Chinese Science Bulletin

Detecting positive darwinian selection in brain-expressed genes during human evolution

QI XueBin[1,2*], YANG Su[1,2,5*], ZHENG HongKun[3,7*], WANG YinQiu[1,2,5*], LIAO ChengHong[1,2,5], LIU Ying[1,2,5], CHEN XiaoHua[1,2], SHI Hong[1,2,5], YU XiaoJing[1,2,5], Alice A. LIN[4], Luca L. CAVALLI-SFORZA[4], WANG Jun[3,6,7†] & SU Bing[1,2†]

[1] Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China;
[2] Kunming Primate Research Center, Chinese Academy of Sciences, Kunming 650223, China;
[3] Beijing Institute of Genomics of Chinese Academy of Sciences, Beijing Genomics Institute, Beijing Proteomics Institute, Beijing 101300, China;
[4] Department of Genetics, Stanford University, Stanford, USA;
[5] Graduate School of the Chinese Academy of Sciences, Beijing 100039, China;
[6] The Institute of Human Genetics, University of Aarhus, Aarhus C, DK-8000, Denmark;
[7] Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, DK-5230, Denmark

To understand the genetic basis that underlies the phenotypic divergence between human and non-human primates, we screened a total of 7176 protein-coding genes expressed in the human brain and compared them with the chimpanzee orthologs to identify genes that show evidence of rapid evolution in the human lineage. Our results showed that the nonsynonymous/synonymous substitution (K_a/K_s) ratio for genes expressed in the brain of human and chimpanzee is 0.3854, suggesting that the brain-expressed genes are under functional constraint. The X-linked human brain-expressed genes evolved more rapidly than autosomal ones. We further dissected the molecular evolutionary patterns of 34 candidate genes by sequencing representative primate species to identify lineage-specific adaptive evolution. Fifteen out of the 34 candidate genes showed evidence of positive Darwinian selection in human and/or chimpanzee lineages. These genes are predicted to play diverse functional roles in embryonic development, spermatogenesis and male fertility, signal transduction, sensory nociception, and neural function. This study together with others demonstrated the usefulness and power of phylogenetic comparison of multiple closely related species in detecting lineage-specific adaptive evolution, and the identification of the positively selected brain-expressed genes may add new knowledge to the understanding of molecular mechanism of human origin.

positive selection, adaptive evolution, brain-expressed gene, hominoids

In spite of only ~1.23% difference in genomic DNA sequence[1], human and our closest living relative, chimpanzee, differ considerably in many biological characteristics, especially the enlarged brain and superior cognitive ability in humans[2]. This observation leads to a long-lasting question of what makes human unique.

The recent study has shown that the genes involved in diverse functions of nervous system exhibit significantly

Received June 27, 2006; accepted November 21, 2006
doi: 10.1007/s11434-007-0062-y
*Contributed equally to this study
†Corresponding authors (email: sub@mail.kiz.ac.cn and wangj@genomics.org.cn)
Supported by the Chinese Academy of Sciences (Grant Nos. KSCX2-SW-121 and GJHZ0518), the National Natural Science Foundation of China (Grant Nos. 30370755, 30440018, 30525028, 90208019, 90403130 and 30221004), the National 973 Project of China (Grant No. 2006CB701506), the Post-doctoral Fellowship of China (Grant No. 2005038271), Ministry of Science and Technology under program (Grant No. CNGI-04-15-7A), the Natural Science Foundation of Yunnan Province of China, China National Grid, and Danish Platform for Integrative Biology and Ole Rømer grants from the Danish Natural Science Research Council

accelerated evolution in primates than in rodents, and within primates, this acceleration of protein evolution is most pronounced in the lineage leading to human[3]. The brain also shows more gene expression changes than other tissues in the human lineage compared to the chimpanzee lineage[4,5]. Therefore, it is reasonable to speculate that the accelerated evolution of protein-coding genes and gene expression in the human lineage may play important roles in improvement of cognitive abilities in human (e.g. refs. [6−11]).

To understand the genetic basis underlying the phenotypic traits that set human apart from nonhuman primates, an important approach is to study the candidate genes that underwent positive Darwinian selection during human evolution, because these genes likely contribute to the human-specific features such as the reproductive system and cognition[12]. A number of genes that have shown evidence of positive Darwinian selection during human and/or primate evolution have been identified in the past decade[13], and some of which are known to be important in development and function of the human brain[12]. For example, the gene *ASPM* (abnormal spindle-like microcephaly associated) and *MCPH1* (microcephalin) are two genes that regulate brain size and have evolved under strong positive selection during human and great ape evolution[8−11,14,15], and continued to evolve adaptively in anatomically modern humans, suggesting that the human brain is still under rapid adaptive evolution[16,17].

With the availability of the chimpanzee draft genome sequence[1], it is now possible to perform a genome-wide sequence comparison of human and chimpanzee in order to identify candidate genes that are potentially responsible for human-specific characteristics. The recent genome-wide comparisons of protein-coding genes between human and chimpanzee suggested that 2.8%−4.4% of human genes are likely under positive selection[1,18]. Although the nervous system genes showed an accelerated evolution during human evolution[3,19], genes with maximal expression in the brain showed little or no evidence for positive selection[20]. Recently, a pairwise human-chimpanzee selective pressure (K_a/K_s) screen of primate seminal fluid proteins has successfully identified previously unidentified positive selection in seven primate seminal proteins[21]. In present study, we screened 7176 protein-coding genes expressed in the human brain and compared them with their chimpanzee orthologs to identify genes that show evidence of adaptive evolution in the human lineage.

1 Material and methods

1.1 DNA sequencing and human-chimpanzee alignment

A total of 10379 brain-expressed genes were identified by searching the *NCBI EST* database (http://www.ncbi.nlm.nih.gov/) and the published human brain expression data[22−25], of which, 2172 genes were also analyzed in ref. [19] and 4764 genes in ref. [20]. As both human adult and fetal brain samples were used in the published data, we believe that most of the genes involved in development and function of the human brain are included in our dataset. The putative chimpanzee gene sequences were obtained from the November 13th 2003 draft genome assembly from the Chimpanzee Sequencing Consortium (http://www.ensembl.org/Pan_troglodytes/), which contains 361782 contigs covering about 90% of the chimpanzee genome.

The gene structure information was extracted from the NCBI (http://www.ncbi.nlm.nih.gov/) and Ensembl (http://www.ensembl.org/) databases, and the phases of the exons were defined for each gene. To avoid false hits, a 500-bp intron sequence at both sides of the exon was retained, and the intron sequences longer than 1 kb were masked for more efficient search of homologues in the chimpanzee genome. We mapped the human exons to the chimpanzee assembled contigs using SIM4, a computer program for aligning cDNA sequences with genomic DNA sequences[26]. According to the exon phase, the aligned sequences of human and chimpanzee were trimmed with only coding sequences retained, and then subject to calculation of synonymous (silent) and nonsynonymous (replacement) substitution rates (K_s and K_a) using Pamilo-Bianchi-Li's method[27,28], in which the transition/transversion bias was considered. After eliminating 3203 genes with sequence length of less than 200 bp, a total of 7176 genes were found to have homologues in the chimpanzee genome (details for all genes are available on request). Of the 7176 orthologous gene pairs analyzed, 64163 nucleotide substitutions were detected between 6762 aligned gene pairs, and 2118 indels between 479 gene pairs.

1.2 The expression profiles of genes

The expression profiles of genes were obtained in GNF

SymAtlas v1.1.1[24]. The expression values used here were based on data generated from applying the GC content adjusted-robust multi-array (gcRMA) condensed algorithm, which computes expression values from probe intensity values incorporating probe sequence information (more information is available at http://www.bepress.com/jhubiostat/paper1.2004). We took the average expression value over the probe sets for those genes with more than one probe set, and used an expression value of 200 as the threshold for calling a gene "expressed in a given tissue"[24]. We simply compared the expression values across 79 tissue types to find the tissue of maximal expression for each gene.

1.3 Functional category classification

The functional categories of genes were classified in a PANTHER (Protein ANalysis THrough Evolutionary Relationships) classification system version 5.0[29], which contains over 6683 protein families, divided into 31705 functionally distinct protein subfamilies. The functional categories of genes were classified according to the function of the protein by itself or with directly interacting proteins at a biochemical level (molecular function), and the function of the protein in the context of a larger network of proteins that interact to accomplish a process at the level of the cell or organism (biological process).

1.4 Maximum likelihood ratio test

We used likelihood ratio tests to determine whether any specific lineages or codon positions were associated with K_a/K_s ratios significantly great than 1.0 and hence possibly subject to positive Darwinian selection, and the analyses were performed using CODEML program in PAML 3.14[30]. Two types of likelihood ratio tests were performed: (1) Lineage-specific models in which all codon sites are assumed to be under the same selective pressure but the selective pressure can vary among different lineages. The analysis consists of two likelihood ratio tests: the first test compares one-ratio model, which assumes an equal ω ratio for all branches in the phylogeny, with free-ratio model, which assumes an independent ω ratio for each branch estimated from the data. Because one-ratio model involves one parameter ω whereas free-ratio model involves N ωs, where N is the number of branches, twice the log-likelihood difference ($2\Delta l$) between the two models is compared with a χ^2 distribution with d.f.=N–1. This is a test for heterogeneity of ω among branches, but when estimates of ω under free-ratio model are >1, positive selection is implicated. The second compares one-ratio model with two-ratio model, which assumes two different ω ratios for specified branches and the rest of branches, because two-ratio model involves two ωs, twice the log-likelihood difference is compared with χ2 distribution with d.f.=1 to determine whether ω on the specified branches are significantly >1. (2) Site-specific models in which selective pressure varies among different sites but the site-specific pattern is identical across all lineages. Six different site-specific models were implemented: model M0 (null model) assumes one ω ratio among all sites, M1 (neutral model) assumes two categories of sites: the conserved sites (ω=0) and the neutral sites (ω=1), M2 (selection model) adds an extra category of site allowing for sites with estimated ω>1, M3 (discrete model) assumes a general discrete distribution with three categories of site with ω ratio free to vary for each site, M7 (beta model) assumes a beta distribution with eight categories of site with ω ratios ranging in the interval (0, 1) depending on parameters p and q, M8 (beta and ω model) adds an extra category of site to the beta model (M7), with the proportion and the estimated ω ratio allowing for sites with ω>1. From these models, three likelihood ratio tests were performed by comparing twice the log-likelihood difference between models M0 and M3 (d.f.=4), between M1 and M2 (d.f.=2), and between M7 and M8 (d.f.=2) with χ^2 distribution with the appropriate degrees of freedom to determine whether particular models provided a significantly better fit to the data[31].

2 Results and discussion

In this study, we compared 7176 protein-coding genes expressed in the human brain with their chimpanzee orthologs to identify genes showing evidence of positive Darwinian selection in the human and/or chimpanzee lineages. Those genes that are potentially under positive Darwinian selection can be determined using the ratios of the number of nonsynonymous substitutions per nonsynonymous site (K_a) and the number of synonymous substitutions per synonymous site (K_s). The K_a/K_s>1 is generally taken as evidence for adaptive or positive Darwinian selection, whereas K_a/K_s=1 suggest neutral evolution, and K_a/K_s<1 indicate negative or purifying selection[32]. The K_a, K_s and other related values were computed for each gene by the method of Pa-

milo-Bianchi-Li[27,28] in MEGA3.0[33], and the details for all genes are available on request. The K_a/K_s ratio distribution of the 7176 brain-expressed genes is illustrated in Figure 1. A similar distribution was obtained when using the YN00 method[34] in PAML 3.14[30] (data not shown).

Of the 7176 genes studied, 849 (11.8%) genes have K_a/K_s ratios of ∞ ($K_a>0$ and $K_s=0$) or >1. These genes may be enriched for genes that show adaptive evolution in the human and/or chimpanzee lineages, and were subject to further analysis. The proportion (11.8%) of the putatively positively selected genes in our dataset is roughly the half of that (20.2%) calculated from 7645 human-chimpanzee-mouse orthologs[19], while it is slightly higher than the proportion (9.1%—10.5%) obtained from a genome-wide collection of >13000 genes[1,20]. The majority (72.5%) of the brain-expressed genes studied have a K_a/K_s ratio of zero or less than 0.5, suggesting that the majority of genes expressed in the brain of human and chimpanzee tend to be under strong purifying selection or functional constraint[35]. When those genes with a K_a/K_s ratio of ∞ (the 414 genes with $K_a=0$, $K_s=0$ and 359 genes with $K_a>0$, $K_s=0$) were excluded from the calculation, the remaining genes yielded an average K_a/K_s ratio of 0.3854. Our estimation of K_a/K_s ratio for the brain-expressed genes is higher than the value (0.23) estimated from a genome-wide collection of 13454 human-chimpanzee orthologous genes[1] although such simple comparison based on different algorithm and evolutionary model may introduce bias, implying that the human brain-expressed genes and their chimpanzee orthologs have diverged rapidly since their common ancestor, which is consistent with previous findings that primate nervous system genes underwent accelerated evolution during primate evolution[3,19].

2.1 Positive selection in the human brain

Human evolution is characterized by a dramatic increase in brain size and functional complexity since we shared common ancestor with chimpanzees 5—6 million years ago[2]. Therefore, it is reasonable to speculate that some of the genes expressed in the brain of human and chimpanzee might have undergone adaptive evolution in the human lineage that resulted in the superior cognitive ability in humans. To test whether human brain displays a significant excess of positively selected genes comparing with other tissues, we compared the K_a/K_s ratio distribution in genes with their maximal expression in different human tissues. Of the 7176 genes studied, 5973 of them had expression profile in GNF SymAtlas v1.1.1[24]. A total of 79 tissue types covering the major physiological systems were considered in the analysis.

A total of 282, 100 and 126 genes are found to have their maximal expression in brain tissues of prefrontal cortex, whole brain and fetal brain, respectively. A significant excess of the putatively positively selected genes were not seen in these genes that highly expressed in the human brain. Instead, we found a significant deficiency of the putatively positively selected genes in prefrontal cortex (one-tailed Fisher's exact test $P<0.05$) and whole brain (one-tailed Fisher's exact test $P<0.01$). These findings are consistent with previous results[20], suggesting that the majority of genes that have their maximal expression in brains are under strong functional constraint and the distinct brain functional divergence

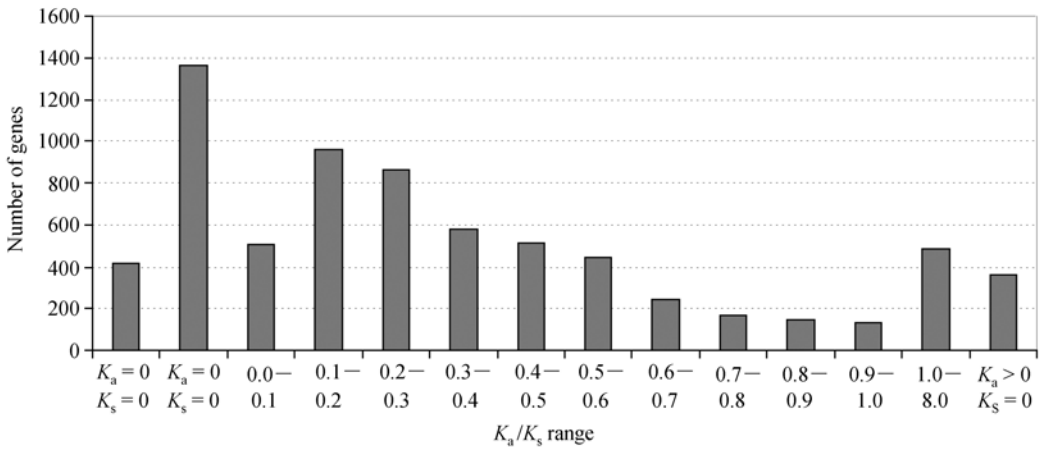

Figure 1 K_a/K_s distribution in 7176 human-chimpanzee orthologous genes pairs. The K_a/K_s distribution was not statistically different between methods of Pamilo-Bianchi-Li[27,28] and YN00[34] (Mann-Whitney U test, $P=0.8175$).

between human and chimpanzee is unlikely caused by a large-scale adaptive evolution of coding regions of brain-expressed genes. An increased gene expression and accelerated adaptive changes in a small fraction of genes involved in diverse functions may have caused the extensive modifications of brain physiology and functions underlying the superior cognitive powers in humans[4,12,22,36−39].

In contrast with brain tissues, testis-specific genes are found to have been evolving under positive selection in multiple organisms during the course of evolution[4,20,40−42], and may have contributed to the process of speciation. In this study, a total of 367 genes are found to have their maximal expression in five types of testis tissue: testis (105 genes), testis germ cell (76 genes), testis interstitial (166 genes), testis leydig cell (13 genes), and testis seminiferous tubule (7 genes), while the putatively positively selected genes are only significantly overrepresented in testis interstitial (one-tailed Fisher's exact test P=0.0025), suggesting that genes expressed in different testis tissues are under differential selective pressures.

In addition, we also observed a significant excess of putatively positively selected genes in liver (one-tailed Fisher's exact test P=0.0319) and pancreatic islets (one-tailed Fisher's exact test P=0.0012), this may also be the result of relaxation of selective constraints for genes maximally expressed in liver and pancreatic islets, as seen in gene expression profiles[4].

2.2 Positive selection among functional categories of genes

To identify functional categories of genes that are potentially targeted by positive Darwinian selection, we applied a Panther classification system (ver. 5.0)[29] to partition the 7176 brain-expressed genes into 51 molecular function (MF) categories (Supplementary Table S1 shown in online version only) and 50 biological process (BP) categories (Supplementary Table S2 shown in online version only). The number of putatively positively selected genes are significantly different among MF categories (χ^2=163.84; d.f.=50; $P<4.99\times10^{-13}$) and among BP categories (χ^2=132.33; d.f.=49; $P<1.36\times10^{-8}$). Further analyses identify a significant excess of putatively positively selected genes in three MF categories (Figure 2) and two BP categories (Figure 3). In particular, the putatively positively selected genes are significantly overrepresented in MF categories of defense/immunity protein (one-tailed Fisher's exact test $P<0.01$) and signaling molecules (one-tailed Fisher's exact test $P<0.05$), and in BP category of immunity and defense (one-tailed Fisher's exact test $P<0.01$). The defense/immunity related genes showed a significantly elevated K_a/K_s ratio when compared with the rest of genes (Mann-Whitney U-test $P<10^{-8}$ for MF defense/immunity protein; $P<10^{-9}$ for BP immunity and defense; and $P<0.05$ for BP T-cell mediated immunity). These observations are consistent with previous findings using gene collections that randomly selected in the genome[18,20] and other case studies (e.g. *IgA*[43], *APOBEC3G*[44−46], *TRIM5α*[47−49], and *β-defensin*[50]). This overrepresented positively selected genes and elevated K_a/K_s ratios in immunity/defense related genes may suggest that the human immunity/defense related genes have undergone rapid adaptive evolution in defending the pathogen attacks[51] or relaxation of functional constraint due to a low or lack of pathogens and viral challenge allowing the deleterious mutations to persist[49].

Using mouse genes as outgroup, Clark et al.[19] have demonstrated that genes involving in a variety of functional categories have undergone the accelerated evolution in human lineage and/or in chimpanzee lineage. Similarly, Nielsen et al.[20] identified 12 BP categories (including genes whose biological processes are unclassified) showing an excess of putatively positively selected genes between humans and chimpanzees. However, our comparisons showed that only those human brain-expressed genes involving in defense/immunity and signaling transduction are likely under positive selection during human evolution.

In addition, the putatively positively selected genes are significantly overrepresented in a large collection of genes with their molecular functions and biological processes unclassified, which call for further studies. Some of these genes may have played important role in forming human cognition. The annotation of these function-unknown genes in the future would contribute to a better understanding of human origin.

2.3 Positive selection on the X chromosome

The previous studies demonstrated that the effects of positive selection might be more pronounced on the X chromosome than on the autosomes in human[20,52,53]. To test whether the action of positive Darwinian selection

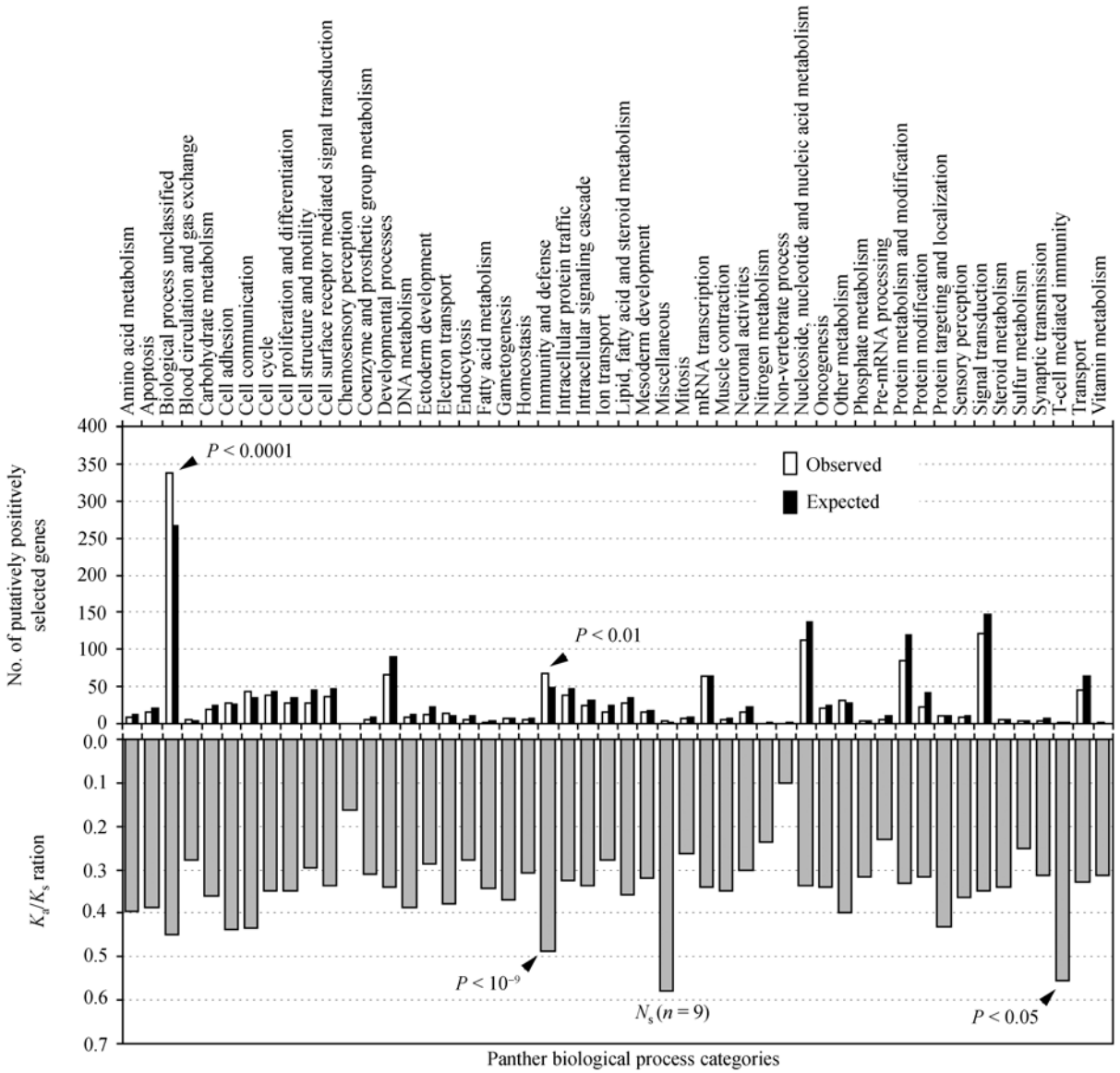

Figure 2 Positive selection and mean K_a/K_s ratios in genes involved in the inferred biological process categories.

on the human brain-expressed genes was also different among chromosomes, we examined the distribution of the putatively positively selected genes and the K_a/K_s ratio for each chromosome, and the results were presented in Supplementary Table S3 (shown in online version only). When the X chromosome was compared with the sum of all autosomes, the frequency of the putatively positively selected genes was significantly higher on the X chromosome (19.6%) than on the autosomes (8.2%) (one-tailed Fisher's exact test $P<0.001$). A similar result was reached when all single chromosomes were compared ($\chi^2=34.3$; d.f.=22; $P<0.05$), which was caused by the excess of positively selected genes on the X chromosome, but no difference was observed between the autosomal chromosomes ($\chi^2=19.2$; d.f.=21; $P>0.05$).

Similarly, the X-chromosome-linked genes exhibited a significantly elevated K_a/K_s ratio compared with all autosomes (0.5623 vs. 0.3972, Mann-Whitney *U*-test $P<0.01$) which is consistent with previous result[53], while there is no significant difference between autosomes (Kruskal-Wallis *H*-test $P=0.67$). The further dissection of the molecular function categories of these positively selected genes reveals that the X chromosome, in comparison with the autosomes, harbors a substantial excess of the putatively positively selected genes with unknown molecular functions (one-tailed Fisher's exact test $P<0.01$), which may play diverse functional roles in human-specific traits because X chromosome is found to

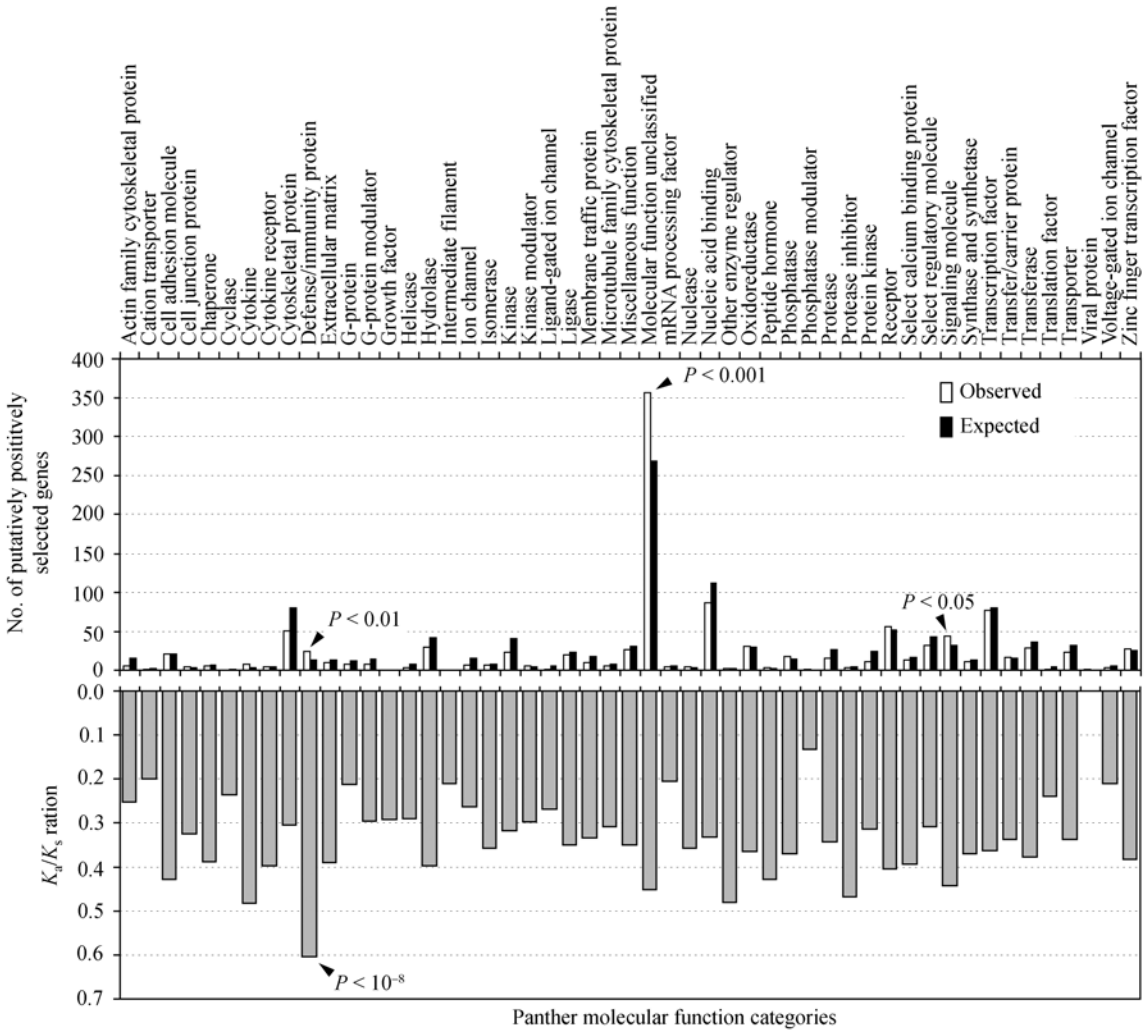

Figure 3 Positive selection and mean K_a/K_s ratios in genes involved in the inferred molecular function categories.

enrich genes involved in male reproduction[54−56] and genes involved in cognitive ability[57]. The male reproductive proteins generally evolve exceptionally rapidly[41,42,58,59], and the positive Darwinian selection is often the driving force behind this rapid evolution[41,60]. These X-linked rapidly evolved genes may involve in sex-specific functional specialization of primate reproductive systems and cognition, and further studies of these positively selected genes will shed light on the evolution of primate reproductive system and human cognition.

In addition, three Y-chromosome-linked genes (*TSPY1*, *USP9Y* and *NLGN4Y*) are found to have their expression in human brain, and two (*TSPY1* and *USP9Y*) of which showed $K_a/K_s>1$ (this study and ref. [61]). They function in spermatogenesis and testicular tumorigenesis, and mutations in these genes can cause spermatogenic failure and male infertility[62,63]. The accelerated evolution of *TSPY1* and *USP9Y* are likely a result of relaxation of selective constraint and/or positive selection as also seen in other Y-chromosome-linked genes[61].

2.4 Analysis of positive selection in 34 candidate genes

The pairwise comparisons of genes expressed in human brains with their chimpanzee orthologs were performed to predict candidate genes under positive selection, and 34 genes were chosen for further analysis with criteria that K_a/K_s ratio significantly >1 (Z-test, $P<0.05$) between human and chimpanzee (Supplementary Table S4 shown in online version only). The maximum likelihood ratio tests, which account for variable selective pressures among sites or branches by allowing different categories of sites or branches in a gene having different ω (K_a/K_s)

ratios[32], have been proved to be useful in detecting positive selection affecting protein evolution[60,64]. In this study, we used likelihood ratio tests to determine whether any lineages or codon positions were associated with K_a/K_s ratios significantly >1 and hence possibly subject to positive Darwinian selection.

In order to conduct the likelihood ratio test, we sequenced other two primate species (white-browed gibbon, *Hylobates hoolock* and rhesus monkey, *Macaca mulatta*) as outgroup(s), which diverged from humans about 18 and 20—25 million years ago respectively[65]. The gibbon and macaque will serve as better outgroup to infer ancestral sequences for human and chimpanzee than rodents, because primates and rodents shared their last common ancestor about 80 million years ago and genes may diverged enormously between primates and rodents[66]. Therefore, the ancestral sequence will be not reliable if rodents are used as outgroup for human and nonhuman primates. We also sequenced 15 human samples (30 chromosomes) from the major continents in order to detect positive selection in the human lineage. Of the 34 candidate genes analyzed, 10 showed evidence of positive selection in the human lineage, nine in the chimpanzee lineage, and four in both lineages when gibbon and/or rhesus macaque used as outgroup (Table 1). Except for four candidate genes (*C3orf18*, *C14orf131*, *MGC33309*, *MGC35118*) which molecular functions are unknown, 11 genes are predicted to play diverse functional roles in embryonic development including brain development (*SCML1* and *RFPL3*), spermatogenesis and male fertility (*GMCL1L*, *SP4* and *SPATA16*), signal transduction (*CXCR1* and *ADCYAP1*), nociception (*MRGX2*), and neurodegenerative disorders (*TKTL1*, *Cx40.1* and *FLJ35808*).

It should be noted that although some of these candidate genes failed to show positive selection, genes with $K_a/K_s<1.0$ might still contain sites under positive selection. This is because the K_a/K_s analyses are not sensitive in identifying specific domains under positive selection, especially when the rest of the gene is subject to purifying selection. Hence, the K_a/K_s analysis is conservative in identifying positively selected genes.

SCML1 and *RFPL3* showed significant K_a/K_s elevation in the chimpanzee lineage ($P<0.05$) and may played important roles in the early embryonic development during primate evolution. For example, *SCML1* is a repressor of expression of Hox genes in mammal and play an important role in the control of embryonal development[67]. This gene showed a pattern of repeated selective sweeps driving divergence between species while eliminating variation within species, and a candidate gene for explaining developmental differences between humans and chimpanzees[20]. *RFPL3* shares close homology with a family of human Ret Finger Protein-Like genes (*RFPL1*, *RFPL2* and *RFPL4*) that contain RING finger-like domains[68]. A majority of RING finger-containing proteins have been shown to function as E3 ubiquitin protein ligases. The expression and protein-protein interaction studies suggested that *RFPL4* functions as an E3 ubiquitin protein ligase to mediate protein degradation pathways important for gametogenesis or early embryonic development[68,69]. Like *RFPL4*, *RFPL3* also functions as ubiquitin-protein ligase and may also play an important role in human early embryonic development.

There are three genes showing dominant expression in testis and involved in spermatogenesis and male fertility. Human germ cell-less homolog 1 (*Drosophila*)-like (*GMCL1L*) gene has 15 fixed differences (of which 13 are nonsynonymous) between humans and chimpanzees (data not shown). The Drosophila ortholog (germ cell-less) of *GMCL1* was shown to be involved in the early events during the formation of pole cells, which are the germ cell precursors in the fly[70]. The mouse ortholog of *GMCL1* is dominantly expressed in testis[71] and has been shown to function as a transcriptional repressor[72]. Transcription factor *SP4* is homologous to the *Drosophila buttonhead* (*btd*) gene and highly expressed in mouse embryos in the developing central nervous system. The studies in mice suggest that *SP4* gene is required for normal male reproductive behavior, and might play an important role in growth, viability and male fertility[73,74]. Similarly, *SPATA16* is expressed exclusively in testis and involved in spermatogenesis and male fertility[75]. In summary, adaptive evolution of *GMCL1L*, *SP4* and *SPATA16* in the human and/or chimpanzee lineages might have contributed to the evolution of human/chimpanzee male reproductive system and possibly the speciation process.

Two genes are involved in signal transduction (*ADCYAP1* and *CXCR1*). The further analyses in divergent primate species pointed out that the K_a/K_s ratios of these two genes are significantly elevated in the primate lineage leading to humans, and this amino acid changes may contribute to functional modification of the pathway of signal transduction in the immune system

Table 1 Likelihood ratio test (LRT) to detect adaptive evolution in 34 genes[a)]

Gene	*N*	Lineage-specific models over all sitesss								Site-specific models over all lineages		
		one ratio vs. free ratio		one ratio vs. two ratios								
				human lineage			chimpanzee lineage					
		d.f.	2Δ*l*	$\omega 0$	ω_H	2Δ*l*	$\omega 0$	ω_C	2Δ*l*	2Δ*l* M3 vs. M0	2Δ*l* M1 vs. M2	2Δ*l* M7 vs. M8
FLJ3187	3	2	0.54	1.23	1.06	0.04	1.11	2.06	0.54	17.32**	16.52**	16.56**
ZNF229	6	8	1.28	∞	∞	1.16	∞	∞	1.28	0	13.18**	13.18**
TXNRD1	3	2	1.70	1.17	∞	0.92	1.127	2.02	0.26	0	0.28	0.28
Hs.548058	4	4	2.92	0.44	∞	1.24	0.44	∞	1.24	1013.14**	1.42	0
MGC3522	3	2	4.90	1.17	∞	2.26	1.34	∞	1.64	2.38	3.32	3.32
GMCL1L	6	8	18.72*	0.53	3.95	5.39*	0.55	2.86	3.28	7.82	1.62	2.02
IL8RA	5	6	9.96	0.44	1.16	1.08	0.41	2.15	3.60	5.62	0	1.44
SP4	5	6	8.94	0.07	0.89	6.66**	0.22	0	0.96	8.8	0	1.66
DEFA3	5	6	3.08	1.36	∞	1.02	1.27	∞	1.64	9.48	9.46**	10.26**
RFPL3	6	8	14.42	0.68	1.39	0.80	0.65	∞	5.28*	7.98	3.2	0
SCML1	4	4	9.62*	1.45	∞	1.08	1.05	6.49	3.88*	0.24	1.84	1.84
LGTN	5	6	10.64	0.32	1.73	2.64	0.32	∞	4.76*	0	0.82	0
TKTL1	5	6	31.34**	0.24	1.46	5.48*	0.23	∞	12.84**	0	2.2	0
dJ222E13.1	7	10	12.32	0.63	∞	3.58	0.65	∞	3.24	27.09**	14.30**	14.48**
C3orf18	4	4	10.32*	0.17	∞	3.96*	0.17	∞	3.94*	0	0.64	0
FLJ20449	4	4	13.56**	0.41	1.89	3.78	0.47	∞	3.00	0.9	0	0.02
FLJ10439	5	6	6.98	0.29	0.74	0.58	0.28	1.09	1.52	0.52	0.18	0
C14orf131	7	10	11.96	0.74	∞	4.38*	0.73	1.97	1.68	3.16	0	1.36
TREML2	4	4	2.04	1.01	1.87	0.60	1.08	∞	0.72	3.62	3.52	3.52
NET-5	5	6	5.90	0.03	0	0.44	0.02	0.17	1.86	0	0.6	0
NRIP2	4	4	4.80	0.70	∞	0.90	0.65	∞	1.84	0	0.18	0
SPATA16	6	8	20.18**	0.65	4.26	4.24*	0.63	∞	8.42**	3.8	2.22	2.34
PEPP-2	6	8	5.38	1.84	2.11	0.00	1.69	4.69	1.12	0.04	6.68*	6.68
MED8	5	6	9.94	0.63	0	0.00	0.55	∞	2.34	2.02	0	1.26
GBP5	4	4	4.36	0.48	1.26	0.70	0.42	2.45	3.28	0.56	0.02	0.16
MRGX2	6	8	12.46	0.65	∞	6.52*	0.73	0.83	0.02	12*	4.78	4.96
MGC33309	4	4	9.40	0.57	0.59	0.00	0.46	∞	6.10*	0	0.9	0
MGC35118	5	6	15.80*	0.80	2.02	0.74	0.61	∞	7.94**	1.64	1	1.06
FLJ30313	4	4	4.86	1.23	∞	1.20	1.19	∞	1.86	0.1	0	0.76
C20orf96	7	10	11.38	0.62	2.49	2.64	0.71	∞	1.32	10.46*	5.46	5.64
CX40.1	4	4	10.62*	0.26	∞	3.96*	0.23	0.89	4.16*	13.88**	2.54	0.64
FLJ34790	5	6	4.56	1.65	∞	2.52	2.21	1.32	0.28	1	4.44	4.44
FLJ35808	6	8	17.56*	0.51	∞	8.40**	0.55	1.08	0.68	1.12	0.64	0
ADCYAP1	6	8	12.54	0.12	0.65	7.90**	0.29	0	1.48	4.40	0.06	0

a) *N*, Number of sequences; d.f., degree of freedom; *, $P<0.05$, **, $P<0.01$.

(*CXCR1*)[76] and the formation of human cognition (*ADCYAP1*)[77] during human evolution. Hence, these two gene cases further demonstrated the power of using multiple closely related species and phylogenetic comparison in detecting lineage-specific adaptive evolution[78].

In addition, *MRGX2*, a G-protein-coupled receptor (GPCR), is expressed specifically in nociceptive neurons of the human peripheral nervous system, and implicated in the modulation of nociception[79]. Like other *MRG* (mass-related gene) genes in which the significant excess of amino acid changes are located in the extracellular domains[79], three of the four human-specific substitutions of *MRGX2* gene were also located in its extracellular domains[80]. As the extracellular domain usually serves as the recognition site for ligand of many GPCRs, the adaptive sequence changes of human *MRGX2* gene may contribute to the functional modifica-

tion of human sensory system, especially the sensory pathway of detecting pain stimuli, which could be a refined self-protection mechanism developed during human evolution[80].

The genes *TKTL1*, *Cx40.1* and *FLJ35808* (*TTL.6*) involved in neurodegenerative disorders or apoptosis[81−83] also showed evidence of positive selection in human and/or chimpanzee lineages, suggesting that these genes might have played roles in shaping up human neural function.

We also identified four candidate genes (*C3orf18*, *C14orf131*, *MGC33309*, *MGC35118*) that showed evidence of positive selection in human and/or chimpanzee lineages. Although the functions of these genes are unknown, these genes may contribute to a better understanding of origin of our own species.

In conclusion, this study demonstrated the usefulness and power of selective pressure (K_a/K_s) screen of orthologous genes and phylogenetic comparison of multiple closely related species in detecting lineage-specific adaptive evolution in the absence of a priori functional knowledge of a gene although this K_a/K_s ratio screen of whole coding region of a gene may leave out those genes that positive selection only acts on a few crucial amino acids in the coding region, or it only occurs in a short time interval during evolution. Furthermore, the identification of these positively selected brain-expressed genes may add new knowledge to the understanding of molecular mechanism of the origin of human cognition.

The authors thank Fan XiaoNa and Yu YiChuan for technical assistance.

1 The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature, 2005, 437: 69−87

2 Carroll S B. Genetics and the making of Homo sapiens. Nature, 2003, 422: 849−857[DOI]

3 Dorus S, Vallender E J, Evans P D, et al. Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. Cell, 2004, 119: 1027−1040[DOI]

4 Khaitovich P, Hellmann I, Enard W, et al. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science, 2005, 309: 1850−1854[DOI]

5 Khaitovich P, Tang K, Franz H, et al. Positive selection on gene expression in the human brain. Curr Biol, 2006, 16: R356−358[DOI]

6 Enard W, Przeworski M, Fisher S E, et al. Molecular evolution of FOXP2, a gene involved in speech and language. Nature, 2002, 418: 869−872[DOI]

7 Ferland R J, Eyaid W, Collura R V, et al. Abnormal cerebellar development and axonal decussation due to mutations in AHI1 in Joubert syndrome. Nat Genet, 2004, 36: 1008−1013[DOI]

8 Kouprina N, Pavlicek A, Mochida G H, et al. Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. PLoS Biol, 2004, 2: E126[DOI]

9 Evans P D, Anderson J R, Vallender E J, et al. Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. Hum Mol Genet, 2004, 13: 489−494[DOI]q

10 Evans P D, Anderson J R, Vallender E J, et al. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. Hum Mol Genet, 2004, 13: 1139−1145[DOI]

11 Wang Y Q, Su B. Molecular evolution of microcephalin, a gene determining human brain size. Hum Mol Genet, 2004, 13: 1131−1137[DOI]

12 Varki A. How to make an ape brain. Nat Genet, 2004, 36: 1034−1036[DOI]

13 Vallender E J, Lahn B T. Positive selection on the human genome. Hum Mol Genet, 2004, 13(Spec. 2): R245−254[DOI]

14 Woods C G, Bond J, Enard W. Autosomal recessive primary microcephaly (MCPH): A review of clinical, molecular, and evolutionary findings. Am J Hum Genet, 2005, 76: 717−728[DOI]

15 Zhang J. Evolution of the human ASPM gene, a major determinant of brain size. Genetics, 2003, 165: 2063−2070

16 Mekel-Bobrov N, Gilbert S L, Evans P D, et al. Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. Science, 2005, 309: 1720−1722[DOI]

17 Evans P D, Gilbert S L, Mekel-Bobrov N, et al. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science, 2005, 309: 1717−1720[DOI]

18 Bustamante C D, Fledel-Alon A, Williamson S, et al. Natural selection on protein-coding genes in the human genome. Nature, 2005, 437: 1153−1157[DOI]

19 Clark A G, Glanowski S, Nielsen R, et al. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science, 2003, 302: 1960−1963[DOI]

20 Nielsen R, Bustamante C, Clark A G, et al. A Scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol, 2005, 3: e170[DOI]

21 Clark N L, Swanson W J. Pervasive adaptive evolution in primate seminal proteins. PLoS Genet, 2005, 1: e35

22 Enard W, Khaitovich P, Klose J, et al. Intra- and interspecific variation in primate gene expression patterns. Science, 2002, 296: 340−343[DOI]

23 Marvanova M, Menager J, Bezard E, et al. Microarray analysis of nonhuman primates: validation of experimental models in neurological disorders. Faseb J, 2003, 17: 929−931

24 Su A I, Cooke M P, Ching K A, et al. Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci USA, 2002, 99: 4465−4470[DOI]

25 Su A I, Wiltshire T, Batalov S, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA, 2004, 101: 6062－6067[DOI]

26 Florea L, Hartzell G, Zhang Z, et al. A computer program for aligning a cDNA sequence with a genomic DNA sequence. Genome Res, 1998, 8: 967－974

27 Pamilo P, Bianchi N O. Evolution of the Zfx and Zfy genes: Rates and interdependence between the genes. Mol Biol Evol, 1993, 10: 271－281

28 Li W H. Unbiased estimation of the rates of synonymous and non-synonymous substitution. J Mol Evol, 1993, 36: 96－99[DOI]

29 Mi H, Lazareva-Ulitsky B, Loo R, et al. The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Res, 2005, 33: D284－288[DOI]

30 Yang Z. PAML: A program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci, 1997, 13: 555－556

31 Yang Z, Nielsen R, Goldman N, et al. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics, 2000, 155: 431－449

32 Yang Z, Bielawski J P. Statistical methods for detecting molecular adaptation. Trends Ecol Evol, 2000, 15: 496－503[DOI]

33 Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform, 2004, 5: 150－163[DOI]

34 Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol, 2000, 17: 32－43

35 Hurst L D. The K_a/K_s ratio: Diagnosing the form of sequence evolution. Trends Genet, 2002, 18: 486

36 Caceres M, Lachuer J, Zapala M A, et al. Elevated gene expression levels distinguish human from non-human primate brains. Proc Natl Acad Sci USA, 2003, 100: 13030－13035[DOI]

37 Khaitovich P, Paabo S, Weiss G. Towards a neutral evolutionary model of gene expression. Genetics, 2005, 170: 929－939[DOI]

38 Gu J, Gu X. Further statistical analysis for genome-wide expression evolution in primate brain/liver/fibroblast tissues. Hum Genomics, 2004, 1: 247－254

39 Hsieh W P, Chu T M, Wolfinger R D, et al. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. Genetics, 2003, 165: 747－757

40 Jagadeeshan S, Singh R S. Rapidly evolving genes of *Drosophila*: Differing levels of selective pressure in testis, ovary, and head tissues between sibling species. Mol Biol Evol, 2005, 22: 1793－1801[DOI]

41 Wyckoff G J, Wang W, Wu C I. Rapid evolution of male reproductive genes in the descent of man. Nature, 2000, 403: 304－309[DOI]

42 Torgerson D G, Kulathinal R J, Singh R S. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol Biol Evol, 2002, 19: 1973－1980

43 Sumiyama K, Saitou N, Ueda S. Adaptive evolution of the IgA hinge region in primates. Mol Biol Evol, 2002, 19: 1093－1099

44 Sawyer S L, Emerman M, Malik H S. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol, 2004, 2: E275[DOI]

45 Zhang J, Webb D M. Rapid evolution of primate antiviral enzyme APOBEC3G. Hum Mol Genet, 2004, 13: 1785－1791[DOI]

46 Holmes E C. Adaptation and immunity. PLoS Biol, 2004, 2: E307[DOI]

47 Sawyer S L, Wu L I, Emerman M, et al. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proc Natl Acad Sci USA, 2005, 102: 2832－2837[DOI]

48 Liu H L, Wang Y Q, Liao C H, et al. Adaptive evolution of primate TRIM5alpha, a gene restricting HIV-1 infection. Gene, 2005, 362: 109－116[DOI]

49 Sawyer S L, Wu L I, Akey J M, et al. High-frequency persistence of an impaired allele of the retroviral defense gene trim5alpha in humans. Curr Biol, 2006, 16: 95－100[DOI]

50 Semple C A, Maxwell A, Gautier P, et al. The complexity of selection at the major primate beta-defensin locus. BMC Evol Biol, 2005, 5: 32[DOI]

51 Prugnolle F, Manica A, Charpentier M, et al. Pathogen-driven selection and worldwide HLA class I diversity. Curr Biol, 2005, 15: 1022－1027[DOI]

52 Payseur B A, Cutter A D, Nachman M W. Searching for evidence of positive selection in the human genome using patterns of microsatellite variability. Mol Biol Evol, 2002, 19: 1143－1153

53 Lu J, Wu C I. Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. Proc Natl Acad Sci USA, 2005, 102: 4063－4067[DOI]

54 Saifi G M, Chandra H S. An apparent excess of sex- and reproduction-related genes on the human X chromosome. Proc Biol Sci, 1999, 266: 203－209[DOI]

55 Wang P J, McCarrey J R, Yang F, et al. An abundance of X-linked genes expressed in spermatogonia. Nat Genet, 2001, 27: 422－426[DOI]

56 Lercher M J, Urrutia A O, Hurst L D. Evidence that the human X chromosome is enriched for male-specific but not female-specific genes. Mol Biol Evol, 2003, 20: 1113－1116[DOI]

57 Zechner U, Wilda M, Kehrer-Sawatzki H, et al. A high density of X-linked genes for general cognitive ability: A run-away process shaping human evolution? Trends Genet, 2001, 17: 697－701[DOI]

58 Torgerson D G, Singh R S. Sex-linked mammalian sperm proteins evolve faster than autosomal ones. Mol Biol Evol, 2003, 20: 1705－1709[DOI]

59 Wang X, Zhang J. Rapid evolution of mammalian X-linked testis-expressed homeobox genes. Genetics, 2004, 167: 879－888[DOI]

60 Swanson W J, Yang Z, Wolfner M F, et al. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci USA, 2001, 98: 2509－2514[DOI]

61 Kuroki Y, Toyoda A, Noguchi H, et al. Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway. Nat Genet, 2006, 38: 158－167[DOI]

62 Schnieders F, Dork T, Arnemann J, et al. Testis-specific protein, Y-encoded (TSPY) expression in testicular tissues. Hum Mol Genet, 1996, 5: 1801－1807[DOI]

63 Sun C, Skaletsky H, Birren B, et al. An azoospermic man with a *de novo* point mutation in the Y-chromosomal gene USP9Y. Nat Genet, 1999, 23: 429－432[DOI]

64 Wong W S, Yang Z, Goldman N, et al. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics, 2004, 168: 1041－1051[DOI]

65 Goodman M, Porter C A, Czelusniak J, et al. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol, 1998, 9: 585－598[DOI]

66 Yu X J, Zheng H K, Wang J, et al. Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. Genomics, 2006, 80: 745－751[DOI]

67 van de Vosse E, Walpole S M, Nicolaou A, et al. Characterization of SCML1, a new gene in Xp22, with homology to developmental polycomb genes. Genomics, 1998, 49: 96－102[DOI]

68 Rajkovic A, Lee J H, Yan C, et al. The ret finger protein-like 4 gene, Rfpl4, encodes a putative E3 ubiquitin-protein ligase expressed in adult germ cells. Mech Dev, 2002, 112: 173－177[DOI]

69 Suzumori N, Burns K H, Yan W, et al. RFPL4 interacts with oocyte proteins of the ubiquitin-proteasome degradation pathway. Proc Natl Acad Sci USA, 2003, 100: 550－555[DOI]

70 Leatherman J L, Levin L, Boero J, et al. germ cell-less acts to repress transcription during the establishment of the Drosophila germ cell lineage. Curr Biol, 2002, 12: 1681－1685[DOI]

71 Leatherman J L, Kaestner K H, Jongens T A. Identification of a mouse germ cell-less homologue with conserved activity in *Drosophila*. Mech Dev, 2000, 92: 145－153[DOI]

72 Masuhara M, Nagao K, Nishikawa M, et al. Enhanced degradation of MDM2 by a nuclear envelope component, mouse germ cell-less. Biochem Biophys Res Commun, 2003, 308: 927－932[DOI]

73 Song J, Mangold M, Suske G, et al. Characterization and promoter analysis of the mouse gene for transcription factor Sp4. Gene, 2001, 264: 19－27[DOI]

74 Supp D M, Witte D P, Branford W W, et al. Sp4, a member of the Sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility. Dev Biol, 1996, 176: 284－299[DOI]

75 Xu M, Xiao J, Chen J, et al. Identification and characterization of a novel human testis-specific Golgi protein, NYD-SP12. Mol Hum Reprod, 2003, 9: 9－17[DOI]

76 Liu Y, Yang S, Lin A A, et al. Molecular evolution of CXCR1, a G protein-coupled receptor involved in signal transduction of neutrophils. J Mol Evol, 2005, 61: 691－696[DOI]

77 Wang Y Q, Qian Y P, Yang S, et al. Accelerated evolution of the pituitary adenylate cyclase-activating polypeptide precursor gene during human origin. Genetics, 2005, 170: 801－806[DOI]

78 Yang Z. The power of phylogenetic comparison in revealing protein function. Proc Natl Acad Sci USA, 2005, 102: 3179－3180[DOI]

79 Choi S S, Lahn B T. Adaptive evolution of MRG, a neuron-specific gene family implicated in nociception. Genome Res, 2003, 13: 2252－2259[DOI]

80 Yang S, Liu Y, Lin A A, et al. Adaptive evolution of MRGX2, a human sensory neuron specific gene involved in nociception. Gene, 2005, 352: 30－35[DOI]

81 Coy J F, Dressler D, Wilde J, et al. Mutations in the transketolase-like gene TKTL1: Clinical implications for neurodegenerative diseases, diabetes and cancer. Clin Lab, 2005, 51: 257－273

82 Wei C J, Xu X, Lo C W. Connexins and cell signaling in development and disease. Annu Rev Cell Dev Biol, 2004, 20: 811－838[DOI]

83 Chen X H, Shi H, Liu X L, et al. The testis-specific apoptosis related gene TTL.6 underwent adaptive evolution in the lineage leading to humans. Gene, 2005, 370: 58－63[DOI]

Supplementary Table S1 The summary statistics for 51 molecular function categories

MF	Panther Mol function category	No. total genes	No. genes of K_a=0, K_s =0	No. genes of K_a>0, K_s=0	No. genes of positively selected_obs	No. genes of positively selected_exp	*P*-values (Fisher's exact test, one tailed)	K_a/K_s	Std Error
1	Actin family cytoskeletal protein	130	7	3	6	15	*0.0043*	0.2525	0.0311
2	Cation transporter	23	2	0	1	3	0.2263	0.1986	0.0758
3	Cell adhesion molecule	173	5	8	21	20	0.4865	0.4285	0.0482
4	Cell junction protein	32	0	2	4	4	0.5344	0.3239	0.0954
5	Chaperone	56	6	3	6	7	0.5012	0.3888	0.1006
6	Cyclase	6	0	0	0	1	0.4699	0.2355	0.0923
7	Cytokine	30	3	4	8	4	*0.0208*	0.4816	0.1557
8	Cytokine receptor	35	2	2	4	4	0.5999	0.3966	0.0671
9	Cytoskeletal protein	674	46	25	51	80	*0.0003*	0.3054	0.0171
10	Defense/immunity protein	107	1	7	24	13	0.0016	0.6038	0.0550
11	Extracellular matrix	113	2	4	10	13	0.2069	0.3905	0.0328
12	G-protein	99	17	6	8	12	0.1601	0.2127	0.0745
13	G-protein modulator	125	3	4	8	15	*0.0344*	0.2963	0.0302
14	Growth factor	1	0	0	0	0	0.8817	0.2924	----
15	Helicase	62	1	1	3	7	0.0556	0.2900	0.0379
16	Hydrolase	352	11	9	30	42	*0.0319*	0.3972	0.0353
17	Intermediate filament	1	0	0	0	0	0.8817	0.2111	----
18	Ion channel	133	8	2	7	16	*0.0088*	0.2626	0.0419
19	Isomerase	62	1	4	7	7	0.5462	0.3579	0.0448
20	Kinase	340	18	8	23	40	*0.0017*	0.3173	0.0196
21	Kinase modulator	37	3	3	6	4	0.2704	0.2966	0.0593
22	Ligand-gated ion channel	44	1	0	1	5	*0.0274*	0.2684	0.1021
23	Ligase	196	6	7	20	23	0.2856	0.3500	0.0295
24	Membrane traffic protein	147	11	3	10	17	*0.0337*	0.3337	0.0469
25	Microtubule family cytoskeletal protein	64	1	2	5	8	0.2178	0.3086	0.0463
26	Miscellaneous function	263	18	12	26	31	0.1953	0.3503	0.0299
27	Molecular function unclassified	2265	145	144	356	268	0.000001	0.4518	0.0140
28	mRNA processing factor	46	6	1	4	5	0.3529	0.2058	0.0574
29	Nuclease	26	1	3	4	3	0.3714	0.3574	0.0717
30	Nucleic acid binding	951	61	40	87	113	*0.0074*	0.3311	0.0169
31	Other enzyme regulator	20	2	0	2	2	0.5728	0.4814	0.1165
32	Oxidoreductase	254	12	14	31	30	0.4579	0.3640	0.0324
33	Peptide hormone	16	0	2	3	2	0.2928	0.4275	0.1275
34	Phosphatase	118	7	9	18	14	0.1591	0.3702	0.0474
35	Phosphatase modulator	3	1	1	1	0	0.3148	0.1327	----
36	Protease	221	5	9	15	26	*0.0103*	0.3427	0.0313
37	Protease inhibitor	37	0	0	3	4	0.3488	0.4673	0.0841
38	Protein kinase	201	10	3	11	24	*0.0019*	0.3134	0.0256
39	Receptor	432	12	18	56	51	0.2612	0.4045	0.0273
40	Select calcium binding protein	140	10	5	13	17	0.2175	0.3942	0.0525
41	Select regulatory molecule	363	28	16	32	43	*0.0445*	0.3075	0.0258
42	Signaling molecule	271	18	22	44	32	0.0211	0.4425	0.0432
43	Synthase and synthetase	108	4	4	11	13	0.3658	0.3693	0.0412
44	Transcription factor	675	49	40	77	80	0.4007	0.3623	0.0239
45	Transfer/carrier protein	134	5	7	17	16	0.4204	0.3367	0.0385
46	Transferase	303	13	14	29	36	0.1331	0.3769	0.0317
47	Translation factor	40	2	0	1	5	*0.0418*	0.2403	0.0479
48	Transporter	266	10	6	23	31	0.0641	0.3375	0.0237
49	Viral protein	1	0	1	1	0	0.1184	----	----
50	Voltage-gated ion channel	45	4	2	3	5	0.2057	0.2109	0.0455
51	Zinc finger transcription factor	210	9	15	28	25	0.2833	0.3830	0.0356
	overall	10451	587	495	1159	1236		0.3711	0.0054

Supplementary Table S2 The summary statistics for 50 biological process categories

BP	Panther biol process category	No. total genes	No. genes excluding K_a=0, K_s=0	No. genes K_a=0, K_s=0	No. genes K_a>0, K_s=0	No. genes of positively selected_obs	No. genes of positively selected_exp	*P*-values (Fisher's exact test, one tailed)	K_a/K_s	Std Error
1	Amino acid metabolism	105	105	0	1	9	12	0.1927	0.3956	0.0521
2	Apoptosis	180	166	14	4	16	21	0.1355	0.3858	0.0283
3	Biological process unclassified	2256	2114	142	135	337	267	0.000076	0.4499	0.0143
4	Blood circulation and gas exchange	23	23	0	4	6	3	0.0473	0.2764	0.0788
5	Carbohydrate metabolism	211	207	4	10	19	25	0.1233	0.3608	0.0362
6	Cell adhesion	216	210	6	11	27	26	0.4139	0.4364	0.0427
7	Cell communication	299	286	13	21	43	35	0.1089	0.4328	0.0414
8	Cell cycle	362	345	17	20	38	43	0.2499	0.3485	0.0265
9	Cell proliferation and differentiation	295	272	23	15	27	35	0.0922	0.3484	0.0277
10	Cell structure and motility	376	354	22	16	27	44	0.0025	0.2963	0.0211
11	Cell surface receptor mediated signal transduction	394	375	19	14	37	47	0.0799	0.3363	0.0254
12	Chemosensory perception	3	2	1	0	0	0	0.6855	0.1616	0.1616
13	Coenzyme and prosthetic group metabolism	69	65	4	2	5	8	0.1612	0.3107	0.0502
14	Developmental processes	767	718	49	33	66	91	0.0037	0.3396	0.0212
15	DNA metabolism	103	101	2	3	9	12	0.2119	0.3875	0.0412
16	Ectoderm development	192	175	17	5	12	23	0.008	0.2862	0.0327
17	Electron transport	90	85	5	6	13	11	0.2663	0.3779	0.0667
18	Endocytosis	87	81	6	3	6	10	0.0996	0.2774	0.0426
19	Fatty acid metabolism	30	27	3	1	1	4	0.1156	0.3421	0.0510
20	Gametogenesis	55	53	2	3	7	7	0.4805	0.3706	0.0504
21	Homeostasis	63	61	2	3	6	7	0.3728	0.3059	0.0485
22	Immunity and defense	411	397	14	20	67	49	0.0056	0.4888	0.0292
23	Intracellular protein traffic	396	364	32	15	38	47	0.1005	0.3254	0.0283
24	Intracellular signaling cascade	270	248	22	12	24	32	0.0803	0.3359	0.0307
25	Ion transport	204	191	13	7	15	24	0.0268	0.2780	0.0320
26	Lipid, fatty acid and steroid metabolism	293	275	18	12	28	35	0.136	0.3560	0.0229
27	Mesoderm development	153	146	7	8	15	18	0.2663	0.3181	0.0321
28	Miscellaneous	10	10	0	1	3	1	0.1055	0.5795	0.2203
29	Mitosis	79	79	0	5	7	9	0.2704	0.2640	0.0333
30	mRNA transcription	544	501	43	35	64	64	0.515	0.3409	0.0222
31	Muscle contraction	61	59	2	3	6	7	0.4076	0.3485	0.0664
32	Neuronal activities	195	182	13	7	15	23	0.0429	0.3025	0.0435
33	Nitrogen metabolism	13	11	2	0	0	2	0.1949	0.2361	0.0599
34	Non-vertebrate process	11	11	0	0	0	1	0.2506	0.1005	0.0521
35	Nucleoside, nucleotide and nucleic acid metabolism	1158	1087	71	52	113	137	0.0212	0.3371	0.0140
36	Oncogenesis	207	203	4	11	21	24	0.2688	0.3392	0.0365
37	Other metabolism	231	215	16	14	32	27	0.2007	0.3996	0.0363
38	Phosphate metabolism	23	22	1	2	3	3	0.5231	0.3151	0.0637
39	Pre-mRNA processing	94	86	8	1	5	11	0.0282	0.2304	0.0342
40	Protein metabolism and modification	1012	964	48	40	84	120	0.0004	0.3301	0.0148
41	Protein modification	357	343	14	8	22	42	0.0003	0.3151	0.0202
42	Protein targeting and localization	84	77	7	3	11	10	0.4101	0.4323	0.0688
43	Sensory perception	91	87	4	5	9	11	0.3551	0.3624	0.0652
44	Signal transduction	1240	1171	69	54	121	147	0.0182	0.3498	0.0150
45	Steroid metabolism	39	38	1	2	5	5	0.4973	0.3410	0.0737
46	Sulfur metabolism	31	30	1	4	4	4	0.5085	0.2514	0.0485
47	Synaptic transmission	59	55	4	0	3	7	0.0711	0.3128	0.0965
48	T-cell mediated immunity	14	13	1	0	2	2	0.5066	0.5557	0.1041
49	Transport	539	514	25	14	45	64	0.0072	0.3271	0.0188
50	Vitamin metabolism	4	4	0	1	1	0	0.3959	0.3121	0.0572
	overall	13999	13208	791	646	1474	1656		0.3621	0.0046

Supplementary Table S3 Number and relative frequency of brain-expressed genes that are potentially under positive Darwinian selection on each chromosome

Chr	No. of total genes	No. genes of K_a=0, K_s=0	No. genes of K_a=0, K_s>0	No. genes of K_a>0, K_s=0	Percentage of genes that are potentially under positive selection[a)]	No. of Observed positively selected	No. of expected positively selected	*P*-value (Fisher's exact test, one-tailed)	Mean K_a/K_s[b)]	SE_K_a/K_s
1	830	56	173	36	11.9	99	98	0.4855	0.3750	0.0187
2	476	28	99	23	12.2	58	56	0.4309	0.3924	0.0271
3	420	22	72	24	13.6	57	50	0.1604	0.4150	0.0296
4	273	16	55	9	9.9	27	32	0.1903	0.3715	0.0300
5	323	12	64	22	14.6	47	38	0.0853	0.3762	0.0270
6	374	23	65	24	11.8	44	44	0.5322	0.3717	0.0265
7	328	23	74	17	12.2	40	39	0.4474	0.3783	0.0332
8	228	20	51	9	9.6	22	27	0.1846	0.3425	0.0318
9	263	13	53	14	12.2	32	31	0.4636	0.3673	0.0265
10	227	11	47	7	11.5	26	27	0.4825	0.4139	0.0442
11	449	25	75	22	9.8	44	53	0.1088	0.3470	0.0198
12	375	27	74	26	14.7	55	44	0.0614	0.4230	0.0361
13	111	4	25	5	11.7	13	13	0.5579	0.4116	0.0837
14	204	19	47	9	11.8	24	24	0.5427	0.3561	0.0374
15	230	12	39	13	10.0	23	27	0.2318	0.3633	0.0356
16	341	14	59	10	8.2	28	40	0.022	0.3501	0.0241
17	451	26	80	22	11.1	50	53	0.3497	0.3928	0.0241
18	101	3	14	5	11.9	12	12	0.5403	0.4245	0.0857
19	456	12	78	20	11.8	54	54	0.5203	0.3788	0.0220
20	208	11	40	11	9.1	19	25	0.1385	0.3036	0.0293
21	66	0	12	3	10.6	7	8	0.4733	0.3580	0.0534
22	182	12	25	4	8.2	15	22	0.0809	0.4434	0.0613
X	255	25	44	23	19.6	50	30	0.0003	0.5623	0.0615
Y	3	0	0	0	33.3	1	0.4	0.3148	0.5980	0.3964
Unknown	2	0	0	1	100.0	2	0.2	0.014	1.4522	
All autosomes	6916	389	1321	335	11.5	796	818	0.2781	0.3792	0.0067
Total	7176	414	1365	359		849	849		0.3854	0.0068

a) Positively selected genes include the genes with K_a/K_s>1 and $K_a/K_s = \infty$ (K_a>0, K_s=0); b) Mean K_a/K_s was calculated after eliminating the genes with K_a=0, K_s=0, and K_a >0, K_s=0.

Supplementary Table S4 The 34 candidate genes analyzed in this study[a)]

Gene	Accession No.	Description	L_C	K_a/K_s	Chr	Tissue	Molecular Funciton	Biological Process
FLJ3187	AK056437		438	1.45	NA	T-lymphocytes	DNA binding, Zinc ion binding	regulation of transcription, DNA-dependent
ZNF229	AK091541	zinc finger protein 229	124	∞	19	embryonic stem cells	DNA binding, zinc ion binding	regulation of transcription, DNA-dependent transcription
TXNRD1	AK098698	thioredoxin reductase 1	331	3.84	12	smoothMuscle	oxidoreductase	electron transport/Other metabolism
Hs.548058	AY030239	transcribed locus, strongly similar to NP_006197.1 platelet-derived growth factor receptor alpha precurso	88	∞	1	no hit	no panther hits	no panther hits
MGC3522	BC029877		110	∞	NA	no hit	no panther hits	no panther hits
GMCL1L	BC033886	germ cell-less homolog 1 (*Drosophila*)-like	524	1.82	5	testis	protein binding	spermatogenesis, cell differentiation
IL8RA	NM_000634	interleukin 8 receptor, alpha	351	2.54	2	whole blood	receptor	cell structure and motility/ cell surface receptor mediated signal transduction/ signal transduction
SP4	NM_003112	Sp4 transcription factor	743	3.48	7	no Hit	transcription factor/ zinc finger transcription factor	mRNA transcription/ nucleoside, nucleotide and nucleic acid metabolism
DEFA3	NM_005217	defensin, alpha 3, neutrophil-specific	94	∞	8	bone marrow	defense/immunity protein	immunity and defense
RFPL3	NM_006604	ret finger protein-like 3	288	3.7	22	brain	ubiquitin-protein, ligase activity	protein-ubiquitination
SCML1	NM_006746	sex comb on midleg-like 1 (Drosophila)	208	7.12	X	colorectal adenocarcinoma	nucleic acid binding/ transcription factor	developmental processes
LGTN	NM_006893	ligatin	584	4.18	1	721_B_lymphoblasts	receptor	intracellular protein traffic
TKTL1	NM_012253	transketolase-like 1	540	5.35	X	testis interstitial	transferase	carbohydrate metabolism
dJ222E13.1	NM_014509	kraken-like	321	7.71	22	embryonic stem cells	hydrolase activity	aromatic compound metabolism
C3orf18	NM_016210	chromosome 3 open reading frame 18	162	∞	3	pituitary	molecular function unclassified	biological process unclassified
FLJ20449	NM_017826	hypothetical protein FLJ20449	416	7.38	13	testis interstitial	no panther hits	no panther hits
FLJ10439	NM_018093	hypothetical protein FLJ10439	386	2.17	11	721_B_lymphoblasts	molecular function unclassified	biological process unclassified
C14orf131	NM_018335	chromosome 14 open reading frame 131	402	3.41	14	testis	molecular function unclassified	biological process unclassified
TREML2	NM_024807	triggering receptor expressed on myeloid cells-like 2	385	5.14	6	placenta	molecular function unclassified	biological process unclassified
NET-5	NM_031285	transmembrane 4 superfamily member tetraspan NET-5	239	4.57	12	heart	signaling molecule/ cell adhesion molecule	cell adhesion/Cell communication/ signal transduction

(*To be continued on the next page*)

(*Continued*)

Gene	Accession No.	Description	L_C	K_a/K_s	Chr	Tissue	Molecular Funciton	Biological Process
NRIP2	NM_031474	nuclear receptor interacting protein 2	167	∞	12	OlfactoryBulb	hydrolase/protease/ transcription factor	mRNA transcription/ nucleoside, nucleotide and nucleic acid metabolism/ other metabolism/protein metabolism and modification
SPATA16	NM_031955	spermatogenesis associated 16	569	2.88	3	testis Interstitial	molecular function unclassified	biological process unclassified
PEPP-2	NM_032498	PEPP subfamily gene 2	294	4.14	X	testis	nucleic acid binding/ transcription factor	developmental processes/ mRNA transcription/ nucleoside, nucleotide and nucleic acid metabolism
MED8	NM_052877	mediator of RNA polymerase II transcription, subunit 8 homolog (yeast)	212	∞	1	whole brain	receptor activity	regulation of transcrition, DNA-dependent transcription
GBP5	NM_052942	guanylate binding protein 5	586	5.44	1	T-lymphocytes	select regulatory molecule/Cytoskeletal protein/G-protein	immunity and defense
MRGX2	NM_054030	G protein-coupled receptor MRGX	330	3.32	11	unclassified	no panther hits	no panther hits
MGC33309	NM_152413	hypothetical protein MGC33309	420	4.22	8	testis	transferase	amino acid metabolism
MGC35118	NM_152453	hypothetical protein MGC35118	271	3.37	15	testis	molecular function unclassified	biological process unclassified
FLJ30313	NM_152757	hypothetical protein FLJ30313	134	∞	20	muscle	molecular function unclassified	biological process unclassified
C20orf96	NM_153269	chromosome 20 open reading frame 96	363	3.06	20	prostate	molecular function unclassified	biological process unclassified
CX40.1	NM_153368	connexin40.1	370	2.55	10	brain	cell junction protein	signal transduction
FLJ34790	NM_173621	hypothetical protein FLJ34790	148	3.05	17	peripheral-nervous system	molecular function unclassified	biological process unclassified
FLJ35808	NM_173623	hypothetical protein FLJ35808	569	3.47	17	testis	tublin-tyrosine ligase activity	protein modification
ADCYAP1	NM_001117	adenylate cyclase activating polypeptide 1 (pituitary)	176	1.83	18	cerebrum	peptide hormone/ signaling molecule	signal transduction

a) *L*c, Number of codons; K_a/K_s, ratio between human and chimpanzee calculated with yn00 method.