Nature. 2013; 501:412-415.

38. Cooper JM, Mann VM and Schapira AHV. Analyses of mitochondrial respiratory chain function and mitochondrial DNA deletion in human skeletal muscle: Effect of ageing. Journal of the Neurological Sciences. 1992; 113:91-98.

39. Salvioli S, Bonafè M, Capri M, Monti D and Franceschi C. Mitochondria, aging and longevity--a new perspective. FEBS Lett. 2001; 492:9-13.

40. Short KR, Bigelow ML, Kahl J, Singh R, Coenen-Schimke J, Raghavakaimal S and Nair KS. Decline in skeletal muscle mitochondrial function with aging in humans. Proc Natl Acad Sci USA. 2005; 102:5618-5623.

41. Wei YH, Ma YS, Lee HC, Lee CF and Lu CY. Mitochondrial theory of aging matures--roles of mtDNA mutation and oxidative stress in human aging. Zhonghua Yi Xue Za Zhi (Taipei). 2001; 64:259-270.

42. Elson JL, Samuels DC, Turnbull DM and Chinnery PF. Random Intracellular Drift Explains the Clonal Expansion of Mitochondrial DNA Mutations with Age. Am J Hum Genet. 2001; 68:802-806.

43. Salvioli S, Capri M, Santoro A, Raule N, Sevini F, Lukas S, Lanzarini C, Monti D, Passarino G, Rose G, De Benedictis G and Franceschi C. The impact of mitochondrial DNA on human lifespan: a view from studies on centenarians. Biotechnol J. 2008; 3:740-749.

44. Atzmon G, Rincon M, Schechter CB, Shuldiner AR, Lipton RB, Bergman A and Barzilai N. Lipoprotein genotype and conserved pathway for exceptional longevity in humans. PLoS Biol. 2006; 4:e113.

45. Perls TT, Wilmoth J, Levenson R, Drinkwater M, Cohen M, Bogan H, Joyce E, Brewster S, Kunkel L and Puca A. Life-long sustained mortality advantage of siblings of centenarians. Proc Natl Acad Sci USA. 2002; 99:8442-8447.

46. Rajpathak SN, Liu Y, Ben-David O, Reddy S, Atzmon G, Crandall J and Barzilai N. Lifestyle factors of people with exceptional longevity. J Am Geriatr Soc. 2011; 59:1509-1512.

47. Samuels DC, Li C, Li B, Song Z, Torstenson E, Boyd Clay H, Rokas A, Thornton-Wells TA, Moore JH, Hughes TM, Hoffman RD, Haines JL, Murdock DG, et al. Recurrent Tissue-Specific mtDNA Mutations Are Common in Humans. PLoS Genet. 2013; 9:e1003929.

48. Brooks-Wilson AR. Genetics of healthy aging and longevity. Hum Genet. 2013; 132:1323-1338.

49. Stewart JB, Freyer C, Elson JL, Wredenberg A, Cansu Z, Trifunovic A and Larsson N-G. Strong purifying selection in transmission of mammalian mitochondrial DNA. PLoS Biol. 2008; 6:e10.

50. Freyer C, Cree LM, Mourier A, Stewart JB, Koolmeister C, Milenkovic D, Wai T, Floros VI, Hagström E, Chatzidaki EE, Wiesner RJ, Samuels DC, Larsson N-G, et al. Variation in germline mtDNA heteroplasmy is determined prenatally but modified during subsequent transmission. Nat Genet. 2012; 44:1282-1285.

51. Marchington DR, Macaulay V, Hartshorne GM, Barlow D and Poulton J. Evidence from human oocytes for a genetic bottleneck in an mtDNA disease. Am J Hum Genet. 1998; 63:769-775.

52. Keogh M and Chinnery PF. Hereditary mtDNA Heteroplasmy: A Baseline for Aging? Cell Metabolism. 2013; 18:463-464.

53. Terry DF, Wilcox MA, McCormick MA, Pennington JY, Schoenhofen EA, Andersen SL and Perls TT. Lower all-cause, cardiovascular, and cancer mortality in centenarians' offspring. J Am Geriatr Soc. 2004; 52:2074-2076.

54. Meyer M and Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010; 2010.

55. Kircher M, Stenzel U and Kelso J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. Genome biology. 2009; 10:R83.

56. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754-1760.

57. Altman DG and Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ. 1994; 308:6943.

58. Altman DG and Bland JM. Diagnostic tests 2: Predictive values. BMJ. 1994; 309:6947.

59. Kuhn. Building Predictive Models in R Using the caret Package. Journal of Statistical Software. 2008; Vol. 28(Issue 5).

60. Sarkar D. 2008. Lattice: Multivariate Data Visualization with R: Springer.

61. van Oven M and Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat. 2009; 30:E386-394.

SUPPORTING INFORMATION

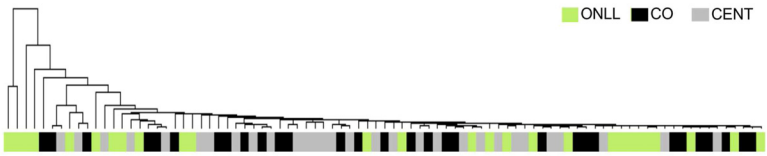


Figure 1S. Unsupervised hierarchical clustering using heteroplasmy levels. CENT: gray bar; CO: black bar; ONLL: green bar.

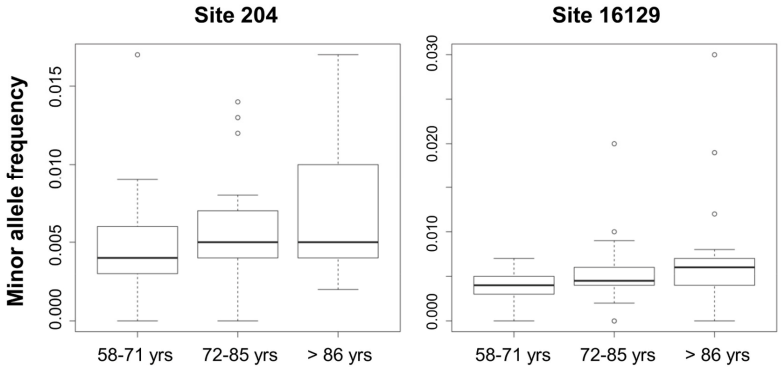


Figure 2S. Boxplots of minor allele frequencies across 3 different age classes. Age < 72 yrs; N=31, age range: 72-86 yrs; N=26 and age > 86 yrs; N=29 respectively for two heteroplasmic sites

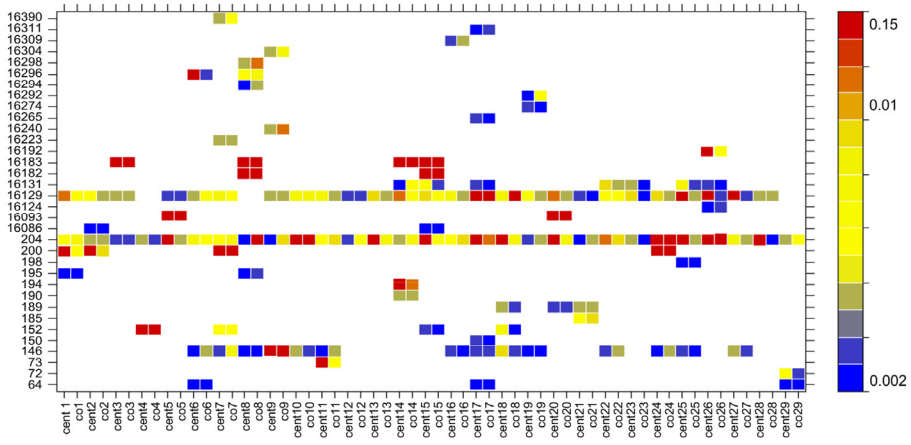


Figure 3S. Heatplot of minor allele frequencies, colored according to the scale bar, for each mother-offspring pair. Rows are positions and columns are individuals.

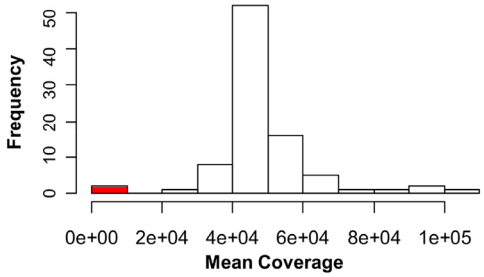


Figure 4S. Mean coverage distribution. Two samples with low coverage (indicated in red) were excluded from the analyses.

Table 1S. Number of occurences observed for each possible outcome in the replicates. TP: true positives, FP: false positives, FN: false negatives, TN: true negatives.

Replicate concordance	Events	Number of occurences
3X heteroplasmy	TP	381
2x heteroplasmy	TP	40
1x heteroplasmy	FP	59
1x homoplasmy	FN	20
2x homoplasmy	TN	118
3x homoplasmy	TN	30090

Table 2S. List of common heteroplasmic positions in each CENT-CO pair. Private heteroplasmic sites (*i.e.*, not shared by other families) are marked in gray.

ID_family	Heteroplasmic sites in CENT and their CO							
Family1	195	200	204	16129				
Family2	200	204	16086	16129				
Family3	204	16129	16183					
Family4	152	204						
Family5	204	16093	16129					
Family6	64	146	204	16129	16296			
Family7	146	152	200	204	16129	16223	16390	
Family8	146	195	204	16182	16183	16294	16296	16298
Family9	146	204	16129	16240	16304			
Family10	146	204	16129					
Family11	73	146	204	16129				
Family12	204	16129	16131					
Family13	146	204	16129	16274				
Family14	190	194	204	16129	16131	16183		
Family15	152	204	16086	16129	16131	16182	16183	
Family16	146	204	16129	16309				
Family17	64	146	150	204	16129	16131	16265	16311
Family18	146	152	189	204	16129			
Family19	146	204	16129	16274	16292			
Family20	189	204	16093	16129				
Family21	185	189	204	16129				
Family22	146	204	16129	16131				
Family23	204	16129	16131					
Family24	146	200	204	16129				
Family25	146	198	204	16129	16131			
Family26	204	16124	16129	16131	16192			
Family27	204	16124	16129	16131	16192			
Family28	204	16129						
Family29	64	72	189	204				

Experimental Gerontology 56 (2014) 234–244

Contents lists available at ScienceDirect

Experimental Gerontology

journal homepage: www.elsevier.com/locate/expgero

Review

mtDNA mutations in human aging and longevity: Controversies and new perspectives opened by high-throughput technologies

Federica Sevini [a,b,*], Cristina Giuliani [c,d], Dario Vianello [a], Enrico Giampieri [e], Aurelia Santoro [a], Fiammetta Biondi [b], Paolo Garagnani [a,b], Giuseppe Passarino [f], Donata Luiselli [c,d], Miriam Capri [a,b], Claudio Franceschi [a,b,g,h], Stefano Salvioli [a,b]

[a] Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, via S. Giacomo 12, 40126 Bologna, Italy
[b] C.I.G. Interdepartmental Centre L. Galvani for Integrated Studies on Bioinformatics, Biophysics and Biocomplexity, University of Bologna, via S. Giacomo 12, 40126, Bologna, Italy
[c] Department of Biological, Geological and Environmental Sciences, Laboratory of Anthropology, University of Bologna, Via Selmi 3, 40126 Bologna, Italy
[d] Department of Biological, Geological and Environmental Sciences, Centre for Genome Biology, University of Bologna, Via Selmi 3, 40126 Bologna, Italy
[e] Department of Physics and Astronomy, Viale Berti Pichat 6/2, 40126 Bologna, Italy
[f] Department of Biology, Ecology and Earth Science, University of Calabria, 87036 Rende, Italy
[g] IRCCS, Institute of Neurological Sciences of Bologna, Ospedale Bellaria, Via Altura 3, 40139 Bologna, Italy
[h] CNR, Institute of Organic Synthesis and Photoreactivity (ISOF), Via P. Gobetti 101, 40129 Bologna, Italy

ARTICLE INFO

Article history:
Received 23 December 2013
Received in revised form 14 March 2014
Accepted 26 March 2014
Available online 5 April 2014

Section Editor: Natascia Ventura

Keywords:
Mitochondrial DNA
Longevity
mtDNA mutations
Heteroplasmy
Next generation sequencing
Epistasis

ABSTRACT

The last 30 years of research greatly contributed to shed light on the role of mitochondrial DNA (mtDNA) variability in aging, although contrasting results have been reported, mainly due to bias regarding the population size and stratification, and to the use of analysis methods (haplogroup classification) that resulted to be not sufficiently adequate to grasp the complexity of the phenomenon.
A 5-years European study (the GEHA EU project) collected and analyzed data on mtDNA variability on an unprecedented number of long-living subjects (enriched for longevity genes) and a comparable number of controls (matched for gender and ethnicity) in Europe. This very large study allowed a reappraisal of the role of both the inherited and the somatic mtDNA variability in aging, as an association with longevity emerged only when mtDNA variants in OXPHOS complexes co-occurred. Moreover, the availability of data from both nuclear and mitochondrial genomes on a large number of subjects paves the way for an evaluation at a very large scale of the epistatic interactions at a higher level of complexity.
This scenario is expected to be even more clarified in the next future with the use of next generation sequencing (NGS) techniques, which are becoming applicable to evaluate mtDNA variability and, then, new mathematical/bioinformatic analysis methods are urgently needed. Recent advances of association studies on age-related diseases and mtDNA variability will also be discussed in this review, taking into account the bias hidden by population stratification. Finally, very recent findings in terms of mtDNA heteroplasmy (i.e. the coexistence of wild type and mutated copies of mtDNA) and aging as well as mitochondrial epigenetic mechanisms will also be discussed.

1. Introduction

In the past 30 years, a number of evidences have delineated the central role of mitochondria in aging (Capri et al., 2013; Salvioli et al., 2008; Wallace, 2005) and today GWAS studies on age-related diseases, such as type 2 diabetes and obesity, have identified some nuclear loci implicated in mitochondrial bioenergetics (PPARγ, PGC-1α, UCPs), providing further support to this idea (Jia et al., 2010; Manning et al., 2012; Morris et al., 2012; Wallace, 2013).

Mitochondria are involved in energy production (oxidative phosphorylation, OXPHOS) and in various cellular processes such as heat production, lipid metabolism, apoptosis regulation, and cellular differentiation. They are also responsible for the production and regulation of the most important byproduct of the cellular metabolism, reactive oxygen species (ROS), which have been posed as the pillar of the free radical theory of aging (Harman, 1957; for a review see Passarino et al., 2010; Wallace, 2010). A critical analysis of the history of this theory is not the purpose of this review and it has been recently well described elsewhere (Lagouge and Larsson, 2013). However the latest theories and hypotheses state that ROS generation is not a cause of aging, but rather a consequence of age-dependent damage. In fact, a series of experimental results suggests that ROS could be more likely

Abbreviations: GWAS, genome wide association study; OXPHOS, oxidative phosphorylation; ROS, reactive oxygen species; mtDNA, mitochondrial DNA; polγ, polymerase gamma; GEHA, Genetic of Healthy Aging; NGS, next generation sequencing; nDNA, nuclear DNA; 90+, nonagenarians.

* Corresponding author at: Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, via S. Giacomo 12, 40126 Bologna, Italy. Tel.: +39 0512094775; fax: +39 0512094747.
E-mail address: federica.sevini@unibo.it (F. Sevini).

http://dx.doi.org/10.1016/j.exger.2014.03.022
0531-5565/

"mediators" of the stress response to age-dependent damage acting as signaling molecules, by an adaptive response commonly defined "mitohormesis" (Hekimi et al., 2011; Lagouge and Larsson, 2013; Ristow and Zarse, 2010; Tapia, 2006). Although there is an open debate on this issue, it is well known that mtDNA mutations accumulate with age and recent findings show that this accumulation could be due to errors in replication machinery or to unrepaired damages, placing the accumulation of damage by ROS in the background. Two studies support this hypothesis: (1) a study by Trifunovic et al. (2004) showed that defects in the proofreading function of mtDNA polymerase (polγ) lead to a high accumulation of mtDNA mutations; (2) more recently, a study by Kennedy et al. (2013) showed that in human brain the mtDNA mutations during aging are comprised predominantly of transition mutations, consistent with misincorporation by polγ or deamination of cytidine and adenosine as the primary mutagenic events in mtDNA.

In this review, we will focus on the results obtained by genotyping the mtDNA in an unprecedented number of samples of long living (>90 years) sibpairs, and a comparable number of controls, recruited in the frame of the GEHA (Genetics of Healthy Aging) European Project (Skytthe et al., 2011). This very large study contributed to a reappraisal of the role of both the inherited (Raule et al., 2013) and the somatic mtDNA variability (Rose et al., 2010) in aging, and paves the way for an evaluation at a very large scale of the epistatic interactions between different mtDNA loci as well as between mitochondrial and nuclear genomes. A focus on next generation sequencing (NGS) techniques applied to mitochondrial variability evaluation and, then, on the need for new mathematical/bioinformatic methods will also be discussed. Recent advances of association studies on age-related diseases and mtDNA variability will also be discussed, taking into account the bias hidden by population stratification. Finally, very recent findings in terms of mtDNA heteroplasmy (i.e. the coexistence of wild type and mutated copies of mtDNA) and aging as well as mitochondrial epigenetic mechanisms will also be reviewed.

2. Inherited and somatic mtDNA mutations in aging: the GEHA study

It is worthwhile to note that various components contribute to mtDNA variability in aging and longevity. In the first section of the review we will try to elucidate (1) how certain variants present at high frequency or low frequency in individuals are inherited in human populations and how these are associated positively or negatively with aging (2) the link between aging and somatic variants and (3) the role played by combinations of these variants (somatic and inherited) in healthy aging.

2.1. Inherited mtDNA variability

Most of the about 1500 proteins shaping the biology of mitochondria are encoded by the nuclear genome (nDNA), but 13 subunits of the OXPHOS are encoded by the mtDNA. Thus, many studies conducted in model organisms and in humans have addressed the effects of nuclear and mtDNA variability (and of the interaction between the two genomes) on mitochondrial functioning and aging (Greaves and Turnbull, 2009; for a review see Lagouge and Larsson, 2013). In particular in humans, the role played by the mtDNA common variability in modulating the processes influenced by mitochondrial activity, and their effects on degenerative diseases and aging, has been addressed by many studies that have been carried out mainly by taking advantage of the mtDNA inherited variants (Bellizzi et al., 2006; Brown et al., 1997; De Benedictis et al., 1999; Ivanova et al., 1998; Montiel-Sosa et al., 2006; Niemi et al., 2003; Rose et al., 2002; Ross et al., 2001; Ruiz-Pesini et al., 2000; van der Walt et al., 2003; Santoro et al., 2006; Wallace, 2005). An association of haplogroup J with longevity has been reported for different European populations, being this haplogroup more frequent in centenarians than in ethnically matched younger controls in Northern Italy (De Benedictis et al., 1999), Ireland (Ross et al., 2001), and Finland (Niemi et al., 2003). Recently, cells with mtDNA molecules belonging to haplogroup J were found to produce less ATP and less ROS than cells with different mtDNA molecules (Bellizzi et al., 2012). Interestingly, haplogroup J was also reported to be strongly associated with Leber's Hereditary Optic Neuropathy (LHON), increasing the penetrance of specific mutations that are significantly less harmful when occurring in mtDNA molecules of different haplogroups (Brown et al., 1997; Man et al., 2004; Torroni et al., 1997). Therefore, it seems that the mutations found in the haplogroup J (laying in the genes for subunits of OXPHOS complex I), put the cell in a situation of instability, where further simultaneous mutations can determine a disease (such as LHON) or, on the contrary, healthy aging (Rose et al., 2001).

In agreement with this hypothesis, it has been reported that the Asian haplogroup D (characterized as haplogroup J by mutations affecting complex I) is overrepresented in Japanese centenarians (Tanaka et al., 1998, 2000). However, these results were not confirmed when the studies were replicated in other geographic areas (Collerton et al., 2013; Dato et al., 2004; Pinós et al., 2012), suggesting not only that the possible effect of mtDNA on longevity might be population-specific, but also that the interaction with the nuclear genetic background can hide or minimize it. If this should be the case, it is likely that the association between mtDNA variability and longevity can reach statistical significance only when a very large number of samples is analyzed. The GEHA project, studying over 2200 subjects older than 90 years coming from different European populations and a similar number of younger controls matched for gender and geographic origins, allowed us to determine that recurrent or sporadic mutations accumulated in specific genes, not detected by haplogroup analysis, may influence longevity. These findings derived from the haplogroup classification (through D-loop sequencing) for nearly 3000 samples and from the complete sequencing of other 1200 samples [650 ultra nonagenarians (90+) and a comparable number of controls, coming from Denmark, Finland, Southern Italy and Greece]. The analysis of a large number of complete mtDNA sequences from European 90+ and geographically matched younger controls collected in the frame of GEHA, has allowed us to shed light on this subject (Raule et al., 2013). Indeed, the analysis of complete sequences, by means of specific analyses based on evaluating the burden of non-synonymous mutations in different mtDNA regions (Wu et al., 2011), made possible to appreciate the aggregate effects of rare or repeated mutations unaccounted for in haplogroup analysis. More specifically, the mutations falling on mtDNA molecules of long lived subjects have been aggregated and compared with those harbored by mtDNA molecules of younger subjects, by means of SKAT, a supervised and flexible regression method to test for the association between rare variants in a gene or genetic region with a continuous/dichotomous trait (Wu et al., 2011). This analysis allowed us to emphasize the effects of mutations (including rare mutations) falling in a specific region, as each of those mutations can have an effect (in most cases a small effect) on longevity but none of them can reach the statistical significance in a case control comparison. This has pointed out that the burden of non-significant mutations in mtDNA genes coding for subunits of OXPHOS is different in 90+ subjects and controls for genes of mitochondrial subunits of the OXPHOS complexes I, III and V. In particular, it has emerged that the presence of mutations on complex I may be beneficial for longevity, while the co-occurrence of mutations on both complexes I and III or on both I and V might lower the individual's chances to reach longevity. As haplogroup J is characterized by mutations in complex I genes, this result might explain previous contrasting findings emerged from association studies on J haplogroup and longevity, as discussed above. Indeed, it can be hypothesized that populations where no association between haplogroup J and longevity was found can be characterized by higher frequency of non-synonymous mutations in complexes III and V. This is indeed the case for Danish and Southern European populations (Raule et al., 2013). It is worth mentioning that the analysis of mtDNA sequences in 90+ and controls has also shown that many mtDNA mutations associated

with degenerative diseases are as frequent in 90+ as in younger controls, supporting the idea that the effect of mtDNA mutations is greatly influenced by the individual-specific genetic background (nuclear genome and combination of mtDNA mutations), as well as by stochastic events (Moilanen et al., 2003; Rose et al., 2001; Wallace, 2010). On the whole, the analysis of mtDNA sequences in GEHA samples has shown that the effect of mtDNA variability on longevity is mainly due to the co-occurrence of rare, private mutations which cannot be fully analyzed by haplogroup analysis, but which cannot be analyzed one by one either. An example of this variability is reported in Fig. 1. Only the aggregate analysis of the mutations falling in a given mtDNA region may help to understand the effect of these mutations on longevity. It is also important to say that the effect of these mutations is quite small and then only a very large sample can give the statistical power to understand the effects of these mutations.

Very recently, a study showed that cybrids carrying a mtDNA belonging to J haplogroup have significantly altered the expression of eight nuclear genes of the alternative complement, inflammation and apoptosis pathways with respect to cybrids carrying haplogroup H (Kenney et al., 2013, 2014). This result, combined with the reported differences in ATP and lactate production between J and H haplogroup cybrids, suggests that J haplogroup mtDNA variants within cells can affect not only energy pathways but also other pathways not directly related to mitochondrial metabolism such as complement activation, inflammation and apoptosis. Furthermore, these data suggest that mitochondria can affect a number of cellular functions previously thought to be independent from them (Kenney et al., 2014).

The study of heteroplasmy is a challenge, as heteroplasmy patterns observed in adulthood are a mixture of both inherited and somatic mtDNA mutations, making them hard to detect because of the relatively low frequencies of each mutation and in consideration that the different tissues can harbor (and produce) different patterns of mutations as exemplified in Fig. 2 (Wallace, 2010). Regarding inherited mutations, it should also be considered that a low level of heteroplasmy can be inherited, meaning that some variants can be found in heteroplasmic state in both mother and child (Payne et al., 2013). This could favor further accumulation of heteroplasmy with age.

2.2. Somatic mtDNA variability

Many studies (Corral-Debrinski et al., 1992; Michikawa et al., 1999; Wang et al., 2001) show an accumulation of heteroplasmy in aging in different tissues (such as muscle and brain). A study of Sondheimer et al. (2011) analyzed the variations of heteroplasmy in 138 sites on a cohort of more than 2400 subjects of different ages (from 0 to 60 years old). This study identifies two types of sites, some tend to become more heteroplasmic with aging, and others on the contrary display a loss of variability and thus an increase of homoplasmy. This is the first study that considers a wide range of ages and highlights that the level of heteroplasmy increases with age, and not only with senescence.

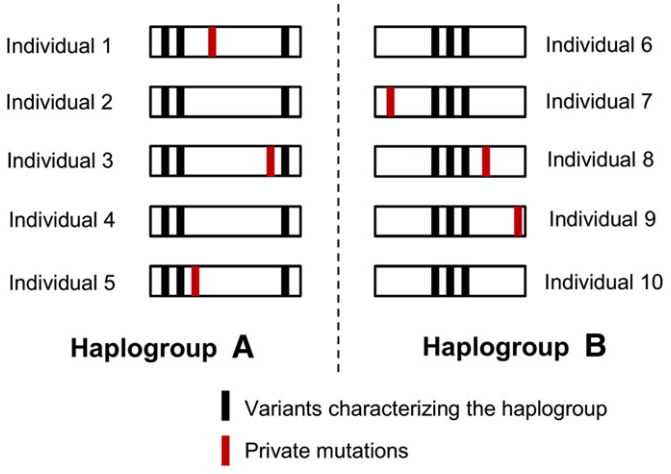


Fig. 1. Common (haplogroup) and private mutations in a population. Each individual is characterized by (i) variants that define the haplogroup –here marked in black – and (ii) private variants, that are rare in the population – here marked in red.

Some studies, such as the above mentioned GEHA EU project, have linked heteroplasmy to healthy aging and longevity (Castrì et al., 2009; Rose et al., 2007, 2010). Previous studies had shown a high incidence of the C150T transition in leukocytes from centenarians, thus hypothesizing a remodeling of mtDNA replication origin associated with this mutation (Niemi et al., 2005; Zhang et al., 2003). The GEHA EU Project results showed that the statistically significant high incidence of DLoop somatic point mutations, and the consequent heteroplasmy, in the mtDNA control region of long-lived subjects was not only restricted to the C150T transition but also affected additional positions correlated with mtDNA replication as well as the C150T transition (Rose et al., 2007, 2010). Moreover, these studies, carried out on sibships or on parent offspring groups, were allowed by permutation analysis to find out that the levels of heteroplasmy at position 150 and at other Dloop positions are similar among relatives and correlate in parent–offspring pairs, thus suggesting that they are modulated by a genetic component (Rose et al., 2007, 2010).

Moreover recent studies showed that the frequency of heteroplasmic variants may be different from one generation to another because of the bottleneck that occurs during postnatal folliculogenesis (Wai et al., 2008) and the somatic events that occur during life (Wallace, 2010; Sondheimer et al., 2011). The transmission of mtDNA heteroplasmy seems to be a central point in the study of aging. In fact, a recent study (Ross et al., 2013) supports the notion that a combination of maternally transmitted and somatic mtDNA mutations can lead to disorders (such as Alzheimer's disease and Parkinson's disease). This study is very interesting if combined with the innovative notion that even low level of heteroplasmy can be inherited as previously mentioned.

It is important to note that the dynamic of accumulation and the type of mutations seem to vary according to tissues and cell types. Moreover, recent studies indicate that the response to a mitochondrial dysfunction can be communicated to the nucleus via yet unidentified mediators and propagated to the whole organism via one or more putative signaling molecules indicated as mitokines (Nunnari and Suomalainen, 2012) creating a sort of crosstalk between mitochondrial genome of different tissues and organs of the same individual.

2.3. New methodological insights into mtDNA-related studies

2.3.1. The "explosion" of complete mtDNA sequences

The significant drop in costs and time commitment required to obtain a complete mtDNA sequence caused a burst in the number of available samples in both private and public databases. Genbank (Benson et al., 2013) currently hosts ~22,000 complete human mtDNA sequences, of which ~5000 were published after 1st January 2013. This accounts for about the 22% of the whole database, which was deposited in roughly last 12 months, with a submission rate that is way greater than the one observed in the previous years. This allows researchers to leverage an unprecedented number of samples that can be exploited in many different ways, for example to strengthen studies somewhat missing statistical power, or to assemble cohorts that no real project could even imagine to have. Such methods, however, must be preceded by a careful inspection for errors, both in the sequences themselves and in their annotation (Yao et al., 2009).

Often not all the biological questions can find an answer looking inside the database of previously published data. For such cases, sequencing of new samples is needed, and in these days this can be achieve in two ways: using NGS to specifically target mtDNA (Sosa et al., 2012) or extracting off-target reads obtained in experiments targeting other

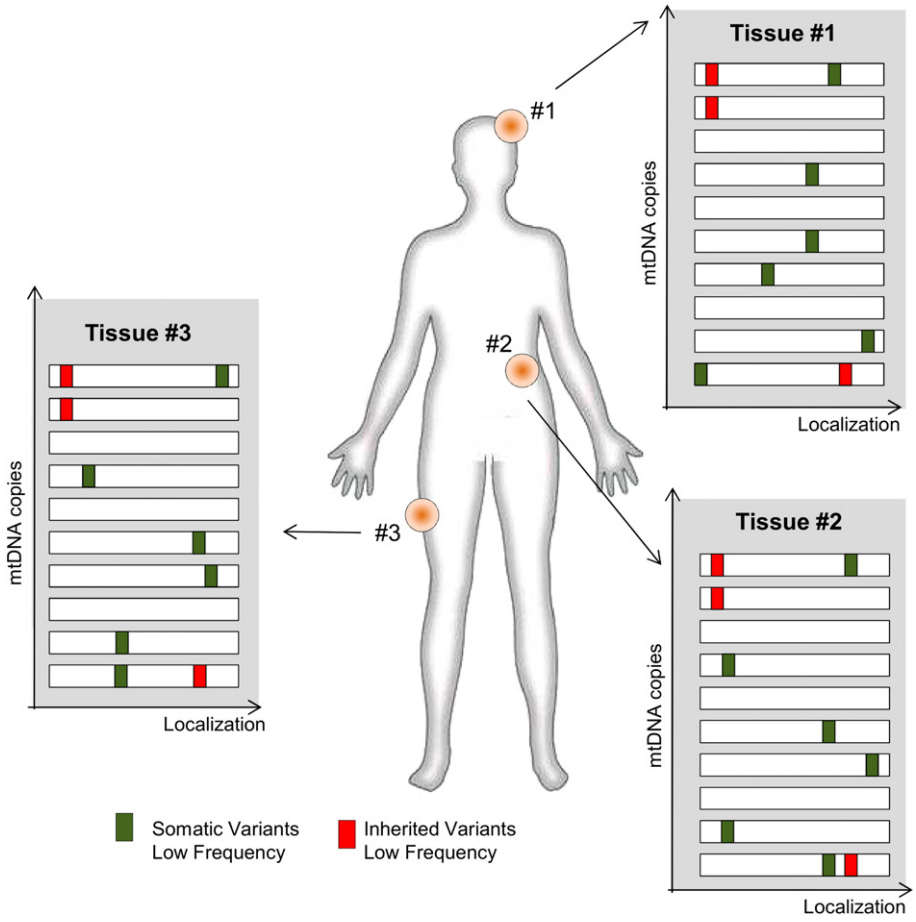


Fig. 2. Different scenarios of mtDNA heteroplasmy. One heteroplasmic mutation can be somatic or inherited – depending on the cell/tissue in which it originates. Somatic variants are marked in green while inherited variants are marked in red.

molecules (i.e. whole genome sequencing, exome sequencing). The latter approach, first suggested by Picardi and Pesole (2012), is now routinely adopted in the field: obtained coverage level allows to easily call variants, albeit a precise estimation of low heteroplasmies is out of reach (Sosa et al., 2012). Fictitious detection of low percentage heteroplasmies may also be caused by NUMTs (Hazkani-Covo et al., 2010; Li et al., 2012) that must be filtered out to obtain high-quality results. MitoSeek (Guo et al., 2013) represents the best answer to this problem to date.

2.3.2. Variants calling and annotation

Independently from the pipeline or software used to obtain them, sequences must be further analyzed to identify and annotate variants. Specialized software called variant callers are used to detect mutations from the aligned reads. The two most famous, freely accessible and open source packages in the field are GATK (DePristo et al., 2011; McKenna et al., 2010) and SAM tools (Li et al., 2009). Identified variation must then be extensively filtered to reduce the number of candidate variants (or genes) that may answer the biological question under investigation. mtDNA analysis thus encounters mainly the same issues that a nuclear-based study has: researchers must find good data sources (e.g. primary databases) for annotating variants, crunch them together and leverage this augmented information to pull out from the plethora of mutations which detected only the handful of them that is relevant to the problem.

Fortunately, much of the toolset and resources routinely adopted in nuclear genomics can also be directly applied to mtDNA. dbSNP (Sherry et al., 2001) is the NCBI-provided database which hosts the known human (and non-human) variability, which hosts 2000 mtDNA variants in the latest release. OMIM (Online Mendelian Inheritance in Man, Cusick-Nathans Institute of Genetic Medicine, Johns Hopkins University) represents the most widely recognized source about manually curated information regarding mutations causing human diseases. Mitochondrial-focused databases, such as MITOMAP (Ruiz-Pesini et al., 2007) or mtDB (Ingman and Gyllensten, 2006) indeed exist, however they may be difficult to fully exploit programmatically due to the lack of an API-based querying system. Independent on the source used to obtain the raw annotations, software such as ANNOVAR (Wang et al., 2010) or SnpEff (Cingolani et al., 2012) can help in disentangling the problem of merging all this information together in a single place and use this new, combined database to provide an efficient and as much comprehensive annotation as possible to the mutations under analysis.

2.3.3. Statistical methods

Relating the mutation on the mtDNA to the aging process requires some carefulness, as the ideal analysis would be a longitudinal study that correlates the survival time to the presence of mutations. Being this kind of study impractical for the long time and costs involved, one is forced to used transversal studies in a study-control setting. This kind of analysis is performed with a logistic regression, linking the presence of SNPs to the longevity of the individual via a regression of the log of the odds of being long-lived. This approach is as powerful as dangerous, as several confounding effect should be taken into consideration. First of all, this classification problem suffers from the base rate fallacy, where trying to identify a category (long-lived versus normal) can introduce a bias in the power of each regressor if one of the two categories is far less common than the other (Pepe et al., 2004).

Issues may also arise from the strategy applied to study single SNPs: assigning an odd ratio to each of them and selecting the most relevant one on the basis of this value (Gail, 2008) is known to be a method prone to selection of non-relevant SNP. On the other side, as the aging process is probably regulated slightly by several interacting factors, the little amount of information can be easily confounded by the huge amount of different features tested, especially given the typically small dimension of the sample (from a statistical point of view) (Greenland, 2000; Marchini et al., 2005; Tranah, 2011). This problem is especially relevant for the interaction between SNPs, as the number of possible interactions grows quadratically with the number of mutations under study. Additionally, the small number of patients involved can also lead to an overestimation of the correlation between mutations, forcing a re-evaluation of the estimated risk factors.

Several methods are available to select the relevant set of SNPs on which performing the statistical analysis for the prediction, each exploiting various strategies to select a plausible optimal subset of SNPs (Chen et al., 2013). These subsets are explored employing different search strategies and the most promising one are selected, typically using an associated based scoring (Chuang et al., 2013; Piriyapongsa et al., 2012; Zhang and Liu, 2007). To help this selection process and correct for spurious observation previous biological knowledge can be injected into the statistical analysis, for example via interaction maps and pathway analysis (Hua et al., 2012; Jamshidi and Palsson, 2006; Liu et al., 2012). This is especially helpful in the interaction studies with the nuclear genome, where the number of possible interactions would otherwise explode. Other selection methods are obviously available, as for example the flux balance analysis which can be useful to infer how sensible a metabolic pathway is to changes in the structure or amount of a protein (Orth et al., 2010). This allows to consider not only the statistical relevance but also the biological one, due to the different non-linearity of the network kinetic.

2.3.4. The missing bricks

mtDNA analysis can adopt the nuclear genome toolbox for its own needs, but these tools do not allow taking mtDNA peculiarities into consideration (i.e. heteroplasmy) in the analysis. Thus, a new generation of dedicated algorithms is needed to successfully investigate the mitochondrial genome. In the last years this gap progressively started to close with the release of a series of algorithms mainly focused on the haplogrouping process and, only in part, on the identification and annotation of the variability contained in the submitted samples. Successful examples in this field are MITOMASTER (Brandon et al., 2009), HaploGrep (Kloss-Brandstätter et al., 2011), HaploFind (Vianello et al., 2013) and HmtDB (Rubino et al., 2012). Nonetheless, these applications usually answer only to the first part of the problem: promising indentified mutations must then be analyzed to identify statistical significance with the phenotype under investigation. This probably represents the most “desert” area in the mitochondrial genome field: de-novo algorithms, or even re-implementations of nuclear-derived ones, which analyze the mtDNA data in innovative ways, going further down the road of identifying the processes or phenotypes where mitochondria have a role.

3. Age-related diseases and mtDNA

Defects of OXPHOS compromising mitochondrial functionality are indeed key players in the aging process and in the pathophysiology of different age-associated diseases (Wallace et al., 2010; Chinnery and Hudson, 2013). The triggering factors of these damages are copious, and also mtDNA sequence variation could contribute to a shift of energy balance facilitating aging and disease susceptibility (Wallace, 2005). Besides aging, the involvement of mtDNA mutations and haplogroups has also been investigated in several age-related neurodegenerative diseases such as Parkinson's Disease (Hudson et al., 2013; Pyle et al., 2005) and Alzheimer's disease (AD) (Lakatos et al., 2010; Santoro et al., 2010), cardiovascular diseases (Strauss et al., 2013), cancer (Lam et al., 2012; Weigl et al., 2013), metabolic diseases like type 2 diabetes (T2D) (Achilli et al., 2011; Nardelli et al., 2013) and age-related macular degeneration (Udar et al., 2009) among others. All these studies reported controversial results and no consensus has been reached so far. Hereafter we will discuss the main data obtained for AD and T2D, as they are not only among the most frequent age-related diseases, but also because of the prominent pathogenic role played by mitochondria in these pathologies. Actually, mitochondria produce ROS and are β-amyloid targets within neuronal cells, while in metabolic diseases the energy metabolism is deeply deranged including mitochondrial dysfunction.

The sporadic form of AD (SAD) is the leading cause of dementia among elderly. SAD has a complex etiology due to environmental and genetic factors that can interact with each other in different combinations to prime the disease (Sosa-Ortiz et al., 2012). Also, mitochondrial dysfunction plays a consistent role and mitochondrial variants, both mtDNA haplogroups or specific mtDNA polimorphysms, have been reported to be associated to AD (Chinnery and Hudson, 2013). Among the 9 European haplogroups, haplogroups T and J (Chagnon et al., 1999), U and K (Fesahat et al., 2007; Lakatos et al., 2010; van der Walt et al., 2004) and H (Coto et al., 2011; Fesahat et al., 2007; Maruszak et al., 2009; Ridge et al., 2012; Santoro et al., 2010) have been found to be associated to a higher or lower risk of SAD, while many other studies failed to find any association (Chinnery et al., 2000; Elson et al., 2006; Hudson et al., 2012; Krüger et al., 2010; Mancuso et al., 2007; Pyle et al., 2005; van der Walt et al., 2005). Table 1 summarizes these results. The reasons for these differences are presently unclear but are at least in part ascribable to methodological issues such as sample size, the ethnicity and the number of markers used to define haplogroups. Recently, we identified in a large cohort of Italian patients and controls subhaplogroup H5 as a risk factor for AD (Santoro et al., 2010) independently from the APOE4 status. The strength of this study was the resolution level (analysis of complete D-loop sequencing and specific markers along the coding sequence) and the quite large number of subjects (1712) of homogeneous ethnicity (central-northern Italy). Another group found that H haplogroup is a risk factor for AD in a Spanish population (Coto et al., 2011), and it is associated with APOE4 risk, while Ridge et al. (2012) reported that two sub-branches of haplogroup H (H6a1a and H6a1b) are protective factors for AD and are not related to APOE4 allele. In the attempt to confirm these results, Hudson et al. (2012) performed the analysis of 138 mtDNA SNPs in 3 Caucasian populations followed by a meta-analysis of the available data failing to replicate any association. Also the correlation with APOE needs further investigation. In fact, some studies found a significant protective effect of haplogroups U and K in patients carrying the APOE4 allele (Carrieri et al., 2001), while others identified a risk effect of haplogroup H (Ridge et al., 2012), and others found that the effect of mtDNA haplogroups on AD risk is independent of APOE status.

T2D is a common metabolic disorder affecting a large part of the aged population and its risk is increasing in the last years. Several variants in the nuclear genome have been identified (De Silva and Frayling, 2010; Dimas et al., 2013), however given the small effect of these variants in this complex phenotype, many other susceptibility factors are likely present. Mutations in mtDNA cause several syndromes that involve hyperglycemia (Amaral et al., 2008). Many studies tried to find a correlation between mtDNA haplogroups and T2D, with controversial results (see Table 1). Yet, most studies did not find associations between T2D and mtDNA variability, or reported conflicting results (Chinnery et al., 2007; Hsouna et al., 2013; Saxena et al., 2006). Interestingly, our group, analyzing the mtDNA variations in a large cohort of Italian patients and controls (Achilli et al., 2011), found significant associations between haplogroups H, H3, U3 and V and diabetes complications such as neuropathy, retinopathy, nephropathy and renal failure. On the same line a previous study found that haplogroup T was associated with retinopathy in Europeans (Kofler et al., 2009).

Table 1
Studies on mtDNA haplogroups and complex traits (type 2 diabetes and Alzheimer's disease). mtDNA haplogroups and/or clusters having higher frequency in cases than controls (over-represented) are considered to have a positive association with the trait, while those under-represented are considered to have a negative association with the trait. D, H, I, J, K, T, U, W, and X refer to haplogroup nomenclature as reported in MITOMAP (http://www.mitomap.org/).

Complex trait	N. cases vs controls	Results	Ethnicity	Reference
Type 2 diabetes	762 vs 402	Over-represented: J	Finnish	Mohlke et al. (2005)
	347 vs 350	Over-represented: J and T	Caucasian-Brazilian	Crispim et al. (2006)
	3304 vs 3304	No association	Swedish, Scandinavian, Canadians, Polish and American	Saxena et al. (2006)
	897 vs 1010	No association	English	Chinnery et al. (2007)
	1289 vs 1617	Under-represented: N9a	Japanese	Fuku et al. (2007)
	732 vs 633		Korean	
	227 vs 1527	Over-represented: T (complications)	Europe Caucasian	Kofler et al. (2009)
	466 vs 438	Under-represented: H1 Over-represented: H, H3, U3 and V (complications)	Italian	Achilli et al. (2011)
	184 vs 288	Over-represented: J Under-represented: T	Spanish (Canary Islands)	González et al. (2012)
	859 vs 1151	Over-represented: B Under-represented: D	Chinese	Liou et al. (2012)
	64 vs 77	No association	Tunisian	Hsouna et al. (2013)
Alzheimer's disease	69 vs 83	Under-represented: T Over-represented: J	French-Canadian	Chagnon et al. (1999)
	185 vs 179	No association	English	Chinnery et al. (2000)
	644 vs 180	Under-represented: U	Caucasian females	van der Walt et al. (2004)
	345 vs 148	Over-represented: U	Caucasian males	van der Walt et al. (2004)
	190 vs 340	No association	Old Order Amish	van der Walt et al. (2005)
	185 vs 447	No association	English	Pyle et al. (2005)
	75 vs 64	No association	US Americans; English	Elson et al. (2006)
	70 vs 64	No association		
	30 vs 100	Over-represented: H and U	Iranian	Fesahat et al. (2007)
	209 vs191	No association	Italian	Mancuso et al. (2007)
	222 vs 252	Over-represented: HV	Polish	Maruszak et al. (2009)
	128 vs 99	No association	Finnish	Krüger et al. (2010)
	936 vs 776	Over-represented: H5	Italian	Santoro et al. (2010)
	170 vs 188	Over-represented: UK	Caucasian	Lakatos et al. (2010)
	500 vs 500	Over-represented: H	Spanish	Coto et al. (2011)
	632 vs 101	Under-represented: H6a1a and H6a1b	US Americans, Utah	Ridge et al. (2012)
	4133 vs 1602	No association	English; US Americans; German	Hudson et al. (2012)

Summarizing, like in AD, the scenario regarding the involvement of mtDNA haplogroups in T2D is far from being elucidated and additional studies at a high level of phylogenetic resolution in large populations with similar haplogroup/subhaplogroup profiles are required. Also the stratification for complications should be included in the analyses. Moreover functional analyses could help to unravel the role of mtDNA mutation in a specific molecular pathway of the pathogenesis. To this purpose Hwang et al. (2011) recently explored the protective effect of Asiatic haplogroup N9a and the risk effect of haplogroup D5 and F using cybrid cell lines having the same nuclear background (Hwang et al., 2011). Notably, the risky haplogroups were associated with a down-regulation of oxidative phosphorylation and an up-regulation of glycolysis compared to the protective haplogroup. These results suggest that variations in mtDNA can affect the expression of nuclear genes regulating mitochondrial functions or cellular energetics.

Since AD and T2D are common age-related disorders, even a small risk associated with mtDNA haplotype could have major causative implications at a population level. It is therefore critically important to determine the role of mtDNA polymorphisms by studying large cohorts of subjects at a level of molecular and phylogenetic resolution as fine as possible. In this attempt, the combination of new NGS methodologies such as GWAS, high-throughput whole genome and whole exome sequencing on nuclear genome and the complete resequencing of the mitochondrial genome applied to very large cohorts of patients and carefully matched controls will provide an efficient approach to detect common and rare variation contributing to disease risk. In this frame, we sequenced a high number of ultra-nonagenarians (90+) and younger controls and found that the frequency of mutations previously reported to be associated with degenerative diseases is higher in 90+ than in controls (Raule et al., 2013). These results support the idea that certain mtDNA inherited mutations could contribute to longevity or disease according to population- and individual-specific genetic backgrounds (nuclear genome and combination of mtDNA mutations), as well as stochastic events (Moilanen et al., 2003; Rose et al., 2001; Wallace, 2010).

Finally, an aspect that has to be taken into account when performing association studies between mtDNA variants and age-related diseases is that certain diseases can be more frequent in distinct populations than in others, and it is well known that mtDNA exhibits, along with Y chromosome, strong signals of geographic genetic differentiation, due to the demographic history, sex biased human migrations and other cultural factors.

4. mtDNA Damage Repair and Aging

mtDNA is packaged into multimeric complexes called nucleoids made up of proteins and nucleic acids (Chen and Butow, 2005). These proteins protect the mtDNA molecules that are otherwise particularly exposed to ROS, as mentioned, and also include factors that help mtDNA to repair and replicate. This review is not focused on these processes, nevertheless from a mechanistic point of view it is becoming more and more clear that the accumulation of mutations with age occurs via alterations of the systems that ensure mtDNA integrity to be maintained over time, rather than increased level of damage, e.g. oxidative stress (Itsara et al., 2014). Moreover, it is worth to underline the importance of these factors to better understand the epistatic phenomena that can take place between nuclear and mitochondrial genomes. As an example, a key molecule for mtDNA stability is polγ, which is the sole mitochondrial polymerase and plays an essential role in both replication and repair of the mtDNA (Kaguni, 2004). As previously mentioned, a polγ defective in proofreading activity has been associated to a phenotype of accelerated aging in a murine model (Kujoth et al., 2005; Trifunovic et al., 2004). Alterations in the polymerase and exonuclease activity of polγ induce an increase in the frequency of mtDNA mutations (Van Goethem et al., 2002: Copeland et al., 2003) and mutations of POLG1 gene are associated with a number of diseases (Cohen et al.,

2010; Longley et al., 2005). Data also suggest that polγ is likely involved in the aging process (Kujoth et al., 2007) and in the pathogenesis of a series of diseases in which mtDNA instability plays a pathogenic role, such as age-associated sarcopenia and Parkinson's Disease (Fayet et al., 2002; Luoma et al., 2004). Nevertheless, the molecular details by which polγ ensures mtDNA integrity are still to be elucidated. TFAM is another very important factor for mtDNA stability, since it is part of the nucleoid packaging and it is involved in mtDNA transcription and replication (Kang and Hamasaki, 2005). Moreover, TFAM is responsible for mtDNA replication and therefore controls mtDNA copy number. However, it has been reported that TFAM can bind preferentially to damaged mtDNA and inhibits the in vitro incision activity of 8-oxoguanine DNA glycosylase (OGG1), uracil-DNA glycosylase (UDG), apurinic endonuclease 1 (APE1), and nucleotide incorporation by pol γ (Canugovi et al., 2010). Its role in mtDNA maintenance in aging has not yet been clarified, but some studies in animals suggest that it could be implicated in accumulating mtDNA mutations (Zhong et al., 2011). It has been reported that p53 also plays a role in the maintenance of mtDNA integrity (Achanta et al., 2005; Bakhanashvili et al., 2008; Canugovi et al., 2010; de Souza-Pinto et al., 2004). p53 can translocate at mitochondrial level in response to oxidative stress (Bonafé et al., 2004; Marchenko et al., 2000; Mihara et al., 2003) where it can bind to polγ, TFAM and other factors such as GSK3b modulating the amount of mtDNA copy number and mutations (Achanta et al., 2005; Altilia et al., 2012; de Souza-Pinto et al., 2004; Vadrot et al., 2012; Wong et al., 2009; Yoshida et al., 2003). Moreover, studies in p53 knock out animals collectively demonstrated a reduced number of mtDNA copies, a reduced base excision repair activity and, as a whole, a reduced mitochondrial activity (Achanta et al., 2005; Bartlett et al., 2014; Chen et al., 2006; de Souza-Pinto et al., 2004). Interestingly, the interaction of p53 with TFAM alleviates TFAM-induced inhibition of BER proteins (Canugovi et al., 2010). In our lab we were able to demonstrate that different p53 isoforms, p53P72 and p53R72 (resulting from a common polymorphism that causes a Proline-to-Arginine substitution at position 72), have a different capability to localize at mitochondrial level and bind to polγ (Altilia et al., 2012). Accordingly, it was found that subjects carrying different p53 codon 72 genotypes were characterized by different amounts of heteroplasmy at the level of D-loop, as detected by DHPLC techniques. However, while a plethora of data suggests that mtDNA mutations accumulate with age, it is still a matter of debate whether such an accumulation is a cause of aging. It has been reported that mtDNA deletions, but not point mutations, are associated with premature aging and lifespan in mice with defective polγ (Vermulst et al., 2007, 2008). However, a recent report suggests that germline-transmitted mtDNA mutations may aggravate the aging phenotype of adult animals (Ross et al., 2013). These authors speculate that this phenomenon may also be relevant for human aging. It is however to note that while cells containing clonally expanded mtDNA mutations appear to be ubiquitous in postmitotic tissues of aged humans, even in the very old they only represent a small minority of cells (<5%). It is therefore not clear whether the mutations actually contribute to the aging process or are minor bystanders (Keogh and Chinnery, 2013). However, it is possible that small levels of inherited heteroplasmy while not compromising germ line function, can prime and potentiate the effect of mtDNA mutations on aging (Keogh and Chinnery, 2013).

5. Perspectives

Strong evidences support the idea that mitochondria are involved in the aging process and in longevity. As discussed, it is not totally clear whether this is also the case for mtDNA. Actually, findings that emerged in the framework of the GEHA EU project have highlighted that the contribution of mtDNA variability is strongly dependent on the interaction between different loci of the mtDNA itself (Raule et al., 2013). In particular, the analysis of the mutations occurring in mtDNA OXPHOS genes shows a complex scenario with a different mutation burden in ultranonagenarian subjects with respect to younger controls. Quite unexpectedly, mutations in the subunits of the OXPHOS complex I seem to have a beneficial effect on longevity, but if they occur simultaneously with mutations in complexes III and V seem to be detrimental. It can be further hypothesized that the interaction between the mitochondrial and the nuclear genomes is even more important. The availability of data for both genomes obtained from a large number of long living subjects within the GEHA project will allow the reaching of a sufficient statistical power needed for the evaluation of such mito-nuclear genetic interaction (epistasis). Moreover, we surmise that, in order to get more insights on longevity, it will be useful to study the subjects that reached the extreme limits of human age, i.e. centenarians, semisupercentenarians and super-centenarians, who are likely provided with a genetic background predisposing to long life (Garagnani et al., 2013; Perls and Terry, 2003; Tan et al., 2008). Therefore, the analysis of mtDNA variability and of the interactions between mitochondrial and nuclear genomes performed on those exceptional subjects will contribute significantly to the identification of the combination of genetic determinants (from both genomes) that lead to longevity. At variance, the analysis of patients suffering by AD or T2D could help in identifying patterns associated with an increased risk for these diseases.

Two new research topics that could have functional effects on aging and longevity are emerging in the field of the regulation of mtDNA transcription: the mitochondrial epigenetics and the mitochondrial small and long non coding RNAs.

The correlation between mtDNA and epigenetics has been investigated for the central role of mitochondria in energy production. Indeed, this has suggested that the variability of mtDNA affecting energy production may also affect epigenetic changes, as energy production modulates the activity of methionine adenosyltransferase (MAT), the enzyme responsible for S-adenosylmethionine (SAM) synthesis from L-methionine and ATP (Wallace and Fan, 2010). Indeed, the analysis of a population sample revealed that methylation levels were higher in the subjects carrying the J haplogroup than in non-J carriers. Consistently, cybrids with J haplogroup mtDNA, where over expression of the MAT1A gene and low ATP were observed, showed higher methylation levels than other cybrids. These findings confirmed that specific interactions between mitochondria and the nucleus, partly modulated by mtDNA variations, regulate epigenetic changes, possibly through its effects on oxidative phosphorylation efficiency (Bellizzi et al., 2012).

Although some authors suggested the absence of methylation in mtDNA (Hong et al., 2013 and references therein), data indicating that also mtDNA can undergo methylation are emerging. Indeed, different amounts of methyl cytosines have been observed in mtDNA of many organisms, although the distribution of the methylated cytosines has not been determined in any of these species (Nass, 1973; Pollack et al., 1984; Rebelo et al., 2009; Shmookler Reis and Goldstein, 1983a, 1983b; Šimková, 1998). Then, Shock et al. (2011) demonstrated an enrichment of mtDNA sequences by immunoprecipitation against 5methylcytosine and 5hydroxy methylcytosine and the translocation of DNA (cytosine-5-)-methyltransferase 1 (DNMT1) into the mitochondria. More recently Bellizzi et al. (2013) reported evidences for epigenetic modifications in mtDNA, as methylated and hydroxymethylated cytosines were detected within the mitochondrial D-loop in DNA samples extracted from blood and cultured cells of both humans and mouse. A peculiar aspect of these data is that the majority of the methylated cytosines were located outside of CpG nucleotides in an asymmetrical pattern previously reported for plants and fungi. Of note is also the existence of intercellular variability for -both CpG and nonCpG methylation, with higher methylation levels in human blood and HeLa cells with respect to fibroblasts and osteosarcoma cells. The same pattern was observed in mouse, supporting the notion of tissue specificity of DNA methylation. In both human and mouse samples methylated cytosines were identified in the promoter region of the heavy filament (PH) and within conserved sequence blocks (CSBI-III), which are highly conserved sequences located at the 5′ end of the D-loop and considered

to be implicated in the processing of the RNA primer during the replication of the H filament. Therefore, it is possible to hypothesize that D-loop methylation might play an important role in modulating either replication or transcription of mtDNA, two processes widely described as physically and functionally correlated. Indeed, the regions found methylated are involved in forming the DNA-RNA hybrid due to the transcription of the leading (H)-strand origin. It is of note that the formation of this hybrid occurs to initiate the H strand replication. Likely, the methylation of L-strand, displaying complementary sequence to RNA primer, may be involved in regulating the formation of the hybrid or stabilizing its persistence. As to the mechanisms responsible for the methylation of mtDNA molecules, DNA methyltransferases were observed within the mitochondria, but the inactivation of Dnmt1, Dnmt3a, and Dnmt3b in mouse embryonic stem (ES) cells resulted in a reduction of the methylation of CpG, while the methylation of non-CpG resulted not affected (Bellizzi et al., 2013). This suggests that D-loop epigenetic modification is only partially established by these enzymes. Further studies are needed to clarify this point.

Recently it has been observed that certain microRNA (miRs) can translocate from cytosol into the mitochondria, for this reason they are called mito-miRs, and their role in physiological function such as apoptosis, proliferation and cell senescence, is reviewed elsewhere in the present issue. Many studies suggest that mitochondria are one of the destinations of pre- and mature miRs, and recently a unique population of small RNAs (sRNAs) including miRs that are associated with mitochondria has been identified in two human cell lines by deep sequencing (Sripada et al., 2012). Interestingly, it was observed that, among the many hundreds of miRs identified in this study, 4 known and 24 putative novel miRs aligned to mitochondrial genome at different positions corresponding to 16S rRNA, tRNA and subunits of complex I. Therefore, the authors suggest that these miRs may regulate mitochondrial transcripts involved in ribosome assembly and electron transport chain. Authors also do not exclude that these miRs could be encoded by mtDNA, but this speculation has yet to be demonstrated. However, should these data be confirmed, they will open a new perspective on the fine tuning of mtDNA expression, much more similar to nuclear gene regulation than previously thought. Accordingly, not only small RNA, but also long non coding RNAs (lncRNAs) are generated from the mitochondrial genome and regulated by nuclear-encoded proteins as shown by Rackham et al. (2011) in human cell lines, thus suggesting a much more complex nuclear-mitochondrial cross-talk. This cross-talk needs to be more deeply investigated by comparing different tissues with massive parallel sequencing analysis in order to evaluate the differences among organs with various levels of metabolism or energy consumption, and further, to assess age-dependent effects on such a cross-talk.

In conclusion, as it happened for the studies of genetics of aging and longevity regarding nuclear genes, also in the field of studies of mitochondrial genetics and aging there is still a lack of understanding due to the fact that mitochondrial genome acts in tight connection with the nuclear one and, perhaps, like the nuclear one, is regulated also at an epigenetic level. Therefore, it is envisaged that in order to shed light and disentangle this complex scenario, studies of nuclear/mitochondrial epistasis and mtDNA epigenetic regulation (methylation, binding with regulatory elements) will be necessary.

Conflict of interest

The authors have no conflicts of interests.

Acknowledgments

This work was supported by the GEHA Project, funded through Priority 1 (Life Sciences, Genomics and Biotechnology for Health) of European Union's FP6, Project Number LSHM-CT-2004-503270 and by the European Union's Seventh Framework Programme (FP7/2007–2011) under grant agreement no 259679 (IDEAL project).

References

Achanta, G., Sasaki, R., Feng, L., Carew, J.S., Lu, W., Pelicano, H., Keating, M.J., Huang, P., 2005. Novel role of p53 in maintaining mitochondrial genetic stability through interaction with DNA Pol gamma. EMBO J. 24, 3482–3492.

Achilli, A., Olivieri, A., Pala, M., Kashani, B.H., Carossa, V., Perego, U.A., Gandini, F., Santoro, A., Battaglia, V., Grugni, V., Lancioni, H., Sirolla, C., Bonfigli, A.R., Cormio, A., Boemi, M., Testa, I., Semino, O., Ceriello, A., Spazzafumo, L., Gadaleta, M.N., Marra, M., Testa, R., Franceschi, C., Torroni, A., 2011. Mitochondrial DNA backgrounds might modulate diabetes complications rather than T2DM as a whole. PLoS ONE 6, e21029.

Altilia, S., Santoro, A., Malagoli, D., Lanzarini, C., Ballesteros Álvarez, J.A., Galazzo, G., Porter, D.C., Crocco, P., Rose, G., Passarino, G., Roninson, I.B., Franceschi, C., Salvioli, S., 2012. TP53 codon 72 polymorphism affects accumulation of mtDNA damage in human cells. Aging (Albany NY) 4, 28–39.

Amaral, S., Oliveira, P.J., Ramalho-Santos, J., 2008. Diabetes and the impairment of reproductive function: possible role of mitochondria and reactive oxygen species. Curr. Diabetes Rev. 4, 46–54.

Bakhanashvili, M., Grinberg, S., Bonda, E., Simon, A.J., Moshitch-Moshkovitz, S., Rahav, G., 2008. p53 in mitochondria enhances the accuracy of DNA synthesis. Cell Death Differ. 15, 1865–1874.

Bartlett, J.D., Close, G.L., Drust, B., Morton, J.P., 2014. The emerging role of p53 in exercise metabolism. Sports Med. 44, 303–309.

Bellizzi, D., Cavalcante, P., Taverna, D., Rose, G., Passarino, G., Salvioli, S., Franceschi, C., De Benedictis, G., 2006. Gene expression of cytokines and cytokine receptors is modulated by the common variability of the mitochondrial DNA in cybrid cell lines. Genes Cells 11, 883–891.

Bellizzi, D., D'Aquila, P., Giordano, M., Montesanto, A., Passarino, G., 2012. Global DNA methylation levels are modulated by mitochondrial DNA variants. Epigenomics 4, 17–27.

Bellizzi, D., D'Aquila, P., Scafone, T., Giordano, M., Riso, V., Riccio, A., Passarino, G., 2013. The control region of mitochondrial DNA shows an unusual CpG and non-CpG methylation pattern. DNA Res. 20, 537–547.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E. W., 2013. GenBank. Nucleic Acids Res. 41, D36–D42.

Bonafé, M., Salvioli, S., Barbi, C., Trapassi, C., Tocco, F., Storci, G., Invidia, L., Vannini, I., Rossi, M., Marzi, E., Mishto, M., Capri, M., Olivieri, F., Antonicelli, R., Memo, M., Uberti, D., Nacmias, B., Sorbi, S., Monti, D., Franceschi, C., 2004. The different apoptotic potential of the p53 codon 72 alleles increases with age and modulates in vivo ischaemia-induced cell death. Cell Death Differ. 11, 962–973.

Brandon, M.C., Ruiz-Pesini, E., Mishmar, D., Procaccio, V., Lott, M.T., Nguyen, K.C., Spolim, S., Patil, U., Baldi, P., Wallace, D.C., 2009. MITOMASTER: a bioinformatics tool for the analysis of mitochondrial DNA sequences. Hum. Mutat. 30, 1–6.

Brown, M.D., Sun, F., Wallace, D.C., 1997. Clustering of Caucasian Leber hereditary optic neuropathy patients containing the 11778 or 14484 mutations on an mtDNA lineage. Am. J. Hum. Genet. 60, 381–387.

Canugovi, C., Maynard, S., Bayne, A.C.V., Sykora, P., Tian, J., de Souza-Pinto, N.C., Croteau, D. L., Bohr, V.A., 2010. The mitochondrial transcription factor A functions in mitochondrial base excision repair. DNA Repair (Amst) 9, 1080–1089.

Capri, M., Santoro, A., Garagnani, P., Bacalini, M.G., Pirazzini, C., Olivieri, F., Procopio, A., Salvioli, S., Franceschi, C., 2013. Genes of human longevity: an endless quest? Curr. Vasc. Pharmacol. (Epub ahead of print).

Carrieri, G., Bonafè, M., De Luca, M., Rose, G., Varcasia, O., Bruni, A., Maletta, R., Nacmias, B., Sorbi, S., Corsonello, F., Feraco, E., Andreev, K.F., Yashin, A.I., Franceschi, C., De Benedictis, G., 2001. Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease. Hum. Genet. 108, 194–198.

Castrì, L., Melendez-Obando, M., Villegas-Palma, R., Barrantes, R., Raventos, H., Pereira, R., Luiselli, D., Pettener, D., Madrigal, L., 2009. Mitochondrial polymorphisms are associated both with increased and decreased longevity. Hum. Hered. 67, 147–153.

Chagnon, P., Gee, M., Filion, M., Robitaille, Y., Belouchi, M., Gauvreau, D., 1999. Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population. Am. J. Med. Genet. 85, 20–30.

Chen, X.J., Butow, R.A., 2005. The organization and inheritance of the mitochondrial genome. Nat. Rev. Genet. 6, 815–825.

Chen, D., Yu, Z., Zhu, Z., Lopez, C.D., 2006. The p53 pathway promotes efficient mitochondrial DNA base excision repair in colorectal cancer cells. Cancer Res. 66, 3485–3494.

Chen, J.B., Chuang, L.Y., Lin, Y. Da, Liou, C.W., Lin, T.K., Lee, W.C., Cheng, B.C., Chang, H.W., Yang, C.H., 2013. Preventive SNP–SNP interactions in the mitochondrial displacement loop (D-loop) from chronic dialysis patients. Mitochondrion 13, 698–704.

Chinnery, P.F., Hudson, G., 2013. Mitochondrial genetics. Br. Med. Bull. 106, 135–159.

Chinnery, P.F., Taylor, G.A., Howell, N., Andrews, R.M., Morris, C.M., Taylor, R.W., McKeith, I. G., Perry, R.H., Edwardson, J.A., Turnbull, D.M., 2000. Mitochondrial DNA haplogroups and susceptibility to AD and dementia with Lewy bodies. Neurology 55, 302–304.

Chinnery, P.F., Mowbray, C., Patel, S.K., Elson, J.L., Sampson, M., Hitman, G.A., McCarthy, M. I., Hattersley, A.T., Walker, M., 2007. Mitochondrial DNA haplogroups and type 2 diabetes: a study of 897 cases and 1010 controls. J. Med. Genet. 44, e80.

Chuang, L.-Y., Chang, H.-W., Lin, M.-C., Yang, C.-H., 2013. Improved branch and bound algorithm for detecting SNP–SNP interactions in breast cancer. J. Clin. Bioinforma. 3, 4.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6, 80–92.

Cohen, B.H., Chinnery, P.F., Copeland, W.C., 2010. POLG-related disorders. In: Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. (Eds.), GeneReviews [Internet]. University of Washington, Seattle, Seattle (WA) (Available from http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&part=alpers).

Collerton, J., Ashok, D., Martin-Ruiz, C., Pyle, A., Hudson, G., Yadegarfar, M., Davies, K., Jagger, C., von Zglinicki, T., Kirkwood, T.B.L., Chinnery, P.F., 2013. Frailty and mortality are not influenced by mitochondrial DNA haplotypes in the very old. Neurobiol. Aging 34, 2889.e1–2889.e4.

Copeland, W.C., Ponamarev, M.V., Nguyen, D., Kunkel, T.A., Longley, M.J., 2003. Mutations in DNA polymerase gamma cause error prone DNA synthesis in human mitochondrial disorders. Acta Biochim. Pol. 50, 155–167.

Corral-Debrinski, M., Horton, T., Lott, M.T., Shoffner, J.M., Beal, M.F., Wallace, D.C., 1992. Mitochondrial DNA deletions in human brain: regional variability and increase with advanced age. Nat. Genet. 2, 324–329.

Coto, E., Gómez, J., Alonso, B., Corao, A.I., Díaz, M., Menéndez, M., Martínez, C., Calatayud, M.T., Morís, G., Álvarez, V., 2011. Late-onset Alzheimer's disease is associated with mitochondrial DNA 7028C/haplogroup H and D310 poly-C tract heteroplasmy. Neurogenetics 12, 345–346.

Crispim, D., Canani, L.H., Gross, J.L., Tschiedel, B., Souto, K.E.P., Roisenberg, I., 2006. The European-specific mitochondrial cluster J/T could confer an increased risk of insulin-resistance and type 2 diabetes: an analysis of the m.4216T>C and m. 4917A>G variants. Ann. Hum. Genet. 70, 488–495.

Dato, S., Passarino, G., Rose, G., Altomare, K., Bellizzi, D., Mari, V., Feraco, E., Franceschi, C., De Benedictis, G., 2004. Association of the mitochondrial DNA haplogroup J with longevity is population specific. Eur. J. Hum. Genet. 12, 1080–1082.

De Benedictis, G., Rose, G., Carrieri, G., De Luca, M., Falcone, E., Passarino, G., Bonafe, M., Monti, D., Baggio, G., Bertolini, S., Mari, D., Mattace, R., Franceschi, C., 1999. Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans. FASEB J. 13, 1532–1536.

De Silva, N.M.G., Frayling, T.M., 2010. Novel biological insights emerging from genetic studies of type 2 diabetes and related metabolic traits. Curr. Opin. Lipidol. 21, 44–50.

De Souza-Pinto, N.C., Harris, C.C., Bohr, V.A., 2004. p53 functions in the incorporation step in DNA base excision repair in mouse liver mitochondria. Oncogene 23, 6559–6568.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.

Dimas, A.S., Lagou, V., Barker, A., Knowles, J.W., Mägi, R., et al., 2013. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes (epub ahead of print).

Elson, J.L., Herrnstadt, C., Preston, G., Thal, L., Morris, C.M., Edwardson, J.A., Beal, M.F., Turnbull, D.M., Howell, N., 2006. Does the mitochondrial genome play a role in the etiology of Alzheimer's disease? Hum. Genet. 119, 241–254.

Fayet, G., Jansson, M., Sternberg, D., Moslemi, A.R., Blondy, P., Lombès, A., Fardeau, M., Oldfors, A., 2002. Ageing muscle: clonal expansions of mitochondrial DNA point mutations and deletions cause focal impairment of mitochondrial function. Neuromuscul. Disord. 12, 484–493.

Fesahat, F., Houshmand, M., Panahi, M.S.S., Gharagozli, K., Mirzajani, F., 2007. Do haplogroups H and U act to increase the penetrance of Alzheimer's disease? Cell. Mol. Neurobiol. 27, 329–334.

Fuku, N., Park, K.S., Yamada, Y., Nishigaki, Y., Cho, Y.M., Matsuo, H., Segawa, T., Watanabe, S., Kato, K., Yokoi, K., Nozawa, Y., Lee, H.K., Tanaka, M., 2007. Mitochondrial haplogroup N9a confers resistance against type 2 diabetes in Asians. Am. J. Hum. Genet. 80, 407–415.

Gail, M.H., 2008. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. J. Natl. Cancer Inst. 100, 1037–1041.

Garagnani, P., Giuliani, C., Pirazzini, C., Olivieri, F., Bacalini, M.G., Ostan, R., Mari, D., Passarino, G., Monti, D., Bonfigli, A.R., Boemi, M., Ceriello, A., Genovese, S., Sevini, F., Luiselli, D., Tieri, P., Capri, M., Salvioli, S., Vijg, J., Suh, Y., Delledonne, M., Testa, R., Franceschi, C., 2013. Centenarians as super-controls to assess the biological relevance of genetic risk factors for common age-related diseases: a proof of principle on type 2 diabetes. Aging (Albany NY) 5, 373–385.

González, A.M., Maceira, B.M., Pérez, E., Cabrera, V.M., López, A.J., Larruga, J.M., 2012. Genetics, environment, and diabetes-related end-stage renal disease in the Canary Islands. Genet. Test. Mol. Biomark. 16, 859–864.

Greaves, L.C., Turnbull, D.M., 2009. Mitochondrial DNA mutations and ageing. Biochim. Biophys. Acta 1790, 1015–1020.

Greenland, S., 2000. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. Biostatistics 1, 113–122.

Guo, Y., Li, J., Li, C.-I., Shyr, Y., Samuels, D.C., 2013. MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. Bioinformatics 29, 1210–1211.

Harman, D., 1957. Prolongation of the normal life span by radiation protection chemicals. J. Gerontol. 12, 257–263.

Hazkani-Covo, E., Zeller, R.M., Martin, W., 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 6, e1000834.

Hekimi, S., Lapointe, J., Wen, Y., 2011. Taking a "good" look at free radicals in the aging process. Trends Cell Biol. 21, 569–576.

Hong, E.E., Okitsu, C.Y., Smith, A.D., Hsieh, C.-L., 2013. Regionally specific and genome-wide analyses conclusively demonstrate the absence of CpG methylation in human mitochondrial DNA. Mol. Cell. Biol. 33, 2683–2690.

Hsouna, S., Ben Halim, N., Lasram, K., Arfa, I., Jamoussi, H., Bahri, S., Ammar, S. Ben, Miladi, N., Abid, A., Abdelhak, S., Kefi, R., 2013. Association study of mitochondrial DNA polymorphisms with type 2 diabetes in Tunisian population. Mitochondrial DNA 1736, 1–6.

Hua, L., Lin, H., Li, D., Li, L., Liu, Z., 2012. Mining functional gene modules linked with rheumatoid arthritis using a SNP–SNP network. Genomics Proteomics Bioinforma. 10, 23–34.

Hudson, G., Sims, R., Harold, D., Chapman, J., Hollingworth, P., Gerrish, A., Russo, G., Hamshere, M., Moskvina, V., Jones, N., Thomas, C., Stretton, A., Holmans, P.A., O'Donovan, M.C., Owen, M.J., Williams, J., Chinnery, P.F., 2012. No consistent evidence for association between mtDNA variants and Alzheimer disease. Neurology 78, 1038–1042.

Hudson, G., Nalls, M., Evans, J.R., Breen, D.P., Winder-Rhodes, S., Morrison, K.E., Morris, H. R., Williams-Gray, C.H., Barker, R.A., Singleton, A.B., Hardy, J., Wood, N.E., Burn, D.J., Chinnery, P.F., 2013. Two-stage association study and meta-analysis of mitochondrial DNA variants in Parkinson disease. Neurology 80, 2042–2048.

Hwang, S., Kwak, S.H., Bhak, J., Kang, H.S., Lee, Y.R., Koo, B.K., Park, K.S., Lee, H.K., Cho, Y.M., 2011. Gene expression pattern in transmitochondrial cytoplasmic hybrid cells harboring type 2 diabetes-associated mitochondrial DNA haplogroups. PLoS ONE 6, e22116.

Ingman, M., Gyllensten, U., 2006. mtDB: human mitochondrial genome database, a resource for population genetics and medical sciences. Nucleic Acids Res. 34, D749–D751.

Itsara, L.S., Kennedy, S.R., Fox, E.J., Yu, S., Hewitt, J.J., Sanchez-Contreras, M., Cardozo-Pelaez, F., Pallanck, L.J., 2014. Oxidative stress is not a major contributor to somatic mitochondrial DNA mutations. PLoS Genet. 10, e1003974.

Ivanova, R., Lepage, V., Charron, D., Schächter, F., 1998. Mitochondrial genotype associated with French Caucasian centenarians. Gerontology 44, 349.

Jamshidi, N., Palsson, B.Ø., 2006. Systems biology of SNPs. Mol. Syst. Biol. 2, 38.

Jia, J., Tian, Y., Cao, Z., Tao, L., Zhang, X., Gao, S., Ge, C., Lin, Q.-Y., Jois, M., 2010. The polymorphisms of UCP1 genes associated with fat metabolism, obesity and diabetes. Mol. Biol. Rep. 37, 1513–1522.

Kaguni, L.S., 2004. DNA polymerase gamma, the mitochondrial replicase. Annu. Rev. Biochem. 73, 293–320.

Kang, D., Hamasaki, N., 2005. Mitochondrial transcription factor A in the maintenance of mitochondrial DNA: overview of its multiple roles. Ann. N. Y. Acad. Sci. 1042, 101–108.

Kennedy, S.R., Salk, J.J., Schmitt, M.W., Loeb, L.A., 2013. Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. PLoS Genet. 9, e1003794.

Kenney, M.C., Chwa, M., Atilano, S.R., Pavlis, J.M., Falatoonzadeh, P., Ramirez, C., Malik, D., Hsu, T., Woo, G., Soe, K., Nesburn, A.B., Boyer, D.S., Kuppermann, B.D., Jazwinski, S.M., Miceli, M.V., Wallace, D.C., Udar, N., 2013. Mitochondrial DNA variants mediate energy production and expression levels for CFH, C3 and EFEMP1 genes: implications for age-related macular degeneration. PLoS ONE 8, e54339.

Kenney, M.C., Chwa, M., Atilano, S.R., Falatoonzadeh, P., Ramirez, C., Malik, D., Tarek, M., Del Carpio, J.C., Nesburn, A.B., Boyer, D.S., Kuppermann, B.D., Vawter, M.P., Jazwinski, S.M., Miceli, M.V., Wallace, D.C., Udar, N., 2014. Molecular and bioenergetic differences between cells with African versus European inherited mitochondrial DNA haplogroups: implications for population susceptibility to diseases. Biochim. Biophys. Acta 1842, 208–219.

Keogh, M., Chinnery, P.F., 2013. Hereditary mtDNA heteroplasmy: a baseline for aging? Cell Metab. 18, 463–464.

Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G., Kronenberg, F., 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum. Mutat. 32, 25–32.

Kofler, B., Mueller, E.E., Eder, W., Stanger, O., Maier, R., Weger, M., Haas, A., Winker, R., Schmut, O., Paulweber, B., Iglseder, B., Renner, W., Wiesbauer, M., Aigner, I., Santic, D., Zimmermann, F.A., Mayr, J.A., Sperl, W., 2009. Mitochondrial DNA haplogroup T is associated with coronary artery disease and diabetic retinopathy: a case control study. BMC Med. Genet. 10, 35.

Krüger, J., Hinttala, R., Majamaa, K., Remes, A.M., 2010. Mitochondrial DNA haplogroups in early-onset Alzheimer's disease and frontotemporal lobar degeneration. Mol. Neurodegener. 5, 8.

Kujoth, G.C., Hiona, A., Pugh, T.D., Someya, S., Panzer, K., Wohlgemuth, S.E., Hofer, T., Seo, A.Y., Sullivan, R., Jobling, W.A., Morrow, J.D., Van Remmen, H., Sedivy, J.M., Yamasoba, T., Tanokura, M., Weindruch, R., Leeuwenburgh, C., Prolla, T.A., 2005. Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. Science 309, 481–484.

Kujoth, G.C., Bradshaw, P.C., Haroon, S., Prolla, T.A., 2007. The role of mitochondrial DNA mutations in mammalian aging. PLoS Genet. 3, e24.

Lagouge, M., Larsson, N.-G., 2013. The role of mitochondrial DNA mutations and free radicals in disease and ageing. J. Intern. Med. 273, 529–543.

Lakatos, A., Derbeneva, O., Younes, D., Keator, D., Bakken, T., Lvova, M., Brandon, M., Guffanti, G., Reglodi, D., Saykin, A., Weiner, M., Macciardi, F., Schork, N., Wallace, D. C., Potkin, S.G., 2010. Association between mitochondrial DNA variations and Alzheimer's disease in the ADNI cohort. Neurobiol. Aging 31, 1355–1363.

Lam, E.T., Bracci, P.M., Holly, E.A., Chu, C., Poon, A., Wan, E., White, K., Kwok, P.-Y., Pawlikowska, L., Tranah, G.J., 2012. Mitochondrial DNA sequence variation and risk of pancreatic cancer. Cancer Res. 72, 686–695.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Data, G.P., Sam, T., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, M., Schroeder, R., Ko, A., Stoneking, M., 2012. Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. Nucleic Acids Res. 40, e137.

Liou, C.-W., Chen, J.-B., Tiao, M.-M., Weng, S.-W., Huang, T.-L., Chuang, J.-H., Chen, S.-D., Chuang, Y.-C., Lee, W.-C., Lin, T.-K., Wang, P.-W., 2012. Mitochondrial DNA coding and control region variants as genetic risk factors for type 2 diabetes. Diabetes 61, 2642–2651.

Liu, Y., Maxwell, S., Feng, T., Zhu, X., Elston, R.C., Koyutürk, M., Chance, M.R., 2012. Gene, pathway and network frameworks to identify epistatic interactions of single nucleotide polymorphisms derived from GWAS data. BMC Syst. Biol. 6, S15.

Longley, M.J., Graziewicz, M.A., Bienstock, R.J., Copeland, W.C., 2005. Consequences of mutations in human DNA polymerase gamma. Gene 354, 125–131.

Luoma, P., Melberg, A., Rinne, J.O., Kaukonen, J.A., Nupponen, N.N., Chalmers, R.M., Oldfors, P.A., Rautakorpi, I., Peltonen, P.L., Majamaa, P.K., Somer, H., Suomalainen, A., 2004.

Parkinsonism, premature menopause, and mitochondrial DNA polymerase γ mutations: clinical and molecular genetic study. Lancet 364, 875–882.
Man, P.Y.W., Howell, N., Mackey, D.A., Nørby, S., Rosenberg, T., Turnbull, D.M., Chinnery, P. F., 2004. Mitochondrial DNA haplogroup distribution within Leber hereditary optic neuropathy pedigrees. J. Med. Genet. 41, e41.
Mancuso, M., Nardini, M., Micheli, D., Rocchi, A., Nesti, C., Giglioli, N.J., Petrozzi, L., Rossi, C., Ceravolo, R., Bacci, A., Choub, A., Ricci, G., Tognoni, G., Manca, M.L., Siciliano, G., Murri, L., 2007. Lack of association between mtDNA haplogroups and Alzheimer's disease in Tuscany. Neurol. Sci. 28, 142–147.
Manning, A.K., Hivert, M.-F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., et al., 2012. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat. Genet. 44, 659–669.
Marchenko, N.D., Zaika, A., Moll, U.M., 2000. Death signal-induced localization of p53 protein to mitochondria. A potential role in apoptotic signaling. J. Biol. Chem. 275, 16202–16212.
Marchini, J., Donnelly, P., Cardon, L.R., 2005. Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat. Genet. 37, 413–417.
Maruszak, A., Canter, J.A., Styczyńska, M., Zekanowski, C., Barcikowska, M., 2009. Mitochondrial haplogroup H and Alzheimer's disease—is there a connection? Neurobiol. Aging 30, 1749–1755.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K. , Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.
Michikawa, Y., Mazzucchelli, F., Bresolin, N., Scarlato, G., Attardi, G., 1999. Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. Science 286, 774–779.
Mihara, M., Erster, S., Zaika, A., Petrenko, O., Chittenden, T., Pancoska, P., Moll, U.M., 2003. p53 has a direct apoptogenic role at the mitochondria. Mol. Cell 11, 577–590.
Mohlke, K.L., Jackson, A.U., Scott, L.J., Peck, E.C., Suh, Y.D., Chines, P.S., Watanabe, R.M., Buchanan, T.A., Conneely, K.N., Erdos, M.R., Narisu, N., Enloe, S., Valle, T.T., Tuomilehto, J., Bergman, R.N., Boehnke, M., Collins, F.S., 2005. Mitochondrial polymorphisms and susceptibility to type 2 diabetes-related traits in Finns. Hum. Genet. 118, 245–254.
Moilanen, J.S., Finnila, S., Majamaa, K., 2003. Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup J. Mol. Biol. Evol. 20, 2132–2142.
Montiel-Sosa, F., Ruiz-Pesini, E., Enríquez, J.A., Marcuello, A., Díez-Sánchez, C., Montoya, J., Wallace, D.C., López-Pérez, M.J., 2006. Differences of sperm motility in mitochondrial DNA haplogroup U sublineages. Gene 368, 21–27.
Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segrè, A.V., et al., 2012. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. 44, 981–990.
Nardelli, C., Labruna, G., Liguori, R., Mazzaccara, C., Ferrigno, M., Capobianco, V., Pezzuti, M., Castaldo, G., Farinaro, E., Contaldo, F., Buono, P., Sacchetti, L., Pasanisi, F., 2013. Haplogroup T is an obesity risk factor: mitochondrial DNA haplotyping in a morbid obese population from southern Italy. Biomed. Res. Int. 2013, 631082.
Nass, M.M., 1973. Differential methylation of mitochondrial and nuclear DNA in cultured mouse, hamster and virus-transformed hamster cells. In vivo and in vitro methylation. J. Mol. Biol. 80, 155–175.
Niemi, A.-K., Hervonen, A., Hurme, M., Karhunen, P.J., Jylhä, M., Majamaa, K., 2003. Mitochondrial DNA polymorphisms associated with longevity in a Finnish population. Hum. Genet. 112, 29–33.
Niemi, A.-K., Moilanen, J.S., Tanaka, M., Hervonen, A., Hurme, M., Lehtimäki, T., Arai, Y., Hirose, N., Majamaa, K., 2005. A combination of three common inherited mitochondrial DNA polymorphisms promotes longevity in Finnish and Japanese subjects. Eur. J. Hum. Genet. 13, 166–170.
Nunnari, J., Suomalainen, A., 2012. Mitochondria: in sickness and in health. Cell 148, 1145–1159.
Orth, J.D., Thiele, I., Palsson, B., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245–248.
Passarino, G., Rose, G., Bellizzi, D., 2010. Mitochondrial function, mitochondrial DNA and ageing: a reappraisal. Biogerontology 11, 575–588.
Payne, B.A.I., Wilson, I.J., Yu-Wai-Man, P., Coxhead, J., Deehan, D., Horvath, R., Taylor, R.W., Samuels, D.C., Santibanez-Koref, M., Chinnery, P.F., 2013. Universal heteroplasmy of human mitochondrial DNA. Hum. Mol. Genet. 22, 384–390.
Pepe, M.S., Janes, H., Longton, G., Leisenring, W., Newcomb, P., 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am. J. Epidemiol. 159, 882–890.
Perls, T., Terry, D., 2003. Genetics of exceptional longevity. Exp. Gerontol. 38, 725–730.
Picardi, E., Pesole, G., 2012. Mitochondrial genomes gleaned from human whole-exome sequencing. Nat. Methods 9, 523–524.
Pinós, T., Nogales-Gadea, G., Ruiz, J.R., Rodríguez-Romo, G., Santiago-Dorrego, C., Fiuza-Luces, C., Gómez-Gallego, F., Cano-Nieto, A., Garatachea, N., Morán, M., Angel Martín, M., Arenas, J., Andreu, A.L., Lucia, A., 2012. Are mitochondrial haplogroups associated with extreme longevity? A study on a Spanish cohort. Age (Dordr.) 34, 227–233.
Piriyapongsa, J., Ngamphiw, C., Intarapanich, A., Kulawonganunchai, S., Assawamakin, A., Bootchai, C., Shaw, P.J., Tongsima, S., 2012. iLOCi: a SNP interaction prioritization technique for detecting epistasis in genome-wide association studies. BMC Genomics 13, S2.
Pollack, Y., Kasir, J., Shemer, R., Metzger, S., Szyf, M., 1984. Methylation pattern of mouse mitochondrial DNA. Nucleic Acids Res. 12, 4811–4824.
Pyle, A., Foltynie, T., Tiangyou, W., Lambert, C., Keers, S.M., Allcock, L.M., Davison, J., Lewis, S.J., Perry, R.H., Barker, R., Burn, D.J., Chinnery, P.F., 2005. Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. Ann. Neurol. 57, 564–567.
Rackham, O., Shearwood, A.-M.J., Mercer, T.R., Davies, S.M.K., Mattick, J.S., Filipovska, A., 2011. Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. RNA 17, 2085–2093.
Raule, N., Sevini, F., Li, S., Barbieri, A., Tallaro, F., Lomartire, L., Vianello, D., Montesanto, A., Moilanen, J.S., Bezrukov, V., Blanché, H., Hervonen, A., Christensen, K., Deiana, L., Gonos, E.S., Kirkwood, T.B.L., Kristensen, P., Leon, A., Pelicci, P.G., Poulain, M., Rea, I. M., Remacle, J., Robine, J.M., Schreiber, S., Sikora, E., Eline Slagboom, P., Spazzafumo, L., Antonietta Stazi, M., Toussaint, O., Vaupel, J.W., Rose, G., Majamaa, K., Perola, M., Johnson, T.E., Bolund, L., Yang, H., Passarino, G., Franceschi, C., 2013. The co-occurrence of mtDNA mutations on different oxidative phosphorylation subunits, not detected by haplogroup analysis, affects human longevity and is population specific. Aging Cell (Epub ahead of print).
Rebelo, A.P., Williams, S.L., Moraes, C.T., 2009. In vivo methylation of mtDNA reveals the dynamics of protein-mtDNA interactions. Nucleic Acids Res. 37, 6701–6715.
Ridge, P.G., Maxwell, T.J., Corcoran, C.D., Norton, M.C., Tschanz, J.T., O'Brien, E., Kerber, R.A., Cawthon, R.M., Munger, R.G., Kauwe, J.S.K., 2012. Mitochondrial genomic analysis of late onset Alzheimer's disease reveals protective haplogroups H6A1A/H6A1B: the Cache County Study on Memory in Aging. PLoS ONE 7, e45134.
Ristow, M., Zarse, K., 2010. How increased oxidative stress promotes longevity and metabolic health: the concept of mitochondrial hormesis (mitohormesis). Exp. Gerontol. 45, 410–418.
Rose, G., Passarino, G., Carrieri, G., Altomare, K., Greco, V., Bertolini, S., Bonafè, M., Franceschi, C., De Benedictis, G., 2001. Paradoxes in longevity: sequence analysis of mtDNA haplogroup J in centenarians. Eur. J. Hum. Genet. 9, 701–707.
Rose, G., Passarino, G., Franceschi, C., De Benedictis, G., 2002. The variability of the mitochondrial genome in human aging: A key for life and death? Int. J. Biochem. Cell Biol. 34, 1449–1460.
Rose, G., Passarino, G., Scornaienchi, V., Romeo, G., Dato, S., Bellizzi, D., Mari, V., Feraco, E., Maletta, R., Bruni, A., Franceschi, C., De Benedictis, G., 2007. The mitochondrial DNA control region shows genetically correlated levels of heteroplasmy in leukocytes of centenarians and their offspring. BMC Genomics 8, 293.
Rose, G., Romeo, G., Dato, S., Crocco, P., Bruni, A.C., Hervonen, A., Majamaa, K., Sevini, F., Franceschi, C., Passarino, G., 2010. Somatic point mutations in mtDNA control region are influenced by genetic background and associated with healthy aging: a GEHA study. PLoS ONE 5, e13395.
Ross, O.A., McCormack, R., Curran, M.D., Alistair Duguid, R., Barnett, Y.A., Maeve Rea, I., Middleton, D., 2001. Mitochondrial DNA polymorphism: its role in longevity of the Irish population. Exp. Gerontol. 36, 1161–1178.
Ross, J.M., Stewart, J.B., Hagström, E., Brené, S., Mourier, A., Coppotelli, G., Freyer, C., Lagouge, M., Hoffer, B.J., Olson, L., Larsson, N.-G., 2013. Germline mitochondrial DNA mutations aggravate ageing and can impair brain development. Nature 501, 412–415.
Rubino, F., Piredda, R., Calabrese, F.M., Simone, D., Lang, M., Calabrese, C., Petruzzella, V., Tommaseo-Ponzetta, M., Gasparre, G., Attimonelli, M., 2012. HmtDB, a genomic resource for mitochondrion-based human variability studies. Nucleic Acids Res. 40, D1150–D1159.
Ruiz-Pesini, E., Lapeña, A.C., Díez-Sánchez, C., Pérez-Martos, A., Montoya, J., Alvarez, E., Díaz, M., Urriés, A., Montoro, L., López-Pérez, M.J., Enríquez, J.A., 2000. Human mtDNA haplogroups associated with high or reduced spermatozoa motility. Am. J. Hum. Genet. 67, 682–696.
Ruiz-Pesini, E., Lott, M.T., Procaccio, V., Poole, J.C., Brandon, M.C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P., Wallace, D.C., 2007. An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Res. 35, D823–D828.
Salvioli, S., Capri, M., Santoro, A., Raule, N., Sevini, F., Lukas, S., Lanzarini, C., Monti, D., Passarino, G., Rose, G., De Benedictis, G., Franceschi, C., 2008. The impact of mitochondrial DNA on human lifespan: a view from studies on centenarians. Biotechnol. J. 3, 740–749.
Santoro, A., Salvioli, S., Raule, N., Capri, M., Sevini, F., Valensin, S., Monti, D., Bellizzi, D., Passarino, G., Rose, G., De Benedictis, G., Franceschi, C., 2006. Mitochondrial DNA involvement in human longevity. Biochim Biophys Acta. Sep-Oct;1757 (9-10), 1388-99.
Santoro, A., Balbi, V., Balducci, E., Pirazzini, C., Rosini, F., Tavano, F., Achilli, A., Siviero, P., Minicuci, N., Bellavista, E., Mishto, M., Salvioli, S., Marchegiani, F., Cardelli, M., Olivieri, F., Nacmias, B., Chiamenti, A.M., Benussi, L., Ghidoni, R., Rose, G., Gabelli, C., Binetti, G., Sorbi, S., Crepaldi, G., Passarino, G., Torroni, A., Franceschi, C., 2010. Evidence for sub-haplogroup H5 of mitochondrial DNA as a risk factor for late onset Alzheimer's disease. PLoS ONE 5, e12037.
Saxena, R., de Bakker, P.I.W., Singer, K., Mootha, V., Burtt, N., Hirschhorn, J.N., Gaudet, D., Isomaa, B., Daly, M.J., Groop, L., Ardlie, K.G., Altshuler, D., 2006. Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. Am. J. Hum. Genet. 79, 54–61.
Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., Sirotkin, K., 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311.
Shmookler Reis, R.J., Goldstein, S., 1983a. Mitochondrial DNA in mortal and immortal human cells. Genome number, integrity, and methylation. J. Biol. Chem. 258, 9078–9085.
Shmookler Reis, R.J., Goldstein, S., 1983b. Mitochondrial DNA in mortal and immortal human cells. Genome number, integrity, and methylation. J. Biol. Chem. 258, 9078–9085.
Shock, L.S., Thakkar, P.V., Peterson, E.J., Moran, R.G., Taylor, S.M., 2011. DNA methyltransferase 1, cytosine methylation, and cytosine hydroxymethylation in mammalian mitochondria. Proc. Natl. Acad. Sci. U. S. A. 108, 3630–3635.
Îimková, H., 1998. Methylation of mitochondrial DNA in carrot (*Daucus carota* L.). Plant Cell Rep. 17, 220–224.
Skytthe, A., Valensin, S., Jeune, B., Cevenini, E., Balard, F., et al., 2011. Design, recruitment, logistics, and data management of the GEHA (Genetics of Healthy Ageing) project. Exp. Gerontol. 46, 934–945.

Sondheimer, N., Glatz, C.E., Tirone, J.E., Deardorff, M.A., Krieger, A.M., Hakonarson, H., 2011. Neutral mitochondrial heteroplasmy and the influence of aging. Hum. Mol. Genet. 20, 1653–1659.

Sosa, M.X., Sivakumar, I.K.A., Maragh, S., Veeramachaneni, V., Hariharan, R., Parulekar, M., Fredrikson, K.M., Harkins, T.T., Lin, J., Feldman, A.B., Tata, P., Ehret, G.B., Chakravarti, A., 2012. Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency. PLoS Comput. Biol. 8, e1002737.

Sosa-Ortiz, A.L., Acosta-Castillo, I., Prince, M.J., 2012. Epidemiology of dementias and Alzheimer's disease. Arch. Med. Res. 43, 600–608.

Sripada, L., Tomar, D., Prajapati, P., Singh, R., Singh, A.K., Singh, R., 2012. Systematic analysis of small RNAs associated with human mitochondria by deep sequencing: detailed analysis of mitochondrial associated miRNA. PLoS ONE 7, e44873.

Strauss, K.A., DuBiner, L., Simon, M., Zaragoza, M., Sengupta, P.P., Li, P., Narula, N., Dreike, S., Platt, J., Procaccio, V., Ortiz-González, X.R., Puffenberger, E.G., Kelley, R.I., Morton, D. H., Narula, J., Wallace, D.C., 2013. Severity of cardiomyopathy associated with adenine nucleotide translocator-1 deficiency correlates with mtDNA haplogroup. Proc. Natl. Acad. Sci. U. S. A. 110, 3453–3458.

Tan, Q., Zhao, J.H., Zhang, D., Kruse, T.A., Christensen, K., 2008. Power for genetic association study of human longevity using the case–control design. Am. J. Epidemiol. 168, 890–896.

Tanaka, M., Gong, J.S., Zhang, J., Yoneda, M., Yagi, K., 1998. Mitochondrial genotype associated with longevity. Lancet 351, 185–186.

Tanaka, M., Gong, J.S., Zhang, J., Yamada, Y., Borgeld, H.J., Yagi, K., 2000. Mitochondrial genotype associated with longevity and its inhibitory effect on mutagenesis. Mech. Ageing Dev. 116, 65–76.

Tapia, P.C., 2006. Sublethal mitochondrial stress with an attendant stoichiometric augmentation of reactive oxygen species may precipitate many of the beneficial alterations in cellular physiology produced by caloric restriction, intermittent fasting, exercise and dietary p. Med. Hypotheses 66, 832–843.

Torroni, A., Petrozzi, M., D'Urbano, L., Sellitto, D., Zeviani, M., Carrara, F., Carducci, C., Leuzzi, V., Carelli, V., Barboni, P., De Negri, A., Scozzari, R., 1997. Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. Am. J. Hum. Genet. 60, 1107–1121.

Tranah, G.J., 2011. Mitochondrial-nuclear epistasis: implications for human aging and longevity. Ageing Res. Rev. 10, 238–252.

Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J.N., Rovio, A.T., Bruder, C.E., Bohlooly-Y, M., Gidlöf, S., Oldfors, A., Wibom, R., Törnell, J., Jacobs, H.T., Larsson, N.-G., 2004. Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature 429, 417–423.

Udar, N., Atilano, S.R., Memarzadeh, M., Boyer, D.S., Chwa, M., Lu, S., Maguen, B., Langberg, J., Coskun, P., Wallace, D.C., Nesburn, A.B., Khatibi, N., Hertzog, D., Le, K., Hwang, D., Kenney, M.C., 2009. Mitochondrial DNA haplogroups associated with age-related macular degeneration. Invest. Ophthalmol. Vis. Sci. 50, 2966–2974.

Vadrot, N., Ghanem, S., Braut, F., Gavrilescu, L., Pilard, N., Mansouri, A., Moreau, R., Reyl-Desmars, F., 2012. Mitochondrial DNA maintenance is regulated in human hepatoma cells by glycogen synthase kinase 3β and p53 in response to tumor necrosis factor α. PLoS ONE 7, e40879.

Van der Walt, J.M., Nicodemus, K.K., Martin, E.R., Scott, W.K., Nance, M.A., Watts, R.L., Hubble, J.P., Haines, J.L., Koller, W.C., Lyons, K., Pahwa, R., Stern, M.B., Colcher, A., Hiner, B.C., Jankovic, J., Ondo, W.G., Allen, F.H., Goetz, C.G., Small, G.W., Mastaglia, F., Stajich, J.M., McLaurin, A.C., Middleton, L.T., Scott, B.L., Schmechel, D.E., Pericak-Vance, M.A., Vance, J.M., 2003. Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. Am. J. Hum. Genet. 72, 804–811.

Van Der Walt, J.M., Dementieva, Y.A., Martin, E.R., Scott, W.K., Nicodemus, K.K., Kroner, C. C., Welsh-Bohmer, K.A., Saunders, A.M., Roses, A.D., Small, G.W., Schmechel, D.E., Murali Doraiswamy, P., Gilbert, J.R., Haines, J.L., Vance, J.M., Pericak-Vance, M.A., 2004. Analysis of European mitochondrial haplogroups with Alzheimer disease risk. Neurosci. Lett. 365, 28–32.

Van der Walt, J.M., Scott, W.K., Slifer, S., Gaskell, P.C., Martin, E.R., Welsh-Bohmer, K., Creason, M., Crunk, A., Fuzzell, D., McFarland, L., Kroner, C.C., Jackson, C.E., Haines, J. L., Pericak-Vance, M.A., 2005. Maternal lineages and Alzheimer disease risk in the Old Order Amish. Hum. Genet. 118, 115–122.

Van Goethem, G., Martin, J.-J., Van Broeckhoven, C., 2002. Progressive external ophthalmoplegia and multiple mitochondrial DNA deletions. Acta Neurol. Belg. 102, 39–42.

Vermulst, M., Bielas, J.H., Kujoth, G.C., Ladiges, W.C., Rabinovitch, P.S., Prolla, T.A., Loeb, L.A. , 2007. Mitochondrial point mutations do not limit the natural lifespan of mice. Nat. Genet. 39, 540–543.

Vermulst, M., Wanagat, J., Kujoth, G.C., Bielas, J.H., Rabinovitch, P.S., Prolla, T.A., Loeb, L.A., 2008. DNA deletions and clonal mutations drive premature aging in mitochondrial mutator mice. Nat. Genet. 40, 392–394.

Vianello, D., Sevini, F., Castellani, G., Lomartire, L., Capri, M., Franceschi, C., 2013. HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. Hum. Mutat. 34, 1189–1194.

Wai, T., Teoli, D., Shoubridge, E.A., 2008. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. Nat. Genet. 40, 1484–1488.

Wallace, D.C., 2005. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. Annu. Rev. Genet. 39, 359–407.

Wallace, D.C., 2010. Review article mitochondrial DNA mutations in disease and aging. Environ. Mol. Mutagen. 450, 440–450.

Wallace, D.C., 2013. Bioenergetics in human evolution and disease: implications for the origins of biological complexity and the missing genetic variation of common diseases. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368, 20120267.

Wallace, D.C., Fan, W., 2010. Energetics, epigenetics, mitochondrial genetics. Mitochondrion 10, 12–31.

Wallace, D.C., Fan, W., Procaccio, V., 2010. Mitochondrial energetics and therapeutics. Annu. Rev. Pathol. 5, 297–348.

Wang, Y., Michikawa, Y., Mallidis, C., Bai, Y., Woodhouse, L., Yarasheski, K.E., Miller, C.A., Askanas, V., Engel, W.K., Bhasin, S., Attardi, G., 2001. Muscle-specific mutations accumulate with aging in critical human mtDNA control sites for replication. Proc. Natl. Acad. Sci. U. S. A. 98, 4022–4027.

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164.

Weigl, S., Paradiso, A., Tommasi, S., 2013. Mitochondria and familial predisposition to breast cancer. Curr. Genomics 14, 195–203.

Wong, T.S., Rajagopalan, S., Freund, S.M., Rutherford, T.J., Andreeva, A., Townsley, F.M., Petrovich, M., Fersht, A.R., 2009. Biophysical characterizations of human mitochondrial transcription factor A and its binding to tumor suppressor p53. Nucleic Acids Res. 37, 6765–6783.

Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., Lin, X., 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. 89, 82–93.

Yao, Y.-G., Salas, A., Logan, I., Bandelt, H.-J., 2009. mtDNA data mining in GenBank needs surveying. Am. J. Hum. Genet. 85, 929–933.

Yoshida, Y., Izumi, H., Torigoe, T., Ishiguchi, H., Itoh, H., Kang, D., Kohno, K., 2003. P53 physically interacts with mitochondrial transcription factor A and differentially regulates binding to damaged DNA. Cancer Res. 63, 3729–3734.

Zhang, Y., Liu, J.S., 2007. Bayesian inference of epistatic interactions in case–control studies. Nat. Genet. 39, 1167–1173.

Zhang, J., Asin-Cayuela, J., Fish, J., Michikawa, Y., Bonafe, M., Olivieri, F., Passarino, G., De Benedictis, G., Franceschi, C., Attardi, G., 2003. Strikingly higher frequency in centenarians and twins of mtDNA mutation causing remodeling of replication origin in leukocytes. Proc. Natl. Acad. Sci. U. S. A. 100, 1116–1121.

Zhong, Y., Hu, Y.-J., Chen, B., Peng, W., Sun, Y., Yang, Y., Zhao, X.-Y., Fan, G., Huang, X., Kong, W.-J., 2011. Mitochondrial transcription factor A overexpression and base excision repair deficiency in the inner ear of rats with D-galactose-induced aging. FEBS J. 278, 2500–2510.

REFERENCE

Adler, C. J., W. Haak, D. Donlon, and A. Cooper. 2011. "Survival and Recovery of DNA from Ancient Teeth and Bones." *Journal of Archaeological Science* 38 (5): 956–64. doi:10.1016/j.jas.2010.11.010.

Agyemang, Charles, Juliet Addo, Raj Bhopal, Ama de Graft Aikins, and Karien Stronks. 2009. "Cardiovascular Disease, Diabetes and Established Risk Factors among Populations of Sub-Saharan African Descent in Europe: A Literature Review." *Globalization and Health* 5 (August): 7. doi:10.1186/1744-8603-5-7.

Alexander, David H., John Novembre, and Kenneth Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19 (9): 1655–64. doi:10.1101/gr.094052.109.

Alisch, Reid S., Benjamin G. Barwick, Pankaj Chopra, Leila K. Myrick, Glen A. Satten, Karen N. Conneely, and Stephen T. Warren. 2012. "Age-Associated DNA Methylation in Pediatric Populations." *Genome Research* 22 (4): 623–32. doi:10.1101/gr.125187.111.

AlQahtani, S. J., M. P. Hector, and H. M. Liversidge. 2014. "Accuracy of Dental Age Estimation Charts: Schour and Massler, Ubelaker and the London Atlas." *American Journal of Physical Anthropology* 154 (1): 70–78. doi:10.1002/ajpa.22473.

Anderson, Alexander R. A., Alissa M. Weaver, Peter T. Cummings, and Vito Quaranta. 2006. "Tumor Morphology and Phenotypic Evolution Driven by Selective Pressure from the Microenvironment." *Cell* 127 (5): 905–15. doi:10.1016/j.cell.2006.09.042.

Antequera, F. 2003. "Structure, Function and Evolution of CpG Island Promoters." *Cellular and Molecular Life Sciences: CMLS* 60 (8): 1647–58. doi:10.1007/s00018-003-3088-6.

Bacalini, Maria Giulia, Alessio Boattini, Davide Gentilini, Enrico Giampieri, Chiara Pirazzini, Cristina Giuliani, Elisa Fontanesi, et al. 2015. "A Meta-Analysis on Age-Associated Change in Blood DNA Methylation: Results from an Original Analysis Pipeline for Infinium 450k Data." *Aging*, January.

Bacalini, Maria Giulia, Simonetta Friso, Fabiola Olivieri, Chiara Pirazzini, Cristina Giuliani, Miriam Capri, Aurelia Santoro, Claudio Franceschi, and Paolo Garagnani. 2014. "Present and Future of Anti-Ageing Epigenetic Diets." *Mechanisms of Ageing and Development*. doi:10.1016/j.mad.2013.12.006.

Bell, Jordana T., Pei-Chien Tsai, Tsun-Po Yang, Ruth Pidsley, James Nisbet, Daniel Glass, Massimo Mangino, et al. 2012. "Epigenome-Wide Scans Identify Differentially Methylated Regions for Age and Age-Related Phenotypes in a Healthy Ageing Population." *PLoS Genetics* 8 (4): e1002629. doi:10.1371/journal.pgen.1002629.

Berdyshev, G. D., G. K. Korotaev, G. V. Boiarskikh, and B. F. Vaniushin. 1967. "[Nucleotide composition of DNA and RNA from somatic tissues of humpback and its changes during spawning]." *Biokhimiia (Moscow, Russia)* 32 (5): 988–93.

Bibikova, Marina, Bret Barnes, Chan Tsan, Vincent Ho, Brandy Klotzle, Jennie M. Le, David Delano, et al. 2011. "High Density DNA Methylation Array with Single CpG Site Resolution." *Genomics* 98 (4): 288–95. doi:10.1016/j.ygeno.2011.07.007.

Bierne, Hélène, Mélanie Hamon, and Pascale Cossart. 2012. "Epigenetics and Bacterial Infections." *Cold Spring Harbor Perspectives in Medicine* 2 (12): a010272. doi:10.1101/cshperspect.a010272.

Boattini, Alessio, Antonella Lisa, Ornella Fiorani, Gianna Zei, Davide Pettener, and Franz Manni. 2012. "General Method to Unravel Ancient Population Structures through Surnames, Final Validation on Italian Data." *Human Biology* 84 (3). http://digitalcommons.wayne.edu/humbiol/vol84/iss3/2.

Boattini, Alessio, Begoña Martinez-Cruz, Stefania Sarno, Christine Harmant, Antonella Useli, Paula Sanz, Daniele Yang-Yao, et al. 2013. "Uniparental Markers in Italy Reveal a Sex-Biased Genetic Structure and Different Historical Strata." *PloS One* 8 (5): e65441. doi:10.1371/journal.pone.0065441.

Bocklandt, Sven, Wen Lin, Mary E. Sehl, Francisco J. Sánchez, Janet S. Sinsheimer, Steve Horvath, and Eric Vilain. 2011. "Epigenetic Predictor of Age." *PloS One* 6 (6): e14821. doi:10.1371/journal.pone.0014821.

Borza, Corina M., and Ambra Pozzi. 2014. "Discoidin Domain Receptors in Disease." *Matrix Biology: Journal of the International Society for Matrix Biology* 34 (February): 185–92. doi:10.1016/j.matbio.2013.12.002.

Breton, Carrie V., and Amy N. Marutani. 2014. "Air Pollution and Epigenetics: Recent Findings." *Current Environmental Health Reports* 1 (1): 35–45. doi:10.1007/s40572-013-0001-9.

Brooks, S., and J. M. Suchey. 1990. "Skeletal Age Determination Based on the Os Pubis: A Comparison of the Acsádi-Nemeskéri and Suchey-Brooks Methods." *Human Evolution* 5 (3): 227–38. doi:10.1007/BF02437238.

Campanella, Gianluca, Silvia Polidoro, Cornelia Di Gaetano, Giovanni Fiorito, Simonetta Guarrera, Vittorio Krogh, Domenico Palli, et al. 2014. "Epigenetic Signatures of Internal Migration in Italy." *International Journal of Epidemiology*, October, dyu198. doi:10.1093/ije/dyu198.

Chan, A. O.-O., S.-K. Lam, B. C.-Y. Wong, W.-M. Wong, M.-F. Yuen, Y.-H. Yeung, W.-M. Hui, A. Rashid, and Y.-L. Kwong. 2003. "Promoter Methylation of E-Cadherin Gene in Gastric Mucosa Associated with Helicobacter Pylori Infection and in Gastric Cancer." *Gut* 52 (4): 502–6. doi:10.1136/gut.52.4.502.

Chen, Xin-Ya, Xiu-Ting Gu, Hexige Saiyin, Bo Wan, Yu-Jing Zhang, Jing Li, Ying-Li Wang, et al. 2012. "Brain-Selective Kinase 2 (BRSK2) Phosphorylation on PCTAIRE1 Negatively Regulates Glucose-Stimulated Insulin Secretion in Pancreatic B-Cells." *The Journal of Biological Chemistry* 287 (36): 30368–75. doi:10.1074/jbc.M112.375618.

Christensen, Brock C., E. Andres Houseman, Carmen J. Marsit, Shichun Zheng, Margaret R. Wrensch, Joseph L. Wiemels, Heather H. Nelson, et al. 2009. "Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context." *PLoS Genetics* 5 (8): e1000602. doi:10.1371/journal.pgen.1000602.

Cockerham, C. C., and B. S. Weir. 1986. "Estimation of Inbreeding Parameters in Stratified Populations." *Annals of Human Genetics* 50 (Pt 3): 271–81.

Consortium, The 1000 Genomes Project. 2010. "A Map of Human Genome Variation from Population-Scale Sequencing." *Nature* 467 (7319): 1061–73. doi:10.1038/nature09534.

Dahly, D. L., P. Gordon-Larsen, M. Emch, J. Borja, and L. S. Adair. 2013. "The Spatial Distribution of Overweight and Obesity among a Birth Cohort of Young Adult Filipinos (Cebu Philippines, 2005): An Application of the Kulldorff Spatial Scan Statistic." *Nutrition & Diabetes* 3: e80. doi:10.1038/nutd.2013.21.

Dedeurwaerder, Sarah, Matthieu Defrance, Emilie Calonne, Hélène Denis, Christos Sotiriou, and François Fuks. 2011. "Evaluation of the Infinium Methylation 450K Technology." *Epigenomics* 3 (6): 771–84. doi:10.2217/epi.11.105.

Doi, Akiko, In-Hyun Park, Bo Wen, Peter Murakami, Martin J. Aryee, Rafael Irizarry, Brian Herb, et al. 2009. "Differential Methylation of Tissue- and Cancer-Specific CpG Island Shores Distinguishes Human Induced Pluripotent Stem Cells, Embryonic Stem Cells and Fibroblasts." *Nature Genetics* 41 (12): 1350–53. doi:10.1038/ng.471.

Du, Pan, Warren A. Kibbe, and Simon M. Lin. 2008. "Lumi: A Pipeline for Processing Illumina Microarray." *Bioinformatics (Oxford, England)* 24 (13): 1547–48. doi:10.1093/bioinformatics/btn224.

Du, Pan, Xiao Zhang, Chiang-Ching Huang, Nadereh Jafari, Warren A Kibbe, Lifang Hou, and Simon M Lin. 2010. "Comparison of Beta-Value and M-Value Methods for Quantifying Methylation Levels by Microarray Analysis." *BMC Bioinformatics* 11 (November): 587. doi:10.1186/1471-2105-11-587.

Eden, Eran, Roy Navon, Israel Steinfeld, Doron Lipson, and Zohar Yakhini. 2009. "GOrilla: A Tool for Discovery and Visualization of Enriched GO Terms in Ranked Gene Lists." *BMC Bioinformatics* 10: 48. doi:10.1186/1471-2105-10-48.

Elango, Navin, and Soojin V Yi. 2008. "DNA Methylation and Structural and Functional Bimodality of Vertebrate Promoters." *Molecular Biology and Evolution* 25 (8): 1602–8. doi:10.1093/molbev/msn110.

Emery, Edward C., Gareth T. Young, and Peter A. McNaughton. 2012. "HCN2 Ion Channels: An Emerging Role as the Pacemakers of Pain." *Trends in Pharmacological Sciences* 33 (8): 456–63. doi:10.1016/j.tips.2012.04.004.

Feinberg, Andrew P., and Rafael A. Irizarry. 2009. "Stochastic Epigenetic Variation as a Driving Force of Development, Evolutionary Adaptation, and Disease." *Proceedings of the National Academy of Sciences*, December, 200906183. doi:10.1073/pnas.0906183107.

Feng, Suhua, Shawn J. Cokus, Xiaoyu Zhang, Pao-Yang Chen, Magnolia Bostick, Mary G. Goll, Jonathan Hetzel, et al. 2010. "Conservation and Divergence of Methylation Patterning in Plants and Animals." *Proceedings of the National Academy of Sciences of the United States of America* 107 (19): 8689–94. doi:10.1073/pnas.1002720107.

Franklin, Daniel. 2010. "Forensic Age Estimation in Human Skeletal Remains: Current Concepts and Future Directions." *Legal Medicine (Tokyo, Japan)* 12 (1): 1–7. doi:10.1016/j.legalmed.2009.09.001.

Fumagalli, Matteo, Manuela Sironi, Uberto Pozzoli, Anna Ferrer-Admettla, Linda Pattini, and Rasmus Nielsen. 2011. "Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution." *PLoS Genet* 7 (11): e1002355. doi:10.1371/journal.pgen.1002355.

Galloway, James N., William H. Schlesinger, Hiram Levy, Anthony Michaels, and Jerald L. Schnoor. 1995. "Nitrogen Fixation: Anthropogenic Enhancement-Environmental Response." *Global Biogeochemical Cycles* 9 (2): 235–52. doi:10.1029/95GB00158.

Galloway, James N., Alan R. Townsend, Jan Willem Erisman, Mateete Bekunda, Zucong Cai, John R. Freney, Luiz A. Martinelli, Sybil P. Seitzinger, and Mark A. Sutton. 2008. "Transformation of the Nitrogen Cycle: Recent Trends, Questions, and Potential Solutions." *Science* 320 (5878): 889–92. doi:10.1126/science.1136674.

Garagnani, Paolo, Maria G Bacalini, Chiara Pirazzini, Davide Gori, Cristina Giuliani, Daniela Mari, Anna M Di Blasio, et al. 2012. "Methylation of ELOVL2 Gene as a New Epigenetic Marker of Age." *Aging Cell* 11 (6): 1132–34. doi:10.1111/acel.12005.

Giuliani, Cristina, Maria Giulia Bacalini, Marco Sazzini, Chiara Pirazzini, Claudio Franceschi, Paolo Garagnani, and Donata Luiselli. 2015. "The Epigenetic Side of Human Adaptation: Hypotheses, Evidences and Theories." *Annals of Human Biology* 42 (1): 1–9. doi:10.3109/03014460.2014.961960.

Gluckman, Peter D, Mark A Hanson, and Hamish G Spencer. 2005. "Predictive Adaptive Responses and Human Evolution." *Trends in Ecology & Evolution* 20 (10): 527–33. doi:10.1016/j.tree.2005.08.001.

Goldberg, Michel, and Anthony J. Smith. 2004. "CELLS AND EXTRACELLULAR MATRICES OF DENTIN AND PULP: A BIOLOGICAL BASIS FOR REPAIR AND TISSUE ENGINEERING." *Critical Reviews in Oral Biology and Medicine: An Official Publication of the American Association of Oral Biologists* 15 (1): 13–27.

Guintivano, Jerry, Martin J. Aryee, and Zachary A. Kaminsky. 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: Official Journal of the DNA Methylation Society* 8 (3): 290–302. doi:10.4161/epi.23924.

Hannum, Gregory, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, SriniVas Sadda, Brandy Klotzle, et al. 2013. "Genome-Wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates." *Molecular Cell* 49 (2): 359–67. doi:10.1016/j.molcel.2012.10.016.

Heijmans, Bastiaan T., Dennis Kremer, Elmar W. Tobi, Dorret I. Boomsma, and P. Eline Slagboom. 2007. "Heritable rather than Age-Related Environmental and Stochastic Factors Dominate Variation in DNA Methylation of the Human IGF2/H19 Locus." *Human Molecular Genetics* 16 (5): 547–54. doi:10.1093/hmg/ddm010.

Henderson, Ian R., and Steven E. Jacobsen. 2007. "Epigenetic Inheritance in Plants." *Nature* 447 (7143): 418–24. doi:10.1038/nature05917.

Hernando-Herraez, Irene, Javier Prado-Martinez, Paras Garg, Marcos Fernandez-Callejo, Holger Heyn, Christina Hvilsom, Arcadi Navarro, Manel Esteller, Andrew J. Sharp, and Tomas Marques-Bonet. 2013. "Dynamics of DNA Methylation in Recent Human and Great Ape Evolution." *PLoS Genetics* 9 (9): e1003763. doi:10.1371/journal.pgen.1003763.

Heyn, Holger, Sebastian Moran, Irene Hernando-Herraez, Sergi Sayols, Antonio Gomez, Juan Sandoval, Dave Monk, et al. 2013. "DNA Methylation Contributes to Natural Human Variation." *Genome Research* 23 (9): 1363–72. doi:10.1101/gr.154187.112.

Higgins, Denice, and Jeremy J. Austin. 2013. "Teeth as a Source of DNA for Forensic Identification of Human Remains: A Review." *Science & Justice: Journal of the Forensic Science Society* 53 (4): 433–41. doi:10.1016/j.scijus.2013.06.001.

Higgins, Denice, John Kaidonis, Grant Townsend, Toby Hughes, and Jeremy J Austin. 2013. "Targeted Sampling of Cementum for Recovery of Nuclear DNA from Human Teeth and the Impact of Common Decontamination

Measures." *Investigative Genetics* 4 (October): 18. doi:10.1186/2041-2223-4-18.

Hodges, Emily, Antoine Molaro, Camila O. Dos Santos, Pramod Thekkat, Qiang Song, Philip J. Uren, Jin Park, et al. 2011. "Directional DNA Methylation Changes and Complex Intermediate States Accompany Lineage Specificity in the Adult Hematopoietic Compartment." *Molecular Cell* 44 (1): 17–28. doi:10.1016/j.molcel.2011.08.026.

Holliday, Robin. 2006. "Epigenetics: A Historical Overview." *Epigenetics* 1 (2): 76–80. doi:10.4161/epi.1.2.2762.

Horvath, Steve. 2013. "DNA Methylation Age of Human Tissues and Cell Types." *Genome Biology* 14 (10): R115. doi:10.1186/gb-2013-14-10-r115.

Hossain, Parvez, Bisher Kawar, and Meguid El Nahas. 2007. "Obesity and Diabetes in the Developing World — A Growing Challenge." *New England Journal of Medicine* 356 (3): 213–15. doi:10.1056/NEJMp068177.

Houseman, Eugene A., William P. Accomando, Devin C. Koestler, Brock C. Christensen, Carmen J. Marsit, Heather H. Nelson, John K. Wiencke, and Karl T. Kelsey. 2012. "DNA Methylation Arrays as Surrogate Measures of Cell Mixture Distribution." *BMC Bioinformatics* 13 (1): 86. doi:10.1186/1471-2105-13-86.

Hubisz, Melissa J, and Katherine S Pollard. 2014. "Exploring the Genesis and Functions of Human Accelerated Regions Sheds Light on Their Role in Human Evolution." *Current Opinion in Genetics & Development*, Genetics of human evolution, 29: 15–21. doi:10.1016/j.gde.2014.07.005.

Irizarry, Rafael A., Christine Ladd-Acosta, Bo Wen, Zhijin Wu, Carolina Montano, Patrick Onyango, Hengmi Cui, et al. 2009. "The Human Colon Cancer Methylome Shows Similar Hypo- and Hypermethylation at Conserved Tissue-Specific CpG Island Shores." *Nature Genetics* 41 (2): 178–86. doi:10.1038/ng.298.

Jablonka, Eva, and Marion J. Lamb. 2005. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. MIT Press.

Jaffe, Andrew E., and Rafael A. Irizarry. 2014. "Accounting for Cellular Heterogeneity Is Critical in Epigenome-Wide Association Studies." *Genome Biology* 15 (2): R31. doi:10.1186/gb-2014-15-2-r31.

Jandt, Fred E. 2004. *Intercultural Communication: A Global Reader*. SAGE Publications.

———. 2007. *An Introduction to Intercultural Communication: Identities in a Global Community*. SAGE Publications.

Jombart, Thibaut, and Ismaïl Ahmed. 2011. "Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data." *Bioinformatics* 27 (21): 3070–71. doi:10.1093/bioinformatics/btr521.

Jombart, Thibaut, Sébastien Devillard, and François Balloux. 2010. "Discriminant Analysis of Principal Components: A New Method for the

Analysis of Genetically Structured Populations." *BMC Genetics* 11 (1): 94. doi:10.1186/1471-2156-11-94.

Jones, P. A., and D. Takai. 2001. "The Role of DNA Methylation in Mammalian Epigenetics." *Science (New York, N.Y.)* 293 (5532): 1068–70. doi:10.1126/science.1063852.

Jones, Peter A. 2012. "Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and beyond." *Nature Reviews. Genetics* 13 (7): 484–92. doi:10.1038/nrg3230.

Jones, Peter A., and Stephen B. Baylin. 2002. "The Fundamental Role of Epigenetic Events in Cancer." *Nature Reviews. Genetics* 3 (6): 415–28. doi:10.1038/nrg816.

Klironomos, Filippos D, Johannes Berg, and Sinéad Collins. 2013. "How Epigenetic Mutations Can Affect Genetic Evolution: Model and Mechanism." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 35 (6): 571–78. doi:10.1002/bies.201200169.

Koch, Carmen M., and Wolfgang Wagner. 2011. "Epigenetic-Aging-Signature to Determine Age in Different Tissues." *Aging (Albany NY)* 3 (10): 1018–27.

Lander, E S, L M Linton, B Birren, C Nusbaum, M C Zody, J Baldwin, K Devon, et al. 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature* 409 (6822): 860–921. doi:10.1038/35057062.

Li, Chengtao, Suhua Zhang, Tingzhi Que, Li Li, and Shumin Zhao. 2011. "Identical but Not the Same: The Value of DNA Methylation Profiling in Forensic Discrimination within Monozygotic Twins." *Forensic Science International: Genetics Supplement Series*, Progress in Forensic Genetics 14 Proceedings of the 24th International ISFG Congress, 3 (1): e337–38. doi:10.1016/j.fsigss.2011.09.031.

Lister, Ryan, Mattia Pelizzola, Robert H. Dowen, R. David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R. Nery, et al. 2009. "Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences." *Nature* 462 (7271): 315–22. doi:10.1038/nature08514.

Lokk, Kaie, Vijayachitra Modhukur, Balaji Rajashekar, Kaspar Märtens, Reedik Mägi, Raivo Kolde, Marina Koltšina, et al. 2014. "DNA Methylome Profiling of Human Tissues Identifies Global and Tissue-Specific Methylation Patterns." *Genome Biology* 15 (4): r54. doi:10.1186/gb-2014-15-4-r54.

Marr, Alexandra K., Julia L. MacIsaac, Ruiwei Jiang, Adriana M. Airo, Michael S. Kobor, and W. Robert McMaster. 2014. "Leishmania Donovani Infection Causes Distinct Epigenetic DNA Methylation Changes in Host Macrophages." *PLoS Pathog* 10 (10): e1004419. doi:10.1371/journal.ppat.1004419.

Martin, David I. K., Meromit Singer, Joseph Dhahbi, Guanxiong Mao, Lu Zhang, Gary P. Schroth, Lior Pachter, and Dario Boffelli. 2011.

"Phyloepigenomic Comparison of Great Apes Reveals a Correlation between Somatic and Germline Methylation States." *Genome Research* 21 (12): 2049–57. doi:10.1101/gr.122721.111.

Meissner, Christoph, and Stefanie Ritz-Timme. 2010. "Molecular Pathology and Age Estimation." *Forensic Science International* 203 (1-3): 34–43. doi:10.1016/j.forsciint.2010.07.010.

Mendizabal, I., T. E. Keller, J. Zeng, and Soojin V. Yi. 2014. "Epigenetics and Evolution." *Integrative and Comparative Biology* 54 (1): 31–42. doi:10.1093/icb/icu040.

Molaro, Antoine, Emily Hodges, Fang Fang, Qiang Song, W. Richard McCombie, Gregory J. Hannon, and Andrew D. Smith. 2011. "Sperm Methylation Profiles Reveal Features of Epigenetic Inheritance and Evolution in Primates." *Cell* 146 (6): 1029–41. doi:10.1016/j.cell.2011.08.016.

Naito, E., K. Dewa, H. Yamanouchi, S. Takagi, and R. Kominami. 1993. "Sex Determination Using the Hypomethylation of a Human Macro-Satellite DXZ4 in Female Cells." *Nucleic Acids Research* 21 (10): 2533–34.

Nei, M. 1973. "Analysis of Gene Diversity in Subdivided Populations." *Proceedings of the National Academy of Sciences of the United States of America* 70 (12): 3321–23.

Ollikainen, Miina, Katherine R. Smith, Eric Ji-Hoon Joo, Hong Kiat Ng, Roberta Andronikos, Boris Novakovic, Nur Khairunnisa Abdul Aziz, et al. 2010. "DNA Methylation Analysis of Multiple Tissues from Newborn Twins Reveals Both Genetic and Intrauterine Components to Variation in the Human Neonatal Epigenome." *Human Molecular Genetics* 19 (21): 4176–88. doi:10.1093/hmg/ddq336.

Pai, Athma A., Jordana T. Bell, John C. Marioni, Jonathan K. Pritchard, and Yoav Gilad. 2011. "A Genome-Wide Study of DNA Methylation Patterns and Gene Expression Levels in Multiple Human and Chimpanzee Tissues." *PLoS Genetics* 7 (2): e1001316. doi:10.1371/journal.pgen.1001316.

Pedersen, Jakob Skou, Eivind Valen, Amhed M Vargas Velazquez, Brian J Parker, Morten Rasmussen, Stinus Lindgreen, Berit Lilje, et al. 2014. "Genome-Wide Nucleosome Map and Cytosine Methylation Levels of an Ancient Human Genome." *Genome Research* 24 (3): 454–66. doi:10.1101/gr.163592.113.

Phan, Trong Nhat, Ee Lin Wong, Xiaoyan Sun, Geunwoong Kim, Seung Hee Jung, Chang No Yoon, and Beom-Seok Yang. 2013. "Low Stability and a Conserved N-Glycosylation Site Are Associated with Regulation of the Discoidin Domain Receptor Family by Glucose via Post-Translational N-Glycosylation." *Bioscience, Biotechnology, and Biochemistry* 77 (9): 1907–16. doi:10.1271/bbb.130351.

Pilli, Elena, Alessandra Modi, Ciro Serpico, Alessandro Achilli, Hovirag Lancioni, Barbara Lippi, Francesca Bertoldi, Sauro Gelichi, Martina Lari, and David Caramelli. 2013. "Monitoring DNA Contamination in Handled vs. Directly Excavated Ancient Human Skeletal Remains." *PLoS ONE* 8 (1): e52524. doi:10.1371/journal.pone.0052524.

Pirazzini, Chiara, Cristina Giuliani, Maria G Bacalini, Alessio Boattini, Miriam Capri, Elisa Fontanesi, Elena Marasco, et al. 2012. "Space/population and Time/age in DNA Methylation Variability in Humans: A Study on IGF2/H19 Locus in Different Italian Populations and in Mono- and Di-Zygotic Twins of Different Age." *Aging* 4 (7): 509–20.

Popkin, B. M., and P. Gordon-Larsen. 2004. "The Nutrition Transition: Worldwide Obesity Dynamics and Their Determinants." *International Journal of Obesity* 28 (S3): S2–9. doi:10.1038/sj.ijo.0802804.

Rakyan, Vardhman K., Thomas A. Down, Natalie P. Thorne, Paul Flicek, Eugene Kulesha, Stefan Gräf, Eleni M. Tomazou, et al. 2008. "An Integrated Resource for Genome-Wide Identification and Analysis of Human Tissue-Specific Differentially Methylated Regions (tDMRs)." *Genome Research* 18 (9): 1518–29. doi:10.1101/gr.077479.108.

Rando, Oliver J, and Kevin J Verstrepen. 2007. "Timescales of Genetic and Epigenetic Inheritance." *Cell* 128 (4): 655–68. doi:10.1016/j.cell.2007.01.023.

Razin, A., and A. D. Riggs. 1980. "DNA Methylation and Gene Function." *Science* 210 (4470): 604–10. doi:10.1126/science.6254144.

Rodríguez-Rodero, Sandra, Juan Luis Fernández-Morera, Agustin F. Fernandez, Edelmiro Menéndez-Torre, and Mario F. Fraga. 2010. "Epigenetic Regulation of Aging." *Discovery Medicine* 10 (52): 225–33.

Sarda, Shrutii, Jia Zeng, Brendan G. Hunt, and Soojin V. Yi. 2012. "The Evolution of Invertebrate Gene Body Methylation." *Molecular Biology and Evolution* 29 (8): 1907–16. doi:10.1093/molbev/mss062.

Sarno, Stefania, Alessio Boattini, Marilisa Carta, Gianmarco Ferri, Milena Alù, Daniele Yang Yao, Graziella Ciani, Davide Pettener, and Donata Luiselli. 2014. "An Ancient Mediterranean Melting Pot: Investigating the Uniparental Genetic Structure and Population History of Sicily and Southern Italy." *PLoS ONE* 9 (4): e96074. doi:10.1371/journal.pone.0096074.

Saygin, Nazan E., William V. Giannobile, and Martha J. Somerman. 2000. "Molecular and Cell Biology of Cementum." *Periodontology 2000* 24 (1): 73–98. doi:10.1034/j.1600-0757.2000.2240105.x.

Schellenberg, E. Glenn, and Sandra E. Trehub. 2008. "Is There an Asian Advantage for Pitch Memory?" *Music Perception: An Interdisciplinary Journal* 25 (3): 241–52. doi:10.1525/mp.2008.25.3.241.

Schmitz, Robert J, Matthew D Schultz, Mathew G Lewsey, Ronan C O'Malley, Mark A Urich, Ondrej Libiger, Nicholas J Schork, and Joseph R Ecker. 2011. "Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants." *Science (New York, N.Y.)* 334 (6054): 369–73. doi:10.1126/science.1212959.

Schroder, Kate, Paul J. Hertzog, Timothy Ravasi, and David A. Hume. 2004. "Interferon-Γ: An Overview of Signals, Mechanisms and Functions." *Journal of Leukocyte Biology* 75 (2): 163–89. doi:10.1189/jlb.0603252.

Shea, N, I Pen, and T Uller. 2011. "Three Epigenetic Information Channels and Their Different Roles in Evolution." *Journal of Evolutionary Biology* 24 (6): 1178–87. doi:10.1111/j.1420-9101.2011.02235.x.

Slieker, Roderick C., Steffan D. Bos, Jelle J. Goeman, Judith Vmg Bovée, Rudolf P. Talens, Ruud van der Breggen, H. Eka D. Suchiman, et al. 2013. "Identification and Systematic Annotation of Tissue-Specific Differentially Methylated Regions Using the Illumina 450k Array." *Epigenetics & Chromatin* 6 (1): 26. doi:10.1186/1756-8935-6-26.

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms." *PloS One* 6 (7): e21800. doi:10.1371/journal.pone.0021800.

Teschendorff, Andrew E., James West, and Stephan Beck. 2013. "Age-Associated Epigenetic Drift: Implications, and a Case of Epigenetic Thrift?" *Human Molecular Genetics* 22 (R1): R7–15. doi:10.1093/hmg/ddt375.

Thompson, Reid F., Gil Atzmon, Ciprian Gheorghe, Hong Qian Liang, Christina Lowes, John M. Greally, and Nir Barzilai. 2010. "Tissue-Specific Dysregulation of DNA Methylation in Aging." *Aging Cell* 9 (4): 506–18. doi:10.1111/j.1474-9726.2010.00577.x.

Thompson, Reid F., Masako Suzuki, Kevin W. Lau, and John M. Greally. 2009. "A Pipeline for the Quantitative Analysis of CG Dinucleotide Methylation Using Mass Spectrometry." *Bioinformatics* 25 (17): 2164–70. doi:10.1093/bioinformatics/btp382.

Tolg, Cornelia, Nesrin Sabha, Rene Cortese, Trupti Panchal, Alya Ahsan, Ashraf Soliman, Karen J. Aitken, Arturas Petronis, and Darius J. Bägli. 2011. "Uropathogenic E. Coli Infection Provokes Epigenetic Downregulation of CDKN2A (p16INK4A) in Uroepithelial Cells." *Laboratory Investigation; a Journal of Technical Methods and Pathology* 91 (6): 825–36. doi:10.1038/labinvest.2010.197.

Traurig, Michael T., Julieanna I. Orczewska, Daniel J. Ortiz, Li Bian, Alejandra M. Marinelarena, Sayuko Kobes, Alka Malhotra, et al. 2013. "Evidence for a Role of LPGAT1 in Influencing BMI and Percent Body Fat in Native Americans." *Obesity (Silver Spring, Md.)* 21 (1): 193–202. doi:10.1002/oby.20243.

Tsukasaki, K., C. W. Miller, E. Greenspun, S. Eshaghian, H. Kawabata, T. Fujimoto, M. Tomonaga, C. Sawyers, J. W. Said, and H. P. Koeffler. 2001. "Mutations in the Mitotic Check Point Gene, MAD1L1, in Human Cancers." *Oncogene* 20 (25): 3301–5. doi:10.1038/sj.onc.1204421.

Tukey, R. H., and C. P. Strassburg. 2000. "Human UDP-Glucuronosyltransferases: Metabolism, Expression, and Disease." *Annual Review of Pharmacology and Toxicology* 40: 581–616. doi:10.1146/annurev.pharmtox.40.1.581.

UNEP, United Nations Environment Programme. 2007. "Reactive Nitrogen in the Environment."

Vanyushin, B. F., L. E. Nemirovsky, V. V. Klimenko, V. K. Vasiliev, and A. N. Belozersky. 1973. "The 5-Methylcytosine in DNA of Rats. Tissue and Age Specificity and the Changes Induced by Hydrocortisone and Other Agents." *Gerontologia* 19 (3): 138–52.

Vidaki, Athina, Barbara Daniel, and Denise Syndercombe Court. 2013. "Forensic DNA Methylation Profiling--Potential Opportunities and Challenges." *Forensic Science International. Genetics* 7 (5): 499–507. doi:10.1016/j.fsigen.2013.05.004.

Voight, Benjamin F, Sridhar Kudaravalli, Xiaoquan Wen, and Jonathan K Pritchard. 2006. "A Map of Recent Positive Selection in the Human Genome." *PLoS Biol* 4 (3): e72. doi:10.1371/journal.pbio.0040072.

Waddington, C. H. 2012. "The Epigenotype." *International Journal of Epidemiology* 41 (1): 10–13. doi:10.1093/ije/dyr184.

Wang, Jinkai, Xiangyu Cao, Yanfeng Zhang, and Bing Su. 2012. "Genome-Wide DNA Methylation Analyses in the Brain Reveal Four Differentially Methylated Regions between Humans and Non-Human Primates." *BMC Evolutionary Biology* 12: 144. doi:10.1186/1471-2148-12-144.

Yao, Yuan, Hong Tao, Dong Il Park, Jorge L. Sepulveda, and Antonia R. Sepulveda. 2006. "Demonstration and Characterization of Mutations Induced by Helicobacter Pylori Organisms in Gastric Epithelial Cells." *Helicobacter* 11 (4): 272–86. doi:10.1111/j.1523-5378.2006.00408.x.

Yoder, J. A., C. P. Walsh, and T. H. Bestor. 1997. "Cytosine Methylation and the Ecology of Intragenomic Parasites." *Trends in Genetics: TIG* 13 (8): 335–40.

Zemach, Assaf, Ivy E. McDaniel, Pedro Silva, and Daniel Zilberman. 2010. "Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation." *Science (New York, N.Y.)* 328 (5980): 916–19. doi:10.1126/science.1186366.

Zeng, Jia, Genevieve Konopka, Brendan G. Hunt, Todd M. Preuss, Dan Geschwind, and Soojin V. Yi. 2012. "Divergent Whole-Genome Methylation Maps of Human and Chimpanzee Brains Reveal Epigenetic

Basis of Human Regulatory Evolution." *American Journal of Human Genetics* 91 (3): 455–65. doi:10.1016/j.ajhg.2012.07.024.

INDEX

Acknowledgments

I would like to express my gratitude to every person who helped me during the 3 years of my Ph.D. programme. I think that the best way to thank them would be through colours that remind me of their particular qualities.

To my supervisor, prof. Donata Luiselli, I would like to dedicate an "ORANGE THANKS". Orange is the mixture of red and yellow, it is the colour of curiosity and exploration of new things, and combines the creative passion of red with the energy and joy of yellow. Thank you for giving me the opportunity to carry out this project and for conveying the passion for the big questions regarding human evolution.

To the "epigenetic team", Paolo Garagnani, Chiara Pirazzini, Maria Giulia Bacalini, Elisa Fontanesi, Elena Marasco and prof. Claudio Franceschi I would like to dedicate a "PURPLE THANKS". Purple is a mixture of red and blue. Red is the colour of passion, creativity and love. Love and passion for scientific research that, day by day, fills our mind and continually amazes us. Red is also the colour of blood, the symbol of effort: this work is about passion but also about dedication. Thank you for being the perfect example that demonstrates how good results require both hard work and passion. Blue is the opposite of red, blue is a cool, calming colour that evokes peace, balance and harmony in working together. Thank you for supporting me so many times and providing me with the right suggestion, since my first steps in the academic world.

To the entire Molecular Anthropology team in Bologna and Ravenna, Marco Sazzini, Alessio Boattini, Elisabetta Cilli, Sara De Fanti, Stefania Sarno, Andrea Quagliariello, Luca Pagani and prof. Davide Pettener I would like to dedicate a "BROWN THANKS". Brown is the colour of wood and stone and evokes the strength for daily research; brown is also the colour of respect, reliability and humility, which are the qualities that I always found in these extraordinary colleagues.

To my European supervisors, Mark Stoneking (Max Planck in Leipzig) and Pedro Moral (University of Barcelona) and all their research team, I would like to dedicate a

"WHITE THANKS". White is the colour of light: thank you for the advice and your scientific inspiration that shed light on some aspects of my work. White is also the colour of comfort: thank you for making me feel at home in your laboratories.

To my husband Valerio, I would like to dedicate a "GREEN THANKS". Green is the colour of nature and harmony. It helps to renewing and restoring depleted energies. Thank you for being with me through both happy and hard times: you always give me a sense of well-being. Being a combination of yellow and blue, green encompasses the mental clarity and optimism of yellow with the emotional calm and insight of blue. Thank you for giving me the gift of these qualities which help me through each moment of my life.

To my family Vendeane, Aurelio and Andrea I would like to dedicate a "BLUE THANKS". Blue is the colour of trust, honesty and peace. The colour of stability, the anchor that helps me to face my insecurities. Thank you for having supported my choices and for your wonderful, sweet and never intrusive presence during these years, with the certainty that your support will never fail.

Finally I would also like to express my thanks to all the people not mentioned here, but who also contributed to my Ph.D and who shared with me their precious expertise.