Rapid Evolution of DNA Methylation in Primates Tend to Occur in Conserved Sequences

Fang Wang, Hui Yang, Min Zhang, Shaojun Zhang

Abstract DNA methylation is an important epigenetic modification, which is involved in developmental and regulatory process, such as genome imprint, X-chromosome inactivation and transcriptional silencing. Though there are more and more studies on DNA methylation, it remains unclear about the differences of DNA methylation between human and its closely related primate species. Differentially methylated genes (DMGs) were defined by comparing two models in our paper then some following analysis were done. We found that the patterns of DNA methylation are divergent in promoter and gene body. The functional enrichment of DMG showed association with developmental process, which demonstrated that the changes of DNA methylation might be frequent in the evolution of primate species. Furthermore we found the negative relationship between DNA methylation and gene expression by integrating the DNA methylation data and gene expression data in prefrontal cortex in human and chimpanzee. We compared the DNA methylation of human and chimpanzee relative to their common ancestor to analyze the special pattern of DNA methylation in human and the results demonstrated the rapid evolution of hypermethylation in promoter. We also found the conversation in protein coding regions in human and chimpanzee while the epigenetic modification changes occurred, this might explained the important role of epigenetic modification in the divergence between human and chimpanzee. Our study would be helpful to reveal the pattern and evolution of DNA methylation in human, chimpanzee and macaque.

Keywords DNA methylation; Primate; Evolution; Promoter; Gene body

Introduction

Studies on the theory of natural selection demonstrate that the variation of phenotype among species results from the mutation of DNA sequence. DNA mutation may get different transcripts and then result in more kinds of protein, further lead to the phenotypic differences (Wittkopf and Kalay, 2012). But the genomic comparison between human and primate species indicate that the variations of sequence among these species are not as significant as expected. A most outstanding example is the variation of sequence between human and chimpanzee, as the divergent sequence only accounts for 1.2% of whole genome, which demonstrates that DNA mutation is not enough for explaining the variable phenotypes in different species (Sequencing and Consortium, 2005). With the application and development of gene chip, people find some species-specific or tissue-specific genes by comparing gene expression profiles, which could make some explanations for the species-specific phenotypes (Cáceres et al., 2003; Fraser et al., 2005). However the rate of differentially expressed genes between human and primate species is very lower, which demonstrates that gene expression is not enough to fully explain the species-specific phenotypes (Chan et al., 2009). The important role of the transcriptional regulation of phenotypic variation, although it was suggested as early as 1970s (King and Wilson, 1975), until the last decade, along with the rapid development of the next generation sequencing technology , the driving role of transcriptional and post-transcriptional regulation on phenotypic differences has been recognized gradually recognized, including alternative splicing, the regulation of transcription factor , epigenetic modifications and so on (Hobert, 2008; Huffman, 2000; Ishii et al., 2002).
Epigenetics is not the traditional genetics, and the difference is that epigenetics don’t change the DNA sequence, but it will lead to changes in gene expression, and obtain some heritable phenotypes. Up to now, histone modifications, non-coding RNA, DNA methylation are the epigenetic phenomena studied more and more. Histone modifications could affect many life processes by altering the chromatin structure (Tost, 2008). Studies have found the important role of non-coding RNA in transcription inhibition, mRNA degradation, and alternative splicing (Morris, 2012). About DNA methylation, the researchers found its role in silencing gene expression (Zeng et al., 2012), and the demethylation and re-methylation of DNA in embryonic development reflects its unique role and mechanisms in the embryonic development process (Kim et al., 2009).

As a relatively conservative chemical modification, DNA methylation is prevalent in different species, but the patterns of DNA methylation exhibit specific biological characteristics. For example, in fungus, only 0.1-0.5% of CpG dinucleotides are methylated, while in mammals up to 60-90% of CpGs are methylated. In addition, the comparison of vertebrate and invertebrate genomes exhibits great differences in the level, functional role and genomic distribution of DNA methylation (Zeng et al., 2012).

There are many species studies on the differences of DNA methylation between human and chimpanzee. The comparison of DNA methylation profiles between species can make some explanations for human-specific phenotypes, and confirm the presence of transcriptional silencing and its regulatory role in the development process (Hernando-Herraez et al., 2013; Zeng et al., 2012). However these studies are mostly based on the mean value of DNA methylation, which would ignore the effect of CpG sites, so they may mask some CpG sites with important biological significance. Therefore, we propose the integration of mixed linear models and residual maximum likelihood estimation, and utilize the information of CpG sites to investigate the DNA methylation of human and chimpanzee.

Our study utilized the whole genomic DNA methylation of human, chimpanzee and macaque downloaded from GEO (Barrett et al., 2013) and fitted the data by combing mixed linear models and residual maximum likelihood estimation. The differentially methylated genes in the promoter and gene body were screened by comparing two models, and then we investigated the evolution of DNA methylation in primate species from some aspects. This paper proposes a new idea for comparing DNA methylation, which considers the information of CpG sites and could make up for past deficiencies raised by averaging method, and the idea can also be applied to other problems. Comparative study of DNA methylation in promoter and gene body reveals the different mechanisms of DNA methylation in the two regions. This study will be helpful to understanding the role DNA methylation in the evolution of primate species, and making some explanations for the special traits in human.

1 Methods

1.1 Data and processing.

DNA methylation data were download from Gene Expression Omnibus with accession no.[GSE37202] (Zeng et al., 2012) which included 3 human samples and 3 chimpanzee samples, and no.[GSE34124] (Tung et al., 2012) which included rhesus macaques. All of the methylation maps were got from whole-genome sequencing of bisulfite-converted DNA. In order to compare the methylation levels between different species, we downloaded the orthologous genes genomic coordinates from Ensembl (Kinsella et al., 2011). Promoters were defined as regions 1.5kb upstream and 0.5 kb downstream from transcription start sites. Gene bodies were defined as regions began from transcription start sites to transcription end sites. 22 675 orthologous genes were utilized to analysis. We split DNA methylation data into single promoter or gene body using JAVA program. Gene expression data used in this study were downloaded from the Gene Expression Omnibus (GEO) repository with accession number GSE33587 (Konopka et al., 2012).
Transcription factor binding sites (TFBS) data were downloaded from UCSC (http://eqtl.uchicago.edu/Home.html) in which the TFBS were identified by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq) from cell lines of brain (Pique-Regi et al., 2011).

1.2 Screen Differentially Methylated Genes by Mixed Linear Models

Single gene’s DNA methylation data both in promoter and gene body region were got in the last step, and then differentially methylated genes were defined by comparing two models. We considered two models as follows:

\[ \text{Model A: } y_i = u_{ij} + \alpha_i + \beta_k + \epsilon_i \]
\[ \text{Model B: } y_i = u_i + \alpha_i + \beta_k + \epsilon_i \]

(1)

where \( y_i \) is the methylation of gene \( i \), \( u_{ij} \) is the fixed effect which refers species, \( \alpha_i \), \( \beta_k \), \( u_i \) are the random effects which refers the CG site in gene \( i \), the sample \( k \), species \( t \) respectively.

Model A considered species as the fixed effect which means there were differences on the level of methylation between species while Model B considers species as well as the CG sites in gene, the sample as random effects, which meant there were no difference on the level of methylation between species. DNA methylation of single gene in promoter and gene body were fitted using lmer() in R package lme4 (Bates et al., 2012), where the criteria for fitting was Restricted Maximum Likelihood, REML (Corbeil and Searle, 1976). Compared to Maximum Likelihood, REML considered the influence to variance which was caused by the decrease of freedom when estimating the fixed effect, so the deviation of estimating would be smaller.

We investigated the DNA methylation in promoter and gene body in two respects according the differentially methylated genes. Firstly, we built two distance matrices of the number of DMGs in promoter and gene body, and then we got the dendrograms using hclust in R. Secondly, functional enrichment was done using the DMGs in promoter and gene body. We used the online tools DAVID (Dennis Jr. et al., 2003), where we chose GO Ontology and KEGG to analyze and the threshold of Benjamini was 0.01.

1.3 Comparative DNA Methylation Patterns between Promoter and Region

The model is better if its \( AIC \) value is smaller, and \( \Delta_i(AIC) \) is the differences in \( AIC \) with respect to the \( AIC \) of the best candidate model. The value of \( w_i(AIC) \) ranges from 0 to 1, if the \( w_i(AIC) \) is near 1 means \( AIC \) is smaller significantly and the model with smaller \( AIC \) is supported by the data. We defined the differentially methylated genes as the one whose methylation data supported model A. That is to say, if the \( w_i(AIC) = 1 \) for gene \( i \), gene \( i \) was defined as DMG.

1.4 The Enrichment of Transcription Factors

We downloaded 13 transcription factor binding sites data from UCSC, including CTCF, POLR2A, EZH2, FOXP2, TAF1, SIN3AK20, NRSF, GATA2, GATA3, EP300, RAD21, USF1, YY1. These TFBS were got from 10 brain-related cell lines by ChIP-Seq experiments, which included BE2_C, Gliobla, HAc, NH-A, PFSK-1, SK-N-MC, SH-SY5Y, SK-N-SH, SK- N-SH RA, U87. Enrichment analysis of transcription factors were conducted using hypergeometric test to test if a particular transcription factor appeared in DMGs between species in promoter region and the formula is as follows:

\[ p(x \geq k) = 1 - \frac{C_M^k C_n^{n-k} C_{n-M}^k}{C_n^n} \]

(5)

where \( k \) is the number of binding sites of a particular transcription factor in a differentially methylated gene set, \( M \) is the number of binding sites of a particular transcription factor in all genes, \( N \) is the number of binding sites of 13 transcription factors in all genes, \( n \)
is the number of binding sites of 13 transcription factors in differentially methylated gene set. Equation 5 represents the probability of there are at least \( k \) binding sites of a particular transcription factor in differentially methylated genes. The gene set is enriched on a particular transcription factor if the \( p \) value of hypergeometric test was less than 0.01.

Besides, in order to reflect the degree of enrichment of transcription factor across species differentially methylated gene sets, dominance ratio of a transcription factor is defined as the proportion of the transcription factor binding sites in the differentially methylated gene set divided by that of transcription factor binding sites in all binding sites, and the formula is as follows:

\[
\text{ratio} = \log\left(\frac{k/n}{M/N}\right)^{2}
\]  

(6)

1.5 The Relationship between DNA Methylation and Gene Expression

DNA methylation and gene expression levels were analyzed in DMGs between human and chimpanzee. DNA methylation data with no GSE37202 were downloaded from GEO which included three human samples and three chimpanzee samples. Gene expression data were from GEO's GSE33587, including RNA-Seq data from six human samples and six chimpanzee samples. Both data about DNA methylation and genome data are from the brain's prefrontal cortex. After data processing, only the genes with both DNA methylation and gene expression was reserved, which included 720 genes in promoter and 3910 genes in gene body. According to the DNA methylation levels, genes were divided into hypermethylated genes in human and hypermethylated genes in chimpanzee. Then we statistic the expression level in each case and plotted a bar graph to demonstrate.

1.6 The Divergence between Human and Chimpanzee relative to the Common Ancestor

We invested the differentially methylated genes between human and chimpanzee to find the special DNA methylation patter in human, so the concept of ancestor’s DNA methylation was defined. If the DNA methylation of macaque was between that of human and chimpanzee, it was defined as the ancestor’s DNA methylation. Otherwise, we chose the species which was closer to the DNA methylation of macaque and then evaluated the median of the chosen species and macaque’s DNA methylation as the common ancestor’s DNA methylation. In the next step, a fold-change value was given to measure the divergence between human and chimpanzee relative to the common ancestor as follows:

\[
\text{fold-change} = \log\left(\frac{\text{methyl}_{\text{human}}/\text{methyl}_{\text{ancestor}}}{\text{methyl}_{\text{chimp}}/\text{methyl}_{\text{ancestor}}}\right)^{2}
\]  

(7)

Where \( \text{methyl}_{\text{human}} \), \( \text{methyl}_{\text{chimp}} \), \( \text{methyl}_{\text{ancestor}} \) represent the DNA methylation of human, chimpanzee and common ancestor respectively.

Fold change value were defined to evaluate the difference (Formula 7) and four situations in human were got:

1. Specially hypo: \( \text{methyl}_{\text{human}} < \text{methyl}_{\text{ancestor}}, \text{methyl}_{\text{chimp}} > \text{methyl}_{\text{ancestor}} \)
2. Specially hyper: \( \text{methyl}_{\text{human}} > \text{methyl}_{\text{ancestor}}, \text{methyl}_{\text{chimp}} < \text{methyl}_{\text{ancestor}} \)
3. Common hypo: \( \text{methyl}_{\text{human}} < \text{methyl}_{\text{ancestor}}, \text{methyl}_{\text{chimp}} < \text{methyl}_{\text{ancestor}} \)
4. Common hyper: \( \text{methyl}_{\text{human}} > \text{methyl}_{\text{ancestor}}, \text{methyl}_{\text{chimp}} > \text{methyl}_{\text{ancestor}} \)

Then we got the situations of DNA methylation of human and chimpanzee relative to their common ancestor and conduct functional enrichment of genes with special DNA methylation levels in human by DAVID.

1.7 Relationship between DNA methylation and sequence about evolution

The homologous genes of human and chimpanzee are split into three sets. Gene set A contains genes that is directional selection in human, gene set B contains stable selected genes in human and chimpanzee which has the sample size with gene set A; gene set C refers
the rest of the homologous genes. dN/dS values are calculated according to the dN, dS downloaded from the Ensembl database, then we can obtain gene sets A (525 genes), B (525 stable genes), C (11653 genes). The density distribution of three gene sets of dN/dS are drawn using R, and we statistic the percentage of genes whose dN/dS are less than 0.25, more than 1 respectively.

2 Results
2.1 Differential DNA methylation Genes between species using Mixed Linear Model.
Our study utilized the whole genomic DNA methylation of human, chimpanzee and macaque downloaded from GEO and fitted the data by combining mixed linear models and residual maximum likelihood estimation. The differentially methylated genes in the promoter and gene body were screened by comparing two models. The numbers of differentially methylated genes between human and chimpanzee, human and macaque, chimpanzee and macaque are 877 2163, 1 605 respectively in promoter while the numbers are 4 407, 3 023, 4 598 respectively in gene body (Table 1).

<table>
<thead>
<tr>
<th>Species pairs</th>
<th>Human vs.chimp</th>
<th>Human vs. macaque</th>
<th>Chimp vs. macaque</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Promoter</td>
<td>Gene body</td>
<td>Promoter</td>
</tr>
<tr>
<td>DMGS</td>
<td>877</td>
<td>4407</td>
<td>2163</td>
</tr>
<tr>
<td>Hyper</td>
<td>65</td>
<td>1176</td>
<td>1268</td>
</tr>
<tr>
<td>Hypo</td>
<td>813</td>
<td>3231</td>
<td>895</td>
</tr>
</tbody>
</table>

Hyper/hypo- represents the methylation in the former species is higher/lower than the latter.

Then we plot the dendrograms according to the numbers in table 1 as Figure 1A, 1B.

The divergence between two clusters (Figure 1A,1B) indicated the patterns of promoter and gene body are different. In promoter region, human and chimpanzee were clustered into one class firstly then is the macaque while in gene body, human and macaque were clustered into one class firstly then is the chimpanzee. Measuring DNA methylation patterns using the numbers of differentially methylated genes, indicated that in promoter region, the DNA methylation patterns were more similar between human and chimpanzee while in gene body, the DNA methylation patterns were more similar between human and macaque. This phenomenon demonstrated the divergence between promoter and gene body on the evolutional patterns of DNA methylation.

In addition, we also filtered differentially methylated genes using T Test in R according to the mean of DNA methylation in each gene. A differentially methylated gene were defined if the fold change was more than 2 or less than 0.5, FDR was less 0.01, DNA methylation level was more than 0.8 and another was less than 0.2. The clusters got by these numbers of differentially methylated genes were Figure 1C, and 1D.

Can be seen that, using the mean values would get more species differentially methylated genes in...
promoter region, but obtain less differentially methylated genes in gene body region. The dendrograms conduct DMGS according to means of DNA methylation couldn’t reflect the history of evolution. The reason leading to this phenomenon may be due to not consider the impact of CpG sites and samples on DNA methylation.

2.2 Functional Enrichment of DMGs in Promoter and Gene Body
DMGs between three species pairs were used to do functional enrichment through DAVID. We chose the biological process of Gene Ontology and KEGG PATHWAY to do analysis, the results were fitted by p value of Benjamini < 0.01. Figure 2A shows the functional enrichment of differentially methylated genes in promoter region. We can see that the DMGs using the method of mixed linear model and using T test are both enriched in the process of development, cell process, tissue cellular components, metabolic processes, and death. In addition, mixed linear model method are further enriched on regulatory process but this function was not enriched using the T Test. Combined with the results already obtained, namely biological regulation of DNA methylation to transcriptional silencing. It can be concluded that using the average value of DNA methylation may mask some real biological phenomena. Figure 2B is the enrichment of DMGs in promoter about KEGG pathways. Mixed linear model method is enriched on Oocyte meiosis and MAPK signaling pathway, but don't appear in the average methylation method. This phenomenon also can be seen in Figure 2C, 2D.

From Figure 2, differences of functional enrichment in promoter and gene body can be observed. GO functional enrichment results shown in Figure 2A, 2C reflects the two regions were involved in some fundamental life processes including developmental process, cellular process, cellular component organization, metabolic process, biological regulation. Besides the DMGs between chimpanzee and macaque in promoter are enriched in death while those in gene body are enriched in processes associated with localization. KEGG pathway enrichment results show that differentially methylated species in promoter are associated with some basic life processes, such as Oocyte meiosis and cell cycle, whereas DMGs in gene body are involved in some concrete processes, including axon guidance, adherens junction. This may imply DMGs in promoter is helpful to explaining the evolution in primate species while DMGs in gene body may be associated with individual differences.

2.3 Transcription factor enrichment of differential methylation genes between species
TFBS were downloaded from UCSC, then we conduct enrichment of 13 transcription factors using the hypergeometric test, and dominance ratios were defined to reveal the results of enrichment (Figure 3).
Transcription factor enrichment results show differentially methylated genes among three species pairs are enriched to SIN3A. Previous studies have shown that, SIN3A is a transcriptional repressor and tends to interact with MeCP2, HDAC1, N-coR and so on to play a role in inhibition. Besides, inhibition in different environments are important for mammalian development (McDonel et al., 2009). Interaction between Sin3A and MeCP2 would cause transcription inhibition which may provide evidence for the negative regulation of DNA methylation to gene expression (Nan et al., 1998). TAF1 is significantly enriched on the DMGs between human and chimpanzee, human and macaque, but there is no significant accumulation in DMGs between chimpanzee and macaque. This reflects the TAF1’s specific binding in human. TAF1 is part of the TFIID multiprotein complex. TFIID includes the TATA binding protein (TBP) and 12 TBP associated factors (TAFs). TAF1 is known as the largest one of TAFs. Studies have shown that mutations in TAF1 are associated with some new symptoms of severe mental retardation (O’Rawe et al., 2015). DMGs between macaque and human, macaque and chimpanzee are significantly enriched on EZH2 transcription factors, but DMGs between human and chimpanzee is not enriched on EZH2. Studies have shown that DNA methylation in the promoter region of EZH2 target gene can't miss EZH2. Mutation of EZH2 is associated with a rare cognitive disorders --Weaver syndrome (Gibson et al., 2012; Li et al., 2013). Besides, EZH2 relates to the occurrence of neurological diseases. NRSF tends to bind to DMGs between human and chimpanzee. Studies show that NRSF plays a role in the formation and plasticity of neuron, and is associated with Huntington, Alzheimer and other neurological diseases (Lu et al., 2014).

2.4 DNA methylation and gene expression in the DMG between human and chimpanzee

Gene expression data were downloaded from GEO with accession no GSE33587, which included 6 human samples and 6 chimpanzee samples. Differential methylation genes between human and chimpanzee were used as DNA methylation data. There were 877 genes in promoter region and 4987 genes in gene body region, but only 720/877 genes in promoter and 3911/4987 genes in gene body had both DNA methylation data and gene expression data which could be used to investigate the relation between methylation and expression.

Levels of gene expression were negative correlated with DNA methylation in promoter region as well as gene body region (Figure 4A; 4B). For example, Among 42 genes whose methylation levels were greater in human than chimpanzee, 17 genes were greater expression in human. Our conclusion is same to previous studies that there is a weak negative correlation between gene expression and DNA methylation in prefrontal cortex. And it is consistent with DNA methylation’s role in transcriptional silencing.

There were 877 genes in promoter region and 4987 genes in gene body region, but only 720/877 genes in promoter and 3911/4987 genes in gene body had both DNA methylation data and gene expression data which could be used to investigate the relation between methylation and expression.

Levels of gene expression were negative correlated with DNA methylation in promoter region as well as gene body region (Figure 4A; 4B). For example, Among 42 genes whose methylation levels were greater in human than chimpanzee, 17 genes were greater expression in human. Our conclusion is same to previous studies that there is a weak negative correlation between gene expression and DNA methylation in prefrontal cortex. And it is consistent with DNA methylation’s role in transcriptional silencing.

Figure 4 DNA methylation is negatively correlated with gene expression level in both promoters and gene bodies in the DMG between human and chimpanzee. (A) In the promoter region. (B) In the gene body region. X-axis represents the situation of differential methylation between human and chimpanzee, all genes means all of the differential methylation genes in promoter (Figure 4A) or gene body (Figure 4B). Greater human methylation in the right represents those genes’ methylation level in human are higher than that of chimpanzee while greater chimpanzee in the left represents those genes’ methylation level in chimpanzee are higher than that of human.

There were 877 genes in promoter region and 4987 genes in gene body region, but only 720/877 genes in promoter and 3911/4987 genes in gene body had both DNA methylation data and gene expression data which could be used to investigate the relation between methylation and expression.

Levels of gene expression were negative correlated with DNA methylation in promoter region as well as gene body region (Figure 4A; 4B). For example, Among 42 genes whose methylation levels were greater in human than chimpanzee, 17 genes were greater expression in human. Our conclusion is same to previous studies that there is a weak negative correlation between gene expression and DNA methylation in prefrontal cortex. And it is consistent with DNA methylation’s role in transcriptional silencing.

Figure 4 DNA methylation is negatively correlated with gene expression level in both promoters and gene bodies in the DMG between human and chimpanzee. (A) In the promoter region. (B) In the gene body region. X-axis represents the situation of differential methylation between human and chimpanzee, all genes means all of the differential methylation genes in promoter (Figure 4A) or gene body (Figure 4B). Greater human methylation in the right represents those genes’ methylation level in human are higher than that of chimpanzee while greater chimpanzee in the left represents those genes’ methylation level in chimpanzee are higher than that of human.
2.5 Divergence of DNA methylation between human and chimpanzee relative to macaque

The differential methylation genes in promoter and gene body were used to investigate the divergence of DNA methylation between human and chimpanzee relative to macaque.

Investigation was done in promoter and gene body respectively. Among 877 genes in promoter regions, 544 (62%) genes’ methylation in human were specially less than that of chimpanzee while only 36 (4%) genes’ methylation in human were specially greater than that of chimpanzee (Figure 4C). 1453 (33%) genes’ methylation in human were specially less than that of chimpanzee while 412 (9%) genes’ methylation in human were specially greater than that of chimpanzee. Notice more than 50% of differential genes between human and chimpanzee had the same trend relative to macaque in gene body region (Figure 4D). DNA methylation in gene body maybe more stable in the evolution of primate species while that of promoter is more divergent. As the negative relation between DNA methylation and gene expression, the low level of DNA methylation in human might have greater gene expression which had the chance to result in the special phenotype in human, so we utilized the special hypo genes to do the functional enrichment using DAVID. The result was showed that the special hypo genes in human is enriched on the cellular component organization, cellular process and developmental process while we didn’t get any enrichment item when using the special hyper genes in human.

2.6 The Role of DNA Methylation in the Evolution of Human

We divided the orthologous genes between human and chimpanzee into three gene sets: the 525 genes for directional selection in human (set A), the first 525 genes for stabilizing selection (set B) and the remaining genes (set C). dN, dS were downloaded from Ensembl to evaluate the natural selection on coding regions and 12,703 genes were got after intersection of genes with dN/dS and genes with DNA methylation. The statistics of the dN/dS demonstrated that most of the values were between 0 and 1, and only a few were more than 1. Furthermore, density distribution was got of the genes whose dN/dS were between 0 and 1 (Figure 5). As can be seen, the peak of dN/dS were increasing gradually from the directional selection gene set in human to the stabilizing selection gene sets. The smallest value of dN/dS in directional selection genes in human demonstrated that if mutation occurred these genes would be eliminated, so the coding region of these genes were inclined to be conservative. However the DNA methylation were found to be specifically changed in these genes which reflected that the evolution of human might be explained by the modification in epigenetics.

![Figure 5 The distribution of dN/dS in three gene sets](image)

3 Discussion

In this paper, we screen differentially methylated genes by comparing two mixed linear models, in which we consider the influence of individual, CG sites and other random factors. Analysis indicts that more truly differentially methylated genes would be screened by our method than the direct use of average methylation. The dendogram obtained according to the number of differentially methylated genes by mixed linear model is consistent to the generally believed evolution while that got by average methylation is different. This indicates that the average methylation is susceptible to methylation levels, individual differences, CpG sites and so can not truly reflect certain biological processes.
Dendrograms conduct by the number of differentially methylated genes in promoter and gene body between humans, chimpanzees, macaque species pairs, indicates that evolutionary mechanisms of DNA methylation are different in these two regions. And the DNA methylation in promoter region can reflect the history of evolution better. DMGs Between human and chimpanzee in promoter tend to hypomethylate in human, which reflects the rapid evolution of hypomethylation in the promoter, in order to reduce the effect of silencing transcription. Then more gene expression would get which promotes the diversity of life and then is conducive to the evolution of the process.

Enrichment of transcription factor demonstrates that the differentially methylated genes between species have the tendency to be bound by some specific transcription factor, which provides a perspective of transcription factors to analyze species-specific features. In this paper, the concept of ancestor methylation analysis was proposed to analyze the human-specific methylation patterns relative to chimpanzees. And this idea is helpful to identify the underlying mechanisms underlying the human-specific phenotypes.

This article mainly analyzed the DNA methylation data from human, chimpanzee, and macaque. Though we reveal divergent DNA methylation patterns in promoter and gene body, the underlying mechanism remains unclear. We need more studies about these two regions' methylation status in other species. DNA methylation can make some explanations for the evolution of certain species, but only DNA methylation data is still not sufficient to reveal more phenomena so more epigenetic phenomena and factors would be necessary for further analysis.

Acknowledge
We thank the support of the Natural Scientific Research Fund of Heilongjiang Provincial (grant number QC2011C061), Natural Science Foundation of China (grant number 61402139), the Health Department Project of Heilongjiang Provincial (grant number 2012-799), and Students Research and Training Program of Heilongjiang Province (grant number 201410226023).

References
Bates, Douglas, Martin Maechler, and Ben Bolker, 2012, lme4: Linear mixed-effects models using S4 classes
Cáceres, Mario, et al., 2003, Elevated gene expression levels distinguish human from non-human primate brains, Proceedings of the National Academy of Sciences,100: 13030-13035
Chan, Esther T., et al., 2009, Conservation of core gene expression in vertebrate tissues, J. Biol., 8: 33
Dennis Jr., Glynn, et al., 2003, DAVID: database for annotation, visualization, and integrated discovery, Genome Biol.,4: P3
Gibson, William T.; et al., 2012, Mutations in EZH2 cause Weaver syndrome, The American Journal of Human Genetics, 90: 110-118
Hernando-Herraez, Irene, et al., 2013, Dynamics of DNA methylation in recent human and great ape evolution, PLoS Genetics, 9: e1003763
King, Mary-Claire, and Allan C Wilson, 1975, Evolution at two levels in humans and chimpanzees, Science, 188: 107-116
http://dx.doi.org/10.1016/j.biocel.2008.07.022
http://dx.doi.org/10.1038/30764
O’Rawe, Jason, et al., 2015, A variant in TAF1 is associated with a new syndrome with severe intellectual disability and characteristic dysmorphic features, BioRxiv, 014050
Pique-Regi, Roger, et al., 2011, Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data, Genome Research, 21: 447-455
Tost, Jörg, 2008, Epigenetics: Horizon Scientific Press
http://dx.doi.org/10.1073/pnas.1202734109
http://dx.doi.org/10.3758/BF03196482
http://dx.doi.org/10.1016/j.ajhg.2012.07.024