Evolutionary insights into human DNA methylation

Irene Hernando Herráez

TESI DOCTORAL UPF / 2015

DIRECTOR DE LA TESI

Dr. Tomàs Marquès Bonet

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

upf. Universitat Pompeu Fabra Barcelona

A Nacho,

The important thing is science is not so much to obtain new facts as to discover new ways of thinking about them.

Sir William Bragg

Acknowledgements

En primer lugar me gustaría expresar mi más sincero agradecimiento a la Agencia de Gestió D'Ajuts Universitaris I de Recerca (AGAUR) por el soporte económico recibido para la realización de esta tesis. En segundo lugar, quiero dar las gracias a mi supervisor Dr.Tomàs Marquès Bonet por su confianza y por haberme guiado durante estos cuatro años.

I am grateful to Dr. Andrew Sharp for hosting and helping me during my research stay in New York. Moreover for his support and fruitful discussion during all these years. I also would like to thank to Dr. Holger Heyn for his help and motivation for this project.

A mis compañeros de la celda 421 antigua 407, a los que están y a los que se fueron: Xavi, Marta, Belen, Marc Dabad, Tiago, Raquel, Lukas, Marc, Lobon... En especial me gustaría agradecer a Marcos por su ayuda, paciencia y sabios consejos de informático. Sin él esta tesis no hubiera sido igual. A Jéssica y a Javi, gracias por todos los buenos momentos y por haber hecho que los momentos difíciles fuesen mas llevaderos. También me gustaría agradecer a Jose, Ramon, Guillem e Ignasi por los interminables cafés, interesantes conversaciones científicas y no tan científicas.

En general a todos los bioevos por hacer que este grupo sea un lugar fantástico para realizar una tesis doctoral.

En un plano más personal me gustaría agradecer a toda mi familia, en especial a mis padres, sin su apoyo no hubiera conseguido nada de lo que me he propuesto. Por ultimo, mi mas especial agradecimiento a Nacho, por estar a mi lado, escucharme, comprenderme y por todo lo que hemos vivido durante estos cuatro años.

Abstract

DNA methylation is a crucial epigenetic modification involved in numerous biological processes. However, despite its functional importance, the evolutionary history of this modification and the mechanisms diving such changes are poorly understood. The aim of this thesis is to provide a better understanding of DNA methylation in the context of human recent evolution. We identified and described hundreds of regions presenting a human-specific DNA methylation pattern compared to great apes. We also analyzed for the first time the relationship between DNA methylation changes and sequence evolution at both nucleotide and protein level. In summary, this research reveals new insights into the evolutionary properties of DNA methylation and the interpretation of inter-species non-coding variation.

Resumen

La metilación del ADN es una modificación epigenética implicada en numerosos procesos biológicos. Sin embargo, a pesar de su relevancia funcional, se sabe muy poco sobre su historia evolutiva y los mecanismos que generan estos cambios. El objetivo de esta tesis es proporcionar una mejor compresión de la metilación del ADN en el contexto de la evolución humana reciente. Hemos identificado y descrito cientos de regiones que presentan un patrón de metilación especifico de humanos. Así mismo, hemos analizado por primera vez la relación entre los cambios en metilación y la evolución de la secuencia tanto a nivel nucleotídico como proteico. En resumen, esta investigación revela nuevos conocimientos sobre las propiedades evolutivas de la metilación del ADN y la interpretación de la variación no codificante entre especies.

Preface

A major goal of biology is to understand the molecular basis of phenotypes, specially those of the human phenotype. Since the evolutionary relationship between humans and primates was proposed by Huxley and Darwin (Huxley 1863; Darwin 1871), many efforts have been made to compare our species with its closest living relatives. In particular there is an increasing interest in understanding how human specific traits arose after the split from the Pan lineage about 6-7 MYA (de Waal 2005). A large literature has discussed the behavioral, cognitive and anatomical differences existing between humans and chimpanzees (de Waal 2005). However, it was not until the 70s that early molecular comparisons were carried out. King and Wilson showed that human and chimpanzee protein sequences were nearly similar (about 99%) and proposed that human-specific traits could be explained by regulatory mutations rather than changes of protein structure (King and Wilson 1975).

In recent years, due to the rapid advances in technology, the genomes of all great ape species have been sequenced providing unprecedented opportunities for comparative genomics (Sequencing and Consortium 2005; Scally et al. 2012; Locke et al. 2011; Prüfer et al. 2012). Analyses comparing coding and non-coding sequences from human and chimpanzees determined about ~1.2% of sequence divergence, a value based on single nucleotide variants (Sequencing and Consortium 2005). More recently ~5% of divergence was

estimated when considering deletion and duplication events. Nevertheless, despite the important advances in the field, the gap between phenotypic and genotypic studies remains extensive. Evidences in recent years support the hypothesis that species evolution predominantly occurs via regulatory adaptation and subsequent gene expression changes (McLean et al. 2011). In this regard, the epigenetic network is expected to add layers of regulation, through an interplay between the genome and the phenotype. However, to date very few attempts have been made to understand the evolutionary features of our epigenome.

Table of Contents

1. INTRODUCTION

1.1. Epigenetics: DNA methylation

The prefix 'epi' is derived from the Greek and means around, over or beside. In the word epigenetics it refers to the addition of an extra layer of information to the DNA sequence. This term was introduced in 1957 by Conrad Waddington to describe the changes in gene activity during development that ultimately will determine the phenotype. In his model, Waddington compares the cellular differentiation process to an epigenetic landscape **(Waddington 1957)**(Figure 1).

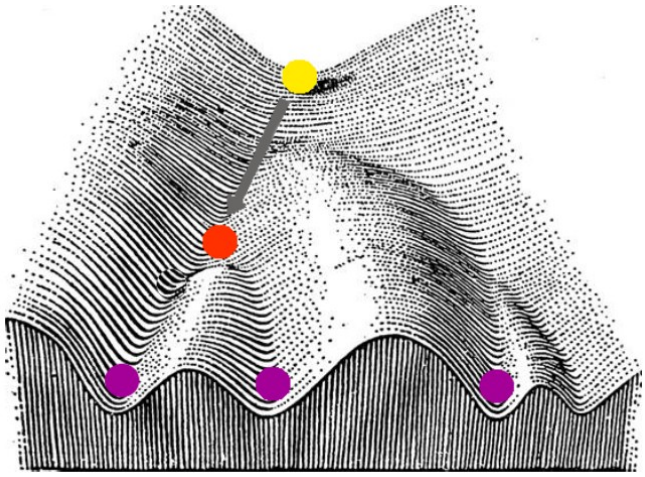

Figure1. Waddington's epigenetic landscape. The yellow ball represents a pluripotent cell and purple balls represent differentiated states. [Adapted form (Waddington 1957)]

Here, a ball situated on top of a hill represents a pluripotent cell. The ball can roll down and go into several hollows which represent differentiated states. As the ball moves down the final destinations become more limited as occurs in the cellular differentiation process. Therefore, the concept of epigenetics appears closely linked to the development. However, nowadays epigenetics is one of the fastest-growing research fields and has become key in many disciplines such as cancer or aging. Epigenetic processes include DNA and histone modifications, non-coding RNAs and chromatin structure. Together, these mechanisms control chromatin accessibility and modulate transcriptional activity. This thesis is focused on one of the best-known epigenetic modification: DNA methylation.

1.1.1. Setting, maintaining and removing DNA methylation patterns

DNA methylation was discovered in a calf thymus by Rollin Hotchkiss in 1948 (Hotchkiss 1948). However, it was not until 1980s where several studies demonstrated that DNA methylation was involved in gene expression and cellular differentiation (Compere and Palmiter 1981; Holliday and Pugh 1975). This modification entails the covalent addition of a methyl group (-CH3) in the fifth carbon of cytosine pirimidine ring. It is found in all vertebrates, plants, many invertebrates, fungi and bacteria and in

mammals, it typically occurs at Cytosine-phosphate-Guanine (CpG) dinucleotides .

The addition of the methyl group is catalyzed by a family of DNA methyltransferases (DNMTs) that transfers the S-Adenosyl methionine to the cytosine (Figure 2). To date, five members of this family have been identified although only three have been shown to be active. Active DNMTs (DNMT1, DNMT3A, DNMT3B) contain a unique variable N-terminal regulatory region and a highly conserved C-terminal catalytic domain (Goll and Bestor 2005). In mice, the knockout of any of these enzymes is lethal and leads to a global depletion of DNA methylation. (Li et al. 1992; Lei et al. 1996; Okano et al. 1999). Not surprisingly, the absence of DNA methylation in some species such as *Saccharomyces cerevisiae* or *Caenorhabditis elegans* is associated with the evolutionary loss of the methyltransferases enzymes (Proffitt et al. 1984; Simpson et al. 1986).

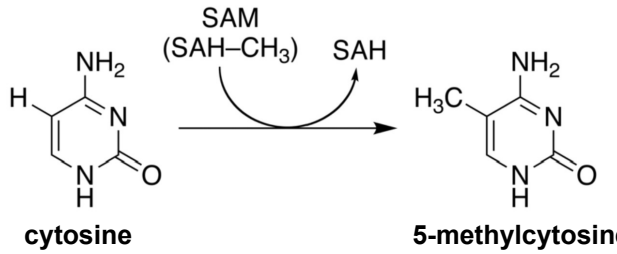

Figure 2. Conversion of cytosine to 5-methylcytosine by DNA methyltranferases.

DNMTs enzymes can be classified in two main groups: maintenance and *de novo* enzymes. DNMT1 is the main maintenance enzyme (Yen et al. 1992). During mitosis, after DNA replication, the methylation pattern of the parental strand creates a profile of hemimethylated sites. DNMT1 enzyme shows high affinity for this motif and introduces methyl groups (Bestor 1992). In this way DNA methylation patterns are efficiently propagated from parent to daughter cells. The recruitment of the DNMT1 enzyme is a complex process highly dependent on the cell cycle (Kishikawa et al. 2003). The association with the replication machinery is mediated by the interaction of the UHFR1 protein, non-coding RNAs and the nuclear PCNA antigen which appears in the replication fork (Chuang et al. 1997; Bostick et al. 2007). In addition, DNMT1 is also present at heterochromatic regions specially in late S-phase (Easwaran et al. 2004).

The establishment of new DNA methylation patterns in early mammalian development is highly controlled by the DNMT3 family members: DNMT3A and DNMT3B, and the regulatory factor DNMT3L. DNMT3A and DNMT3B introduce DNA methylation at palindromic CpG sites in the absence of a template, creating a pattern of methylated and unmethylated sites (Okano et al. 1999). DNMT3L lacks a catalytic domain however it has the ability to compete with DNMT3A and DNMT3B, what reduces their activity and consequently methylation levels (Hata et al. 2002; Aapola et al. 2000). The regulation of DNMT3 enzymes has been

subject of intense research (Law and Jacobsen 2010). Several mechanisms have been proposed including specific interactions with proteins and histone tail modifications. These interactions are partially facilitated by the regulatory domains located on the N-terminal of DNMT3 enzymes. The two major domains are the PWWP domain and the ATRX-DNMT3-DNMT3L (ADD) domain (also known as the plant homeodomain, PHD). The PWWP domain guides the enzyme to H3K36 di- and trimethylation which is located in gene bodies and heterochromatin regions (Dhayalan et al. 2010). The ADD domain prevent the enzymes from binding to H3K4 di- or trimethylation, a characteristic of active chromatin (Zhang et al. 2010). In addition, DNMT3A and DNMT3B interact with transcription factors such as MYC (Brenner et al. 2005), microRNAs and transcriptional repressors (Fuks et al. 2001). It is important to remark that the classification of demethyltransferases in these two categories is a simplification of the process. It has been shown that DNMT3 enzymes are implicated in the maintenance of DNA methylation patterns and DNMT1 in the *de novo* methylation process, particularly at repeats elements (Liang et al. 2002).

DNA methylation landscape is also influenced by DNA demethylation mechanisms. This process can occur passively after replication if remethylation at hemimethylated sites does not take place, or actively. Active demethylation is directed by three members of the ten-eleven translocation (TET) protein family. TET proteins convert 5-methylcytosine (5mC) into 5-

hydroxymethylcytosine (5hmC). Two consecutive oxidative reactions catalyzed also by TET result in 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Tahiliani et al. 2009; Ito et al. 2011). 5fC and 5caC marks are recognized by the enzyme thymine DNA glycosylase (TDG) which replaces the modified cytosine with an unmodified cytosine (Maiti and Drohat 2011)(Figure 3).

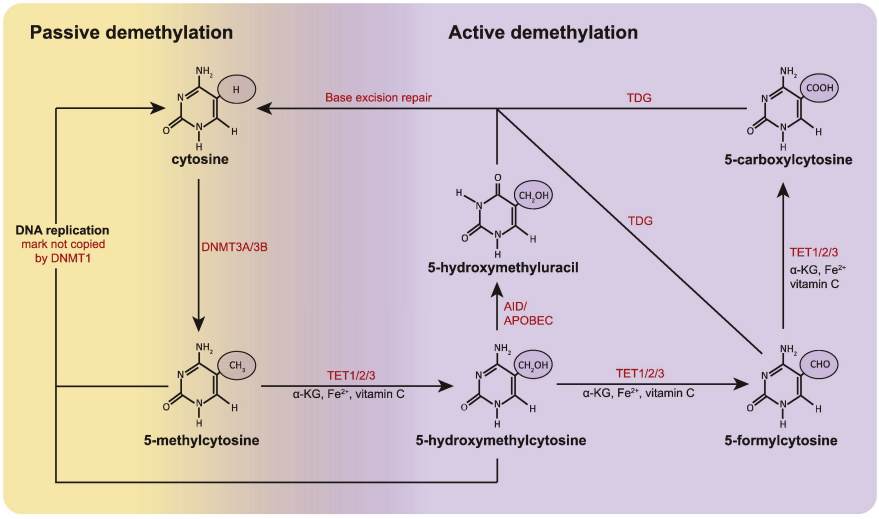

Figure 3. DNA methylation and demethylation pathways. DNMT1 is mainly involved in maintaining 5mC during DNA replication, while DNMT3A and DNMT·B regulate *de novo* DNA methylation. After replication if remethylation at hemimethylated sites does not take place this can lead to passive demethylation. Active demethylation is mediated by TET proteins that convert 5mC into 5hmC, 5fC and 5caC through three consecutive oxidative reactions (Kroeze et al. 2015)

The activity of TET enzymes and the intermediate states of DNA methylation are still far from understood. TET1 enzyme is mainly expressed in the early development, whereas TET2 and TET3 are

more broadly expressed, however the process is (Branco et al. 2011). Recent studies have shown that active regulatory regions present high levels of 5hmC probably indicating a high turnover of DNA methylation(Stroud et al. 2011; Song et al. 2013).

1.1.2. Distribution and function of 5mC in mammalian genomes

The distribution of 5mC throughout mammalian genomes is not random. The majority of CpG sites are methylated, whereas unmethylated CpGs are principally concentrated at CpG rich regions called CpG islands (CGIs). In humans about 70% of genes promoters (mostly housekeeping genes) are embedded into a CGI (Saxonov et al. 2006). These promoters are rarely methylated and can be either transcriptionally active or inactive depending on other factors such as H3K4me3 and Polycomb proteins respectively (Taberlay et al. 2011; Kelly et al. 2010). Methylation of CGI at transcription start sites (TSS) causes robust transcriptional repression and it is associated with long-term silencing. Classical examples include imprinted genes or genes located on the inactive X chromosome in mammalian females. Therefore, CGI promoters rarely show tissue-specific methylation patterns. In contrast methylation levels of CpG poor promoters tend to be more variable.

Recent genome-wide techniques have revealed a regulatory role of DNA methylation in CpG poor regions located outside promoters. It

has been shown that certain enhancer regions acquire intermediate methylation levels when occupied by transcription factors (Stadler et al. 2011). It is also likely that other transcription factors are influenced by DNA methylation, in particular those containing a CG dinucleotide in their binding motif (Chen et al. 2011).

Transposable elements are extensively methylated, as are gene bodies and intergenic regions. However, the functional consequences are poorly understood. Some studies have shown found a positive correlation between gene body DNA methylation and transcriptional levels (Ball et al. 2009; Aran et al. 2011). In addition, gene body DNA methylation has been associated with alternative splicing. (Shukla et al. 2011). In transposable elements, DNA methylation acts a suppressor mechanism avoiding transposition and ensuring gene expression stability (Slotkin and Martienssen 2007). All these facts highlight the complexity of DNA methylation machinery and indicate that DNA methylation function is highly dependent on the genomic context.

1.1.3. Importance of DNA methylation

a) Development

The role of DNA methylation in embryonic development and cellular differentiation has been studied for more than three decades**.** In general, adult somatic cells are highly specialized and

their methylation patterns are relatively stable (Hon et al. 2013; Ziller et al. 2013). Mammalian sperm and egg show high levels of DNA methylation similar to levels of somatic cells. This epigenetic state represent a major barrier to sexual reproduction since the preparation for the next generation requires a complete reset of the epigenome. This resetting process is called epigenetic reprogramming and it is a crucial step during mammalian development (Lee et al. 2014).

Genome-wide epigenetic reprogramming occurs at two different stages (Figure 4). The first stage begins after fertilization and entails the erasure of gamete epigenetic marks including DNA methylation and histone modifications. Paternal and maternal genomes undergo active and passive demethylation respectively. Consequently, at the time of implantation the embryo genome is broadly unmethylated. (Eilertsen et al. 2007; Sasaki and Matsui 2008). Afterwards, the genome is rapidly remethylated followed by further methylation remodeling during cell differentiation and tissue development (Figure 4B). The second wave of reprogramming takes place at embryonic primordial germ cells (PGCs), which are the precursors of gametes and founders of the following generation.

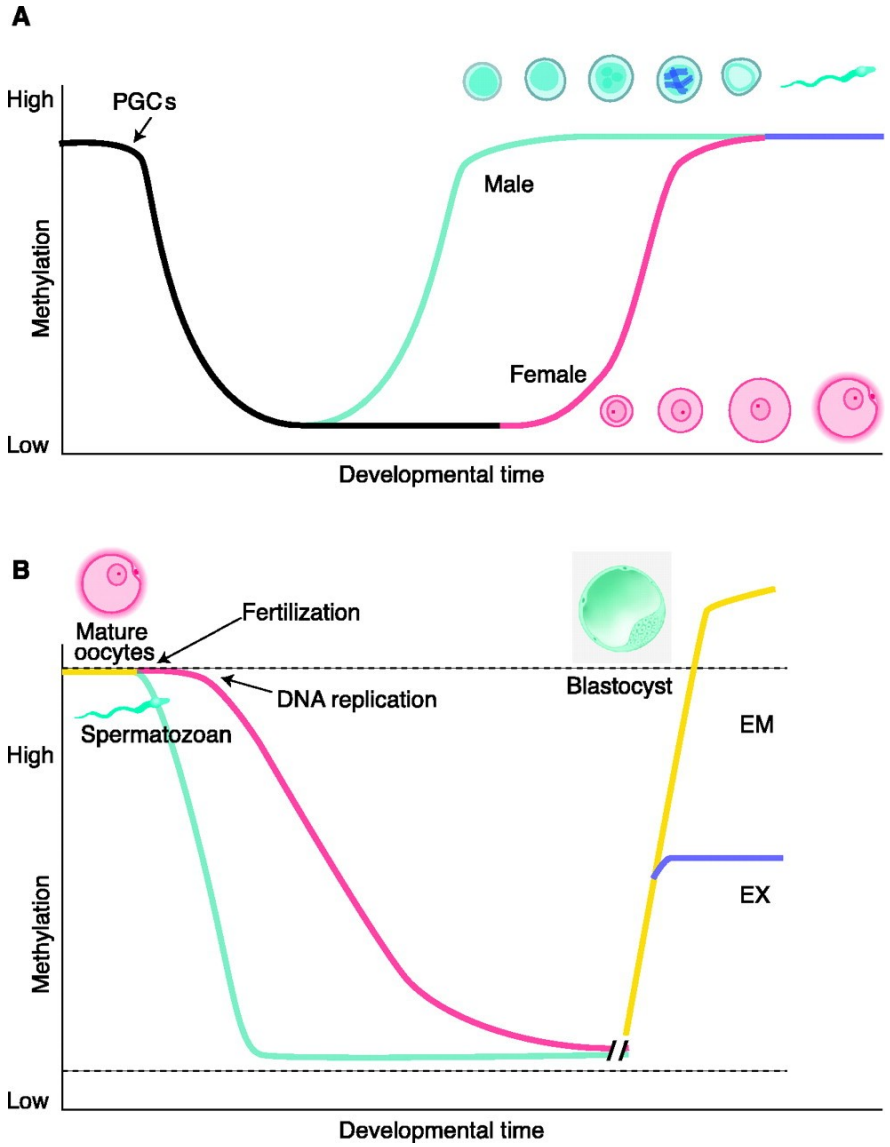

Figure 4. Two waves of epigenetic reprogramming (A) Methylation reprogramming in the germ line. (B) Methylation reprogramming in embryos. Paternal (green) and maternal (pink) genomes are remethylated around the time of implantation in embryonic (EM) and extra-embryonic (EX) tissues. (Reik et al. 2001)

PGCs undergo global demethylation during migration to genital ridges and when they arrive their genome is mostly unmethylated. Subsequently PGCs acquire their correspondent methylation patterns which would be further reshape during gametogenesis in adulthood (Seki et al. 2005) (Figure 4). Imprinted regions, whose methylation pattern depends on the parental origin, are also reprogrammed in PGCs. The original methylation pattern must be lost and re-established based on the sex of the embryo.

Despite during reprogramming the vast majority of epigenetic marks are removed, some if them persist. The best-known examples are the intracisternal A particle (IAP) retrotransposon family and adjacent CGIs in mice (Guibert et al. 2012; Hackett et al. 2013). This incomplete reprogramming may provide a potential mechanism for trans-generational transmission of epigenetic marks (Ferguson-Smith and Patti 2011).

b) Aging and disease

Although to less extent, DNA methylation patterns also vary across lifespan. In a recent study, the methylomes of a newborn and a centenarian individual were compared. The centenarian methylome presented lower levels of methylation and worse correlation between neighboring CpG sites in terms of methylation (Heyn et al. 2012b). Other studies have calculated the age of an individual using exclusively methylation information (Horvath 2013; Hannum et al. 2013). The difference between the predicted age and the

chronological age has been associated with age-related diseases and mortality risk (Marioni et al. 2015; Hannum et al. 2013; Horvath 2013). In fact, alterations of DNA methylation patterns have been observed in certain diseases, particularly in cancer (Portela and Esteller 2010). The importance of DNA methylation in human cancer was evident with the identification of *TET2* and *DNMT3* mutations in several leukemias (Ko et al. 2010; Ley et al. 2010). Despite the underlying mechanisms remain largely unclear, the loss of function of these proteins seems to be an initial event. In addition, promoters of several tumor suppressor genes have been found extensively methylated in many different tumors types (Jones 2012).

c) DNA mutation

DNA methylation levels can indirectly modulate nucleotide substitution rates. Mammalian genomes present an overall depletion of CpG sites. They represent only ~0.8% of all dinucleotides while the expected frequency is 4.4%. This can be attributed to the high mutability of methylated cytosines, which spontaneously deaminate to thymine generating a T•G mismatch (Duncan and Miller 1980). Non methylated cytosines also deaminate into uracil, however uracil can be removed by uracil-DNA glycosylase whereas the T•G mis-pair can not be corrected by this enzyme (Neddermann and Jiricny 1994). Consequently, the mutation rate of 5mC is between 4-fold and 15-fold higher than other transition mutation rates

(considering polymorphisms, evolutionary divergence and disease mutations) (Li et al. 2009; Jiang and Zhao 2006; Kondrashov 2003). For instance, the human-chimpanzee divergence at CpG sites is around 15% in sharp contrast with barely 1% for any other nucleotide substitution type (Duncan and Miller 1980). Moreover, C to T transitions at CpG sites is the most frequent mutation observed in human inherited diseases (Cooper et al. 2010).

1.1.4. Epigenetic variation and inheritance

It is generally accepted that DNA methylation patterns that lay down during embryonic development are afterwards dictated by the underlying genetic code. However, epigenetic patterns can be altered in a manner that is independent of the DNA sequence. On the one hand, epigenetic variation can be produced by stochastic processes such as small errors in copying DNA methylation patterns during successive cell divisions (Feinberg and Irizarry 2010). On the other hand, epigenetic variation can be triggered by environmental factors. Honeybees provide an impressive example where the epigenome and ultimately the phenotype are affected by the diet. In a colony, thousands of larvae are genetically identical and only those fed with royal jelly will become queens. Recent studies have demonstrated that the knock down of *DMNT3* causes an increased number of queens from larvae not fed with royal jelly (Lyko et al. 2010). Furthermore, brains of queens and workers have

different DNA methylation patterns. Interestingly, queens that were induced via *DMNT3* down-regulation showed similar patterns to normal queens, suggesting that the effects of royal jelly occurred via DNA methylation (Kucharski et al. 2008).

In humans, for example, the consequences of smoking on the epigenome have been examined in several studies (Shenker et al. 2013; Richmond et al. 2014) . It is also well-known the effect of maternal diet during pregnancy. An interesting study showed that individuals who were prenatally exposed to famine during the Dutch Hunger Winter in 1944 showed altered patterns of DNA methylation. Moreover, they presented a higher risk for developing diabetes and obesity when compared to their unexposed siblings. Interestingly, these changes were detected sixty years after the famine was experienced (Heijmans et al. 2008; Tobi et al. 2014). However, how these changes influence future generations remains largely unclear.

It is important to distinguish intergenerational effects, such as the above-described case of in utero-exposure, from true transgenerational inheritance. During pregnancy, not only the mother and the embryo are exposed to the same environmental influences but also the embryo primordial germ cells that will eventuality produce the F2 generation. To be considered trans-generational, the consequences must also be observed at least in the

F3 generation in case of a maternal exposure or the F2 in case of the parental exposure. (Figure 6)

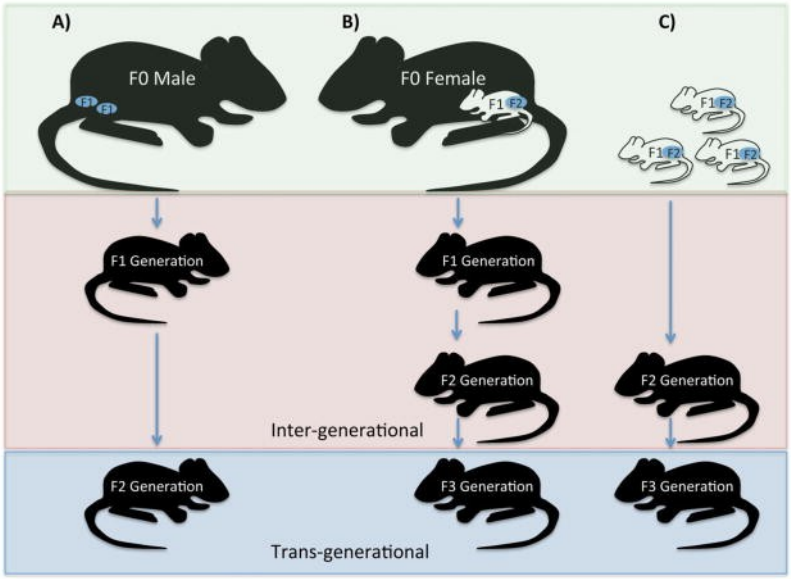

Figure 6. Inter and transgenerational inheritance. A)The exposure of the paternal ancestor to an environment perturbation affects not only the individual but also his sperm and consequently the F1 generation. To be considered transgenerational inheritance the effects should be observed in the F2 generation B) In the case of maternal exposure during pregnancy the embryo and the primordial germ cells are affected, altering F1 and F2 generation. Here, transgenerational effects should survive into the F3 generation as well as C) in the case of early peri-natal exposure (Dias and Ressler 2014a)

There is an increasing interest in epigenetic transgenerational inheritance owing to their possible role in adaptation and evolution (Richards 2008)(see section 1.2.2). However, in mammals epigenetic transgenerational inheritance is a rare phenomenon. This can be explained by the genome-scale erasure of epigenetic marks

during germ cell reprogramming. However, in plants where PGCs do not undergo genome-wide reprogramming, the inheritance of epigenetic marks is more common. A classic example is the study of toadflax (*Linaria vulgaris*) that present two different varieties. The wild type plant produces asymmetric bilateral flowers, whereas the mutant, originally described by Linnaeus more than 200 years ago, produces symmetric radial flowers (Figure 7). These two different phenotypes are caused by DNA hypermethylaion and transcriptional silencing of the cycloidea type gene (*Lcyc*) which controls the formation of dorsal petals. This epialelle has been identified in natural populations during several generations (Cubas et al. 1999; Kalisz and Purugganan 2004) .

Figure7. Wild type (left) and mutant of *Linaria,* which show bilateral and radial symmetry due to differences in DNA methylation.

In mammals, examples of epigenetic transgenerational inheritance have been reported only in model organisms. The best evidence comes from the viable yellow agouti mice (A^{vy}). The A^{vy} allele is the result of an insertion of a intra-cisternal A particle (IAP) retrotransposon about 100 Kb upstream of the *Agouti* gene. Phenotypic variation occurs due to stochastic methyation of the IAP element. When the IAP sequence is unmethylated the gene is transcribed resulting in yellow fur, obesity and increased susceptibility to develop tumors and diabetes (Figure 9). The degree of IAP methylation varies among isogeneic individuals causing a wide distribution in coat colour, ranging from yellow (unmethylated) to brown (methylated) (Morgan et al. 1999; Rakyan et al. 2002; Rakyan and Beck 2006) .

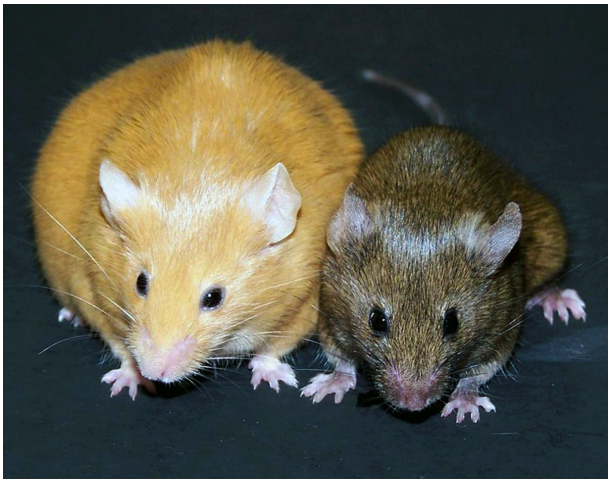

Figure 8. Agouti mice of the same age and sex.

Trans-generational transmission has also been reported in cases of environmentally induced changes. For example, alterations in the DNA methylation patterns were observed in mice after ancestral exposure to fungicides (Anway et al. 2005) and conditioned fear to an odorant (Dias and Ressler 2014b). Despite the underlying molecular mechanisms of trans-generational inheritance are largely unknown, it raises the intriguing question of whether this phenomenon occurs in natural populations and whether it could affect species evolution.

REVIEW

1.2. Comparative epigenetics: Patterns of CpG methylation in human evolution

---

Irene Hernando-Herraez[1], Raquel Garcia-Perez[1], Andrew J. Sharp[2,*], Tomas Marques-Bonet[1,3,4*]

***Plos Genetics,* under review.**

[1]Institute of Evolutionary Biology (UPF-CSIC), PRBB, 08003 Barcelona, Spain.
[2]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai School, New York, NY 10029, USA.
[3]Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain
[4]Centro Nacional de Analisis Genomico (CNAG), PCB, 08023 Barcelona, Spain

**Corresponding authors*

Abstract

Humankind has always been intrigued by the acquisition of the particular traits that define our species. However, after a deluge of primate comparative genomic studies, most phenotypic differences between humans and the rest of great apes remain unexplained. It has long been hypothesized that human specific traits may be largely driven by changes in the regulatory processes rather than in the protein sequence. Over the past few years, evolutionary studies have shifted from mere sequence comparisons to integrative analyses where consideration of regulatory elements is key to understand the basis of species evolution. Epigenetic mechanisms such as DNA methylation might play a central role in understanding phenotypic differences. However, the evolution of the human epigenome and the processes driving such changes are poorly understood.

In this review, we discuss the role and significance of epigenetics in evolution. We also summarize the latest advances in the field of comparative epigenomics, focusing on human specific changes compared to our closest living relatives, the primates. We examine the challenges that need to be faced and encourage the scientific community to approach future research bearing in mind the importance of genome-epigenome interactions to disclose our evolutionary history.

1.2.1. Introduction

A long-standing question for evolutionary biologists is to understand the molecular mechanisms underlying phenotypic diversity. The very first molecular studies focused on protein variability as the main force driving species evolution. Nonetheless, it soon became evident that protein sequence divergence alone could not explain the striking phenotypic differences between closely related species. Humans and chimpanzees, our closest living relative, present many morphological and cognitive differences. However, our proteins are surprisingly similar, and most differences are found in non-coding regions [1]. In this context, regulatory modifications have been proposed as critical in the acquisition of human-specific traits [2] rather than changes in protein-coding genes.

Studies over the last decade support the hypothesis that species evolution predominantly occurs via regulatory adaptation and subsequent gene expression changes [3]. By comparing multiple species genomes, regulatory elements have been identified based on sequence conservation and accelerated mutation rate across coding and non-coding regions of the genome [4–6]. However, epigenetic marks provide an additional layer of complexity that cannot be immediately detected by sequence analyses. Here we focus on DNA methylation. Although this epigenetic mark is a key regulatory mechanism involved in many biological processes such as gene

regulation, cellular differentiation, repression of transposable elements, chromosome X inactivation and genomic imprinting [7], its role in evolution is only beginning to be explored. In this review we survey the incipient field of epigenetics and evolution, mainly the role of DNA methylation during human evolution.

1.2.2. Relevance of DNA methylation in evolution

The importance of DNA methylation as a regulatory mark in cellular differentiation and development has been recognized for more than three decades [8,9]. More recently the discovery of altered DNA methylation patterns in several diseases, particularly in cancer, triggered the emergence of a new field [10,11]. DNA methylation changes have also been identified in association with aging and environmental stimuli such as nutrition and lifestyle. For example, the effects of smoking on the epigenome have been examined in several studies [12–14] and persistent epigenetic alterations associated with maternal smoking have also been detected after prenatal exposure [15]. Despite the impressive growth of epigenetics research, it is important to remark that these studies have been focused on the dynamics of DNA methylation during an individual's lifetime.

In the recent years, DNA methylation has drawn the attention of the evolutionary community. This is an exciting field that can shed light into the principles lying behind human epigenome, but with many

questions remain unanswered. Little is known about how DNA methylation patterns have changed over evolutionary time and whether these have contributed to the speciation process. It has been hypothesized that epigenetic variation could affect species evolution [16]. Similar to genetic mutation, stochastic epigenetic changes could provide a source of variability. Moreover epigenetic modifications could appear as a consequence of environmental factors, affecting simultaneously several individuals within the population that would favor rapid adaptation. However, it is difficult to understand the scope and relevance of epigenetic variation in evolution, principally because of its controversial mechanism of transmission. In order to contribute to the evolution of species, any such epigenetic variants would need to be heritable from one generation to the next. This trans-generational inheritance requires the passage of epigenetic marks through the germline without being erased. In mammals, an efficient reprogramming takes place at early developmental stages that removes almost all epigenetic marks [17]. As a result, it is thought that DNA methylation patterns largely reflect the underlying genetic blueprint and thus, most of the epigenome is encoded in the genome (Figure 1). However, recent studies have demonstrated the transmission of sequence independent DNA methylation changes across generations in model systems reopening the old Lamarckian debate. The most extensively studied case of trans-generational inheritance is the mouse agouti locus in which the stochastic methylation state of IAP

can alter expression of the endogenous genes [18,19]. Trans-generational persistence of epigenetic marks has also been reported in cases of environmentally induced changes. For example, a recent report indicated that alterations in DNA methylation patterns were observed in mice after parental olfactory exposure [20]. These findings have raised the intriguing question of whether this phenomenon occurs in natural populations and whether it could affect species evolution.

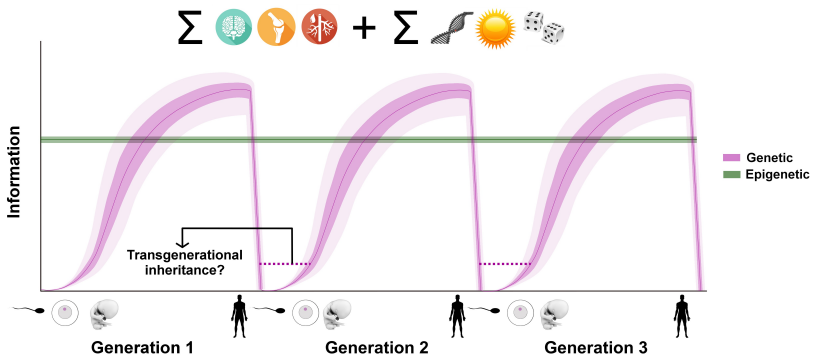

Figure1: Schema of genetic and epigenetic information across generations. The y-axis represents how information varies across generations (x-axis). Epigenetic marks(purple) are erased after fertilization and re-established afterwards. Epigenetic information can be regarded as the sum of the epigenome in every tissue; this information is determined primarily by the genetic sequence, but it can also be altered over anindividual's lifetime by stochastic and environmental processes. Genetic information(green) is stable across generations, with the exception of mutational and recombination effects. The stable transmission of epigenetic information across generations would result in the trans-generational inheritance (dotted line).

In this regard, one of the most controversial aspects is that these epigenetic alterations are unstable over time and the phenotypic effects disappear in few generations. Nonetheless, the field of epigenetic transgenerational inheritance is actively growing and many scenarios, such us the study of long-term environmental exposure (across several generations) [21], have not been explored yet.

1.2.3. The interplay between the genome and the epigenome

It is generally accepted that the basic patterns of DNA methylation laid down during embryonic development are largely dictated by the underlying genetic code. Several lines of evidence support this idea. For example, sequence evolution that alters CpG sites, transposable elements (TE), the binding of transcription factors (TF) or single nucleotide changes have all been shown to shape the landscape of the methylome [22] (Figure 2). CpG sites can arise from point mutations and also as a by-product of biased gene conversion [23]. New CpG sites have the potential to create novel regulatory regions. One such example is the *LHX1* gene, which acts in the transcriptional regulation of vertebrate head and mesodermal configuration. Here, the human lineage acquired additional CpGs adjacent to a CpG island, creating a tissue-specific differentially methylated region (DMR) in humans that is absent in mouse [24].

Another recent study described regions with a high density of CpG sites present exclusively in the human lineage, indicating these as loci with potential species-specific regulatory roles [25].These regions were found to be enriched for genes related to cognition and behavior, including the well-characterized *HAR1A* gene, which plays a crucial role in cortical development [4].

Consistent with the notion that genomic sequence influences the epigenome, a number of studies have identified genetic variants that are associated with cis-linked DNA methylation levels, termed methylation quantitative trait loci (mQTL) [26–29]. In human populations thousands of mQTLs have been identified. In these cases a single mQTL can modulate the methylation state of multiple clustered CpGs. While the influence of mQTLs can extend up to several kilobases, the strongest effects are generally seen for SNPs that are located within or nearby the affected region [29,30]. Further, SNPs that disrupt TF binding sites are more frequently associated with altered methylation levels, suggesting a complex mechanism of epigenetic regulation orchestrated via sequence-specific DNA or chromatin binding factors. Indeed, studies in mice have shown that the presence of specific sequence motifs, potentially representing transcription factor binding sites, acts to specify methylation levels during cellular differentiation [31].

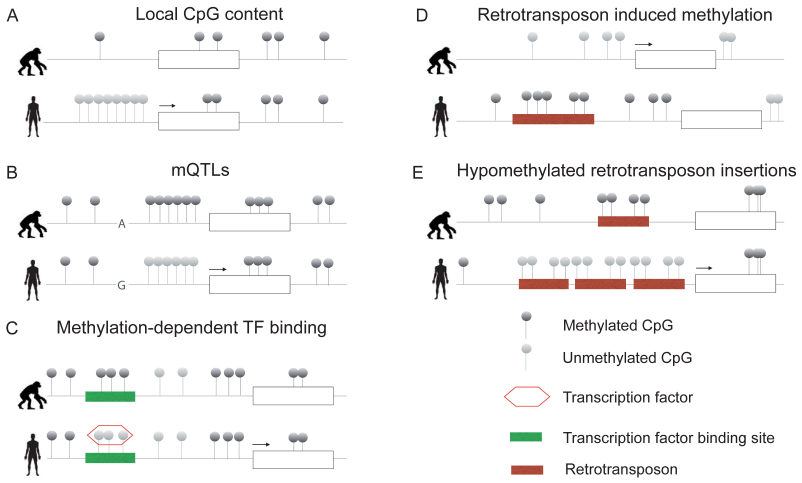

Figure 2: Putative mechanisms of methylation divergence. a) The generation of novel CpG sites can lead to the formation of novel CpG islands, CpG island shores or CpG island shelves, which in turn are susceptible to methylation. b) mQTLs can modulate the methylation levels of adjacent regions. c) The binding of TFs can be influenced by the methylation status of the underlying TF binding site. d) Hypermethylated retrotransposons induce the methylation of nearby CpGs. e) Hypomethylated retrotransposon are mobile and can insert at different genomic locations.

While there is clear evidence for the influence of DNA sequence on epigenetic state, conversely DNA methylation levels have been also shown to modulate nucleotide substitution rates and thus influence genome evolution. Mammalian genomes present an overall depletion of CpG sites that can be attributed to the high mutability of methylated cytosines, which undergo rapid deamination to thymine [32]. For instance, while the human-chimpanzee nucleotide substitution frequency is on average ~1% across the genome, at

CpG sites is approximately 15% [1]. In contrast to high mutation rates associated with methylated CpGs, unmethylated cytosines are much less prone to mutation, allowing for the maintenance of high CpG content in regions with low methylation levels. In addition to influencing the mutation rate, analysis of chromosomal rearrangements in the gibbon genome shows an association with hypomethylation of Alu elements, suggesting that DNA methylation patterns might also play role in karyotype evolution [33]. Further studies are required to understand the interplay between genetic and epigenetic variation and how this influences evolutionary processes.

1.2.4. Comparative epigenomics

Comparison of epigenetic profiles from multiple primate species shows that methylation patterns are able to recapitulate the phylogenetic relationships between species and even between populations [28,34,35]. The most likely explanation for this is the dependence of methylation levels on the underlying genome sequence. Nonetheless other scenarios, such us separate environments, cannot be disregarded. Irrespective of how epigenetic patterns are determined, the analysis of epigenetic marks between species is an expanding and promising research field [36] comparative epigenomics can provide insight into novel human-specific regulatory regions and the molecular mechanisms governing epigenetic patterns.

a) Primates

In general, global patterns of DNA methylation are similar when comparing human and closely related species [35,37], presenting a bimodal distribution of DNA methylation levels along the genomes which are largely methylated except most promoter regions. Methylation levels at promoter-associated CpG islands tend to be highly conserved even when comparing divergent species of vertebrates [38], demonstrating that the unmethylated state is an ancient feature of vertebrate promoters. This unmethylated state could partially explain the absence of CpG erosion at promoter regions instead of being the result of selective pressure [39]. Unmethylated regions outside gene promoters have been associated with distal regulatory elements, non-coding RNAs and unannotated transcription start sites (TSS) [38]. Unmethylated regions shared between species are generally longer, have higher CpG density and tend to be located near TSS compared to the majority of interspecies variation that is found away from gene promoters, complicating interpretation of their functional effects [36,38,40]. Several studies have revealed that regions of methylation divergence coincide not only with tissue-specific function but also with developmental and neurological processes [25,34,37,41]. Although it is unclear how to interpret the latter, it has been speculated that this could be due to vestigial regulatory elements that were functional during development and their methylation state has remain unaltered in adult tissues [42].

The potential relationship between DNA methylation differences and gene expression has been addressed in several evolutionary studies. It has been shown that species-specific DMRs are associated with histone modifications and transcription factor binding sites, suggesting a role in regulating gene expression [28,36,38]. Direct evidence suggests that differences in promoter methylation underlie 12-18% of gene expression differences between human and chimpanzee [41]. This is in agreement with the results obtained by Heyn et al., who explored human population methylomes and determined that gene expression differences were associated with promoter hypermethylation in ~13% of cases [28].

b) Ancient hominids

The genome sequencing of extinct hominids has provided a better understanding of the population history and genome evolution of our species. However additional studies have been limited due to the tiny amounts and degraded nature of DNA that can be extracted from ancient bones. Recently, a novel method was developed to infer DNA methylation patterns from the genome sequence of Neanderthal and Denisovan, opening the possibility to study for the very first time the methylome of extinct species [43]. This method is based on the different spontaneous deamination rates of methylated and unmethylated cytosines taking the advantage of the characteristic substitution pattern accumulated over thousands of years of chemical degradation. Using this approach, they identified

over a thousand DMRs among the three species. Remarkably, one of these DMRs was located in the *HOXD* cluster, a key regulator of limb development [44]: whereas the *HOXD9* promoter and *HOXD10* gene body are hypomethylated in humans, both archaic species were hypermethylated at the very same regions; also, *HOXD9* gene body was hypermethylated in the Denisovan genome. They speculated that differential methylation in the *HOXD* cluster could account for some of the anatomical differences between archaic and present-day humans [43].

Despite the relevance of this study it is also important to note some caveats. Perhaps most importantly, these results are based on analysis of just one extinct individual, and therefore the observed differences could simply represent polymorphisms. Further studies would be required to define the ancestral DNA methylation state since this information would allow us to reconstruct the evolutionary trajectories of these regions.

c) Human populations

Some studies have addressed DNA methylation patterns across human populations [28,45]. To date, Heyn et al. have conducted one the most thorough studies examining the contribution of DNA methylation to human population variation. By analyzing over 400,000 CpG sites in B-lymphocytes isolated from three distinct human populations (Caucasian, African-American and Han

Chinese) they were able to recapitulate the demographic history of each group. They identified 439 population-specific CpGs associated with several histone modifications and transcription factor binding sites, suggesting their involvement in gene regulation. Moreover the affected genes were related to susceptibilities to different diseases, response to drugs and environmental factors. Furthermore, comparison to chimpanzee methylation patterns, they revealed 39 CpG sites that may have evolved under local selective pressure. Interestingly, these included immune and xenobiotic response factors, suggesting that these changes may have been driven by geographic differences in pathogens or environmental factors.

1.2.5. Comparative epigenomics: challenges

In spite of the recent advances in the field of comparative epigenomics, several challenges still need to be faced. One of the most difficult problems relates to the dynamic nature of DNA methylation among cell lineages.

Most studies analyze heterogeneous tissues which are a composition of multiple different cell types. Since methylation patterns are unique to each cell type this can be a major confounding factor, especially as the fraction of each cell type present within a particular tissue can vary between individuals or species. One way to overcome this limitation is to conduct studies in purified single

cell types [35], although accessing the necessary samples that are needed for cellular purification is not always feasible. Alternatively, bioinformatic approaches that attempt to account for cellular heterogeneity in a sample are available, although it is not clear how robustly such approaches would perform when considering data from multiple different species [46,47]. Another potential confounder is posed by inter-individual variability. Some genomic loci show considerable epigenetic polymorphism in the human population (A. Sharp, unpublished data), and similar variability in DNA methylation levels is likely present in most species. As a result, studies that use only a small number of individuals could easily confuse such loci with sites of fixed difference between species.

Other hurdles in this field relate to technical effects, including such factors as assay design or cross-species genome mapping. Multiple different technologies now exist for profiling DNA methylation [48], and these present specific challenges when used on divergent genomes. Probe-based assays, such as microarrays, perform sub-optimally when hybridized to DNA extracted from a species other than that for which the array was designed for, necessitating the removal of data from probes that hybridize to divergent positions. Other assays, such as RRBS, include restriction enzyme digestion, and sequence differences can create altered digestion patterns, creating measurement biases between species that are unrelated to epigenetic state. Finally, although we now have high quality

genome assemblies from dozens of mammals, accurate cross-species comparison can still remain complicated for many regions of the genome.

1.2.6. Conclusions and Future Directions

Although the advent of genome-wide methylome studies has enabled the study of evolution from a new perspective several problems currently exist that hinder progress. Firstly, diverse technologies are used in different studies and often cover different genomic regions, meaning that results obtained from different studies are not directly comparable. The continuing reduction in costs of gold-standard methods for epigenome profiling, such as bisulfite sequencing, should help in this regard. However, standardized analytical approaches are still required to ensure that different datasets can be fully integrated. Further confounders come from the use of very small sample sizes, which can result in polymorphic epigenetic differences within a species being interpreted as fixed species-specific changes.

The evolutionary history of the human epigenome and the phenomena responsible for epigenetic divergence between species remains poorly understood. We expect future work will elucidate the occurrence and significance of trans-generational epigenetic inheritance. However, in order to be effective such studies will require robust experimental design and adequate sample sizes to

ensure accurate conclusions. Another major area of investigation will be the study of the molecular mechanisms underlying how DNA methylation patterns are generated and maintained upon differentiation. In addition, further studies are required to decipher how inter-species methylation differences vary across tissues and their functional relevance. Finally, the significance of non-CpG methylation has not been addressed in evolutionary studies.

We anticipate an upcoming revolution in the field. For example, third-generation sequencing technologies have the capability to directly detect cytosine methylation as well other epigenetic modifications. Future research will likely focus on the integration of different kinds of genomic data, such as novel DNA and/or histone modifications, coding and non- coding RNA and chromatin interactions. We posit that a deeper understanding of the causes and consequences of epigenetic differences between species will inform not only the field of evolution, but will likely also provide important information that will aid in the interpretation of non-coding variation in studies of human disease.

Acknowledgements

T.M-B is supported by ICREA, EMBO YIP 2013 and MICINN BFU2014-55090-P. AJS is supported by NIH grants DA033660, HG006696, HD073731, and MH097018, and research grant 6-FY13-92 from the March of Dimes Foundation. IH-H is supported by AGAUR (Generalitat de Catlunya, Spain) and RG-P by a fellowship from Ministerio de Ciencia e Innovación. We also acknowledge the Barcelona Zoo (Ajuntament de Barcelona) for an award to IHH.

References

1. Sequencing TC, Consortium A. Initial sequence of the chimpanzee genome and comparison with the human genome. 2005;437. doi:10.1038/nature04072
2. King M, Wilson AC. Evolution at Two Levels. 1975;188.
3. McLean CY, Reno PL, Pollen A a, Bassan AI, Capellini TD, Guenther C, et al. Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature. Nature Publishing Group; 2011;471: 216–9. doi:10.1038/nature09774
4. Pollard KS, Salama SR, Lambert N, Lambot M-A, Coppens S, Pedersen JS, et al. An RNA gene expressed during cortical development evolved rapidly in humans. Nature. 2006;443: 167–172. doi:10.1038/nature05113
5. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, et al. Ultraconserved elements in the human genome. Science. 2004;304: 1321–1325. doi:10.1126/science.1098119
6. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, et al. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 2005;3. doi:10.1371/journal.pbio.0030007
7. Jones P a. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. Nature Publishing Group; 2012;13: 484–92. doi:10.1038/nrg3230
8. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature. 2007;447: 425–32. doi:10.1038/nature05918
9. Bird A. DNA methylation patterns and epigenetic memory. Genes and Development. 2002. pp. 6–21. doi:10.1101/gad.947102
10. Robertson KD. DNA methylation and human disease. Nat Rev Genet. 2005;6: 597–610. doi:10.1038/nrg1655
11. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. Nature Publishing Group; 2010;28: 1057–68. doi:10.1038/nbt.1685
12. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. PLoS One. 2013;8. doi:10.1371/journal.pone.0063812
13. Joubert BR, Håberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. Environ Health Perspect. 2012;120: 1425–1431. doi:10.1289/ehp.1205412
14. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation Into Cancer And Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013;22: 843–851. doi:10.1093/hmg/dds488

15. Richmond RC, Simpkin AJ, Woodward G, Gaunt TR, Lyttleton O, McArdle WL, et al. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). Hum Mol Genet. 2014;44: 1–39. doi:10.1093/hmg/ddu739
16. Flores KB, Wolschin F, Amdam G V. The role of methylation of DNA in environmental adaptation. Integrative and Comparative Biology. 2013. pp. 359–372. doi:10.1093/icb/ict019
17. Feng S. Epigenetic reprogramming in plant and animal development. Science (80-). 2010;330: 622–627. doi:10.1126/science.1190614.Epigenetic
18. Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet. 1999;23: 314–318. doi:10.1038/15490
19. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. Trends in Genetics. 2002. pp. 348–351. doi:10.1016/S0168-9525(02)02709-9
20. Dias BG, Ressler KJ. Parental olfactory experience influences behavior and neural structure in subsequent generations. Nat Neurosci. 2014;17: 89–96. doi:10.1038/nn.3594
21. Burggren WW. Dynamics of epigenetic phenomena: inter- and intra-generational phenotype "washout". J Exp Biol. 2014; 80–87. doi:10.1242/jeb.107318
22. Schübeler D. Function and information content of DNA methylation. 2015; doi:10.1038/nature14192
23. Polak P, Arndt PF. Long-range bidirectional strand asymmetries originate at CpG islands in the human genome. Genome Biol Evol. 2009;1: 189–197. doi:10.1093/gbe/evp024
24. Feinberg AP, Irizarry R a. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc Natl Acad Sci U S A. 2010;107 Suppl : 1757–64. doi:10.1073/pnas.0906183107
25. Bell CG, Wilson GA, Butcher LM, Roos C, Walter L, Beck S. Human-specific CpG "beacons" identify loci associated with human-specific traits and disease. Epigenetics. 2012;7: 1188–1199. doi:10.4161/epi.22127
26. Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res. 2010;20: 883–889. doi:10.1101/gr.104695.109
27. Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, et al. Genetic Control of Individual Differences in Gene-Specific Methylation in Human Brain. The American Journal of Human Genetics. 2010. pp. 411–419. doi:10.1016/j.ajhg.2010.02.005
28. Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, et al. DNA methylation contributes to natural human variation. Genome Res. 2013;23: 1363–72. doi:10.1101/gr.154187.112

29. Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, et al. Methylation QTLs Are Associated with Coordinated Changes in Transcription Factor Binding, Histone Modifications, and Gene Expression Levels. Reddy TE, editor. PLoS Genet. 2014;10: e1004663. doi:10.1371/journal.pgen.1004663
30. Zhi D, Aslibekyan S, Irvin MR, Claas SA, Borecki IB, Ordovas JM, et al. SNPs located at CpG sites modulate genome-epigenome interaction. Epigenetics. 2013;8: 802–806. doi:10.4161/epi.25501
31. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011;480: 490–5. doi:10.1038/nature10716
32. Duncan BK, Miller JH. Mutagenic deamination of cytosine residues in DNA. Nature. 1980;287: 560–561. doi:10.1038/287560a0
33. Carbone L, Harris RA, Vessere GM, Mootnick AR, Humphray S, Rogers J, et al. Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. PLoS Genet. 2009;5. doi:10.1371/journal.pgen.1000538
34. Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C, et al. Dynamics of DNA methylation in recent human and great ape evolution. PLoS Genet. 2013;9: e1003763. doi:10.1371/journal.pgen.1003763
35. Martin DIK, Singer M, Dhahbi J, Mao G, Zhang L, Schroth GP, et al. Phyloepigenomic comparison of great apes reveals a correlation between somatic and germline methylation states. Genome Res. 2011;21: 2049–57. doi:10.1101/gr.122721.111
36. Xiao S, Xie D, Cao X, Yu P, Xing X, Chen CC, et al. Comparative epigenomic annotation of regulatory DNA. Cell. 2012;149: 1381–1392. doi:10.1016/j.cell.2012.04.029
37. Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, et al. Sperm Methylation Profiles Reveal Features of Epigenetic Inheritance and Evolution in Primates. Cell. Elsevier Inc.; 2011;146: 1029–1041. doi:10.1016/j.cell.2011.08.016
38. Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, et al. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. Elife. 2013;2013. doi:10.7554/eLife.00348
39. Cohen NM, Kenigsberg E, Tanay A. Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. Cell. 2011;145: 773–786. doi:10.1016/j.cell.2011.04.024
40. Wilson G a, Butcher LM, Foster HR, Feber A, Roos C, Walter L, et al. Human-specific epigenetic variation in the immunological Leukotriene B4 Receptor (LTB4R/BLT1) implicated in common inflammatory diseases. Genome Med. 2014;6: 19. doi:10.1186/gm536
41. Pai A a, Bell JT, Marioni JC, Pritchard JK, Gilad Y. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human

and chimpanzee tissues. PLoS Genet. 2011;7: e1001316. doi:10.1371/journal.pgen.1001316
42. Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat Genet. 2013;45: 1198–206. doi:10.1038/ng.2746
43. Gokhman D, Lavi E, Prüfer K, Fraga MF, Riancho JA, Kelso J, et al. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. Science. 2014;344: 523–7. doi:10.1126/science.1250368
44. Goodman FR. Limb malformations and the human HOX genes. American Journal of Medical Genetics. 2002. pp. 256–265. doi:10.1002/ajmg.10776
45. Fraser HB, Lam LL, Neumann SM, Kobor MS. Population-specificity of human DNA methylation. Genome Biol. 2012;13: R8. doi:10.1186/gb-2012-13-2-r8
46. Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J. Epigenome-wide association studies without the need for cell-type composition. Nat Methods. 2014;11: 309–11. doi:10.1038/nmeth.2815
47. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13: 86. doi:10.1186/1471-2105-13-86
48. Bock C. Analysing and interpreting DNA methylation data. Nature Reviews Genetics. 2012. pp. 705–719. doi:10.1038/nrg3273

1.3. Detecting DNA methylation

First locus-specific studies were based on restriction enzymes. These methods relied on digestion with methylation-sensitive enzymes, followed by gel electrophoresis and southern blot. Over the last decade, the development of techniques for profiling DNA methylation has been particularly active, making possible to obtain genome-scale epigenetic information affordably and efficiently. To date, methylated cytosines cannot be detected directly. In addition, methyl groups are erased during PCR amplification. To overcome these limitations several indirect approaches have emerged. These can be classified into bisulfite conversion, immunoprecipitation and restriction enzyme methods [this section is reviewed in (Laird 2010; Plongthongkum et al. 2014; Bock 2012)

1.3.1. Bisulfite conversion

Bisulfite conversion methods are based on an initial chemical treatment of the DNA. Here, unmethylated cytosines are converted into uracils by deamination, while methylated cytosine residues remain intact. In subsequent PCR reactions uracils are amplified as thymines and methylated cytosines as cytosines. Therefore, the methylation status can be assessed as sequence differences. Nowadays two genome-scale approaches based on bisulfite treatment are widely used: array hybridization and sequencing.

a) Array hybridization

Bisufite treatment followed by array hybridization makes possible to measure methylation levels at a preselected fraction of cytosines of the genome. The low sequence complexity of bisulfite treated DNA requires a specific array designed based on the converted sequence. One of the most popular arrays is the HumanMethylation450 BeadChip that interrogates more than 485,000 CpG sites per sample. Here, after DNA fragmentation, samples are hybridized to specific 50 bp probes. The array contains two types of probes: Infinium I and II. Infinium I technology employs two different probe designs for each CpG site. These probes are identical except the last base that interrogates the CpG site. This last nucleotide can be either a guanine (complementing a cytosine) or an adenine (complementing a thymine). A fluorescent single base extension takes place only when the last nucleotide is hybridized, allowing to determine the methylation state. Infinium II technology utilizes a single probe design and two color extension occurring differentially at thymine or cytosine of converted DNA. The methylation state is determined based on the different color intensities (Figure 5). HumanMethylation450 BeadChip is a powerful technique in terms of reagent costs and time. However only a small fraction of the genome is analyzed (< 1%) and cross hybridization remains a primary source of bias. In addition, there are no commercial arrays for the analysis of non human samples.

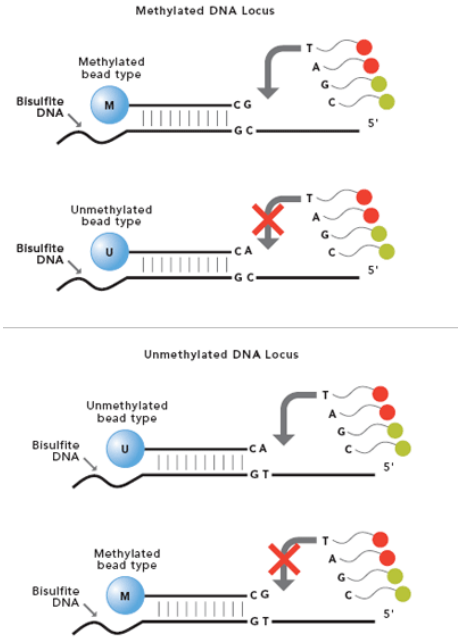

Figure 9. Infinium Methylation Assay. The unmethylated CpG target site matches with the unmethylated probe, enabling single-base extension (top figure). If the CpG locus is methylated, it will matches the methylated probe (bottom figure).

b) Sequencing

Whole genome bisulfite sequencing (WGBS) is considered the gold standard method because it provides single base-pair resolution. It can determine the pattern of DNA methylation across the entire genome, except certain repetitive regions where reads cannot align uniquely. Bisulfite conversion can be performed on prepared

shotgun sequencing libraries, in this case adapters cannot contain unmethylated cytosines. Otherwise, libraries can be prepared after bisulfite treatment, what requires a polymerase able to read uracil nucleotides. After sequencing, reads need to be aligned accounting for nucleotide changes caused by the bisulfite treatment. Two strategies have been adopted to cope with this problem: aligners where Cs can match both Cs and Ts, and three letter aligners. The latter replaces all Cs into Ts and the alignment is exclusively based on three nucleotides. Aligners using this approach (such as Bismark or BS-seeker) are the most widely used due to their high accuracy. Afterwards methylation levels are inferred from the proportion of Cs and Ts at each CpG site.

1.3.2. Methylated DNA immunoprecipitation

This assay is based in the identification of methylated cytosines. It consists of isolating DNA fragments using a monoclonal antibody specific for 5mC. Immunoprecipitation is followed by hybridization (MeDIP-chip) or sequencing (MeDIP-seq). Disadvantages of this strategy include the need for substantial bioinformatic adjustments and the lack of base pair resolution.

1.3.3. Future directions.

Over the past years, DNA methylation technology has undergone an impressive transformation. Recent advances in the field are based

on increasing accuracy and throughput, and lowering sample input and costs. Single cell technology is one of the most significant advances due to the dynamic nature of cytosine methylation. It has been shown that methylation patterns differ from cell type to cell type and they can even vary within the same cell population. Therefore, cellular composition can contribute to methylation differences between samples making difficult to interpret the results. The analysis of DNA methylation at single cell level is a promising strategy to understand cell heterogeneity and epigenetic regulation. In this regard, bisulfite based methods were considered incompatible with this technology because most of the DNA is lost or damaged after bisulfite treatment. However, recent improvements in bisulfite treatment protocols have enabled to obtain genome-wide coverage from one single cell (Smallwood et al. 2014).

Third generation technology is also a promising strategy. It has raised the possibility of detecting DNA modification directly without previous chemical or enzymatic steps. The single-molecule real-time (SMRT) sequencing approach developed by Pacific Biosciences enables to detect 5mC based on nucleotide incorporation time. Moreover, other modifications such as 5hmC can also be detected. Nevertheless, many aspects still need to be improved before this technology can be used in epigenetic research.

2. OBJETIVES

Objetives

1-Determine the accuracy of the Infinium HumanMethylation450 BeadChip for profiling non-human methylomes.

2-Obtain a genome-scale description of DNA methylation patterns in humans and great apes using both HumanMethylation450 BeadChip and whole genome bisulfite sequencing data.

3-Identify and characterize differentially methylated regions, particularly in humans.

4-Explore the relationship between lineage-specific evolutionary changes in the genome and the epigenome.

3. RESULTS

Chapter 1

3.1. Dynamics of DNA methylation in Recent Human and Great Ape Evolution

---

Irene Hernando-Herraez[1], Javier Prado-Martinez[1], Paras Garg[2], Marcos Fernandez-Callejo[1], Holger Heyn[3], Christina Hvilsom[4], Arcadi Navarro[1,6], Manel Esteller[3,5,6], Andrew J. Sharp[2,*], Tomas Marques-Bonet[1,6,*]

Published

Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C, et al. **Dynamics of DNA methylation in recent human and great ape evolution. PLoS Genet. 2013;9: e1003763.**

PLoS Genet. 2013;9(9):e1003763. doi: 10.1371/journal.pgen.1003763

Chapter 2

3.2. The interplay between DNA methylation and sequence divergence in recent human evolution

---

Irene Hernando-Herraez[1#], Holger Heyn[2#], Marcos Fernandez-Callejo[3], Enrique Vidal[2], Hugo Fernandez-Bellon[4], Javier Prado-Martinez[1], Andrew J Sharp[5], Manel Esteller[2,6] and Tomas Marques-Bonet[1,3,6].

***Nucleic Acids Research,* under review.**

[1]Institute oí Evolutionary Biology (UPF-CSIC), PRBB, 08003 Barcelona, Spain.

[2]Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08908 L'Hospitalet de Llobregat, Barcelona, Spain.

[3]Centro Nacional de Análisis Genómico (CNAG), Parc Científic de Barcelona, Barcelona 08028, Spain.

[4]Parc Zoològic de Barcelona, Parc de la Ciutadella s/n, Barcelona 08003, Spain

[5]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai School, New York, NY 10029, USA.

[6]Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain

Correspondence should be addressed to: irene.hernando@upf.edu tomas.marques@upf.edu and mesteller@idibell.cat

Hernando-Herraez I, Heyn H, Fernandez-Callejo M, Vidal E, Fernandez-Bellon H,Prado-Martinez J, Sharp AJ, Esteller M, Marques-Bonet T. The interplay between DNA methylation and sequence divergence in recent human evolution. Nucleic Acids Res. 2015 Sep 30;43(17):8204-14. doi: 10.1093/nar/gkv693.

4. DISCUSSION

Epigenetics is undergoing a true revolution and it is influencing vast areas of biology. The number of publications with the word "epigenetic" in their title has increased 10 fold in the last decade. Ambitious projects, such as the International Human Epigenome Consortium or the NIH Roadmap Epigenomics Mapping Consortium, analyzing human epigenomes for a variety of cell types and disease states, have been launched worldwide. The Nobel prize in Physiology or Medicine 2012 to Gurdon and Yamanaka for their discovery of induced pluripotent cells has been also a recognition to epigenetics. However, some of the most fundamental issues including how specific modifications are established at selected positions of the genome remain largely unknown.

Evolutionary comparisons have demonstrated to be a powerful tool to study genome functions. An excellent example is the identification of human regulatory regions based on sequence conservation among distant taxa. Motivated by the success of this approach the field of comparative epigenomics has recently emerged. In this framework the analysis of epigenetic modifications from different species or populations can be used to provide important information about the functionality of our epigenome. Nonetheless, we first need to determine the general evolutionary properties of our epigenome as well as its relationship with the genome.

Furthermore, from an evolutionary perspective, comparative epigenomic studies can provide new insights into the basis of species phenotypic differences. The idea that regulatory modifications and subsequent gene expression changes might underlie the molecular basis of human-specific traits is not new. This hypothesis has been debated since 1970s (Davidson and Britten 1979; King and Wilson 1975), however due to technological limitations, it could not be thoroughly explored. The advent of high-throughput sequencing technologies has allowed the study of species evolution from a new angle. To date, numerous studies have examined the evolutionary features of the transcriptome (Necsulea and Kaessmann 2014) . As can be expected, expression levels differ more between tissues from the same species (e.g. human heart versus human liver) than between homologous tissues from different species (e.g. human heart versus chimpanzee heart). However, inter-species comparisons have shown that not all tissues present the same level of expression divergence: in neural tissues, expression profiles evolve slowly at both coding and non-coding level while testis exhibit a particular high rate of transcriptome evolution. Overall, although certain genes and pathways have evolved under positive selection in different tissues, mammalian expression levels are largely conserved (Brawand et al. 2011).

While most efforts have focused on characterizing transcriptome evolution, very few studies have investigated regulatory mechanisms from an evolutionary perspective. In particular, DNA

methylation patterns across species are surprisingly unexplored and many questions await future research. Which are the evolutionary traces of the human methylome? Are epigenetic changes a mere consequence of genomic variation? To which extent has DNA methylation influenced the evolution of the genome?

In this thesis I have performed one of the first and most complete methylome comparisons between human and great apes using two different approaches. In the first study, the Infinium HumanMethylation450 BeadChip technology was used to determine DNA methylation levels in a total of 32 individuals, representing all species of great apes. This study was particularly challenging since the 50 bp probes on the Infinium array were designed to interrogate human samples. To overcome this technical limitation, all probes that hybridize to divergent or multiple positions of the species genomes were removed. As a result, only a fourth of the data was retained (~100.000 CpG sites) and most of these sites were located in genomic regions conserved between species. In the second study, I analyzed the entire genomes from one individual per species using to that effect whole genome bisulfite sequencing technology. This reduced sample size was a major constraint that has limited the study. Future research using adequate sample sizes will be required to minimize confounding factors, such as intra-species variability. Taken together, in this thesis I have provided an evolutionary perspective to our epigenome and showed that species-specific epigenetic alterations are an

important source of variation that has been under-explored in evolutionary comparative studies.

One of the first steps to perform accurate inter-species analyses, is to define which regions of the species genomes are comparable. In the first study, probe sequences were mapped to the species reference genomes and only those probes that had either a perfect match or up to 2 mismatches (not affecting the CpG site) were retained. Afterwards, all coordinates and functional annotations were based on the human hg19 reference genome. In the second study, all comparisons were carried out using pre-computed whole genome alignments (6-primate EPO)(Paten et al. 2008). This approach facilitated an unbiased comparison between the four species and, in comparison with previous studies, it allowed to reliably analyze a greater proportion of the species epigenomes. However, due to technical limitations some interesting regions, such as most transposable elements, were not covered. Transposable elements are CpG dense regions and represent 44% of the human genome. Interestingly, although they are presumed to be inactive via DNA methylation, in humans 10% of transposable elements subfamilies are hypomethylated in a tissue-specific manner and present enhancer signatures (Xie et al. 2013). However, since bisulfite treatment reduces the complexity of the DNA to three nucleotide bases, most of transposable elements were covered by multi-mapping reads that were subsequently discarded. In the near future, technological advances will overcome this limitation.

Particularly, the advent of longer reads and even the identification of cytosine modification without chemical treatments from third-generation sequencing technologies (Plongthongkum et al. 2014) . These advances will likely facilitate the analysis of complex and repetitive regions of the genome by improving the rate of uniquely mapped reads .

In this thesis it has been shown that global patterns of DNA methylation were similar between humans and other non-human great apes. DNA methylation levels presented a bimodal distribution along the genomes which are largely methylated except most CpG rich promoter regions. Since unmethylated cytosines mutate less frequently than methylated ones, this unmethylated state could explain the absence of CpG erosion at promoter regions as has been proposed by Cohen et al., instead of being the result of a selective pressure (Cohen et al. 2011). Interestingly, the unmethylated state of CpG rich promoter regions has also been identified by other studies covering larger evolutionary scales (Long et al. 2013), demonstrating that the absence of methylation in these regions is a highly conserved feature in evolution.

The identification of human- specific methylation patterns was one of the main objectives of this work. Despite the global conservation of the species methylomes, I have identified and characterized hundreds of differentially methylated regions (DMRs) in human blood samples. Moreover, most human DMRs were enriched for

species-specific histone modifications and contrary to expectations, they were located distal to transcription start sites, highlighting the importance of regions outside the close regulatory context. Remarkably, studies analyzing different human cell types have shown that tissue regulation is also mediated by distal regions to transcription start sites (Ziller et al. 2013). Therefore, future research should be focused on investigating in depth this important, but unknown, part of the genome. In this regard, Hi-C technology, which determines the spacial organization of the genome, will likely help to elucidate the significance of distal regulatory regions.

In both studies, methylation profiles were obtained from whole blood samples. Since each cell type possesses its own epigenome, it is likely that differences in blood cellular composition between species might introduce a bias. To overcome this limitation, several stringent filters have been used resulting in a significant loss of information. Therefore, although collecting samples from endangered species is a difficult task, comparative studies should be performed in purified single cell types .

Interestingly, I estimated that ~25% of human DMRs in blood were detectable throughout several human tissues. I believe this is a highly relevant finding that should be explored in detail in future research. Differentially methylated regions between species that are conserved across tissues could be regulatory elements active during the embryonic development but dormant in adult tissues. Indeed,

studies in mice have identified certain developmental enhancers that remain unmethylated in adult tissues. Nevertheless, in adult tissues they are inactive and marked with repressive histone modifications (Hon et al. 2013).This epigenetic memory could potentially be used to unravel the developmental decisions made during embryognesis, a period that is key to the acquisition of species-specific traits. Furthermore, it could help to understand why some human DMRs coincide not only with tissue-specific functions but also with developmental and neurological processes. One example of human DMR identified in this work that was detectable in human brain, liver and placenta is the 5'UTR of the TGIF1 gene, a transcription factor of the evolutionarily conserved TALE homeobox family involved in embryonic development. Interestingly, mutations in this gene have been associated with structural anomalies of the brain(El-Jaick et al. 2007). Therefore, future research covering several tissues and developmental stages from different species will be crucial to better understand the significance of inter-species methylation differences, and particularly inter-species differences that are conserved across tissues.

Functional consequences of epigenetic divergence between species are also poorly understood. While histone modifications are directly associated with transcriptional activity, the consequences of DNA methylation divergence are not so clear. It has been found that DNA methylation differences between species mildly influences gene expression levels and in a way that is independent from histone

modifications (Xiao et al. 2012). Indeed, direct evidence suggested that differences in promoter methylation underlie only 12-18% of gene expression differences between human and chimpanzee(Pai et al. 2011). This finding is in agreement with values obtained from human populations (Heyn et al. 2013). Nevertheless it is important to notice that these studies were based on a limited number of CpG sites (~ 10,000) and thus, whole genome datasets are needed to fully understand to which extent methylation differences are involved in regulating expression levels. In addition, RNA sequencing analyses have revealed that most mammalian protein-coding genes have multiple splicing variants, many of which are alternatively spliced in a lineage-specific manner. However, until now no studies have attempted to jointly analyze DNA methylation and alternative splicing patterns.

The analysis of ~30.000 CpG sites in this work, proved that methylation levels recapitulate phylogenetic relationships between species, even at subspecies level. Interestingly, this finding has also been observed when comparing different human populations (Heyn et al. 2013). Although the most likely explanation for this phylogenetic clustering is the dependence of methylation levels on the underlying genome sequence, other scenarios such as separate environments cannot be discarded.

Deciphering the causes of DNA methylation patterns is a central goal not only for evolutionary biology but also for many areas such

as biomedical research. Inter-species epigenetic divergence can be a consequence of sequence-independent mechanisms, including environmental factors, circadian rhythms or stochastic events. Studies in model organisms typically control external or genetic factors minimizing their effects. However, in non-models organisms, notably in primates, external factors are often impossible to exclude and therefore, differences in DNA methylation between species may be a consequence of environmental factors on that lineage (for example the use of cooked food in humans (Finch and Stanford 2004)). In addition, although controversial, epigenetic differences between species could be attributed to pure epigenetic causes, that is, epigenetic modifications that are transmitted across generations. Although to date most evidence suggest that trans-generational epigenetic marks are erased after a few generations many scenarios have not been addressed yet. For example, a poorly explored aspect is to what extent epigenetic alterations can be transmitted when the initial trigger is presence over multiple generations. It is important to note that transgenerational effects of epigenetic changes will be one of the areas with most impact on human health over the coming decades because the effects of pollutants, drugs or food. Moreover, if it is proved that this phenomenon occurs in natural populations, we will face a complete shift of paradigm in evolutionary biology.

In addition, it is assumed that a significant proportion of DNA methylation variation observed among individuals or across species

can be explained by corresponding genetic changes. Genetic influences can arise from key genes maintaining the epigenetic profile or from the neighboring genetic variants affecting the epigenetic state, denominated mQTLs. The association between epigenetic modifications and genetic variants between species or individuals will be crucial to interpret the significance of non-coding variation as well as to understand the evolution of complex phenotypes. In the recent years, several evolutionary studies have begun to investigate these questions. In human populations, thousands of mQTLs have been identified so far and evidence suggests that a single mQTL modulates the methylation state of several CpG sites (Banovich et al. 2014; Heyn et al. 2013). Nevertheless, the functional role of mQTLs is still poorly understood.

In this thesis I have analyzed for the first time the relationship between the genetic and the epigenetic code at two different levels: protein and nucleotide sequence. At the protein level, I have identified a significant positive relationship between the rate of amino acid changes within genes and DNA methylation alteration at promoter regions, suggesting a coupled evolution between gene regulation and protein sequence. Furthermore, this analysis also identified 184 genes that are perfectly conserved at the amino acid level between human and chimpanzees, but show DNA methylation changes at promoter regions between these two species. This finding highlights the importance of evolutionary regulatory

changes, and also indicates that studies based uniquely in DNA or protein sequence are insufficient to capture the complete spectrum of evolutionary variation.

At nucleotide level, I have shown, for the first time, a close relationship between inter-species genetic and epigenetic variation supported by three different analyses. Firstly, regions of incomplete lineage sorting between human, chimpanzee and gorilla presented also strong incomplete lineage sorting at methylation level. Secondly, an excess of human-specific substitutions was observed in transcription factor binding sites located within human differentially methylated regions, suggesting that changes within regulatory motifs underly some human DMRs. Thirdly, the acquisition of DNA hypermethylation in the human lineage is frequently coupled with a rapid evolution at nucleotide level in the neighborhood of these CpG sites. It is important to notice that the association between DNA methylation levels and fixed genetic variants when comparing different species is a challenging task due to the high amount of genomic variation between species. The identification of inter-species meQTLs will require large sample sizes and the combination of individual genetic and epigenetic data sets.

CONCLUDING REMAKRS

This thesis provides important evolutionary insights into the human epigenome. General features of the human methylome have been described and hundreds of human DMRs have been characterized at both epigenetic and genetic level. These regions would be promising candidates for further functional studies. The next step to complete an accurate and detailed evolutionary picture of our epigenome is to obtain and to analyze methylomes from several individuals, tissues and developmental stages. In addition, many types of data such as chromatin modifications, maps of nucleosome positions, different markers of enhancer elements as well as coding and non-coding transcriptional information are required to elucidate the complexity of regulatory mechanisms and to assess the significance of evolutionary changes. Furthermore, parallel surveys of interspecies differences in genetic and regulatory elements will be key to interpret non-coding variation as well as to understand the underlying mechanisms of regulatory changes. Therefore the challenge for the upcoming years is to integrate different sources of genomic, epigenomic and transcriptomic data in a unified model of gene regulation.

Finally, the most appealing goal not only for evolutionary but also for many other biological studies, is to move beyond comparative descriptions to complex phenotypes. To assess the impact of regulatory changes at the organism level direct experimentation is

required. However, due to ethical reasons, the only approach feasible in primates is to perform functional experiments in model organisms. Although a handful of functional examples have been characterized (Prabhakar et al. 2008), these approaches are laborious and their results are often difficult to interpret. In this regard, the generation of iPS cells from primates with the aim of differentiating them into somatic cells will likely represent one of the most promising advances in the field of comparative epigenomics (Romero et al. 2015).

List of communications

Hernando-Herraez I, Prado-Martinez J, Garg P, et al. Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet*. 2013; 9(9):e1003763

Heyn H, Moran S, **Hernando-Herraez I**, et al. DNA methylation contributes to natural human variation. *Genome Res*. 2013 Sep; 23(9):1363-72

Hormozdiari F, Konkel MK, Prado-Martinez J, et al, Rates and patterns of great ape retrotransposition. *Proc Natl Acad Sci U S A. 2013 Aug 13;110(33):13457-62*

Prado-Martinez J, Sudmant PH, Kidd JM, et al. *G*reat ape genetic diversity and population history. *Nature.* 2013 Jul 25;499(7459):471-5

Prado-Martinez J, **Hernando-Herraez I,** Lorente-Galdos B, et al. The genome sequencing of an albino Western lowland gorilla reveals inbreeding in the wild. *BMC Genomics.* 2013 May 31;14:363

Hernando-Herraez I, Heyn H, Fernandez-Callejo M, et al. The interplay between DNA methylation and sequence divergence in recent human evolution. *Nucleic Acids Research,* under review.

Hernando-Herraez I, Garcia-Perez R, Sharp AJ & Marques-Bonet T. Comparative epigenetics: Patterns of CpG methylation in human evolution. *PLoS Genet,* under review.

Gallego-Romero I, Pavlovic JB, **Hernando-Herraez I,** et al. Generation of a Panel of Induced Pluripotent Stem Cells From Chimpanzees: a Resource for Comparative Functional Genomics. *Elife,* under review.

Bibliography

Aapola U, Kawasaki K, Scott HS, Ollila J, Vihinen M, Heino M, Shintani A, Kawasaki K, Minoshima S, Krohn K, et al. 2000. Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics* **65**: 293–298.

Anway MD, Anway MD, Cupp AS, Cupp AS, Uzumcu M, Uzumcu M, Skinner MK, Skinner MK. 2005. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**: 1466–9.

Aran D, Toperoff G, Rosenberg M, Hellman A. 2011. Replication timing-related and gene body-specific methylation of active human genes. *Hum Mol Genet* **20**: 670–680.

Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H, Xie B, Daley GQ, Church GM. 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* **27**: 361–368.

Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, Roux J, Pritchard JK, Gilad Y. 2014. Methylation QTLs Are Associated with Coordinated Changes in Transcription Factor Binding, Histone Modifications, and Gene Expression Levels ed. T.E. Reddy. *PLoS Genet* **10**: e1004663. http://dx.plos.org/10.1371/journal.pgen.1004663 (Accessed September 18, 2014).

Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3439153&tool=pmcentrez&rendertype=abstract (Accessed July 9, 2014).

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. 2006. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**: 315–326.

Bestor TH. 1992. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J* **11**: 2611–2617.

Bock C. 2012. Analysing and interpreting DNA methylation data. *Nat Rev Genet* **13**: 705–719.

Bostick M, Kim JK, Estève P-O, Clark A, Pradhan S, Jacobsen SE. 2007. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**: 1760–1764.

Branco MR, Ficz G, Reik W. 2011. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet*.

Brawand D, Soumillon M, Necsulea a, Julien P, Csardi G, Harrigan P, Weier M, Liechti a, Aximu-Petri a, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* **478**: 343–348. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=22012392.

Brenner C, Deplus R, Didelot C, Loriot A, Viré E, De Smet C, Gutierrez A, Danovi D, Bernard D, Boon T, et al. 2005. Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J* **24**: 336–346.

Chen P-Y, Feng S, Joo JWJ, Jacobsen SE, Pellegrini M. 2011. A comparative analysis of DNA methylation across human embryonic stem cell lines. *Genome Biol* **12**: R62. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3218824&tool=pmcentrez&rendertype=abstract.

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF. 1997. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* **277**: 1996–2000.

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

Compere SJ, Palmiter RD. 1981. DNA methylation controls the inducibility of the mouse metallothionein-I gene lymphoid cells. *Cell* **25**: 233–240.

Cooper DN, Mort M, Stenson PD, Ball E V, Chuzhanova N a. 2010. Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Hum Genomics* **4**: 406–410.

Court F, Tayama C, Romanelli V, Martin-Trujillo A, Iglesias-Platas I, Okamura K, Sugahara N, Simón C, Moore H, Harness J V, et al. 2014. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. *Genome Res* **24**: 554–69. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3975056&tool=pmcentrez&rendertype=abstract (Accessed October 29, 2014).

Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**: 157–161.

Darwin C. 1871. The Descent of Man, and Selection in Relation to Sex. *D Appleton, New York*.

Davidson E, Britten R. 1979. Regulation of gene expression: possible role of repetitive sequences. *Science (80-)* **204**: 1052–1059. http://www.sciencemag.org/content/204/4397/1052.short.

De Waal FBM. 2005. A century of getting to know the chimpanzee. *Nature* **437**: 56–59.

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. 2010. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* **285**: 26114–26120.

Dias BG, Ressler KJ. 2014a. Experimental evidence needed to demonstrate inter- and trans-generational effects of ancestral

experiences in mammals. *BioEssays* n/a–n/a. http://doi.wiley.com/10.1002/bies.201400105.

Dias BG, Ressler KJ. 2014b. Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat Neurosci* **17**: 89–96. http://www.ncbi.nlm.nih.gov/pubmed/24292232.

Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature* **287**: 560–561.

Easwaran HP, Schermelleh L, Leonhardt H, Cardoso MC. 2004. Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep* **5**: 1181–1186.

Eilertsen KJ, Power RA, Harkins LL, Misica P. 2007. Targeting cellular memory to reprogram the epigenome, restore potential, and improve somatic cell nuclear transfer. *Anim Reprod Sci* **98**: 129–146.

El-Jaick KB, Powers SE, Bartholin L, Myers KR, Hahn J, Orioli IM, Ouspenskaia M, Lacbawan F, Roessler E, Wotton D, et al. 2007. Functional analysis of mutations in TGIF associated with holoprosencephaly. *Mol Genet Metab* **90**: 97–111. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1820763&tool=pmcentrez&rendertype=abstract (Accessed October 29, 2014).

Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**: 215–6. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3577932&tool=pmcentrez&rendertype=abstract (Accessed July 11, 2014).

Farcas R, Schneider E, Frauenknecht K, Kondova I, Bontrop R, Bohl J, Navarro B, Metzler M, Zischler H, Zechner U, et al. 2009. Differences in DNA methylation patterns and expression of the CCRK gene in human and nonhuman primate cortices. *Mol Biol Evol* **26**: 1379–89. http://www.ncbi.nlm.nih.gov/pubmed/19282513 (Accessed December 3, 2014).

Feil R, Fraga MF. 2012. Epigenetics and the environment: emerging patterns and implications. *Nat Publ Gr* **13**: 97–109. http://dx.doi.org/10.1038/nrg3142.

Feinberg AP, Irizarry R a. 2010. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A* **107 Suppl** : 1757–64. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2868296&tool=pmcentrez&rendertype=abstract (Accessed July 11, 2014).

Ferguson-Smith AC, Patti ME. 2011. You are what your dad ate. *Cell Metab* **13**: 115–117.

Finch CE, Stanford CB. 2004. Meat-adaptive genes and the evolution of slower aging in humans. *Q Rev Biol* **79**: 3–50.

Fuks F, Burgers WA, Godin N, Kasai M, Kouzarides T. 2001. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J* **20**: 2536–2544.

Gokhman D, Lavi E, Prüfer K, Fraga MF, Riancho JA, Kelso J, Pääbo S, Meshorer E, Carmel L. 2014. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* **344**: 523–7. http://www.ncbi.nlm.nih.gov/pubmed/24786081 (Accessed December 3, 2014).

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74**: 481–514.

Guibert S, Forné T, Weber M. 2012. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res* **22**: 633–641.

Hackett J a, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down T a, Surani MA. 2013. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**: 448–52.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3847602&tool=pmcentrez&rendertype=abstract.

Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, et al. 2013. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol Cell* **49**: 359–367.

Hata K, Okano M, Lei H, Li E. 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129**: 1983–1993.

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 U S A* **105**: 17046–17049.

Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C, Navarro A, Esteller M, Sharp AJ, Marques-Bonet T. 2013. Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet* **9**: e1003763. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3764194&tool=pmcentrez&rendertype=abstract (Accessed October 28, 2014).

Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-mut J V, Setien F, Carmona FJ, et al. 2012a. Distinct DNA methylomes of newborns and centenarians.

Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut J V., Setien F, Carmona FJ, et al. 2012b. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci* **109**: 10522–10527.

Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, Monk D, Hata K, Marques-Bonet T, Wang L, et al. 2013. DNA methylation contributes to natural human variation. *Genome Res* **23**: 1363–72. http://www.pubmedcentral.nih.gov/articlerender.fcgi?

artid=3759714&tool=pmcentrez&rendertype=abstract (Accessed July 16, 2014).

Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, Park J, Butler J, Rafii S, McCombie WR, et al. 2011. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell* **44**: 17–28. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3412369&tool=pmcentrez&rendertype=abstract (Accessed December 1, 2014).

Holliday R, Pugh JE. 1975. DNA modification mechanisms and gene activity during development. *Science* **187**: 226–232.

Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, Ren B. 2013. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* **45**: 1198–206. http://www.ncbi.nlm.nih.gov/pubmed/23995138.

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

Hotchkiss RD. 1948. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem* **175**: 315–332.

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**: 86. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3532182&tool=pmcentrez&rendertype=abstract (Accessed December 5, 2014).

Huang W, Loganantharaj R, Schroeder B, Fargo D, Li L. 2013. PAVIS: A tool for Peak Annotation and Visualization. *Bioinformatics* **29**: 3097–3099.

Huxley TH. 1863. Evidence as to Man's place in Nature. *Williams and Norgate, London,*.

http://www.ncbi.nlm.nih.gov/pubmed/18329365 (Accessed September 13, 2014).

Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP. 2010. Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome. *PLoS Genet* **6**.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**: 1300–1303.

Jiang C, Zhao Z. 2006. Directionality of point mutation and 5-methylcytosine deamination rates in the chimpanzee genome. *BMC Genomics* **7**: 316.

Jones P a. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**: 484–92. http://www.ncbi.nlm.nih.gov/pubmed/22641018 (Accessed July 10, 2014).

Kalisz S, Purugganan MD. 2004. Epialleles via DNA methylation: Consequences for plant evolution. *Trends Ecol Evol* **19**: 309–314.

Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, et al. 2014. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* **42**.

Kelly TK, Miranda TB, Liang G, Berman BP, Lin JC, Tanay A, Jones PA. 2010. H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. *Mol Cell* **39**: 901–911.

King M, Wilson AC. 1975. Evolution at Two Levels. **188**.

Kino K, Sugiyama H. 2005. UVR-induced G-C to C-G transversions from oxidative DNA damage. *Mutat Res - Fundam Mol Mech Mutagen* **571**: 33–42.

Kishikawa S, Murata T, Ugai H, Yamazaki T, Yokoyama KK. 2003. Control elements of Dnmt1 gene are regulated in cell-cycle dependent manner. *Nucleic Acids Res Suppl* 307–308.

Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R, et al. 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**: 839–843.

Kondrashov AS. 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing mendelian diseases. *Hum Mutat* **21**: 12–27.

Kroeze LI, van der Reijden B a, Jansen JH. 2015. 5-hydroxymethylcytosine: An epigenetic mark frequently deregulated in cancer. *Biochim Biophys Acta* **1855**: 144–154. http://www.sciencedirect.com/science/article/pii/S0304419X15000025.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**: 1571–2. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3102221&tool=pmcentrez&rendertype=abstract (Accessed July 17, 2014).

Kucharski R, Maleszka J, Foret S, Maleszka R. 2008. Nutritional control of reproductive status in noneybees via DNA Methylation. *Science (80-)* **319**: 1153069. http://www.sciencemag.org/cgi/content/abstract/1153069v1.

Laird PW. 2010. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* **11**: 191–203.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**: 204–220.

Lee HJ, Hore TA, Reik W. 2014. Reprogramming the methylome: Erasing memory and creating diversity. *Cell Stem Cell* **14**: 710–719.

Lei H, Oh SP, Okano M, Jüttermann R, Goss KA, Jaenisch R, Li E. 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**: 3195–3205.

Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, et al. 2010. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**: 2424–2433.

Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**: 915–926.

Li JB, Gao Y, Aach J, Zhang K, Kryukov G V., Xie B, Ahlford A, Yoon JK, Rosenbaum AM, Zaranek AW, et al. 2009. Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. *Genome Res* **19**: 1606–1615.

Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, Laird PW, Jones PA. 2002. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol Cell Biol* **22**: 480–491.

Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schübeler D. 2011. Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* **43**: 1091–7. http://www.ncbi.nlm.nih.gov/pubmed/21964573 (Accessed July 12, 2014).

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth L V, Muzny DM, Yang S-P, Wang Z, Chinwalla AT, Minx P, et al. 2011. Comparative and demographic analysis of orang-utan genomes. *Nature* **469**: 529–533.

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

Lu X, Sachs F, Ramsay L, Jacques P-É, Göke J, Bourque G, Ng H-H. 2014. The retrovirus HERVH is a long noncoding RNA

required for human embryonic stem cell identity. *Nat Struct Mol Biol* **21**: 423–5. http://www.ncbi.nlm.nih.gov/pubmed/24681886.

Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010. The honey bee epigenomes: Differential methylation of brain DNA in queens and workers. *PLoS Biol* **8**.

Maiti A, Drohat AC. 2011. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *J Biol Chem* **286**: 35334–35338.

Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, et al. 2015. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* **16**: 1–12. http://genomebiology.com/2015/16/1/25.

Martin DIK, Singer M, Dhahbi J, Mao G, Zhang L, Schroth GP, Pachter L, Boffelli D. 2011. Phyloepigenomic comparison of great apes reveals a correlation between somatic and germline methylation states. *Genome Res* **21**: 2049–57. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3227095&tool=pmcentrez&rendertype=abstract (Accessed October 28, 2014).

McLean CY, Reno PL, Pollen A a, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT, et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**: 216–9. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3071156&tool=pmcentrez&rendertype=abstract (Accessed July 11, 2014).

Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. 2011. Sperm Methylation Profiles Reveal Features of Epigenetic Inheritance and Evolution in Primates. *Cell* **146**: 1029–1041. http://linkinghub.elsevier.com/retrieve/pii/S0092867411009421 (Accessed August 27, 2014).

Morgan HD, Sutherland HG, Martin DI, Whitelaw E. 1999. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* **23**: 314–318.

Necsulea A, Kaessmann H. 2014. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat Publ Gr* **23**: 349–357.

Neddermann P, Jiricny J. 1994. Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc Natl Acad Sci U S A* **91**: 1642–1646.

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.

Pai A a, Bell JT, Marioni JC, Pritchard JK, Gilad Y. 2011. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet* **7**: e1001316. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3044686&tool=pmcentrez&rendertype=abstract (Accessed October 8, 2014).

Paten B, Herrero J, Beal K, Fitzgerald S, Birney E. 2008. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res* **18**: 1814–28. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2577869&tool=pmcentrez&rendertype=abstract (Accessed September 30, 2014).

Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. 2011. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res* **21**: 447–455.

Plongthongkum N, Diep DH, Zhang K. 2014. Advances in the profiling of DNA and beyond. *Nat Publ Gr* **15**: 647–661. http://dx.doi.org/10.1038/nrg3772.

Portela A, Esteller M. 2010. Epigenetic modifications and human disease. *Nat Biotechnol* **28**: 1057–68. http://www.ncbi.nlm.nih.gov/pubmed/20944598 (Accessed July 9, 2014).

Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Morrison H, Fitzpatrick DR, Afzal V, et al. 2008. Human-specific gain of function in a developmental enhancer. *Science* **321**: 1346–1350.

Proffitt JH, Davie JR, Swinton D, Hattman S. 1984. 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. *Mol Cell Biol* **4**: 985–988.

Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G, Kodira C, Winer R, et al. 2012. The bonobo genome compared with the chimpanzee and human genomes. *Nature*.

Rakyan VK, Beck S. 2006. Epigenetic variation and inheritance in mammals. *Curr Opin Genet Dev* **16**: 573–577.

Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. 2002. Metastable epialleles in mammals. *Trends Genet* **18**: 348–351.

Ranta S, Zhang Y, Ross B, Takkunen E, Hirvasniemi A, de la Chapelle A, Gilliam TC, Lehesjoki AE. 2000. Positional cloning and characterisation of the human DLGAP2 gene and its exclusion in progressive epilepsy with mental retardation. *Eur J Hum Genet* **8**: 381–4. http://www.ncbi.nlm.nih.gov/pubmed/10854099 (Accessed December 19, 2014).

Reik W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**: 425–32. http://www.ncbi.nlm.nih.gov/pubmed/17522676 (Accessed July 14, 2014).

Reik W, Dean W, Walter J. 2001. Epigenetic reprogramming in mammalian development. *Science* **293**: 1089–1093.

Richards EJ. 2008. Population epigenetics. *Curr Opin Genet Dev* **18**: 221–226.

Richmond RC, Simpkin AJ, Woodward G, Gaunt TR, Lyttleton O, McArdle WL, Ring SM, Smith ADAC, Timpson NJ, Tilling K, et al. 2014. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Hum Mol Genet* **44**: 1–39. http://www.hmg.oxfordjournals.org/cgi/doi/10.1093/hmg/ddu739.

Romero IG, Pavlovic BJ, Hernando-herraez I, Banovich NE, Kagan CL, Burnett JE, Huang CH, Mitrano A, Claudia I, Ben-nun IF, et al. 2014. Generation of a Panel of Induced Pluripotent Stem Cells From Chimpanzees: a Resource for Comparative Functional Genomics.

Sasaki H, Matsui Y. 2008. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* **9**: 129–140.

Saxonov S, Berg P, Brutlag DL. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**: 1412–1417.

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**: 169–75. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3303130&tool=pmcentrez&rendertype=abstract (Accessed July 16, 2014).

Schlesinger F, Smith AD, Gingeras TR, Hannon GJ, Hodges E. 2013. De novo DNA demethylation and noncoding transcription define active intergenic regulatory elements. 1601–1614.

Schneider E, El Hajj N, Richter S, Roche-Santiago J, Nanda I, Schempp W, Riederer P, Navarro B, Bontrop RE, Kondova I, et al. 2014. Widespread differences in cortex DNA methylation of the "language gene" CNTNAP2 between humans and chimpanzees. *Epigenetics* **9**: 533–45. http://www.ncbi.nlm.nih.gov/pubmed/24434791 (Accessed December 3, 2014).

Schübeler D. 2015. Function and information content of DNA methylation.

Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y. 2005. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* **278**: 440–458.

Sequencing TC, Consortium A. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. **437**.

Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, Belvisi MG, Brown R, Vineis P, Flanagan JM. 2013. Epigenome-wide association study in the European Prospective Investigation Into Cancer And Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet* **22**: 843–851.

Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**: 74–9. http://www.ncbi.nlm.nih.gov/pubmed/21964334 (Accessed July 16, 2014).

Simpson VJ, Johnson TE, Hammen RF. 1986. Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res* **14**: 6711–6719.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**: 272–285.

Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G. 2014. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* **11**: 817–20. http://www.ncbi.nlm.nih.gov/pubmed/25042786.

Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu H, et al. 2013. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**: 678–691.

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**: 490–5. http://www.ncbi.nlm.nih.gov/pubmed/22170606 (Accessed July 15, 2014).

Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 2011. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol* **12**: R54.

Taberlay PC, Kelly TK, Liu CC, You JS, De Carvalho DD, Miranda TB, Zhou XJ, Liang G, Jones PA. 2011. Polycomb-repressed genes have permissive enhancers that initiate reprogramming. *Cell* **147**: 1283–1294.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, et al. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930–935.

Teare JA. 2002. 1954- and Int'l Species Information System (Firm) ISIS reference ranges for physiological values in captive wildlife. http://trove.nla.gov.au/work/29865672.

Tobi EW, Goeman JJ, Monajemi R, Gu H, Putter H, Zhang Y, Zwet EW Van, Bock C, Slieker RC, Stok AP, et al. 2014. Exposure To Growth and Metabolism.

Tsankov AM, Gu H, Akopian V, Ziller MJ, Donaghey J, Amit I, Gnirke A, Meissner A. 2015. Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**: 344–349. http://dx.doi.org/10.1038/nature14233.

Waddington CH. 1957. The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. *Strateg genes A Discuss some ...* ix +–262 . http://www.cabdirect.org/abstracts/19580101706.html\npapers3://publication/uuid/E3C5A230-84D0-4CB4-BA78-4B7DBF839E2C.

Wakeman L, Al-Ismail S, Benton A, Beddall A, Gibbs A, Hartnell S, Morris K, Munro R. 2007. Robust, routine haematology reference ranges for healthy adults. *Int J Lab Hematol* **29**: 279–283.

Wilson G a, Butcher LM, Foster HR, Feber A, Roos C, Walter L, Woszczek G, Beck S, Bell CG. 2014. Human-specific epigenetic variation in the immunological Leukotriene B4 Receptor (LTB4R/BLT1) implicated in common inflammatory diseases. *Genome Med* **6**: 19. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4062055&tool=pmcentrez&rendertype=abstract (Accessed December 3, 2014).

Xiao S, Xie D, Cao X, Yu P, Xing X, Chen CC, Musselman M, Xie M, West FD, Lewin HA, et al. 2012. Comparative epigenomic annotation of regulatory DNA. *Cell* **149**: 1381–1392.

Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL, et al. 2013. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* **45**: 836–41. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3695047&tool=pmcentrez&rendertype=abstract.

Yen RW, Vertino PM, Nelkin BD, Yu JJ, el-Deiry W, Cumaraswamy A, Lennon GG, Trask BJ, Celano P, Baylin SB. 1992. Isolation and characterization of the cDNA encoding human DNA methyltransferase. *Nucleic Acids Res* **20**: 2287–2291.

Yoshida Y, Han B, Mendelsohn M, Jessell TM. 2006. PlexinA1 signaling directs the segregation of proprioceptive sensory axons in the developing spinal cord. *Neuron* **52**: 775–88. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1847564&tool=pmcentrez&rendertype=abstract (Accessed December 3, 2014).

Zeng J, Konopka G, Hunt BG, Preuss TM, Geschwind D, Yi S V. 2012. Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am J Hum Genet* **91**: 455–65. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3511995&tool=pmcentrez&rendertype=abstract (Accessed September 30, 2014).

Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W, et al. 2010. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res* **38**: 4246–4253.

Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, De Jager PL, Rosen ED, Bennett D a, Bernstein BE, et al. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**: 477–81. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3821869&tool=pmcentrez&rendertype=abstract (Accessed July 11, 2014).