論文の内容の要旨

Computational and experimental analysis of evolutionary changes at Hominidae and Hominoidea specific coding and conserved noncoding genomic elements

（日本語:ヒト科とヒト上科特有のコード・非コードゲノム要素の進化的変化に関するコンピュータ解析および実験的解析）

Mahmoudisaber Morteza（マフモウディサベール モルテザ）

Superfamily Hominoidea, the name of which originates from the Latin word meaning 'Human-like', includes humans and apes and is one of the two living superfamilies of Catarrhini parvorder. Taxonomically this superfamily belongs to the order of primates and is comprised of two families of Hominidae (humans and great apes) and Hylobatidae (lesser apes) (Fig. 1). All members of this superfamily have large brains, well known for their complex social behavior and intellectual abilities.

Despite the increasing genome data, the genetic factors underlying the phenotypic uniqueness of Hominoidea have remained elusive. A phenotype is the result of a collective network of genes along with other regulatory elements. Clade-specific genes and highly conserved noncoding sequences (HCNSs) are among the high-potential evolutionary candidates involved in driving clade-specific characters and phenotypes which are comprehensively investigated in the superfamily Hominoidea and family Hominidae in this study.

1) Highly Conserved Noncoding Sequences (HCNSs):

For identification of HCNSs, first the neutral evolution rates were set by comparing noncoding genomic conservation and coding synonymous site conservation level in Hominidae and Hominoidea along with closest outgroup species. Using the determined neutral evolution thresholds 1,658 and 679 HCNSs were identified with 100 percent sequence similarity in Hominidae and Hominoidea, respectively, which have no orthologs in outgroup species with conservation above neutral threshold.

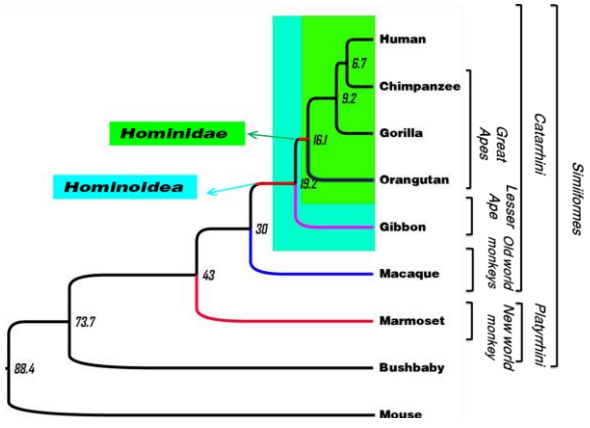

Fig. 1. The phylogenetic and taxonomic classification of Simians.

Genetic variation is a suitable indicator of selective constraint on a sequence and indicates functionality. HCNSs do have significantly lower overlap with genomic polymorphisms compared to random coordinates and their flanking regions indicating the functionality of these conserved elements. Derived allele frequency (DAF) analysis of HCNSs, further confirmed that the lower mutation rate within these elements is not due to them being located on mutation cold spots, but rather due to the evolutionary constraint that prevents spread of mutations occurring within these elements in human populations.

To figure out the evolutionary forces leading to the formation of HCNSs, the properties of orthologs of HCNSs in outgroup species were investigated. With assumption of molecular clock, it has been shown that substitution rates in Hominidae and Hominoidea ancestral branches are respectively 5 and 2.3 times higher than that under neutral evolution (Fig. 2), suggesting that these elements may have emerged through some kind of positive selection, and then purifying selection started to operate to keep the functions served by these elements.

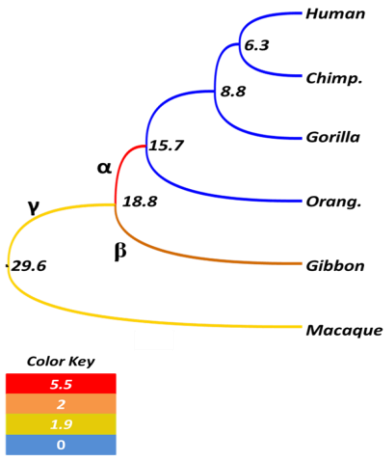

Fig. 2. HCNS ancestral sequences' nucleotide substitution rate across catarrhini (From Saber et al. 2016)

Having established that the HCNSs are under purifying selection, genomic distribution of HCNSs were investigated. HCNSs are significantly overrepresented in proximity of protein-coding genes' transcription start sites, especially at distance of 5-50 kb at both upstream and downstream directions and underrepresented at farther distances. This pattern of distribution is similar to that of silencer elements and contradictory to enhancer elements distribution which tend to be located far from protein coding genes transcription start sites.

HCNSs tend to cluster around genes involved in nervous system, transcription regulation and development. HCNS target genes were also shown to have significantly higher non-coding portion compared to the total human protein coding genes suggesting the action of evolutionary forces to prevent loss of conserved noncoding regions in these genes. HCNS target genes were also shown to be located in isolation with higher absolute distance to the next protein coding genes compared to average human genome protein coding genes.

Ancestral CNSs in primates have been shown to be playing role within their genomes as expression-activator enhancer elements (Babarinde and Saitou, 2016). Analyzing the production of enhancer RNA (eRNA) and the overlap with H3K4me1 and H3K27ac epigenomic marker, however, revealed depletion of enhancer markers in young Hominoidea and Hominidae HCNSs that have evolved less than 30 million years ago. Expression analysis of HCNS target genes within human tissues also suggest inhibitory effect of HCNSs on their target genes, especially in fetal brain tissue, the tissue in which HCNSs are supposed to be in their most active form considering Gene Ontology data.

2) De novo protein coding genes:

For identification of coding genes restricted to superfamily Hominoidea and family Hominidae, major DNA, protein and gene orthology databases were analyzed. Setting strict thresholds in order to avoid false positives, only one protein coding gene named Down Syndrome Critical Region 4 (DSCR4) showed strong evidence of being bona fide Hominidae specific gene. DSCR4 has expression mainly in human placenta and is located on a medically important region called Down syndrome critical region which has long been proposed to be involved in higher brain functions.

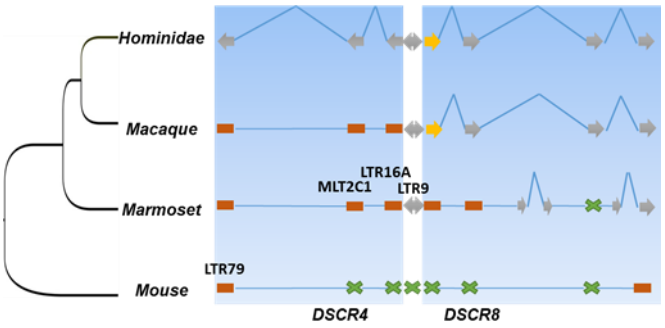

Fig. 3. Evolutionary origin of DSCR4 gene (From Saber et al. 2016)

Evolution of DSCR4 can be mainly classified in three evolutionary periods. Period (1) LTR79 retrotransposition, that took place in the common ancestor of mammals >100 Mya. This transposition formed DSCR4's exon 3 ancestral sequences (Fig. 3). Period (2) During the evolution of common ancestor of primates

at 29– 45 Mya, three independent retrotranspositions by MLT2C1, LTR16A and LTR9 led to the formation of DSCR4's exon 2, exon 1 along with DSCR4/8 shared bidirectional promoter (Fig. 3) (3) The final required mutation was a GC transversion that formed the stop codon and completed the formation of DSCR4 ORF.

There are numerous computational evidences at RNA and protein level indicating the functionality of DSCR4 gene, however, due to the absence of known protein domains within the DSCR4-coded protein, the function of DSCR4 within the cell could not be determined without experimental verifications.

3) Functional analysis of DSCR4:

For experimental verification of the functionality of DSCR4, I conducted DSCR4 gene perturbation analysis followed by transcriptome profiling in primary human bone marrow cells which provided strong evidence for the involvement of this gene in regulation of cell migration, motility and locomotion. This hypothesis, is further supported by tissue-specific expression of DSCR4 in human cells in which migration is important for proper functioning (Fig. 4).

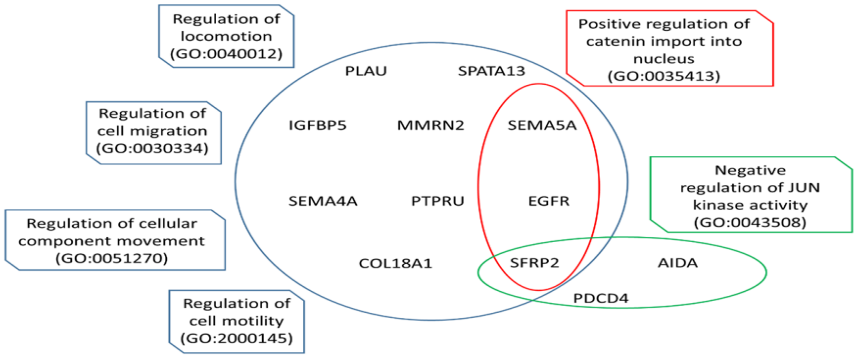

Fig. 4. Biological processes significantly affected by over-expression of DSCR4 in human HS27a cells