

**Comparative DNA sequence and methylation
analyses of orthologous genes in humans and
non-human primates:
Genetic and Epigenetic Footprints of
Evolution**

Dissertation

Zur Erlangung des Grades
„Doktor der Naturwissenschaften“

am Fachbereich Biologie
der Johannes Gutenberg-Universität
in Mainz

Ruxandra Denisa Farcas

geb. am 24.09.1980 in Klausenburg, Rumänien

Mainz, 2009

Dekan: _____

1. Berichterstatter: _____

2. Berichterstatter: _____

Tag der mündlichen Prüfung: _____

1. INTRODUCTION	1
1.1 An excursion into genome evolution.....	1
1.2 Understanding the role of positive and negative selection in evolution.....	6
1.3 Gene expression and the importance in primate evolution	10
1.4 Insight into Epigenetics	13
1.4.1 DNA Methylation	15
1.4.2 DNA Methylation and Evolution.....	19
2. MATERIALS AND METHODS.....	21
2.1 MATERIALS	21
2.1.1 Tissues Samples.....	21
2.1.2 Oligonucleotides	23
2.2 METHODS	25
2.2.1 Isolation of nucleic acids	25
2.2.1.1 Isolation of genomic DNA from tissue samples	25
2.2.1.2 Plasmid DNA isolation	26
2.2.1.3 RNA isolation from tissue sample	27
2.2.2 Spectrophotometric quantification of nucleic acid concentration.....	29
2.2.3 Agarose gel electrophoresis.....	29
2.2.4 Gel extraction of PCR fragments.....	30
2.2.5 DNA Cleanup	31
2.2.6 Cloning of PCR products.....	33
2.2.7 Bisulfite conversion and cleanup of DNA for methylation analysis.....	37
2.2.8 PCR (Polymerase Chain Reaction).....	41
2.2.8.1 Standard PCR.....	42
2.2.8.2 RT (Reverse Transcription) PCR.....	43
2.2.8.3 Quantitative Real-time PCR	44
2.2.9 DNA Sequencing.....	47

INDEX

2.2.9.1 Bisulfite-sequencing	49
2.2.9.2 Pyrosequencing	50
2.2.10 Bioinformatic Tools	53
2.2.10.1 In silico analysis of the coding sequences in the breakpoint regions	53
2.2.10.2 BioEdit version 5.0.6	55
2.2.10.3 MEGA 3.1 version (Molecular Evolutionary Genetics Analysis software)	56
2.2.10.4 CpG Island search	57
2.2.10.5 Primer Design	57
2.2.10.6 RepeatMasker (http://www.repeatmasker.org)	58
2.2.10.7 Giri (http://www.girinst.org)	58
2.2.10.8 Transcriptional Regulatory Element Database (http://rulai.cshl.edu)	58
2.2.10.9 Panther Database (http://www.pantherdb.org)	58
2.2.10.10 Novartis gene expression Atlas (http://symatlas.gnf.org/SymAtlas)	59
3. RESULTS	60
3.1 Evolution of coding sequences in the evolutionary breakpoint regions	60
3.1.1 Evolution of the coding sequences in the breakpoint regions of chromosome 1, 4, 5, 9, 12, 17 and 18	61
3.1.2 Summarized data of the coding sequence comparison between human and chimpanzee	69
3.2 Bioinformatical comparison of gene expression in the evolutionary breakpoint regions of human and chimpanzee	76
3.3 DNA-Methylation analysis in human and non-human primates	79
3.3.1 Comparison of CGI promoter methylation in human and chimpanzee cortex	80
3.3.2 Variation of <i>CCRK</i> CGI promoter methylation among humans and non-human primates	85
3.4 Expression of <i>CCRK</i> gene in the frontal cortex of human and non-human primates	94
3.5 Analysis of DNA methylation patterns using bisulfite Pyrosequencing	97
4. DISCUSSION	102

4.1 Natural Selection of genes at evolutionary breakpoint regions	102
4.2 Plasticity of CpG islands methylation patterns in human and non-human primates cortices	107
4.3 Comparison of DNA-Methylation patterns in human and chimpanzee cortices by bisulfite Pyrosequencing	113
5. SUMMARY.....	115
6. REFERENCES	118
7 ANNEX	128
7.1 Abbreviations	128
7.2 Figure Index	131
7.3 Table Index	134
7.4 Acknowledgments	136
7.5 Curriculum vitae	137
7.6 Publication	138

1. INTRODUCTION

1.1 An excursion into genome evolution

Hominid evolution and human speciation are the most fascinating topics in evolutionary biology.

As it was mentioned by McConkey and Goodman[1997]: “ Comparative analysis of human and ape genomes is far more than an excursion into natural history at the molecular level. Until we have a detailed understanding of genetic differences between ourselves and our closest evolutionary relatives, we cannot really know what we are (p.351).”

Since the divergence of humans (*Homo sapiens*, HSA) and chimpanzees (*Pan troglodytes*, PTR) about 4.6-6.2 million years ago, these species have undergone a remarkable evolution with drastic divergence in anatomy and cognitive abilities.

As humans, we have an inherent interest for understanding the genetic basis of these physical and behavioral traits that distinguish human beings from each other, and the human species from other primates. “Understanding humanity is like”Searching for needles in a haystack” (Varki and Tasha , 2005).

The evident differences between humans and chimpanzees are assumed to have been strongly influenced by lineage and species-specific genomic changes that gave rise to differences in gene expression, gains or losses of genes and changes in protein function.

What we really want to explore and understand is actually a complex puzzle of multiple genetic differences, interacting with diverse environmental and cultural factors, resulting in the observed phenotypic differences.

Examination of genomic differences between humans and chimpanzees began when Yunis and coworkers explored differences in chromosome banding patterns between humans and great apes (Yunis J.J and Prakash O. 1982), and defined large lineage-specific chromosome rearrangements.

Only a few years ago, the knowledge of the structural divergence between the human and chimpanzee genomes was restricted to those large rearrangements that are visible at the cytogenetic level. The fusion of the ancestral chromosomes homologues to chimpanzee chromosomes 12 and 13 gave rise to human chromosome 2 reducing the human chromosome number relative to the other hominid species (Fan et al.2002). Nine pericentric inversions distinguish human and chimpanzee chromosomes (Eichler et al. 1996, Nickerson and Nelson 1998) and the additions of heterochromatin to the subtelomeric regions of great ape chromosomes. Some possible effects of these chromosome rearrangements were suggested: they might interfere with gene function(s) directly by disrupting a gene(s), by formation of fusion genes or otherwise causing expression differences and, thus, act as a driving force in evolution. Many of this chromosomal rearrangements have been intensively studied in the last years by molecular cytogenetic methods and the underlying breakpoints have been characterized at the DNA sequence level (reviewed in Wienberg 2005 and Szamalek et al. 2006 and references therein), so that a few of the proposed effects of this inversions could be excluded. In none of the mapped inversion breakpoint regions a novel gene was generated by the rearrangement, nor was a gene disrupted leading to loss of function. Although the *SLCO1B3* gene is located at the site of the inversion breakpoint at chromosome 12p, the complete *SLCO1B3* gene was restored by an 86 kb chimpanzee-specific duplication (Kehrer-Sawatzki et al. 2005).

But still it remained unclear whether any of these rearrangements played a role during the process of speciation. Only the possibility of a systematic comparison of these breakpoint regions at the nucleotide level will provide insights into the mechanisms and consequences of chromosome reshuffling during primate evolution.

The release of the final version of the human genome (International Human Genome Sequencing Consortium 2004) and one year later of the initial nucleotide sequence of the chimpanzee (The chimpanzee Sequencing and Analysis Consortium 2005; Varki and Altheide 2005) was real revolution in the whole of biology and a new era, the "postgenomic era" has started. Genome comparisons of human and chimpanzee could reveal the molecular basis of the human-specific traits as well as the evolutionary forces that have formed our species. With the two genome sequences "in hand" it was possible to begin a systematic identification of specific genes or genomic regions that differentiate between humans and chimpanzees.

But which are these genes or regions?

The large amount of information made it also difficult to decide where to focus further efforts on. At the nucleotide sequence level, the difference between humans and chimpanzees was found to be surprisingly small. These genetic differences include approximately 1% fixed single-nucleotide substitutions and 3% euchromatic divergence due to insertion and deletion (indels) events (reviewed Portin P. 2007). The nucleotide divergence rates are not constant across the genome, variation in the divergence rate is evident even at the level of whole chromosomes (The Chimpanzee Sequencing and Analysis Consortium). The most dramatic outliers are the sex chromosomes, with a mean divergence of 1,8% for the Y and 0,94% for the X, but significant variation in divergence rates was also seen among the autosomes (Hughes et al. 2005, Kuroki et al. 2006, Patterson et al. 2006).

The sequence divergence rate is influenced by conserved factors (stable across mammalian evolution) and lineage-specific factors which may change with chromosomal rearrangements.

To elucidate the question whether or not the chromosomal rearrangements contributed to the human-chimpanzee speciation, a systematic *in silico* analysis of the coding sequences flanking the pericentric inversion breakpoints (BP) specific for human and chimpanzee on chromosomes 1, 4, 5, 9, 12, 17 and 18 was done in this thesis. All coding sequences surrounding the mentioned breakpoint regions were analyzed.

Progress in comparing the human and chimpanzee genomes has revealed that beside the gross chromosomal rearrangements there are a considerable number of submicroscopic structural differences (Newman et al. 2005). These include insertions and deletions and involve a variety of different sequences like microsatellites, high copy number repeats and transposons. Comparison of the two genomes also indicated that the sites containing CpG dinucleotides in either species showed a substantially elevated divergence rate. The Alu elements have been threefold more active in humans than in chimpanzees. Most chimpanzee-specific elements belong to a subfamily (AluYc1) and the human-specific Alu elements belong to two new subfamilies (AluYa5 and AluYb8) (The Chimpanzee Sequencing and Analysis Consortium 2005).

The draft sequence of the chimpanzee genome is also important for studying human population genetics with a high relevance to human medical genetics. The chimpanzee sequence allows recognition of those human alleles that represent the ancestral state or the derived state. In case of some genes, it was shown that the disease-associated allele in humans is the ancestral allele, which is the wildtype allele in chimpanzee and other species.

At the same time it became more and more clear that additional genome sequences (i.e. rhesus macaque and orangutan) are necessary for understanding whether a certain change occurred on the chimpanzee (PTR) or on the human (HSA) lineage. The sequence shared with a more distant species is likely to be ancestral.

1. Introduction

One or more outgroups are needed to determine the ancestral state at any given locus. In 2007 a new important milestone was reached with publication of the rhesus macaque (an Old World monkey) genome (Rhesus Macaque Genome Seq. Anal. Consortium 2007).

The genome sequencing projects of orangutan, gorilla and a New World monkey (*Callithrix jacchus*) are in the initial stage.

Comparative analysis of the genomes and transcriptomes of humans and non human primates give insights into the formation of our species, indicating a complex speciation process and providing a picture of the events that have shaped our genomes.

1.2 Understanding the role of positive and negative selection in evolution

The interest in detecting and understanding selection arose with the demand to understand the forces that govern how populations and species evolve. Populations evolve as mutations that appear in individual organisms, and through hereditary transmission gradually became fixed in the population. Many mutations have no effect on organismal fitness, however some mutations produce selective advantages that increase their chances of reaching fixation.

There are two known types of selective forces that shape the evolution of species: the Negative (purifying) selection which acts to eliminate deleterious mutations and favors the conservation of existing phenotypes and the Positive selection (also known as Darwinian selection), which favors advantageous mutations once promotes the emergence of new phenotypes.

Comparative genomics within the human species as well as between us and our closest living relatives can shed light on the roles of positive and negative natural selection during human evolution. Genes, or genomic regions, that are affected by natural selection may show an excess of functionally important molecular changes, beyond what would be expected in the absence of selection. Genomic sequence with such an excess of changes are believed to have experienced positive selection, i.e. selection in favour of new genetic variants. In this context questions like which genes have been positively selected, when did it occur, and what were the functional consequences have to be answered. Identifying genes or genomic regions that have been influenced by positive selection may provide a key to understand the processes that lead to differences among species.

Positive Darwinian selection is an important source of evolutionary innovation and a major force behind the divergence of species. However, positive selection is sometimes difficult to detect because it often appears episodically just on a few amino acid sites, and the signal may be masked by negative selection. A gene that experiences a “selective sweep” as a result of only a few changes, because those changes are strongly advantageous, would not leave a significant signal of its positive selection. Also a novel Bayesian analysis of the possible “selection histories” of each gene indicated that most positively selected genes (PSGs) have switched multiple times between positive selection and non-selection.

Recent studies (Zhang J. 2004, CSAC 2005, Zhang et al. 2005) discuss the importance of distinguishing between positive selection and the relaxation of purifying selection or “relaxation of selective constraints (RSC)” to exclude false positive results that will obscure our understanding of this complex process of selection.

A wide variety of methods for detecting positively selected genes exists, including comparative or phylogenetic methods, which use patterns of substitutions between species, and population genetic methods, which rely on patterns of intraspecific polymorphisms.

While the phylogenetic methods are suitable for detecting selection that operates over relatively long periods of time in evolution, the population genetic methods are used for detecting more recent selection. In the last years there has been a “big debate” on which of the phylogenetic methods has the largest power and at the same time sensitiveness to give the most reliable result. Each method has his own advantages / disadvantages that have to be taken in consideration.

One method for the detection of positive selection acting on a protein-coding gene relies on the comparison of the rate of nonsynonymous “dN” (amino acid altering) changes with the rate of synonymous (silent) “dS”, by calculating the ratio of dN/dS (Ka/Ks).

If $dN/dS > 1$, then the gene or analyzed genomic sequence may be considered to have evolved under positive selection. Neutrally evolving genes generally show a $dN/dS = 1$, whereas a ratio of $dN/dS < 1$ indicate the presence of purifying or negative selection. The dN/dS ratio can thus be used to gain insight into the extent and mode of selection affecting a particular gene.

One example of an extreme activity of positive selection is found in the protamine 1 (*PRM1*) gene, a sperm-specific protein that compacts sperm DNA. *PRM1* shows 13 nonsynonymous and 1 synonymous differences between human and chimpanzee (Sabeti et al. 2006).

Since it first became possible to compare the sequences of complete mammalian genomes a number of genome-wide scans for PSGs have been done, providing an review of the landscape of positive selection, but some important questions remain unanswered.

It has been reported that PSGs are enriched for roles in sensory perception, immunity, apoptosis, tumor suppression and spermatogenesis (Clark et al. 2003, Nielsen et al. 2005) and that more genes have undergone positive selection in the chimpanzee than in the human lineage (Bakewell, Shi, Zhang 2007). PSGs often coincide with segmental duplications (SDs). Segmental duplications are large nearly identical copies of genomic DNA that map to two or more genomic locations and can either be organized in tandem or are interspersed in the genome. Comparisons between and within species indicated that SDs have played an important role in primate evolution, by creating new genes or providing the possibility of evolution of new gene functions, those shaping the human genetic variation (Bailey and Eichler 2006). At the same time it was claimed that the evolutionary breakpoints are often associated with segmental duplications. Comparisons of the great ape and human genomes show that most evolutionary chromosome rearrangement events are associated specifically with SD (Locke et al. 2003, Stankiewicz 2004, Kehrer-Sawatzki 2005, Bailey J.A and Eichler E.E 2006)

1. Introduction

It has been proposed that positive selection operates more intensely on rearranged than on collinear chromosomes during speciation.

This hypothesis was first published by Navarro and Barton paper (2003), who analyzed the coding regions of 115 annotated autosomal genes using the ratio dN/dS . They concluded that the mutation rate is much higher in rearranged than in collinear chromosomes, and that the rearranged chromosomes had experienced notably more positive selection during human-chimpanzee speciation.

Reexamination of these conclusions represents one part of the research in this thesis.

1.3 Gene expression and the importance in primate evolution

Despite only around 1% difference in genomic DNA sequence, humans and chimpanzees differ in many phenotypic traits but also in the susceptibility for some diseases i.e. neurodegenerative disorders and cancer. Some differences in the organization of the neocortex, i.e. enlargement of the neocortex during hominoid evolution was important for the development of specific cognitive abilities in humans, but little is known about the underlying changes at the molecular level. It is assumed that many of these species differences, i.e. in cognitive abilities, must be due to changes in gene regulation rather than structural changes in the gene products and protein structure.

The pioneer study of Enard and Pääbo 2002 presented a first view on the gene expression differences between humans and non-human primates. They compared mRNA levels in brain and liver of humans, chimpanzees and orangutan using Affymetrix arrays which contain oligonucleotides for approximately 12,000 human genes. They observed more expression differences in the brain compared to liver, implying that gene expression pattern changed more dramatically in the brain during recent human evolution, and that the rate of evolutionary change of gene expression levels in the brain is accelerated in the human evolutionary lineage relative to the chimpanzee.

Other comparative transcriptome analyses also revealed substantial expression differences between humans and chimpanzees, in particular in brain tissue. A subset of genes showed elevated expression in the human brain after the split from the chimpanzee (Caceres et al. 2003; Gu and Gu 2003; Khaitovich et al. 2006).

Another large Affymetrix screen for differences in gene expression pattern between human and chimpanzee brain showed that out of 2014 genes that are differently expressed between

the two species 1270 genes are more highly expressed in humans than in chimpanzees and that the ratio of differentially expressed genes varied among the chromosomes, ranging from 9.4% on chromosome 19 to 17.3% on chromosome 9 (Kaitovich et al. 2004).

Generally, mutations can alter the phenotype either by changing primary coding sequences of proteins and other gene products, or by changing regulatory DNA sequences that control transcription and translation. A few studies have been able to establish a link between DNA sequences and phenotype differences (Gilbert et al. 2005), the identified DNA sequence changes between humans and chimpanzees are likely to be relevant for functional differences between the species. One intriguing example is *FOXP2*, a gene which is important for the human ability to develop language. A point mutation in this gene that encodes a known transcription factor segregates in families with severe articulation impairment (Fisher et al. 1998). The human *FOXP2* protein differs in two amino acid changes in a functional domain from chimpanzee (Enard et al. 2002, Zhang et al. 2002). This may be one of the genes in which specific nucleotide changes occurring in the human lineage contributed to the human phenotype.

Most differences in gene regulation are believed to occur at the transcription level.

The processes involved are likely to include: sequence divergence of regulatory regions, differences in the control of transcriptional initiation, modification of chromatin structure and differences in DNA methylation (Wray et al. 2003).

Sequence divergence in gene promoter regions between humans and chimpanzees has been shown to be responsible for the modulation of promoter activity, influencing the transcription levels. One example is the apolipoprotein(a) (LPA) gene promoter (Huby et al. 2001).

1. Introduction

To better understand the molecular mechanisms that govern the expression patterns it is also important to identify the regulatory elements (enhancers, repressors, transcriptional factor binding sites) associated with each gene. Even single-nucleotide differences in the regulatory sequence can have significant effects on gene expression. Therefore, these elements are usually located in evolutionary conserved non-coding regions

1.4 Insight into Epigenetics

Although the DNA sequence provides the basic blueprint for life, this code is subject to a number of “epigenetic” modifications that provide another complex layer of information.

The term “Epigenetics” literally “above genetics” was in 1942 by Conrad Waddington as “the interactions of genes with their environment that bring the phenotype into being”. However, the meaning has changed over the years Wu and Morris (2001) defined epigenetic as: “a study of the changes in gene expression that are mitotically and/or meiotically heritable and do not involve a change in the DNA sequence.”

Two fundamental components of the epigenome (the overall epigenetic state of a cell) are chromatin structure and the covalent modification by methylation of the DNA molecule itself, termed “DNA methylation”. These components are interrelated.

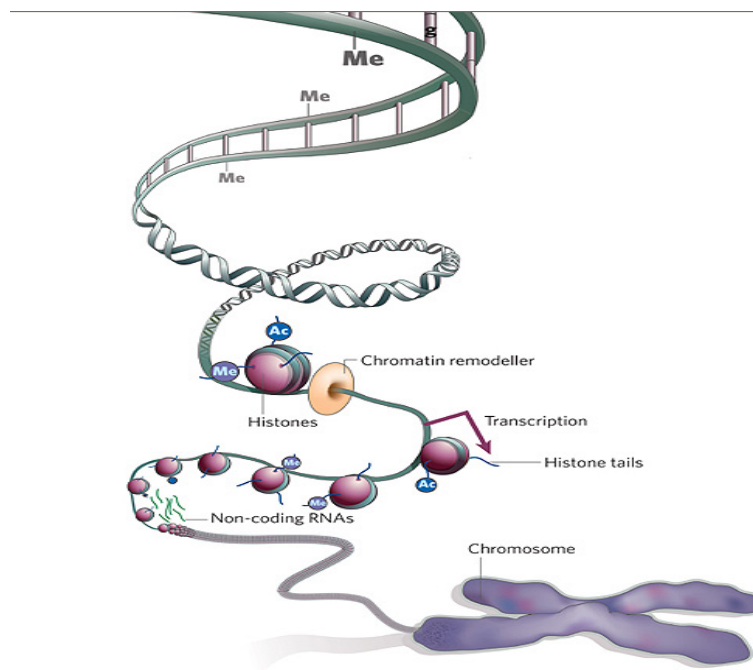


Figure 1: The Epigenetic mechanisms include histone modification, positioning of histone variants, nucleosome remodelling, DNA methylation (adapted from *Nature*, 7 Aug 2008)

1. Introduction

Chromatin structure includes the higher order of chromatin folding and attachment to the nuclear matrix, packaging of DNA around nucleosomes and the covalent modifications of histone tails (acetylation, methylation, phosphorylation, ubiquitination).

Chromatin can be found in an open form (euchromatin) that allows access of the transcriptional machinery, and a closed form (heterochromatin) which restricts the access of RNA polymerase II to DNA thereby repressing gene transcription. The transition between these two states of chromatin can be induced through epigenetic modifications of histones and the DNA itself.

The modification of histones, particularly acetylation and methylation, play a crucial role in this process. It is known that histone acetylation increases the transcriptional activity/potential of chromatin (Stral and Allis 2000). The same effect is seen for the methylation of lysin 4 of histone 3 (H3), whereas a lower transcriptional activity is associated with methylation of lysin 9 of H3 (Jenuwein and Allis 2001, Bird 2002).

DNA methylation in mammals is a post-replicative modification that is predominantly found in cytosines of the CpG diucleotide and is due to the enzymatic attachment of a methyl group to carbon-5 atom of the pyrimidine ring, creating a 5-methylcytosine. DNA methylation establishes and/or maintains an inactive chromatin structure by posttranslational histone modifications (Wolffe and Matzke 1999; Jaenisch and Bird 2003).

Epigenetic mechanisms are responsible for several phenomena including: X-inactivation – the random silencing of one of the X chromosomes in somatic cells of female mammals (reviewed by Park and Kuroda, 2001), genomic imprinting – the expression or repression of certain genes or genetic loci according to their parental origin (reviewed by Ferguson-Smith and Surani, 2001), embryonic development (Li, Bestor and Jaenisch 1992), transcriptional regulation and cancer development.

1.4.1 DNA Methylation

DNA methylation is the most thoroughly studied epigenetic modification of the genome that is involved in the regulation of many cellular processes.

The mammalian DNA methylation machinery comprises the DNA methyltransferases (DNMTs) that establish and maintain DNA methylation patterns, and the methyl-CpG binding proteins (MBDs), which are involved in ‘reading’ the methylation mark. There are five known DNA methyltransferase (DNMT) family members, DNMT1, 2, 3a, 3b, and 3l in mammalian cells that establish genome-wide DNA methylation patterns during embryonic development and maintain them in somatic cells (Bestor 2000; Robertson 2002).

DNMT1 is the most abundant and catalytically active enzyme with primary role in maintaining the established methylation pattern through cell division; it copies the pre-existing methylation patterns onto the newly synthesized DNA strand during DNA replication (Bestor et al. 1996; Bestor 2000). *DNMT3a* and *DNMT3b* are responsible for introducing methylated cytosines at previously unmethylated sites. They are essential for *de novo* methylation, which establishes methylation pattern during embryogenesis or differentiation processes in adult cells (Razin and Cedar 1993). *DNMT3L* is not a functional enzyme due to lacks of critical catalytic site motifs but may directly stimulate other DNMTs. The function of *DNMT2* remains unclear.

In addition to their catalytic role, DNMTs also have a non-enzymatic function in the modification of chromatin, because they can interact biochemically with histone methyltransferases and histone deacetylases (Bai et al. 2005).

DNMT enzymes can be recruited to the target DNA in different ways: (1) The *DNMT3* enzymes may recognize DNA or chromatin via specific domains, (2) *DNMT3a* and *DNMT3b*

may be recruited through protein–protein interactions with transcriptional repressors or other factors, (3) the RNA-mediated interference (RNAi) system may target de novo methylation to specific DNA sequences (Klose and Bird 2006).

To alter the established methylation pattern demethylation processes must exist that remove the methyl groups from methylated DNA. Passive demethylation occurs when DNMT1 fails to maintain the existing methylation pattern, whereas active demethylation cleaves methyl group or excises methyl cytosins from non-replicating DNA.

In mammals, methylation is largely restricted to CpG dinucleotides, which are depleted from the genome except at short genomic regions called CpG islands, which are often found in promoters (Jaenisch and Bird 2003). CpG islands (CGIs) are 500-2,000 bp long genomic DNA segments enriched for the dinucleotide 5'-CpG-3', and associated with cis-regulatory sequences in most mammalian genes. Whereas most regions of the genome are constantly methylated, most CGIs in promoters and first exons are protected from methylation in somatic tissues (Rollins et al. 2006) thereby facilitating the establishment of the transcription initiation. The methylation status of CpG islands is controlled at several levels: the location (promoter versus non-promoter), the local chromatin structure, regulatory motifs that target DNMTs indirectly through DNA binding proteins, and the overall structure and sequence composition of the CpG island itself. Based on CpG frequency, three different classes can be distinguished. High CpG island promoters are largely unmethylated, even when inactive. Low and intermediate CpG island promoters are predisposed to de novo methylation during development and differentiation (Weber et al. 2007).

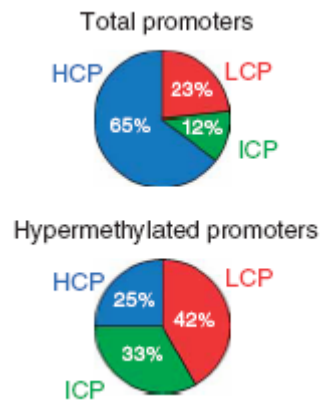


Figure 2: Pie charts showing the relative frequency of CGI classes among total promoters and hypermethylated promoters defined by \log_2 ratio > 0.4 (Weber et al. Nat.Genetics 2007)

Methylation of critical CpG dinucleotides in cis-regulatory regions, in particular promoters, is generally thought to act as epigenetic signal that regulates the temporally, spatially, and parent-specifically appropriate gene expression patterns. This regulation can be explained by the “critical site” model which suggests that the methylation of specific cytosines in transcription-factor binding sites reduces binding affinity directly or via recruitment of methyl-CpG binding domain (MBD) proteins inducing chromatin changes.

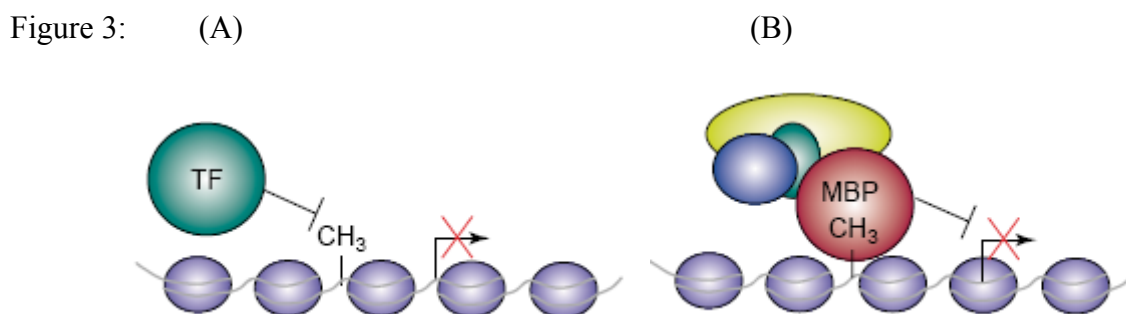


Figure 3 (A) and (B): Mechanisms of DNA methylation mediating repression
(A) By blocking activators from binding targets sites, DNA methylation directly inhibits transcriptional activation. TF = transcription factors **(B)** Methyl-CpG-binding proteins (MBPs) directly recognize methylated DNA and recruit co-repressor molecules to silence transcription and to modify surrounding chromatin.

According to the “methylation density” model the proportion of methylated cytosines across a region controls chromatin conformation and transcriptional potential.

Another remarkable compartment enriched with CpGs are the repetitive DNA elements in particular the Alu transposons which are the most prominent short interspersed nuclear elements (SINEs) adding up to roughly 10% of the human genome. Alus are nonautonomous, small elements that require the enzymatic machinery provided by LINE-1 expression for retrotransposition (Hagan et al. 2003). Although the period(s) of rapid expansion of Alus and other retrotransposons lies in the past of human evolution, Alus can still be actively retrotransposing in the human genome.

Primate Alu elements display a dimeric structure and originated from tRNA or the 7SL RNA which is an integral part of the signal recognition particle.

The transposable elements (TEs) like Alus have been proposed to be major players in shaping the primate genome and transcriptome. Many genes contain one or several Alus in close proximity 5' to their CGIs (International Human Genome Sequencing Consortium 2004), placing of the repeats in stretches of sequence most probably relevant to gene regulation.

To prevent retrotransposition of TE, the genome uses epigenetic “defence” mechanisms to suppress their activity. The silencing of Alus can be achieved by transcriptional (by DNA methylation) and post-transcriptional (by RNAi) mechanisms. Most Alu transposons are highly methylated. DNA methylation is thought to be a major mechanism for preventing the genome from transcription and transposition of these elements. In somatic tissues and mature germ cells, most retroelements are densely methylated and consequently transcriptionally inactive (Schulz, Steinhoff and Florl 2006).

1.4.2 DNA Methylation and Evolution

During the last 5-7 million years of human evolution, the brain has evolved dramatically, giving rise to our unique cognitive abilities. It was already mentioned in the previous chapters that many of the human-specific traits are caused by changes in gene regulation and that the methylation at CpG dinucleotides represent one important factor for gene regulation in mammals.

However, little is known about the evolutionary conservation of DNA methylation patterns and the evolutionary impact of epigenetic differences between closely related species.

In humans and non-human primates, DNA methylation is restricted to CpG dinucleotides, which are largely depleted from the genome because of their inherent mutability (deamination of methylated cytosines causing C to T transitions) (Shen et al. 1994).

Comparative bisulfite DNA sequencing revealed a remarkable conservation of methylation profiles between human and mouse orthologous genes in four different tissues (skin, liver, skeletal and heart muscles). Less than 5% of the analyzed loci were differentially methylated between these two evolutionarily distant species (Eckhardt et al. 2006). An array-based comparison of 145 CpG sites from 36 different genes identified 18 CpGs (12 genes) that have significantly changed methylation between humans and chimpanzees (Enard et al. 2004). The observation that many more CpGs were differentially methylated in the brain than in liver or lymphocytes promotes the idea that a change in methylation patterns has contributed to the evolution of the human brain. It is tempting to speculate that changes in the methylation patterns of key genes during evolution may have preceded or dictated functional changes involving chromatin organization and /or transcription.

1. Introduction

Enlargement of the neocortex during hominoid evolution was important for the development of specific cognitive abilities in humans. Genes that have changed their regulation during evolution of the human brain may also contribute to the intraspecific variation and pathology of cognitive abilities among humans. In this light, differences in DNA methylation patterns between humans and chimpanzees can be viewed as an epigenetic footprint of genes that are crucial for human brain development and function.

One aim of this thesis is to compare CGI promoter methylation of representative genes between human and chimpanzee cortex. One gene, the cell cycle related kinase (*CCRK*), exhibited differences in its methylation pattern in humans and chimpanzees and therefore was studied further in two Old World monkeys.

2. MATERIALS AND METHODS

2.1 MATERIALS

All common chemicals, materials, medias and stock solutions were ordered from different companis or prepared according to Sambrook et al. (1989), and will not be listed here.

2.1.1 Tissues Samples

Brain samples were obtained between 1-3 days post-mortem from 12 unrelated humans (*Homo sapiens*, HSA), three chimpanzees (*Pan troglodytes*, PTR), one rhesus macaque (*Macaca mulata*, MMU), and three baboons (*Papio hamadryas*, PHA). Frontal cortex tissue (area A10) was dissected by experienced neuropathologists and immediately frozen and stored at -80°C . Human brain autopsy samples (excess material) were from the Institute of Legal Medicine, the Department of Neuropathology and the Department of Child Pathology at the Mainz University Medical Center. Primate samples were obtained from the Biomedical Primate Research Centre, Rijswijk, Netherlands and the Primate Centre, Göttingen, Germany. More information about the tissue samples are presented in the table 1.

The frozen tissue sections were stored at -80°C prior to DNA and RNA extraction. Extraction of all DNAs and RNAs was performed using standard protocols (see Methods). The quality and quantity of the templates was assessed by spectrophotometry and agarose-gel analysis, and were subsequently stored at -20°C respectively at -80°C until further use.

2. Materials and Methods

Individual	Sex	Age	Time post-mortem	Cause of death
HSA1	female	66 years	1 day	pneumonia
HSA2	female	80 years	2 days	acute heart failure
HSA3	female	83 years	1 day	pulmonary embolism
HSA4	female	82 years	1-2 days	chronic heart insufficiency
HSA5	female	81 years	1-2 days	multiple organ failure
HSA6	male	84 years	1 day	pneumonia
HSA7	male	54 years	2 days	traffic accident
HSA8	male	31 years	1 day	suicide
HSA9	male	45 years	1-2 days	tractor accident
HSA10	male	59 years	3 days	acute heart failure
HSA11	male	40 years	1-2 days	motorcycle accident
HSA12	female	17 gestat.	1-2 days	chorion amnionitis
PTR1	male	14 years	12 h	haemolytic anemia
PTR2	male	7 years	12 h	anaesthesia accident
PTR3	female	40 years	12 h	drown in pond
MMU1	female	24 years	6 d	septicaemia after injury
PHA1	male	30 years	12 h	terminated for experimental brain
PHA2	male	9 years	1 day	terminated
PHA3	male	9 years	12 h	terminated

Table 1. Brain tissue samples used in this study

2.1.2 Oligonucleotides

The synthetic oligonucleotide primers used in these study were purchased from OPERON or MWG and dissolved in water to a final concentration of 100 pmol/ μ l. The primer sequence are listed in Table 2.A, B, C

Gene	Forwerd Primer	Revers Primer
CCRK	AGAAGGTGGCCCTAAGGCGGTTGG	GCTACCTGCAGATGCTGCTCAAG
B2M	QT 00088935, Qiagen	

Table 2. (A) Primers used for the qRT-PCR

2. Materials and Methods

Gene Symbol		Forward Primer	Reverse Primer	External Reverse Primer
CCRK		AGGCCTACTCTGTCTGTCTGGG	ATGACAGCAGGCACCTCCTATC	GCCCAGGATGCAGTACTGGTCCATCC
	Bisulfite	AGGTTTATTTGTTTTGTTTGGG	ATAACAACAACACCTCCTATC	ACCCAAAATACAATACTAATCCATCC
ALDH1B1		GTGCCTTAGGAGCCTAGATCTGAG	AGCCAGGAGCTGCGGCTTCCC	CGGGGCTCCTGTAGCGTCACC
	Bisulfite	GTGTTTTAGGAGTTTAGATTTGAG	AACCAAAAACATACTTCCC	CAAACTCCTATAACATCACC
IGFBPL1		GCATACTATCTGGAGCCAGCCTTGG	ACGTCGCGGATCCCAAGGCTC	CCGGGCAGGGCGCAGGGGCC
	Bisulfite	GTATATTATTTGGAGTTAGTTTTGG	ACATCACAATCCCAAACTC	CCAAACAAAACACAAACACC
ZNF519		TGAAGCYGAGAAATGGTGAGT	CCTTTCTTTCAAGACCTTCTAGT	
	Bisulfite	TGAAGTRGAGAAATGGTGAGT	CCTTTCTTTCAAAACCTTCTAAT	
SHC3 CGI-1		GCCTTCCAAGTCAAGATTCATCAGG	ACCCTTCTCAACAAAGGCTC	
	Bisulfite	GTTTTTAAGTTAAGATTTATTAGG	ACCCTTCTCAACAAAACCTC	
SHC3 CGI-2		GTGAGCAGCCCCAAGCAGCTGGG	TGCRGTGAAGCATGCCCTCC	
	Bisulfite	GTGAGTAGTTTTAAGTAGTTGGG	TACRTAAAACATACCCCTCC	
NTRK2		ACTCTGCGGGTAGATCAGTG	CAAGCCTTGTCTGAGAATCC	
	Bisulfite	ATTTTGYGGGTAGATTAGTG	CAAACCTTATCTAAAATCC	
MGMT		GGGTCAGGCGCACAGGGCAG	GGTCAGGGCRGCCACACCC	
	Bisulfite	GGGTTAGGYGTATAGGGTAG	AATCAAAACRACCCACACCC	

Table 2.B Classical Bisulfite Sequencing Primer

Gene Sybol		Forward Primer	Reverse Primer	Sequencing Primer
MGMT		GGATATGCTGGGACAGCC	AGGCTGGGCAACACCTGGG	
	Bisulfite	GGATATGTTGGGATAGTT-biotin	AAACTAAACAACACCTAAA	CCCAAACACTCACAAA
GJB2		ACCCGGGAAGCTCTGAGGA	TCTGCGCTGGGGCTCCTGC	
	Bisulfite	ATTYGGGAAGTTTTGAGGA	TCTACRCTAAAACCTAC-biotin	TTGAGGATTTAGAGG
NESP_AS		GATGAAGGGGTGGCCAGCA	CCAGGGGTACCTTCTTGACCTTG	
	Bisulfite	GATGAAGGGGTGGTTAGTA-biotin	CCAAAATACCTTCTTAACCTTAA	TAAACTAAAACCTCTCAAAT
GNAS		CTCTCTGCAGAGCCAGAGGGCAGGC	GGGAGGGACAGCTCAAGGTCTGCC	
	Bisulfite	TTTTTTGTAGAGTTAGAGGGTAGGT	AAAAAAAACAACCTCAAAATCACC-biotin	GTGTTTAAGAGGATGGAT
MEG3		GATCCCCCACACATTGTGTTTG	CTCATTTCTCTAAAAGTGATTGGCC	
	Bisulfite	GATTTTTTTATATATTGTGTTTG	CTCATTTCTCTAAAATAATTAACC-biotin	AATTTATTTTGTGTTGG
HELT		AGTGTGCATGGAATGAAATGTGGT	CCCTCCCAGGTTGCTCTGCCA	
	Bisulfite	AGTGTGTATGGAATGAAATGTGGT-biotin	CCCTCCCAATACTCTACCA	CCCACTCCCATTTTTA

Table 2.C Bisulfite Pyrosequencing Primers

2.2 METHODS

2.2.1 Isolation of nucleic acids

2.2.1.1 Isolation of genomic DNA from tissue samples

The genomic DNA isolation from brain tissue was done with the QIAmp DNA Mini Kit from Qiagen.

◆ **Standard protocol for DNA isolation from tissue sample**

- Homogenization: Cut up to 25 mg of tissue into a 1.5 ml microcentrifuge tube containing 80 µl PBS. Homogenize the sample and then add 100 µl Buffer ATL.
- Lysis: Add 20 µl Proteinase K, mix by vortexing, and incubate at 56°C with shaking until the tissue is completely lysed. Lysis time varies depending on the type of tissue, lysis overnight is possible and does not influence the preparation.
- RNA precipitation: Add 200 µl Buffer AL to the sample, mix by pulse- vortexing and incubate at 70°C for 10 min.
- Add 200 µl ethanol (96-100%), vortex shortly and centrifuge the tube to remove drops from inside the lid.
- Apply the mixture (including the precipitate) to the Spin Column and centrifuge at 6000 x g for 1 min. Discard the filtrate.
- Add 500 µl Buffer AW1, centrifuge at 6000 x g for 1 min. Discard the filtrate.

2. Material and Methods

- Add 500 µl Buffer AW2, centrifuge at 20,000 x g for 3 min. Discard the filtrate.
- Optional: Additional centrifugation for 1 min. to eliminate any remnants of Buffer AW2.
- DNA Elution: Place the Spin Column containing the DNA in a clean 1.5 ml microcentrifuge tube, add 50 µl Buffer AE or distilled water. Incubate at room temperature for 1 min. and then centrifuge at 6000 x g for 1 min.

2.2.1.2 Plasmid DNA isolation

The plasmid DNA isolation was done with the NucleoSpin[®] plasmid Kit from Macherey-Nagel. The bacterial pellets were resuspended and plasmid DNA was released from the *E.coli* host cells by SDS/alkaline lysis. The resulting lysate was neutralized with buffer A3, creating appropriate conditions for binding of plasmid DNA to the silica membrane of the provided column.

Contaminations such as salts, metabolites and soluble macromolecular cellular components were removed by simple washing with ethanol containing buffer A4. Pure plasmid DNA was finally eluted under low ionic strength conditions with slightly alkaline buffer AE (5 mM Tris-Cl, pH8.5). This standard protocol is suited for plasmids up to 15 kb length.

◆ Standard protocol

- Cultivate and harvest bacterial cells: Centrifuge 3 ml of a saturated overnight *E. coli* LB culture for 1 min. at 11,000x g. All centrifugation steps are at room temperature.) Discard the supernatant.

2. Material and Methods

- Cell lysis:
 - Add 250 μ l buffer A1 (with RNase A). Vortex to resuspend the cell pellet.
 - Add 250 μ l buffer A2. Mix gently by inverting the tube 6-8 times. Incubate at room temperature for <5 min.
 - Add 300 μ l buffer A3. Mix gently by inverting the tube 6-8 times.
- Purification of lysate: Centrifuge for 5-10 min at 11,000x g
- Bind DNA: Load the supernatant from last step onto the column. Centrifuge for 1 min at 11,000x g. Discard flowthrough.
- Wash silica membrane:
 - Optional: Add 500 μ l prewarmed buffer AW (50°C). Centrifuge for 1 min at 11,000x g.
 - Add 600 μ l buffer A4 (with ethanol). Centrifuge for 1 min at 11,000x g. Discard flowthrough and repeat once.
- Dry silica membrane: Reinsert the column into the 2 ml collecting tube. Centrifuge for 2 min at 11,000x g.
- Elute clean DNA: Place the column in a 1.5 ml microcentrifuge tube and add 25-50 μ l buffer AE. Incubate for 1 min at room temperature. Centrifuge for 1 min at 11,000x g.

2.2.1.3 RNA isolation from tissue sample

TRIzol Reagent[®] was used to isolate total RNA from brain tissues. TRIzol is a ready-to-use mono-phasic solution of phenol and guanidine isothiocyanate. This method is an improvement of

2. Material and Methods

the single-step RNA isolation method of Chomczynski and Sacchi (1987). TRIzol maintains the integrity of RNA, while disrupting cells and dissolving cell components during sample homogenization or lysis. Addition of chloroform followed by centrifugation separates the solution into an aqueous phase and an organic phase. RNA remains exclusively in the aqueous phase. After transfer of the aqueous phase, the RNA is recovered by precipitation with isopropanol.

◆ **Standard protocol for RNA isolation from tissue sample**

- Homogenization: Homogenize tissue sample in 1 ml of TRIzol Reagent[®] per 50 mg of tissue using a power homogenizer or sonicator.
- Phase separation: Incubate the homogenized sample for 5 min at room temperature. Add 200 µl chloroform (100%). Shake tubes vigorously by hand for 15 sec and incubate at room temperature for 2-3 min. Centrifuge the sample at 12,000x g for 15 min at 4°C.
- RNA precipitation: Pipet 500 µl chilled isopropanol in a new eppendorf tub and then transfer the aqueous phase from previous step. Swing the tub carefully. Incubate sample at room temperature for 10 min and centrifuge at 12,000x g for 10 min at 4°C. Remove supernatant. Wash the RNA pellet with 1 ml of 75% ethanol per 1 ml of TRIzol and centrifuge at 12,000x g for 5 min at 4°C. Remove supernatant and then repeat the wash step again.
- Redissolving the RNA: Air-dry RNA pellet for 5-10 min till it become glassy. Dissolve the RNA in RNase-free H₂O.

2.2.2 Spectrophotometric quantification of nucleic acid concentration

The nucleic acid (DNA and RNA) concentration and purity can be determined by spectrophotometric analysis. Absorbance measurement made on any spectrophotometer will include the absorbance of all molecules in the sample that absorb at the wavelength of interest. Since nucleotides, RNA, ssDNA (single stranded), and dsDNA (double stranded) all absorb at 260 nm, they will contribute to the total absorbance of the sample. The ratio of absorbance at 260 nm and 280 nm is used to assess the purity of DNA and RNA. A ratio of ~1.8 is generally accepted as “pure” for DNA; a ratio of ~2.0 is generally accepted as “pure” for RNA. If the ratio is appreciably lower in either case, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm.

2.2.3 Agarose gel electrophoresis

All RNA and DNA probes were tested for their quality by agarose gel electrophoresis. Agarose gels are used to separate nucleic acid molecules from as small as 50 bases to more than 20 kb, depending on the concentration of the agarose. Usually 0.8-2% gels were used.

In case of RNA, single-stranded RNA molecules often have complementary regions that can form secondary structures. Therefore, before loading the gel, 1 µl of RNA, 2 µl of DEPC water and 3 µl of FDE-Loading Dye, are denatured together for 5 min. at 95°C and chilled on ice for 1-2 min. The electrophoresis of RNA is at ~ 80V for ~ 40 min. and for DNA at 100-135V till the markers are well separated.

2.2.4 Gel extraction of PCR fragments

The band with the expected product size was extracted from the gel using the NucleoSpin[®] Extract Kit from Macherey-Nagel. DNA binds to a silica membrane in the presence of chaotropic salts (buffers NT1 and NT2). Buffer NT1 contains additional components in order to dissolve agarose gel slices. Afterwards, the dissolved agarose mixtures were loaded onto columns. Contaminations such as salts and soluble macromolecular components were removed by washing with ethanolic buffer NT3. Purified DNA was eluted under low ionic strength conditions with slightly alkaline buffer NE (5 mM Tris-Cl, pH8.5).

◆ Standard protocol for DNA extraction from agarose gel

- Excise DNA fragment with a clean scalpel from agarose gel.
- Gel lysis: Add 300 µl NT1 buffer per 100 mg of agarose gel. Incubate sample at 50°C for 5-10 min until the gel slices are dissolved. Vortex the sample every 2-3 min to help to dissolve the gel slices.
- Bind DNA: Load the sample on a NucleoSpin[®] Extract column and centrifuge for 1 min at 8,000x g. Discard flowthrough.
- Wash silica membrane:
 - Add 500 µl buffer NT2. Centrifuge for 1 min at 11,000x g and discard flowthrough.
 - Add 600 µl buffer NT3. Centrifuge for 1 min at 11,000x g and discard flowthrough.
 - Add 200 µl buffer NT3. Centrifuge for 2 min at 11,000x g to remove buffer NT3 and dry the silica membrane.

2. Material and Methods

- Elute DNA: Place the column into a clean 1.5 ml microcentrifuge tube. Add 25 μ l prewarmed (70°C) elution buffer NE (for fragments >5-10 kb) and incubate at room temperature for 1 min. Centrifuge for 1 min at 11,000x g.

2.2.5 DNA Cleanup

DNA cleanup is necessary before using the DNA for nucleotide sequencing, microarray analysis, PCR, restriction endonuclease digestion and other molecular processes.

Cleaning up DNA from aqueous solutions to remove buffer salts, enzymes or other substances can be done either using column-based kits, through ethanol precipitation or by performing an EXO/SAP digestion.

Column-based kits: The principle is that chaotrophic salts are added to the sample to denature the DNA by disrupting its hydrogen bonds. Under these conditions, the DNA will selectively bind to the silica resin in the column, allowing it to be separated from the rest of the sample. After washing the DNA is eluted from the column with a low salt solution. This allows the re-naturing of the DNA, which reduces its affinity for the silica.

The kit used was DNA Clean & ConcentratorTM-5 from Zymo Research.

2. Material and Methods

Standard protocol for DNA Clean& Concentrator™-5

- In a 1.5 ml microcentrifuge tube, add 2 volumes of DNA Binding Buffer to each volume of DNA sample. Mix briefly by vortexing.
- Transfer mixture to Zymo-Spin Column.
- Centrifuge at > 10,000 rpm for 45 sec. Discard flow-through.
- Add 200 µl Wash Buffer (with 24 ml 100% ethanol) to the column.
Centrifuge at > 10,000 rpm for 45 sec. Repeat wash step.
- Transfer the column to a 1,5 ml tube and add 6-10 µl water directly to the column matrix, centrifuge at > 10,000rpm for 45 sec to elute the DNA. Repeat the step once.

Ethanol precipitation: Is a method for de-salting and concentrating DNA. 0.1 to 0.5 M monovalent cations (normally in the form of the acetate salt of sodium) is added to the DNA, along with ethanol to a final concentration of 70%. Ethanol changes the DNA structure so that the DNA molecules aggregate and precipitate from solution. Since most salts and small organic molecules are soluble in 70% ethanol they stay in solution and the precipitated DNA can be separated from them by centrifugation.

DNA cleanup through EXO/SAP digestion:

- Mastermix for one reaction: 0.225 µl EXOI (20U/µl), 0.9 SAP (1U/µl) and 4.375 µl water
- Pipet 5.5 µl from the master mix in each 20.5 µl PCR product.
- Cycler program: 37°C for 25 min and 72°C for 15 min the DNA clean up is ready.

2.2.6 Cloning of PCR products

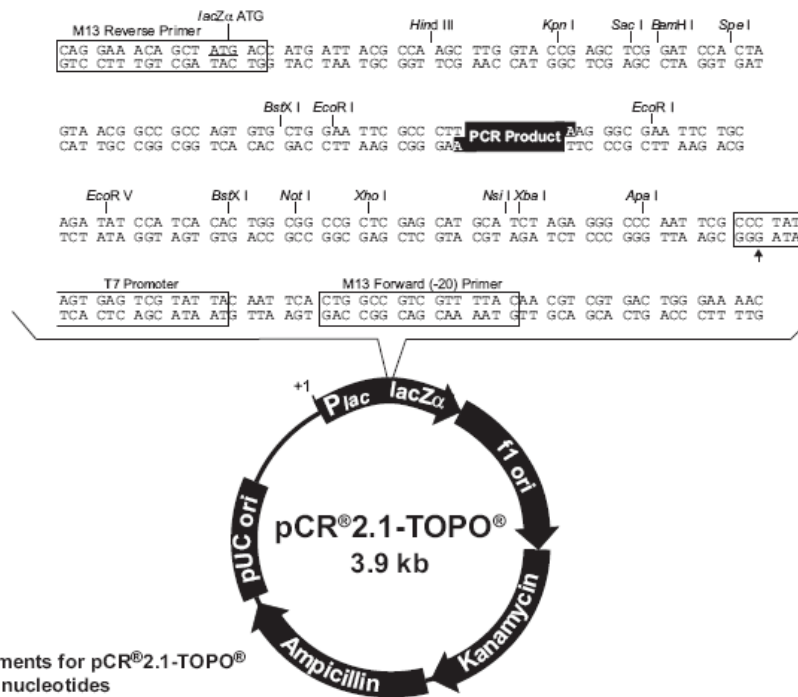
TOPO[®] TA cloning (Invitrogen)

TOPO[®] TA cloning provides a highly efficient one-step cloning strategy for the direct insertion of *Taq* polymerase-amplified PCR products into a plasmid vector.

Plasmid vectors pCR[®]II-TOPO[®] and pCR[®]2.1- TOPO[®] are supplied in linearized form with a single 3'-thymidine (T) overhang and topoisomerase I covalently bound to the vector. *Taq* polymerase has a nontemplate-dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3' ends of PCR products. This allows efficient ligation of PCR products into the vector. Topoisomerase binds to duplex DNA at specific sites and cleaves the phosphodiester backbone after 5'-CCCTT in one strand. The energy from the broken phosphodiester backbone is conserved by formation of a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue (Tyr-274) of topoisomerase I.

The phosphotyrosyl bond between DNA and enzyme can subsequently be attacked by the 5' hydroxyl of the original cleaved strand, reversing the reaction and releasing topoisomerase. The PCR products to be cloned should have a final extension step for at least 10 min. The cloning site on the vector (pCR[®]II-TOPO[®] or pCR[®]2.1- TOPO[®]) lies within the *lacZα* gene which codes for beta galactosidase. The ligation of PCR product disrupts the *lacZα* gene so that the substrate of beta-galactosidase, X-gal, can not be converted into an insoluble blue dye. This allows blue/white screening for the selection of insert-containing clones.

2. Material and Methods



◆ Standard protocol for TOPO[®] cloning

- PCR products: fresh PCR products are purified by gel extraction if more than one band is amplified or by using a DNA Clean-up Kit.
- TOPO[®] cloning reaction (Ligation reaction)
 - Set up reaction as follows:
 - 1) 4 μl H₂O
 - 2) 2 μl fresh PCR products
 - 3) 1 μl Ligation Buffer
 - 4) 2 μl TOPO[®] vector
 - 5) 1 μl T4 DNA Ligase
 - Mix gently and incubate at 14°C for at least 4 hours (preferably overnight).
- One Shot[®] chemical transformation

2. Material and Methods

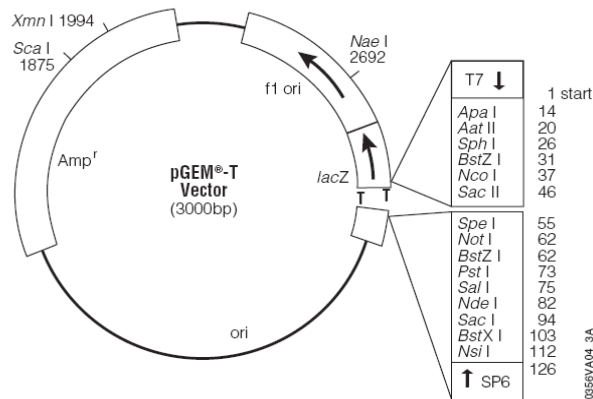
- Thaw One Shot[®] competent cells (supplied with the kit) on ice for 5 min.
- Add 1-2 μ l of ligation reaction to a vial of One Shot[®] TOPO10 *E. coli* competent cells and mix gently.
- Incubate on ice for 30 min.
- Heat-shock at 42°C for 30 sec without shaking. Transfer vials to ice.
- Recovery and plating
 - Add 250 μ l of S.O.C. medium at room temperature and recover the bacteria by shaking horizontally at 225 rpm for 1 hr at 37°C.
 - Spread 100-150 μ l bacterial suspension onto a selective plate. LB-agar plates should contain X-gal and 50 ng/ml kanamycin or 100 μ g/ml ampicilin and IPTG when using TOP10F' cells.
 - Incubate over night at 37°C.
 - Pick up white clones and transfer them in LB liquid medium with the adequate antibiotica and incubate them at 37°C overnight with shaking (~225 rpm).

pGEM[®]-T and pGEM[®]-T Easy Vector Systems (Promega)

These are convenient systems also for the cloning of PCR products and works on the same principal like the TOPO[®] TA cloning kit.

2. Material and Methods

pGEM[®]-T Vector Map and Sequence Reference Points



Bacterial cultures

Bacteria can be cultivated on agar plates or in liquid medium.

➤ Culture on agar plates

Streak the bacteria from glycerol stocks or LB agar stabs on a selective agar plate (with selective antibiotic, X-Gal and IPTG) and incubate at 37°C for 12-16 hours, overnight.

➤ Culture in liquid medium

Inoculate a single colony from agar plate in the corresponding selective LB liquid medium: 3 ml for mini-prep. Incubate at 37°C for 12-16 hours or over- night with shaking (~225 rpm).

Bacterial storage

➤ Short term storage: Streaked plate can be stored at 4°C for some months.

➤ Long term storage: Mix 850 µl of overnight cultured bacteria with 150 µl of 100% glycerol and store at -80°C for years.

2.2.7 Bisulfite conversion and cleanup of DNA for methylation analysis

Methylation of DNA occurs on cytosine residues, especially on CpG dinucleotides enriched in small regions of DNA (<500 bp). The methylation status of a DNA sequence can be determined using sodium bisulfite. Incubation of the target DNA with sodium bisulfite results in conversion of unmethylated cytosine residues into uracil, leaving the methylated cytosines unchanged. Therefore, bisulfite treatment yields to different DNA sequences for methylated and unmethylated DNA (see below).

	Original sequence	After bisulfite treatment
Unmethylated DNA	N-C-G-N-C-G-N-C-G-N	N-U-G-N-U-G-N-U-G-N
Methylated DNA	N-C-G-N-C-G-N-C-G-N	N-C-G-N-C-G-N-C-G-N

Conversion of unmethylated cytosines is achieved by incubating the DNA in high bisulfite salt concentrations at high temperature and low pH. These harsh conditions usually lead to a high degree of DNA fragmentation and a substantial loss of DNA during purification.

The EpiTect Bisulfite Kit provides a fast procedure for efficient DNA conversion and purification of as little as 1 ng DNA. DNA fragmentation is mostly prevented during the bisulfite conversion reaction by DNA Protect Buffer that contains a pH-indicator dye as a mixing control in the reaction setup, ensuring the correct pH for cytosine conversion.

2. Material and Methods

The bisulfite thermal cycling program provides an optimized series of incubation steps necessary for thermal DNA denaturation and subsequent sulfonation and cytosine deamination, enabling high cytosine conversion rates of over 99%. Desulfonation, the final step in chemical conversion of cytosines, is achieved by a convenient one-column step included in the purification procedure.

For the bisulfite conversion ~ 500 ng of genomic DNA were used.

◆ **Standard protocol for bisulfite conversion and cleanup of DNA**

Bisulfite conversion:

- Thaw DNA and dissolve the required number of aliquots of Bisulfite Mix by adding 800 μ l RNase-free water to each aliquot. Vortex until the Bisulfite Mix is completely dissolved.
- Prepare bisulfite reactions in 200 μ l PCR tubes as follow:

Component	Volume per reaction (μl)
DNA solution (1 ng – 2 μ g)	Variable*(maximum 20)
RNase-free water	Variable*
Bisulfite Mix (dissolved), see step 1	85
DNA Protect Buffer	35
Total volume	140

- Perform the bisulfite DNA conversion on a thermal cycler

2. Material and Methods

Bisulfite Conversion Thermal Cycler Conditions

Step	Time	Temperature
Denaturation	5 min	99°C
Incubation	25 min	60°C
Denaturation	5 min	99°C
Incubation	85 min (1h 25 min)	60°C
Denaturation	5 min	99°C
Incubation	175 min (2h 55 min)	60°C
Hold	Indefinite [†]	20°C

Cleanup of bisulfite converted DNA:

- Once the bisulfite conversion is complete, centrifuge the PCR tubes containing the bisulfite reactions briefly, and then transfer the complete bisulfite reactions to 1.5 ml microcentrifuge tubes.
- Add 560 µl freshly prepared Buffer BL, mix the solution by vortexing and centrifuge. 10 µg/ml carrier RNA must be added when using low (<100 ng) amount of DNA. Transfer the whole mixture into the EpiTect spin column. Centrifuge the column at maximum speed for 1 min. Discard the flow-through.
- Add 500 µl washing buffer BW (with 30 µl 96-100% ethanol) and centrifuge at maximum speed for 1 min. Discard the flow-through.
- Add 500 µl desulfonation buffer BD (with 27 µl 96-100% ethanol) to the spin column and incubate for 15 min at room temperature.
- Centrifuge the column at maximum speed for 1 min, discard the flow- through.
- Add 500 µl buffer BW and centrifuge at maximum speed for 1 min. Repeat the step once.

2. Material and Methods

- Place the spin column into a new collection tube and centrifuge the spin column at maximum speed for 1 min to remove any residual liquid.
- Place the spin column into a 1.5 ml tube add 20 μ l elution buffer EB to the centre of the membrane. Elute the purified DNA by centrifugation for 1 min at 15,000 x g.

2.2.8 PCR (Polymerase Chain Reaction)

PCR is an enzymatic method for exponential amplification of specific DNA fragments *in vitro* (Saiki et al., 1985). Since the amplified products from the previous cycle serve as templates for the next cycle, the amplification is an exponential process and a highly sensitive technique for nucleic acid detection.

PCR depends on a pair of oligonucleotide primers that are designed so that a forward or sense primer directs the synthesis of DNA towards a reverse or antisense primer, and vice versa. The heat stable Taq DNA polymerase (Chien et al., 1976) catalyses synthesis of a new DNA strand that is complementary to the template DNA from the 5' to 3' direction by a primer extension reaction, producing a DNA molecule that is flanked by the two primers. The amplification reaction requires the four deoxyribonucleoside triphosphates (dNTPs) as substrates, a buffer and salts including MgCl₂. The PCR reagents have been standardized and appropriate conditions such as temperature and concentrations have been defined for certain primer amplifications.

Each amplification cycle requires **1.** denaturation of DNA molecules at 95°C, **2.** annealing that is hybridization of DNA primers in a temperature range between of 40 °C – 68 °C, and **3.** extension, that is the synthesis (replication) of new DNA strands by DNA polymerase (which catalyzes growth of the new strand from the 5' → 3') between the primers at 72 °C. To obtain a good yield of the final product 25-35 PCR cycles are usually sufficient.

2. Material and Methods

2.2.8.1 Standard PCR

The annealing temperature for standard PCR depends on the melting temperature of the primers used. The FastStart Taq DNA Polymerase, dNTPack from Roche was used for amplification. Standard PCR was performed in total volume of 25 μ l or 50 μ l.

Reagents for standard PCR

10x PCR buffer (with MgCl ₂)	5 μ l
10 mM dNTPs	1 μ l
10 μ M forward primer	2,5 μ l
10 μ M reverse primer	2,5 μ l
2 U/ μ l Taq Polymerase	0.4 μ l
DNA template	Variable (1 ng – 100 ng)
H ₂ O	add up to 50 μ l

◆ **Standard PCR program:**

- 1) First denaturation 95°C for 4 min
- 2) Denaturation 95°C for 30 sec
- 3) Annealing Variable (Primer specific) for 30 sec
- 4) Elongation 72°C for 45 sec[#]
- 5) Cycles Go to step 2) for 34 cycles
- 6) Final extension 72 °C for 7 min
- 7) Hold 4 °C for unlimited time

If the predicted size of PCR is > 1kb, prolong elongation time 1 min per kb.

2.2.8.2 RT (Reverse Transcription) PCR

RNA cannot serve as a template for PCR, so it must first be reversely transcribed into cDNA. RT-PCR is a combined technique in which reverse transcription (RT) is coupled with PCR amplification of the resulting cDNA and is very useful for determining the expression of genes in specific tissues. cDNAs were synthesized from varying amounts of total RNA using the SuperScript™ III Reverse Transcriptase from Invitrogen. Usually 1-2 µl total RNA were used for cDNA synthesis.

◆ First-strand cDNA synthesis

➤ Denaturation:

- Mix samples as follows:
 - 1) 1 µl of 10 mM dNTP mix
 - 2) 1 µl of oligo (dT)₂₀ primer (50 µM)
 - 3) 1 µl of random primer (50-250 ng)
 - 4) 0.5-5 µg total RNA
 - 5) Add H₂O up to 13 µl
- To avoid possible secondary structures of RNA, which might interfere with the synthesis, heat the reaction mixture at 65°C for 5 min, and then chill it on ice for >1 min.

➤ cDNA synthesis:

- Add the following reagents to the mixture
 - 1) 4 µl of 5x First-stand buffer
 - 2) 1 µl 0.1 M DTT

2. Material and Methods

- 3) 1 μ l RNaseOUTTM Recombinant RNase inhibitor
 - 4) 1 μ l SuperScriptTM III RT
- Mix by pipetting gently up and down and then incubate the mixture at 25°C for 5 min
 - Incubate at 50°C for 40-60 min
- Inactivation of the reaction and removal of RNA:
- Incubate at 70°C for 15 min to inactivate the reaction.

2.2.8.3 Quantitative Real-time PCR

Real time PCR is based on the detection and quantification of a fluorescent reporter molecule. This signal increases in direct proportion to the amount of PCR product in the reaction. By recording the amount of fluorescence emission at each cycle, it is possible to monitor the PCR reaction during the exponential phase where the first significant increase in the amount of PCR product correlates with the starting amount of target DNA template.

In order to compare the relative *CCRK* mRNA levels in human and non-human primates, quantitative real time RT-PCR analyses of total RNAs were performed using the QuantiTect SYBER Green PCR Kit on an Applied Biosystems 7500 Fast Real-Time PCR system.

Measurements of all cDNA samples were done in triplicates in a 25 μ l reaction volume. For each run the so called “No Template Control or Negative Control“ (NTC) was added. All assays were forecast at the end with the melting curves analyses.

2. Material and Methods

The beta-2-microglobulin *B2M* gene (QT 00088935; Qiagen), was chosen as endogenous control because it showed rather constant expression levels in human and non-human primate cortex.

Because the commercially available QuantiTect Primer Assays did not work for all primates we had to design our own RT-PCR primers in evolutionarily conserved regions (without nucleotide changes between the analyzed species). The forward primer is localized in *CCRK* exon 2 and the reverse primer in 3, spanning 2825 bp intronic sequence. The primers were tested for their hybridization specificity first by a BlastN-search in Ensembl and NCBI and second through a standard PCR where the expected specific 239 bp amplification product was obtained.

For each template and gene (*CCRK* and *B2M*) a standard curve was generated by serial dilution steps (1:1, 1:2, 1:4, 1:8, 1:16, 1:32, 1:64, 1:128) of the corresponding cDNA. This is a “must” when inter-species quantitative real-time is performed.

The standard curve method does not only simplify calculations but also avoids problems associated with PCR efficiency assessment. Using this standard curve, the efficiency of each PCR reaction was derived from the slope (that was between -3.19 and -3.32) of each standard curve using the formula: $E = 10^{(-1 / \text{slope}) - 1}$. The calculated efficiencies were $\geq 96\%$.

The results of the real time RT-PCR were analyzed by using Applied Biosystems 7500 Fast System SDS Software version 1.3 and the Qty values that were calculated by extrapolation from the standard curve of each sample.

2. Material and Methods

Reagents for Quantitative RT-PCR for a 25 μ l reaction volume

2x QuantiTect SYBER Green PCR Master Mix	12,5 μ l
Forward Primer 10 pM	1,25 μ l
Reverse Primer 10 pM	1,25 μ l
Template cDNA	10 μ l (25 ng template/reaction)

◆ Quantitative Real-time PCR program:

- 1) PCR Initial activation step 95°C for 15 min
- 2) Denaturation 94°C for 15 sec
- 3) Annealing Variable (50-60°C) for 30 sec
- 4) Extension 72°C for 40 sec
- 5) Cycles Go to step 2) for 40 cycles

2.2.9 DNA Sequencing

DNA sequencing was performed with the CEQ 2000 Dye Terminator Cycle Sequencing (DTCS) Quick Start Kit and analyzed on CEQ™ 8000 Genetic Analysis System from Beckman Coulter™. DNA sequencing was done with Fluorophor-labeled Dideoxynucleotide-Threephosphates (ddNTPs), which were mixed with the DNA and the sequencing primer in the terminal sequencing reaction. The ddNTPs are one component of the used CEQ 2000 Dye Terminator Cycle Sequencing (DTCS) Quick Start Kit. Via the laser of the CEQ™ 8000 Genetic Analysis System from Beckman Coulter the coupled fluorochromes are stimulated and detected.

◆ **Standard protocol for DNA or plasmid DNA sequencing**

- Set up sequencing reaction as follows:
 - 1) 5 µl H₂O
 - 2) Cleanup 2,5 µl PCR product or 300-500 ng of plasmid DNA template
 - 3) For plasmid DNA: Denature the template at 96°C for 2 min, chill on ice for at least 2 min.
 - 4) Add 0.5 µl 10 µM primer.
 - 5) Add 2 µl DTCS Quick Start Master Mix.
- Thermal cycling program:
 - 1) 96°C for 20 sec
 - 2) 50°C for 20 sec
 - 3) 60°C for 4 min
 - 4) Go to step 1) and repeat 34 cycles

2. Material and Methods

- 5) Hold at 4°C.
- Ethanol precipitation:
 - In a 1,5 ml centrifuge tube, transfer 10 µl of the sequencing reaction, 10 µl water, 4 µl precipitation mixture (containing 2 µl 3M sodiumacetat and 2 µl 100 mM EDTA) and 60 µl 95% ethanol (-20°C). Mix by vortexing.
 - Centrifuge for 15 min at 14,000 rpm at 4°C.
 - Carefully remove the supernatant.
 - Add 100 µl 70% ethanol (-20°C) and centrifuge for 10 min at 14,000 rpm at 4°C.
 - Remove the supernatant and repeat the last step again.
 - Dry the DNA pellet for 8 min in a speed-vac.
 - Resuspend the pellet in 30 µl SLS (Sample Loading Solution) (provided with the kit). The sample can be kept in SLS at -20°C for up to one week.
 - Sample preparation for loading onto the sequencer:
 - Transfer the resuspended samples to the appropriate wells of the CEQ sample plate.
 - Overlay each of the resuspended samples with one drop of light mineral oil (provided with the kit).
 - Load the sample plate onto the CEQ and start the desired method.

Sequencing electropherograms were viewed with FinchTV version 1.4.0

2.2.9.1 Bisulfit-sequencing

Many techniques exist for measuring the methylation state of a single CpG but for analysis of an entire region, cloning and sequencing remains the standard. This method allows cytosine and 5-methylcytosine to be distinguished, because of the selective deamination of the unmethylated cytosine to uracil following sodium bisulfite treatment. After conversion unmethylated regions of DNA contain no cytosine. The double-stranded DNA product after amplification is abnormal in two aspects: it contains only three nucleotide types in each strand (A, G, T versus A, C, T), and each strand has an excess of one nucleotide (T or A) that can be sometimes a problem in the direct sequencing method. Other important factors that can influence the direct sequencing method are the quality of the primers and longer stretches of a certain nucleotide.

For the selected regions, bisulfite sequencing of direct PCR Products or cloned PCR products combined with a limiting dilution approach was used to confirm methylation status of CpG sites within the CGI promoters. For this analysis, we cloned the PCR products into the pGEM-T (Promega) or pCR[®]2.1- TOPO[®] (Invitrogen) vector, extracted plasmid DNA from the resulting clones and sequenced it. One important problem of classic bisulfite sequencing is an amplification bias in the PCR reaction that leads to preferential amplification of only a few DNA molecules from the heavily degraded bisulfite-treated DNA in the starting sample and consequently to overrepresentation of certain alleles in the PCR product.

2. Material and Methods

To avoid such an amplification bias, the bisulfite-treated DNA was diluted 1:50 and 6-8 independent PCR products were generated from each starting samples. Approximately 6-8 plasmid clones from different PCR replicates were sequenced to obtain a representative view on the methylation status of the analyzed CpG island. As a control, it was not just looked at the critical CpG sites but also at the conversion of non-CpG cytosines in each plasmid. This allowed one to distinguish between different alleles.

2.2.9.2 Pyrosequencing

Pyrosequencing is a quantitative method to assess DNA methylation levels at specific CpG sites. It is a real-time DNA sequencing-by-synthesis method that relies on an enzymatic cascade that quantitatively converts the pyrophosphate released upon nucleotide incorporation into a luminometric signal.

The pyrophosphate released in the DNA synthesis reaction is quantified by monitoring a luciferase reaction and this produces a signal proportional to the number of pyrophosphate molecules released. It is ideally used for DNA methylation analysis after bisulfite treatment, as it combines the capability for direct quantitative sequencing with reproducibility and speed.

Pyrosequencing is extremely useful for methylation analysis of rather short DNA sequences containing critical CpG sites with key roles, i.e. important transcription-factor binding sites.

2. Material and Methods

Limitations of this method are first the length of the sequence read, and thereby the number of CpGs that can be analyzed in one sequencing reaction. So the method is not suitable for methylation analysis of complete CpG islands.

If the goal is to compare the methylation status of CpGs between different species, it is very difficult to find pyrosequencing primers without nucleotide changes between the analyzed species (commercial available assays are designed for the human and mouse but not for chimpanzee, baboon or other species). The analysis of one fragment requires three primers, the PCR primers (one of them is biotinylated) and one pyro-sequencing primer.

The sequencing reactions were performed using an automated PSQ 96MA System from Biotage. This technique was chosen for the analysis of methylation status of specific CpG sites in a group of candidate genes that are important for imprinting, or associated with human disease.

◆ **Standard protocol for Pyrosequencing**

- In the PCR plate pipet

PCR Product	40 μ l
Binding Buffer	40 μ l
Sepharose Beads	5 μ l

Using a mixer/shaker incubate the plate at room temperature for 5 min agitating constantly to keep the beads dispersed

- In the Probe plate of the pyro pipet

Pyro-primer	1,6 μ l (10 pmol/ μ l)
Annealing Buffer	38,4 μ l

2. Material and Methods

➤ Vacuum Prep Tool

- 1) Wash the probe tips of the Vacuum Prep Tool by lowering it into high purity water (Parking Position) for 20 sec and then let it dry for 1 min.
- 2) Capture the beads containing the immobilized templates on the filter probes by slowly lowering the Vacuum Prep Tool into the PCR plate, for 10-20 sec. Make sure that the liquid has been aspirated from all wells and that all beads have been captured onto the probe tips.
- 3) Move the Vacuum Prep Tool to the cuvet containing 70% ethanol and let the solution flush through the filters for 5 sec.
- 4) Move to the denaturation solution and flush through the filters for 5 sec.
- 5) Move to the washing buffer and flush the filters for 5 sec.
- 6) Close the vacuum.
- 7) Release the beads in the Probe plate of the pyro containing the pyro primer and the annealing buffer by intense shaking.
- 8) Incubate the Probe plate at 80°C for 2 min and then let it for 2-3 min at room temperature.
- 9) Load the Dispensing Cartridge with the enzyme, substrate and the nucleotides.
- 10) Load the pyro plate and the cartridge in the PSQ 96MA System.

The Enzyme, Substrate (light sensitive) and the nucleotides are pipeted into the pyrosequencing cartridge according to the instructions manual.

2.2.10 Bioinformatic Tools

2.2.10.1 In silico analysis of the coding sequences in the breakpoint regions

To elucidate the question whether chromosomal rearrangements contributed to the human-chimpanzee speciation in silico analyses of the coding sequences flanking the pericentric inversion breakpoints (BP) between human (HSA) on chromosomes 1, 4, 5, 9, 12, 17 and 18 and the homologues chimpanzee (PTR) chromosomes were performed. Two inversions are human specific - chromosomes 1 and 18 - and five are chimpanzee specific - chromosomes 4, 5, 9, 12 and 17. Chromosome 15 and 16 were excluded because of low quality of the chimpanzee draft sequence at that time, to avoid “in silico evolution” results.

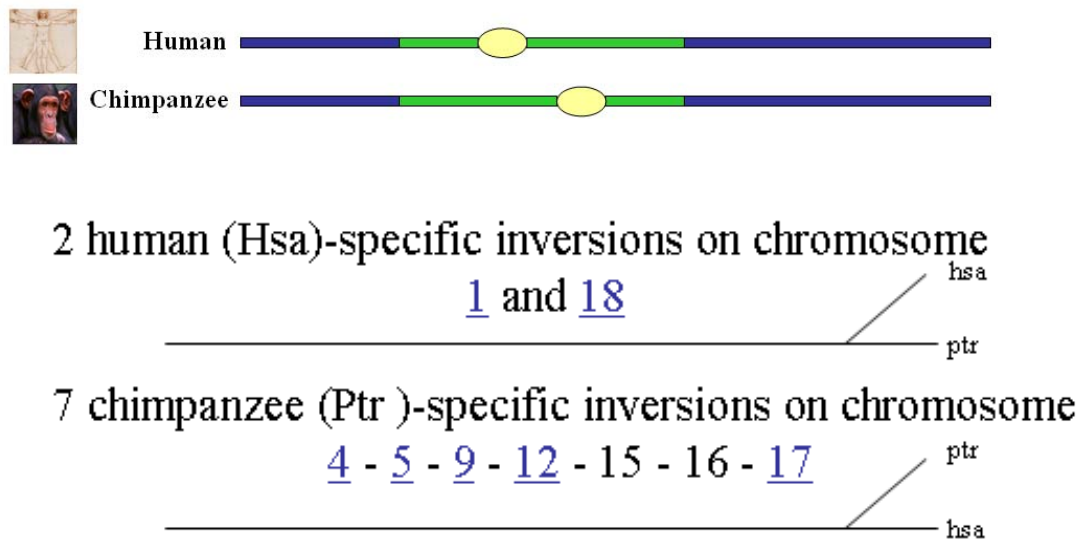


Figure 4. The nine pericentric inversions that distinguish human and chimpanzee karyotypes.

2. Material and Methods

The sizes of the breakpoint flanking regions (BP-FR) were chosen to be 2Mb on either side of the break and for comparison a control region (CR) of 4Mb at 10Mb distance to the breakpoint region.

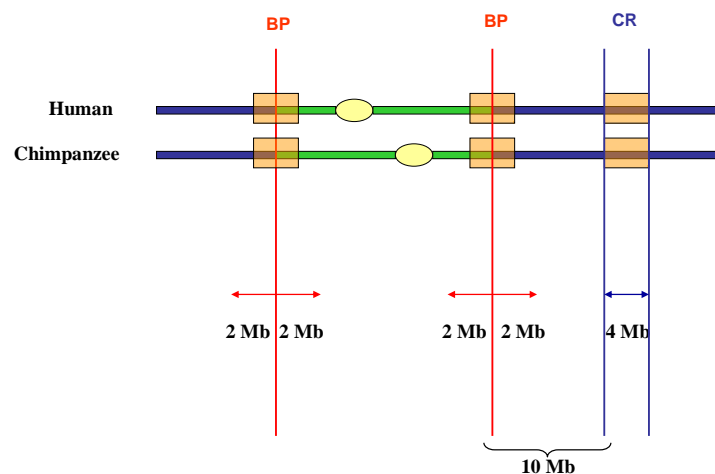


Figure 5. Breakpoint and control region for in silico sequence analysis

Human and chimpanzee coding sequences of known and predicted genes in the breakpoint flanking and control regions (BP-FR and CR) were downloaded from the Ensembl release 39-42 and NCBI database, and aligned in BioEdit with ClustelW. If a sequence or sequence region from chimpanzee was missing in the database, the homologues human sequence was used for an Ensembl- BLAST against the chimpanzee genome. The resulting chimpanzee sequence was taken for “Gap removal“. At the end the HSA and PTR sequences were again aligned.

2. Material and Methods

2.2.10.2 BioEdit version 5.0.6

BioEdit is a biological sequence editor that runs on Windows 95/98/NT/2000 and provides basic tools for protein and nucleic sequence editing, alignment, manipulation and analysis. BioEdit version 5.0.6 was used for all the in silico coding sequence analyses of the breakpoint region genes and other analyzed sequences between human and non-human primates.

The alignments of the sequences were done with the ClustalW multiple Alignment tool. ClustalW is a program which was designed by Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) to construct multiple alignments of biological sequences. ClustalW will automatically align many sequences with a profile- based progressive alignment procedure.

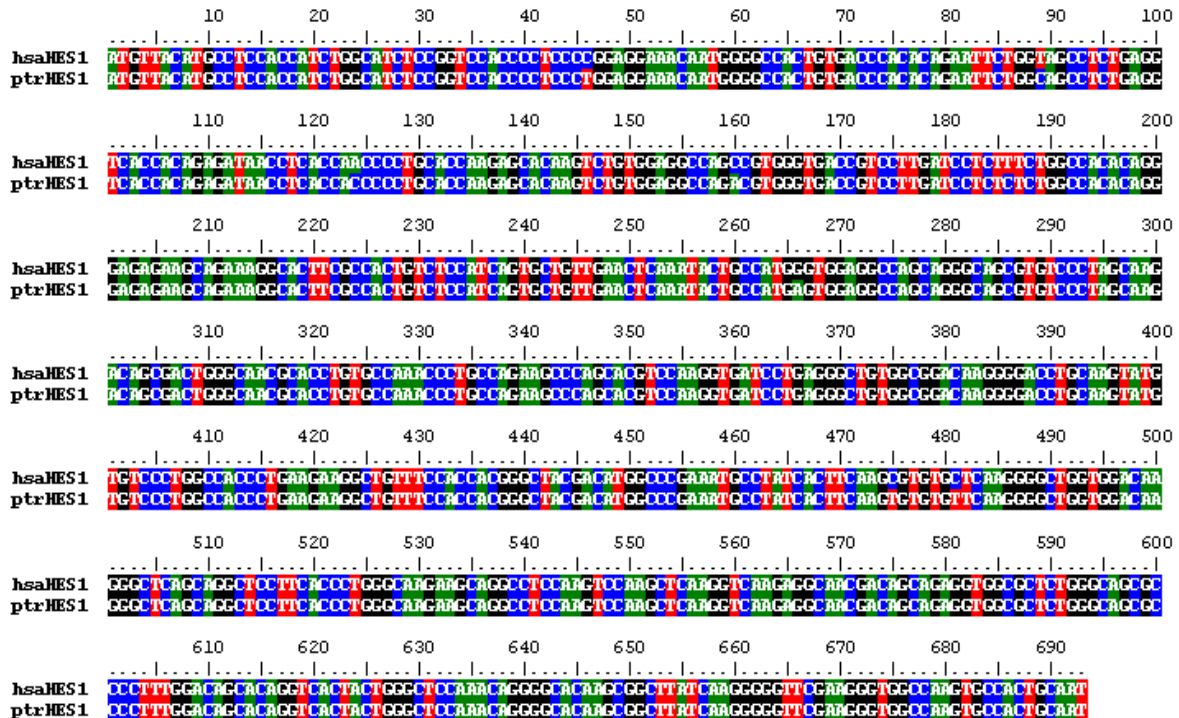


Figure 6. Example of an alignment for the coding sequence of the *HES1* gene

2. Material and Methods

This program was used to design PCR primer for interspecies comparisons. The sequences of interest from the human and the non-human primates were first aligned and then the primers were chosen in evolutionarily conserved regions (without nucleotide changes between the analysed species).

2.2.10.3 MEGA 3.1 version (Molecular Evolutionary Genetics Analysis software)

The software was designed for comparative analysis of orthologous gene sequences from different species with emphasis on the evolutionary relationships and patterns of DNA and protein evolution.

MEGA 3.1 was applied for the calculation of positive selection of the coding sequences in the breakpoint region. As measure for positive selection i.e. adaptive evolution on a protein – coding sequence the dN/dS (Ka/Ks) relative rate test was used. A ratio of nonsynonymous (dN, amino acid substitutions) to synonymous (dS) substitution rates >1 indicates the action of positive selection. If the ratio $dN/dS = < 1$, we assume a negative (or purifying) selection process.

$dN/dS = 1$ corresponds to a neutral selection.

The modified Nei-Gojobori with Jukes-Cantor-Correction- Model was chosen for pairwise sequences comparison. This method computes the numbers of synonymous and nonsynonymous substitutions and the numbers of potentially synonymous and potentially nonsynonymous sites (Nei and Gojobori 1986). Based on the correction of Jukes-Cantor the p-distances computed above can be corrected to account for multiple substitutions at the same site. This model takes transitions/ transversions and multiple substitutions into consideration.

2. Material and Methods

2.2.10.4 CpG Island search

The programs CpG Island Searcher or EBI Tools CpG Islands were used to identify CpG islands in putative cis-regulatory regions of genes of interest 500 bp downstream to 10 kb upstream of the transcription start site (TSS). When a gene contained two or more CpG rich segments, usually the CpG island nearest to the TSS was analyzed. Based on the CpG ratio, GC content and length of CpG-rich region and using the formula **(number of CpGs x number of bp) / (number of Cs x number of Gs)** for the CpG ratio two classes of CGI were defined: High-CpG promoters (HCPs) and Intermediate CpG content promoters (ICPs).

2.2.10.5 Primer Design

For all analyzed gene sequences, the well known Ensembl (<http://www.ensembl.org>) and NCBI (<http://www.ncbi.nlm.nih.gov>) database were used to find the transcript sequences that are without nucleotide changes between human and non-human primates. Sequences of interest were aligned with BioEdit and the primers were selected from evolutionarily conserved regions that did not exhibit nucleotide changes among humans and non-human primates. The number of hits for each individual primer after BLAST (Basic Local Alignment Search Tool) should be one.

BLAST offers the following programs: **BLASTN**- search a nucleotide database with a nucleotide query, **BLASTP**- search protein database with a protein query, **BLASTX**- search a protein database with a six-frame translation of a nucleotide query sequence and others.

2.2.10.6 RepeatMasker (<http://www.repeatmasker.org>)

The Program screens DNA sequences against a library of repetitive elements and creates a masked query sequence which can be used for database searches. It also generates a table annotating the masked regions.

2.2.10.7 Giri (<http://www.girinst.org>)

This software tool screens query sequences against a collection of repeats and generates a report classifying all identified repeats in the query sequence. Using this web site, an Alu-Sg1 repeat was detected in the *CCRK* gene of humans and non-human primates.

2.2.10.8 Transcriptional Regulatory Element Database (<http://rulai.cshl.edu>)

The database gives information on cis- and trans- regulatory elements in mammals, including transcription factor binding motifs of promoter regions. The program was used for detecting the transcription factor binding sites in the putative promoter regions of the analysed genes.

2.2.10.9 Panther Database (<http://www.pantherdb.org>)

The database is extremely useful for the classification of genes into different biological processes or molecular functions. It offers also information on genes pathways. It has to be noted that a gene may belong to more than one processes or pathway.

The Panther database was used for the functional classification of the positively selected genes and genes under purifying selection in the evolutionary breakpoint regions of human and chimpanzee.

2.2.10.10 Novartis gene expression Atlas (<http://symatlas.gnf.org/SymAtlas>)

The GeneAtlas data provides expression data for all protein-encoding transcripts across more than 100 human tissues and hundreds of microarray experiments.

3. RESULTS

3.1 Evolution of coding sequences in the evolutionary breakpoint regions

The human and chimpanzee karyotypes are distinguished by nine pericentric inversions involving the homologues of human (HSA) 1, 4, 5, 9, 12, 15, 16, 17 and 18. It was proposed that these nine pericentric inversions facilitated the speciation process that separated the human and chimpanzee lineages (Navarro and Barton, 2003; Rieseberg and Livingstone, 2003; Valender and Lahn, 2004; Zhang et al., 2004). The “recombination suppression model” of speciation suggests that positive selection operated more intensely on rearranged than on collinear chromosomes during chromosomal speciation. In this thesis I analyzed the potential relevance of the nine pericentric inversions to the human/chimpanzee speciation process (the impact of these inversions on selection) by comparing DNA divergence rates in the coding sequences of 2 Mb regions flanking the evolutionary breakpoints (BP-FR) on chromosomes 1, 4, 5, 9, 12, 17 and 18. The control regions (CR), collinear regions on rearranged chromosomes, are represented by a 4 Mb region localized at 10 Mb distance of the breakpoints. To avoid “in silico evolution” results, chromosome 15 and 16 were excluded because of poor sequence quality at the time of analysis. The rate of protein evolution of a gene can be used to uncover the footprints of past positive selection and can be measured as the ratio of non-synonymous (dN) to synonymous (dS) substitutions rates.

3.1.1 Evolution of the coding sequences in the breakpoint regions of chromosome 1, 4, 5, 9, 12, 17 and 18

The availability of the full genome sequences of human and chimpanzee, the existence of computational methods for comparing them and the possibility of an precise mapping of the inversion boundaries at the DNA sequence level are prerequisites for the identification of differences in the rates of protein evolution between the rearranged and collinear regions.

The total number of genes analyzed was 334 in the breakpoint flanking regions (BP-FR) and 72 genes in the control regions (CR). The 334 genes in the analyzed BP-FRs exhibit different patterns of distribution. For instance the examined BP-FR of chromosome 18 was the most gene poor region while the BP-FR of chromosome 17 was very gene rich. Therefore the number of genes studied in the BP-FRs of chromosome 17 was limited to around 30 on either sides of the break.

It is important to note that the studied pericentric inversions differ in their complexity. Six of the nine pericentric inversions map to regions with a complex genomic structure due to segmental duplications (SD). The human and chimpanzee coding sequences of the analyzed genes were downloaded from the Ensembl Genome Browser using the latest versions available (release 39 up to 42) during the timespan of investigation and aligned. Existing results were randomly rechecked to rule out biased results.

In Ensembl release 39 and 40, the HSA-PTR inversion breakpoints on chromosome 1 were on the q-arm, so the centromer was not directly involved. The quality sequence information of both species in the region was generally of lower quality compared to the other regions under investigation. Of 31 genes in the BP-FR on chromosome 1, 3 were found to be positively selected (*NUD17*, *ZNF364* and *ACP6*) between human and chimpanzee, representing 9.6 %.

3. Results

In the CR 24 genes were analyzed and 1 gene (*IFI16*) (4.1%) was found to be positively selected.

5 genes (16.1%) (*POLR3C*, *CD160*, *HIST2H3C*, *HIST2H2AC* and *SF3B4*) were negatively selected in the BP-FR. The same numbers of negatively selected genes (*IGSF4B*, *TAGLN2*, *KCNJ10*, *WDR42A* and *PEX19*) were found also in the control region (20.8%).

These results confirm that on chromosome 1 there are no differences between the two analysed regions with respect to positive or negative selection, there is more or less a random distributions of the selected genes.

The human-specific pericentric inversion of chromosome 18 was in Ensembl release 40 detectable and included the complete human 18p arm. The BP-FR on chromosome 18 contained one positively selected gene (*CLUL1*) (2.9%) and 6 negatively selected genes (17.6%) of 34 genes analyzed.

The chimpanzee specific pericentric inversion of chromosome 4 was in Ensembl release 39 and 40 detectable. The chromosomal position of the breakpoint on HSA 4p maps at 44.505.120-44.505.140 Mb and on HSA 4q at 86.177.195-86.177.215 Mb. 9 out of 23 genes showed no non-synonymous changes between human and chimpanzee. Four of these 9 genes (*GABRG1*, *GABRA2*, *GABRA4*, *GABRB1*) belong to the *GABA*-receptors family. This class of receptors respond to the neurotransmitter gamma-aminobutyric acid (*GABA*), the main inhibitory neurotransmitter in the vertebrate central nervous system. In the BP-FRs of chromosome 4 no gene was found to be positively selected. In the gene poor CR, one of 8 analysed genes (*NM_777*) (12.5%) was as positively selected and 2 genes (*TSPAN5* and *EIF4E*) (25%) had a ratio of $dN/dS = 0$. For more details of the analysis of the pericentric inversion see Figure 7. The three cases representing $dN/dS=0$ are marked as follows:

Genes with $dN=0$ are marked with a star

Genes with $dS=0$ are marked by a blue arrow

Genes with $dN=0$ and $dS=0$ are marked by a yellow arrow

3. Results

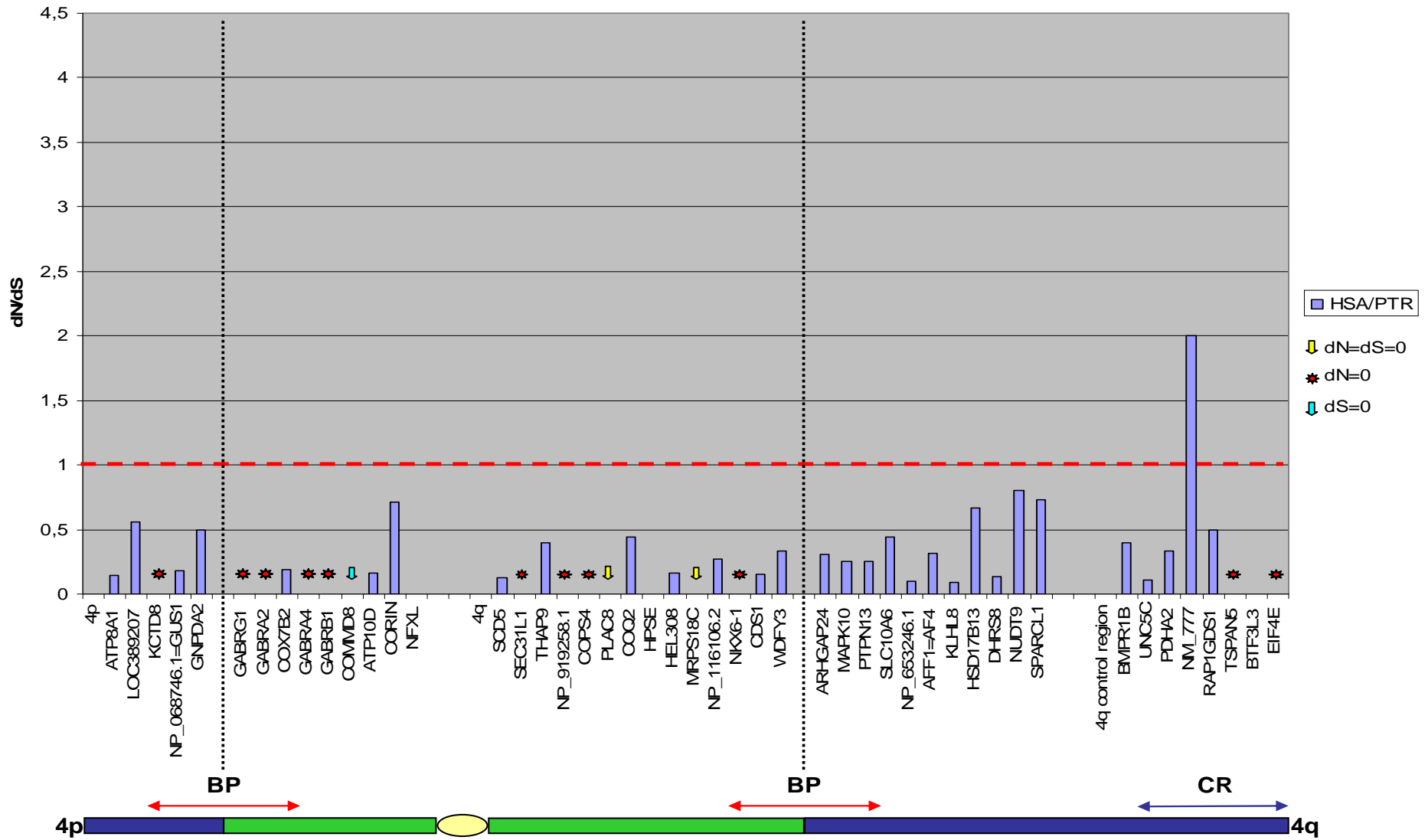


Figure 7. Chromosome 4 pericentric inversion HSA-PTR

3. Results

The human chromosome 5 represents the ancestral state. The pericentric inversion breakpoint on 5q maps between the *PCSK1* and *CAST* genes and on 5p between the *BASP1* and *CDH18* genes. Chromosome 5 is endowed with numerous relatively recent intra-chromosomal duplications (Schmutz et al., 2004). However, these duplications are several megabases distant to the inversion breakpoints.

With 2 out of 30, chromosome 5 showed the lowest percentage of selected genes (Figure 8). Two genes (*ANKRD32* and *FAM81B*) were positively selected (6.7%) and one (*RFESD*) (3.3%) was negatively selected between human and chimpanzee. The same pattern was observed also for the CR with 6 analyzed genes.

A rather complex pericentric inversion is on chromosome 12. According to Sawatzki et. al., (2005) the pericentric inversion of chromosome 12 was defined as exceptional, since the inversion is associated with two large duplications and disrupts the structure of the gene *SLCO1B3* on 12p. One duplication includes the functional *SLCO1B3* locus, which is thus retained in the chimpanzee, whereas the second duplication does not contain any expressed sequences. The BP on 12q maps between the *DYRK2* and *IFNG* genes. The precise positions of the breakpoints are 20.854.544-20.854.476 Mb on HSA 12p and 66.667.781-66.667.800 Mb on HSA 12q. In the BP-FRs of chromosome 12, two out of 57 analyzed genes were positively selected (3.5%) and 15 (26.3%) were negatively selected. The control region contains two positively and two negatively selected genes.

3. Results

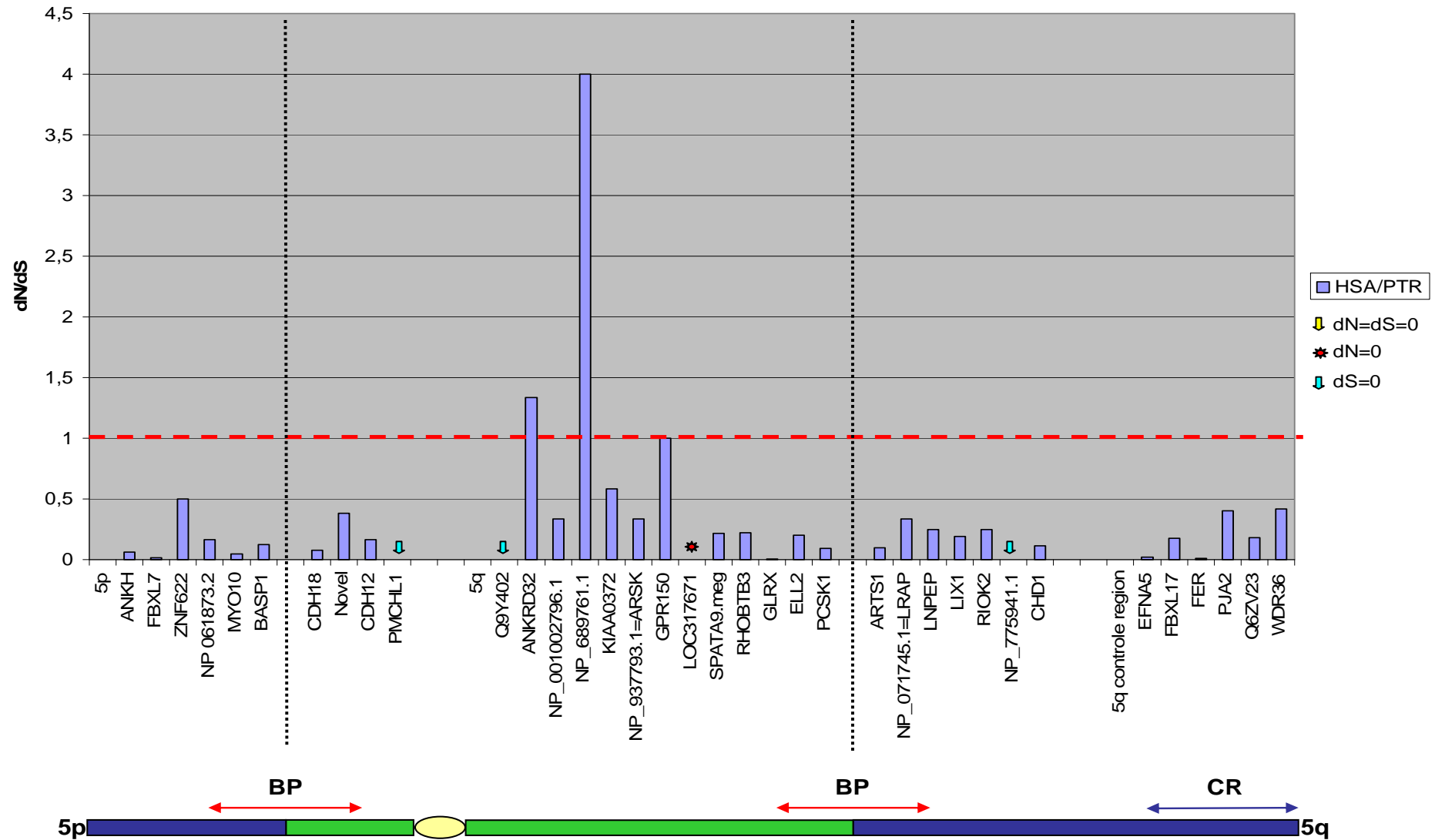


Figure 8. Chromosome 5 pericentric inversion HSA-PTR

3. Results

Molecular cytogenetic studies have identified SD in the breakpoint region of chromosome 9p, while the breakpoint in 9q mapped close to the centromer alpha-satellite DNA, suggesting that in addition to the pericentric inversion, centromer repositioning or a second smaller pericentric inversion has occurred. The inversion of chromosome 9 was not detectable in Ensembl release 39 and 40.

In the figure 9 below, the gray region represent the BLAST hits with DQ000185.1 and DQ000183.1 which are the PTR fusion sequence (5' and 3') ends as deposited in the NCBI database. Consequently the regions should be on HSA 9p and 9q arms, which is obviously not the case.

```
>chromosome:NCBI36:9:88036401:88041775:1
88036401 TTTGTCACATTTGTCACTACTGAGATCACTATATTGATATACAGGACTAGAAAAGCAAT 88036460
88036461 ACAGTGCAGCATGCCTGCCACAGACCTGTTTATCTCTAGGCAATTACAATGAGACTTGAT 88036520
88036521 TGTGCTTTCTCCTCCTGCAAAGGAAGGAAATTTTCTGCAAAGGAAAATTTCTCTAATG 88036580
88036581 TGCTTTACCTAAAACCTGTCAATGTCTTCCAGTCAATCATTGCTATTGTCTTTCAGACAG 88036640
88036641 CTTCAGAGTTGTCCACCATAGGAAGGTTGCCCTTCTCCCATCACCAAGGTTAATAAATTT 88036700
88036701 TCTCCTCCCTTCCCTCAGAGGAGGAAAAGATGTAATCTCTCGCCTCTCTCCCTTTTCTC 88036760
88036761 TCTCTCTCCCTGCCACCTCCCAATTTAATCAGCAGTGCCAATATCCAACCTCAATGGCC 88036820
88036821 AATCTCCAACCTCTATACATGGAGTCAAAGCAATAGCTAAGAAATATGTCCTCATGTTA 88036880
88036881 TTACAATAGTATTGATGATGACAAGGAATGGAGCACAAAATAAGTTAGATTCTTGCTCAT 88036940
88036941 TCTCTTTGCTAAACTAC CGTTGGGTGCATTTCCCTCTCACTGAAGTGTGAGGATCCGAA 88037000
88037001 TGCTGAAACAGATCACTAGCCGAAAATGCTTTTAGACACTTAAAGGTTGACTTGACTAGC 88037060
88037061 AAACCTTTAACCTCAGTTTTGTGCTGATCCAAGTCAATGCCTAAATCTAGACCAGCCCTT 88037120
88037121 TGACCCAATCTCTCATGGAGAGTGTCTCTGGTGAAGTCATGGAAATATGGCCAATGGCCT 88037180
88037181 GTTTACAGATTTGTGTGAGTAAAGTGAAGCTTTTGCAGGGGACTGTGGCACTTCTCTTCA 88037240
88037241 TCGTACTCCCACAGATTCCTACAGTAAATTTTCTTGTACTGTGAAGAGCAAGGGCGTC 88037300
88037301 TCACAGAAACTCTCGGGGATGGTATCCTCTTCTCCATTTGCACTACCCCCAACTTTTCA 88037360
88037361 CCCTTCCCTCAAATCAGAAGGAAAAAAGTTCTCTCTGTCATCTAATGTTAATTTTTC 88037420
88037421 TCTCTCCCACCCAGGTTGTATATTTATCCTCCTCGTAATCTTTTCTTTTCTTTCTT 88037480
88037481 TTTTGTAGATAGAGTCTCACTCTGTCAACCAGGCTGGAGTACAGTGGCACAGTCTCAGCT 88037540
88037541 CACTGCAACCTCCAGTTCCTGGGTTCAAGCCATTTCTCTGCTCAGCCTCCAAGCAGCT 88037600
88037601 GGGACTACAGGCAAGCGCCACCACACCCGGCTAATTTTTTATTTTATTTAGTAGAGACAGG 88037660
88037661 TTTTGCATGTTGCCAGGCTGGTCTCAAACCTCCTGACCTCAAGTGATCCACCCACCTCA 88037720
88037721 GCCTCCCAAAGTGTGGGATTACAAGCGTGCACCATTGCGCCCGCCCTCCTCATAACT 88037780
88037781 CTTGTGGAATATTTCCCTCAGTTATCCTCCCTTCCCTTCCAACAACA ACTCCTCTGTC 88037840
88037841 TTCCTTCTGCTTCTATCATTCTTATCTTTCAGTCTCATGTTCCCAATGGCTCCTTC 88037900
88037901 CTTTCTTTGCAAATATGTAATAATGTTTCCACTTGGGGTAATAATGCAACTTCTCTTGAC 88037960
88037961 TGTGATCCTACCCTAACCTGGACTCTTGTCTTCTTTACCTTGAACCTGCTGTATT 88038020
88038021 TATATTTTCATCATGTAATCTTCAAGTATAATGACATTTAGCTTCCACCTGGCTGATGAAC 88038080
88038081 TAAAATTGAATCTTA AAAACAATAGTACTTCTTAATTATCAATTCTAGTGTAGTTTT 88038140
88038141 ACAGTCTCACTTTTCTTTTCTTTTGGAGACAGAGTCTGGTCTCTCACCCAGGCTG 88038200
88038201 GAGTGCAGTGGCGCAGTCTCGGCTCACTGCAACCTCCGTCTCCGGGTTCAAGCAATTCT 88038260
```

Figure 9. PTR fusion sequence of chromosome 9

3. Results

This observation was proven experimentally by the amplification of the NCBI fusion sequence of the chimpanzee (primers sequences in the figure above are marked in red) in humans (Figure 10).

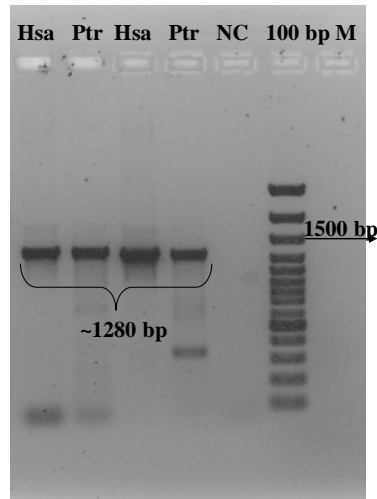


Figure 10. Amplification of the chimpanzee fusion sequence from human chromosome 9
A ~ 1280 bp PCR product was obtained for both species, human (HSA) and chimpanzee
(PTR)

However, when taking the genes from Ensembl releases 39 and 40 in that region, chromosome 9 appeared to be the most “ active “ chromosome. 11 out of 27 (40.8%) from the breakpoint region and 6 out of 20 (30%) genes from the control region were under selection (Figure 11). Of these 11 genes, just one (*ZNF658*) (3.7%) was positively selected whereas the other genes (37%) were negatively selected. Although chromosome 9 seems to be under stronger selection than the other analyzed chromosomes, it does not show a significant difference between BP-FR and CR.

The pericentric inversion of chromosome 17 is a very gene rich region, but the inversion BP doesn't affect the genomic structure of a gene(s). No addition or deletion of any sequence element was detected at the breakpoints or in the neighboring sequences. The BP on HSA 17p is at 7.871.912-7.871.935 Mb and on HSA 17q at 44.975.608-44.975.626 Mb. Eight out of 116 (6,8%) genes were positively selected and 23 genes (20%) were negatively selected.

3. Results

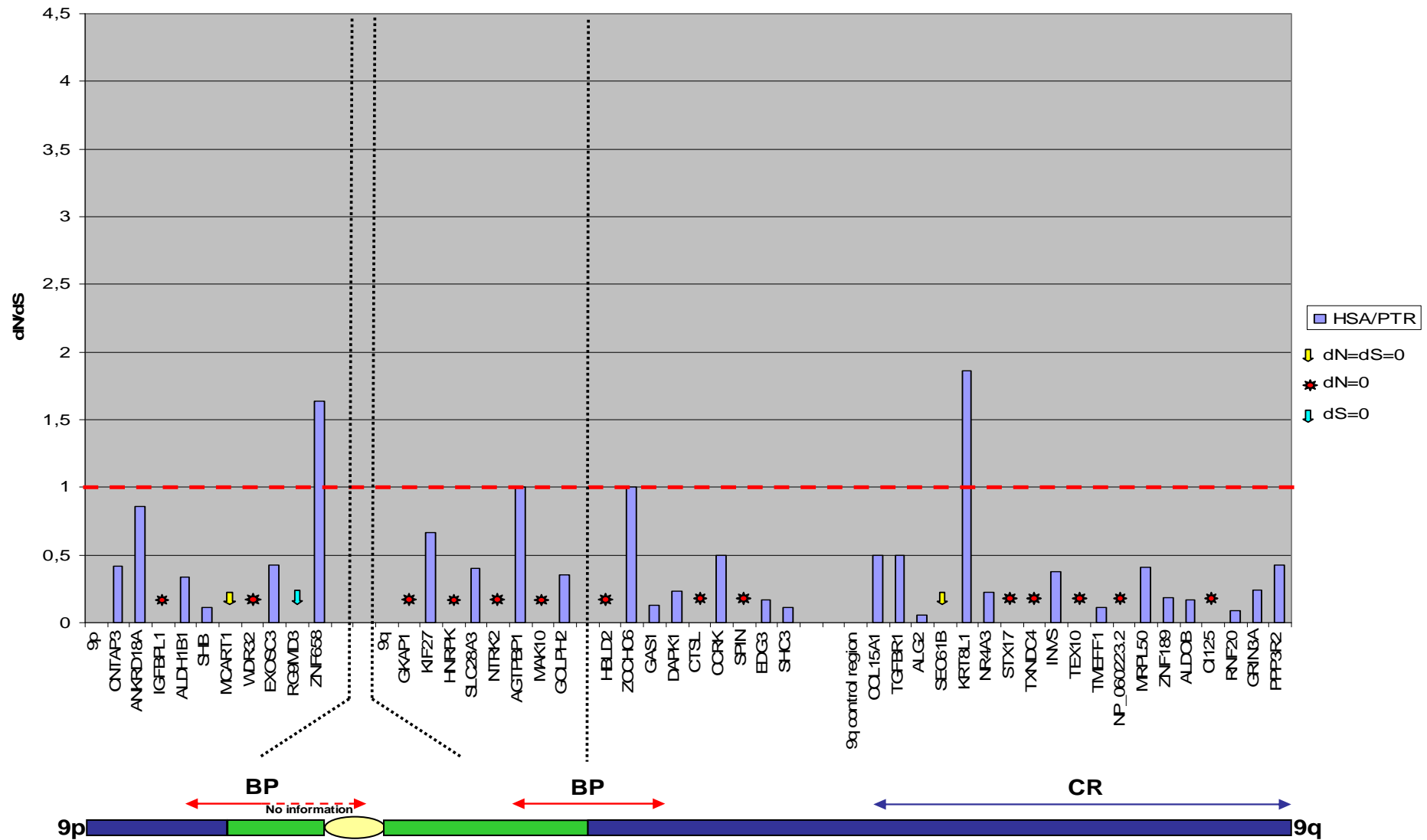


Figure 11. Chromosome 9 pericentric inversion HSA-PTR

3.1.2 Summarized data of the coding sequence comparison between human and chimpanzee

The total number of analyzed genes was 334 in the BP-FRs and 72 in the CR. The rates of positively and negatively selected genes are summarised in Table 3A.

Chromosome	# of genes in BP-FRs	positively selected	%	negatively selected	%	# of genes in CR	positively selected	%	negatively selected	%
1	31	3	9.67%	5	16.13%	24	1	4.17%	5	20.83%
4	39	0	0%	9	23.07%	8	1	12.5%	2	25%
5	30	2	7%	1	3.34%	6	0	0	0	0
9	27	1	3.7%	9	33.34%	20	1	5%	5	25%
12	57	2	3.5%	15	26.31%	14	2	14.28%	2	14.28%
17	116	8	6.80%	23	19.82%	n.d.				
18	34	1	2.94%	6	17.64%	n.d.				

TABLE 3(A) Number of genes analysed on chromosome 1, 4, 5, 9, 12, 17, 18
Genes in the BP-FRs are marked in pink and genes in the CR are marked in blue

In all analyzed regions a random distributions of selected and non-selected genes were observed. Two tendencies were observed when comparing the numbers of positively and negatively selected genes per chromosome (see Table 3 A and B). Chromosome 9 contained the highest percentage of selected genes in both BP-FR and CR, whereas chromosome 5 had the lowest number of selected genes. Interestingly, the BP-FR of chromosome 4 did not contain any gene under positive selection, but 9 genes under negative selection.

3. Results

Chromosome	# of genes in BP-FRs	positively selected	%	negatively selected	%	% non-neutral select	# genes in CR	positively selected	%	negatively selected	%	% non-neutral select
1	31	3	9.67%	5	16.13%	26%	24	1	4.17%	5	20.83%	25%
4	39	0	0%	9	23.07%	23%	8	1	12.5%	2	25%	38%
5	30	2	7%	1	3.34%	10%	6	0	0	0	0	0%
9	27	1	3.7%	9	33.34%	37%	20	1	5%	5	25%	30%
12	57	2	3.5%	15	26.31%	30%	14	2	14.28%	2	14.28%	29%
17	116	8	6.80%	23	19.82%	27%	n.d.					
18	34	1	2.94%	6	17.64%	21%	n.d.					

TABLE 3(B) Percentage of genes under non-neutral selection on chromosome 1, 4, 5, 9, 12, 17, 18 for the analyzed regions

Genes in the BP-FRs are marked in pink and genes in the CR are marked in blue

The majority of non-neutrally selected genes in the BP-FRs (85 out of 334 genes) are under purifying selection (68 genes) while 17 genes are positively selected. (Table 4 A and B).

gene symbol positively selected	chromosome	region	dN/dS	gene description
NUD17	1:144,297,502-144,300,792	inv	5,34	Nucleoside diphosphate-linked moiety X motif 17
ZNF364	1:144,322,393-144,402,018	inv	1,25	Zinc finger protein 364
ACP6	1:145,585,807-145,609,258	3' FR	1,34	Lysophosphatidic acid phosphatase type 6 precursor
none	4			
ANKRD32	5:93,979,808-94,059,082	inv	1,34	Ankyrin repeat domain-containing protein 32
FAM81B	5:94,754,247-94,811,897	inv	4	Protein FAM81B
ZNF658	9:40,750,700-40,826,415	5' FR	1,64	Zinc finger protein 658
SLCO1B3	12:20,859,895-20,960,923	break	2,67	Solute carrier organic anion transporter family member 1B3
Novel Gene			1,63	
SAT2	17:7,470,277-7,471,912	5' FR	4,75	Diamine acetyltransferase 2
C17orf44	17:8,064,710-8,068,086	inv	2,67	uncharacterized protein
LRRC46	17:43,264,099-43,269,689	inv	1,6	Leucine-rich repeat-containing protein 46
COPZ2	17:43,458,535-43,470,138	inv	1,5	Coatomer subunit zeta-2
TTLL6	17:44,194,602-44,226,685	inv	2,75	Tubulin tyrosine ligase-like family, member 6
TAC4	17:45,270,670-45,280,378	3' FR	1,7	Tachykinin 4 isoform delta
HILS1	17:45,603,873-45,604,836	3' FR	2,8	Spermatid-specific linker histone H1-like protein
EME1	17:45,805,589-45,813,817	3' FR	1,1	Crossover junction endonuclease EME1
CLUL1	18: 588,521-640,291	inv	2,67	Clusterin-like protein 1 retinal

Table 4 (A) Positively selected genes in the analyzed breakpoint flanking regions

3. Results

gene symbol negatively selected	chromosome	region	dN/dS	gene description
POLR3C	1:144,303,963-144,322,294	inv	0	DNA-directed RNA polymerase III subunit C
CD160	1:144,407,155-144,426,971	inv	0	CD160 antigen precursor
HIST2H3C	1:148,077,734-148,079,389	3' FR	0	Histone H3.2
HIST2H2AC	1:148,125,149-148,125,585	3' FR	0	Histone H2A type 2-C
SF3B4	1:148,161,833-148,167,146	3' FR	0	Splicing factor 3 subunit4
KCTD8	4:43,870,683-44,145,581	5' FR	0	BTB/POZ domain-containing protein KCTD8
GABRG1	4:45,732,543-45,820,855	inv	0	Gamma-aminobutyric-acid receptor subunit gamma-1 precursor
GABRA2	4:45,946,341-46,086,702	inv	0	Gamma-aminobutyric-acid receptor subunit alpha-2 precursor
GABRA4	4:46,615,674-46,691,181	inv	0	Gamma-aminobutyric-acid receptor subunit alpha-4 precursor
GABRB1	4:46,728,134-47,123,218	inv	0	Gamma-aminobutyric-acid receptor subunit beta-1 precursor
SEC31L1	4:83,958,839-84,040,715	inv	0	SEC31 homolog A isoform 4
LIN54	4:84,065,167-84,151,006	inv	0	lin-54 homolog
COPS4	4:84,175,263-84,215,994	inv	0	COP9 signalosome complex subunit 4
NKX6-1	4:85,633,460-85,638,411	inv	0	Homeobox protein Nkx-6.1
RFESD	5:95,008,239-95,019,542	inv	0	Rieske (Fe-S) domain containing
IGFBPL1	9:38,398,991-38,414,444	5' FR	0	insulin-like growth factor binding protein-like 1
WDR32	9:37,790,864-37,853,322	5' FR	0	WD repeat domain 32
GKAP1	9:85,544,156-85,622,477	inv	0	G kinase anchoring protein 1
HNRPK	9:85,772,818-85,785,339	inv	0	Heterogeneous nuclear ribonucleoprotein K
NTRK2	9:86,473,286-86,828,325	inv	0	BDNF/NT-3 growth factors receptor precursor
MAK10	9:87,745,881-87,827,033	inv	0	corneal wound healing-related protein
HBLD2=ISCA1	9:88,069,281-88,087,310	3' FR	0	Iron-sulfur cluster assembly 1 homolog
CTSL	9:89,530,254-89,536,127	3' FR	0	Cathepsin L precursor
SPIN1	9:90,193,154-90,283,429	3' FR	0	Spindlin
LDHB	12:21,679,543-21,702,043	inv	0	L-lactate dehydrogenase B chain
NP_079006.1	12:18,125,073-18,134,381	5' FR	0	uncharacterized
IRK8=KCNJ8	12:21,809,156-21,819,014	inv	0	ATP-sensitive inward rectifier potassium channel 8
ABCC9	12:21,845,245-21,985,434	inv	0	ATP-binding cassette transporter sub-family C member 9
Q6ZQU9	12:64,535,994-64,562,088	inv	0	uncharacterized
C12orf31	12:64,803,109-64,810,800	inv	0	uncharacterized
CAND1	12:65,949,426-65,994,658	inv	0	Cullin-associated NEDD8-dissociated protein 1
DYRK2	12:66,329,021-66,340,410	inv	0	Dual specificity tyrosine-phosphorylation-regulated kinase 2
IL26	12:66,881,396-66,905,838	3' FR	0	Interleukin 26
RAP1B	12:67,290,919-67,340,641	3' FR	0	Ras-related protein Rap-1b precursor
CPSF6	12:67,919,663-67,951,290	3' FR	0	Cleavage and polyadenylation specificity factor 6
LYSC	12:68,028,401-68,034,280	3' FR	0	Lysozyme C precursor
YEATS4	12:68,039,799-68,070,842	3' FR	0	YEATS domain-containing protein 4
TCPB	12:68,265,475-68,281,624	3' FR	0	T-complex protein 1 subunit beta
RAB3IP	12:68,418,923-68,497,421	3' FR	0	RAB3A-interacting protein

3. Results

gene symbol negatively selected	chromosome	region	dN/dS	gene description
EIF4A1	17:7,416,748-7,423,048	5' FR	0	Eukaryotic initiation factor 4A-I
MPDU1	17:7,427,698-7,432,532	5' FR	0	Mannose-P-dolichol utilization defect 1 protein
ATP1B2	17:7,494,979-7,501,812	5' FR	0	Sodium/potassium-transporting ATPase subunit beta-2
TP53	17:7,512,464-7,531,642	5' FR	0	Cellular tumor antigen p53
EFNB3	17:7,549,245-7,555,421	5' FR	0	Ephrin-B3 precursor
RPL26	17:8,221,559-8,227,293	inv	0	60S ribosomal protein L26
NDEL1	17:8,279,895-8,324,667	inv	0	Nuclear distribution protein nudE-like 1
MYH10	17:8,318,248-8,474,804	inv	0	Myosin heavy chain 10
STX8	17:9,094,514-9,420,000	inv	0	Syntaxin-8
DHRS7C	17:9,615,536-9,635,371	inv	0	Dehydrogenase/reductase (SDR family) member 7C
RCVRN	17:9,741,752-9,749,409	inv	0	Recoverin
GAS7	17:9,754,651-10,042,593	inv	0	Growth-arrest-specific protein 7
IMB1=KPNB1	17:43,082,274-43,116,003	inv	0	Karyopherin (importin) beta 1
TBX21	17:43,165,609-43,178,484	inv	0	T-box transcription factor TBX21
PNPO	17:43,373,936-43,380,653	inv	0	Pyridoxine-5'-phosphate oxidase
HXB5	17:44,023,619-44,026,322	inv	0	Homeobox protein Hox-B5
HXB13	17:44,157,132-44,161,119	inv	0	Homeobox protein Hox-B13
PHB	17:44,836,419-44,847,241	inv	0	Prohibitin
SPOP	17:45,031,247-45,110,524	3' FR	0	Speckle-type POZ protein
SLC35B1	17:45,133,339-45,140,527	3' FR	0	Solute carrier family 35 member B1
MYST2	17:45,221,057-45,261,455	3' FR	0	Histone acetyltransferase MYST2
COL1A1	17:45,616,456-45,633,992	3' FR	0	Collagen alpha-1(I) chain precursor
LRRC59	17:45,813,598-45,957,115	3' FR	0	Leucine-rich repeat-containing protein 59
YES1	18:711,747-802,547	inv	0	Proto-oncogene tyrosine-protein kinase Yes
MLRM	18:3,237,528-3,246,226	inv	0	Myosin regulatory light chain 2, nonsarcomeric
DLGAP1	18:3,488,837-3,870,135	inv	0	Disks large-associated protein 1
C18orf19	18:13,655,741-13,672,108	inv	0	uncharacterized
SNRPD1	18:17,446,235-17,464,342	3' FR	0	Small nuclear ribonucleoprotein Sm D1

Table 4 (B) Genes under purifying selection in the analyzed breakpoint flanking regions

3. Results

A small number of genes (10) in the BP-FRs, that have neither synonymous nor non-synonymous differences between human and chimpanzee, are summarized in the table 5.

gene	chromosome	region	dN/dS	gene description
none	chromosome 1			
PLAC8	4:84,230,239-84,254,935	inv	0/0	Placenta-specific gene 8 protein
MRPS18C	4:84,596,196-84,601,900	inv	0/0	28S ribosomal protein S18c, mitochondrial
none	chromosome 5			
MCART1	9:37,875,368-37,878,596	5' FR	0/0	Mitochondrial carrier triple repeat protein 1
AEBP2	12:19,448,267-19,564,424.	5' FR	0/0	AE binding protein 2
GOLT1B	12:21,546,016-21,562,358	inv	0/0	Vesicle transport protein GOT1B
TNFSF12/13	17:7,392,932-7,405,649	5' FR	0/0	Tumor necrosis factor ligand superfamily
CBX1	17:43,502,414-43,533,806	inv	0/0	Chromobox protein homolog 1
HOXB6	17:44,028,113-44,037,333	inv	0/0	Homeobox protein Hox-B6
GNGT2	17:44,638,596-44,641,742	inv	0/0	Guanine nucleotide-binding protein
DLX3	17:45,422,388-45,427,587	inv	0/0	Homeobox protein DLX-3
none	chromosome 18			

Table 5. Hyperconserved genes dN = 0 and dS = 0

To determine whether the genes under non-neutral selection are associated with human diseases, a search was performed in the online Mendelian Inheritance in Man (NCBI -OMIM) database. The disease associated genes are listed in table 6.

gene	Chr.	dN/dS	gene description	OMIM ID	disease association in human
GABRA2	4	0	Gamma-aminobutyric-acid receptor subunit alpha-2 precursor	137140	alcoholism
GABRA4	4	0	Gamma-aminobutyric-acid receptor subunit alpha-4 precursor	137141	autism
NTRK2	9	0	BDNF/NT-3 growth factors receptor precursor	600456	obesity, hyperphagia, developmental delay
LDHB	12	0	L-lactate dehydrogenase B chain	150100	lactate dehydrogenase B deficiency
ABCC9	12	0	ATP-binding cassette transporter sub-family C member 9	601439	dilated cardiomyopathy
MPDU1	17	0	Mannose-P-dolichol utilization defect 1 protein	604041	glycosylation disorder
TP53	17	0	Cellular tumor antigen p53	191170	cancer
RCVRN	17	0	Recoverin	179618	cancer associated retinopathy, autoantigen
PNPO	17	0	Pyridoxine-5'-phosphate oxidase	603287	encephalopathy
COL1A1	17	0	Collagen alpha-1(I) chain precursor	120150	osteogenesis imperfecta
SNRPD1	18	0	Small nuclear ribonucleoprotein Sm D1	601063	SLE, autoantigen
TNFSF13	17	0/0	Tumor necrosis factor ligand superfamily member 12	604472	rheumatoid arthritis
HOXB6	17	0/0	Homeobox protein Hox-B6	142961	severe developmental abnormalities
DLX3	17	0/0	Homeobox protein DLX-3	600525	AIHHT, TDO

No positiv selected gene was found to be directly disease associated

Table 6. Positively and negatively selected genes with disease association

3. Results

The genes under non-neutral selection (85 genes) were classified according to the tissue of expression in the Novartis Gene Expression Atlas -GNF 1H, MAS5. Five out of 13 positively selected genes (with known expression patterns) showed the maximum expression level in the testis. Genes with maximum expression in the brain didn't exhibit a tendency towards positive selection, in fact they tended to be negatively selected (25%).

To find out whether a certain biological processes predominately occur in the group of the positively and negatively selected genes, the two groups of genes were analyzed in the Panther database. From the positively selected group, 11 genes (57.9%) couldn't be assigned to a specific biological process. The most representative biological processes in the group of genes under purifying selection were: nucleoside, nucleotide and nucleic acid metabolism (13 genes =16,4%), protein metabolism and signal transduction (11 genes =20,8% each) (Figure 12. A and B).

3. Results

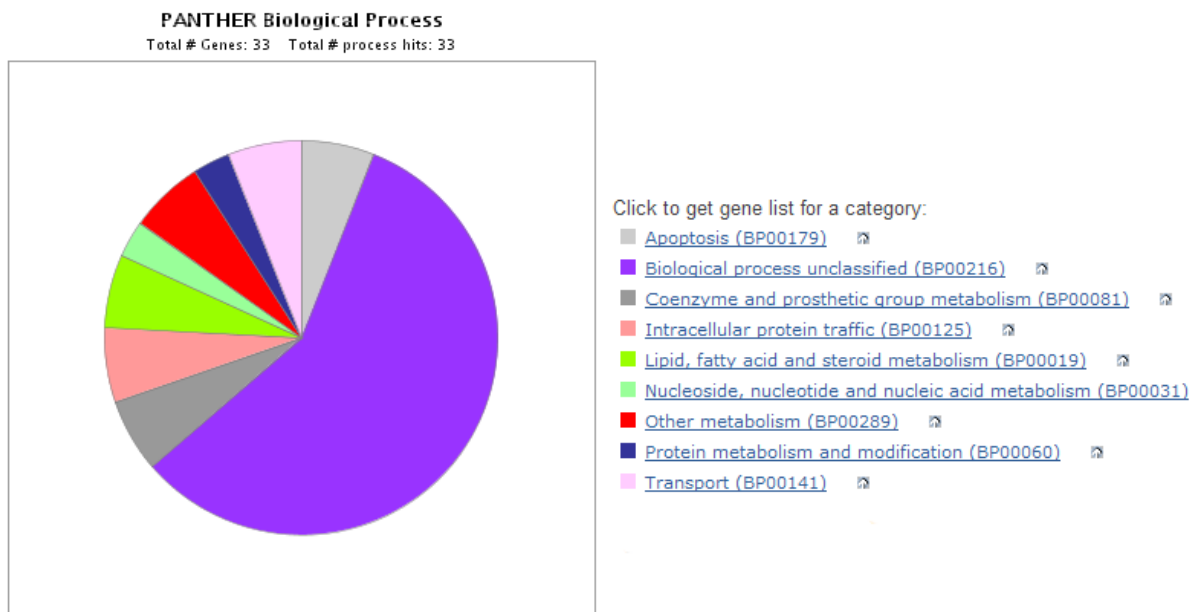


Figure 12. (A) Biological Process classification of the positively selected genes

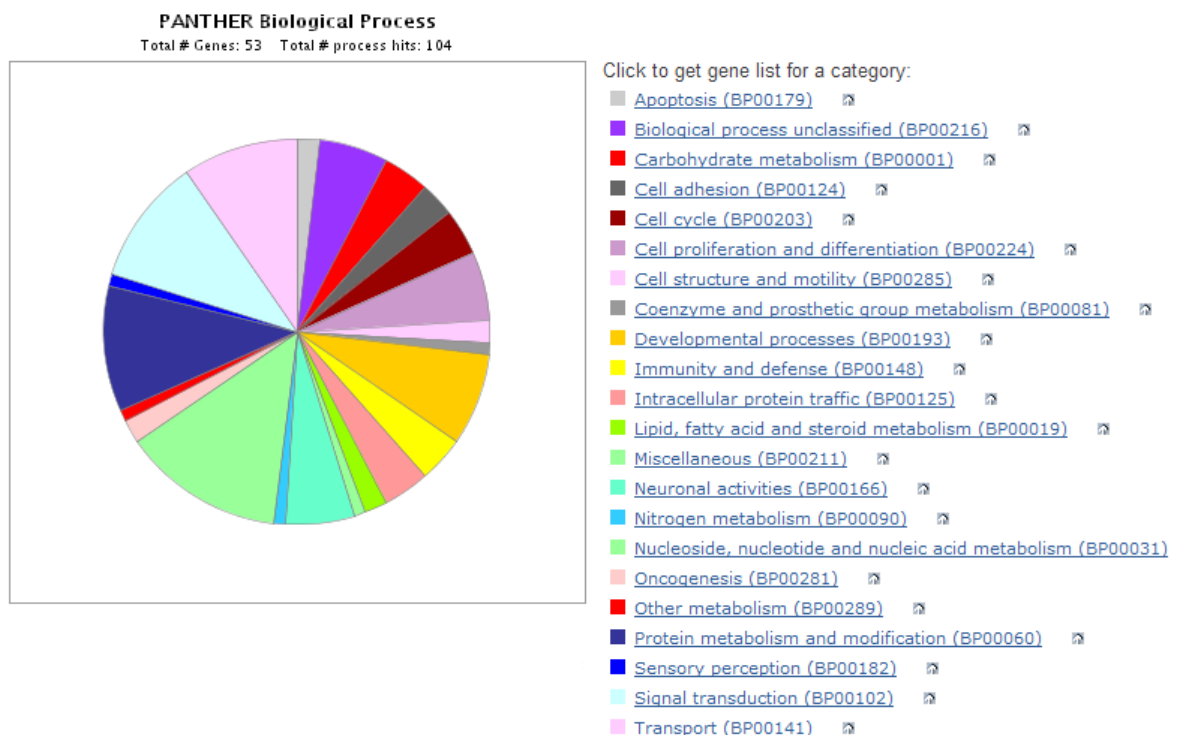


Figure 12. (B) Biological Process classification of genes under purifying selection

3.2 Bioinformatical comparison of gene expression in the evolutionary breakpoint regions of human and chimpanzee

This part of the work was done in collaboration with the Institute of Medical Biostatistics, Epidemiology and Informatics Mainz.

Since the pericentric inversions are associated with changes in gene order, they could possibly have given rise to expression differences through position effects.

Therefore the 334 genes were analyzed *in silico* for differential gene expression between human and chimpanzee, using published data from affimetrix chip U95 and U133v2 microarrays (Gene Expression Omnibus GEO and ArrayExpress EBI databases).

When inter-species studies are performed using a microarray designed for a single species, sequence mismatches between the target template (chimpanzee) and the probes on the array (human sequences) or the presence of multiple copies in the human or chimpanzee genome can affect the hybridisation efficiency. This leads to biased estimates of gene expression data. In order to circumvent this problem a BLAST of all affimetrix chip U95 and U133v2 probes against the human and chimpanzee genome was performed. The BLAST- result is a so called “Maskfile“ containing the uncalled probes - those that were absent or found more than twice in one of the genomes, and those with sequence mismatches. All these probes were then excluded (masked) from the analysis. The unprocessed and the filtered (masked) data of the analyzed genes were compared, using the software-package “R“. Different normalization methods (Quantile-Normalization and Variance Stabilizing Normalization) were used for both the processed and unprocessed data sets.

3. Results

The different Normalization methods had a smaller effect on the result than the filtering process.

Using relatively low stringency conditions, 31 differentially expressed genes between human and chimpanzee were identified (Table 7. A).

Data analysis using higher stringency and Bonferroni adjustment (adjustment to the various experiments), revealed 6 differentially expressed genes with a significance level of ≤ 0.05 and 11 differentially expressed genes with a significance level of ≤ 0.20 (red and yellow lines Table 7.A and B).

3. Results

Symbol	Q (E-TABM-20)	Q (E-AFMX-11)	Bonferoni_ad	Location
ALDH1B1	0,00000	0,00000	0,01003	Chr.9
C17orf68	0,00332	0,00503	0,07370	Chr.17
CCRK	0,00083	0,00419	0,01715	Chr.9
CDH18	0,04355	0,05048	1,00000	Chr.5
CORIN	0,01519	0,01408	0,45033	Chr.4
DHRS7C	0,02184	0,02360	0,93765	Chr.17
EPN3	0,04534	0,05065	1,00000	Chr.17
EXOSC3	0,00782	0,01144	0,24155	Chr.9
FBXL7	0,03545	0,03800	1,00000	Chr.5
GABRB1	0,00000	0,00096	0,01003	Chr.4
GABRB3	0,00181	0,00427	0,85900	Chr.17
IGFBPL1	0,00181	0,00427	0,04596	Chr.9
JMJD3	0,03617	0,04615	1,00000	Chr.17
LIX1	0,00000	0,00000	0,01003	Chr.5
LIX1L	0,00782	0,01090	0,22292	Chr.1
MAK10	0,04355	0,04924	1,00000	Chr.9
MTMR11	0,01611	0,02012	0,16755	Chr.1
MYH10	0,03545	0,03800	1,00000	Chr.17
MYO10	0,00000	0,00096	0,15299	Chr.5
NDEL1	0,00753	0,01090	0,66471	Chr.17
PCSK1	0,01611	0,02012	1,00000	Chr.5
PDE4DIP	0,00000	0,00000	0,17484	Chr.1
PNPO	0,01154	0,01408	0,36440	Chr.17
PTPN13	0,01611	0,02012	0,65177	Chr.17
RHOBTB3	0,00000	0,00000	0,23862	Chr.5
ROCK1	0,02590	0,03099	1,00000	Chr.18
SAT2	0,03368	0,03800	1,00000	Chr.17
TOB1	0,01363	0,01724	1,00000	Chr.17
UTP18	0,04355	0,05065	1,00000	Chr.17
ZBTB4	0,03545	0,03800	1,00000	Chr.17
ZNF519	0,00181	0,00427	0,04596	Chr.18

Table 7. (A) Data mining of gene expression profiles identifies differentially expressed genes between human and chimpanzee

Symbol	Q (E-TABM-20)	Q (E-AFMX-11)	Bonferoni adjusted	Location	gene name
ALDH1B1	0,00000	0,00000	0,01003	Chr. 9	aldehyde dehydrogenase 1 family, member B1
C17orf68	0,00332	0,00503	0,07370	Chr. 17	hypothetical protein FLJ22170
CCRK	0,00083	0,00419	0,01715	Chr. 9	cell cycle related kinase; cyclin-dependent protein kinase H
GABRB1	0,00000	0,00096	0,01003	Chr. 4	GABA receptor beta1
IGFBPL1	0,00181	0,00427	0,04596	Chr. 9	insulin-like growth factor binding protein-like 1
LIX1	0,00000	0,00000	0,01003	Chr. 5	Lix1 homolog (mouse); limb expression 1
LIX1L	0,00782	0,01090	0,22292	Chr. 1	Lix1 homolog (mouse) like
MTMR11	0,01611	0,02012	0,16755	Chr. 1	Myotubularin related protein 11
MYO10	0,00000	0,00096	0,15299	Chr. 5	Myosin-10
PDE4DIP	0,00000	0,00000	0,17484	Chr. 1	phosphodiesterase4Dinteracting protein (myomegalin)
ZNF519	0,00181	0,00427	0,04596	Chr. 18	zinc finger protein 519

Table 7. (B) Data mining of gene expression profiles with the Bonferroni adjustment

3.3 DNA-Methylation analysis in human and non-human primates

Little is known about the evolutionary conservation of DNA methylation patterns and the evolutionary impact of epigenetic changes between closely related species. The genetic differences between humans and chimpanzees are rather small (Chimpanzee Sequencing and Analysis Consortium, 2005). Therefore, the striking species differences, i.e. in cognitive abilities, must be due to changes in gene regulation rather than structural changes in the gene products. Regulation of gene expression may be achieved by a large number of transcriptional mechanisms including transcriptional initiation, chromatin condensation, and DNA methylation. Methylation of critical CpG dinucleotides in cis-regulatory regions, in particular promoters, are generally thought to act as epigenetic signals that regulate the temporally, spatially, and parent-specifically appropriate gene expression patterns.

Using modified bisulfite sequencing and pyrosequencing techniques, promoter methylation studies were carried out, focusing on differences in DNA methylation patterns between species and secondly on intra-species variation.

Frontal cortices of eleven humans (*Homo sapiens*, HSA), three chimpanzees (*Pan troglodytes*, PTR), three baboons (*Papio hamadryas*, PHA) and one rhesus monkey (*Macaca mulata*, MMU) were available for comparative methylation analysis of candidate genes.

3.3.1 Comparison of CGI promoter methylation in human and chimpanzee cortex

To get a general view on how DNA methylation patterns change during primate evolution, classic bisulfite sequencing was used to compare the methylation pattern of putative promoter CpG islands of six candidate genes in human (*Homo sapiens*, HSA) and chimpanzee (*Pan troglodytes*, PTR) cortex. The cortex was chosen, because in a previous microarray-based study the DNA methylation differences between human and chimpanzee were more pronounced in the brain than in other tissues (Enard et al. 2004).

One group of candidate genes selected because they showed differential expression patterns between human and chimpanzee in an silico analyses (see chapter 3.2) or found as differentially expressed genes in other microarrays experiments. Another important criteria for choosing these genes was of their location in close proximity (<2 Mb) to the evolutionary inversion breakpoints. (Kehrer-Sawatzki and Cooper 2008).

Cell cycle related kinase (*CCRK*), aldehyde dehydrogenase 1 family member B1 (*ALDH1B1*), insulin-like growth factor binding protein-like1 (*IGFBPL1*), and zing finger protein 519 (*ZNF519*) , Src homology 2 domain containing transforming protein 3 (*SHC3*) and neurotrophic tyrosine kinase receptor type 2 (*NTRK2*) were the selected genes. Each analysis started with the search for orthologous transcripts of the respective genes in chimpanzee and other primates. CpG Island Searcher and EBI Tools CpG Islands were used to identify CpG islands in putative cis-regulatory regions of genes of interest 500 bp downstream to 10 kb upstream of the transcription start site (TSS). When a gene contained two or more CpG rich segments, usually the CpG island nearest to the transcription start site was analyzed.

3. Results

Compared was the methylation status of 205 CpGs from 7 CpG islands (Table 8. A). The CpG ratio (ObservedCpG/ExpectedCpG) varied from 0.64 to 1.03 (Table 8. B), representing intermediate (*ALDH1B1*, *CCRK*, *SHC3* and *ZNF519*) and high CpG (*IGFBPL1* and *NTRK2*) promoters (Weber et al., 2007). The transcription start site (TSS) of *ALDH1B1*, *IGFBPL1*, *NTRK2* and *SHC3* lie in the analyzed CGIs, and those of *CCRK* and *ZNF519* in close proximity (<200bp).

Genes	Length of analyzed CpG island	Number of CpGs	Chromosomal position	Position of the TSS	Distance to TSS
<i>ALDH1B1</i>	330 bp	32	Chr. 9, 38,382,506-38,382,836 bp	38,382,660 bp	Contains the TSS
<i>CCRK</i>	600 bp	33	Chr. 9, 89,780,105-89,779,503 bp	89,779,444 bp	59 bp
<i>IGFBPL1</i>	380 bp	48	Chr. 9, 38,414,728-38,414,350 bp	38,414,444 bp	Contains the TSS
<i>NTRK2</i>	370 bp	36	Chr. 9, 86,472,937-86,473,307 bp	86,473,285 bp	Contains the TSS
<i>SHC3 CGI1</i>	380 bp	9	Chr. 9, 90,983,905-90,983,522 bp	90,983,502 bp	Contains the TSS
<i>SHC3 CGI2</i>	260 bp	21	Chr. 9 90,983,470-90,983,203 bp	90,983,502 bp	32 bp
<i>ZNF519</i>	440 bp	26	Chr. 18 14,122,269-14,121,809 bp	14,122,430 bp	161 bp

Table 8. (A) Candidate genes and their analyzed CpG islands

Gene	Length of CpG island	%GC	CpG ratio	Length of CpG island analyzed	%GC	CpG ratio
<i>ALDH1B1</i>	656 bp	59,5	0,66	330 bp	72,7	0,75
<i>CCRK</i>	895 bp	60	0,73	600 bp	56,3	0,69
<i>IGFBPL1</i>	528 bp	56,2	1,03	380 bp	72,2	0,97
<i>MGMT</i>	732 bp	73,8	0,88	460 bp	76,4	1
<i>NTRK2</i>	595 bp	63,2	1,01	370 bp	65,4	0,99
<i>SHC3 CGI 1</i>	600 bp	55	0,65	380 bp	41,2	0,64
<i>SHC3 CGI 2</i>	780 bp	72,5	0,88	260 bp	69,5	0,64
<i>ZNF 519</i>	535 bp	56,1	0,65	440 bp	61,2	0,64

Table 8. (B) Candidate genes and their %GC and CpG ratio

3. Results

Alignment of the orthologous human and chimpanzee CGI sequence, with the assistance of BioEdit revealed a human-chimpanzee divergence of 0.55% to 1.3%, which corresponds to the average sequence difference between the human and chimpanzee genomes (Chimpanzee Sequencing and Analysis Consortium, 2005). According to the Transcriptional Regulatory Element Database (<http://rulai.cshl.edu>), all known and predicted transcription factor binding sites in the analyzed regions were very similar between human and chimpanzee.

When the methylation profiles between one human (HSA1) and one chimpanzee (PTR1) cortex were compared, 5 of the 7 analyzed CpG islands were completely or almost completely demethylated. Only *CCRK* and *SHC3* CGI-1 were partially methylated.

Two of 32 analyzed CpG sites in *ALDH1B1*, 9 of 33 CpGs in *CCRK*, 5 of 48 in *IGFBPL1*, 1 of 36 in *NTRK2*, 3 of 30 in *SHC3*, and none of 26 in *ZNF519* were differentially methylated in human and chimpanzee. With the notable exception of a few (12 of 205, 6%) individual CpG sites, the overall CGI methylation patterns of 5 genes, *ALDH1B1*, *IGFBPL1*, *NTRK2*, *SHC3*, and *ZNF519*, appeared to be well conserved between species (Figure 13.).

3. Results

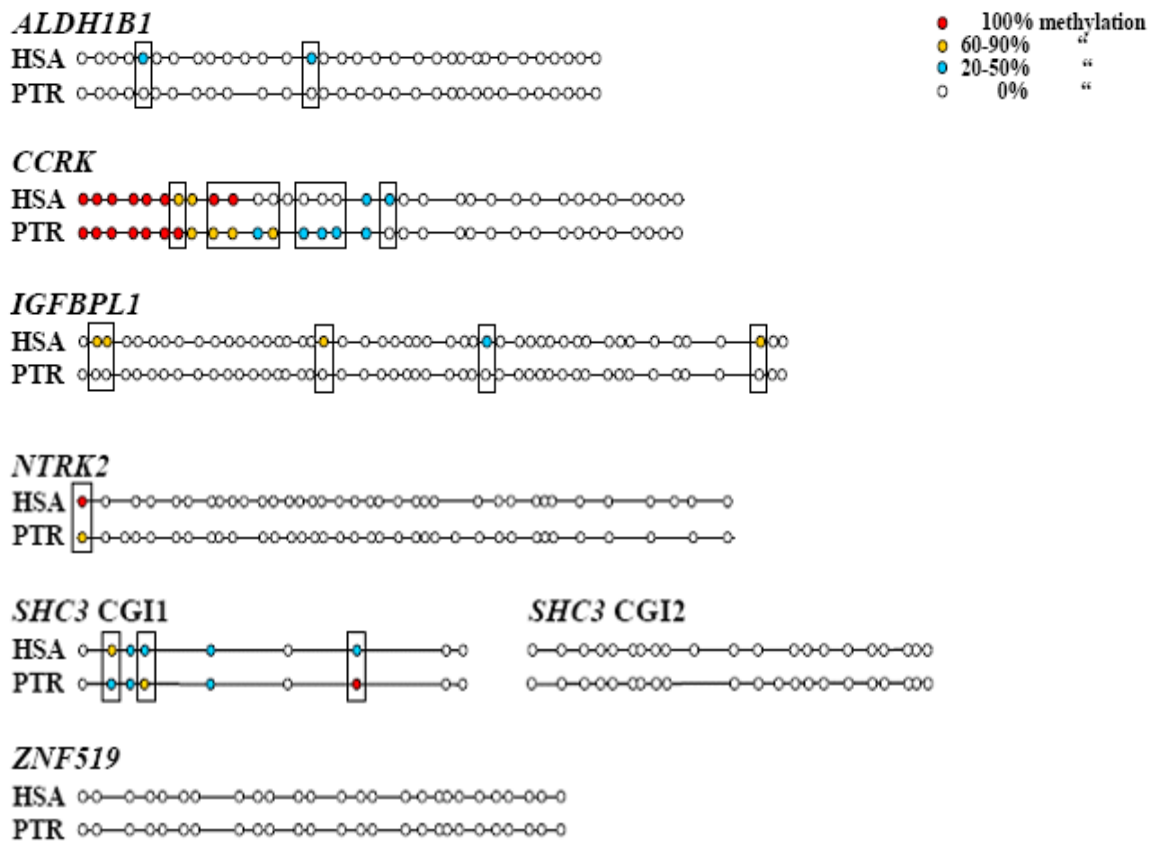


Figure 13. Methylation patterns of orthologous CpG island promoters in human and chimpanzee cortex

5-7 different alleles were sequenced for each of the human (HSA1) and chimpanzee (PTR1) sample. Each circle represents a CpG site and their spacing reflects the CpG density of the region. White circles represent CpG sites that were completely unmethylated and red circles CpGs that were completely methylated in all analyzed plasmids. Blue circle indicate CpGs with 20-50% methylation and yellow CpGs with 60-85% methylation. Differentially methylated CpG sites are framed.

3. Results

Only one gene, *CCRK*, appeared to exhibit a larger cluster of species-specifically methylated CpG sites in its promoter region and, therefore, was analyzed in more detail.

CCRK is a member of the cyclin-dependent kinase (CDK) family that are important for cell cycle control and transcriptional regulation. It appears to be indispensable for cell growth and may also act as a negative regulator of apoptosis (Liu et al. 2004, MacKaigan et al. 2005). Transcriptional upregulation of *CCRK* has been implicated in human glioblastoma tumorigenesis, whereas knockdown of *CCRK* expression by siRNA significantly suppressed cell proliferation (Ng et al., 2007).

The analyzed human *CCRK* CGI contains 33 CpGs. Because of the human-chimpanzee sequence divergence CpG numbers 13 and 31 are not present in the chimpanzee, whereas CpG 23A (between human CpGs 23 and 24) is only found in the chimpanzee sequence. The first half of the island (CpGs 1-16) corresponds to an Alu-Sg1 repeat. The first five CpG sites in the Alu repeat were hypermethylated in both human and chimpanzee, followed by a number of differentially methylated CpGs. The second half of the *CCRK* CpG island that means all CpG sites after the Alu repeat (CpGs 19-33) were completely demethylated in both species.

3.3.2 Variation of *CCRK* CGI promoter methylation among humans and non-human primates

DNA methylation screening of the six candidate genes, revealed interesting methylation differences between human and chimpanzee in one gene, *CCRK* (cell cycle-related kinase).

One important question is how dose this DNA methylation pattern is maintained/changed in other primate species diverged earlier from the human lineage. Therefore two Old World monkeys the rhesus macaque (*Macaca mulata*, *MMU*) and baboon (*Papio hamadryas*, *PHA*) were included in the methylation analysis. Because there were no database informations on *CCRK* promoter sequences in baboon the genomic DNA of the baboon CGI region was sequenced.

Although the *CCRK* CGI promoter sequence including the Alu-Sg1 repeat were well conserved in all analyzed primate species, it was necessary to introduce in addition to the chimpanzee CpG position 23A (between human CpGs 23 and 24) five new CpG positions into the CGI consensus sequence for interspecific comparisons. CpGs 9A and 16A are specific for the baboon, CpGs 12A and 22A are specific for the rhesus monkey, whereas 10A is found in baboon and rhesus monkey but not in human and chimpanzee (Figure 14.).

3. Results

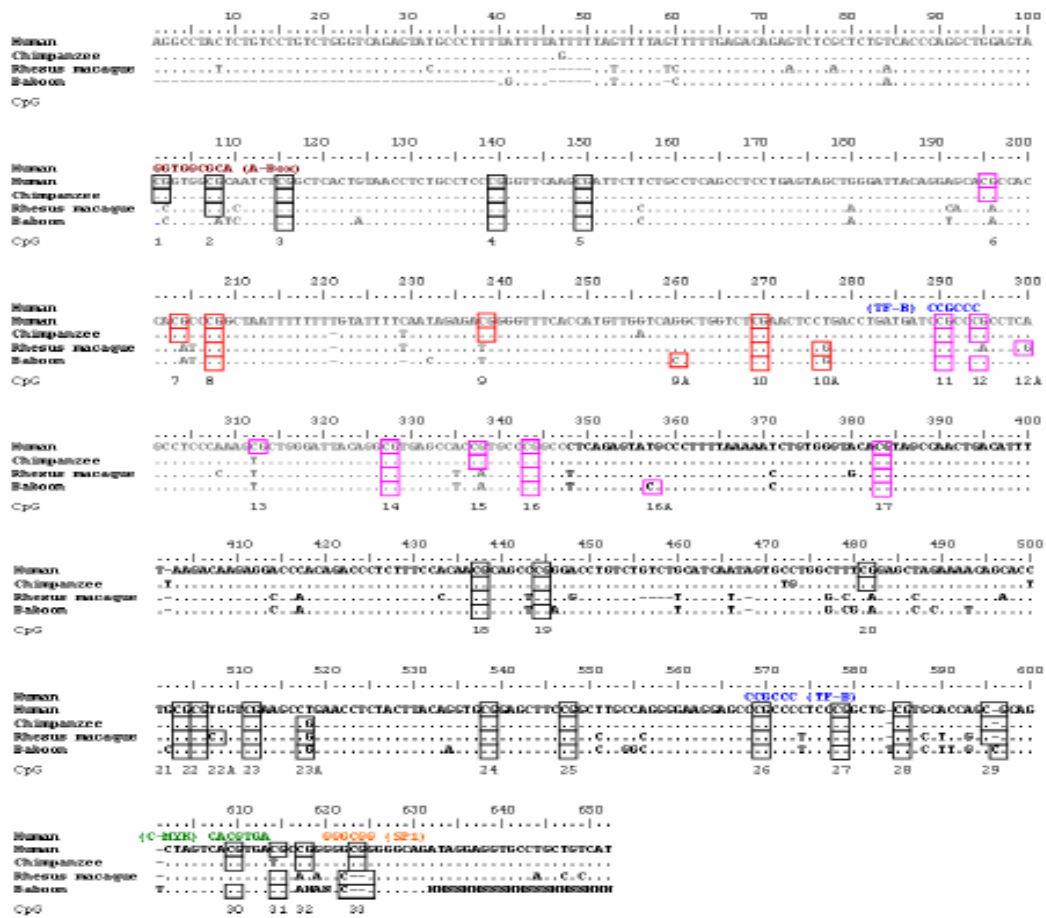


Figure 14. Nucleotide sequence alignment of the *CCRK* CGI promoter for human, chimpanzee, rhesus macaque and baboon.

Important sequence motifs, i.e. an Alu A-box as well as TF-B, C-MYK and SP1 binding sites are marked in different colors. In the non-human primate sequences only positions that differ from the human sequence are indicated. Deletions are indicated by gaps (“-”). Nucleotide positions 1-348 highlighted in gray represent the Alu-Sg1 repeat. The positions of the human CpG sites 1-33 as well as the primate-specific CpGs 9A, 10A, 12A, 16A, 22A, and 23A are indicated below the sequence. CpGs 6-10A which show high intra- and interspecific methylation differences are framed in red; CpGs 11-17 which display less variation are framed in pink. CpGs which are framed in black are either almost completely methylated (CpGs 1-5) or completely demethylated (CpGs 18-33) in all four analyzed species.

3. Results

Methylation analysis showed that *CCRK* gene is endowed with a CGI that exhibits considerable interspecific and intraspecific variation in the degree of methylation. Three distinct regions with a specific methylation pattern can be distinguished in the analyzed CGI of *CCRK* gene (Figure15. A, B, C, D).

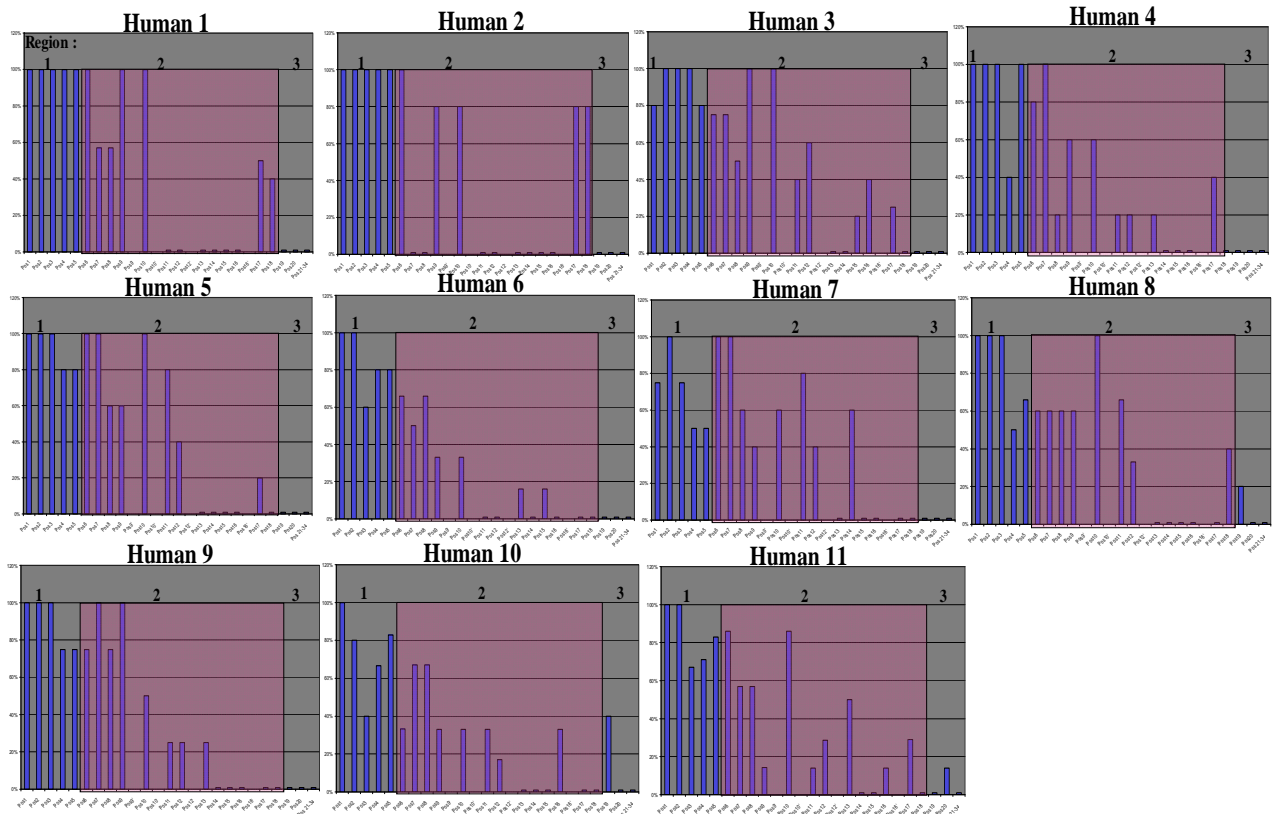


Figure 15. (A) Methylation pattern of the CGI promoter of *CCRK* in the 11 humans

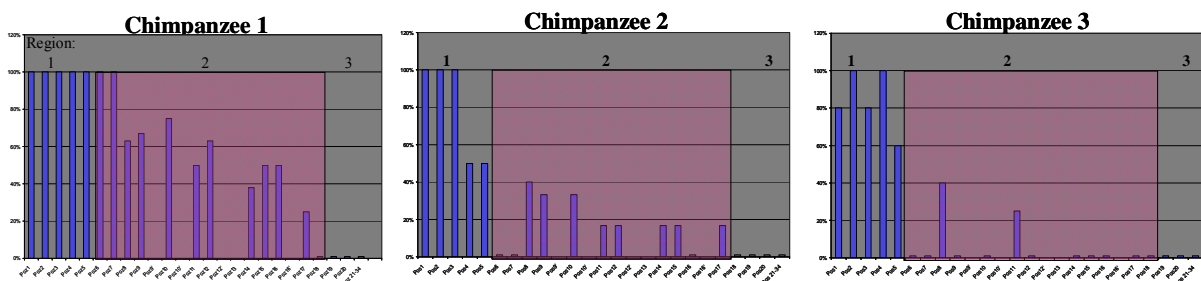


Figure 15. (B) Methylation pattern of the CGI promoter of *CCRK* in 3 chimpanzees

3. Results

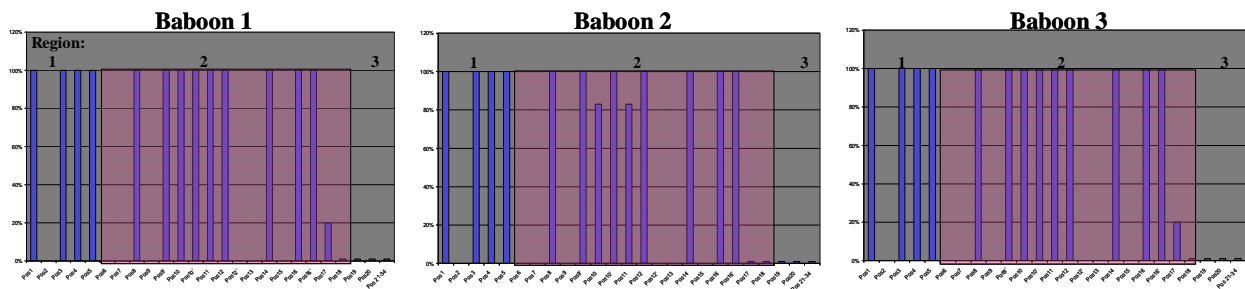


Figure 15. (C) Methylation pattern of the CGI promoter of *CCRK* in 3 baboons

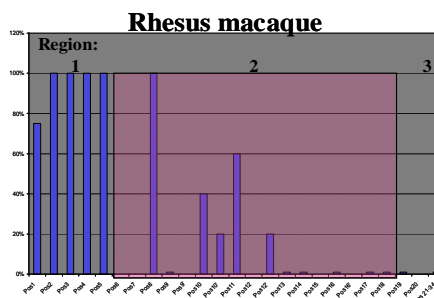


Figure 15. (D) Methylation pattern of the CGI promoter of *CCRK* in 1 rhesus macaque

Figure 15. A, B, C, D Methylation patterns of *CCRK* CpG island promoter in 11 human, 3 chimpanzee, 3 baboon and 1 rhesus macaque cortices. Each bar represents the DNA methylation percentage of the corresponding CpG position. To distinguish between the CpG positions that are missing in one of the analyzed species and those that are completely unmethylated, a hypothetical value of 1% methylation was given to “unmethylated” CpGs. Because the last part of the CGI (positions 21-33) was completely unmethylated in all species, only one bar is shown for CpGs 21-33. Region 1 is hypermethylated, region 2 differentially methylated and region 3 demethylated within and between species.

3. Results

The first region (1) containing CpG positions 1-5 was highly methylated in all human and primate samples. The same conservation pattern was detected also for the last region (3), position 19-33, which was found to be unmethylated in all individuals and all species.

CpG positions 6-18 which represents the second region (2) in the Alu-Sg1 repeat displays a significant inter-species and also an intra-species variations (in humans and chimpanzees) (Figure 16.).

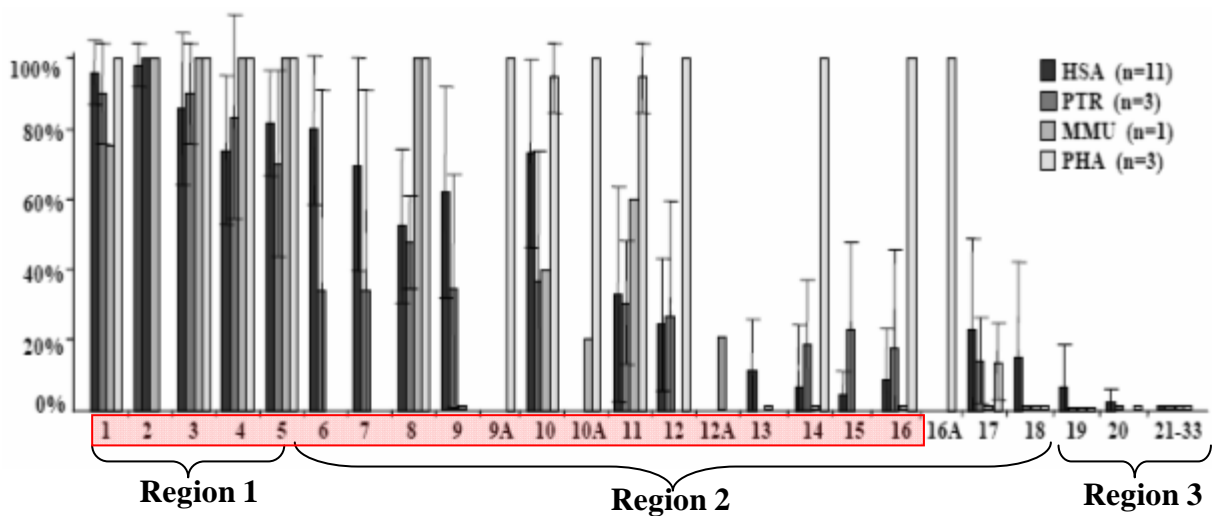


Figure 16. Methylation percentages (mean \pm standard deviation) in human cortex (n=11), chimpanzee (n=3), rhesus monkey (n=1), and baboon (n=3).

Marked are the three regions of *CCRK* CGI and with the red bar the Alu-Sg1 repeat.

Some CpG sites (6, 7, 8, 9, 10) in the second region show a higher average methylation in humans than in chimpanzees whereas CpGs (14, 15, 16) are more methylated in chimpanzees (Figure 17).

3. Results

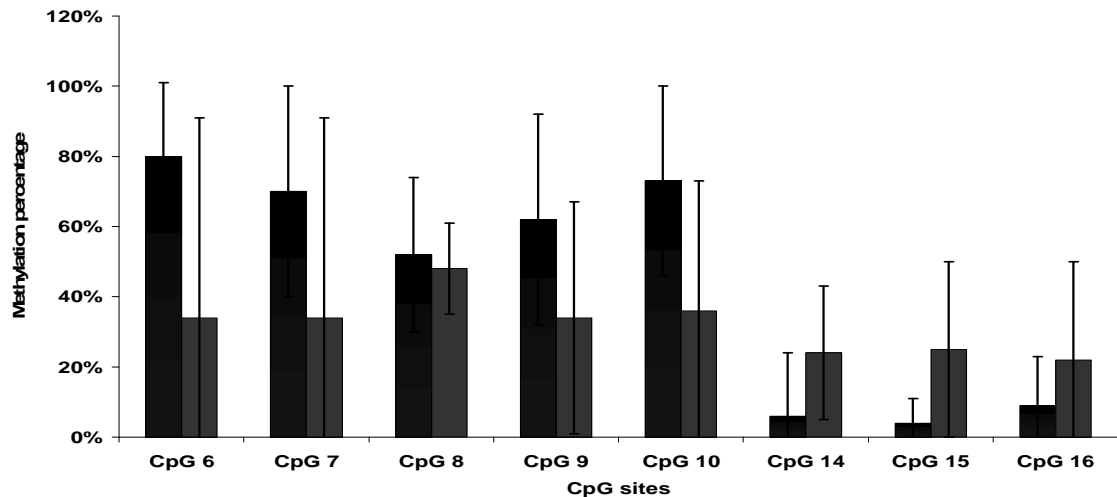


Figure 17. Example of CpGs that are more or less methylated in human compared to chimpanzee. Black bars represent the average methylation in HSA (n=11) while gray the average methylation in PTR (n=3).

Despite inter-species differences, considerable intraspecific variation of the *CCRK* CGI methylation patterns was also observed in the adult cortex samples of the 11 unrelated human individuals and 1 fetal cortex. CpG sites 1-5 were hypermethylated in all or essentially all analyzed adult cortices and displayed much less variation than the following CpG sites (6-20) in the Alu-Sg1 repeat (Figure 18. B). CpG sites 21-33 were completely demethylated in all analyzed human cortex samples. The entire Alu repeat (CpG1-16) appeared to be also somewhat less methylated in fetal cortex (Figure 18. A, white bars) than in the average adult human cortex (Figure 18. A, gray bars).

3. Results

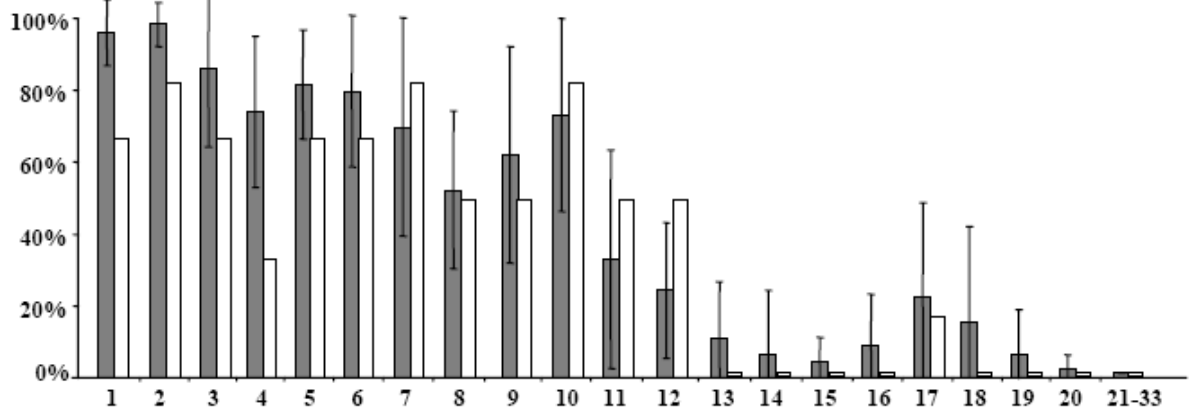


Figure 18.(A) Variation of *CCRK* CGI promoter methylation among humans.

Figure 18. (A) dark gray bars indicate the methylation percentage (mean \pm standard deviation) at each of the 33 analyzed CpG sites in 11 independent human cortex samples. The white bars show relative hypomethylation at most CpG sites in a fetal brain sample.

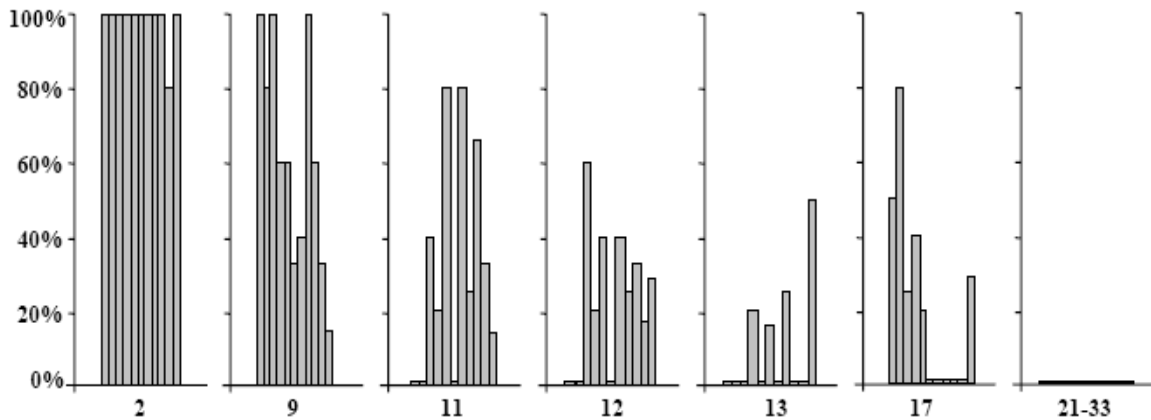


Figure 18. (B) Variation of *CCRK* CGI promoter methylation among humans

Figure 18. (B) present the methylation percentages at selected CpG sites in the 11 analyzed adult cortex samples. Site 2 is completely methylated, whereas sites 21-33 are completely demethylated in all individuals. CpGs 9, 11, 12, 13 and 17 are representative examples of the intraspecific variation of the DNA methylation.

3. Results

The three analyzed chimpanzees also displayed significant high intraspecific methylation differences. Chimpanzee PTR1, who died of an immune-mediated haemolytic anemia, showed higher CpG methylation levels in the Alu-Sg1 repeat than the two other analyzed chimpanzee individuals, who had accidents (Figure 20.). However, it appears unlikely that different causes of death have dramatic effects on methylation patterns. All of the chimpanzee cortex samples were obtained within 12 hours post-mortem.

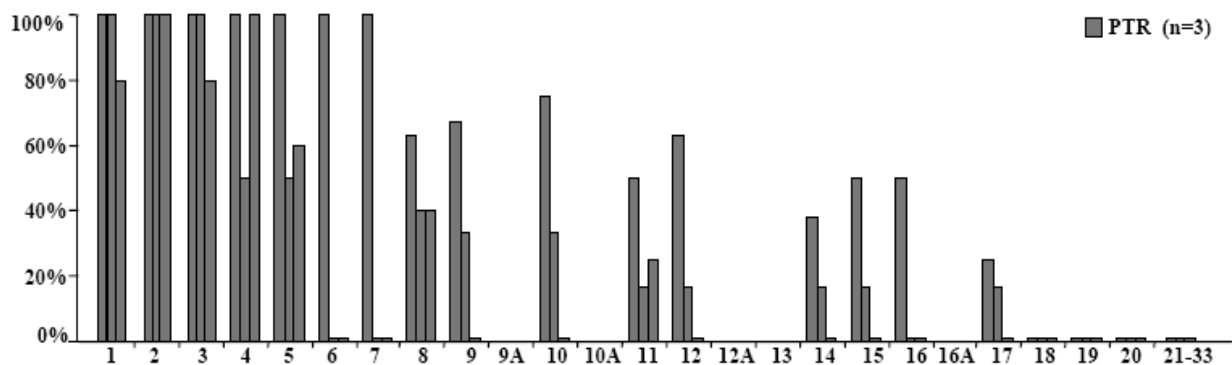


Figure 19. Variation concerning the methylation percentages of individual CpG sites in the adult cortices of three unrelated chimpanzees (PTR).

In contrast to humans and chimpanzees, the methylation patterns of the three analyzed baboon cortices were very similar among each other and clearly different from human and chimpanzee. CpG positions 1-16A were completely methylated, whereas CpGs 18-33 were completely demethylated in all three samples (Figure 21.). This implies that the Alu-Sg1 repeat is completely methylated in baboon and that there is a sharp boundary between the hypermethylated first half and the hypomethylated second half of the CpG island.

3. Results

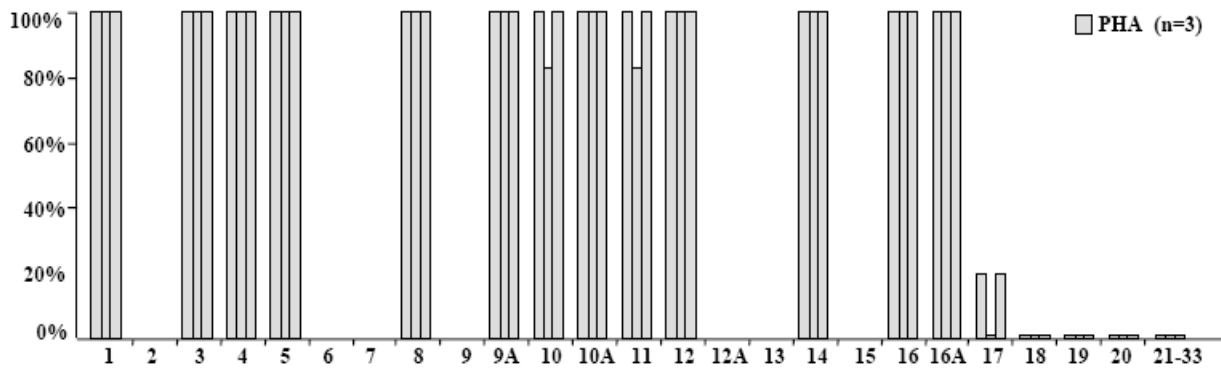


Figure 20. Methylation percentages of individual CpG sites in the adult cortices of three unrelated baboons (PHA).

It is difficult to draw conclusions from only one analyzed cortex, but the CGI methylation pattern of rhesus macaque was more similar to the pattern seen in humans and chimpanzees than to the pattern in baboons. CpGs 1-5 were hypermethylated, CpGs 8-12A showed intermediate methylation percentages and CpGs 13-33 were completely demethylated. (Fig 16. D and 17.).

This study provides a comprehensive analysis of important intra- and interspecific variability of DNA methylation of an Alu insertion in a regulatory region of *CCRK* gene. The most dramatic species differences were observed in a region, comprising CpGs 6- 10 in humans and chimpanzees, CpGs 8-10A in rhesus macaque, and CpGs 8- 10A in baboons.

3.4 Expression of *CCRK* gene in the frontal cortex of human and non-human primates

Transcriptional silencing by promoter methylation is one important mechanisms for tumor suppressor gene silencing. However, much less is known about the correlation between promoter methylation and the transcriptional silencing of other genes. To find out whether the intraspecies differences in *CCRK* promoter methylation affect transcription, the relative *CCRK* mRNA levels in human and non-human primates were compared by quantitative realtime RT-PCR analyses of total RNAs from four human (HSA7, 8, 10 and 11), three chimpanzee, one rhesus macaque, and three baboon cortex samples.

The relative expression in the humans and the baboons was approximately two times higher than that in chimpanzees and the rhesus macaque (Figure 22. B). Humans and chimpanzees who showed more intraspecific variation in *CCRK* promoter methylation than baboons also exhibited more variation in their gene expression levels. It is noteworthy that high *CCRK* expression in humans and baboon is associated with hypermethylated CpGs, in the differentially methylated “region 2” whereas low gene expression in the chimpanzees and the rhesus macaque is associated with hypomethylation (Figure 21. A and B).

When looking to each individual separately, there is no evidence for a correlation between the percentage of the species-specific methylated CpGs and the level of *CCRK* expression within the human or the baboon species. This may be largely due to the small number of individuals analyzed per species.

3. Results

In chimpanzees, the one individual (PTR1 in Figure 21.C) which was intra-species hypermethylated had significantly lower *CCRK* mRNA levels than the two other analysed individuals.

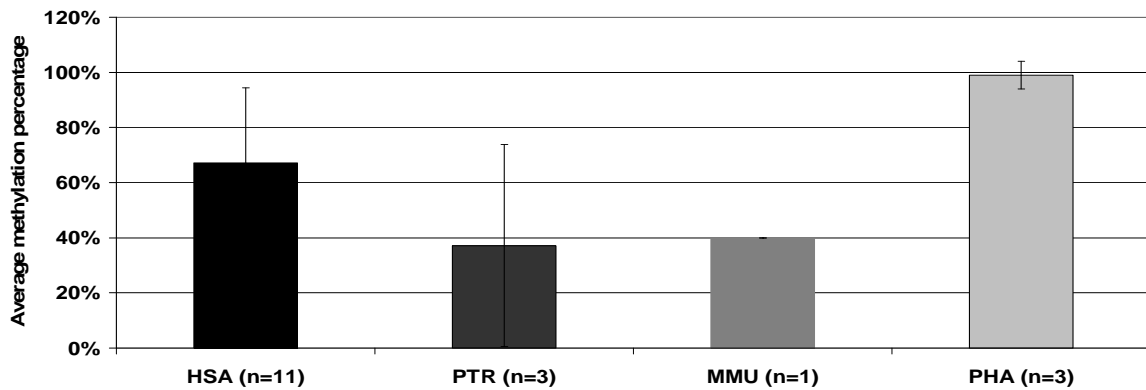


Figure 21. (A) Average methylation percentage (and standard deviation) of the differentially methylated CpG sites (6-10A) in human and non-human primates cortices.

Numbers in parentheses indicate the number of individuals analysed.

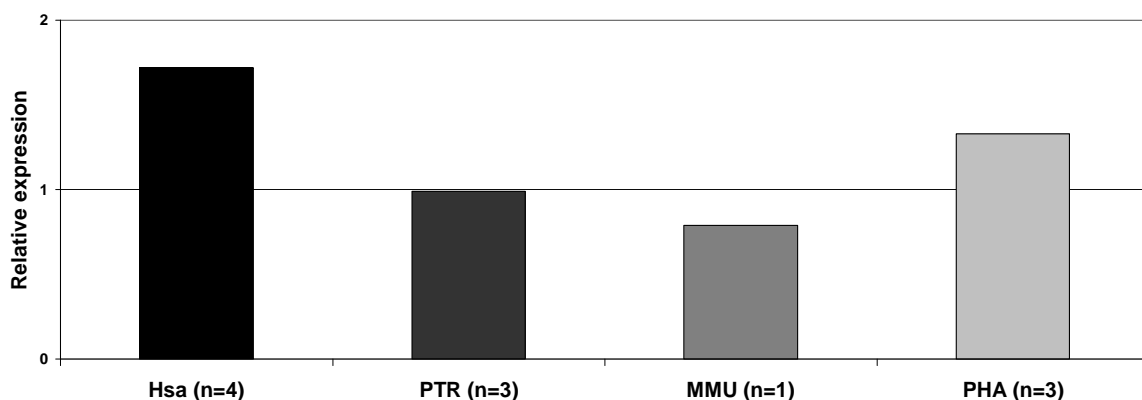


Figure 21. (B) Expression of *CCRK* gene in human and non-human primates cortices.

Relative expression of the *CCRK* gene in human, chimpanzee, rhesus and baboon cortex.

Numbers in parentheses indicate the number of individuals analysed.

3. Results

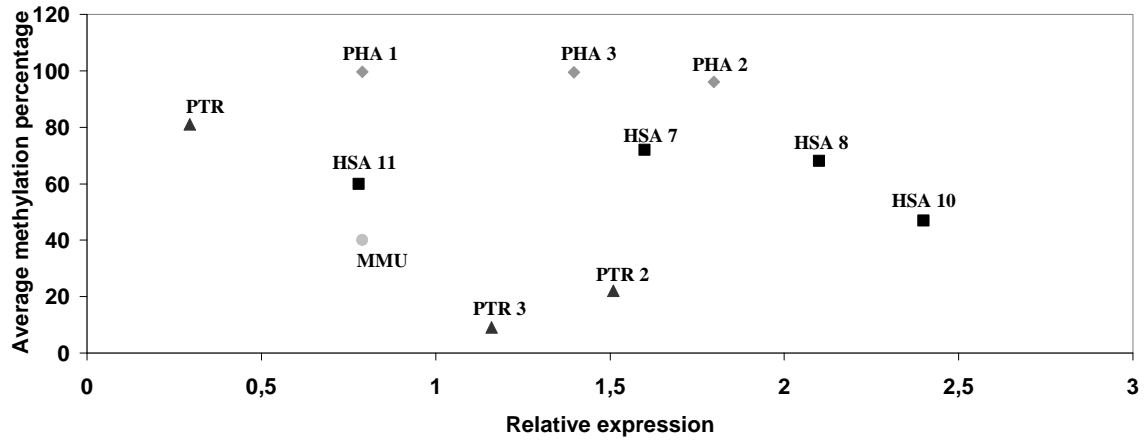


Figure 21. (C) Correlation of the average methylation percentage of the species-specifically methylated CpG sites and the relative mRNA expression of *CCRK* .

The individual humans are marked by squares, chimpanzees by triangles, rhesus macaque by circle, and baboons by diamonds.

3.5 Analysis of DNA methylation patterns using bisulfite Pyrosequencing

In addition to classic bisulfite sequencing (see chapter 3.3.1), the Pyrosequencing technique was used for quantitative DNA methylation analysis of specific CpG sites of six candidate genes in human and chimpanzee cortex (Table 9.). The analyzed genes are associated with human disease, imprinting and/or display a methylation-dependent regulation.

Genes	Number of CpG sites	Localisation
<i>MGMT</i>	4	Chr. 10
<i>GJB2</i>	8	Chr. 14
<i>HELT</i>	5	Chr. 4
<i>MEG3</i>	3	Chr. 14
<i>NESP55</i>	3	Chr. 20
<i>NESPAS</i>	3	Chr. 20

Table 9. Candidate genes and the number of CpG sites analyzed by Pyrosequencing.

Pyrosequencing is extremely useful for methylation analysis of short DNA sequences (30-50 bp) containing functionally important CpG sites. The analysed sequences of the six candidate genes contained either transcription-factor binding sites, (that are important for gene expression) or represent stable indicators of the global CGI methylation level. The genes *MGMT* (O⁶-Methylguanine-DNA Methyltransferase), *HELT* (*HES/HEY*-like transcription factor), *GJB2* (Connexin 26), *MEG3* (Maternally expressed 3), *GNAS* complex locus: (*NESP55* and *NESPAS*) were selected because of their association with human disease, their imprinting or their methylation-dependent regulation. Comparative pyrosequencing revealed that most (5 out of 6) analyzed genes showed very similar methylation patterns of the analyzed CpG sites in the human and the chimpanzee cortex (Figure 23. A and B).

3. Results

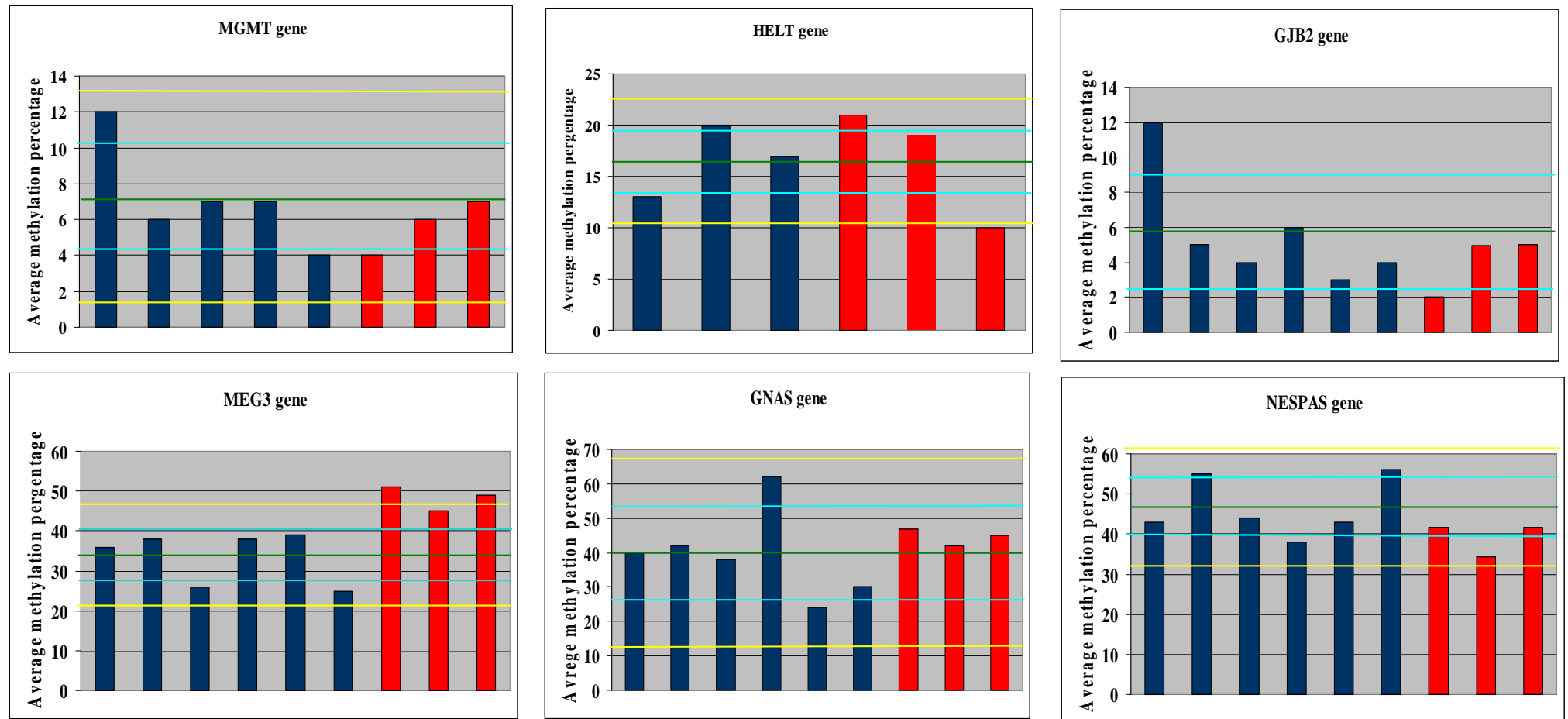


Figure 22. (A) Pyrosequencing data of six genes in human and chimpanzee cortices (mean methylation percentages)

The blue bars represent the human samples and the red bars for the chimpanzee. The green horizontal lines indicate the mean methylation of all human samples. The blue and yellow lines represent first respectively second standard deviation.

3. Results

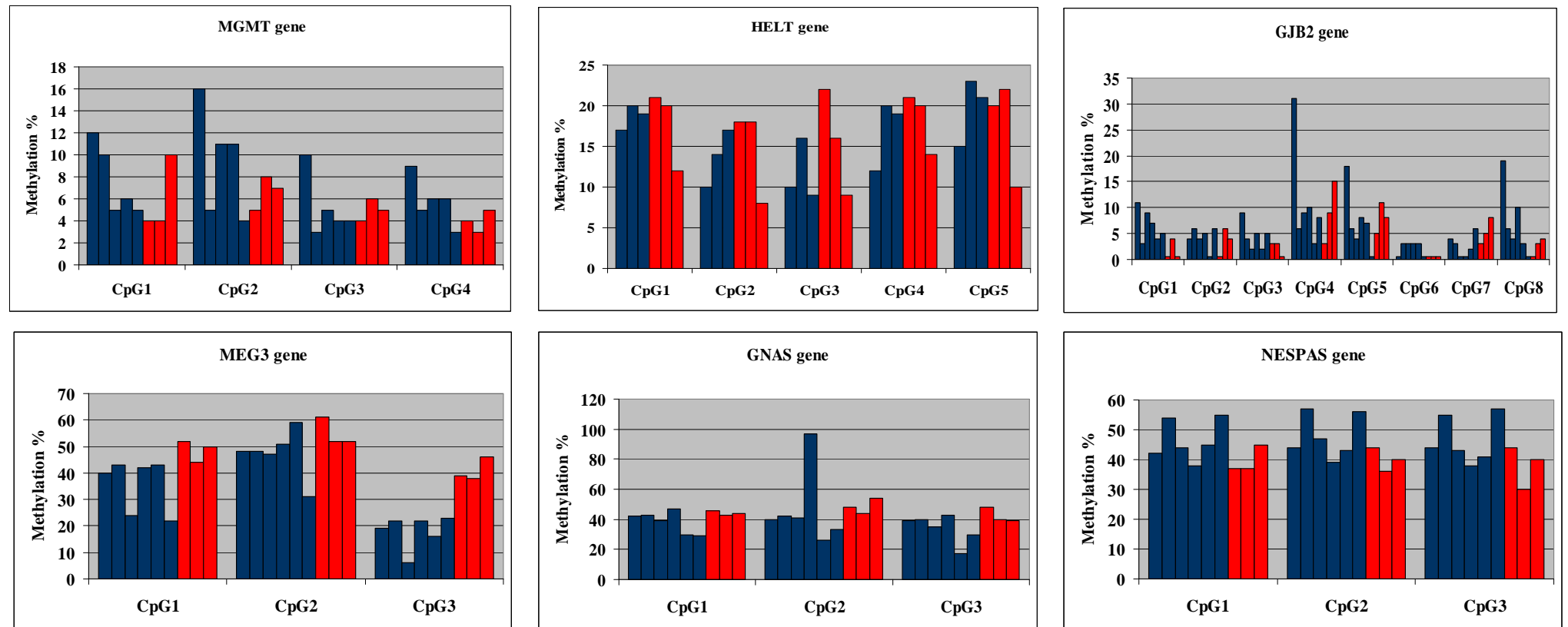


Figure 22. (B) Methylation status of individual CpGs in the six analyzed genes in human and chimpanzee cortices.

The blue bars represent the human samples and the red bars the chimpanzee.

3. Results

For the DNA repair gene *MGMT*, the mean methylation of the four CpGs analyzed in five human and three chimpanzee cortices samples was about the same 6±1%. Interestingly, one of the human brain samples exhibited a slightly higher methylation of all 4 CpG sites.

It is known that in normal human brain tissue *MGMT* shows a stochastic increase in methylation starting around age 50. The human sample (HSA10) with slightly higher methylation was from a 58 year old individual, whereas the other three samples (HSA 8, 9, 11) were from younger individual (< 45 years).

The human *HES/HEY*-like transcription factor has been implicated in schizophrenia and bipolar disorder. It may affect human cognitive functions, i.e. the capacity for complex social relations and language. Higher methylation levels of the CGI promoter region of this gene was found in cortex samples from patients with major psychosis, compared to controls (Mill J. et al., 2008). There was no interspecies variation of *HES* methylation between the analyzed human and chimpanzee samples.

The same conservation of the methylation pattern was also observed for the *GJB2* gene. The gene was unmethylated (< 12%) in all six human and three chimpanzee samples. The methylation pattern of the CpGs in the imprinted genes *GNAS (NESP55)* and *NESPAS* were also conserved between the human and chimpanzee cortices. The interspecific differences between the average mean methylation percentages of the five mentioned genes ranged between 0-6%.

Only one gene *MEG3*, showed a higher inter-species variation. The mean methylation of the six human samples was 33.7% and that of the three chimpanzee samples 48.3%.

Looking at the individual CpG sites in *MEG3*, there was a significant variation in the methylation status between the first two CpG sites and the last one. This difference was more pronounced in humans (~24%) than in chimpanzees(~7%).

3. Results

Another example of inter-CpG variation was seen *GJB2*. CpGs 4, 5 and 8 represent variable sites in both human and chimpanzee, whereas CpG 6 was the most stable site within and between the two species. The three variable CpG sites are part of known transcription binding factor site.

Taken together, these results suggest that even if for most genes there is no significant difference concerning the mean methylation percentages between the two analyzed species the existence of some CpG sites with higher inter- and intraspecific variations, that could contribute to interindividual differences in brain development, function, and disease.

4. DISCUSSION

4.1 Natural Selection of genes at evolutionary breakpoint regions

The interest in selection arose with the demand to understand the forces that shape the evolution of species. Identifying genes or genomic regions that have been influenced by natural selection may provide a key to understand the processes that lead to differences among species. The nine pericentric inversions (on chromosomes 1, 4, 5, 9, 12, 15, 16, 17 and 18) that distinguish the human and chimpanzee karyotypes have attracted considerable attention. These regions might have driven the speciation process that separated the human and chimpanzee lineages. According to the “recombination suppression model, these chromosomal rearrangements, in their heterozygous state, could have acted as a partial barrier to prevent gene flow between early populations of emerging humans and chimpanzees at a time when interbreeding was still possible. Across collinear chromosomes the gene flow remains undisturbed. As a consequence, selected differences are predicted to accumulate more readily in rearranged than in collinear chromosomes.

Considering the structural variation of the human genome, it was an important observation that positively selected genes (PSGs) often coincide with segmental duplications (SD) and that SD are often associated with evolutionary chromosome rearrangement events.(Fan et al. 2002, Locke et al., 2003; Stankiewicz, 2004; Bailey J.A and Eichler E.E, 2006).

It is also known that many PSGs are clustered in the genome, probably reflecting a phenomenon known as “genetic hitchhiking”.

Taken together, these findings promoted the idea that positive selection may operate more intensely on rearranged than on collinear chromosomes during chromosomal speciation.

This hypothesis was first studied by Navarro et. al (2003) who compared 115 pairs of human-chimpanzee orthologs and found that the dN/dS which is- a measure for the pace of protein evolution scaled to mutation rate- is higher in rearranged than in collinear chromosomes. The rearranged chromosomes had indeed experienced more positive selection during human-chimpanzee speciation.

Because the conclusions of the Navaro et. al (2003) would have an important impact for understanding the speciation process, I reexamined this phenomenon in a more systematic way.

My strategy for screening positive selection was to analyze the coding sequences of 2 Mb regions flanking the evolutionary breakpoints on chromosomes 1, 4, 5, 9, 12, 17, 18.

The control region (collinear regions on rearranged chromosomes) is represented by a 4 Mb region localized in 10 Mb distance of the breakpoints. The total number of analyzed genes was 334 in the breakpoint flanking regions (BP-FRs) and 72 genes in the control regions (see results, Table 3.A and B). In all investigated regions a random distributions of selected and non-selected genes was observed except for chromosome 4 where in the inversion region 35 % of the analyzed genes had no non-synonymous changes between human and chimpanzee.

The number of genes under purifying selection (68 genes) in the BP-FR was much higher than that of positively selected genes (17). This result is consistent with the prediction of Kimura and Ohta (1974) that over the evolution of life purifying selection and chance fixation of neutral variants have occurred far more frequently than fixation by positive selection.

Reexamination of the positively selected genes with the newest sequence information (dN/dS ratios) of Ensembl release 52 from 9 Dec. 2008 confirmed 14 of the 17 PSGs.

Collectively, I could not observe a higher rate of protein evolution in BP-FR compared to the control region. My data set does not support the chromosomal speciation hypothesis for humans and chimps.

Vallender and Lahn (2004) examined over 7000 genes using alternative methods and did not find a higher number of PSG on rearranged chromosomes than on collinear chromosomes. Three years later Szamalek et al (2007) compared the sequence divergence of non-coding DNA between humans and chimpanzees and concluded that the divergence is even a little bit lower in inverted chromosomal regions than in non-inverted regions.

By comparing the amounts of positively and negatively selected genes per chromosome, I found that chromosome 9 harbors the highest number of PSG, whereas chromosome 5 has the lowest percentage of selected genes. This is consistent with the study of Szamalek et al., 2007, that found a significantly lower divergence rate within inverted regions, compared to non-inverted regions on chromosome 5.

The modest overlap between genes reported to have signatures of positive selection in different studies can be explained in multiple ways.

First, the different data sets used and the non-uniform distribution of genes as well as the variable quality of the chimpanzee sequence assembly are some factors that could influence the results. This is particularly true for studies published before 2005 when the complete genome sequence of the chimpanzee was not yet available. Even individual low-quality nucleotide positions can result in sites that appear to be falsely divergent between the analyzed species. In addition, it is difficult to exclude that apparent genomic differences between human and chimpanzee represent polymorphisms in one of the species (Ruvolo, 2004).

There are controversial discussions on which of the statistical tests has the biggest power and at the same time the highest sensitivity to detect selection. Each test has its own advantage/disadvantage that has to be taken in consideration. Depending on the dataset to be analysed and the working hypothesis it is essential to choose the most powerful statistical test. Positive selection often acts just on a few sites and in a short period of evolutionary time, and the signal can be masked by the “ubiquitous” negative selection. In general, it can be difficult to identify PSGs because they switched multiple times between selection and nonselection. Moreover, positive selection may act at some sites and negative selection on others, implying that dN/dS is less than 1, even though positive selection has occurred. The idea to distinguish between positive selection and the relaxation of purifying selection is relatively new and hard to prove with the existing statistical tests. Much remains to be learned about positive selection of protein-coding regions in the mammalian genome to understand all the underlying factors that contribute to these results.

It has been argued (by Bleckham R.M. et al. unpublished observations) that positively selected genes are more likely than non-PSGs to cause Mendelian disorders in humans because the environmental factors of humans are much more different from that of primates.

My data do not support this hypothesis because none of the positively selected genes in this thesis has been associated with a human disease in the OMIM database (Online Mendelian Inheritance in Man). In this context it has to be noted that the set of genes under purifying selection (68) is also much larger than the set of PSGs (17).

Since high expression levels and functional importance in a given tissue are often correlated, I examined which tissues show maximum expression of the non-neutrally selected gene (using public data from Novartis Gene Expression Atlas-GNF 1H, MAS5).

Of 13 PSGs with known expression data, I found 6 genes with a maximal expression level in the testis. It is plausible to conclude that many PSGs play a role in reproduction.

4. Discussion

One example is the histone linker H1 domain, spermatid-specific 1 gene (*HILS1*) on chromosome 17 with a dN/dS ratio >2 and a very high expression in the testis. Other studies (Nielsen R. et al., 2005 and Wyckoff G.J. et al, 2000) have also found an accelerated evolution of genes involved in spermatogenesis. One explanation may be the “sperm competition” between males that has driven positive selection of genes involved in male reproductive traits.

In contrary 25% (14 of 57) genes under purifying selection showed their highest expression level in the brain. The predominant biological processes of genes under purifying selection are the nucleoside, nucleotide and nucleic acid metabolism, the protein metabolism and signal transduction.

The selective pressures acting on regulatory processes in general and gene expression levels in particular remain to be elucidated. Studies in model organisms suggest that gene expression levels mostly evolve under stabilizing selection (negative selection), and just a few are due to adaptive evolution (positive selection). Gilad Y. et al., 2006 concluded that in both primates and model organisms, negative selection is the primary force determining expression changes. In this study, I analyzed the dN/dS ratio of 31 genes differential expression between human and chimpanzee (see results, Table 7.A). Seven of these genes were under purifying selection (dN/dS = 0) and just one was positively selected (dN/dS >1).

This supports the idea that genes expression do indeed levels evolve under stabilizing selection. Much greater mechanistic knowledge of gene expression evolution is required because it is difficult to understand and explain the changes in expression simply by genomic sequences.

4.2 Plasticity of CpG islands methylation patterns in human and non-human primates cortices

During the last 5-7 million years of human evolution, the brain has evolved dramatically, giving rise to our unique cognitive abilities. Still, it is currently unknown which genetic differences are responsible for the phenotypic differences that differentiate us from the other primates.

A gene is not only defined by its DNA sequence but also by its expression pattern and by its epigenetic marks through DNA methylation and chromatin modification.

One possibility for the identification of human-specific traits is to compare gene expression patterns between human and chimpanzee. Another useful approach is to compare methylation patterns especially in regulatory DNA sequences (such as promoters), because the methylation status can be viewed as a “footprint” of the chromatin structures that is crucial for gene regulation (Enard et. al., 2004).

Mammalian genomes are globally methylated in the sense that all types of DNA sequences (genes, transposons, intergenic DNA) are targets for CpG methylation.

A recent study of human chromosomes 6, 20 and 22 showed that the majority of the analyzed regions were either hypomethylated (less than 30% of the CpG sites) or hypermethylated (more than 70% of CpG sites). The hypomethylated sites usually represent promoter CpG islands (CGIs) that are associated with 75% of human genes. This promoter hypomethylation might be necessary for the efficient regulation of gene expression. Even if most CGIs remain unmethylated throughout development, a minority become methylated during development and this correlates with transcriptional silencing of the associated gene.

The most prominent examples for this phenomenon are X chromosome inactivation and imprinting. There is no clear idea about the biological significance of DNA methylation changes for regular genes.

Illingworth and colleagues (2008) noted that differentially methylated CGIs are preferentially found in genes that have central roles in development i.e. homeobox (*HOX*) and paired box (*PAX*) genes.

Little is known about the evolutionary conservation of DNA methylation patterns and the evolutionary impact of epigenetic differences between closely related species.

Towards a better understanding how DNA methylation patterns in brain have changed during primate evolution, I analyzed DNA methylation in CpG island promoters of 6 differentially expressed genes between human and chimpanzee.

Comparative bisulphite sequence revealed that most (5 of 6) analyzed genes showed very similar CGI promoter methylation patterns in human and chimpanzee.

With the exception of a few individual sites, the CGI promoters of *ALDH1B1*, *IGFBPL1*, *NTRK2*, *SHC31* and *ZNF519* were completely demethylated. The presence of some partially or fully methylated sites in otherwise demethylated CpG islands may represent stochastic methylation of individual CpG sites.

In agreement with an earlier study (Enard et al., 2004) in my study the majority (11 of 14) of these stochastically individual methylated CpGs were hypermethylated in the human brain, compared to the chimpanzee brain. This suggests a general tendency towards a slightly higher degree of CGI promoter methylation in human lineage.

Even if I can not entirely exclude the formal possibility that methylation changes of individual CpGs contribute to intraspecific variation and/or interspecies differences in brain structure and function, it is unlikely that upmethylation of a single or few sites especially in large CpG islands directly correlate with changes in gene regulation and expression.

Only one of the study genes, *CCRK* (cell cycle related kinase) was endowed with a promoter CGI that appeared to exhibit considerable intraspecific and interspecific variation in the degree of methylation of a cluster of CpG sites. These differentially methylated CpGs lie within an Alu-Sg1 repeat comprising the first half of the CpG island. Genomic sequences alignment of the *CCRK* promoter region between the four analyzed species (human, chimpanzee, rhesus macaque and baboon) revealed a good conservation of the Alu-Sg1 repeat.

The average methylation percentage of the species specifically methylated CpGs in the *CCRK* promoter ranged from 35% in chimpanzees and 40% in the rhesus macaque to 70% in humans and 100% in baboons.

The interindividual variation in the DNA methylation pattern between the 11 unrelated humans is most likely not an effect of the widely different ages of the donors. According to the literature (Siegmund et al., 2007), the methylation of Alus and other repetitive elements in cortex shows a significant decrease during the first decade of human life (0-10 years), followed by relatively little changes during maturation and aging (10-90 years). Instead the dramatic epigenetic differences between individuals can have resulted from stochastic processes, endogenous mechanism and/or environmental perturbations.

Consistent with my findings, bisulphite sequencing of Alu insertion/deletion polymorphisms in three-generation families revealed a significant interindividual variability in the methylation level of specific Alu elements (Sandovici et al., 2005).

The Alu repeat of the *CCRK* promoter CGI is relatively dynamic in its methylation pattern in the brain. This intra/interspecies variation in DNA methylation may contribute to phenotypic variation and potentially to “speciation” itself.

Alus are the most prominent short interspersed nuclear elements (SINEs) adding up to roughly 10% of the human genome (International Human Genome Sequencing Consortium, 2004). More than 20 Alu subfamilies can be distinguished that all have been active with regard to retrotransposition during different periods of primate evolution (Batzer and Deininger, 2002; Xing et al., 2007).

In addition, phylogenetic studies have identified a relation between lineage divergence and increased rates of transposition in primates, promoting the idea that Alu expansions have played role in speciation (Kim et al., 2004).

Alus are nonautonomous, small elements that require the enzymatic machinery provided by LINE-1 expression for retrotransposition.

Although the period(s) of rapid expansion of Alus and other retrotransposons lies in the past of human evolution, the Alus can still be active in the human genome and de novo Alu insertions are an important source of human mutations (Ostertag and Kazazian, 2001).

It is interesting to note that similar to *CCRK* many genes contain one or several Alus in close proximity 5' to their CpG islands (International Human Genome Sequencing Consortium 2004). The neutralist or selectionist “marker” model explains this Alu enrichment by their preferential preservation in gene-flanking regions (Urrutia et al. 2008). In contrast, the “expression modifier” model assumes that Alu insertions near or in promoter regions can influence gene expression (Britten, 1996). For several genes, including the IgE receptor FCER1 γ chain (FCER1G) (Brini et al. 1993), the parathyroid hormone (PTH) (McHaffie and Ralston, 1995) and the non-coding brain cytoplasmic RNA1 (BCYRN1) (Ludwig et al. 2005), it has been demonstrated experimentally that Alu sequences are indeed involved in the regulation of transcription.

The fact that Alus and other retrotransposons are susceptible to epigenetic modification by DNA methylation is consistent with their likely role in gene regulation. Alus can be seen as “centers of the novo methylation” from which methylation spreads into adjacent promoters to induce gene silencing (Turker, 2002).

Especially during cancer development and progression aberrant CGI hypermethylation extends from upstream Alus into their CpG islands, leading to gene silencing (Levine et al., 2003; Stirzaker et al., 2004).

Another intriguing example is the regulation of *KIR* gene selection in natural killer cells.

Different clones express different subsets of genes from the *KIR* cluster at chromosome 10q13, determining the specificity of NK cells. In the non-selected genes methylation appears to spread from upstream Alus into the promoter region (Santourlidis et al., 2002).

DNA methylation is thought to be a major mechanism for preventing retrotransposon activity in human phylogeny and ontogeny (Yoder et al., 1997).

In somatic tissues and mature germ cells, most retroelements are densely methylated and consequently transcriptionally inactive. With the notable exception of some Alus, most retrotransposon-derived elements become methylated at later stages of germ-cell development (Li, 2002). In the zygote and early embryo genome-wide demethylation erases most germline methylation patterns, followed by de novo methylation and establishment of the somatic methylation patterns around the time of implantation (Mayer et al., 2000; Reik et al., 2001).

Interestingly, Alus and some other retroelements show heterogeneous methylation patterns in somatic tissues, with a substantial fraction (10-15%) of Alus being undermethylated (Schmid, 1998; Yang et al., 2004). This suggests that the silencing process is imperfect and incomplete. It has been shown for at least some genes that the epigenetic state of retrotransposon inserted into regulatory regions can be transmitted through the germline (Raykan et al., 2003).

Taken together, their epigenetic properties and ability to influence gene transcription supports the role for retrotransposon inserted into genes (especially in promoters) for phenotypic variation within a species and between species.

By analysing the expression levels of *CCRK* gene, I did not find a clear relationship between the differentially methylated CpGs and gene expression, within a species. This may be largely due to the small number of individuals analyzed per species. Between species the observed methylation differences appeared to be correlated with changes in gene expression.

The relative *CCRK* mRNA levels in human and baboon cortex were approximately two times higher than in chimpanzee and rhesus macaque.

This is consistent with the idea that at the species level global methylation of the species-specifically methylated CpGs abrogates a repressor activity and/or confers an enhancer-like activity.

The common knowledge that DNA hypermethylation is positively coupled with repression of gene expression is not always true. A few reports have suggested that some genes can be activated by CpG methylation. One of the most thoroughly studied models for this phenomenon is the imprinted mouse *Igf2* gene. By preventing binding of trans-acting factors (such as the GCF2 repressor protein), DNA methylation blocks an upstream silencing element(s) in the paternally expressed *Igf2* allele, whereas the inactive maternal allele shows under-methylation of these upstream sequences (Constancia et al. 2000; Eden et al. 2001).

Another example is the *Rad9* oncogene, where specific CpG sites of an intronic silencer are hypermethylated in tumor cell lines that express high *Rad9* mRNA levels and experimental demethylation reduced *Rad9* gene expression (Cheng et al. 2005).

4.3 Comparison of DNA-Methylation patterns in human and chimpanzee cortices by bisulfite Pyrosequencing

The Pyrosequencing technique was used for quantitative DNA methylation analysis of specific CpG sites of six candidate genes in human and chimpanzee cortex.

The analysed CpG sites in *MGMT* gene belong to a known 59 bp enhancer, located at the first exon/intron boundary, which is required for efficient *MGMT* promoter function. Methylation of these CpGs has been associated with gene silencing in cell lines and certain human cancers especially brain tumors (Mikeska et. al., 2007).

In case of *GJB2* gene, the two GC boxes (CCGCCC) in the proximal promoter region of the gene play an important role in the regulation of *GJB2* gene expression (Zheng J.T. and Ktang D.T., 1998). Although these GC box are entirely conserved between human and chimpanzee, the hypermethylation in one of the species may influence the expression level of this gene.

Genomic imprinting is the epigenetic marking of a gene subsets resulting in monoallelic or predominant expression of one of the two parental alleles.

There are currently approximately 80 known imprinted genes, including three of our study genes. A large number of the imprinted genes is expressed in the brain where they may influence neurodevelopment and behaviour. Imprinted genes exhibit parent-of-origin-specific patterns of methylation. It is known that, differentially methylated regions (DMR) are regulatory mechanisms controlling the allele-specific expression of imprinted genes.

Comparative bisulphite pyrosequencing revealed that most (5 of 6) analyzed genes showed very similar methylation patterns of the analyzed CpG sites in human and chimpanzee. Only *MEG3* showed a higher inter-species variation.

Whether the interspecies differences in methylation of the *MEG3* imprinting control region has an effect on the stability of the imprint and/or parent-specific expression remains to be shown.

Imprinting, i.e. of the *IGF2R* genes is not always conserved between mammalian species.

Epigenetic alterations at the *MEG3* locus have been already reported in neuroblastoma, pheochromocytoma, and Wilms tumor. The imprinted domain on 14q32 (including the DMR-*MEG3*) is also a critical region for the upd(14)mat phenotype.

5. SUMMARY

What genetic changes make us different from our closest relative, the chimpanzee, and on the other hand at the same time similar?

What we really want to explore and understand is actually a complex relationship between multiple genetic and epigenetic differences which interact with diverse environmental and cultural factors resulting in the observed phenotypic differences.

In order to elucidate whether or not chromosomal rearrangements contribute to the human-chimpanzee divergence and which are the selective forces that have shaped their evolution, I have analyzed the coding sequences of 2 Mb regions flanking the pericentric inversion breakpoints on chromosomes 1, 4, 5, 9, 12, 17, 18. Collinear regions of 4 Mb on rearranged chromosomes that are separated by at least 10 Mb from the breakpoint regions served as controls.

Collectively, I did not observe a higher rate of protein evolution in the breakpoint flanking regions compared to the control region.

My results do not support the chromosomal speciation hypothesis for humans and chimps because the proportion of positively selected genes (5.1% in breakpoints flanking regions and 7% in the control region), was similar in the two regions.

By comparing the numbers of positively and negatively selected genes per chromosome, I found that chromosome 9 contains the highest number of PSGs in both breakpoint and control regions, whereas chromosome 5 carries the lowest number of selected genes.

The number of genes under purifying selection (68) was found to be much higher than the number of positively selected genes (17).

Bioinformatic analysis of published microarray expression data (affimetrix chip U95 and U133v2) revealed 31 genes which are differentially expressed between human and chimpanzee. By analysing the dN/dS ratio of these 31 genes, I found 7 genes under purifying selection but only one positively selected gene. This is consistent with the idea that genes expression levels evolve under the stabilizing selection. Most of the PSGs play a role in reproduction.

Many of these species differences may be due to changes in gene regulation rather than structural changes in the gene products. Most differences in gene regulation are believed to occur at the transcriptional level. In this study I focused on the differences in the DNA methylation pattern between species.

By classical bisulfite sequencing and pyrosequencing, I analyzed the CpG island promoter methylation patterns of 12 genes in human and chimpanzee cortex.

The candidate genes were selected because of their differential expression pattern between human and chimpanzee or because of their association with human disease or imprinting. With the exception of a few individual sites, the majority of the analysed genes did not present a high intra or interspecific variation of DNA methylation between these two species.

Only one gene *CCRK*, appeared to exhibit considerable intraspecific and interspecific variation in the degree of methylation. This differentially methylated CpG sites lie within an Alu-Sg1 repeat. The study of *CCRK* gene provides a comprehensive analysis of the intra- and interspecific variability of DNA methylation of an Alu insertion in a regulatory region. The observed species differences suggest that the *CCRK* methylation patterns evolve under positive selection, probably in adaptation to specific requirements for fine-tuning of *CCRK* regulation.

The *CCRK* promoter is susceptible to epigenetic modification by DNA methylation, which could result in complex patterns of transcription.

Because of their genomic mobility, high CpG content and their effect on gene expression Alu insertions are extremely attractive candidates for promoting changes in the developmental regulation of primates genes. The intra and interspecific methylation comparisons of specific Alu insertions in other genes and tissues is a promising strategy.

6. REFERENCES

Bai S. et al. (2005) DNA methyltransferase 3b regulates nerve growth factor-induced differentiation of PC12 cells by recruiting histone deacetylase 2. *Mol. Cell. Biol.* 25, 751–766.

Bailey J.A. and Eichler E.E (2006) Primate segmental duplications: crucibles of evolution, diversity and disease. *Nature Rev. Genet.* 7, 552-564.

Bakewell M.A., Shi P., Zhang J. (2007) More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc Natl Acad Sci USA* 104: 7489-7494.

Batzler M.A., Deininger P.L. (2002) Alu repeats and human genomic diversity. *Nat Rev Genet* 3:370-379.

Bird A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6-21.

Brini A.T., Lee G.M., Kinet J.P. (1993) Involvement of Alu sequences in the cell-specific regulation of transcription of the gamma chain of Fc and T cell receptors. *J Biol Chem* 268: 1355-1361.

Britten R.J.(1996) DNA sequence insertion and evolutionary variation in gene regulation. *Proc Natl Acad Sci USA* 93:9374-9377.

Bustamante C.D. et al. (2005) Natural selection on protein-coding genes in the human genome. *Nature* 437, 1153-1157.

Cáceres, M., Lachuer J., Zapala M.A., Redmond J.C., Kudo L., Geschwind D.H., Lockhart D.J., Preuss T.M., Barlow C. (2003) Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl. Acad. Sci. USA* 100: 13030-13035.

6. References

Chen Y., Hühn D., Knösel T., Pacyna-Gengelbach M., Deutschmann N., Petersen I. (2005) Downregulation of Connexin 26 in human lung cancer is related to promoter methylation. *Int. J. Cancer*: 113, 14-21.

Cheng C.K., Chow L.W., Loo W.T., Chan T.K. Chan V. (2005) The cell cycle chrckpoint gene Rad9 is a novel oncogene activated by 11q13 amplification and DNA methylation in breast cancer. *Cancer Res* 65: 8646-8654.

Citron M., Graver M., Schoenhaus M., Chen S., Decker R., Kleynerman L., Kahn L.B., White A., Fornace A.J. and Yarosh D. (1992) Detection of messenger RNA from O⁶ – methylguanine-DNA methyltransferase gene *MGMT* in human normal and tumor tissues. *J.Natl. Cancer Inst.* 84: 337-340.

Clark A.G., Glanowski S, Nielson R, Thomas P.D., Kejariwal A, et al. (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* 302: 1960-1963.

Constancia M., Dean W., Lopes S., Mooret T., Kelsey G., Reik W. (2000) Deletion of a silencer element in *Igf2* results in loss of imprinting independent of *H19*. *Nat Genet* 26: 203-206.

CSAC (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437: 69-87.

Eckhardt F., Lewin J., Cortese R., Rakyan V.K., Attwood J., Burger M., Burton J., Cox T.V., Davies R., Down T.A., et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* 38:1378-1385.

Eden S., Constancia M., Hashimshony T., Dean W., Goldstein B., Johnson A.C., Krshet I., Reik W., Cedar H. (2001) An upstream repressor element plays a role in *Igf2* imprinting. *EMBO J* 20: 3518-3525

Eichler E.E., Lu F., Shen Y., Antonacci R., Jurecic V., Doggett N.A., Moyzis R.K., Baldini A., Gibbs R.A., Nelson D.L., (1996) Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum Mol Genet* 5: 899-912.

6. References

Enard W., Fassbender A., Model F., Adorjan P., Pääbo S., Olek A. (2004) Differences in DNA methylation patterns between humans and chimpanzees. *Curr. Biol.* 14: R148-149.

Enard W., Przeworski M., Fisher S.E., Lai C.S., Wiebe V., et al. (2002) Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418: 869-72.

Fan Y., Linardopoulou E., Friedman C., Williams E., Trask B.J. (2002) Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes. *Genome Res* 12: 1651-1662.

Fan Y., Linardopoulou E., Friedman C., Williams E., Trask B.J. (2002) Genomic structure and evolution of the ancestral chromosome fusion site in human chromosome 2q13-2q14.1 and paralogous regions on other human chromosome. *Genome Res* 12:1651-1662.

Ferguson-Smith, A.C., and Surani, M.A. (2001) Imprinting and the epigenetic asymmetry between parental genomes. *Science*. 293: 1086-1089.

Fisher S.E., Vargha-Khadem F., Watkins K.E., et al. (1998) Localization of a gene implicated in a severe speech and language disorder. *Nat. Genet.* 18: 168-170.

Gilbert, S.L., Ddobyns, W.B. and Lahn, B.T. (2005) Genetic links between brain development and brain evolution. *Nature Rev. Genet.* 6, 581-590.

Gu J., Gu, X. (2003) Induced gene expression in human brain after the split from chimpanzee. *Trends Genet.* 19: 63-65.

Hagan C.R., Sheffield R.F., Rudin C.M. (2003) Human Alu element retrotransposition induced by genotoxic stress. *Nat. Genet.* 35: 219-220.

Huby T., Datchet C., Lawn R.M., Wickings J., Chapman M.J., Thillet J. (2001) Functional analysis of the chimpanzees and human apo(a) promoter sequences: identification of sequence variations responsible for elevated transcriptional activity in chimpanzee. *J Biol Chem.* 276: 22209-22214.

6. References

Hughes J.F., Skaletsky H., Pyntikova T., Minx P.J., Graves T., Rozen S., Wilson R.K., Page D.C. (2005) Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* 437: 100-103.

Illingworth R. et al. (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* 6, e22.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-941.

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-945.

Jaenisch R., Bird A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33: 245-254.

Kehrer-Sawatzki H, Sandig C.A., Goidts V, Hameister H (2005) Breakpoint analysis of the pericentric inversion between chimpanzee chromosome 10 and the homologous chromosome 12 in humans. *Cytogenet Genome Res* 108:91-97.

Kehrer-Sawatzki H., Cooper D.N. (2008) Molecular mechanisms of chromosomal rearrangement during primate evolution. *Chromosome Res.* 16: 41-56.

Khaitovich P., Enard W., Lachmann M., Pääbo S. (2006) Evolution of primate gene expression. *Nat. Rev. Genet.* 7: 693-702.

Khaitovich P., Muetzel B., She X., Lachmann M., Hellmann I., Dietzsch J., Steigele S., Do H-H., Weiss G., Enard W., Heissig F., Arendt T., Nieselt-Struwe K., Eichler E.E and Pääbo S. (2004) Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* 14: 1462-1473.

Kim T.M., Hong S.J., Rhyu M.G. (2004) Periodic explosive expansion of human retroelements associated with the evolution of the hominoid primate. *J Korean Med Sci* 19:177-185.

6. References

Kimura M., (1983) *The Neutral Theory of Molecular Evolution*. (Cambridge Univ. Press, Cambridge, UK), pp34-54.

Klose, R.J. and Bird, A.P. (2006) Genomic DNA methylation: the mark and its mediators. *TRENDS in Biochemical Sciences* Vol 31 No.2.

Kuroki Y., Toyoda A., Noguchi H., Taylor T.D., Itoh T., Kim D.S., Kim D.W., Choi S.H., Kim I.C., Choi H.H., Kim Y.S., Satta Y., Saitou N., Yamada T., Morishita S., Hattori M., Sakaki Y., Park H.S., Fujiyama A. (2006) Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway. *Nat. Genet.* 38: 158-167.

Levine J.J., Stimson-Crider K.M., Vertino P.M. (2003) Effects of methylation on expression of TMS1/ASC in human breast cancer cells. *Oncogene* 22:3475-3488.

Li E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 3:662-673.

Li, E., Bestor, T.H., and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell.* 69: 915-926.

Liu Y., Wu C., Galaktionov K. (2004) p42, a novel cyclin-dependent kinase-activating kinase in mammalian cells. *J. Biol. Chem.* 279: 4507-4514.

Locke, D.P. et al. (2003) Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. *Genome Biol.* 4, R50.

Ludwig A., Rozhdestvensky T.S., Kuryshv V.Y., Schmitz J., Brosius J. (2005) An unusual primate locus that attracted two independent Alu insertions and facilitates their transcription. *J Mol Biol* 350: 200-214.

MacKeigan J.P., Murphy L.O., Blenis J. (2005) Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. *Nat. Cell Biol.* 7: 591-600.

6. References

Mayer W., Niveleau A., Walter J., Fundele R., Haaf T. (2000) Demethylation of the zygotic paternal genome. *Nature* 403:501-502.

McHaffie G.S., Ralston S.H. (1995) Origin of a negative calcium response element in an ALU-repeat: implications for regulation of gene expression by extracellular calcium. *Bone* 17: 11-14.

Mikeska T., Bock C., El-Maarri O., Hübner A., Ehrentraut D., Schramm J., Felsberg J., Kahl P., Büttner R., Pietsch T. and Waha A. (2007) Optimization of quantitative *MGMT* promoter methylation analysis using pyrosequencing and combined bisulfite restriction analysis. *J. Molec. Diagnostics* 9: 3.

Mill J., Tang T., Kaminsky Z., Khare T., Yazdanpanah S., Bouchard L., Jia P., Assadzadeh A., Flanagan J., Schumacher A., Wang S-C and Petronis A. (2008) Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am. J. Hum. Genet.* 82(3): 696-711.

Navarro A., Barton N.H. (2003) Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes. *Science* 300: 321-324.

Newman T.L., Tuzun E., Morrison V.A., Hayden K.E., Ventura M., McGrath S.D., Rocchi M., Eichler E.E., (2005) A genome - wide survey of structural variation between human and chimpanzee. *Genome Res.* 15: 1344-1356.

Ng S.S., Cheung Y.T., An X.M., Chen Y.C., Li M., Li G.H., Cheung W., Sze J., Lai L., Peng Y., et al. (2007) Cell cycle-related kinase: a novel candidate oncogene in human glioblastoma. *J. Natl. Cancer Inst.* 99: 936-948.

Nickerson E., Nelson D.L., (1998) Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees. *Genomics* 50: 368-372.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, et al. (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PloS Biol* 3: e170.

6. References

Ostertag E.M., Kazazian H.H. (2001) Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 35:501-538.

Park, Y., and Kuroda, M.I. (2001) Epigenetic aspects of X chromosome dosage compensation. *Science*. 293: 1083-1085.

Patterson N., Richter D.J., Gnerre S., Lander E.S., Reich D. (2006) Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441: 1103-1108.

Petter Portin (2007) Evolution of man in the light of molecular genetics: a review. Part I. Our evolutionary history and genomics. *Hereditas* 144: 80-95.

Raykan V.K. Chong S., Champ M.E. Cuthbert P.C., Morgean H.D., Luu K.V., Whitelaw E. (2003) Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci USA* 100:2538-2543.

Razin, A. and Cedar, H. (1993) DNA methylation and embryogenesis. *EXS* 64, 343-57.

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

Rhesus Macaque Genome Sequencing and Analysis Consortium (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 13: 222-234.

Robertson, K. D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F. A., and Jones, P. A. (1999) The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Res* 27, 2291-8.

Rollins R.A., Haghghi F., Edwards J.R., Das R., Zhang M.Q., Ju J., Bestor T.H. (2006) Large-scale structure of genomic methylation patterns. *Genome Res*. 16: 157-163.

6. References

Ruvolo M., (2004) Comparative primate genomics: The year of the chimpanzee. *Curr. Opin. Genet. Dev.* 14: 650-656.

Sabeti P.C., Schaffner S.F., Fry B., Lohmueller J., Varilly P., Shamovsky O., Palma A., Mikkelsen T.S., Altshuler D., Lander E.S. (2006) Positive Natural Selection in the human lineage. *Science* vol 312.

Sandovici I., Kassovska-Bratinova S., Loredó-Osti J.C., Leppert M., Suarez A., Stewart R., Bautista F.D., Schiraldi M., Sapienza C. (2005) Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. *Hum Mol Genet* 14: 2135-2143.

Santourlidis S., Trompeter H.I., Weinhold S., Eisermann B., Meyer K.L., Wernet P., Uhrberg M. (2002) Crucial role of DANN methylation in determination of clonally distributed killer cell Ig-like receptor expression patterns in NK cells. *J Immunol* 169:4253-4261.

Schmid C.W. (1998) Does SINE evolution preclude Alu function? *Nucleic Acids* 26:4541-4550.

Schulz W.A., Steinhoff C., Florl A.R. (2006) Methylation of endogenous human retroelements in health and disease. *CTMI* 310: 211-250.

Shen J.C., Rideout W.M., Jones P.A. (1994) The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* 22: 972-976.

Siegmund K.D., Connor C.M., Campan M., Long T.I., Weisenberger D.J., Biniszkiwicz D., Jaenisch R., Laird P.W., Akbarian S. (2007) DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. *PLoS One* v.2(9):e895.

Stankiewicz P., Shaw C.J., Withers M., Inoue K., Lupski J.R. (2004) Serial segmental duplication during primate evolution result in complex human genome architecture. *Genome Res.* 14: 2209-2220.

6. References

Stirzaker C., Song J.Z., Davidson B., Clark S.J. (2004) Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res* 64:3871-3877.

Szamalek J.M., Cooper D.N., Hoegel J., Hameister H., Kehrer-Sawatzki H. (2007) Chromosomal speciation of humans and chimpanzees revisited: studies of DNA divergence within inverted regions. *Cytogenet Genome Res* 116:53-60.

Szamalek J.M., Goidts V, Searle J.B., Cooper D.N., Hameister H, Kehrer-Sawatzki H (2006) The chimpanzee-specific pericentric inversions that distinguish humans and chimpanzees have identical breakpoints in *Pan troglodytes* and *Pan paniscus*. *Genomics* 87:39-45.

The Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequencing of the chimpanzee genome and comparison with the human genome. *Nature* 437: 69-87.

Turker M.S. (2002) Gene silencing in mammalian cells and the spread of DNA methylation. *Oncogene* 21:5388-5393.

Urrutia A.O., Ocana L.B., Hurst L.D. (2008) Do Alu repeats drive the evolution of the primate transcriptome? *Genome Biol.* 9:R25.

Vallender E.J. and Lahn B.T. (2004) Effects of chromosomal rearrangements on human-chimpanzee molecular evolution. *Genomics* 84:757-761.

Varki A. and Tasha K.A. (2005) Comparing the human and chimpanzee genomes: Searching for needles in a haystack. *Genome Res.* 15: 1746-1758.

Weber M., Hellmann I., Stadler M.B., Ramos L., Pääbo S., Rebhan M., Schübeler D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* 39: 457-466.

Wienberg J. (2005) Fluorescence in situ hybridization to chromosomes as a tool to understand human and primate genome evolution. *Cytogenet Genome Res* 108:139-160.

6. References

- Wolffe, A.P., Matzke, M.A. (1999) Epigenetics: regulation through repression. *Science* 286:481–486.
- Wray G.A., Hahn M.W., Abouheif E., Balhoff J.P., Pizer M., Rockman M.V., Romano L.A. (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20: 1377-1419.
- Wyckoff G.J., Wang W., and Wu C.I. (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature* 403, 304-309.
- Xing J., Witherspoon D.J. Ray D.A., Batzer M.A., Jorde L.B. (2007) Mobile DNA elements in primate and human evolution. *Am J Phys Anthropol Suppl* 45, 2-19.
- Yang A.S., Estecio M.R., Doshi K., Kondo Y., Tajjry E.H. Issa J.P. (2004) A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 32:e38.
- Yoder J.A., Walsh C.P., Bestor T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13:335-340.
- Yunis J.J and Prakash O. (1982) The origin of man: a chromosomal pictorial legacy. *Science* 215: 1525-30.
- Zhang J., Nielsen R., Yang Z. (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22: 2472-2479.
- Zhang J (2004) Frequent false detection of positive selection by the likelihood method with branch-site models. *Mol Biol Evol* 21: 1332-1339.
- Zhang J., Wang X., Podlaha O. (2004) Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Res* 14:845-851.
- Zhang J., Webb D.M., Podlaha O. (2002) Accelerated protein evolution and origins of human-specific features: *Foxp2* as an example. *Genetics* 162: 1825-35.

7 ANNEX

7.1 Abbreviations

A	Adenine
Alu	Are the most prominent short interspersed nuclear elements (SINEs)
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BP	Evolutionary breakpoint
BP-FR	Breakpoint flanking region
°C	Degree Celsius
C	Cytosine
cDNA	Complementary deoxyribonucleic acid
CGI	CpG island
CR	Control region
CSAC	The Chimpanzee Sequencing and Analysis Consortium
DEPC	Diethylpyrocarbonate
dN	Non-synonymous substitution rate
DNA	Desoxyribonucleic acid
DNMTs	DNA methyltransferases
dNTP	Deoxynucleoside triphosphate
dS	Synonymous substitution rate
EDTA	Ethylenediaminetetracetic acid
F	Female

7. Annex

G	Guanine
HCP	High CpG promoter
hr	Hour
HSA	<i>Homo sapiens</i>
ICP	Intermediate CpG promoter
IPTG	Isopropyl- β -thiogalactopyranoside
-K	Negativ control
kb	Kilobase (1000bp)
LB	Luria Bertani (medium)
LINE	Long interspersed nuclear element
M	Male
M	Molar
Mb	Megabase
MBDs	Methyl-CpG binding proteins
mg	Milligram
min	Minute
μ l	Microliter
ml	Milliliter
mM	Milimolar
MMU	<i>Macaca mulata</i>
mRNA	Messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
nm	Nanometer
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PHA	<i>Papio hamadryas</i>

7. Annex

PMI	Post-mortem interval
PSGs	Positive selected genes
PTR	<i>Pan troglodytes</i>
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Rounds per minute
RSC	Relaxation of selective constraints
RT	Reverse transcriptase
RT-PCR	Reverse transcriptase-PCR
SD	Segmental duplication
SDS	Sodiumdodecyl sulfate
sec	Second
SINE	Short interspersed nuclear element
T	Thymine
TAE	Tris Acetate EDTA buffer
Taq	<i>Thermus aquaticus</i> (DNA polymerase)
TE	Tris/EDTA (buffer)
TEs	Transposable elements
TF	Transcription factor
TSS	Transcription start site
U	Unit
V	Voltage
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactoside

7.2 Figure Index

Figure 1. The Epigenetic mechanisms (adapted from <i>Nature</i> , 7 Aug 2008)	13
Figure 2. Pie charts showing the relative frequency of CGI classes among total promoters and hypermethylated promoters	17
Figure 3. Mechanisms of DNA methylation mediating repression	17
Figure 4. The nine pericentric inversions that distinguish human and chimpanzee karyotypes	53
Figure 5. Breakpoint and control region for in silico sequence analysis	54
Figure 6. Example of an alignment for the coding sequence of the <i>HES1</i> gene	55
Figure 7. Chromosome 4 pericentric inversion HSA-PTR	63
Figure 8. Chromosome 5 pericentric inversion HSA-PTR	65
Figure 9. PTR fusion sequence of chromosome 9	66
Figure 10. Amplification of the chimpanzee fusion sequence from human chromosome 9	67
Figure 11. Chromosome 9 pericentric inversion HSA-PTR	68
Figure 12. (A) Biological Process classification of the positive selected genes	75
Figure 12. (B) Biological Process classification of genes under purifying selection	75

Figure 13. Methylation patterns of orthologous CpG island promoters in human and chimpanzee cortex	83
Figure 14. Nucleotide sequence alignment of the <i>CCRK</i> CGI promoter for human, chimpanzee, rhesus macaque and baboon.	86
Figure 15. (A) Methylation pattern of the CGI promoter of <i>CCRK</i> in the 11 humans	87
Figure 15. (B) Methylation pattern of the CGI promoter of <i>CCRK</i> in 3 chimpanzees.	87
Figure 15. (C) Methylation pattern of the CGI promoter of <i>CCRK</i> in 3 baboons.	88
Figure 15. (D) Methylation pattern of the CGI promoter of <i>CCRK</i> in 1 rhesus macaque.	88
Figure 16. Methylation percentiges (mean \pm standard deviation) in human cortex (n=11), chimpanzee (n=3), rhesus monkey (n=1) and baboon (n=3).	89
Figure 17. Example of CpGs that are more or less methylated in human compared to chimpanzee.	90
Figure 18. (A) Variation of <i>CCRK</i> CGI promoter methylation among humans.	91
Figure 18. (B) Variation of <i>CCRK</i> CGI promoter methylation among humans	91
Figure 19. Variation concerning the methylation percentages of individual CpG sites in the adult cortices of three unrelated chimpanzees (PTR).	92

Figure 20. Methylation percentages of individual CpG sites in the adult cortices of three unrelated baboons (PHA).	93
Figure 21. (A) Average methylation percentage (and standard deviation) of the differentially methylated CpG sites (6-10A) in human and non-human primates cortices.	95
Figure 21. (B) Expression of <i>CCRK</i> gene in human and non-human primates cortices.	95
Figure 21. (C) Corelation of the average methylation percentage of the species-specifically methylated CpG sites and the relative mRNA expression of <i>CCRK</i> .	96
Figure 22. (A) Pyrosequencing data of six genes in human and chimpanzee cortices (mean methylation percentiges).	98
Figure 22. (B) Methylation status of individual CpGs in the six analyzed genes in human and chimpanzee cortices.	99

7.3 Table Index

Table 1. Brain tissue samples used in this study.	22
Table 2. (A) Primers used for the qRT-PCR.	23
Table 2.B Classical Bisulfite Sequencing Primer.	24
Table 2.C Bisulfite Pyrosequencing Primers.	24
Table 3. (A) Number of genes analysed on chromosome 1, 4, 5, 9, 12, 17, 18.	69
Table 3. (B) Percentage of genes under non-neutral selection on chromosome 1, 4, 5, 9, 12, 17, 18 for the analyzed regions.	70
Table 4. (A) Positively selected genes in the analyzed breakpoint flanking regions.	70
Table 4. (B) Genes under purifying selection in the analyzed breakpoint flanking regions.	71
Table 5. Hyperconserved genes $dN = 0$ and $dS = 0$.	73
Table 6. Positively and negatively selected genes with disease association.	73
Table 7. (A) Data mining of gene expression profiles identifies differentially expressed genes between human and chimpanzee.	78
Table 7. (B) Data mining of gene expression profiles with the Bonferroni adjustment.	78

Table 8. (A) Candidate genes and their analyzed CpG. 81

Table 8. (B) Candidate genes and their %GC and CpG ratio. 81

Table 9. Candidate genes and the number of CpG sites analyzed by Pyrosequencing. 97

7.4 Acknowledgments

7.5 Curriculum vitae

7.6 Publication

1: [Mol Biol Evol.](#) 2009 Mar 12. [Epub ahead of print]

[Related Articles](#), [Links](#)



Differences in DNA Methylation Patterns and Expression of the CCRK Gene in Human and Non-human Primate Cortices.

[Farcas R.](#), [Schneider E.](#), [Frauenknecht K.](#), [Kondova I.](#), [Bontrop R.](#), [Bohl J.](#), [Navarro B.](#), [Metzler M.](#), [Zischler H.](#), [Zechner U.](#), [Daser A.](#), [Haaf T.](#)

Institute for Human Genetics, Johannes Gutenberg University, 55101 Mainz, Germany.

Changes in DNA methylation patterns during embryo development and differentiation processes are linked to the transcriptional plasticity of our genome. However, little is known about the evolutionary conservation of DNA methylation patterns and the evolutionary impact of epigenetic differences between closely related species. Here we compared the methylation patterns of CpG islands (CGIs) in the promoter regions of 7 genes in humans and chimpanzees. We identified a block of CpGs in the cell cycle related kinase (CCRK) gene that is more methylated in the adult human cortex than in the chimpanzee cortex and, in addition, it exhibits considerable intraspecific variation both in humans and chimpanzees. The species-specifically methylated region (SMR) lies between the almost completely methylated 5' region and the completely demethylated 3' region of the presumed CCRK CGI promoter. It is part of an Alu-Sg1 repeat that has been integrated into the promoter region in a common ancestor of humans and New World monkeys. This SMR is relatively hypomethylated in the rhesus monkey cortex and more or less completely methylated in the baboon cortex, indicating extraordinary methylation dynamics during primate evolution. The mRNA expression level of CCRK has also changed during the course of primate evolution. CCRK is expressed at much higher levels in human and baboon cortices, which display an average SMR methylation of 70% and 100%, respectively, than in chimpanzee and rhesus macaque cortices with an average SMR methylation of 35% and 40%, respectively. The observed evolutionary dynamics suggests a possibility that CCRK has been important for evolution of the primate brain.

PMID: 19282513 [PubMed - as supplied by publisher]