

# A High Density of Human Communication-Associated Genes in Chromosome 7q31-q36: Differential Expression in Human and Non-Human Primate Cortices

E. Schneider<sup>a</sup> L.R. Jensen<sup>b</sup> R. Farcas<sup>c</sup> I. Kondova<sup>f</sup> R.E. Bontrop<sup>f</sup> B. Navarro<sup>d</sup>  
E. Fuchs<sup>e</sup> A.W. Kuss<sup>b</sup> T. Haaf<sup>a</sup>

<sup>a</sup>Institute of Human Genetics, Julius Maximilians University, Würzburg, <sup>b</sup>Interfaculty Institute for Genetics and Functional Genomics, Ernst Moritz Arndt University, Greifswald, <sup>c</sup>Institute for Cardiovascular Regeneration, Goethe University, Frankfurt/Main, <sup>d</sup>Institute of Legal Medicine, University Medical Center, Mainz, and <sup>e</sup>German Primate Center, Göttingen, Germany; <sup>f</sup>Biomedical Primate Research Center, Rijswijk, The Netherlands

## Key Words

Baboon · Chimpanzee · Chromosome 7q31-q36 · Communication abilities · Cortex · Gene expression · Language evolution · Marmoset · T cell receptor beta locus

## Abstract

The human brain is distinguished by its remarkable size, high energy consumption, and cognitive abilities compared to all other mammals and non-human primates. However, little is known about what has accelerated brain evolution in the human lineage. One possible explanation is that the appearance of advanced communication skills and language has been a driving force of human brain development. The phenotypic adaptations in brain structure and function which occurred on the way to modern humans may be associated with specific molecular signatures in today's human genome and/or transcriptome. Genes that have been linked to language, reading, and/or autism spectrum disorders are prime candidates when searching for genes for human-specific communication abilities. The database and genome-wide expression analyses we present here revealed a clustering of such communication-associated genes (COAG) on human

chromosomes X and 7, in particular chromosome 7q31-q36. Compared to the rest of the genome, we found a high number of COAG to be differentially expressed in the cortices of humans and non-human primates (chimpanzee, baboon, and/or marmoset). The role of X-linked genes for the development of human-specific cognitive abilities is well known. We now propose that chromosome 7q31-q36 also represents a hot spot for the evolution of human-specific communication abilities. Selective pressure on the T cell receptor beta locus on chromosome 7q34, which plays a pivotal role in the immune system, could have led to rapid dissemination of positive gene variants in hitchhiking COAG.

Copyright © 2012 S. Karger AG, Basel

Many studies on the unique position of humans among the clade of primates did not identify a sharp boundary between human and non-human primates but a more or less gradient transition. Nevertheless, 3 distinctive traits have been exclusively attributed to humans: bipedal locomotion, modified teeth and jaw morphology, and most importantly an extraordinary brain size [Jones et al., 2000]. The latter holds true in absolute as well as in

relative terms. Related to body size, humans are endowed with the largest brain among all mammals. In general, this encephalization is assumed to reflect intelligence and cognitive abilities [Alba, 2010]. Human communication, in particular spoken language, is a major cognitive ability which has its structural basis in the enlarged human brain. At the same time, it may have played an essential role in human brain evolution that has led to a 3 times higher encephalization quotient (EQ) than in chimpanzees [Jerison, 1976].

Having been constant in Australopithecines for at least 2 million years, absolute cranial capacity started to increase about 2–3 million years ago. *Homo habilis* was the first *Homo* showing an enhanced EQ in the fossil record [Ruff et al., 1997]. A higher EQ is associated with increased energy consumption. In fact, the adaptive changes to increase human brain metabolism may be at their limits [Khaitovich et al., 2008]. In modern humans the adult brain takes up 20–25% of the total resting metabolic rate (RMR), mostly in form of about 130 g glucose per day [Hitze et al., 2010]. In newborns the brain even requires up to 60% of the RMR [Gibbons, 1998]. This is outstanding, compared to non-human primates with an average RMR of 8–9%. According to the ‘expensive tissue hypothesis’ [Aiello and Wheeler, 1995], the reduced size of the human gastrointestinal tract, which compares to only 60% of a similar-sized primate, along with a dietary shift from low to high energy-dense food, makes good for the high energy demand of the human brain. Consumption of animal meat, as proven by cut marks of stone tools on fossilized bones, may have resulted in a higher dietary quality and supplied early humans with fatty acids for optimum brain metabolism [Cordain et al., 2001]. It is generally accepted that the increase in brain size, which took place somewhere along the way from late Australopithecines to the early *Homo*, was largely influenced by bipedalism and remodeling of the jaw apparatus that led to changed feeding habits and higher energy turnover [Jones et al., 2000].

Little is known about what has pushed the increase in brain size in the human lineage. We propose a 2-step process with an initial spark causing selective pressure on a critical region in the genome, followed by adaptations for fine adjustments. It is reasonable to assume that the dramatic evolutionary changes in human brain structure and function are still reflected in today’s human genome. There are numerous studies on (positive) selection of brain-specific genes which are usually defined as genes only expressed in the brain and/or the nervous systems [Enard et al., 2002]. The results do not provide evidence

for positive selection in the human lineage [Kosiol et al., 2008] or even show an excess of positively selected genes in the chimpanzee [Shi et al., 2006]. Thus, changes in brain-specific genes cannot satisfactorily explain the mechanism underlying human brain evolution. As outlined by Jobling et al. [2004], one plausible approach towards identification of the genetic changes underlying human evolution is ‘defining important human specific phenotypes (e.g., language, disease phenotypes) and researching their genetic bases’.

Any intentional thinking of human beings is structured and influenced through language [Lakatos and Janka, 2008]. If language acquisition is hindered, i.e., in untreated children with congenital deafness, mental development is impaired. On the other hand, once language is acquired, even deaf, mute, and/or blind people can communicate with artificial aids. Language, reading, and autism spectrum disorders may help to identify a group of communication-associated genes (COAG). Autism spectrum disorders are characterized by impairment of social interaction, communication, and language development [Tuchmann, 2003; Scherer and Dawson, 2011]. In this study, we performed *in silico* analyses on a panel of COAG as well as comparative gene expression analyses in frontal cortices of human and non-human primates. Our goal was to identify possible evolutionary mechanisms which may have contributed to the development of human-specific communication abilities.

## Materials and Methods

### Database Analyses

An intense literature [Cody et al., 2007; Smith, 2007; Gibson and Gruen, 2008; Fisher and Scharff, 2009; Caglayan, 2010; Newbury and Monaco, 2010; Scherer and Dawson, 2011] and database (<http://www.mindspec.org/autdb.html>) research revealed 244 genes associated with human-specific communication (table 1). These genes have been associated with autism spectrum, reading, speech, and language disorders. Stuttering genes were excluded, because stuttering is thought to be fairly different from the above-mentioned communication disorders [Newbury and Monaco, 2010].

Data mining including searches for information on genes were performed with BioMart in the Ensembl database (Ensembl genome browser release 60, November 2010). Chromosomal synteny were delineated with Compare Genome (<http://genomeevolution.org/CoGe/index.pl>). Biological gene functions were assigned with the Panther database (<http://www.pantherdb.org>). Both COAG and the entire human Ensembl gene set were analyzed using the Panther batch search tool. Statistical analyses were performed with Fisher’s exact test for count data.

**Table 1.** Catalogue of communication-associated genes and chromosomal localization in the human genome

Chromosome	Genes (localization)
1	<i>GSTM1</i> (1p13.3), <i>NTNG1</i> (1p13.3), <i>DPYD</i> (1p22), <i>DAB1</i> (1p32-p31), <i>MTF1</i> (1p33), <i>CA6</i> (1p36.2), <i>RIMS3</i> (1pter-p22.2), <i>NTRK1</i> (1q21-q22), <i>NOS1AP</i> (1q23.3), <i>RFPWD2</i> (1q25.1-q25.2), <i>PTGS2</i> (1q25.2-q25.3), <i>HSD11B1</i> (1q32-q41), <i>MARK1</i> (1q41), <i>DISC1</i> (1q42.1), <i>OR1C1</i> (1q44)
2	<i>NRXN1</i> (2p16.3), <i>NPAS</i> (2q11.2), <i>DPP10</i> (2q14.1), <i>CNTNAP5</i> (2q14.3), <i>TSN</i> (2q21.1), <i>GALNT13</i> (2q24.1), <i>SCN2QA</i> (2q23-q24), <i>SLC4A10</i> (2q23-q24), <i>SLC25A12</i> (2q24), <i>SCN1A</i> (2q24.3), <i>STK39</i> (2q24.3), <i>ITGA4</i> (2q31.3), <i>RAPGEF4</i> (2q31-q32), <i>DLX1</i> (2q32), <i>DLX2</i> (2q32), <i>INPP1</i> (2q32), <i>NRP2</i> (2q33.3), <i>MAP2</i> (2q34-q35), <i>AGAPI</i> (2q37)
3	<i>ROBO1</i> (3p12), <i>SUCLG2</i> (3p14.1), <i>FHIT</i> (3p14.2), <i>CACNA1D</i> (3p14.3), <i>FEZF2</i> (3p21.1), <i>GPX1</i> (3p21.3), <i>OXTR</i> (3p25), <i>XPC</i> (3p25), <i>CNTN4</i> (3p26-p25), <i>DRD3</i> (3q13.3), <i>FBXO40</i> (3q21.1), <i>MBD4</i> (3q21.3), <i>C3orf58</i> (3q24), <i>SLC9A9</i> (3q24), <i>SCHIP1</i> (3q25.33), <i>DCUN1D1</i> (3q26.3), <i>NLGN1</i> (3q26.32), <i>HTR3C</i> (3q27.1)
4	<i>GABRA4</i> (4p12), <i>GABRB1</i> (4p12), <i>EIF4E</i> (4q21-q25), <i>PCDH10</i> (4q28.3), <i>TDO2</i> (4q31-q32)
5	<i>CDH9</i> (5p14), <i>CDH10</i> (5p14.2), <i>SEMA5A</i> (5p15.2), <i>MEF2C</i> (5q14), <i>APC</i> (5q21-q22), <i>PITX1</i> (5q31), <i>ADRB2</i> (5q31-q32)
6	<i>KIAA1586</i> (6p12.1), <i>LRRC1</i> (6p12.2), <i>C4B</i> (6p21.3), <i>HLA-A</i> (6p21.3), <i>HLA-DRB1</i> (6p21.3), <i>GLO1</i> (6p21.3-p21.1), <i>DCDC2</i> (6p22.1), <i>ALDH5A1</i> (6p22), <i>KIAA0319</i> (6p22.3-p22.2), <i>F13A1</i> (6p25.3-p24.3), <i>HTR1B</i> (6q13), <i>CNR1</i> (6q14-q15), <i>GRIK2</i> (6q16.3-q21), <i>FRK</i> (6q21-q22.3), <i>PLN</i> (6q22.1), <i>HS3ST5</i> (6q22.31), <i>FABP7</i> (6q22-q23), <i>AHI1</i> (6q23.3), <i>OPRM1</i> (6q24-q25), <i>ESR1</i> (6q25.1), <i>PARK2</i> (6q25.2-q27), <i>RPS6KA2</i> (6q27)
7	<i>DCC</i> (7p11), <i>FIGNL1</i> (7p12.2), <i>COBL</i> (7p12.2-p12.1), <i>ABCA13</i> (7p12.3), <i>ADCY1</i> (7p13-p12), <i>HOXA1</i> (7p15.3), <i>NPY</i> (7p15.3), <i>TMEM185</i> (7p21.1), <i>AUTS2</i> (7q11.22), <i>PON1</i> (7q21.3), <i>NPTX2</i> (7q21.3-q22.1), <i>PFTK1</i> (7q21-q22), <i>LAMB1</i> (7q22), <i>RELN</i> (7q22), <i>PIK3CG</i> (7q22.3), <i>CTTNBP2</i> (7q31), <i>FOXP2</i> (7q31), <i>GPR85</i> (7q31), <i>IMMP2L</i> (7q31), <i>LEP</i> (7q31), <i>MET</i> (7q31), <i>WNT2</i> (7q31), <i>LRRN3</i> (7q31.1), <i>NTCAM</i> (7q31.1-q31.2), <i>ST7</i> (7q31.2), <i>CADPS2</i> (7q31.3), <i>GRM8</i> (7q31.3-q32.1), <i>UBE2H</i> (7q32), <i>MEST</i> (7q32), <i>FASTK</i> (7q35), <i>CNTNAP2</i> (7q35-q36), <i>EN2</i> (7q36), <i>DPP6</i> (7q36.2)
8	<i>DLGAP2</i> (8p23), <i>MCPHI</i> (8p23.1), <i>RBICCI</i> (8q11), <i>CYP11B1</i> (8q21), <i>FABP5</i> (8q21.13), <i>SDC2</i> (8q22-q23)
9	<i>SLC1A1</i> (9p24), <i>ASTN2</i> (9q33), <i>TSC1</i> (9q34), <i>ASS</i> (9q34.1), <i>DAPK1</i> (9q34.1)
10	<i>CTNNA3</i> (10q21), <i>EGR2</i> (10q21.1), <i>JMJD1C</i> (10q21.2), <i>REEP3</i> (10q21.3), <i>HTR7</i> (10q21-q24), <i>KCNMA1</i> (10q22.3), <i>PTEN</i> (10q23.3), <i>LZTS2</i> (10q24)
11	<i>CD44</i> (11p13), <i>HRAS</i> (11p15.5), <i>TH</i> (11p15.5), <i>DHCR7</i> (11q13.4), <i>SHANK2</i> (11q13.2), <i>HTR3A</i> (11q23.1), <i>CADMI</i> (11q23.2)
12	<i>DDX11</i> (12p21.1), <i>CACNA1C</i> (12p13.3), <i>ITGB7</i> (12q13.13), <i>AVPR1A</i> (12q14-q15), <i>TPH2</i> (12q21.1), <i>CCDC64</i> (12q24.23)
13	<i>ALOX5AP</i> (13q12), <i>FLT1</i> (13q12), <i>NBEA</i> (13q13), <i>PCDH9</i> (13q21.32), <i>HTR2A</i> (13q14-q21), <i>DIAPH3</i> (13q21.2), <i>MYO16</i> (13q33.3)
14	<i>FBXO33</i> (14q13.3), <i>LRFN5</i> (14q21.1), <i>MDGA2</i> (14q21.3), <i>ESR2</i> (14q23.2), <i>ESRRB</i> (14q24.3), <i>VASH1</i> (14q24.3)
15	<i>ATP10A</i> (15q11.2), <i>GABRB3</i> (15q11.2-q12), <i>UBE3A</i> (15q11-q13), <i>NDNL2</i> (15q13.1), <i>DYX1C1</i> (15q21.1), <i>ARNT2</i> (15q24), <i>NTRK3</i> (15q25)
16	<i>PRKCB1</i> (16p11.2), <i>SEZ6L2</i> (16p11.2), <i>SYT17</i> (16p12.3), <i>ABAT</i> (16p13.2), <i>GRIN2A</i> (16p13.2), <i>CACNA1H</i> (16p13.3), <i>TSC2</i> (16p13), <i>RBFOX1</i> (16p13.3), <i>ANKRD11</i> (16q24.3)
17	<i>RAI1</i> (17p11.2), <i>PER1</i> (17p13.1-p12), <i>SLC6A4</i> (17q11.1-q12), <i>NF1</i> (17q11.2), <i>NOS2A</i> (17q11.2-q12), <i>HOXB1</i> (17q21.3), <i>ITGB3</i> (17q21.32), <i>CACNA1G</i> (17q22), <i>BZRAP1</i> (17q22-q23)
18	<i>RIT2</i> (18q12.3), <i>ST8SIA5</i> (18q12.3), <i>SYT4</i> (18q12.3), <i>ATP5A1</i> (18q21), <i>MBD1</i> (18q21), <i>SETBP1</i> (18q21.1)
19	<i>MBD3</i> (19p13.3), <i>ODF3L2</i> (19p13.3), <i>DMPK</i> (19q13.3)
20	<i>MACROD2</i> (20p12.1), <i>ADA</i> (20q13.12), <i>CDH22</i> (20q13.1), <i>GNAS</i> (20q13.3)
21	<i>CNS</i> (21q22.3)

**Table 1** (continued)

Chromosome	Genes (localization)
22	<i>COMT</i> (22q11.21), <i>TBX1</i> (22q11.21), <i>ADORA2A</i> (22q11.23), <i>ADSL</i> (22q13.2), <i>SHANK3</i> (22q13.3)
X	<i>CACNA1F</i> (Xp11.23), <i>PPP1R3F</i> (Xp11.23), <i>WNK3</i> (Xp11.22), <i>MAOA</i> (Xp11.3), <i>TSPAN7</i> (Xp11.4), <i>DMD</i> (Xp21.2), <i>CDKL5</i> (Xp22), <i>PTCHD1</i> (Xp22.11), <i>ARX</i> (Xp21.3), <i>GLRA2</i> (Xp22.1-p21.3), <i>IL1RAPL1</i> (Xp22.1-p21.3), <i>SH3KBP1</i> (Xp22.1-p21.3), <i>FRMPD4</i> (Xp22.2), <i>GRPR</i> (Xp22.2), <i>ASMT</i> (Xp22.3), <i>NLGN4X</i> (Xp22.33), <i>AR</i> (Xq12), <i>OPHN1</i> (Xq12), <i>CXCR3</i> (Xq13), <i>MED12</i> (Xq13), <i>NLGN3</i> (Xq13.1), <i>PCDH19</i> (Xq13.3), <i>SRPX2</i> (Xq21.33-q23), <i>HNRNPH2</i> (Xq22), <i>PSMD10</i> (Xq22.3), <i>DCX</i> (Xq22.3-q23), <i>AGTR2</i> (Xq22-q23), <i>RHOXF1</i> (Xq24), <i>SLC9A6</i> (Xq26.3), <i>FMR1</i> (Xq27.3), <i>MECP2</i> (Xq28), <i>PDZD4</i> (Xq28), <i>RAB39B</i> (Xq28), <i>RPL10</i> (Xq28), <i>SLC6A8</i> (Xq28)
Y	<i>NLGN4Y</i> (Yq11.22)

#### Brain Samples, RNA Isolation, and cDNA Synthesis

Human brain samples (excess material from autopsies) were obtained from the Department of Legal Medicine, University Medical Center Mainz, Germany and primate samples from the Biomedical Primate Research Centre, Rijswijk, The Netherlands and the German Primate Center, Göttingen. Male brain samples were prepared between 1–2 days post-mortem from 3 humans (*Homo sapiens*), 1 chimpanzee (*Pan troglodytes*), 3 baboons (*Papio hamadryas*) (Old World monkey), and 5 marmosets (*Callithrix jacchus*) (New World monkey). The human brains were from two 40-year-old victims of an accident and a 59-year-old male who died of acute heart failure. The chimpanzee brain was from a 14-year-old (late adolescent) animal which died of hemolytic anemia. The baboon brains came from two 9-year-old (subadult) animals and one 30-year-old (adult) animal which were terminated for experimental brain surgery. The marmoset brains came from five 2-year-old (adult) animals which were terminated for this study. Age classes of chimpanzee [Goodall, 1983], baboons [Sigg et al., 1982], and marmosets [Aroujo et al., 2000] were determined according to established standards.

Because of similar brain architecture in humans and great apes [Semendeferi et al., 2001; Sherwood et al., 2003], area A10 was excised from the frontal pole of human and chimpanzee brains. The baboon and marmoset samples were taken from the corresponding topological region. Frontal cortex tissue was immediately frozen and stored at  $-80^{\circ}\text{C}$  after dissection. Total RNA was isolated using TRIzol Reagent (Invitrogen, Darmstadt, Germany). The absorbance ratio at 260 and 280 nm was determined with a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, Del., USA). With one single exception, the measured ratios were between 1.92 and 2.04, indicative of pure high-quality RNA. RNA was amplified with the Illumina TotalPrep RNA Amplification Kit (Ambion, Austin, Tex., USA). T7 promoter containing cDNA was produced by reverse transcription of 400 ng total RNA each. This cDNA was subsequently transcribed into cRNA in the presence of biotinylated nucleotides.

#### Array Hybridization and Analysis

The labeled cDNA samples of different individuals from the same species (3 humans, 1 chimpanzee, 3 baboons, and 5 marmosets) were pooled. The 4 cDNA pools were hybridized under the same stringency conditions to 4 (of the 6) arrays on a Sentrix Hu-

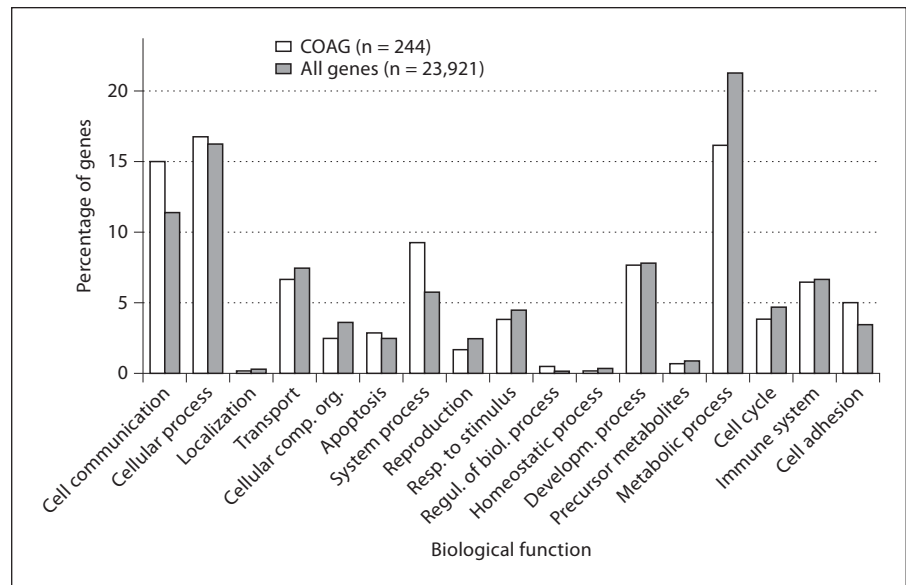
man-6 Expression BeadChip (Illumina, San Diego, Calif., USA) that contain probes for >48,000 transcripts. Following washing, the arrays were stained with Cy3-streptavidin and scanned with an Illumina BeadStation 500. The Illumina BeadStudio software was used for data analysis. The 'rank invariant normalization' algorithm was used for normalization of the data. Only genes passing the quality thresholds of diffScores of lower than -13 and higher than +13 with detection p values <0.05 were considered. A positive diffScore represents upregulation, while a negative diffScore represents downregulation.

When non-human primate samples are hybridized on a human array, sequence mismatches between the primate sample and the human oligonucleotide sequences on the array may reduce the signal intensity and, thus, be misinterpreted as reduced gene expression [Dannemann et al., 2009]. In order to avoid this mismatch problem, we only considered genes that appeared to be upregulated in baboon and marmoset compared to humans. It is very unlikely that stronger hybridization of the primate sample to a human array is due to sequence divergence. Because of high conservation of coding sequences between humans and chimpanzees [Chimpanzee Sequencing and Analysis Consortium, 2005; Varki and Altheide, 2005], both upregulated and downregulated genes were analyzed in this species.

## Results

### Distribution of COAG in the Human Genome

We performed an extensive literature and database research to select a panel of 244 communication-associated genes (table 1). Eight COAG were associated with language disorders, 4 with dyslexia, 2 with schizophrenia and language impairment, and 230 with autism spectrum disorders (ASD). This asymmetry reflects the enormous research efforts on autism and, on the other hand, our so far limited knowledge on the genetic basis of language development. In a first step we assigned biological functions to the 244 COAG and all genes (23,921) with



**Fig. 1.** Assignment of biological functions to COAG (white bars) and all genes (gray bars) listed in Ensembl (release 60, November 2010) using the Panther database. There is no between-group difference in gene functions.

chromosomal assignments in the Ensembl database. Of 17 possible functions in the Panther database, most genes of both the COAG and the Ensembl group were involved in cell communication, cellular and metabolic processes. There were no statistically significant between-group differences (fig. 1).

We then compared the chromosomal distribution of the 244 COAG with that of all Ensembl genes with chromosomal assignments (table 2). The expected number of COAG on a particular chromosome was calculated based on the ratio of the number of Ensembl genes on this chromosome divided by the number (23,921) of all Ensembl genes. Using Fisher's exact test, COAG were considered to be enriched/depleted on a particular chromosome when the observed number of COAG was significantly higher/lower than the expected number. Chromosomes 7 with 33 COAG ( $p = 0.0001$ , 95% CI 1.89–3.86) and X with 35 COAG ( $p = 0.003$ , 95% CI 1.16–2.32) were enriched with COAG, whereas chromosomes 1 (15 genes;  $p = 0.004$ , 95% CI 0.26–0.77) and 19 (3 genes;  $p = 0.003$ , 95% CI 0.04–0.58) contained less than expected COAG. When applying strict Bonferroni adjustment to correct for 24-fold parallel testing, the 5% criterion of significance has to be lowered accordingly to a  $p$  value of 0.002 being significant. Thus, even after correcting for multiple testing, the enrichment of chromosome 7 with COAG remained significant. In this context it is worth emphasizing that 17 of the 33 COAG on chromosome 7, including *FOXP2* and *CNTNAP2*, were clustered in 7q31-q36.

**Table 2.** Enrichment of COAG on human chromosomes 7 and X

Chromosome	Number of known genes <sup>a</sup>	Number of COAG	Enrichment factor	p value
1	3,141	15	0.5	0.004*
2	1,346	19	1.4	0.741
3	1,463	18	1.2	0.445
4	796	5	0.6	0.280
5	923	7	0.7	0.440
6	1,557	22	1.4	0.144
7	1,150	33	2.8	0.0001**
8	484	6	1.2	0.639
9	904	5	0.5	0.170
10	819	8	1.0	0.905
11	1,368	7	0.5	0.067
12	1,069	6	0.6	0.143
13	356	7	1.9	0.084
14	662	6	0.9	0.776
15	634	7	1.1	0.837
16	902	9	1.0	0.948
17	1,217	9	0.7	0.343
18	289	6	2.0	0.082
19	1,427	3	0.2	0.003*
20	603	4	0.7	0.391
21	225	1	0.4	0.395
22	508	5	1.0	0.937
X	2,000	35	1.7	0.003**
Y	78	1	1.3	0.820
Total	23,921	244		

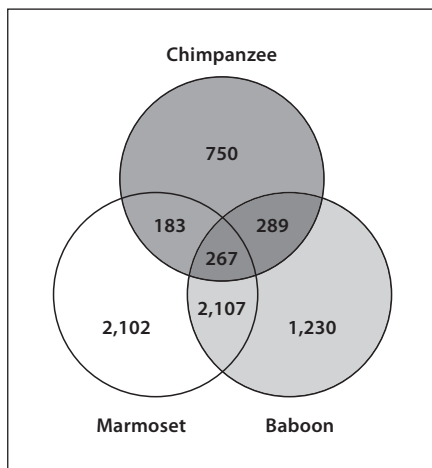
<sup>a</sup> According to Ensembl release 60, November 2010.

\* Significant underrepresentation of COAG.

\*\* Significant enrichment with COAG.

**Table 3.** COAG that are differentially expressed in human and non-human primate cortices

Expression level	Genes
Higher in chimpanzee and baboon	<i>ALOX5AP, CD44, OPHN1, SLC9A9</i>
Higher in chimpanzee and marmoset	<i>GPXI, SLC9A9</i>
Higher in baboon and marmoset	<i>ADSL, AUTS2, CACNA1G, CADPS2, CTTNBP2, FASTK, FBXO33, FRMPD4, GLRA2, GNAS, GRIK2, GRM8, HS3ST5, ITGB3, JMJD1C, KCNMA1, MED12, NBEA, NLGN1, NOS1AP, PCDH10, PCDH9, PTEN, RAI1, RFW2, SCHIP1, SCN1A, SH3KBP1, SLC9A9, ST7, SYT17, TBX1, UBE3A, VASH1</i>
Lower in chimpanzee	<i>AH11, ALDH5A1, ASS, BZRAP1, C3orf58, CDH22, CNTNAP2, DHCR7, DLGAP2, DPP6, GLO1, HTR2A, IMMP2L, LAMB1, LRRN3, MCPH1, NLGN4X, OXTR, PARK2, PDZD4, RAPGEF4, SETBP1, UBE2H</i>

**Fig. 2.** Venn diagram showing the number of genes with higher expression levels in chimpanzee, baboon, and marmoset cortices, respectively, compared to the human cortex.

#### *Expression Differences in Human and Non-Human Primate Cortices*

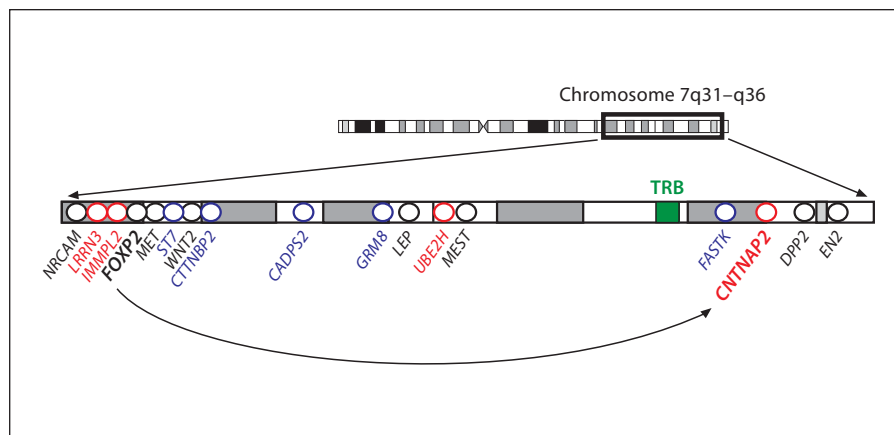
COAG expression in 3 human, 1 chimpanzee, 3 baboon, and 5 marmoset cortex samples was analyzed with an Illumina Expression BeadChip. In total 24,385 transcripts produced detectable hybridization signals on this array. Our first goal was to identify genes with higher expression levels in primate cortices than in the human cortex. This largely excludes interspecific hybridization artifacts due to sequence divergence. Compared to humans, 1,489 genes appeared to be upregulated in chimpanzee

cortex, 3,893 in baboon cortex, and 4,659 in marmoset cortex (fig. 2). Seventy-seven (2.0%) of 3,893 upregulated genes in baboon and 79 (1.7%) of 4,659 in marmoset belong to the COAG group, implying that the set of genes that is upregulated in Old and New World monkeys is significantly ( $p < 0.001$ , 95% CI 1.55–2.47 in baboon and 1.33–2.09 in marmoset) enriched with COAG. In contrast, only 11 COAG showed higher expression levels in the chimpanzee than in the human brain.

For the identification of genes that are upregulated in the human cortex, we only relied on the human-chimpanzee comparison. Due to the high degree of coding sequence conservation between humans and chimpanzees, most detected interspecies hybridization differences can be assumed to reflect true expression differences. Altogether, we found 2,750 genes with higher expression levels in human than in the chimpanzee cortex. A conceptually related study [Nowick et al., 2009] revealed 2,182 upregulated genes in the human cortex. In sum, 508 genes were identified in both studies. Fifty-nine (2.2%) of the 2,750 human upregulated genes in our study represent COAG, which is a significant enrichment ( $p < 0.001$ , 95% CI 1.87–3.17).

Because we were mainly interested in genes with conserved expression in non-human primates and human-specific up- or downregulation, we filtered out genes showing opposite expression changes in different non-human primates. For example, *GABRB3* showed higher expression levels in baboon and marmoset, but a reduced expression level in chimpanzee, compared to humans. Table 3 presents the remaining 23 COAG with human-specific upregulation and 38 genes with human-specific

**Fig. 3.** Enrichment of chromosome 7q31-q36 with 17 COAG. Genes with similar expression levels in human and primate cortices are indicated by black circles, genes that are downregulated in human cortex (compared to chimpanzee, baboon, and/or marmoset) are indicated by blue circles, and genes that are upregulated in human cortex (compared to chimpanzee) by red circles. Gene names are written below the circles. *FOXP2* binding to *CNTNAP2* (indicated by an arrow) leads to downregulation of *CNTNAP2*. *TRB* (indicated by a green box) represents the T cell receptor beta locus.



downregulation. Of the latter group, 34 genes showed increased expression levels in baboon and marmoset cortex, 2 in chimpanzee and marmoset cortex, and 4 in chimpanzee and baboon cortex. One gene encoding the sodium hydrogen exchanger, *SLC9A9*, was more highly expressed in all 3 non-human primates compared to humans. When only considering COAG which are differentially regulated in human and non-human primate cortices, chromosome 7 also shows a highly significant ( $p < 0.001$ ) enrichment. Six (*CNTNAP2*, *DPP6*, *IMMP2L*, *LAMB1*, *LRRN3*, and *UBE2H*) of 23 genes with human-specific upregulation and 6 (*AUTS2*, *CADPS2*, *CTTNBP2*, *FASTK*, *GRM8*, and *ST7*) of 38 genes with human-specific downregulation are located on chromosome 7.

#### Chromosome 7q31-q36

Both in silico and expression analyses identified chromosome region 7q31-q36 as a hot spot for COAG. It contains 17 (7%) of 244 COAG, including the 2 most prominent language genes *FOXP2* and *CNTNAP2* (fig. 3). *FOXP2* and *CNTNAP2* are not only physically but also functionally linked [Vernes et al., 2008]. Five COAG on 7q31-q36, namely *ST7*, *CTTNBP2*, *CADPS2*, *GRM8*, and *FASTK*, were downregulated in the human cortex compared to non-human primates. Three genes, *LRRN3*, *IMMP2L*, and *CNTNAP2*, were upregulated in the human cortex compared to chimpanzee. Thus, 8 (13%) of our 61 top candidate genes for human-specific regulation (table 2) are clustered in this region which represents only approximately 1.5% of the entire genome. Consistent with the results of Nowick et al. [2009], *FOXP2* was not differentially regulated in the human and non-human primate cortex. In addition to the COAG cluster, 7q31-q36 contains the T cell receptor beta locus, which is important for T cell activation in response to specific antigens.

#### Discussion

There are many hypotheses about the evolution of language and the human brain. One widely accepted theory claims that language was the main driving force for human brain evolution [Jerison, 1976]. On the other hand, even non-human primates appear to exhibit a predisposition to language [Cooper, 2006]. Thus, the genetic basis of language may have arisen long before language appeared [Chater et al., 2009]. Two or 3 million years ago, an Australopithecine species may have acquired communication abilities which were still far away from spoken language but already distinct from ape-like communication. This predisposition may have paved the way for the rapid evolution of a large brain and spoken language. Because it is not possible to analyze the genomes of Australopithecines, inferential methods, i.e., comparisons between human and non-human primate species, must be used to identify molecular signatures of language evolution in today's human genome and transcriptome.

First, we delineated a set of COAG and looked for their biological functions. Because language is a very complex trait, it is not unexpected that COAG are involved in many different biological processes, and there is no obvious enrichment for genes with a particular function compared to the entire human gene catalogue. Probably due to the high energy requirements of the human brain [Peters and Langemann, 2009], a large proportion of both COAG (16%) and all genes (21%) is associated with metabolic processes. There is a significant enrichment of COAG on chromosomes 7 (33 genes) and X (35 genes). The latter is not unexpected, because the X chromosome has accumulated a disproportionate number of genes for mental functioning and social cognition [Zechner et al., 2001; Skuse, 2005]. With the exception of *RPL10*, which

resides on chimpanzee chromosome 19, 34 of 35 human X-linked COAG are located on the orthologous primate X chromosomes. The hypothetical last common ancestor of primates was endowed with 2 human chromosome 7 homologs, called 7a (syntenic to 7p21-q11.21, 7q11.23-q21.3, and 7q22.1-qter) and 7b (syntenic to 7pter-p22, 7q11.21-q11.23, and 7q21.3-q22.1). A centric fusion of 7a and 7b in a common ancestor of recent great apes and humans generated the ancestral chromosome 7 of Homiidae [Müller et al., 2004; Froenicke, 2005]. This implies that the 33 COAG on human chromosome 7 have travelled together on the same chromosome for at least 15–20 million years (split estimate of humans and great apes) [Goodman et al., 1998; Enard and Pääbo, 2004]. The cluster of 17 (7%) COAG on human chromosome 7q31-q36 was already present on the ancestral primate chromosome 7a more than 80 million years ago, although the gene order on this chromosome may have been extensively reshuffled in the course of evolution. We conclude that highly conserved arrays of genes were selected during primate evolution, long before these genes were recruited for advanced communication abilities. Similar to transcriptional operons in bacteria, functionally cooperating genes also tend to cluster in higher eukaryotic genomes. This clustering could facilitate their coordinated regulation, i.e., by reducing the expenditure of chromatin unpackaging for transcription [Lee and Sonnhammer, 2003]. In this light, it is tempting to speculate that co-expression and fine-regulation of the 7q31-q36 COAG cluster in the human cortex has been important for the acquisition of human-specific communication abilities and language. Indeed, at least 2 genes in this cluster are known to functionally interact. The transcription factor FOXP2 binds to specific sites in the first intron of *CNTNAP2* and downregulates it [Vernes et al., 2008].

The extremely high similarity of the human and chimpanzee genomes [Chimpanzee Sequencing and Analysis Consortium, 2005; Varki and Altheide, 2005] argues in favor of the notion that phenotypic differences between humans and chimpanzees are based on differences in gene regulation. Indeed, comparative transcriptome analyses revealed subsets of genes with different expression levels in human and chimpanzee brains [Caceres et al., 2003; Gilad et al., 2006; Khaitovich et al., 2006; Nowick et al., 2009; Somel et al., 2009]. In this study, we used human expression arrays to compare the COAG expression profiles in human, chimpanzee, baboon, and marmoset cortices. About 12 million years of evolution can be assumed for the human-chimpanzee, 45 million years for the human-baboon, and 80 million years for the hu-

man-marmoset pairwise comparisons [Goodman et al., 1998; Enard and Pääbo, 2004]. One of our main interests was to find genes with similar expression levels in different non-human primates, including chimpanzee and Old and New World monkeys, but changed expression levels in the human brain. When using a human array to measure differential gene expression across primate species, probe sequence mismatches due to sequence divergence render the identification of genes that are downregulated in Old and New World monkeys difficult. In contrast to conceptually related studies [e.g., Somel et al., 2009], we did not mask all probes on the array that did not perfectly match the DNA sequences of the species examined. The Illumina bead technology used in our study is based on rather long oligonucleotides compared to other platforms. Consequently, there are relatively few oligonucleotides without any sequence mismatch in all 4 analyzed species. On the other hand, long oligonucleotides are more robust against hybridization artifacts. Because sequence mismatches are unlikely to improve hybridization efficiency, we focused on genes with higher expression levels in non-human primates. Another unfavorable factor for this type of studies is the limited availability of primate brain samples for high-quality RNA preparation. This makes it difficult to minimize stochastic effects. Previous studies demonstrated considerable intraspecific variation in epigenetic gene regulation both in human and non-human primate brains [Farcas et al., 2009]. Although we have used only male brain samples from subadult and adult individuals, it is unrealizable to control for all possible confounding factors.

A large proportion (20%) of the 244 analyzed COAG was downregulated in the human cortex compared to both baboon and marmoset. Relatively few (5%) COAG were downregulated in humans compared to chimpanzee, and only 1 gene, *SLC9A9*, was downregulated in humans compared to all 3 analyzed non-human primates. This may be partially explained by the fact that the transcriptomes of humans and chimpanzee are more similar to each other than those of humans and Old or New World monkeys, respectively. On the other hand, only 1 chimpanzee brain was available for expression analyses, whereas 3 baboons and 5 marmosets could be analyzed. In contrast to Old and New World monkeys, the coding sequence between humans and chimpanzee is highly conserved exhibiting approximately 1 mismatch in 100 bp [Chimpanzee Sequencing and Analysis Consortium, 2005] which allows one to identify genes that are downregulated in chimpanzee using a human array. In silico analyses confirmed that for the vast majority of chim-



panzee downregulated genes the 50mer oligonucleotides on the array exhibit either no or a single mismatch which is unlikely to interfere with hybridization efficiency. Altogether, 59 (25%) of 244 COAG appeared to be downregulated in the chimpanzee cortex compared to humans. However, as outlined above, results based on 1 cortex sample have to be interpreted with caution. Nevertheless, our study clearly demonstrates an enrichment of COAG among all genes that are differentially regulated in human and non-human primate cortices.

Relying mostly on autism-related genes as source of COAG, the selected set includes genes that may be involved in many sorts of communication, not only language. In addition to impaired language development, autistic persons may exhibit stereotypic, ritualistic, and compulsive behaviors, limited social interactions as well as sensory problems [Tuchmann, 2003; Scherer and Dawson, 2011]. Evidently, non-human primates do not have language, but they cannot afford to indulge in repetitive or self-injurious behaviors. So far there is no general agreement on what is unique to human language. Tomasello [2008] considers a ‘theory of mind’ as basis for the faculty of language. Advanced (intentional rather than reflective) communication abilities in humans require assessment of the mental states of others. This ability is severely impaired in autistic persons, which makes ASD genes prime candidates for COAG. Although some genes, for example *HLA-A* and *HLA-DBR*, that have been associated with ASD are unlikely to be related to communication, we did not eliminate these genes from our COAG set to avoid any bias in selection.

Several observations argue in favor of the notion that human chromosome 7q31-q36 may represent a hot spot for the evolution of advanced communication abilities. Firstly, this region is enriched with COAG that show

human-specific upregulation (*LRRN3*, *IMMPL2*, and *CNTNAP2*) or downregulation (*ST7*, *CTTNBP2*, *CADPS2*, *GRM8*, and *FASTK*). The FOXP2-CNTNAP2 pathway has been related to speech and language development as well as different neurodevelopmental disorders, including Gilles de la Tourette syndrome, schizophrenia, autism, and attention deficit hyperactivity disorder [Fisher and Scharff, 2009; Newbury and Monaco, 2010]. Secondly, this region contains the T cell receptor beta locus which plays an essential role in the immune system. In T cells, antigen receptor genes are assembled by V(D)J recombination from germline V, D, and J segments, creating much of the diversity of the immune system [Bonnet et al., 2009]. Genes involved in immune defense are targets of rapid evolution, most likely because they have a direct impact on the organisms’ fitness in the battle with pathogens [Kosiol et al., 2008]. The T cell receptor beta locus could have conferred selective pressure on a haplotype with favorable COAG variants. Australopithecines most likely improved the protein content of their diet by scavenging, increasing the risk for infections through exposure to rotten corpses [Teaford and Ungar, 2000]. Epidemic diseases may have dramatically reduced the number of breeding individuals ( $N_e$ ) from time to time [Wright, 1931]. Consistent with a bottleneck effect [Nei et al., 1975], individuals with an immune system fitting new challenges had a higher chance to survive and to transmit hitchhiking positive COAG variants into next generations. Such selective pressure could have rapidly increased the frequency of a COAG haplotype predisposing to language acquisition and/or brain development. Future bioinformatic analyses estimating the age of 7q31-q36 haplotypes may help to elucidate the time window when the favorable COAG haplotype arose during human evolution.

## References

- Aiello LC, Wheeler P: The expensive-tissue hypothesis: the brain and the digestive system in human and primate evolution. *Curr Anthropol* 36:199–221 (1995).
- Alba DM: Cognitive inferences in fossil apes (Primates, Hominoidea): does encephalization reflect intelligence? *J Anthropol Sci* 88: 11–48 (2010).
- Araújo A, Arruda MF, Alencar AI, Albuquerque F, Nascimento MC, et al: Body weight of wild and captive common marmosets (*Callithrix jacchus*). *Int J Primatol* 21:317–324 (2000).
- Bonnet M, Ferrier P, Spicuglia S: Molecular genetics at the T-cell receptor beta locus: insights into the regulation of V(D)J recombination. *Adv Exp Med Biol* 650:116–132 (2009).
- Cáceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, et al: Elevated gene expression levels distinguish human from non-human primate brains. *Proc Natl Acad Sci USA* 100: 13030–13035 (2003).
- Caglayan AO: Genetic causes of syndromic and non-syndromic autism. *Dev Med Child Neurol* 52:130–138 (2010).
- Chater N, Reali F, Christiansen MH: Restrictions on biological adaptation in language evolution. *Proc Natl Acad Sci USA* 106:1015–1020 (2009).
- Chimpanzee Sequencing and Analysis Consortium: Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87 (2005).
- Cody JD, Sebold C, Malik A, Heard P, Carter E, et al: Recurrent interstitial deletions of proximal 18q: a new syndrome involving expressive speech delay. *Am J Med Genet Part A* 143A:1181–1190 (2007).

- Cooper DL: Broca's arrow: evolution, prediction, and language in the brain. *Anat Rec B New Anat* 289:9–24 (2006).
- Cordain L, Watkins BA, Mann NJ: Fatty acid composition and energy density of foods available to African hominids. Evolutionary implications for human brain development. *World Rev Nutr Diet* 90:144–161 (2001).
- Dannemann M, Lorenc A, Hellmann I, Khaitovich P, Lachmann M: The effects of probe binding affinity differences on gene expression measurements and how to deal with them. *Bioinformatics* 25:2772–2779 (2009).
- Enard W, Pääbo S: Comparative primate genomics. *Annu Rev Genomics Hum Genet* 5:351–578 (2004).
- Enard W, Khaitovich P, Klose J, Zöllner S, Heisig F, et al: Intra- and interspecific variation in primate gene expression patterns. *Science* 296:340–343 (2002).
- Farcas R, Schneider E, Frauenknecht K, Kondova I, Bontrop R, et al: Differences in DNA methylation patterns and expression of the *CCRK* gene in human and non-human primate cortices. *Mol Biol Evol* 26:1379–1389 (2009).
- Fisher SE, Scharff C: *FOXP2* as a molecular window into speech and language. *Trends Genet* 25:166–177 (2009).
- Froenicke L: Origins of primate chromosomes – as delineated by Zoo-FISH and alignments of human and mouse draft genome sequences. *Cytogenet Genome Res* 108:122–138 (2005).
- Gibbons A: Solving the brain's energy crisis. *Science* 280:1345–1347 (1998).
- Gibson CJ, Gruen JR: The human lexinome: genes of language and reading. *J Commun Disord* 41:409–420 (2008).
- Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP: Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* 440:242–245 (2006).
- Goodall J: Population dynamics during a 15-year period on one community of free-living chimpanzees in the Gombe national park, Tanzania. *Z Tierpsychol* 61:1–60 (1983).
- Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, et al: Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol Phylogenet Evol* 9:585–598 (1998).
- Hitze B, Hubold C, van Dyken R, Schlichting K, Lehnert H, et al: How the selfish brain organizes its supply and demand. *Front Neuroenergetics* 9:2–7 (2010).
- Jerison HJ: The paleoneurology of language. *Ann NY Acad Sci* 280:370–382 (1976).
- Jobling MA, Hurles ME, Tyler-Smith C: *Human Evolutionary Genomics* (Garland Science, New York 2004).
- Jones S, Martin R, Pilbeam D: *The Cambridge Encyclopedia of Human Evolution*. (Cambridge University Press, Cambridge 2000).
- Khaitovich P, Enard W, Lachmann M, Pääbo S: Evolution of primate gene expression. *Nat Rev Genet* 7:693–702 (2006).
- Khaitovich P, Lockstone HE, Wayland MT, Tsang TM, Jayatilaka SD, et al: Metabolic changes in schizophrenia and human brain evolution. *Genome Biol* 9:R124 (2008).
- Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD: Patterns of positive selection in six mammalian genomes. *PLoS Genet* 4:8 (2008).
- Lakatos L, Janka Z: Evolution of human brain and intelligence. *Ideggogy Sz* 61:220–229 (2008).
- Lee JM, Sonnhammer EL: Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res* 13:875–882 (2003).
- Müller S, Finelli P, Neusser M, Wienberg J: The evolutionary history of human chromosome 7. *Genomics* 84:458–467 (2004).
- Nei M, Maruyama T, Chakraborty R: The bottleneck effect and genetic variability of populations. *Evolution* 29:1–10 (1975).
- Newbury DF, Monaco AP: Genetic advances in the study of speech and language disorders. *Neuron* 68:309–320 (2010).
- Nowick K, Gernat T, Almaas E, Stubbs L: Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc Natl Acad Sci USA* 106:22358–22363 (2009).
- Peters A, Langemann D: Build-ups in the supply chain of the brain: on the neuroenergetic cause of obesity and type 2 diabetes mellitus. *Front Neuroenergetics* 1:2 (2009).
- Ruff CB, Trinkhaus E, Holliday TW: Body mass and encephalization in Pleistocene Homo. *Nature* 387:173–176 (1997).
- Scherer SW, Dawson G: Risk factors for autism: translating genomic discoveries into diagnostics. *Hum Genet* 130:123–148 (2011).
- Semendeferi K, Armstrong E, Schleicher A, Zilles K, van Hoesen GW: Prefrontal cortex in humans and apes: a comparative study of area 10. *Am J Phys Anthropol* 114:224–241 (2001).
- Sherwood CC, Broadfield DC, Holloway RL, Gannon PJ, Hof PR: Variability of Broca's area homologue in African great apes: implications for language evolution. *Anat Rec A Discov Mol Cell Evol Biol* 271:276–285 (2003).
- Shi P, Bakewell MA, Zhang J: Did brain-specific genes evolve faster in humans than in chimpanzees? *Trends Genet* 22:608–613 (2006).
- Sigg H, Stolba A, Abegglen JJ, Dasser V: Life history of hamadryas baboons: physical development, infant mortality, reproductive parameters and family relationships. *Primates* 23:473–487 (1982).
- Skuse DH: X-linked genes and mental functioning. *Hum Mol Genet* 14:R27–R32 (2005).
- Smith SD: Genes, language development, and language disorders. *Ment Retard Dev Disabil Res Rev* 13:96–105 (2007).
- Somel M, Franz H, Yan Z, Lorenc A, Guo S, et al: Transcriptional neoteny in the human brain. *Proc Natl Acad Sci USA* 106:5743–5748 (2009).
- Teaford MF, Ungar PS: Diet and the evolution of the earliest human ancestors. *Proc Natl Acad Sci USA* 97:13506–13511 (2000).
- Tomasello M: *Origins of Human Communication* (The MIT Press, Cambridge 2008).
- Tuchman R: Autism. *Neurol Clin* 21:915–932 (2003).
- Varki A, Altheide TK: Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res* 15:1746–1758 (2005).
- Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, et al: A functional genetic link between distinct developmental language disorders. *N Engl J Med* 359:2337–2345 (2008).
- Wright S: Evolution in Mendelian populations. *Genetics* 16:97–159 (1931).
- Zechner U, Wilda M, Kehrer-Sawatzki H, Vogel W, Fundele R, et al: A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? *Trends Genet* 17:697–701 (2001).