doi:10.1038/nature05113



ARTICLES

# An RNA gene expressed during cortical development evolved rapidly in humans

Katherine S. Pollard[1]*†, Sofie R. Salama[1,2]*, Nelle Lambert[4,5], Marie-Alexandra Lambot[4], Sandra Coppens[4], Jakob S. Pedersen[1], Sol Katzman[1], Bryan King[1,2], Courtney Onodera[1], Adam Siepel[1]†, Andrew D. Kern[1], Colette Dehay[6], Haller Igel[3], Manuel Ares Jr[3], Pierre Vanderhaeghen[4] & David Haussler[1,2]

**The developmental and evolutionary mechanisms behind the emergence of human-specific brain features remain largely unknown. However, the recent ability to compare our genome to that of our closest relative, the chimpanzee, provides new avenues to link genetic and phenotypic changes in the evolution of the human brain. We devised a ranking of regions in the human genome that show significant evolutionary acceleration. Here we report that the most dramatic of these 'human accelerated regions', HAR1, is part of a novel RNA gene (*HAR1F*) that is expressed specifically in Cajal–Retzius neurons in the developing human neocortex from 7 to 19 gestational weeks, a crucial period for cortical neuron specification and migration. *HAR1F* is co-expressed with reelin, a product of Cajal–Retzius neurons that is of fundamental importance in specifying the six-layer structure of the human cortex. HAR1 and the other human accelerated regions provide new candidates in the search for uniquely human biology.**

The hallmark of evolutionary shift of function is sudden change in a region of the genome that previously has been highly conserved owing to negative selection. It has been speculated that changes of this type in *FOXP2*[1], a gene involved in speech production, and *ASPM*[2], which affects brain size, have had a significant role in the evolution of the human brain (reviewed in ref. 3). The vast majority of the approximately 15 million changes in our genome since our common ancestor with the chimpanzee are likely to represent neutral drift[4,5], so systematic searches for potentially important evolutionary acceleration have focused exclusively on protein coding regions[5–9], where there is a more favourable signal-to-noise ratio. However, protein coding regions account for only about one-third of the segments in the human genome thought to be under negative selection[10], and thus these searches may be missing the majority of the functional elements in the genome[5,11]. With the availability of nearly complete genome sequences for several vertebrates, comparative genomics can now be used to predict functional elements in the 98.5% of the genome that is non-coding, through patterns of ancestral negative selection[10,12–16]. Here we scan these ancestrally conserved genomic regions to find those that show a significantly accelerated rate of substitution in the human lineage since divergence from our common ancestor with the chimpanzee. Many of the human accelerated regions (HARs) found in this scan are associated with genes known to be involved in transcriptional regulation and neurodevelopment. HAR1, the most dramatically changed element, is part of a novel RNA gene expressed during human cortical development.

## Identification of human accelerated regions

To overcome the very low signal-to-noise ratio among changes in non-coding regions, we first searched independently for maximal-length regions of the chimpanzee genome with at least 96% identity over 100 base pairs (bp) with the orthologous regions in mouse and rat, suggesting a significant level of negative selection in these regions. For each of the approximately 35,000 such mammalian conserved regions (median length, 140 bp), we then examined the orthologous segments in all other available amniote genomes, looking for regions that have a large number of non-adjacent changes in human relative to other species. A likelihood ratio test was used to rank and evaluate the significance of each region. Controlling the genome-wide false discovery rate (FDR) to be less than 5%, we identified 49 regions with a statistically significant substitution rate increase in human. We compared the results of this analysis to those from several complementary approaches (see Supplementary Notes S1) and found that the fastest-evolving regions score uniformly high regardless of the method (Supplementary Notes S2). Of the 49 HARs, 96% are in non-coding segments (Supplementary Table S2). Gene Ontology terms related to DNA binding and transcriptional regulation are significantly enriched among the genes adjacent to the HARs (Supplementary Table S8), and 24% of the HARs are adjacent to a neurodevelopmental gene (Supplementary Table S7), making these regions especially interesting as candidates for brain-specific regulatory elements that may have changed significantly during human evolution.

## HAR1 lies in a pair of novel non-coding RNA genes

The 118-bp HAR1 region showed the most dramatically accelerated change (FDR-adjusted $P < 0.0005$), with an estimated 18 substitutions in the human lineage since the human–chimpanzee ancestor, compared with the expected 0.27 substitutions on the basis of the slow rate of change in this region in other amniotes (Supplementary Notes S3). Only two bases (out of 118) are changed between chimpanzee and chicken, indicating that the region was present

[1]Center for Biomolecular Science & Engineering, [2]Howard Hughes Medical Institute, and [3]Center for Molecular Biology of RNA, Department of Molecular, Cell & Developmental Biology, University of California, Santa Cruz, California 95064, USA. [4]Institut de Recherche Interdisciplinaire en Biologie Humaine et Moléculaire (IRIBHM), Free University of Brussels (ULB), and [5]Department of Psychiatry, Erasme Hospital, Free University of Brussels (ULB), B-1070 Brussels, Belgium. [6]INSERM, U371, Stem Cell & Brain Research Institute, PrimaStem, Université Claude Bernard, Lyon 1, 69575 Bron Cedex, France. †Present addresses: Department of Statistics and UC Davis Genome Center, University of California, Davis, California 95616, USA (K.S.P.); Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853, USA (A.S.).
*These authors contributed equally to this work.

and functional in our ancestor at least 310 million years (Myr) ago. No orthologue of HAR1 was detected in the frog (*Xenopus tropicalis*), any of the available fish genomes (zebrafish, *Takifugu* and *Tetraodon*), or in any invertebrate lineage, indicating that it originated no more than about 400 Myr ago[17]. No paralogues were detected in any amniote genome draft. Resequencing in four primates further confirms that all 18 substitutions are very likely to have occurred in the human lineage (Supplementary Fig. S2). Checking against available single nucleotide polymorphisms (SNPs; http://www.ncbi.nlm.nih.gov/SNP/)[18] and resequencing this region in a 24-person diversity panel[19] indicates that all 18 substitutions are fixed in the human population. Evidence from preliminary resequencing of a 6-kilobase (kb) region containing HAR1 shows levels of polymorphism and a positive skew in the frequency spectrum that are typical of European samples[20], suggesting that a recent selective sweep in this region of the genome is unlikely (Supplementary Notes S4). Thus, the changes in HAR1 clearly occurred on the human lineage, but probably took place more than 1 Myr ago.

HAR1 lies in the last band of chromosome 20q and is part of a pair of overlapping divergently transcribed genes, *HAR1F* and *HAR1R*, supported by expressed sequence tag (EST) evidence. Analysis by polymerase chain reaction with reverse transcription (RT–PCR), and 5′ and 3′ rapid amplification of complementary DNA ends (RACE), revealed a two-exon *HAR1F* gene and an alternatively spliced *HAR1R* gene with isoforms of two and three exons (Fig. 1, Supplementary Notes S5). We sequenced the entire *HAR1F* genomic region in chimpanzee, gorilla, orang-utan and crab-eating macaque (Supplementary Notes S4, Supplementary Fig. S3). The very high divergence of the non-human primates from the human reference found in HAR1 extends through most of the first *HAR1F* exon. In addition to the primates, the dog and cow genomes also contain *HAR1F* orthologues, whereas only the HAR1 region, and not the entire transcript, can be aligned to other available amniote genomes. Most *HAR1R* regions are highly divergent, except in the chimpanzee and macaque genomes.

We have also cloned HAR1-containing transcripts from mice (Supplementary Notes S6). As in human, we found overlapping transcripts that contain HAR1. As expected by the poor conservation in the surrounding genomic region, the homology between the human and mouse HAR1 transcripts is limited to an approximately 180-bp region surrounding the HAR1 segment. Thus, although the region of overlap between human *HAR1F* and *HAR1R* (including the HAR1 segment) was conserved from our common ancestor with chicken to our common ancestor with chimpanzee, the longer *HAR1F* and *HAR1R* transcripts have evolved much more rapidly during this period.

### The *HAR1F* RNA forms a stable structure

Protein coding potential of the mouse and human *HAR1F* and *HAR1R* RNAs is poor or absent (Supplementary Notes S7). However, a phylogenetic stochastic context-free grammar model for RNA evolution[12] gives a strong prediction that the HAR1 region of the *HAR1F* transcript can fold into a stable RNA structure (Fig. 2), and a slightly less-strong prediction for the same region of the *HAR1R* transcript (data not shown). Three of five RNA helices in the predicted secondary structure are statistically confirmed by five pairs of human-specific compensatory changes that preserve Watson–Crick pairing: four pairs of transitions and one pair of transversions ($P = 3.95 \times 10^{-8}$; Supplementary Notes S8). A different compensatory pair of transversions is also found in the platypus genome. No changes inconsistent with the core of any helix, excluding the terminal base pairs and the second pair of the outermost enclosing helix in mouse, were found in any sequence match from the National Center for Biotechnology Information (NCBI) trace repository (one match from each of 11 amniote species; Supplementary Fig. S2).

We tested this secondary structure prediction by performing structure probing experiments using dimethyl sulphate (DMS) treatment and primer extension of *in vitro* synthesized human and chimpanzee *HAR1F* RNAs (Supplementary Notes S9)[21]. The results were largely consistent with the predicted structure, but we observed several differences between the human and chimpanzee RNAs (Fig. 2c and Supplementary Fig. S10). In particular, our results are consistent with a human-specific structure where helix D is extended at the expense of helix C (compare Fig. 2b to Fig. 2c). In contrast, excluding the human sequence from the comparative analysis leads to an evolutionarily conserved structure with three additional base pairs in helix C and two fewer pairs in helix D. This structure agrees with the protection of bases in helix C and the accessibility of bases in the shorter helix D we observe for the chimpanzee RNA. In the human RNA, bases in helix D are strongly protected whereas those in helix C are more accessible by DMS. These results indicate that although selective forces to preserve base pairing interactions shaped this region, the structure of the *HAR1F* RNA may be significantly different between humans and their ancestors, and may even lack helices C and E. This RNA structure is novel; no primary or secondary structural homology was found with any RNA in the Rfam RNA repository[22], and no convincing evidence of microRNA (miRNA) precursors in these transcripts has yet been found (Supplementary Notes S10). On the basis of this analysis, it seems that *HAR1F* and *HAR1R* are novel non-coding RNA genes.

### *HAR1F* is expressed in the developing human and primate brain

To determine if *HAR1F* is expressed in the developing human central nervous system, we performed RNA *in situ* hybridizations on human embryonic brain sections. We found that *HAR1F* is strongly and specifically expressed in the developing neocortex early in human embryonic development (Fig. 3). *HAR1F* was first detected at 7 and 9 gestational weeks (GW), specifically in the dorsal telencephalon (the primordium of the cerebral cortex) and not detected in other parts of

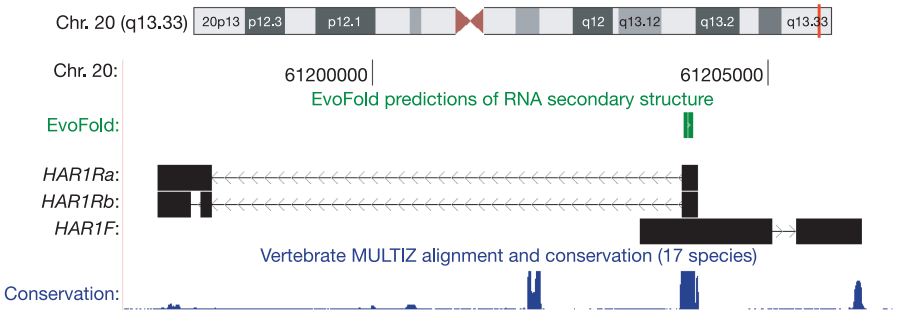


**Figure 1 | HAR1-associated transcripts in genomic context.** Schematic drawing showing the genomic context on chromosome 20q13.33 of the HAR1-associated transcripts *HAR1F* and *HAR1R* (black, with a chevroned line indicating introns), and the predicted RNA structure (green) based on the May 2004 human assembly in the UCSC Genome Browser[41]. The level of conservation in the orthologous region in other vertebrate species (blue) is plotted for this region using the PhastCons program[16]. Both the common and testes-specific splice sites are conserved (data not shown).

the forebrain (Fig. 3a and data not shown). Within the early developing cortex, *HAR1F* was selectively expressed in a subset of cells located close to the pial surface in the marginal zone (Fig. 3a, b), indicating its expression in Cajal–Retzius neurons. This was confirmed by co-detection of reelin protein, a specific marker of these neurons, and *HAR1F* RNA (Fig. 3a, c, d), which showed that both *HAR1F* and reelin are expressed in the same cells. In contrast, *HAR1F* did not seem to be expressed in interneurons expressing vesicular GABA (γ-aminobutyric acid) transporter (vGAT), the other main neuronal population of the marginal zone (Fig. 3d). The expression of HAR1 in reelin-positive neurons was maintained until at least 17–19 GW (Fig. 3c, d), at which stage reelin-positive Cajal–Retzius-like neurons are interspersed within the subpial granular layer (SGL)—a layer thought to contain Cajal–Retzius neurons and their precursors migrating tangentially from outside the neocortex[23–25], and which is uniquely developed in humans[24–27]. *HAR1F* was also expressed in cells of the upper cortical plate (Fig. 3b, c), presumably corresponding to neurons finishing their radial migration. Notably, the human reelin receptor gene, *VLDLR*, also shares this expression pattern in humans, but is not expressed in Cajal–Retzius neurons in the mouse[28]. At later stages (24 GW), expression of *HAR1F* was no longer detectable in Cajal–Retzius cells. At 17–24 GW, expression of *HAR1F* was observed in brain regions other than the cortex: in particular, the hippocampal primordium, the dentate gyrus, the developing cerebellar cortex, and a few hindbrain nuclei such as the olivar complex (Fig. 4). In contrast to the distinct pattern of *HAR1F* expression, probes specific for *HAR1R* show no specific staining in these embryonic sections (data not shown).

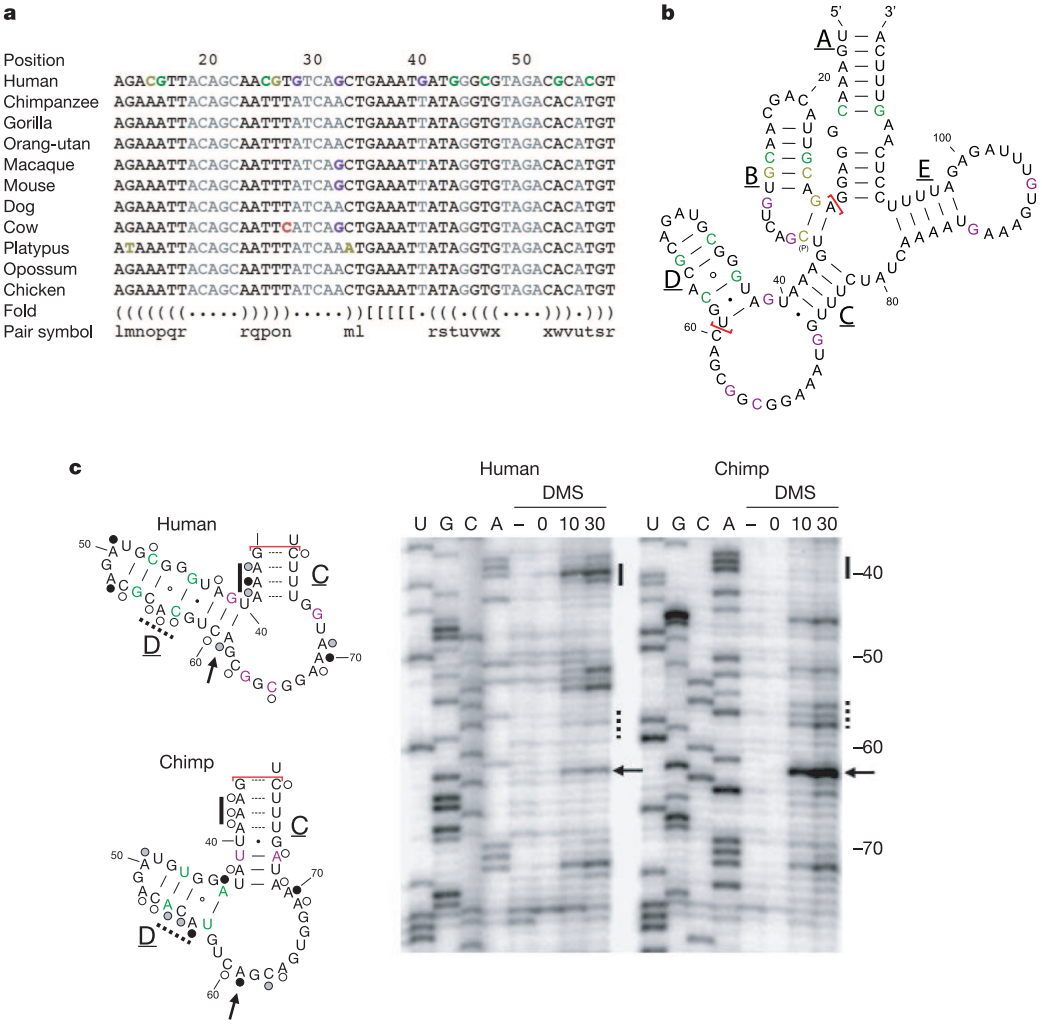


**Figure 2 | Predicted RNA secondary structure for *HAR1F*.** **a**, Section of the multiple alignment of *HAR1F* in various amniote species. The multiple alignment is annotated with the secondary structure (fold) shown in panel **b**. Matching round parentheses indicates pairing bases. Square parentheses are used to indicate bases that are predicted to pair outside the region shown. Unpaired regions are shown in grey and substitutions are shown with the following colour scheme: green denotes compensatory transitions, yellow–green denotes compensatory transversions, purple denotes substitutions in unpaired regions, and red denotes non-compensated changes. **b**, The evolutionarily conserved parts of the RNA secondary structure of the *HAR1F* region as predicted using the EvoFold program[12]. Substitutions are shown using the colour scheme in panel **a**; red bars indicate the region for which the alignment is shown in panel **a**. The structure is supported by substitutions on the human lineage, as well as a pair of changes in platypus (indicated by (P)). Only the helices with the compensatory substitutions can be considered to be supported by evolutionary data. **c**, Distinct secondary structure predictions for helices C and D in the human sequence (Human) versus non-human sequences (Chimp), on the basis of combined comparative sequence analysis and structure probing. Circles indicate the bases that can be modified by DMS and indicate strong (black), weak (grey) or no (white) modification as shown in the gel to the right. For each species, lanes 1–4 show the sequencing reactions labelled according to the corresponding base in the RNA structure; lanes 5–8 show the primer extension reactions of RNA after no (−), 0, 10 or 30 min of DMS treatment at 18 °C. Numbers on the right indicate the position in RNA structure; arrows indicate the modified base at position 61; solid lines indicate the position of bases 37–39; and dashed lines indicate the position of bases 54–57. The presence of a band in the last two lanes indicates modification of the base at the previous position (that is, the arrows at position 62 on the gel indicate a modified A at position 61).

We examined the evolutionary conservation of this expression pattern by performing RNA *in situ* hybridizations on brain sections of another primate, the cynomolgus macaque. For these experiments, we cloned a fragment of macaque DNA corresponding to most of *HAR1F* exon 1 and tested it on embryonic day (E) 79–85 samples, corresponding to midgestation and to stages of active cortical neuron production and radial migration[29]. The *HAR1F* expression pattern was very similar to that observed in the human sections (Fig. 3e). *HAR1F* is co-expressed with reelin in the SGL and is also expressed in cells of the upper cortical plate. Thus, despite extensive sequence changes, the expression pattern of *HAR1F* in the developing cortex has been highly conserved since the divergence of hominoids and Old World monkeys some 25 Myr ago. This suggests that the *HAR1F* expression pattern is functionally significant and is reminiscent of the evolution of *ASPM*, which shows rapid change in its coding sequence but conserved patterns of expression in neural progenitors[2,30].

## HAR1 genes are expressed in the adult brain

We also examined *HAR1F* and *HAR1R* expression in adult human samples by using tissue RNA preparations for northern blot and quantitative real-time PCR (qPCR) analysis (Table 1), as well as RNA *in situ* hybridization. *In situ* hybridization experiments on adult brain sections revealed a diffuse pattern of staining for *HAR1F* in the frontal cortex and hippocampus (Supplementary Fig. S12). Although no specific signal was detected for either transcript by northern blot analysis (data not shown), we detected expression of both transcripts by qPCR in total brain, ovary and testes samples. Neither transcript was detected in other tissues tested, including adrenal, bladder,

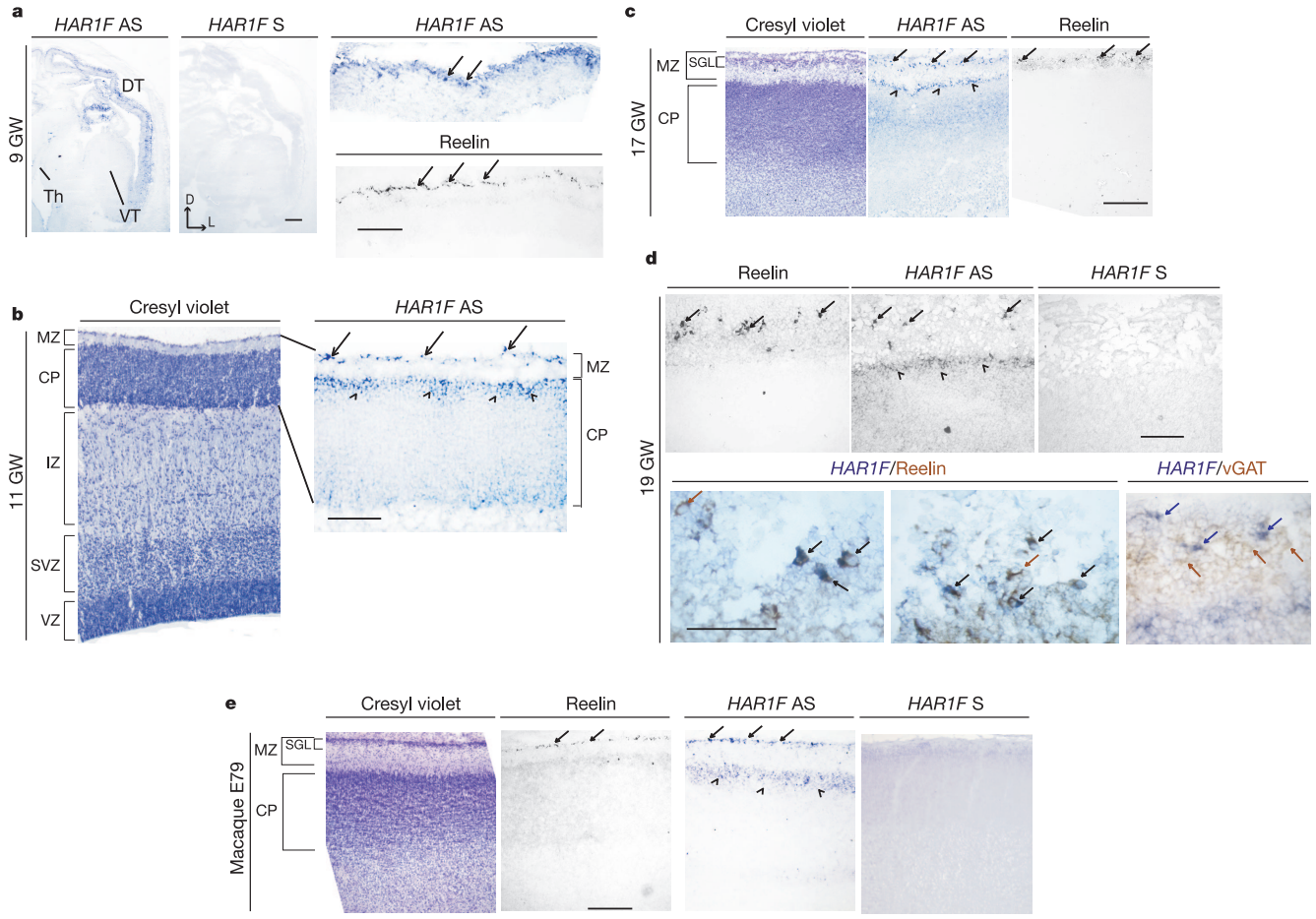


**Figure 3 | Expression of *HAR1F* and *HAR1R* in the developing neocortex.** Pattern of expression of *HAR1F* and reelin in the human and macaque developing brain, revealed by RNA *in situ* hybridization (*HAR1F*) and immunohistochemistry (reelin). **a**, *In situ* hybridization of *HAR1F* antisense (AS) and sense (S) probes on 9-GW human brain coronal sections, illustrating selective expression of *HAR1F* in the dorsal telencephalon (DT), whereas the ventral telencephalon (VT) and thalamus (Th) show no obvious expression. The HAR1 probes contain 112 bp of the region of overlap between *HAR1F* and *HAR1R*. Therefore, the *HAR1F* sense probe has the potential to detect *HAR1R* expression. Within the DT, *HAR1F* is selectively enriched in the most superficial part of the cortex, much like reelin at the same stage (arrows). Scale bars: 500 μm on left panels; 250 μm on right panels. Dorsal (D) is top; lateral (L) is right on left panels. **b**, At 11 GW, *HAR1F* expression is confined to the cerebral cortex, with expression in large cells situated at the border of the marginal zone (MZ) (arrows), corresponding to Cajal–Retzius neurons, and in the upper cortical plate (CP) (arrowheads). Scale bar: 125 μm for the right panel. IZ, intermediate zone; SVZ, subventricular zone; VZ, ventricular zone. **c**, At 17 GW, expression of *HAR1F* and reelin remain confined to Cajal–Retzius cells in the MZ (arrows), intermingled with other cells of the subpial granular layer (SGL). In addition, *HAR1F* is expressed in neurons of the upper CP (arrowheads). Scale bar: 1,000 μm. **d**, At 19 GW, *HAR1F* and reelin show the same pattern in the MZ/SGL (arrows), and *HAR1F* remains expressed in the upper CP (arrowheads), whereas the *HAR1F* sense probe shows no signal. Lower panels centred on the MZ illustrate double *in situ* hybridization/immunohistochemistry experiments. Lower left and middle panels show that *HAR1F* (in dark blue) and reelin (in dark brown) are co-localized in most cells, corresponding to Cajal–Retzius neurons (black arrows), although some reelin-positive cells do not express *HAR1F* (brown arrows). The lower right panel illustrates that *HAR1F* cells (blue arrows) seem to be mainly distinct from interneurons (brown arrows) expressing vesicular GABA transporter (vGAT). Scale bar: 250 μm. **e**, Expression of *HAR1F* in macaque developing cortex. *In situ* hybridization (*HAR1F*) and immunohistochemistry (reelin) on E79 macaque brain parasagittal sections. As in the human, *HAR1F* and reelin are expressed in Cajal–Retzius cells in the MZ (arrows), intermingled with other cells of the SGL. In addition, *HAR1F* is expressed in neurons of the upper CP (arrowheads). Scale bar: 1,000 μm. The *HAR1F* sense probe does not reveal any obvious expression pattern.

breast, colon, liver, pancreas, placenta, skeletal muscle and thymus. To further characterize the expression in brain, we obtained RNA samples from specific adult human brain regions as well as a fetal brain RNA sample. Expression of *HAR1F* is highest in cerebellum and is also prominent in forebrain structures, including the cortex, hippocampus, thalamus and hypothalamus. In general, *HAR1F* expression is higher than *HAR1R*. Testis is an exception; here transcripts are present in small, but roughly equal, amounts. It is interesting to note that *HAR1R* levels are dramatically lower than *HAR1F* (50-fold lower) in the fetal brain sample, consistent with our inability to detect it with *in situ* hybridization. The finding that both transcripts overlap and are expressed in the same tissues raises the possibility of antisense regulation between them. So far, however, we have not detected expression of the two transcripts in the same cells.

In a qPCR analysis of mouse *HAR1F* and *HAR1R* expression (Supplementary Notes S6), we found that both *HAR1F* and *HAR1R* are expressed in embryonic development, and the peak expression of both *HAR1F* and *HAR1R* was at E15. In contrast to our observations in human fetal brain samples, *HAR1F* expression was not significantly higher than *HAR1R* expression. In adult tissues, we found both *HAR1F* and *HAR1R* specifically expressed in brain regions, but again the levels of *HAR1F* and *HAR1R* were similar. These results suggest that the regulation of the HAR1 region transcripts may have changed since the common ancestor of mouse and human.

## HAR1 substitutions show a weak-to-strong bias

A curious feature of the evolution of HAR1 is that all 18 human-specific changes in the predicted RNA structure are from A or T ('W') nucleotides to G or C ('S') nucleotides. This has the effect of strengthening RNA helices in the predicted RNA structure. However, this pattern of highly biased substitutions extends well beyond the predicted RNA structure in a 1.2-kb region overlapping the first exon of *HAR1F* (34 W → S substitutions out of 44 total human–chimpanzee differences, with no S → W substitutions). Notably, a W → S substitution bias is present in many of the human accelerated regions. Among the 49 most accelerated, the frequency of W → S substitutions is 77% higher than that of S → W substitutions, even after a correction for variation in the overall base composition of the ancestral sequence, whereas there is no difference in the larger set of 35,000 conserved regions (Supplementary Notes S11). Compared with an expected <5%, we find that 12% of these 49 regions, including HAR1, lie in the final band of their chromosome arms, where there is an increased rate of recombination[14]. Recombination hotspots are known to differ in location between human and chimpanzee[31], and have been speculated to be associated with increases in W → S substitutions through a process known as biased gene conversion[32]. Selection for increased gene stability or increased expression levels have been proposed as alternative explanations for W → S substitution biases[33,34]. Bias towards W → S substitution through any of these general processes may have combined with selection for specific compensatory changes in the RNA structure to dramatically alter HAR1 during human evolution.

## Conclusion

One primary issue remaining concerns the role of *HAR1F* and *HAR1R* in cortical development. Although *HAR1F* is co-expressed with reelin—a known regulator of cortical development—within Cajal–Retzius cells of the SGL, it remains to be determined whether *HAR1F* influences the expression of reelin or its receptors either directly or indirectly, or whether it acts through a different pathway. Notably, the extra-cortical sites of expression of *HAR1F* (such as cerebellar cortex and olivar nuclei) also correspond to structures that are patterned by the reelin pathway in mouse and human[35]. A suggested association between reelin-expression-related defects in the brain and schizophrenia provides additional motivation for further investigation of this issue[36]. Additionally, the pattern of expression of *HAR1R* suggests, but does not prove, that it may be expressed later in development to downregulate *HAR1F* by antisense inhibition, as is the case for some other developmental genes[37].

Our comparative genomic approach to identify the most dramatically changed segments of DNA in the human genome has identified a number of new candidate regions to test for clues in the attempt to decode the key events in human evolution. The first on this list seems to define a new type of RNA molecule expressed at a very crucial time and place in the development of the neocortex. Many of the other top candidates are associated with genes known to be involved in neurodevelopment, an area where there has been significant divergence since our last common ancestor with chimpanzee. Thus, this seems to be a promising approach to identifying candidate regions involved in neurodevelopmental aspects of our uniquely human biology.

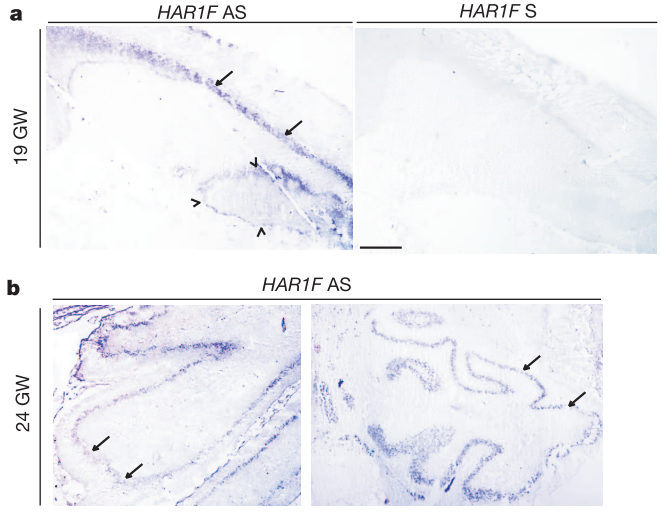


**Figure 4 | Expression of *HAR1F* in other parts of the developing brain.** Pattern of expression of *HAR1F* in the developing brain, revealed by RNA *in situ* hybridization. **a**, *In situ* hybridization on 19-GW coronal brain sections, illustrating expression of *HAR1F* in the hippocampal primordium (arrows) and dentate gyrus (arrowheads). *HAR1F* sense probe detects no obvious signal. Scale bar: 750 μm **b**, *In situ* hybridization on 24-GW coronal brain sections, illustrating expression of *HAR1F* in the cerebellar cortex (arrows in left panel) and olivar complex (arrows in right panel).

**Table 1 | Expression of human *HAR1F* and *HAR1R***

| Sample | *HAR1F** | *HAR1R** |
|---|---|---|
| Total RNA | | |
| Cerebral cortex | 1.0† (0.77, 1.3) | 0.095 (0.054, 0.17) |
| Cerebellum | 4.1 (2.4, 7.2) | 0.041 (0.030, 0.055) |
| Frontal lobe | 1.2 (0.70, 2.0) | 0.12 (0.067, 0.22) |
| Temporal lobe | 0.72 (0.55, 0.95) | 0.049 (0.029, 0.083) |
| Parietal lobe | 0.77 (0.61, 0.98) | 0.083 (0.052, 0.13) |
| Occipital pole | 1.1 (0.92, 1.3) | 0.12 (0.065, 0.21) |
| Insula | 0.91 (0.62, 1.3) | 0.078 (0.045, 0.13) |
| Hippocampus | 0.65 (0.44, 0.96) | 0.051 (0.031, 0.087) |
| Pons | 0.51 (0.35, 0.76) | 0.21 (0.11, 0.40) |
| Medulla oblongata | 0.39 (0.27, 0.56) | 0.043 (0.025, 0.074) |
| Fetal brain | 0.14 (0.11, 0.18) | 0.003 (0.002, 0.005) |
| Brain | 0.96 (0.71, 1.3) | 0.024 (0.015, 0.039) |
| Testes | 0.031 (0.006, 0.16) | 0.12 (0.047, 0.31) |
| Ovary | 0.19 (0.15, 0.25) | 0.009 (0.006, 0.012) |
| PolyA$^+$ RNA‡ | | |
| Thalamus | 4.9 (3.2, 7.5) | 0.51 (0.27, 0.98) |
| Hypothalamus | 4.8 (3.5, 6.6) | 0.54 (0.30, 0.96) |

* Values are expressed as a ratio of the expression in the cerebral cortex sample of *HAR1F*. Entries represent the averages and 95% confidence intervals (in parentheses) of six replicates from two independent experiments, with the exception of brain *HAR1R*, which represents nine replicates from two independent experiments.
† A value of 1.0 is equivalent to 24,014 copies in 100 ng of template cDNA.
‡ Approximately one-third the amount of template was used for polyA$^+$-derived samples. It is expected that expression will be higher in polyA$^+$-enriched samples than in samples derived from total RNA.

## METHODS

Detailed methods are given in Supplementary Information. Using alignments produced by MULTIZ (http://www.bx.psu.edu/miller_lab/), we identified 34,498 conserved regions of the chimpanzee genome that are ≥100 bp long and ≥96% identical with mouse and rat. Each conserved region was evaluated for acceleration in the human lineage using a likelihood ratio test (LRT) using DNA from other species. Statistical significance was assessed by simulation and *P*-values are adjusted for multiple comparisons.

For RT–PCR, RACE and qPCR analysis, RNAs (Ambion and Clontech) were used to synthesize cDNA primed with a combination of oligo-dT and random hexamer primers using Superscript III (Invitrogen) according to the manufacturer's recommendations.

The qPCR analysis used iQSYBR Green Supermix (BioRad) with 2.5 ng $\mu l^{-1}$ (or 0.8 ng $\mu l^{-1}$ for polyA$^+$-derived samples) cDNA template in an i-Cycler thermal cycler (BioRad); RACE analysis used the GeneRacer kit (Invitrogen). Secondary structure models were based on an EvoFold prediction[12], and chemical probing of unpaired A and C residues with dimethyl sulphate (DMS) was carried out using methods described previously[21].

RNA *in situ* hybridization using digoxigenin-labelled RNA probes adapted to human brain specimens was performed as described previously[38]. Immunodetection was performed as described[39] using mouse monoclonal anti-reelin antibody 142 (ref. 40; kindly provided by A. Goffinet) and a rabbit antibody to vGAT (Synaptic Systems). Between 2 and 16 sections from each case were tested.

This study was approved by the three relevant local ethics committees (Erasme Academic Hospital, University of Brussels, and Belgian National Fund for Scientific Research) on research involving human subjects. Written informed consent was given by the parents in each case. All experiments were performed in compliance with the national and European laws, as well as with INSERM institutional guidelines, concerning animal experimentation.

**Received 27 June; accepted 25 July 2006.**
**Published online 16 August 2006.**

1. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418,** 869–872 (2002).
2. Evans, P. D. *et al.* Adaptive evolution of *ASPM*, a major determinant of cerebral cortical size in humans. *Hum. Mol. Genet.* **13,** 489–494 (2004).
3. Hill, R. S. & Walsh, C. A. Molecular insights into human brain evolution. *Nature* **437,** 64–67 (2005).
4. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
5. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437,** 69–87 (2005).
6. Clark, A. G. *et al.* Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. *Science* **302,** 1960–1963 (2003).
7. Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3,** e170 (2005).
8. Dorus, S. *et al.* Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119,** 1027–1040 (2004).
9. Wang, X., Grus, W. E. & Zhang, J. Gene losses during human origins. *PLoS Biol.* **4,** e52 (2006).
10. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
11. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188,** 107–116 (1975).
12. Pedersen, J. S. *et al.* Identification and classification of conserved RNA secondary structures in the human genome. *PLOS Computat. Biol.* **2,** e33 (2006).
13. Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432,** 695–716 (2004).
14. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
15. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302,** 413 (2003).
16. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–1050 (2005).
17. Hedges, S. B. & Kumar, S. Genomics. Vertebrate genomes compared. *Science* **297,** 1283–1285 (2002).
18. Altshuler, D. *et al.* A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
19. Collins, F. S., Brooks, L. D. & Chakravarti, A. A. DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8,** 1229–1231 (1998).
20. Stajich, J. E. & Hahn, M. W. Disentangling the effects of demography and selection in human history. *Mol. Biol. Evol.* **22,** 63–73 (2005).
21. Merryman, C. & Noller, H. in *RNA–Protein Interactions: A Practical Approach* (ed. Smith, C.) 237–253 (Oxford Univ. Press, New York, 1998).
22. Griffiths-Jones, S. *et al.* Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33,** D121–D124 (2005).
23. Meyer, G. & Goffinet, A. M. Prenatal development of reelin-immunoreactive neurons in the human neocortex. *J. Comp. Neurol.* **397,** 29–40 (1998).
24. Meyer, G. & Wahle, P. The paleocortical ventricle is the origin of reelin-expressing neurons in the marginal zone of the foetal human neocortex. *Eur. J. Neurosci.* **11,** 3937–3944 (1999).
25. Zecevic, N. & Rakic, P. Development of layer I neurons in the primate cerebral cortex. *J. Neurosci.* **21,** 5607–5619 (2001).
26. Meyer, G., Soria, J. M., Martinez-Galan, J. R., Martin-Clemente, B. & Fairen, A. Different origins and developmental histories of transient neurons in the marginal zone of the fetal and neonatal rat cortex. *J. Comp. Neurol.* **397,** 493–518 (1998).
27. Rakic, S. & Zecevic, N. Emerging complexity of layer I in human cerebral cortex. *Cereb. Cortex* **13,** 1072–1083 (2003).
28. Perez-Garcia, C. G., Tissir, F., Goffinet, A. M. & Meyer, G. Reelin receptors in developing laminated brain structures of mouse and human. *Eur. J. Neurosci.* **20,** 2827–2832 (2004).
29. Lukaszewicz, A. *et al.* G1 phase regulation, area-specific cell cycle control, and cytoarchitectonics in the primate cortex. *Neuron* **47,** 353–364 (2005).
30. Bond, J. & Woods, C. G. Cytoskeletal genes regulating brain size. *Curr. Opin. Cell Biol.* **18,** 95–101 (2006).
31. Ptak, S. E. *et al.* Fine-scale recombination patterns differ between chimpanzees and humans. *Nature Genet.* **37,** 429–434 (2005).
32. Duret, L., Semon, M., Piganeau, G., Mouchiroud, D. & Galtier, N. Vanishing GC-rich isochores in mammalian genomes. *Genetics* **162,** 1837–1847 (2002).
33. Bernardi, G. Isochores and the evolutionary genomics of vertebrates. *Gene* **241,** 3–17 (2000).
34. Kudla, G., Lipinski, L., Caffin, F., Helwak, A. & Zylicz, M. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol.* **4,** e180 (2006).
35. Tissir, F. & Goffinet, A. M. Reelin and brain development. *Nature Rev. Neurosci.* **4,** 496–505 (2003).
36. Impagnatiello, F. *et al.* A decrease of reelin expression as a putative vulnerability factor in schizophrenia. *Proc. Natl Acad. Sci. USA* **95,** 15718–15723 (1998).
37. Alfano, G. *et al.* Natural antisense transcripts associated with genes involved in eye development. *Hum. Mol. Genet.* **14,** 913–923 (2005).
38. Lambot, M. A., Depasse, F., Noel, J. C. & Vanderhaeghen, P. Mapping labels in the human developing visual system and the evolution of binocular vision. *J. Neurosci.* **25,** 7232–7237 (2005).
39. Depaepe, V. *et al.* Ephrin signalling controls brain size by regulating apoptosis of neural progenitors. *Nature* **435,** 1244–1250 (2005).
40. de Bergeyck, V., Naerhuyzen, B., Goffinet, A. M. & Lambert de Rouvroit, C. A panel of monoclonal antibodies against reelin, the extracellular matrix protein defective in reeler mutant mice. *J. Neurosci. Methods* **82,** 17–24 (1998).
41. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank G. Bejerano, C. Lowe, J. Kent, H. Noller, D. Feldheim, A. Love (UCSC), V. Albert and J.-C. Noel (Erasme Hospital), J.-P. Brion (ULB), A. Goffinet (UCL Louvain), the UCSC Genome Browser Group, Webb Miller (Penn. State), the Macaque Genome Sequencing Consortium, and the Broad Institute Genome Sequencing Platform and Whole-Genome Assembly Team. This work was funded by the Howard Hughes Medical Institute (D.H., S.R.S. and B.K.), the US NHGRI (D.H., A.D.K., S.K. and C.O.), the US National Cancer Institute (J.S.P.), the US NIGMS (K.S.P. and M.A.), the University of California Biotechnology Research and Education Program (A.S.), the Danish Research Council (J.S.P.), the Belgian FNRS, the Belgian FRSM, the Belgian Queen Elizabeth Medical Foundation (FMRE), the Interuniversity Attraction Poles Programme, Belgian State and Federal Office for Scientific, Technical and Cultural Affairs (P.V.), and the Fondation Erasme (M.-A.L. and N.L.). P.V. and M.-A.L. are Research Associate and Research Fellow of the FNRS, respectively.

**Author Contributions** Computational methods developed and applied by K.S.P., J.S.P., A.S. and A.D.K. Experimental analysis by S.R.S., P.V., N.L., M.-A.L., S.C., S.K., B.K., C.O., C.D., H.I. and M.A. Manuscript primarily written by K.S.P., S.R.S., P.V. and D.H.

**Author Information** The sequences of the HAR1 transcripts reported in this study have been submitted to GenBank (accession numbers DQ860409–DQ860415). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.H. (haussler@soe.ucsc.edu).