Hum Genet (2006) 119: 169–178
DOI 10.1007/s00439-005-0125-6

ORIGINAL INVESTIGATION

Yoonsoo Hahn · Byungkook Lee

Human-specific nonsense mutations identified by genome sequence comparisons

Received: 7 October 2005 / Accepted: 10 December 2005 / Published online: 10 February 2006

Abstract The comparative study of the human and chimpanzee genomes may shed light on the genetic ingredients for the evolution of the unique traits of humans. Here, we present a simple procedure to identify human-specific nonsense mutations that might have arisen since the human–chimpanzee divergence. The procedure involves collecting orthologous sequences in which a stop codon of the human sequence is aligned to a non-stop codon in the chimpanzee sequence and verifying that the latter is ancestral by finding homologs in other species without a stop codon. Using this procedure, we identify nine genes (*CML2*, *FLJ14640*, *MT1L*, *NPPA*, *PDE3B*, *SERPINA13*, *TAP2*, *UIP1*, and *ZNF277*) that would produce human-specific truncated proteins resulting in a loss or modification of the function. The premature terminations of *CML2*, *MT1L*, and *SERPINA13* genes appear to abolish the original function of the encoded protein because the mutation removes a major part of the known active site in each case. The other six mutated genes are either known or presumed to produce functionally modified proteins. The mutations of five genes (*CML2*, *FLJ14640*, *MT1L*, *NPPA*, *TAP2*) are known or predicted to be polymorphic in humans. In these cases, the stop codon alleles are more prevalent than the ancestral allele, suggesting that the mutant alleles are approaching fixation since their emergence during the human evolution. The findings support the notion that functional modification or inactivation of genes by nonsense mutation is a part of the process of adaptive evolution and acquisition of species-specific features.

Keywords Nonsense mutation · Human · Chimpanzee · Molecular evolution

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00439-005-0125-6 and is accessible for authorized users.

Y. Hahn · B. Lee (✉)
Laboratory of Molecular Biology, Center for Cancer Research,
National Cancer Institute, National Institutes of Health,
Building 37, MSC 4264, 37 Convent Drive Room 5120A,
Bethesda, MD, 20892-4264, USA
E-mail: bk@nih.gov
Tel.: +1-301-4966580
Fax: +1-301-4804654

Introduction

Genome sequence contains a record of evolutionary history that a species has experienced. By comparing genome sequences of different species, one can detect genetic modifications that must have led to physical, physiological, and behavioral changes that are responsible for the speciation and accumulation of lineage-specific traits. The genome sequence of the modern human (*Homo sapiens*) contains numerous human-specific additions, deletions, alterations, and relocations of genetic materials that have been applied during the 5–7 million years since human and chimpanzee split. Such modifications can result in quantitative and spatiotemporal changes in gene expression and/or structural changes of individual proteins (Ruvolo 2004).

It has been argued that crucial characteristics of a species develop by changes of gene expression pattern rather than by changes in individual protein sequences (Carroll 2005). However, modifications in the coding sequences of genes are clearly important in either case. For example, higher expression levels for many genes in the human brain (Preuss et al. 2004) might be caused by modification of *cis*-regulatory elements associated with the genes but also by accelerated amino acid substitution of *trans*-acting factors that control the expression of the genes. Accelerated amino acid substitution of some genes has been reported to be associated with the evolution of particular human-specific traits; for example, the *FOXP2* gene and the evolution of speech and language (Enard et al. 2002), and the *ASPM* gene and the brain enlargement (Evans et al. 2004). In dogs, breed-specific limb and skull morphology are believed to be associated with variations in the number of amino acid repeats in the Alx4 and the Runx2 proteins, respectively (Fondon and Garner 2004). Comparison of 13,731 human genes with their chimpanzee orthologs reveals that many genes involved in sensory perception or immune defenses are subject to positive selection (Nielsen et al. 2005).

Substitutions and in-frame insertions or deletions of amino acids are commonly observed between orthologous proteins in closely related species. However, frameshift mutations and nonsense mutations, which lead to inactivation or functional modification of the gene products, may also be associated with the acquisition of species-specific features. For example, the loss of the *MYH16* gene function by a two nucleotide deletion was suggested to be linked with a specific change of skull shape by reducing masticatory muscle mass and allowing development of bigger brain during human evolution (Stedman et al. 2004), although a new study suggests that the mutation occurred much earlier (Perry et al. 2005). Likewise, the inactivation of the *CMAH* gene caused by the *Alu*-mediated exon deletion was proposed to be associated with brain expansion of human ancestors (Chou et al. 2002). Other examples of genes the function of which has been lost in humans include *FMO2*, *KRTHAP1*, *EMR4*, and *SIGLEC13* (Angata et al. 2004; Dolphin et al. 1998; Hamann et al. 2003; Winter et al. 2001).

Recent progress of primate genome sequencing projects including chimpanzee (*Pan troglodytes*), orangutan (*Pongo pygmaeus*), and rhesus macaque (*Macaca mulatta*) together with the release of the nearly complete human genome sequence enables one to directly compare the genome sequences to find lineage-specific genetic alterations (Goodman et al. 2005; International Human Genome Sequencing Consortium 2004; Li and Saunders 2005; The Chimpanzee Sequencing and Analysis Consortium 2005). Previously, we reported identification of nine novel human-specific frameshift mutations by comparing the human and the chimpanzee genome sequences (Hahn and Lee 2005). In this report, we applied a similar technique to identify human-specific nonsense mutations that might result in generation of inactive or functionally modified proteins specifically in the human lineage. The procedure involves the collection of the last coding exons of human genes of which stop codons align with non-stop codons in the orthologous chimpanzee exons. Since the current chimpanzee genome assembly (build 1, November 13, 2003 release) has limited sequence accuracy, we assume that many of the mismatches between the two genome sequences are due to errors in the chimpanzee genome sequence. We therefore require that there be at least one known non-human/non-chimpanzee protein homolog for each case that supports the chimpanzee protein as being ancestral. Each human-specific mutation candidate is further scrutinized by comparing the human/chimpanzee sequences with the orthologous orangutan and/or rhesus macaque sequences assembled from the whole genome shotgun (WGS) trace data. Using this procedure, we identify nine highly probable human-specific nonsense mutations in the current human genome sequence (build 35). We discuss the possible functional consequence of each mutation.

Materials and methods

Sequence and alignment data source

The human mRNA to genome alignment data and the genome sequence assemblies (human, build 35; chimpanzee, build 1) were downloaded from the University of California at Santa Cruz (UCSC) web site (http://www.genome.ucsc.edu) in July 2005. The human mRNA to human genome alignments were extracted from the tables ‘hg17.RefSeqAli’ and ‘hg17.all_mrna.’ The human mRNA to chimpanzee genome alignments were found in the tables ‘panTro1.xenoRefSeqAli’ and ‘panTro1.xenoMrna.’ The coding region coordinates were obtained from the tables ‘hg17.gbCdnaInfo’ and ‘hg17.cds’ to define the last codon of each human mRNA sequence. The non-redundant non-human vertebrate protein database was locally prepared by processing the nr (non-redundant) protein database of the National Center for Biotechnology Information (NCBI; ftp://www.ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz).

Collection of the coding exons with the stop codon discrepancy

For a systematic identification of human-specific nonsense mutations, we first collected the human/chimpanzee coding exon pairs, in each of which the human exon contained in-frame stop codon that was aligned with a non-stop codon in the chimpanzee genome sequence. The human mRNA to human genome sequence alignment and the human mRNA to chimpanzee genome sequence alignment data were obtained and filtered by using an in-house Perl program. If a sequence aligned to multiple places in a genome, only the best alignment was considered. The last codon sequence of each human mRNA sequence was examined to ensure that the coding region ended with one of the three stop codons. In order to remove redundancy, only one mRNA sequence (a RefSeq sequence when available) alignment was kept if more than one mRNA sequence were mapped to the same last codon in the human genome sequence. For each human mRNA sequence, the last codon sequences were extracted from the human and the corresponding chimpanzee genome sequences. Only the cases where the human genome sequence matched the human mRNA stop codon and the aligned chimpanzee codon sequence was not a stop codon were selected for further investigation.

Identification of the non-human/non-chimpanzee homologs

In order to find the non-human/non-chimpanzee protein homolog, we generated a hypothetical chimpanzee-like mRNA sequence by replacing the stop codon of the human mRNA sequence with the corresponding codon in the aligned chimpanzee genome sequence. The resulting chimpanzee-like mRNA was re-translated using the start codon annotated in the human mRNA to produce a chimpanzee-like protein sequence. We then performed BLAST searches through a custom-built non-redundant non-human vertebrate protein database using the human and chimpanzee-like protein sequences as queries to find homologs in the third species. The BLAST outputs were parsed to obtain a hit list for each human/chimpanzee-like protein pair. The hit list was sorted by the BLAST score and the top five hits were retained. The cases were noted wherein the third species sequence had a score increase of at least 1 bit when the chimpanzee-like protein was used as the query over when the human protein was used. (This is a change from our previous study (Hahn and Lee 2005) wherein we used a minimum increase in score of 10 bits. The condition of minimal score increase was used here in order to detect mutations that occur near the C-terminus and therefore remove only a small number of amino acids.) The final candidate cases were subjected to the in-depth analysis described below.

Collection of the highly probable human-specific nonsense mutations

In order to verify and select highly reliable human-specific nonsense mutation cases, we manually inspected the final candidates obtained in the previous step. All available sequence data for human, chimpanzee, and other species were obtained by database searches using the NCBI BLAST web server (http://www.ncbi.nlm.nih.gov/BLAST) and analyzed to verify that the nonsense mutations were human-specific. The genuine chimpanzee cDNAs were predicted from the chimpanzee genome sequence by assembling predicted exons. These latter were obtained from an alignment of the human mRNA and the chimpanzee genomic fragment using the GMAP program (Wu and Watanabe 2005). The composite chimpanzee cDNAs were sometimes amended by using the chimpanzee virtual transcripts (Clark et al. 2003) or the whole genome shotgun (WGS) trace data obtained at the Trace Archive database (http://www.ncbi.nlm.nih.gov/blast/mmtrace.shtml). The putative orangutan and the rhesus macaque orthologs were obtained by searching through the Trace Archive database and the matched WGS sequences were assembled by using the cap3 program (Huang and Madan 1999). The chimpanzee, orangutan, and rhesus macaque orthologs were identified by the reciprocal best hit principle: The putative non-human ortholog sequence was aligned to the human genome by using the BLAT server (http://www.genome.ucsc.edu/cgi-bin/hgBlat); if the best hit was the same as the starting human gene, the sequence was regarded as orthologous. Multiple sequence alignment analyses were performed by using the ClustalW program (Thompson et al. 1994). Allele frequencies for single nucleotide polymorphisms (SNPs) were obtained from the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP). The dotplot was prepared using the dotter program (Sonnhammer and Durbin 1995). The MEGA3 program was used for a phylogenetic analysis (Kumar et al. 2004).

To be confirmed as a human-specific nonsense mutation we required that (1) the orthology of the human and the chimpanzee genes is not ambiguous (possible paralogous alignments between similar members of the gene families and pseudogene-to-authentic gene matches are excluded), (2) the nonsense mutation is in a stable exon (stop codons in retained introns or in alternative *Alu* exons are excluded), (3) the orthologous chimpanzee protein sequence shows a significant sequence similarity over the entire length with at least one known non-human protein homolog, and (4) the orthologous orangutan and/or rhesus macaque genes have a non-stop codon as the chimpanzee gene does at the putative human-specific nonsense mutation position.

Results

The procedure for a systematic identification of human-specific nonsense mutations

The genome-wide identification of human-specific nonsense mutations involves a series of filtering steps applied to the publicly-available human mRNA-to-human genome alignment and human mRNA-to-chimpanzee genome alignment data. Table 1 shows notable steps of the procedure and the number of data kept at each step. There were 81,849 human mRNA sequences that aligned in both the human and the chimpanzee genome sequences. When the redundancy was removed, 22,853 stop codons of the human transcripts aligned with the respective orthologous sequences in the human and chimpanzee genomes. As expected, most of these (22,317 cases) had the same stop codon sequence in both human and chimpanzee genome. There were 330 cases in which the human stop codon was aligned to a chimpanzee non-stop codon. The non-human/non-chimpanzee vertebrate protein sequence database was searched for homologs of these sequences using BLAST. Of the 261 cases where a homolog was found in the third species, 96 cases showed an increase in the BLAST score by at least 1 bit when the chimpanzee-like protein sequence was used over when the human protein sequence was used.

As the final step, we performed comprehensive in-depth analysis on each candidate (see Materials and methods for detail) to identify nine human-specific nonsense mutations. These genes, and the nature of

Table 1 Number of sequences at each step of the procedure for identification of human-specific nonsense mutations

Step	Data count
1. Human mRNA sequences that align in both the human and the chimpanzee genomes	81,849
2. Alignments that contain the annotated last codon	69,034
3. Non-redundant last codon	23,274
4. The last codon is a stop codon in the human genome	22,853
The aligned chimpanzee codon is:	
a. the same stop codon as in human	22,317
b. different stop codon	149
c. ambiguous sequence	57
d. non-stop codon (analyzed in the next step)	330
i. 1 nt difference	293
ii. 2 nt differences	29
iii. 3 nt differences	8
5. Non-human/non-chimpanzee protein homolog found	261
6. BLAST score increases by at least 1 bit	96
7. Passed the manual inspection	9

mutations for each case, are listed in Table 2. These are distinct from the previously identified genes with a frameshift mutation. While this latter group of genes also has a premature stop codon, the stop codon sequence also exists in the corresponding chimpanzee sequence, although out of frame. These genes are, therefore, removed at step 4a (see Table 1) of the filtering process.

For each of the three genes, *CML2*, *MT1L*, and *SERPINA13*, the nonsense mutation probably leads to gene inactivation or production of a protein which lacks the original function since the protein produced would lack a major part of the active site (see below). The other six genes have the nonsense mutations near the C-termini and may encode modified but functional proteins. Analysis of all available sequence data suggests that the five genes (*CML2*, *FLJ14640*, *MT1L*, *NPPA*, *TAP2*) are polymorphic in humans. Sequence comparison of the region surrounding the human-specific stop codon mutation in human and orthologous sequences from chimpanzee and other closely related species is presented in Fig. 1. Sequence counts supporting the given sequence from various database sources are also provided in Fig. 1. When the mutation is known to be polymorphic, the estimated allele frequency is also given. Detailed descriptions of each case are given below.

CML2: putative N-acetyltransferase Camello 2

A cDNA sequence (accession no. BC069564) and the genome sequence of the *CML2* gene show two in-frame stop codons. Both of the human stop codons align with non-stop codons in the chimpanzee ortholog (Fig. 1a). Comparison of the human CML2 protein sequence with that of the yeast homolog GNA1, for which the crystal structure is known (Peneff et al. 2001) revealed that the second stop codon mutation occurred in the middle of the *N*-acetyltransferase catalytic domain, destroying the active site of the enzyme. Either of the two mutations will remove the active site and hence the mutant protein will lose the enzymatic function.

Sequence database search revealed a human mRNA sequence (accession no. NM_016347) that has neither of the two stop codons, suggesting that the human-specific nonsense mutation is polymorphic in human population. In fact, analysis of the available genomic sequences shows that the second stop codon position is polymorphic (see Fig. 1a). The polymorphism at the second stop codon is also present in the dbSNP database (accession

Table 2 Human-specific nonsense mutations

Gene	GenBank accession	Human protein[a] (aa)	Chimp protein (aa)	Mutation[b]	Mutated/ total number of exons	Human Chrom	Chimp Chrom[c]	Description
CML2	BC069564	15	227	W16XS, Q168XQ	2/2	2p13.1	2A (12)	Putative *N*-acetyltransferase Camello 2
FLJ14640	NM_032816	783	791	Q784XQ	19/19	19q13.11	19 (20)	Hypothetical protein
MT1L	AF348998	25	61	C26XC	2/3	16q12.2	16 (18)	Metallothionein 1L
NPPA	NM_006172	151	153	R152XR	3/3	1p36.22	1 (1)	Natriuretic peptide precursor A
PDE3B	NM_000922	1112	1113[d]	E1113X	16/16	11p15.2	11 (9)	Phosphodiesterase 3B, cGMP-inhibited
SERPINA13	NM_207378	307	394	R308X	4/5	14q32.13	14 (15)	Serine proteinase inhibitor, clade A, member 13
TAP2	BC002751	686	703	Q687XQ	12/12	6p21.32	6 (5)	Transporter 2, ATP-binding cassette, sub-family B
UIP1	AF267739	358	371	Q359X	10/10	Xq28	X (X)	26S proteasome-associated UCH interacting protein 1
ZNF277	NM_021994	438	444	Q439X	12/12	7q31.1	7 (6)	Zinc finger protein (C2H2 type) 277

[a] Length of the truncated human protein due to the nonsense mutation

[b] Amino acid difference: chimpanzee amino acid followed by the codon number followed by human amino acid (more than one amino acid residues in case of the polymorphism in humans); X, stop codon

[c] Chimpanzee chromosome number is based on the new numbering system (McConkey 2004). The old numbers are in parentheses

[d] Chimpanzee has a species-specific nonsense mutation (R1114X). Ancestral gene would encode 1117 amino acids

Fig. 1 Summary of the human-specific nonsense mutations. Nucleotide sequences and protein sequences surrounding the human-specific mutations of the nine genes are presented: **a** *CML2*, **b** *FLJ14640*, **c** *MT1L*, **d** *NPPA*, **e** *PDE3B*, **f** *SERPINA13*, **g** *TAP2*, **h** *UIP1*, and **i** *ZNF277*. Orthologous sequences from chimpanzee, orangutan, and rhesus macaque, and other homologous sequences are aligned. The human-specific and the ancestral stop codons are underlined. Lower case letters represent nucleotides in the 3′-untranslated region. Sequence counts show the numbers of database entries supporting the given sequence: genomic, genomic clone including bacterial artificial clone (*BAC*); *WGS* whole genome shotgun trace, *EST* expressed sequence tag. A blank means either that no sequence was found or that the analysis was not done. Two codon positions are shown for the genes *CML2* and *UIP1*. The second human sequences of the genes *CML2*, *FLJ14640*, *MT1L*, *NPPA*, and *TAP2* are for the respective minor alleles. Allele frequencies obtained from the dbSNP database, when available, are given in the right-most column. A human *MT1L* mRNA (X97261) was reported as *MT1R* gene

	Species	Gene	Accession	Codon number DNA	Protein	Sequence counts				dbSNP ID Allele frequency
				16		genomic	WGS	mRNA	EST	
a	human	CML2	BC069564	GACCGCAAGTAGGTCGTGGGC	DRK*VVG	2	9	1		
	human	CML2	NM_016347	GACCGCAAGTCGGTCGTGGGC	DRKWVVG			1		
	chimp	CML2	genome	GACCGCAAGTGGGTCGTGGGC	DRKWVVG		9			
	human	NAT8	NM_003960	GACCGCCAGTGGGTTGTGGGC	DRQWVVG					
	chimp	NAT8	genome	GACCGCCAGTGGGTTGTGGGC	DRQWVVG					
	orang	NAT8	CR859753	GACCGCAAGTGGGTCGTGGGC	DRKWVVG					
	rhesus	NAT8	WGS	GACCGCAAGTGGGTCGTGCGC	DRKWVVR					
	mouse	Cml4	NM_023455	GACTACAAACAGGTCGTGGAT	DYKQVVD					
				168		genomic	WGS	mRNA	EST	dbSNP:rs4852974
	human	CML2	BC069564	GCCCGGGACTAGGGCTACAGT	ARD*GYS	2	1	1		not determined
	human	CML2	NM_016347	GCCCGGGACCAGGGCTACAGT	ARDQGYS		2	1	8	not determined
	chimp	CML2	genome	GCCCGGGACCAGGGCTACAGT	ARDQGYS		5			
	human	NAT8	NM_003960	GCCCGGGACCAGGGCTACAGT	ARDQGYS					
	chimp	NAT8	genome	GCCCGGGACCAGGGCTACAGT	ARDQGYS					
	orang	NAT8	CR859753	GCCCGGGACCAGGGCTACAGT	ARDQGYS					
	rhesus	NAT8	WGS	GCCCGGGACCAGGACTGCAGT	ARDQDCS					
	mouse	Cml4	NM_023455	GCAAGGGACCAGGGTTACAGT	ARDQGYS					
				784		genomic	WGS	mRNA	EST	dbSNP:rs745961
b	human	FLJ14640	NM_032816	ACC---TGCTAGAATCTGCGG	T-C*NLR	1	7	2	7	0.754
	human	FLJ14640	WGS	ACC---TGCCAGAATCTGCGG	T-CQNLR		2		1	0.246
	chimp	FLJ14640	genome	ACC---TGCCAGAATCTGCAG	T-CQNLQ		5			
	orang	FLJ14640	WGS	ACC---TGCCAGAATCTGCGG	T-CQNLR		2			
	rhesus	FLJ14640	WGS	ACC---CGCCAGAGTCTGCGG	T-RQSLR		1			
	mouse	2610507L03Rik	NM_028120	ACCAACCGCCAGAGTCTGGGG	TNRQSLG					
				26		genomic	WGS	mRNA	EST	
c	human	MT1L	AF348998	GAGTGCAAATGAACCTCCTGC	ECK*TSC	3	14	2	20	
	human	MT1L (MT1R)	X97261	GAGTGCAAATGTACCTCCTGC	ECKCTSC			1		
	chimp	MT1L	genome	GAGTGCAAATGCACCTCCTGC	ECKCTSC		11			
	orang	MT1L	WGS	GAGTGCAAATGCACCTCCTAC	ECKCTSY		4			
	rhesus	MT1L	WGS	GAGTGCAAATGCACCTCGTGC	ECKCTSC		1			
	human	MT1A	NM_005946	GAGTGCAAATGCAACTCCTGC	ECKCNSC					
				152		genomic	WGS	mRNA	EST	dbSNP:rs5065
d	human	NPPA	NM_006172	TTCCGGTACTGAAGATAAcag	FRY*R*	4		1	61	0.827
	human	NPPA	BC005893	TTCCGGTACCGAAGATAAcag	FRYRR*		1	1	22	0.173
	chimp	NPPA	genome	TTCCGGTACCGAAGATAAcag	FRYRR*		12			
	orang	NPPA	WGS	TTCCGGTACCGAAGATAAcag	FRYRR*		2			
	rhesus	NPPA	WGS	TTCCGGTACCGAAGATAAcag	FRYRR*		4			
	mouse	Nppa	NM_008725	TTCCGGTACCGAAGATAAcag	FRYRR*					
	cat	NPPA	AF298813	TTCCGGTACCGAAGATAAtgg	FRYRR*					
	horse	NPPA	X58563	TTCCGGTACCGAAGATAAcag	FRYRR*					
				1113		genomic	WGS	mRNA	EST	
e	human	PDE3B	NM_000922	GAAGAGGAATAGCGACAGTTT	EEE*RQF	4	1	3	10	
	chimp	PDE3B	genome	GAAGAGGAAGAGTGACAGTTG	EEEE*QL		5			
	orang	PDE3B	WGS	GAAGAGGAAGAGCGACAGTTT	EEEERQF		6			
	rhesus	PDE3B	WGS	GAAGAGGAAGAGCGACAGTTT	EEEERQF		4			
	mouse	Pde3b	NM_011055	GAAGAGGAAGAACAAATGTTT	EEEEQMF					
	chicken	RCJMB04_11f1	AJ851613	GAAGAAGATGAGCGGCAGCTA	EEDERQL					
				308		genomic	WGS	mRNA	EST	
f	human	SERPINA13	NM_207378	CTCCTCCCATGAGTGACTGTG	LLP*VTV	1	4	1		
	chimp	SERPINA13	genome	CTCCTCCCACGAGTGACTGTG	LLPRVTV		9			
	orang	SERPINA13	WGS	CTCCTCGCACGAGTGACTGTG	LLARVTV		2			
	rhesus	SERPINA13	WGS	CTCCTCCCACAAGTGACTGTG	LLPQVTV		2			
				687		genomic	WGS	mRNA	EST	dbSNP:rs241448
g	human	TAP2	BC002751	GCCCAGCTCTAGGAGGGACAG	AQL*EGQ	7		5	20	0.812
	human	TAP2	NM_000544	GCCCAGCTCCAGGAGGGACAG	AQLQEGQ	4	1	7	10	0.188
	chimp	TAP2	genome	GCCCAGCTCCAGGAGGGACAG	AQLQEGQ		4			
	gorilla	TAP2	L49033	GCCCAGCTCCAGGAGGGACAG	AQLQEGQ			4		
	orang	TAP2	WGS	GCCCAGCTCCAGGAGGGACAG	AQLQEGQ		4			
	rhesus	TAP2	AC148669	GCCCAGCTCCAAGAGGGGCAG	AQLQEGQ	2				
	mouse	Tap2	NM_011530	GACCAGCTCAGGGACGGCCAG	DQLRDGQ					
				359		genomic	WGS	mRNA	EST	
h	human	UIP1	NM_017518	ATGGAGAAATAGAAAGTCTTC	MEK*KVF	2	3	10	72	
	chimp	UIP1	genome	ATGGAGAAACAGAAAGTCTTC	MEKQKVF		5			
	rhesus	UIP1	WGS	ATGGAGAAATACAGAGTCTTC	MEKYRVF		4			
	mouse	1110020L19Rik	NM_028633	ACACAGAAGTACAAGATCCTT	TQKYKIL					
				372		genomic	WGS	mRNA	EST	
	human	UIP1	NM_017518	AGCACAGGAT-Ggggcgggc-	STGwggq	2	3	10	72	
	chimp	UIP1	genome	AGCACAGGATAGgggcgggc-	STG*		6			
	orang	UIP1	WGS	AGCACAGGATAGtgggggc-	STG*		1			
	rhesus	UIP1	WGS	AGCACAGGATAGtgg-gggc-	STG*		4			
	mouse	1110020L19Rik	NM_028633	GGCACAAGATGAtgcagggct	GTR*					
				439		genomic	WGS	mRNA	EST	
i	human	ZNF277	NM_021994	TTGCTACTATAAGAGTACTTG	LLL*EYL	2	5	7	66	
	chimp	ZNF277	AC146143	TTGCTACTACAAGAGTACTTG	LLLQEYL	1	8			
	orang	ZNF277	WGS	TTGCTACTACAAGAGTACTTG	LLLQEYL		1			
	rhesus	ZNF277	WGS	TTGCTACTACAAGAGTGCTTG	LLLQECL		4			
	mouse	Zfp277	NM_178845	TTGCTACTCCAGGGGTGCCTG	LLLQGCL					

no. rs4852974), although the allele frequency has not been reported.

The coding region of the *CML2* gene shows 93% nucleotide sequence identity with the *NAT8* gene encoding an *N*-acetyltransferase (also known as *CML1* for Camello 1; accession no. NM_003960). Because the reported *CML2* mRNA sequences are derived from a single exon and contain two stop codons, *CML2* has been proposed to represent a processed pseudogene of *NAT8* (Zhang et al. 2003). However, *CML2* and *NAT8* are closely mapped in the chromosome band 2p13.1 and an examination of the dotplot (Supplementary Fig. S1) indicates that there is a pair of genomic duplicons of about 28 kb in length, one of which contains *NAT8* and the other *CML2*. The *CML2* gene seems to be composed of two exons as *NAT8* is, although no cDNA sequence with both exons of *CML2* has been reported. We collected *NAT8* and *CML2* orthologs from apes and monkeys to determine when the duplication occurred. The orangutan *NAT8* mRNA sequence is available in

GenBank (accession no. CR859753). We successfully identified chimpanzee *CML2* and *NAT8*, and rhesus macaque *NAT8* from the genome assembly and/or WGS trace data. A phylogenetic analysis based on a multiple alignment of the CML2 and NAT8 protein sequences (Supplementary Fig. S2) indicates that the duplication occurred after the ape-Old World monkey split (Supplementary Fig. S3). It implies that the ancestor of all extant great apes, including orangutans, gorillas, chimpanzees, as well as humans, had two copies of *N*-acetyltransferase 8. The exact copy number of *NAT8*-related genes in orangutan and gorilla is yet to be identified. It is expected that both copies are fully active in chimpanzee but one copy is inactive in some (probably most) human individuals.

The human *NAT8* gene encodes a kidney- and liver-specific *N*-acetyltransferase (Ozaki et al. 1998). It is proposed to be involved in drug metabolism in liver. The *Xenopus* camello (Xcml) protein is suggested to play some role in embryogenesis by modifying the cell surface and extracellular matrix proteins in the secretory pathway (Popsueva et al. 2001). Given that *CML2* is a recent genomic duplicate of *NAT8*, the two genes may share similar *cis*-regulatory elements and hence could exhibit a similar expression pattern. The biological significance of the inactivation of one of the two functionally redundant genes is not known.

FLJ14640: hypothetical protein FLJ14640

The human FLJ14640 gene has a human-specific nonsense mutation near the carboxyl terminus (Fig. 1b; Supplementary Fig. S4). The nonsense mutation is associated with T/C single nucleotide polymorphism in human population. The dbSNP record (accession no. rs745961) indicates that the T allele producing a stop codon is prevalent: 0.754 (T) versus 0.246 (C). The mutation removes eight amino acids from the carboxyl terminus: The ancestral allele and the chimpanzee ortholog encode 791 amino acids. The FLJ14640 protein is highly conserved across species including chicken and zebrafish. The protein is predicted to contain a coiled coil domain of about 500 amino acids in length (residues roughly from 220 to 720) and shows a sequence homology with carboxyl half of the human myosin XVIIA (MYO18A) (Supplementary Fig. S5). Its biological function and the phenotypic consequence of the nonsense mutation in human are unclear.

MT1L: metallothionein 1L

Metallothioneins are the heavy-metal binding proteins and function in the regulation of trace metal chemistry and in the detoxification of heavy metal ions (Vallee 1995). At least 13 class 1 metallothionein genes are clustered in the chromosome band 16q12.2. However, four of these genes, *MT1I*, *MT1J*, *MT1K*, and *MT1L* are suggested to be non-functional due to unacceptable amino acid substitutions or nonsense mutations (Stennard et al. 1994). We have compared the human *MT1L* gene with its putative ape and rhesus macaque orthologs and found that the nonsense mutation at codon 26 is human-specific (Fig. 1c). More than 30 sequence data, including 14 WGS and 20 EST clone sequences, agree with the genome sequence. Only one mRNA sequence (accession no. X97261) which was reported as *MT1R* gene has a non-stop codon sequence, raising the possibility of polymorphism in humans (Lambert et al. 1996).

We inferred possible effect on the metal binding capability of the human MT1L protein imposed by the nonsense mutation. As other metallothioneins, the chimpanzee MT1L has two cysteine-rich metal binding domains: β- (N-terminal 9 cysteines) and α-domain (C-terminal 11 cysteines) (Rigby et al. 2005). Each of the two domains can bind to up to three and four metal ions, respectively. The conserved cysteines are critical for the metal binding ability of metallothioneins. The human-specific nonsense mutation in *MT1L* gene removed two cysteines from the β-domain and the whole α-domain. The mutant protein, if produced, can only bind to one instead of seven metal ions, indicating that the original function of the protein is severely damaged in humans. Pseudogenization of some members in an isogenic gene cluster is rather frequently observed (Cooper et al. 2005; Gilad et al. 2003; Hesse et al. 2004). The putative orangutan MT1L turns out to be truncated also due to an orangutan-specific stop codon at a position different from that in the human MT1L (Supplementary Fig. S6). Thus, the presumed functional impairment of *MT1L* gene appears to be not entirely specific to humans only.

NPPA: natriuretic peptide precursor A

The human *NPPA* gene encodes a precursor protein for the atrial natriuretic peptide (ANP) that plays a central role in the regulation of blood pressure by promoting excretion of excessive salt and water (Levin et al. 1998). Many sequence polymorphisms have been reported for this gene. One of them is the "T2238C" mutation which occurs at the first position of the codon 152 (TGA to CGA). The T allele encodes a 151 amino acid-long precursor protein for the human ANP and is known as "normal," whereas the C allele is known as "mutant" which adds two additional arginines to the "normal" sequence. The reason of this designation seems to be due to the prevalence of the T allele in the contemporary human population: 0.827 (T) versus 0.173 (C) (dbSNP accession no. rs5065). However, sequence comparison between the human NPPA and its orthologs from other animals including chimpanzee, orangutan, rhesus macaque, mouse, rat, cat, horse, and cow reveals that the C allele is the ancestral allele (Fig. 1d; Supplementary Fig. S7). In humans, the ancestral CGA allele has been reported to be significantly associated with increased risk of stroke recurrence (Rubattu et al. 2004). The derived TGA allele, which produces a peptide hormone that lacks two highly conserved C-terminal arginines, might be advantageous over the ancestral allele CGA in response to some unknown change in systemic salt physiology in humans.

PDE3B: phosphodiesterase 3B, cGMP-inhibited

The human *PDE3B* gene encodes a cGMP-inhibited cyclic nucleotide phosphodiesterase that modulates cyclic nucleotide signaling in adipose tissue (Miki et al. 1996). Human PDE3B protein is 1,112 amino acids in length. Comparison of the human last exon (exon 16) sequence with orthologous sequences from chimpanzee, orangutan, rhesus macaque, mouse, and chicken reveals the human-specific GAG to TAG mutation at the codon 1,113 (Fig. 1e; Supplementary Fig. S8). If the stop codon is reverted to the ancestral sequence, the human *PDE3B* gene would encode a protein with 1,117 amino acids. Interestingly, the chimpanzee *PDE3B* gene also exhibits a species-specific nonsense mutation at the codon 1,114 (CGA to TGA). All WGS clone sequences currently available supports this mutation in chimpanzee, rejecting a chance of sequencing error in the genome assembly. The biochemical and physiological effect of these species-specific mutations remains to be determined.

SERPINA13: serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 13

The serpin genes encode serine proteinase inhibitors involved in diverse biological functions (Silverman et al. 2001). The human *SERPINA13* (also known as *kallistatin-like* or *KAL-like*) is a member of the serpin cluster at the chromosome band 14q32.13 where 11 serpin genes are found (Marsden and Fournier 2005). The gene encodes a protein with 307 amino acids. It is composed of 5 exons with the stop codon in exon 4. Comparison of the exon 4 sequences of the gene from human, chimpanzee, orangutan, and rhesus macaque indicates that the stop codon is a result of human-specific mutation that might have been fixed in the human lineage after the human-chimpanzee divergence (Fig. 1f). The predicted full-length chimpanzee SERPINA13 is 394 amino acids in length and has a highly conserved amino acid sequence of clade A serpins (Supplementary Fig. S9). The truncated human SERPINA13 would not be functional at least as a proteinase inhibitor since it loses the active loop which is critical for the proteinase inhibition activity. The original biological function, which is presumably lost in humans, of the ancestral SERPINA13 is unknown. It might have been secreted from the liver because mRNA from this gene has been detected in human liver (Marsden and Fournier 2005).

TAP2: transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)

The transporter associated with antigen processing (TAP) is a heterodimer consisting of TAP1 and TAP2 and plays a pivotal role in antigen presentation by translocating processed peptides from cytosol into the endoplasmic reticulum (Kelly et al. 1992). Many sequence polymorphisms have been reported for the human *TAP2* gene, including a well-known stop codon polymorphism (CAG to TAG) at codon 687 (Colonna et al. 1992; Powis et al. 1992). The dbSNP database record (accession no. rs241448) indicates that the stop codon allele is more frequent among humans: 0.812 (T) versus 0.188 (C). Sequence comparison reveals that all other great apes (chimpanzee, gorilla, and orangutan) have non-stop codon sequence CAG at this position (Fig. 1g; Supplementary Fig. S10). It confirms that the stop codon allele arose in a human ancestor and was selected over the non-stop codon allele. To our knowledge, there has been no report of any selective advantage of the stop codon allele.

UIP1: 26S proteasome-associated UCH interacting protein 1

The human UIP1 protein interacts with ubiquitin carboxyl-terminal hydrolase L5 (UCHL5; also known as UCH37) (Li et al. 2001). Its precise role in the biological process is not known but it has been hypothesized that UIP1 may regulate the ubiquitinated protein turnover by hindering the association of UCHL5 with the 26S proteasome (Li et al. 2001). We found two human-specific mutations of the ancestral coding sequence (Fig. 1h; Supplementary Fig. S11). One is the CAG-to-TAG mutation at codon 359 which leads to a premature termination compared with the ancestral protein of 371 amino acids in length. The other is the single adenine nucleotide deletion (TAG to TG) within codon 372, which would abolish the ancestral stop codon and induce out-of-frame translation through the entire 3′ untranslated region if the preceding nonsense mutation did not happen.

ZNF277: zinc finger protein (C2H2 type) 277

ZNF277 is a C2H2 type zinc finger-containing protein which is highly conserved even in *Caenorhabditis elegans* (Supplementary Fig. S12) (Liang et al. 2000). Comparison of the exon 12 sequences from human, chimpanzee, orangutan, rhesus macaque, and mouse reveals that the human ZNF277 has a nonsense mutation at codon 439 (CAA to TAA) (Fig. 1i). The mutation leads to production of six amino acid-less polypeptide in human cells compared with chimpanzee ZNF277 (444 amino acids). The human ZNF277 is predicted to contain 5 C2H2 type zinc fingers and a 30 amino acid coiled coil domain which are preserved in all orthologous proteins. These domains are known to mediate DNA binding and protein–protein interaction, respectively, suggesting that ZNF277 might be a transcriptional regulator. It is probable that the six-residue truncation may not deter the protein from functioning in human cells particularly because the C-terminus is less conserved compared with other parts. However, given that ZNF277 might be a transcriptional regulator, the slight modification of the protein could result in a significant alteration of downstream gene expression.

Discussion

Destructive change such as frameshift and nonsense mutation in a coding sequence of a gene usually results in a genetic disease caused by the production of malfunctioning protein or by the loss of the gene product (Caputi et al. 2002; Kang et al. 1997). However, a substantial number of genes that had been modified by frameshift or nonsense mutation specifically in human lineage have been reported (Angata et al. 2004; Chou et al. 2002; Dolphin et al. 1998; Hahn and Lee 2005; Hamann et al. 2003; Stedman et al. 2004; Winter et al. 2001). Furthermore, a couple of the gene inactivation cases have been proposed to play some role in the development of human-specific traits (Chou et al. 2002; Stedman et al. 2004). These observations indicate that some destructive mutations can be tolerated or even be constructive for evolution of species-specific phenotypes (Olson 1999). In this study, we present nine evolutionarily conserved genes that acquired stop codon mutation in the human genome. The phenotypic effects of most of these truncated gene products are unclear and yet to be examined. These nonsense mutations, together with previously reported human-specific frameshift mutations (Hahn and Lee 2005), provide additional opportunities for studying the effect of genetic alterations that distinguish humans from chimpanzees.

Of the nine cases reported here, the gene is probably inactivated in three due to the loss of the active site (Table 2). These are *CML2*, which is a genomic duplicate of *NAT8*, and *MT1L* and *SERPINA13*, which are members of the metallothionein and the clade A serpin clusters of genes, respectively. Lineage-specific pseudogenization of some members of gene clusters harboring functionally overlapping genes is rather common. For example, humans have much higher fraction of olfactory receptor pseudogenes compared with apes (Gilad et al. 2003). However, in the case of the *CML2*, *MT1L*, and *SERPINA13* genes, we cannot rule out the possibility that these genes had a dosage-dependent or isogene-specific function, which became lost in humans.

Each of the six other genes with a nonsense mutation reported in the present study appears to produce a functional protein since the mutation results in a loss of only a small number of amino acid residues from its C-terminus (Table 2). The number of amino acids removed ranges from 2 (*NPPA*) to 17 (*TAP2*). Although the change is small, the effect could be large. For example, the human atrial natriuretic peptide (ANP) encoded by the *NPPA* gene is shorter by only two amino acids compared with the ancestral form: The mature human and chimpanzee peptides are 28 and 30 amino acids, respectively, in length (Supplementary Fig. S7). The stop codon mutation of the *NPPA* gene is polymorphic, allowing a direct comparison of the physiological influence of the two peptides with different lengths. The ancestral CGA codon allele that produces a 30 amino acid-long ANP has been reported to show a significant association with cardiovascular diseases compared with the human-specific stop codon allele, which is more prevalent in human populations (Gruchala et al. 2003; Rubattu et al. 2004). It is interesting that the shorter ANP is advantageous in humans because all known ANP peptides in other organisms have the additional two arginines at their C-termini. It is not known why other organisms do not suffer from the harmful effect of the longer isoform as humans do.

Five genes described in the present study are known or predicted to show a stop codon polymorphism in humans (Table 2): *CML2*, *FLJ14640*, *MT1L*, *NPPA*, and *TAP2*. Allele frequencies reported in dbSNP database and sequence count data from genomic, mRNA, WGS, and EST clones indicate that the stop codon allele is more prevalent than the ancestral non-stop codon allele in each case (Fig. 1). The stop codon allele must have arisen in a certain human ancestor and must have been spreading across human individuals. The stop codon allele appears to be approaching fixation in the human population, although it is also possible that the polymorphisms are maintained by balancing natural selection. There are emerging examples of very recently arisen genetic changes that are being positively selected in local populations in response to biological, environmental, or cultural influences (Balter 2005). Even some stop codon mutations that lead to a functional gene loss exhibit beneficial effect on the individual carrying the mutant allele. For example, the stop codon polymorphism of Toll-like receptor 5 (TLR5) is associated with protection from the development of systemic lupus erythematosus (Hawn et al. 2005).

We present 10 nonsense mutations in 9 genes in this study (see Fig. 1). Eight of these are C:G to T:A transitions (7 C to T and 1 G to A) and two are C:G to A:T transversions (1 C to A and 1 G to T). The substitution pattern agrees well with the observation that hydrolytic deamination of cytosine generating uracil is the major cause of single nucleotide change (Pearl 2000). Of the eight C:G to T:A transitions, two are associated with CG to TG mutation and four are CAG to TAG mutation, suggesting that the methylation-mediated deamination of 5-methylcytosine of CpG dinucleotide and CpNpG trinucleotide may be responsible for these mutations (Clark et al. 1995; Krawczak et al. 1998).

In the present study, we identified ancestral genes on the basis of protein sequence homology to distantly related species. The existence of the distant homologs makes it likely that the gene has been conserved through a long period of evolution and therefore must play an important role in an essential biological process, which might be specifically modified in human lineage. Genome sequencing of orangutan, rhesus, and other primates are currently under way. For this study, we used orangutan and/or rhesus macaque WGS trace data for verification of the final candidates. When high-quality and high-coverage genome sequences become available, one can directly compare these genomes to find human-specific genetic alterations without the supporting evidence from the distantly related species.

Acknowledgments This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

Angata T, Margulies EH, Green ED, Varki A (2004) Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. Proc Natl Acad Sci USA 101:13251–13256

Balter M (2005) Evolutionary genetics. Are humans still evolving? Science 309:234–237

Caputi M, Kendzior RJ Jr, Beemon KL (2002) A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer. Genes Dev 16:1754–1759

Carroll SB (2005) Evolution at two levels: on genes and form. PLoS Biol 3:e245

Chou HH, Hayakawa T, Diaz S, Krings M, Indriati E, Leakey M, Paabo S, Satta Y, Takahata N, Varki A (2002) Inactivation of CMP-*N*-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. Proc Natl Acad Sci 99:11736–11741

Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferriera S, Wang G, Zheng X, White TJ, Sninsky JJ, Adams MD, Cargill M (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science 302:1960–1963

Clark SJ, Harrison J, Frommer M (1995) CpNpG methylation in mammalian cells. Nat Genet 10:20–27

Colonna M, Bresnahan M, Bahram S, Strominger JL, Spies T (1992) Allelic variants of the human putative peptide transporter involved in antigen processing. Proc Natl Acad Sci 89:3932–3936

Cooper SJ, Wheeler D, De Leo A, Cheng JF, Holland RA, Marshall Graves JA, Hope RM (2005) The mammalian α^D-globin gene lineage and a new model for the molecular evolution of α-globin gene clusters at the stem of the mammalian radiation. Mol Phylogenet Evol (in press)

Dolphin CT, Beckett DJ, Janmohamed A, Cullingford TE, Smith RL, Shephard EA, Phillips IR (1998) The flavin-containing monooxygenase 2 gene (FMO2) of humans, but not of other primates, encodes a truncated, nonfunctional protein. J Biol Chem 273:30599–30607

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418:869–872

Evans PD, Anderson JR, Vallender EJ, Gilbert SL, Malcom CM, Dorus S, Lahn BT (2004) Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. Hum Mol Genet 13:489–494

Fondon JW III, Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. Proc Natl Acad Sci USA 101:18058–18063

Gilad Y, Man O, Paabo S, Lancet D (2003) Human specific loss of olfactory receptor genes. Proc Natl Acad Sci 100:3324–3327

Goodman M, Grossman LI, Wildman DE (2005) Moving primate genomics beyond the chimpanzee genome. Trends Genet 21:511–517

Gruchala M, Ciecwierz D, Wasag B, Targonski R, Dubaniewicz W, Nowak A, Sobiczewski W, Ochman K, Romanowski P, Limon J, Rynkiewicz A (2003) Association of the ScaI atrial natriuretic peptide gene polymorphism with nonfatal myocardial infarction and extent of coronary artery disease. Am Heart J 145:125–131

Hahn Y, Lee B (2005) Identification of nine human-specific frameshift mutations by comparative analysis of the human and the chimpanzee genome sequences. Bioinformatics 21:i186-i194

Hamann J, Kwakkenbos MJ, de Jong EC, Heus H, Olsen AS, van Lier RA (2003) Inactivation of the EGF-TM7 receptor EMR4 after the Pan-Homo divergence. Eur J Immunol 33:1365–1371

Hawn TR, Wu H, Grossman JM, Hahn BH, Tsao BP, Aderem A (2005) A stop codon polymorphism of Toll-like receptor 5 is associated with resistance to systemic lupus erythematosus. Proc Natl Acad Sci

Hesse M, Zimek A, Weber K, Magin TM (2004) Comprehensive analysis of keratin gene clusters in humans and rodents. Eur J Cell Biol 83:19–26

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

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

Kang S, Graham JM Jr, Olney AH, Biesecker LG (1997) GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome. Nat Genet 15:266–268

Kelly A, Powis SH, Kerr LA, Mockridge I, Elliott T, Bastin J, Uchanska-Ziegler B, Ziegler A, Trowsdale J, Townsend A (1992) Assembly and function of the two ABC transporter proteins encoded in the human major histocompatibility complex. Nature 355:641–644

Krawczak M, Ball EV, Cooper DN (1998) Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. Am J Hum Genet 63:474–488

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Lambert E, Kille P, Swaminathan R (1996) Cloning and sequencing a novel metallothionein I isoform expressed in human reticulocytes. FEBS Lett 389:210–212

Levin ER, Gardner DG, Samson WK (1998) Natriuretic peptides. N Engl J Med 339:321–328

Li T, Duan W, Yang H, Lee MK, Bte Mustafa F, Lee BH, Teo TS (2001) Identification of two proteins, S14 and UIP1, that interact with UCH37. FEBS Lett 488:201–205

Li WH, Saunders MA (2005) The chimpanzee and us. Nature 437:50–51

Liang H, Guo W, Nagarajan L (2000) Chromosomal mapping and genomic organization of an evolutionarily conserved zinc finger gene ZNF277. Genomics 66:226–228

Marsden MD, Fournier RE (2005) Organization and expression of the human serpin gene cluster at 14q32.1. Front Biosci 10:1768–1778

McConkey EH (2004) Orthologous numbering of great ape and human chromosomes is essential for comparative genomics. Cytogenet Genome Res 105:157–158

Miki T, Taira M, Hockman S, Shimada F, Lieman J, Napolitano M, Ward D, Taira M, Makino H, Manganiello VC (1996) Characterization of the cDNA and gene encoding human

PDE3B, the cGIP1 isoform of the human cyclic GMP-inhibited cyclic nucleotide phosphodiesterase family. Genomics 36:476–485

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, J JS, Adams MD, Cargill M (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3:e170

Olson MV (1999) When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet 64:18–23

Ozaki K, Fujiwara T, Nakamura Y, Takahashi E (1998) Isolation and mapping of a novel human kidney- and liver-specific gene homologous to the bacterial acetyltransferases. J Hum Genet 43:255–258

Pearl LH (2000) Structure and function in the uracil-DNA glycosylase superfamily. Mutat Res 460:165–181

Peneff C, Mengin-Lecreulx D, Bourne Y (2001) The crystal structures of Apo and complexed *Saccharomyces cerevisiae* GNA1 shed light on the catalytic mechanism of an amino-sugar N-acetyltransferase. J Biol Chem 276:16328–16334

Perry GH, Verrelli BC, Stone AC (2005) Comparative analyses reveal a complex history of molecular evolution for human MYH16. Mol Biol Evol 22:379–382

Popsueva AE, Luchinskaya NN, Ludwig AV, Zinovjeva OY, Poteryaev DA, Feigelman MM, Ponomarev MB, Berekelya L, Belyavsky AV (2001) Overexpression of camello, a member of a novel protein family, reduces blastomere adhesion and inhibits gastrulation in *Xenopus laevis*. Dev Biol 234:483–496

Powis SH, Mockridge I, Kelly A, Kerr LA, Glynne R, Gileadi U, Beck S, Trowsdale J (1992) Polymorphism in a second ABC transporter gene located within the class II region of the human major histocompatibility complex. Proc Natl Acad Sci 89:1463–1467

Preuss TM, Caceres M, Oldham MC, Geschwind DH (2004) Human brain evolution: insights from microarrays. Nat Rev Genet 5:850–860

Rigby KE, Chan J, Mackie J, Stillman MJ (2006) Molecular dynamics study on the folding and metallation of the individual domains of metallothionein. Proteins 62:159–172

Rubattu S, Stanzione R, Di Angelantonio E, Zanda B, Evangelista A, Tarasi D, Gigante B, Pirisi A, Brunetti E, Volpe M (2004) Atrial natriuretic peptide gene polymorphisms and risk of ischemic stroke in humans. Stroke 35:814–818

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

Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin PB, Gettins PG, Irving JA, Lomas DA, Luke CJ, Moyer RW, Pemberton PA, Remold-O'Donnell E, Salvesen GS, Travis J, Whisstock JC (2001) The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. J Biol Chem 276:33293–33296

Sonnhammer EL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167:GC1–10

Stedman HH, Kozyak BW, Nelson A, Thesier DM, Su LT, Low DW, Bridges CR, Shrager JB, Minugh-Purvis N, Mitchell MA (2004) Myosin gene mutation correlates with anatomical changes in the human lineage. Nature 428:415–418

Stennard FA, Holloway AF, Hamilton J, West AK (1994) Characterisation of six additional human metallothionein genes. Biochim Biophys Acta 1218:357–365

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

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Vallee BL (1995) The function of metallothionein. Neurochem Int 27:23–33

Winter H, Langbein L, Krawczak M, Cooper DN, Jave-Suarez LF, Rogers MA, Praetzel S, Heidt PJ, Schweizer J (2001) Human type I hair keratin pseudogene ϕhHaA has functional orthologs in the chimpanzee and gorilla: evidence for recent inactivation of the human gene after the Pan-Homo divergence. Hum Genet 108:37–42

Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859–1875

Zhang Z, Harrison PM, Liu Y, Gerstein M (2003) Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res 13:2541–2558