Human Accelerated Regions and Other Human-Specific Sequence Variations in the Context of Evolution and Their Relevance for Brain Development

Anastasia Levchenko[1,*], Alexander Kanapin[1,2], Anastasia Samsonova[1,2], and Raul R. Gainetdinov[1,3]

[1]Institute of Translational Biomedicine, Saint Petersburg State University, Russia

[2]Department of Oncology, University of Oxford, United Kingdom

[3]Skolkovo Institute of Science and Technology, Skolkovo, Moscow, Russia

*Corresponding author: E-mail: a.levchenko@spbu.ru.

Accepted: November 14, 2017

Abstract

The review discusses, in a format of a timeline, the studies of different types of genetic variants, present in *Homo sapiens*, but absent in all other primate, mammalian, or vertebrate species, tested so far. The main characteristic of these variants is that they are found in regions of high evolutionary conservation. These sequence variations include single nucleotide substitutions (called human accelerated regions), deletions, and segmental duplications. The rationale for finding such variations in the human genome is that they could be responsible for traits, specific to our species, of which the human brain is the most remarkable. As became obvious, the vast majority of human-specific single nucleotide substitutions are found in noncoding, likely regulatory regions. A number of genes, associated with these human-specific alleles, often through novel enhancer activity, were in fact shown to be implicated in human-specific development of certain brain areas, including the prefrontal cortex. Human-specific deletions may remove regulatory sequences, such as enhancers. Segmental duplications, because of their large size, create new coding sequences, like new functional paralogs. Further functional study of these variants will shed light on evolution of our species, as well as on the etiology of neurodevelopmental disorders.

Key words: substitutions, duplications, deletions, neurodevelopmental disorders, psychiatry, genes.

Introduction

Human accelerated regions (HARs) and human-specific genomic rearrangements, including deletions and segmental duplications, are among the most interesting sequences in the human genome to study, because they seem to shed light on the appearance of *Homo sapiens* as a species, especially in the sense of the exceptionally developed human brain (O'Bleness et al. 2012a; Sassa 2013; Hubisz and Pollard 2014; Zhang and Long 2014; Bae et al. 2015; Franchini and Pollard 2015; Dennis and Eichler 2016; Silver 2016; Namba and Huttner 2017). The human brain appeared in evolution probably as a result of natural selection among the intellectually developed members of the Homo genus. Only *Homo sapiens* survived, by presenting an exceptionally developed telencephalon, the prefrontal cortex in particular, that gave us the capacity to invent more sophisticated tools and to plan ahead.

An obvious question is: which evolutionarily new sequence variations determined the new pattern of brain development? The human brain is indeed much more sophisticated when compared with our closest living relative, the chimpanzee, from whom we diverged ~8 Ma and with whom we share ~99% of our genome (The Chimpanzee Sequencing and Analysis Consortium 2005; Moorjani et al. 2016). As concluded by the classic study by King and Wilson (King and Wilson 1975), functionally new protein-coding genes cannot account for all obvious morphological differences between humans and chimpanzees, because human protein-coding genes are very similar to chimpanzee genes and constitute only ~1.5% of the human genome. The genetic basis of the morphological differences must therefore primarily come from noncoding, regulatory sequences. One of the most appealing ways to demonstrate that evolutionarily new regulatory regions play an important role in the function of the human brain was the discovery of HARs of 2006, followed by a number of similar studies (Pollard et al. 2006a; Prabhakar et al. 2006; Bird et al. 2007; Bush and

Lahn 2008; Lindblad-Toh et al. 2011; Gittelman et al. 2015). HARs are short, ~260 bp on an average, stretches of DNA, to 97% noncoding. They are conserved in vertebrates, including Pan troglodytes, but not in *Homo sapiens*, in whom the conserved sequences were subjected to significantly, in many cases dramatically, higher rates of single nucleotide substitutions (Hubisz and Pollard 2014). An example of a HAR is given on figure 1*a*. Because conserved noncoding sequences are most likely regions regulating gene expression, new substitutions in these regions imply new patterns of regulation. A fraction of HARs affect noncoding RNA genes (Pollard et al. 2006b), which are often also implicated in regulation of gene expression. In fact, a meta-analysis (Haygood et al. 2010), that used data from Pollard et al. (2006a), Prabhakar et al. (2006) and a study that investigated positive selection in human promoters (Haygood et al. 2007), indicated that human genes regulating neurodevelopment were subjected, during evolution, to the most prominent positive selection only within their noncoding sequences. This tendency is exclusive to neurodevelopmental genes, whereas positive selection in coding sequences is characteristic for evolution of human genes implicated in other functions, such as immune system and olfaction (Haygood et al. 2010). Similar results were obtained by two other scientific teams that showed that substitutions in the coding sequence of human brain-expressed genes were not numerous enough to indicate positive selection (Shi et al. 2006; Wang et al. 2007). Finally, coding sequences of genes associated with schizophrenia and autism were not subjected to accelerated nonsynonymous substitutions in humans (Ogawa and Vallender 2014).

As became clear starting in 2003 (Locke et al. 2003), the human genome is also subject to an important number of complex genomic rearrangements, including segmental duplications, that can also, because of their large size, affect coding genes (Fortna et al. 2004), in particular, create new functional paralogs (Charrier et al. 2012; Dennis et al. 2012, 2017). An example is given on figure 1*b*. Human-specific deletions may also remove regulatory sequences (McLean et al. 2011). An example is shown on figure 1*c*. The study of the patterns of human-specific large structural variants is currently in its infancy (Dennis and Eichler 2016), but it has already become clear that the impact of these variants in the context of human evolution is paramount.

The majority of single nucleotide substitutions relevant to human brain evolution are in noncoding regions and affected coding genes are subjected to genomic rearrangements. Notwithstanding, it is worth to note that there are examples of individually studied genes, with human-specific coding variants and evidence of positive selection, that also seem to play a role in brain development or function. Examples are the extensively studied transcription factor forkhead box P2 (*FOXP2*) that bears two human-specific nonsynonymous substitutions (Enard et al. 2002, 2009; Spiteri et al. 2007; Konopka et al. 2009; Vernes et al. 2011; Schreiweis et al. 2014), abnormal spindle-like microcephaly associated (*ASPM*) that bears a number of human-specific synonymous and nonsynonymous substitutions (Zhang 2003; Evans et al. 2004; Buchman et al. 2011; Jayaraman et al. 2016) and Abelson helper integration site 1 (*AHI1*) that also underwent positive selection in humans (Ferland et al. 2004; Cheng et al. 2012). The reader is referred to reviews that describe these genes in further detail (Vallender et al. 2008; O'Bleness et al. 2012a; Bae et al. 2015).

Taken together, currently available scientific data indicate that both noncoding regulatory regions, modified by single nucleotide substitutions and deletions, and new coding sequences, created by segmental duplications, are important in human evolution and seem to shape the human brain. Only genomic regions and genes that were discovered in genome-wide studies of human evolution will be dealt with in this review. Table 1 summarizes these regions and genes. Figure 2 presents the timeline of the studies.

Original Studies Describing HARs

The First Discovery of HARs

The two pioneering studies, in which HARs were described, appeared in 2006 (Pollard et al. 2006a, 2006b). The studies used a biostatistical method, which aimed at detecting any significant acceleration in the rate of nucleotide substitutions in human. First, Pollard et al. (2006a) aligned genomes of chimpanzee, mouse, and rat, in order to define regions of at least 96% conservation over 100 bp; second, they aligned the orthologous conserved sequences of the total of 17 vertebrates, including human. Software MultiZ and PhastCons from the PHAST package (Blanchette et al. 2004; Siepel et al. 2005; Hubisz et al. 2011) and a likelihood ratio test were used to this end. This method was different from the approach where the rate of substitutions in the genome of a species was compared with theoretical neutral expectations. In other words, Pollard et al. (2006a) relied only on factual sequences of a number of different vertebrates, to directly compare them with *Homo sapiens*. Pollard et al. (2006a) discovered a more stringent data set of 49 HARs, at a false discovery rate (FDR)-corrected P value <0.05, and a broader data set of 202 HARs in the human genome. As an example, the top five HARs had substitution rates 26 times higher than substitution rates in chimpanzee, compared with mouse. All these substitutions, except one, are fixed in humans. Figure 1*a* illustrates HAR1, as an example of HAR.

Most of the discovered HARs were near telomeres, in regions of high recombination rates. Furthermore, these were regions of high GC content, the two bases that link to each other via three hydrogen bonds, therefore creating strong nucleotide pairs. And the new variants,

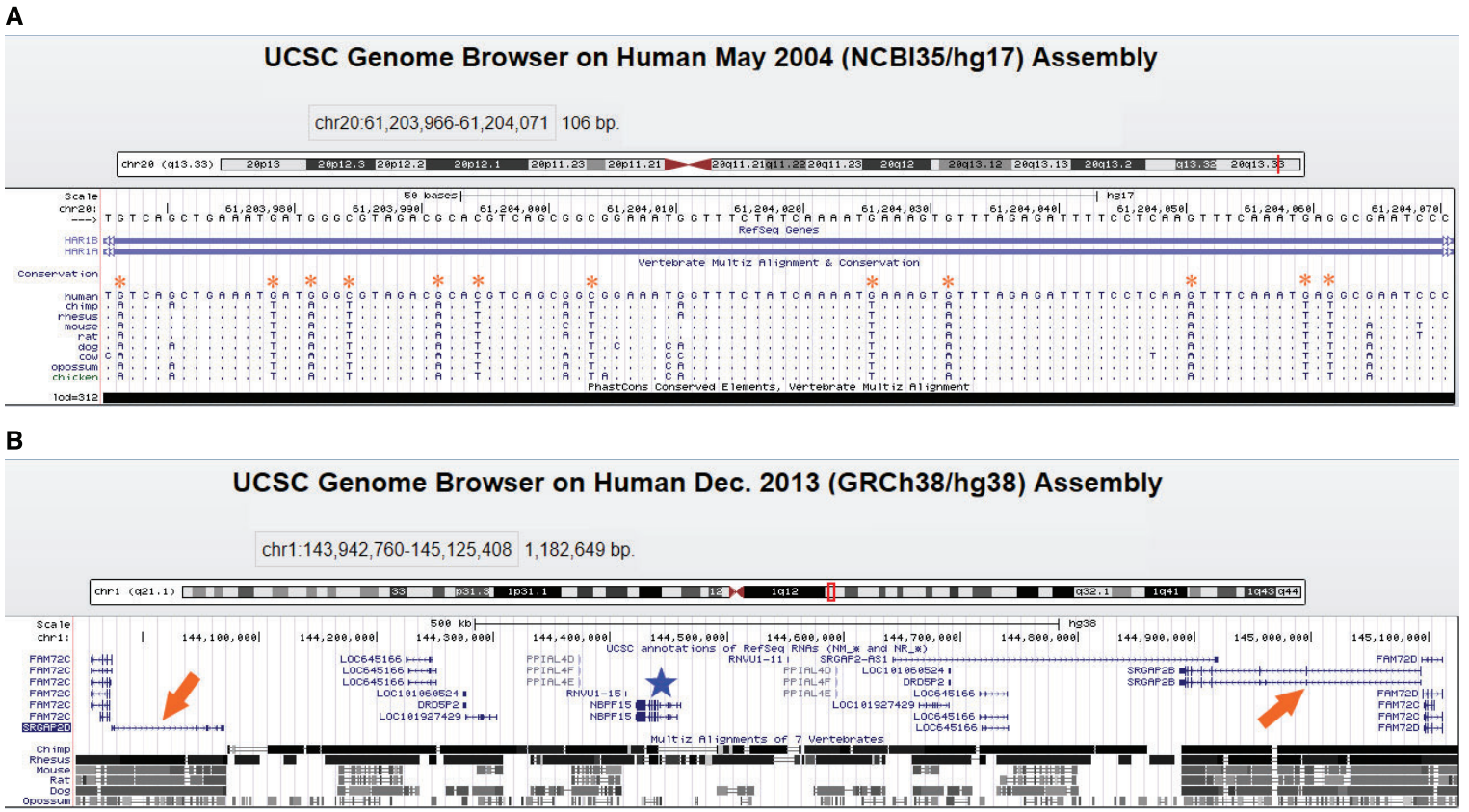

Fig. 1.—Examples of sequence variants described in the present review. (*a*–*c*) Actual examples of a HAR, several duplications, and a deletion. Screen shots from the UCSC Genome Browser are used. The species, genome assembly, chromosomal coordinates, and size of the region in bp are shown. (*a*) HAR1 is depicted. The coordinates of the region in the human NCBI genome assembly 35 are as defined in Pollard et al. (2006a). The twelve human-specific substitutions are denoted with *asterisks*. *Dots* within the alignment denote bases identical in human and other species. (*b*) The paralog *SRGAP2B* and the partially duplicated and inverted pseudogene *SRGAP2D* are pointed by *arrows*. Neighboring genes *FAM72* are depicted; this picture however contradicts results in Dennis et al. (2017): *SRGAP2B* duplicated within the same genomic fragment as *FAM72B* and *SRGAP2D* duplicated within the same genomic fragment as *FAM72D*. This contradiction illustrates the incompleteness of the current human genome assembly. The depicted gene *FAM72C* is actually located on chromosome 1p11.2. The gene *NBPF15* with primate-specific DUF120 domain amplification, denoted with a *star*, is also shown in this genomic fragment. The multispecies alignment depicts factual absence of this gene in nonprimate animals. The picture is similar for the 100 vertebrates' alignment. (*c*) hCONDEL.122 as defined in McLean et al. (2011). Build 2 Version 1 (Oct 2005) of the chimpanzee genome assembly, produced by the Chimpanzee Sequencing and Analysis Consortium, is shown. The corresponding human sequence contains a gap (the deletion). Other species are shown for comparison. *Single lines* signify gaps (sequence absent) and *double lines* signify more complex cases that are more difficult to resolve. In cases where multiple chains of the same species align over a region of the chimpanzee genome, the chains with single-lined gaps are often due to retrotransposed pseudogenes, whereas chains with double-lined gaps are more often due to paralogs and duplicated pseudogenes. (*d* and *e*) Generalized examples of functional variants. The examples do not represent particular regulatory sequences or coding genes, but illustrate principles upon which these sequences function, according to studies, described in the present review. The schematics are not to scale. (*d*) A HAR can act as an enhancer. In this case, transcriptional activity and/or expression pattern in tissues are altered. (*e*) Segmental duplications may result in paralogs with new functions. The paralogs are found in different chromosomal regions. The new paralogs are also often inverted relative to the parental sequence. The fact that the duplications are partial explains differences between the ancestral gene and the paralogs. Additionally, new sequence variants, denoted with a *star*, may be introduced in paralogs and will result in further alterations of the protein sequence. An example is the splicing variant in the gene *ARHGAP11B*. Although pseudogenes are often regarded as "dead genes," there is evidence that these noncoding gens may generate functional RNAs that regulate gene expression (Pink et al. 2011).

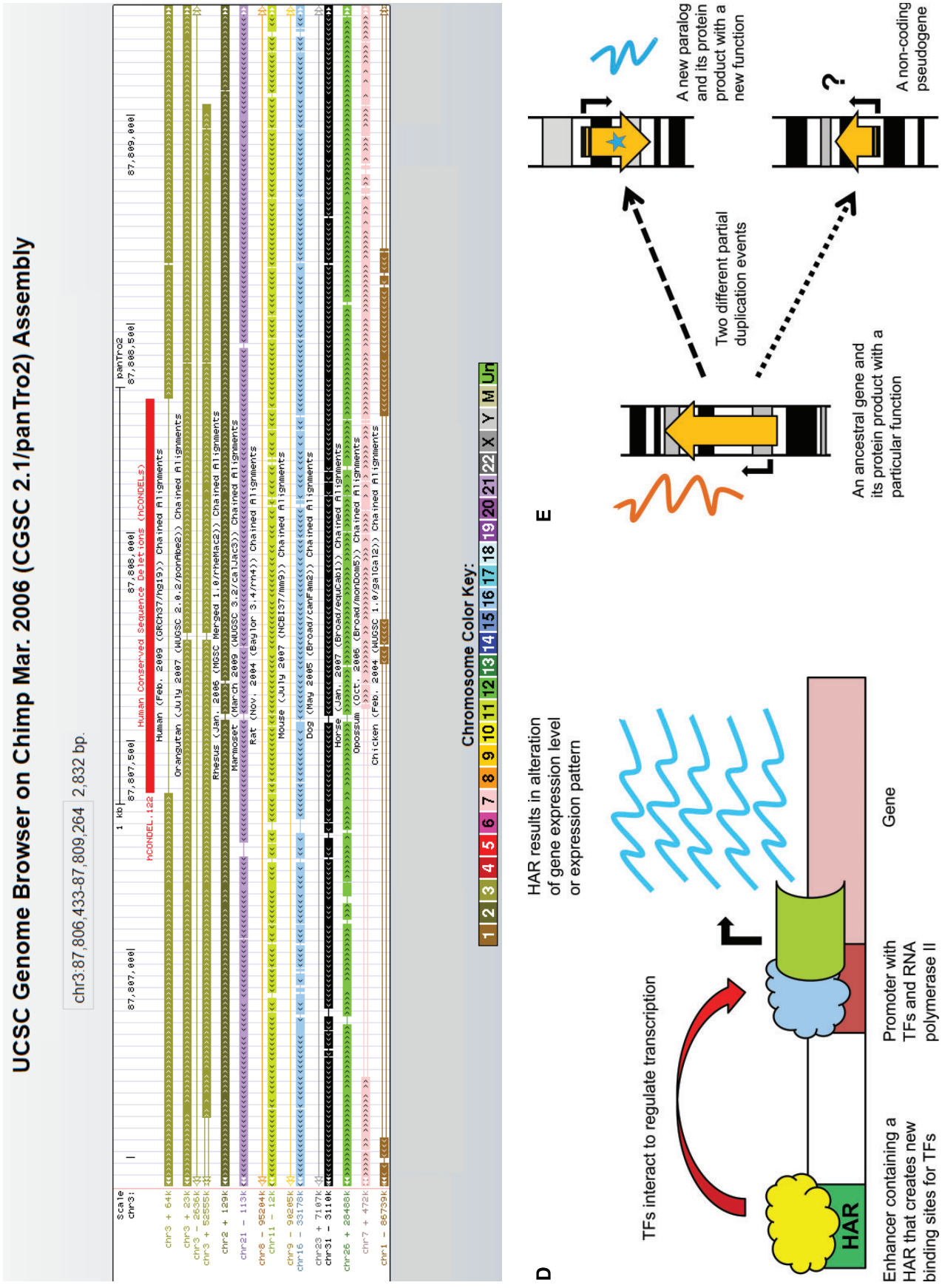

FIG. 1.—Continued

Table 1
Summary of Genomics Regions, Affected by Human-Specific Genetic Variations

Genomic Region Category	Evidence of Conservation	Number of Regions	Associated Genes with Some Functional Evidence[a]	References
HARs (Pollard)[b]	vertebrates	202	*HAR1A, AUTS2, NPAS3*	(Pollard et al. 2006a, 2006b; Kamm, Lopez-Leal, et al. 2013; Kamm, Pisciottano, et al. 2013; Oksenberg et al. 2013)
HACNSs (Prabhakar)	vertebrates	992	*AUTS2, NPAS3, CUX1*	(Prabhakar et al. 2006; Kamm, Lopez-Leal, et al. 2013; Kamm, Pisciottano, et al. 2013; Oksenberg et al. 2013; Doan et al. 2016)
ANCs (Bird)	vertebrates	1,356	*HAR1A, FZD8, PTBP2, GPC4*	(Pollard et al. 2006b; Bird et al. 2007; Boyd et al. 2015; Doan et al. 2016)
HARs (Bush)	mammals	63	unknown	(Bush and Lahn 2008)
2xHARs (Lindblad-Toh)	mammals	563	*NPAS3*	(Lindblad-Toh et al. 2011; Kamm, Lopez-Leal, et al. 2013; Kamm, Pisciottano, et al. 2013)
haDHSs (Gittelman)	primates	524	unknown	(Gittelman et al. 2015)
HSDs	apes	218	*SRGAP2, ARHGAP11*	(Charrier et al. 2012; Dennis et al. 2012, 2017; Florio et al. 2015, 2016)
hCONDELs	vertebrates	510	*GADD45G*	(McLean et al. 2011)

NOTE.—Associated genes with an experimentally confirmed role in brain development are also indicated.
[a]The shown genes are limited to the scope of this review.
[b]To distinguish the different data sets of HARs, the name of the first author of original publication is indicated in parentheses.

introduced in HARs, also followed this pattern, as they mostly consisted in changes from weak AT pairs to strong GC pairs.

Pollard et al. (2006a) proposed that a mechanism, that could explain this observation, is GC biased gene conversion (gBGC). This mechanism takes place during chromosomal recombination: when both strands on one chromosome undergo double-strand breaks that initiate meiotic recombination, a repair process that prefers to rebuild an allele containing guanines and cytosines, rather than adenines and thymines, can result (Mugal et al. 2015). As was shown later, this mechanism does not seem to explain the observed accelerated rate of substitutions to a high GC content in most of HARs (Kostka et al. 2012). In fact, the majority, 76%, of HARs appear to be driven by positive selection and only a small fraction, 14–19%, of HARs can be explained by this molecular mechanism, common in a number of eukaryotes, primarily aimed at preserving a higher GC content, which is otherwise lost given the spontaneous oxidative deamination of cytosines (Kostka et al. 2012; Mugal et al. 2015).

At the same time, Pollard et al. (2006a) excluded relaxation of evolutionary constraint in HARs. When compared with the neutral rate of nucleotide substitutions, 201 of 202 HARs showed an increased rate of substitutions, which indicates positive selection. If the fact that substitutions are fixed in the top five HARs is taken into account, then it is possible to assume that positive selection reverted back to negative selection at some point within the last 8 Myr of human evolution.

As was shown in this (Pollard et al. 2006a) and following studies (Hubisz and Pollard 2014), only ~3% of HARs are found within exons coding for a protein and another 5% are found within exons transcribed into a noncoding RNA. Perhaps surprisingly, promoters constitute only a small, <1%, fraction of HARs. But the vast majority, ~92%, of discovered HARs is found in intergenic regions and introns, and therefore could be regulatory sequences, like enhancers. A subsequent study estimated that ~30% of these sequences could in fact be enhancers (Capra et al. 2013). This, of course, raises the question about the function of the other 62% of HARs, which should constitute a different type of regulatory sequences, like insulators or silencers (Hubisz and Pollard 2014).

Analysis by Pollard et al. (2006a) of the neighboring genes with the Gene Ontology (GO) tool (Ashburner et al. 2000) revealed that the 202 HARs are found mostly near genes coding for transcription factors, DNA-binding proteins and regulators of nucleic acid metabolism.

Human-Accelerated Conserved Noncoding Sequences

The next study was published the same year, where the authors started with conserved noncoding sequences (CNSs), identified in eight vertebrates, including *Homo sapiens* (Prabhakar et al. 2006). The authors used another test statistic, in which they calculated a *P* value for each conserved region with human-specific substitutions. They discovered 992 human-accelerated CNSs (HACNSs) at $P \leq 0.005$, in

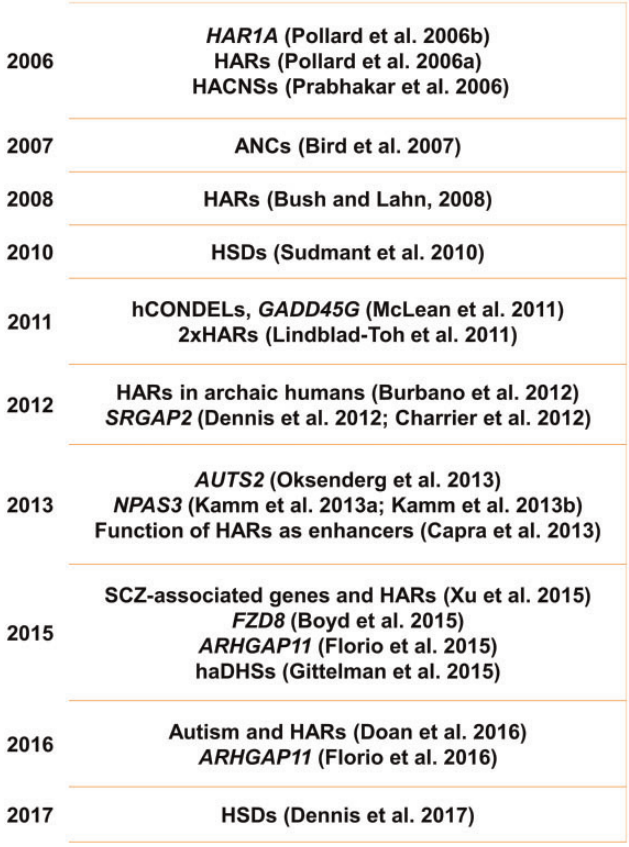

Fig. 2.—The timeline of studies describing human-specific sequence variations, mentioned in the present review.

which novel substitutions were 79% more numerous than what would have happened by chance.

The GO tool revealed that the cellular component, associated with the protein category most significantly enriched in these sequences was basal lamina, whereas the biological process was cell adhesion. Additional analyses showed that the most significant biological process was neuronal cell adhesion, with such examples as cadherins, protocadherins, contactins, and neuroligins.

Accelerated Conserved Noncoding Sequences

The study that followed described sequences that authors called accelerated conserved noncoding (ANC), discovered by leveraging another statistical method (Bird et al. 2007). First, a list of the most conserved, top 5%, noncoding (CNC) sequences, excluding open reading frames, was generated, by using MultiZ and PhastCons over 17 vertebrate genomes, from fish to mammals, including humans. Then, CNC sequences with more than four substitutions in human compared with chimpanzee were also aligned with the corresponding sequences of rhesus macaque. In the next step, Bird et al. (2007) used the χ^2-based relative rate test (Tajima 1993) to discover 1,356 ANC sequences that were accelerated in *Homo sapiens*, at $P \leq 0.08$. This number shrunk to 1,145 ANC sequences when segmental duplications, copy number variants, pseudogenes, and retroposons, as potential alignment artifacts, were excluded.

About 15–19% of the 1,356 ANC sequences were estimated by Bird et al. (2007) to have undergone positive selection. Using a different approach, derived allele frequency (DAF) spectrum of SNPs from the phase II HapMap (International HapMap Consortium et al. 2007), Bird et al. (2007) determined that DAF in ANCs is significantly higher than in all other sequences tested, which, again, indicated that ANC sequences are subject to positive selection. Nevertheless, this implies that the majority, ~80%, of ANC sequences are not subject to it and that these sequences, otherwise highly conserved in other species, lose evolutionary constraint in *Homo sapiens*. This in turn means that the regulatory sequences lose their previous function either because it switched to other conserved regulatory sequences during human evolution, or because this loss of regulation of gene expression was necessary to human evolution. An example of the latter possibility is a deletion of the enhancer of the gene *GADD45G*, which seems associated with expansion of several human brain regions (McLean et al. 2011).

Bird et al. (2007) further investigated whether the ANC sequences, that appear to be mostly due to loss of evolutionary constraint, result from duplications, in a manner, similar to pseudogenes. As was shown, a significantly greater number of ANC sequences are included in segmental duplications and copy number variants (CNVs), compared with nonaccelerated sequences. Furthermore, these duplications are very recent in evolution. This led Bird et al. (2007) to the conclusion that 8% of ANC sequences act like duplicated regulatory elements that degenerated with time.

To study the impact of human-specific substitutions on gene expression, Bird et al. (2007) estimated associations between SNPs from the phase II HapMap within ANC sequences and gene expression levels from the 210 unrelated HapMap individuals. Authors found three, 58 and 458 associations at adjusted P values = 0.0001, 0.001 and 0.01, respectively.

Human Accelerated Regions

The next study appeared a year later, to describe another data set of accelerated regions, discovered by using a likelihood ratio test, aimed at identifying only noncoding regions, by accounting for local sequence variation rates (Bush and Lahn 2008). The authors started by extracting elements, conserved among six eutherian mammalian species, including human, as determined by PhastCons and MultiZ software. Then each conserved element was considered together with nearby repeated sequences that were assumed to have neutral substitutions; conserved elements, that could have undergone positive selection, were therefore compared against this backdrop. The test statistic, at an FDR-corrected P value = 0.1,

indicated 63 extremely rapidly changing regions in *Homo sapiens*. Contrary to results by Pollard et al. (2006a), these 63 HARs do not have a significant bias of AT to GC substitutions; although, they are found more frequently near telomeres.

2xHARs

A further study appeared in 2011 and used 29 mammalian genomes, sequenced to 2× coverage, to assess regions in the human genome, affected by evolutionary constraint (Lindblad-Toh et al. 2011). The conserved regions were defined using MultiZ and PhastCons, excluding the human genome. Then Lindblad-Toh et al. (2011) used the software PhyloP (Pollard et al. 2010), also from the PHAST package, to reveal 563 2xHARs at an FDR-corrected *P* value <0.1. Additionally, Lindblad-Toh et al. (2011) discovered 1,930 2xHARs that were conserved in five primate species, but accelerated in human. Interestingly, 2xHARs were found in 0.2% of primate-conserved elements, but only in 0.04% of mammalian-conserved elements, which suggests that accelerated rate of substitutions happens primarily in regulatory regions with less deep evolutionary conservation (Lindblad-Toh et al. 2011). In accordance with previous results, gBGC is not responsible for 85% of the human-specific substitutions in 2xHARs. Nearly 10% of these regions overlap with enhancers, whereas genes, found nearby or containing the 2xHARs, are involved in extracellular signaling, receptor activity, immunity, cartilage development, and embryonic pattern specification (Lindblad-Toh et al. 2011).

Human-Accelerated DNAse I Hypersensitive Sites

The most recent study describing an original data set of HARs appeared in 2015 (Gittelman et al. 2015), in which authors undertook a different approach: they started with a list of human DNAse I hypersensitive sites (DHS), which could be active regulatory regions, previously determined in the ENCODE and Roadmap Epigenomics projects (Dorschner et al. 2004; Bernstein et al. 2010; Maurano et al. 2012). The reason to use this approach in addition to comparative genomics was that, according to several studies, there is a substantial functional turnover in conserved regulatory sequences in different species: despite conservation, regulatory regions may by active in some species and inactive in others (Dermitzakis and Clark 2002). By selecting DHSs, Gittelman et al. (2015) attempted to limit their investigation to regions with some molecular evidence of regulation. Then, Gittelman et al. (2015) compared these sites among six primate species, including our own, whose genomes were aligned in Ensembl V.70 (Paten et al. 2008; Flicek et al. 2014). It is worth to note that primate-conserved DHSs contained 7% of less conserved sequences, which underlines the difference between the simple sequence comparison and the approach where evidence from functional studies is taken into account (Gittelman et al. 2015). Then human-specific substitutions in DHSs were compared with surrounding 50 kb that were considered to evolve under the neutral evolutionary model. To this end, the PhyloFit software from the PHAST package (Hubisz et al. 2011) was used to compare the neutrally evolving 50 kb of surrounding sequence. Then, PhyloP was used to reveal DHSs conserved in other species, but accelerated in *Homo sapiens*, at an FDR-corrected *P* value = 0.05 (Gittelman et al. 2015). This allowed determining 524 regulatory regions, named human-accelerated DHSs (haDHSs). These regions evolved on an average 4 times faster than the neutral rate; in comparison, in other five primate species, the rate of substitutions was <0.5 of the neutral rate, meaning that in other primates the regions are under negative selection and conserved. Noncoding regions were overrepresented at these sites, at $P = 1.16 \times 10^{-7}$, in comparison with conserved nonaccelerated DHSs, which in turn included a higher proportion of exons. Gittelman et al. (2015) also showed that, in accordance with previous results, the majority, 70%, of substitutions in haDHSs were explained by positive selection. gBGC possibly explains <10% of haDHSs.

Furthermore, the majority, 64%, of haDHSs were found in tissue samples coming from the brain. Gittelman et al. (2015) then went on to demonstrate that haDHSs are actually enhancers: LacZ transgenic mice and luciferase reporter assays in two cell lines were used to reveal the total of 32 enhancers among the 75 haDHSs tested. These enhancers were active mostly in brain regions, midbrain, and forebrain, and in cells derived from neuroepithelioma. In addition, data generated with the chromosome conformation capture (3 C) technique Hi-C (Lieberman-Aiden et al. 2009; Dixon et al. 2012) was used by Gittelman et al. (2015) to identify points of contact between haDHSs, which are putative enhancers, and putative promoters near genes, by studying the 3 D chromatin conformation. Altogether, Gittelman et al. (2015) identified nearly 9,000 such points in the human genome. Interestingly, when compared with conserved nonaccelerated DHSs, haDHSs contacted fewer genes on an average, which suggested that adaptive evolution leads to a more narrowly targeted regulation of gene expression. When using categories from GO, authors found that the majority of haDHSs, at a corrected *P* value <0.05, were contacting promoters of genes implicated in development, brain, and neuron development in particular.

Comparative Analysis of HAR Data Sets

In a number of studies described above authors estimated the degree of intersection between different HAR data sets. Bird et al. (2007) compared their results with the two previous studies, Pollard et al. (2006a) and Prabhakar et al. (2006). This comparison was rather disappointing, because only 15, a minority of sequences, overlapped among the three studies. Furthermore, 51 sequences overlapped between Pollard et al. (2006a) and Prabhakar et al. (2006), 37 between Pollard et al. (2006a) and Bird et al. (2007), and 159 between Prabhakar

et al. (2006) and Bird et al. (2007). Nevertheless, HAR1, the most significant region from Pollard et al. (2006a), is the same as the most significant ANC from Bird et al. (2007), which is certainly encouraging. If these results are compared with subsequent studies, then the picture is similar: three studies (Bush and Lahn 2008; Kamm, Pisciottano, et al. 2013; Gittelman et al. 2015) showed that three, four, and five different data sets of HARs, respectively, excluding the data set of 1,356 ANCs by Bird et al. (2007), had only one HAR common among these data sets, HAR2 (Pollard et al. 2006a) or HACNS1 (Prabhakar et al. 2006).

Bird et al. (2007) also compared DAF from their results with DAF in HARs described by Pollard et al. (2006a) and Prabhakar et al. (2006). In accordance with previous results, DAF in HARs by Pollard et al. (2006a) were indeed compatible with positive selection. As to accelerated regions by Prabhakar et al. (2006), their DAF indicated loss of constraint and evolutionarily neutral substitutions. These comparisons among different studies certainly raise the question about sensitivity and specificity of the different statistical methods employed to discover HARs.

Figure 3 presents the degree of intersection among the six HAR data sets. Following methods were used: Five HAR data sets have been liftovered to the Genome Reference Consortium version 37 of the human genome (GRCh37), using the UCSC LiftOver tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver; last accessed November 20, 2017). The data set by Gittelman et al. (2015) was already in GRCh37 coordinates. Seven HARs from the data set by Lindblad-Toh et al. (2011) were absent in GRCh37. To estimate the degree of overlap between HAR data sets we computed the Jaccard Index using the GenometriCorr R package V. 1.1.17 (http://genometricorr.sourceforge.net; last accessed November 20, 2017) (Favorov et al. 2012). A heatmap showing similarity in pairwise overlap between genomic regions from different data sets demonstrated that HAR collections have little in common, at an FDR-corrected P value <0.01. The maximum degree of overlap is 7% between data sets, generated by Pollard et al. (2006a) and Lindblad-Toh et al. (2011) (fig. 3*a*). We also intersected available HAR data sets using bedtools intersect module from the bedtools software suite (http://quinlanlab.org/#portfolioModal1; last accessed November 20, 2017; http://bedtools.readthedocs.io/en/latest/; last accessed November 20, 2017). The length of overlapping regions was set to ≥ 1 bp. The results were visualized using the UpSetR R package V. 1.3.3 (http://caleydo.org/tools/upset/; last accessed November 20, 2017; https://github.com/hms-dbmi/UpSetR; last accessed November 20, 2017) (Conway et al. 2017). There is one common HAR among the six data sets (fig. 3*b*). Its coordinates in GRCh37 are: chr2: 236774014–236774088. It is also known as HAR2 (Pollard et al. 2006a) or HACNS1 (Prabhakar et al. 2006).

Table 2 summarizes properties of the six different HAR studies. These data allow concluding that different statistical and bioinformatics approaches used by different authors resulted in data sets with different properties. For instance, some studies considered coding and noncoding sequences, whereas others limited their investigation only to noncoding regions. Various categories of species were used for alignments and definition of conserved regions, ranging from only six primates, to 29 mammals to 17 vertebrates.

Comparison reveals HAR data sets with findings that seem more robust than others, thanks to the chosen methodology (table 2). First of all, only data sets that also included coding regions seem to be more representative of the biological reality. This allows selection of data sets by Pollard et al. (2006a), Lindblad-Toh et al. (2011), and Gittelman et al. (2015). In fact, only these three studies reported a high percentage of HARs being explained by positive selection (from 70% to 85%). The question of evolutionary conservation of the regions remains open. The original idea behind HARs was that these sequences had to be conserved deep in evolution. On the other hand, as Lindblad-Toh et al. (2011) showed, HARs actually tend to be found more frequently in sequences conserved in primates, not in sequences conserved in mammals. This observation could be explained by the fact that sequences conserved deep in evolution determine some very important molecular functions that have to be strictly guarded from change. Primates are an evolutionarily younger order and sequences conserved only in primates are relevant to new phenotypic features and seem to be more tolerant toward further change. Both the argument for the use of vertebrate conservation and the argument for the use of primate conservation seem to be convincing. Perhaps, one way to resolve this ambiguity is to undertake the same approach as Lindblad-Toh et al. (2011), who performed separate analyses for mammal-conserved and primate-conserved regions. Finally, functional evidence of enhancer activity is present in only one study, Gittelman et al. (2015).

Further Investigation of the Discovered HAR Data Sets

Further four studies appeared in 2012, 2013, 2015, and 2016 and aimed at integrating the available data sets of HARs, in order to investigate additional aspects of these regions.

HARs in Archaic Humans

The first study (Burbano et al. 2012) aimed to estimate the degree to which HARs, determined in four studies (Pollard et al. 2006a; Prabhakar et al. 2006; Bird et al. 2007; Bush and Lahn 2008), were shared with Neanderthals and Denisovans (Green et al. 2010; Reich et al. 2010; Meyer et al. 2012; Prufer et al. 2013). The results showed that 8.3% of nucleotide substitutions in HARs are not shared with the archaic hominins and are modern human-specific, whereas the rest of substitutions appeared before the split from these hominins that took place 500,000 years ago

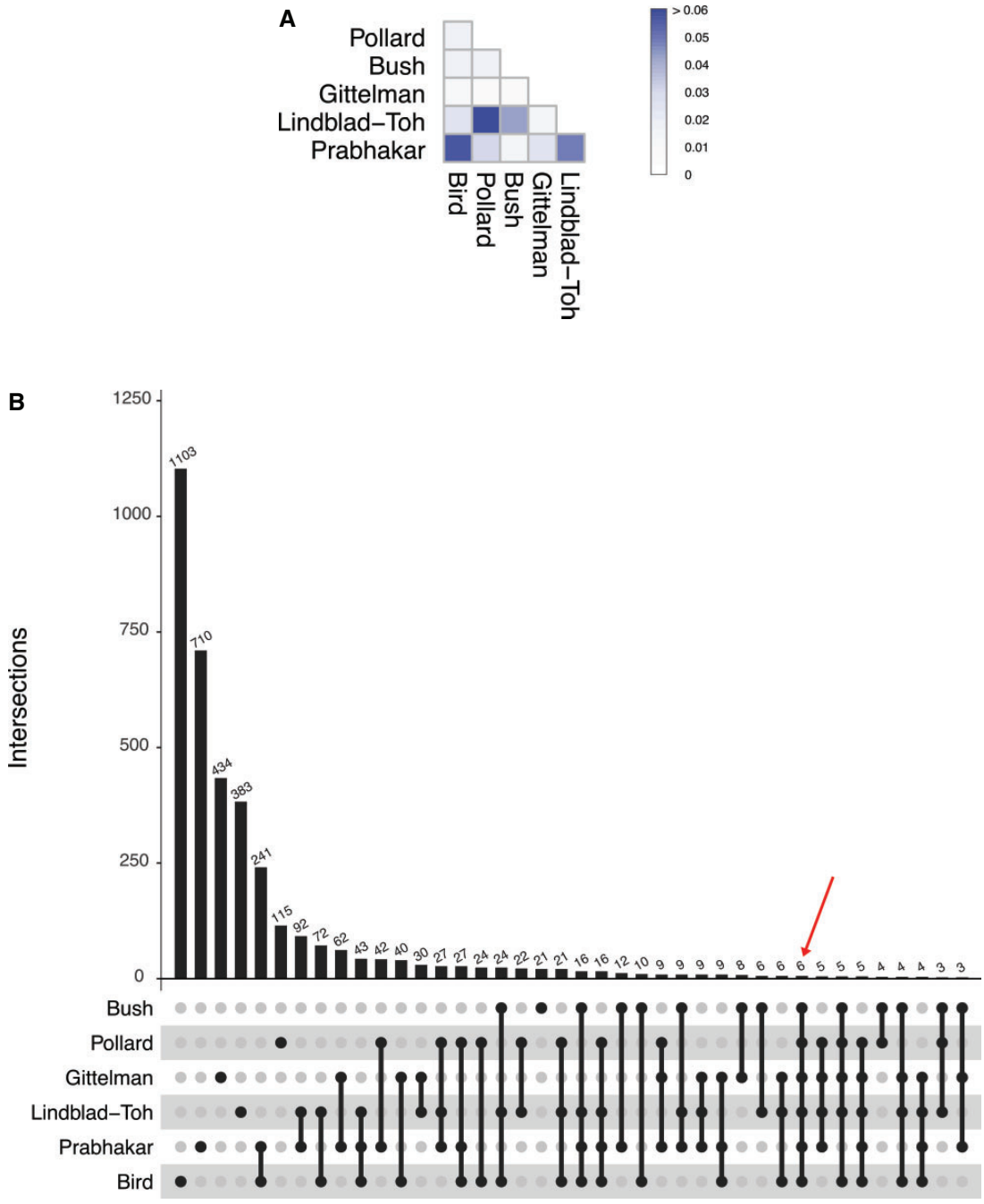

Fig. 3.—Intersection statistics for the six HAR data sets. (*a*) A heatmap showing pairwise comparison between data sets. (*b*) Intersection between data sets. Each data set is represented by a *black filled circle*. A *vertical black line* connects the circles to emphasize intersections between corresponding data sets. The number of intersections between HARs is shown as a bar chart. The *arrow* indicates intersection between the six data sets. *Bush* HARs from Bush and Lahn (2008), *Pollard* HARs from Pollard et al. (2006a), *Gittelman* haDHSs from Gittelman et al. (2015), *Lindblad-Toh* 2xHARs from Lindblad-Toh et al. (2011), *Prabhakar* HACNSs from Prabhakar et al. (2006), and *Bird* ANCs from Bird et al. (2007).

(Prufer et al. 2013). Similar estimates of 7.1% were published in a later study (Hubisz and Pollard 2014).

Function of HARs as Enhancers

The second publication (Capra et al. 2013) integrated data from five previously published studies (Pollard et al. 2006a; Prabhakar et al. 2006; Bird et al. 2007; Bush and Lahn 2008; Lindblad-Toh et al. 2011). The overlapping regions were merged and all protein-coding sequences were omitted, because the main goal of that meta-analysis was to test the hypothesis that HARs are, to a significant degree, enhancers. This resulted in a list of 2,649 noncoding (nc) HARs (Capra

Table 2
Comparison of HARs Methods

HAR Data Set[a]	Only Noncoding Regions Considered?	Human Genome Used to Define Conserved Regions?	Evidence of Conservation	Tools Used for Alignments and Estimation of Conservation	Statistical Method Used to Estimate Acceleration	Percentage of HARs Explained by Positive Selection	Additional Functional Evidence
HARs (Pollard)	No	Yes	17 vertebrates	MultiZ, PhastCons	LRT[b]	76%	no
HACNSs (Prabhakar)	Yes	Yes	8 vertebrates	UCSC Genome Browser (PhastCons)	Human-acceleration *P* value	not estimated	no
ANCs (Bird)	Yes	Yes	17 vertebrates	MultiZ, PhastCons	χ^2-based relative rate test	15–19%	no
HARs (Bush)	Yes	Yes	6 mammals	MultiZ, PhastCons	LRT[b]	not estimated	no
2xHARs (Lindblad-Toh)	No	No	29 mammals	MultiZ, PhastCons	LRT[b] (PhyloP)	~85%	no
haDHSs (Gittelman)	No	Yes	6 primates	Ensembl Genome Browser	LRT[b] (PhyloP)	70%	yes

LRT, likelihood ratio test.
[a]The total numbers of discovered HARs are indicated in table 1.
[b]A likelihood ratio test with different author-defined parameters.

et al. 2013). 95% of these sequences were intronic or intergenic, the average length of ncHARs was 257 bp and the average distance from the transcription start site was 307 kb. These regions were found near genes mostly involved in development, of the brain in particular; furthermore, almost 60% of ncHARs were within 1 Mb of a gene differentially expressed between chimpanzee and human. In addition, as revealed by Capra et al. (2013), nearly 60% of ncHARs overlapped an enhancer, as revealed by H3K4me1 and H3K27ac, histone epigenetic marks, or p300, a transcriptional coactivator indicating the position of enhancers (Asahara et al. 2002; Visel et al. 2009). Further evidence, produced by a machine-learning algorithm called EnhancerFinder (Erwin et al. 2014), indicated that almost 30% of ncHARs are enhancers implicated in development (Capra et al. 2013) (fig. 1*d*). The largest proportion, almost 1/3 of these predicted enhancers is active in the brain. When comparing enhancer activity of ncHARs between human and chimpanzee, Capra et al. (2013) discovered that, again, nearly 60% of these regions bind at least one TF different between the two species. Finally, Capra et al. (2013) studied in detail the enhancer activity of five ncHARs with consistent pattern of expression and differences between chimpanzee and human, by using LacZ transgenic mice. These enhancers drove different patterns of expression in the central nervous system and limbs, underscoring the morphological differences between the two primate species.

Evaluation of Interaction between Schizophrenia-Associated Genes and HARs

The third integrative study (Xu et al. 2015) investigated an overlap between HARs and genome-wide association study (GWAS) SNPs, significantly associated with schizophrenia (SCZ). SCZ is a disorder that affects higher mental functions, proper in its full picture only to *Homo sapiens*; it is therefore reasonable to conclude that genes or regulatory regions, implicated in the pathogenesis of SCZ, are the same ones that are exclusive to the human brain. Authors considered SCZ-associated SNPs from a meta-analysis performed by the Psychiatric Genomics Consortium (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014). Then Xu et al. (2015) imputed other SNPs in linkage disequilibrium (LD) with the GWAS SNPs in order to define 52-kb-long SCZ-associated genomic regions, using LD calculations procedure of the PLINK software (Purcell et al. 2007). Xu et al. (2015) also considered HARs from the study by Lindblad-Toh et al. (2011), in which these regions were determined using separately conservation in mammals (mHARs) and conservation in primates (pHARs). Genes within 100-kb upstream and downstream from every HAR were considered as possibly regulated by that HAR. In the next step, the 100-kb HAR-associated regions and 52-kb SCZ-associated regions were tested for the presence of an overlap, using the INRICH software, which

can account for confounding factors, such as SNP density or overlapping genes. Xu et al. (2015) used a panel of different *P* values; the results showed that the SCZ-associated regions were enriched in genes, associated with pHARs, at all corrected *P* values, ranging from $<10^{-2}$ to 10^{-5}. These pHAR-associated SCZ genes were in addition found to be the most conserved in evolution of primates, because the ratio of nonsynonymous (potentially functional) to synonymous (likely neutral) substitutions during human evolution, denoted as dn/ds, was low in this category of genes, compared with other categories, such as all autosomal genes, SCZ genes, or HAR-associated genes. These results may seem counterintuitive, because HARs are defined as regions that changed in human under positive selection and were not subject to evolutionary constraint, compared with other species. However, HARs are not necessarily genes, but more likely regulatory sequences, somewhere near SCZ genes. In this scenario, evolutionarily conserved and therefore functionally important genes may have acquired new regulation of transcription via new TF-binding sites, which defined the appearance of *Homo sapiens* as a species.

In the next step, the Xu et al. (2015) verified whether pHAR-associated SCZ genes make part of any coexpressed module in the human brain. A previous study described 23 coexpression modules, using 21 SCZ brain samples and 19 controls, of which five were differentially expressed between SCZ cases and controls (Roussos et al. 2012). Of these five modules, one module, enriched in genes, coexpressed in the inhibitory GABAergic neurons, was found to contain the same genes as the pHAR-associated SCZ genes (Xu et al. 2015). This result was revealed by the weighted gene coexpression network analysis (WGCNA), at a corrected *P* value = 0.017 (Zhang and Horvath 2005). The pHARs-associated SCZ genes are implicated in brain development, as was indicated by GO categories in which the genes were enriched (Xu et al. 2015).

Finally, Xu et al. (2015) evaluated the place pHAR-associated SCZ genes occupy in gene networks. To that end, first, gene expression patterns in additional 220 human prefrontal cortex samples, without a psychiatric or neurological disease, were analyzed with WGCNA. Similar to the previous analysis, 36 gene modules were determined. The pHAR-associated SCZ genes, same as all SCZ genes and all pHAR SCZ-unrelated genes were found to be more closely connected to each other, in comparison with all other genes in the networks. Second, an additional sample of dorsolateral prefrontal cortex tissues from 173 individuals without dementia was investigated by Xu et al. (2015), by constructing a gene interaction network with the help of RIMBANET software package and Cytoscape 3.1.1 (Cline et al. 2007; Zhu et al. 2008). The results showed that pHAR-associated SCZ genes were located only inside and at the center of the largest interconnected gene module, active in the human prefrontal cortex (Xu et al. 2015). These results indicate that evolutionarily conserved brain genes that acquired new patterns of regulation in *Homo sapiens*, seem to be interconnected and to form a nucleus in the largest gene network component, expressed in the human prefrontal cortex.

Evaluation of Association between Autism and HARs

A further study revealed an association between rare homozygous single nucleotide variants and indels in several enhancer-containing HARs and autism, another neurodevelopmental disorder, in consanguineous families (Doan et al. 2016). Furthermore, a significantly greater proportion of de novo, rare, and heterozygous, CNVs in other cases of autism was found to hit a HAR, when compared with all other cases of de novo CNVs (Doan et al. 2016). The studied HARs, that authors extracted from five previous studies (Pollard et al. 2006a; Prabhakar et al. 2006; Bird et al. 2007; Bush and Lahn 2008; Lindblad-Toh et al. 2011), were also found to be enriched in TF binding sites active in the brain. The enrichment was found by Doan et al. (2016), in particular, for sites binding SRY-box 2 (SOX2), a TF that regulates neural progenitor renewal, preventing these stem cells from early differentiation into mature neurons, apparently through inhibition of the canonical WNT pathway (Ferri et al. 2004; Kelberman et al. 2008). This pathway is implicated in a great variety of intracellular molecular processes and active virtually in all tissues in ontogeny, that is, from development to adult life (Clevers 2006); in particular, it is important for brain development and function (Freese et al. 2010; Noelanders and Vleminckx 2017). Furthermore, the WNT pathway has been shown to play a role in SCZ and other neurodevelopmental disorders (Mao et al. 2009; Singh 2013; Tucci et al. 2014; Levchenko et al. 2015), which again points to its importance in human brain development.

To prove interaction between HARs, acting as enhancers, and promoters of genes, the Doan et al. (2016) used a combination of 3C techniques: their own 4C-sequencing (Matelot and Noordermeer 2016) and previously published data obtained with ChIA-Pet (Li et al. 2012) and Hi-C (Jin et al. 2013). This allowed determining >500 HARs that physically interact with promoters. Unfortunately, Doan et al. (2016) did not integrate the list of HARs from Gittelman et al. (2015), which was created taking into account enhancers with some functional evidence.

Studies of Individual Genes and Regulatory Elements That Contain HARs

Although a few HARs were found to be enhancers important for limb development, like HAR2 and 2xHAR114 (Prabhakar et al. 2008; Sumiyama and Saitou 2011; Hubisz and Pollard 2014), most attention was given to HARs that are part of or regulate genes important for development of the human brain (table 1). Figure 1*d* presents a theoretical example of a HAR that regulates gene expression level or expression pattern.

HAR1A

The first discovered genes associated with a HAR were *HAR1A* (also, *HAR1F*) and *HAR1B* (also, *HAR1R*), described in the pioneering studies (Pollard et al. 2006a, 2006b). The genes are small, having two to three exons depending on the isoform, coding for functional RNAs, and are transcribed in opposite directions on chromosome 20q. Their respective exons 1 overlap with each other and with HAR1 at an FDR-corrected $P < 5 \times 10^{-4}$ (Pollard et al. 2006b). HAR1 itself is 118 bp long and contains 18 human-specific, fixed substitutions. The substitutions were estimated to occur ~1 Ma. The conserved region was estimated to be 310–400 Myr old, because it is conserved down to the bird (chicken) (Pollard et al. 2006b).

Both genes form secondary RNA structures helix-loop-helix as determined by EvoFold (Pedersen et al. 2006). Pollard et al. (2006b) studied the secondary structure of HAR1A in more detail, to reveal differences with the secondary structure in chimpanzee, which could explain different function in human, compared with other primates. A subsequent study determined a different RNA structure of HAR1A, with even more dramatic differences between human and chimpanzee (Beniaminov et al. 2008).

HAR1A was found to be expressed in Cajal–Retzius neurons, together with reelin, during neocortical development in human, as was demonstrated by RNA in situ hybridization and immunohistochemistry (Pollard et al. 2006b). In particular, *HAR1A* is first expressed at seven to nine gestational weeks (GW) in the dorsal telencephalon, which gives rise to the cerebral cortex; starting at 11 GW, *HAR1A* is expressed in the marginal zone, close to the pial surface, in cells, also expressing reelin (Pollard et al. 2006b). This expression is maintained until 17–19 GW and at 24 GW is no longer observable in Cajal–Retzius neurons. At this point, *HAR1A* is expressed in the developing hippocampus, cerebellum, and medulla. Contrary to this RNA, *HAR1B* is almost nonexistent in the developing human brain, with expression 50-fold lower, as was shown by quantitative real time PCR (qPCR). As to the adult brain, in situ hybridization revealed expression of *HAR1A* in the frontal lobe and hippocampus, whereas qPCR indicated high levels of expression in the cerebellum, diencephalon, and frontal lobe (Pollard et al. 2006b).

An obvious question is how the human-specific substitutions in the conserved region, that includes overlapping exons 1 of the two HAR1 genes, are responsible for human-specific brain phenotype. As mentioned before, all 18 human-specific substitutions in HAR1 are from weak AT to strong GC pairs (Pollard et al. 2006b). This strengthens RNA helices in the 3 D structure, possibly by creating a stable cloverleaf-like structure in human, instead of an extended and unstable hairpin in chimpanzee (Beniaminov et al. 2008). This could determine different functionalities of these RNAs. As listed in the Rfam RNA repository (Griffiths-Jones et al. 2005), both HAR1 RNAs belong to the category of long noncoding (lnc) RNAs [http://rfam.xfam.org/family/har1a accessed on June 13, 2017], which are implicated in a myriad of intracellular processes, from transcriptional and posttranscriptional regulation, to epigenetic regulation and maintenance of chromatin structure, often by acting in complexes with proteins (Wilusz et al. 2009; Chen 2016; Engreitz et al. 2016; Kim and Shiekhattar 2016). In particular, lncRNA are important in the human cerebral cortex development (Ng et al. 2013; Lipovich et al. 2014; Hart and Goff 2016). In this sense, the human version of HAR1A lncRNA may shape, for example, a different transcriptome in human Cajal–Retzius neurons, which are crucial for correct development of the six layers of the human cerebral cortex, the anatomical structure holding the most pronounced differences when compared with that in other species. Interestingly, the haplotype CCCCGC, formed by SNPs s6011613, rs2427496, rs6122371, rs750696, rs750697, and rs2427498, completely covering a region that includes *HAR1A* gene, was found to be associated with auditory hallucinations in 221 psychiatric patients, as confirmed by a Bonferroni-corrected $P = 0.017$ (Tolosa et al. 2008). Hallucinations are a downstream result of erroneous neuronal connections that formed during brain development (Owen et al. 2016); therefore, some functional alleles in *HAR1A* may alter activity of this lncRNA and therefore increase chances of psychosis.

AUTS2

The next studied HAR-associated gene, autism susceptibility candidate 2 (*AUTS2*), was previously found to be disrupted by structural variants in a multitude of psychiatric and neurological disorders, most of them neurodevelopmental (reviewed in Oksenberg and Ahituv 2013). *AUTS2* contains three HARs in its introns: one found by Pollard et al. (2006a), HAR31, and two, by Prabhakar et al. (2006), denoted as HACNS174 and HACNS369 (Oksenberg et al. 2013). Furthermore, roughly the first half of the region containing the first four exons of the gene is the region the most different between human and the Neanderthal (Green et al. 2010). The differences consist in 293 consecutive SNPs, which are variable in human, but have only ancestral alleles in the Neanderthal; neither of the SNPs, except one, changes the amino acids (aa) content, therefore they could have a regulatory role.

This gene is protein coding, contains up to 19 coding exons, depending on the isoform, as listed by GENCODE v.24 (Harrow et al. 2012), in which case the protein is 1,259 aa-long. The function of the protein is still enigmatic, but presence of the proline–tyrosine (PY) motive suggests it may function as a TF (Oksenberg et al. 2013). Interestingly, AUTS2 protein has also the histidine repeat that is associated with localization of a protein in nuclear speckles, a storage inside nucleus, where various transcription, splicing, and mRNA-processing factors are found when not in active use

(Salichs et al. 2009; Spector and Lamond 2011). LncRNAs has also been localized in nuclear speckles (Engreitz et al. 2016), so it is theoretically possible that AUTS2 and HAR1A interact during embryonic brain development, by regulating gene transcription. Furthermore, the T-box brain 1 TF activates the expression of both *Auts2* and Reln (reelin) (Bedogni et al. 2010a) that is coexpressed in Cajal–Retzius neurons with *HAR1A*). Functional in vitro or in vivo studies are necessary to explore this possibility of interaction.

The expression pattern of *AUTS2* in human is not limited to the brain and is quite ubiquitous (Oksenberg and Ahituv 2013) which seems to explain why patients with structural variants in this gene and autism or intellectual disability, are also afflicted with hypotonia, short stature, and urogenital and skeletal abnormalities (Sultana et al. 2002; Kalscheuer et al. 2007). As Oksenberg et al. (2013) demonstrated, *auts2* in zebrafish determines the development of the embryo in general and of its CNS in particular. As other studies demonstrated, *AUTS2* expression in human is localized in various parts of the developing brain, including the prefrontal cortex (Sultana et al. 2002; Lepagnol-Bestel et al. 2008; Zhang et al. 2011). The expression was furthermore detected in the nucleus of murine glutamatergic, GABAergic, and dopaminergic neurons (Bedogni et al. 2010b).

Oksenberg et al. (2013) therefore tested the hypothesis that the three HARs and the region of human-Neanderthal sweep contain enhancers. To that end, they first selected putative human enhancers, based on $\geq$70% sequence identity, for at least 100 bp between human and chicken, and on data from forebrain or hindbrain ChIP-Seq that used p300 as a marker. Next, the authors cloned the enhancers together with a promoter and the green fluorescent protein (GFP) gene in zebrafish, or with a promoter and the LacZ gene in mouse. Oksenberg et al. (2013) discovered 16 enhancers in zebrafish and then confirmed four of them in mouse. The four enhancers were found within HAR31, HACNS369, and the human-Neanderthal sweep region and drove expression of reporter genes exclusively in the CNS.

NPAS3

Just a month later the same year a new publication appeared, describing another TF, neuronal PAS domain-containing protein 3 (*NPAS3*), belonging to the basic helix-loop-helix PAS gene family, this time with 14 HARs within its genomic sequence (Kamm, Pisciottano, et al. 2013). The reason to study that gene was because it contained the highest number of HARs in the genome. To reveal the gene, Kamm, Pisciottano, et al. (2013) first considered the 1,629 nonredundant HARs from four previous studies (Pollard et al. 2006a; Prabhakar et al. 2006; Bush and Lahn 2008; Lindblad-Toh et al. 2011). Then, Kamm, Pisciottano, et al. (2013) determined the largest transcriptional units listed in the RefSeq database (Pruitt et al. 2007; O'Leary et al. 2016) and 1-Mb genomic regions with the highest concentration of HARs, using Galaxy tools (Goecks et al. 2010). The 14 HARs of *NPAS3* are found in introns of the gene, which has seven isoforms, according to GENCODE v.24, and up to 12 exons coding for 938 aa.

The gene, as shown by studies in mice, seems to be important in brain development and maintenance of normal neurosignaling (Erbel-Sieler et al. 2004; Brunskill et al. 2005) and determines normal maturation of the hippocampus (Sha et al. 2012). In humans, *NPAS3* is expressed in the developing neocortex, hippocampus, and cerebellum (Gould and Kamnasaran 2011). Interestingly, the gene was found to be disrupted by a chromosomal translocation in patients with SCZ and learning disability and SNPs in this gene were associated with SCZ and a closely related psychiatric illness, bipolar disorder (Kamnasaran et al. 2003; Pickard et al. 2005, 2006, 2009). These data point to a role of this gene in human brain evolution.

Out of the 14 HARs, 11 were found to act, similar to previous studies, as enhancers in the CNS, as was demonstrated by the expression reporter assay, containing a HAR, a promoter, and GFP, in transgenic zebrafish (Kamm, Pisciottano, et al. 2013). In particular, the 11 HARs drove a stable expression at 24- and 48-h postfertilization in many parts of the fish embryo, but also in the developing brain. Kamm, Pisciottano, et al. (2013) further demonstrated, by using the MatInspector software (http://www.genomatix.de/; last accessed November 20, 2017), that the numerous human-specific single nucleotide substitutions in the enhancers modify the affinity with which TFs bind to them.

Later the same year the same scientific team further studied one of the enhancers described in their previous study, 2xHAR142, located in intron 5 of *NPAS3* (Kamm, Lopez-Leal, et al. 2013). Authors used mice as a model this time. The authors discovered that the human version of 2xHAR142, compared with chimpanzee and mouse versions, drives an extended lacZ expression pattern in the murine nervous system.

FZD8

Another HAR-associated brain-expressed gene is Frizzled 8 (*FZD8*) whose enhancer contains a HAR that authors named HARE5 (Boyd et al. 2015). As a first step of their analysis, the authors build a list of 106 HAREs, segments were HARs, identified in four previous studies (Pollard et al. 2006a; Prabhakar et al. 2006; Bird et al. 2007; Lindblad-Toh et al. 2011), overlap with putative human enhancers. Boyd et al. (2015) established evidence of enhancer function by an in silico meta-analysis of two studies that used mouse embryonic neocortical tissue or neural stem cells to run ChIP-seq (Visel et al. 2009; Creyghton et al. 2010). The overlap between the four data sets of published HARs and mouse enhancers was determined using Exact Pairwise MAF blocks V.1.0.1, MAF to

BED V.1.0.0, and Intersect V.1.0.0 tools from the Galaxy platform (Goecks et al. 2010). HARE5 happened to be a region of overlap between an enhancer that binds p300 and ANC516 from Bird et al. (2007), near *FZD8*. The reason to give a particular importance to this gene was because it codes for a receptor in the WNT pathway.

To study differences in enhancer activity between HARE5 and its chimpanzee ortholog, Boyd et al. (2015) used transgenic mice and either the LacZ, GFP, or tdTomato reporter system. By comparing mice with either human, or chimpanzee enhancer, authors demonstrated that the human HARE5 drives a higher LacZ expression, starting at embryonic day 10. The fluorescent proteins on their turn showed an expression up to 30 times higher with the human enhancer. As Boyd et al. (2015) showed, this expression was specific to the mouse developing neocortex, the same location as for *FZD8*, known to be expressed in neural progenitors in mice and humans (Fischer et al. 2007; Miller et al. 2014). Furthermore, the results in Boyd et al. (2015) indicated that mice with human HARE5-containing constructs had significant acceleration of neural progenitor cell cycle and enlarged brain size. 3 C-qPCR assays (Hagege et al. 2007), further deployed in Boyd et al. (2015), demonstrated that HARE5 and the promoter of *FZD8*, which are 307,758 bp apart, interact physically and specifically in the developing mouse neocortex.

CUX1, PTBP2, and GPC4

The study mentioned in the previous section, by Doan et al. (2016), reported a curated list of eight genes, important in brain development or associated with a number of neurodevelopmental disorders, that interact with HAR-containing brain enhancers, homozygously mutated in the reported cases of autism. Three of the genes were studied in a greater detail. Cut-like homeobox 1 (*CUX1*) interacts with HAR426, reported in Prabhakar et al. (2006). This enhancer gains new TF binding sites because of the substitution G > A, found in recessive cases of autism with intellectual disability (ID) (Doan et al. 2016). The gene is known to play important roles in dendritic development in layer II–III neurons of the cerebral cortex in mice and was already predicted to have a role in autism (Cubelos et al. 2015). It seems to act as a transcriptional repressor (Snyder et al. 2001) and, according to results from Doan et al. (2016), its overexpression and the substitution G > A inhibit normal dendritic pruning.

Polypyrimidine tract binding protein 2 (*PTBP2*) is regulated by HAR169, reported in Bird et al. (2007). The accelerated region is disrupted by a small indel changing GGGTAC to A in two siblings with autism and ID (Doan et al. 2016). This indel creates new and removes existing TF binding sites. The gene regulates splicing in developing murine neurons, by acting as an inhibitor of isoforms proper to mature neurons (Licatalosi et al. 2012; Li et al. 2014), and has been previously shown to be disrupted in cases of autism and ID (Carter et al. 2011; Willemsen et al. 2011). 4 C-sequencing confirmed interaction between the enhancer and the promoter of *PTBP2* and luciferase reporter gene analyses in neuronal cells indicated a 40–50% reduction in enhancer activity (Doan et al. 2016).

Finally, glypican 4 (*GPC4*) is regulated by HAR1325, also reported in Bird et al. (2007). The regulatory region is disrupted by two sequence variations, T > C and a deletion of T, found at different genomic locations and in two unrelated families with autism and ID (Doan et al. 2016). The sequence variations remove several TF binding motifs and reduce enhancer activity in luciferase assays (Doan et al. 2016). The gene product is a signaling molecule, a proteoglycan secreted by astrocytes and inducing glutamatergic synapse formation in the murine hippocampus (Allen 2012).

Human-Specific Segmental Duplications, Deletions, and Other Evolutionarily Novel Sequence Variations

Important to mention that, apart from evolutionary novel regulation through single nucleotides substitutions, found in HARs, human-specific deletions, and duplications can alter gene expression levels or create novel transcripts. A recent paper (Dennis et al. 2017) continued previous efforts (Sudmant et al. 2010) that laid the path for the new direction in the study of human evolution from the point of view of genomic rearrangements, by assessing human-specific segmental duplications (HSDs) (Eichler 2001), absent in other great apes: orangutan, gorilla, and chimpanzee. In addition to whole genome deep sequencing used in these two studies, Dennis et al. (2017) deployed a genome assembly of the CHM1 haploid hydatidiform mole, an innovation not used in previous whole genome studies of human evolution. This method was previously used by the same scientific team (Dennis et al. 2012) and allows identifying complex genomic rearrangements more accurately, because only one allele of highly redundant sequences is present. In Sudmant et al. (2010) authors discovered 23 genes with HSDs, whereas the authors of Dennis et al. (2017) discovered 218 HSDs, ranging in size from 5 to 362 kb. Earlier studies (Locke et al. 2003; Fortna et al. 2004; Dumas et al. 2007) used array comparative genomic hybridization, a technique nowadays outdated, because it is less sensitive (Sudmant et al. 2010). An apparent drawback of these five studies (Locke et al. 2003; Fortna et al. 2004; Dumas et al. 2007; Sudmant et al. 2010; Dennis et al. 2017) is that genomes of other mammals or vertebrates were not included in the comparison (Dumas et al. 2007 compared humans with nine other primate species and the remaining studies compared humans with other apes), obviously given the formidable challenge of correctly discovering large and complex rearrangements in multiple genomes. This comparison seems important, because it is possible that rearrangements in *Homo sapiens*, absent in great apes, are present in some lower species and therefore

their relevance for human evolution will be less convincing. As we wait for further investigation of the relevance of the numerous human-specific genomic rearrangements, there are already three examples of studied genes, with a role in brain development, subject to human-specific duplications and deletions (table 1). Figure 1e presents a theoretical example of segmental duplications that generated at least one gene with new function.

SRGAP2 Paralogs

One example of functional genomic rearrangements specific to humans are three consecutive partial duplications of the ancestral gene Slit-Robo Rho GTPase-activating protein 2 (*SRGAP2*), giving rise to paralogs *SRGAP2A* (also, *SRGAP2*), *SRGAP2B*, *SRGAP2C*, and *SRGAP2D*, revealed by Fortna et al. (2004) and later confirmed by Sudmant et al. (2010) and Dennis et al. (2017). The paralogs' structure and function were then studied in further detail (Charrier et al. 2012; Dennis et al. 2012). *SRGAP2* is not duplicated in 13 mammals, including chimpanzee, and the number of gene copies in humans is constant (Sudmant et al. 2010; Dennis et al. 2012). The copies *SRGAP2B* and *SRGAP2C* were also found to have an accelerated rate of nucleotide substitutions (Dennis et al. 2012). Dennis et al. (2012) estimated the duplications to happen from ~3.4 to 1 Ma, which is after the split from the common ancestor with chimpanzee and corresponds to the transition from Australopithecus to Homo, at the beginning of the human neocortex expansion. In agreement with these data, the duplications are also present in the Neanderthal and the Denisovan (Sudmant et al. 2010; Dennis et al. 2012). The apparent mechanism was mediated through AluS and AluY repeats, found precisely at the boundaries of these duplications (Dennis et al. 2012); Alu repeats are strongly associated with genomic duplications in primates (Bailey et al. 2003; Zhou and Mishra 2005).

According to Dennis et al. (2012), only *SRGAP2A* and *SRGAP2C* code for functional proteins, whereas *SRGAP2B* and *SRGAP2D* could be pseudogenes. Note that the current release 89 of Ensembl (Yates et al. 2016), containing the GRCh38 genome assembly (Schneider et al. 2017), states that *SRGAP2B* is not a pseudogene; although the current release 2017_06 of the UniProt database (The UniProt Consortium 2017) indicates that the evidence for this protein is only inferred from homology; as to *SRGAP2D*, it is indeed listed as a pseudogene in Ensembl and NCBI's RefSeq databases.

SRGAP2C is a 458 aa-long truncated copy of SRGAP2A, containing only a portion of the Fes/CIP4 homology Bin-Amphiphysin-Rvs (F-BAR) domain that is implicated in neuronal migration and morphogenesis through cellular membrane remodeling (Guerrier et al. 2009). SRGAP2C, in transfected COS7 fibroblast-like cell cultures, dimerizes with and inhibits SRGAP2A (Charrier et al. 2012) that contains three functional domains: F-BAR, Rho GTPase-activating protein (RhoGAP) (Barrett et al. 1997), and SRC Homology 3 (Mayer 2001). In fact, the ancestral *SRGAP2A* is known to play a role in cortical development, by negatively regulating neuronal migration (Guerrier et al. 2009; Guo and Bao 2010). It seems that the loss of 49 aa in the truncated F-BAR domain is necessary for the new function of SRGAP2C (Charrier et al. 2012). *SRGAP2A* and the human-specific *SRGAP2C* are both expressed in the embryonic and adult human brain. In particular, expression of the two paralogs was detected in the developing human neocortex: ventricular zone, SVZ, and cortical plate (Charrier et al. 2012). Confirming previous results (Guerrier et al. 2009), Charrier et al. (2012) showed that *SRGAP2A*, when overexpressed by *in utero* electroporation of murine cortical neural progenitors, induces excessive branching of the leading process of neurons, migrating from the ventricular zone to the cortical plate; this limits the number on neurons that finally reach the cortical plate. In addition, SRGAP2A accumulates at the postsynaptic density of excitatory synapses, promotes dendritic spine maturation, and limits spine density. On the contrary, the inhibition of SRGAP2A by SRGAP2C in mice, subjected to *in utero* electroporation, delays maturation of spines in pyramidal neurons, increases spine length and density, leads to a more rapid radial migration of neurons in the developing neocortex, and results in a greater number of neurons reaching the cortical plate (Charrier et al. 2012). These changes correspond to the known evolutionary novelties of the human neocortical pyramidal neurons, especially the ones in the prefrontal cortex, which play an important role in cognitive abilities (Elston et al. 2001; Benavides-Piccione et al. 2002).

Interestingly, several neurodevelopmental disorders are associated with CNVs in these paralogs: deletions and duplications, involving *SRGAP2A* and *SRGAP2C*, were found in several cases of ID and autism (Dennis et al. 2012). These two paralogs are not subject to CNVs in the normal human population; this is not the case of *SRGAP2B*, a paralog often altered by CNVs in the normal population, which again suggests that it is not very active or important in the human brain (Dennis et al. 2012). Especially remarkable is the fact that the evolutionarily recent paralog *SRGAP2C* is the least variable gene in terms of CNVs among the 23 genes that underwent human-specific duplications (Sudmant et al. 2010); this underlines its importance for the human brain function.

ARHGAP11 Paralogs

Another example of human-specific paralogs is Rho GTPase-activating protein 11 A (*ARHGAP11A)* and B (*ARHGAP11B)*, revealed in the genome-wide study of human-specific genomic rearrangements (Sudmant et al. 2010) and in two studies of chromosome 15q13.3 genomic instability (Riley et al. 2002; Antonacci et al. 2014). The segmental duplication was later confirmed by Dennis et al. (2017). The paralogs were further

investigated from the perspective of their function during neurodevelopment by another scientific team (Florio et al. 2015).

Florio et al. (2015) studied, in mice and humans, apical and basal radial glia, cells that guide migrating neurons in the six layers of the developing neocortex and fulfill neural progenitor cell function, as they can differentiate during brain development into intermediate progenitors, neurons, and astrocytes (Lui et al. 2011; Florio and Huttner 2014; Taverna et al. 2014; Wilsch-Brauninger et al. 2016; Namba and Huttner 2017). Neural progenitors, which can be apical or basal, comprise radial glia and intermediate progenitors. Apical and basal radial glia have different morphology and can give rise to different cell types during brain development. In particular, apical radial glia are found in the ventricular zone of the developing neocortex and are connected to both basal lamina and ventricular surface with their plasma membrane. Upon asymmetric cell divisions in human, apical radial glia can give rise to basal radial glia, various types of intermediate progenitors and neurons. Basal radial glia on their turn are found in the SVZ, do not have their plasma membrane attached to the ventricular surface and can differentiate, in human, into intermediate progenitors and neurons. Primates, humans in particular, have an exceptionally developed SVZ, thanks to abundant basal radial glia, which give rise to a larger pool of neurons (Florio and Huttner 2014; Namba and Huttner 2017).

Florio et al. (2015) then studied human genes that have equal or higher expression in basal radial glia relative to apical radial glia. Levels of expression of studied genes in both types of radial glia had to be higher in radial glia, compared with neighboring neurons. In the course of their study, Florio et al. (2015) discovered nearly 400 human genes that showed roughly equal levels of expression in apical and basal radial glia. Furthermore, 56 genes with this pattern of expression in human were absent in the mouse genome. These genes were mostly implicated in DNA repair and telomere maintenance, which indicates that both apical and basal radial glia in humans have stem cell properties, engage mostly in cell proliferation and not in cell differentiation into neurons.

As Florio et al. (2015) showed, *ARHGAP11B* was the only gene, among the 56 genes, with a 10 times higher expression level in both types of human radial glia, relative to neurons. *ARHGAP11B*, coding for a 267 aa-long protein, appeared from a partial duplication of much longer *ARHGAP11A*, also after the split of human from chimpanzee and it is also present in the Neanderthal and the Denisovan (Sudmant et al. 2010; Antonacci et al. 2014; Dennis et al. 2017). This time the paralogs share a part of the RhoGAP domain. The C-terminal portion of the domain, after Lys220, seems to be important for the RhoGAP activity, because neither is present in ARHGAP11B that possesses instead a unique stretch of 47 aa at its C-terminus (Florio et al. 2015). Contrary to the previously described paralogs, ARHGAP11B does not inhibit the function of ARHGAP11A. The novel paralog seems to play a role in the human-specific expansion of the SVZ, where basal radial glia of the developing neocortex reside. Transient transfection, with *in utero* electroporation, of ARHGAP11B-containing construct in mice, led to increased basal progenitor proliferation and self-renewal, as well as an enlarged cortical plate and development of cortical gyri (Florio et al. 2015). The precise molecular function of the truncated and modified RhoGAP domain is still unclear.

The genetic variant that led to the new functional paralog ARHGAP11B was not only the partial duplication of the *ARHGAP11A* gene, that is, of the first eight exons and introns but also the substitution NM_014783.5: c.661 C > G that was introduced in exon 5 of *ARHGAP11A* during human evolution (Florio et al. 2016). As Florio et al. (2016) showed, this variant creates a new and early GU-purine splice site "gta" that leads to removal of a portion of exon 5. This truncation of exon 5 leads to a frameshift in codons and to the new 47 aa-long stretch of the aa sequence of ARHGAP11B that ends with an early stop codon. Only the substitution was shown to be important for inducing proliferation of basal progenitors, that is, basal radial glia and basal intermediate progenitors, in the developing cerebral neocortex. In particular, mice subjected to *in utero* electroporation that induced expression of *ARHGAP11B* in the neocortex during its development had a 2-fold increase in mitotic basal progenitors (Florio et al. 2016). These results recapitulated those in the previous study (Florio et al. 2015). The substitution occurred after the duplication and the version of the gene without the substitution was considered as an ancestral version. Florio et al. (2016) introduced in mice a genetic sequence identical to *ARHGAP11B* except for the substitution C > G, that is, characterized only by the duplication. As Florio et al. (2016) found out, this ancestral version of the gene does not boost basal progenitors' proliferation, which indicates that the substitution C > G alone is responsible for the human-specific proliferation of basal progenitors in the SVZ. Neanderthals and Denisovans have the modern-human version of *ARHGAP11B* (Florio et al. 2016).

Also of interest is the fact that *ARHGAP11A/B* pair is found on 15q13.3, one of the most variable and unstable regions in the human genome, where palindromes of duplicated genes *GOLGA8* fuel further genomic rearrangements (Antonacci et al. 2014). In particular, that region was found to have a ~2-Mb deletion associated with a number of neurodevelopmental disorders: epilepsy, intellectual disability, autism, and SCZ (reviewed in Antonacci et al. 2014).

hCONDELs and *GADD45G*

A further example is a study, published in 2011, in which the authors analyzed the human genome for the presence of deletions in regions otherwise conserved in vertebrates, including chimpanzees (McLean et al. 2011). As a first step,

the authors identified regions of at least 23 bp, conserved between chimpanzee and macaque, but deleted in humans, as determined by University of California Santa Cruz (UCSC) chains and nets (Kent et al. 2003). In a second step, McLean et al. 2011 superimposed these regions with the ones conserved among chimpanzee, macaque, and chicken, using UCSC chains and nets, BLASTZ (Schwartz et al. 2003), and MultiZ. The resulting 583 regions, 6,126 bp long on an average, were termed hCONDELs, 510 of which were later confirmed by McLean et al. (2011) by direct alignment with a human sequence, using BLAT (Kent 2002). Interestingly, 88% of these human deletions were also discovered in the Neanderthal, indicating that the majority of hCONDELs occurred before human-Neanderthal divergence (McLean et al. 2011). In addition, the conserved regions lost in humans are mostly AT-rich in the chimpanzee genome, which may be related to the fact that HARs also tend to present a reduced AT content: both human-specific genetic variant types appear to result in a higher GC-content in the human genome.

Almost all, except one hCONDEL, reside in noncoding regions (McLean et al. 2011). However, it is possible to assess nearby genes that could be otherwise regulated by the sequences lost in humans: by using the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al. 2010), McLean et al. (2011) discovered that the majority of nearby genes are implicated in steroid hormone receptor signaling and brain function. One of those deletions, 3,181 bp long, removed an enhancer of the growth arrest and DNA-damage inducible gamma (*GADD45G*) tumor suppressor gene (McLean et al. 2011). The enhancer, that binds p300, is specific for the forebrain subventricular zone (SVZ) and other parts of the telencephalon and diencephalon in mice (McLean et al. 2011). In this experiment, chimpanzee and mouse enhancers, that were used in a construct containing LacZ, drove the expression of this reporter gene in the ventral telencephalon and diencephalon. Similar results were obtained in human immortalized neural progenitor cells (McLean et al. 2011). This LacZ expression matches the expression pattern of *GADD45G* (Gohlke et al. 2008). The SVZ is particularly interesting in this context, because it contains, during embryonic development, neural progenitors that give rise to neurons populating the neocortex; proliferation of these progenitor cells is thought to drive the neocortex expansion in primates (Kriegstein et al. 2006; Abdel-Mannan et al. 2008; Fish et al. 2008; Rakic 2009; Lui et al. 2011; Florio and Huttner 2014; Taverna et al. 2014; Wilsch-Brauninger et al. 2016; Namba and Huttner 2017). GADD45G normally represses cell proliferation and induces apoptosis, whereas somatic loss of expression was observed in cancer (Zhang et al. 2002; Zerbini et al. 2004). Therefore, the discovered deletion could result in the more dramatic expansion of the neocortex in humans, compared with other primates (McLean et al. 2011).

Further Examples of Human-Specific Genetic Variations That Seem to Be Important in Human Brain Evolution

Rockman et al. described in 2005 evolutionarily new sequences in the promoter of the human prodynorphin gene (*PDYN*), a precursor of a number of endogenous opioid neuropeptides, endorphins, implicated in perception, social behavior, and learning (Rockman et al. 2005). Five fixed single nucleotide substitutions and one site with variable insertions of one to three nucleotides are scattered inside a 68-bp-long sequence, otherwise conserved in seven nonhuman primates. In addition, this regulatory region is duplicated to form one to four tandem repeats only in humans. The sequence is 1,250 bp away from the transcription start site. Rockman et al. (2005) showed that the substitution rate in this human promoter is not neutral and is accelerated, as was determined by Poisson probability, at P values $<10^{-4}$ and $<5\times10^{-3}$, respectively. As was demonstrated in the same study that used the human neural cell line SH-SY5Y, the human promoter significantly increases expression of luciferase in cells transfected with a chimpanzee construct, in comparison with constructs with the chimpanzee promoter. This indicates that prodynorphin acquired higher rates of expression in Homo sapiens, compared with other primates.

The gene family neuroblastoma breakpoint family (*NBPF*) is characterized by domain of unknown function (DUF1220) protein domains that underwent the largest, >300, copy number expansion in humans (Vandepoele et al. 2005; Dumas and Sikela 2009; Sudmant et al. 2010, 2013; O'Bleness et al. 2012a, 2012b; Keeney et al. 2014; Zimmer and Montgomery 2015; Astling et al. 2017). This human-specific tandem repeat expansion was first described in 2006 (Popesco et al. 2006). There are numerous studies showing statistical association between neurodevelopmental disorders—microcephaly, macrocephaly, autism, and SCZ—and alterations in the number of DUF1220 domains in *NBPF* genes (Dumas et al. 2012; Davis et al. 2014, 2015; Searles Quick et al. 2015). In particular, the number of domain repeats statistically correlates with brain size in human patients and in different primate species, underlying relevance of the gene family in human health and evolution (Dumas et al. 2012; Keeney et al. 2014, 2015). Expression and functional studies of this gene family were also undertaken: Popesco et al. (2006) and Keeney et al. (2015) showed that DUF1220 domains are expressed in various parts of the embryonic and adult human brain (Popesco et al. 2006; Keeney et al. 2015). In addition, Keeney et al. (2015) showed that human *NBPF15*, transfected into H9-derived human neural stem cell lines, induces more dynamic cell proliferation. Knowing that the DUF1220 domain expansion is also very significant in apes in general, the drawback of the study by Keeney et al. (2015) was that the authors did not investigate whether other ape-specific repeats act differently on the neural stem cell lines.

Conclusions and Future Directions

The rationale for studying the human-specific single nucleotide substitutions, deletions, and segmental duplications is to define the genetic basis of human-specific morphology, of which the most remarkable is the human brain, the cerebral cortex in particular. The list of discovered variants is incomplete: genome-wide gene conversions and inversions, although assessed for several genes in Dennis et al. (2017), as well as small indels (Mullaney et al. 2010) must play some role in evolution of the human brain too. The major difficulty in the study of large, complex genomic rearrangements is that with short reads, like the ones generated by the Illumina's whole genome sequencing, researchers can miss rearrangements larger than the reads themselves. An answer to this difficulty is application of a technology, allowing much longer reads, such as Single Molecule Real Time Sequencing (SMRT) from Pacific Biosciences that allows reads 10 kb on an average and up to 60 kb (Huddleston et al. 2014) (Dennis et al. 2017 partially used this technology). Despite the technical challenges, it becomes more obvious that human-specific large genomic rearrangements constitute an undeniably important mechanism that drove human evolution.

Another aspect that seems interesting is the actual extent to which HARs, hCONDELs, and HSDs are shared with our closest relatives Neanderthals and Denisovans. The studies showed that 7.1–8.3% of HARs are not shared with Neanderthals or Denisovans (Burbano et al. 2012; Hubisz and Pollard 2014), 12% of hCONDELs are not shared with Neanderthals (McLean et al. 2011) and three gene-containing regions out of the 28 such HSD regions—that is, 10.7%—are not shared with Neanderthals or Denisovans (Dennis et al. 2017). It is worth to note that results of McLean et al. (2011) and Burbano et al. (2012) should be interpreted with caution, because the archaic genomes used at that time were of low quality: each nucleotide position was determined, on an average, once. Only in October 2012 the Denisovan genome at 31× coverage became available (Meyer et al. 2012). Later, in January 2014, the Neanderthal genome at 52× coverage was published (Prufer et al. 2013). The quality of both archaic genomes is currently similar to that of present-day human genomes. Also of note that, unfortunately, Dennis et al. (2017) did not assess human versus Neanderthal/Denisovan differences in the remaining 190 HSDs found in noncoding regions.

Although the presented percentage of genome shared between modern and archaic humans corresponds well to the time scale of divergence of humans from chimpanzees, 8 Ma (Moorjani et al. 2016) to 6 Ma (Patterson et al. 2006), and of humans from Neanderthals and Denisovans, 800,000 (Pennisi 2007) to 500,000 (Prufer et al. 2013) years ago, the actual picture could be different given that the Neanderthal genome is actually a part of the human genome (Vernot and Akey 2014). This Neanderthal DNA is spread all over the human genome, sparing only the Y-chromosome (Mendez et al. 2016) and the mitochondrial DNA (Green et al. 2008). The Neanderthal DNA accounts for 1–3% of the genome in each tested non-African individual (Prufer et al. 2013; Vernot and Akey 2014), but these regions are not the same in different individuals; the variation in different human genomes is such that the total amount of the Neanderthal DNA in humans accounts for as much as 20% of the Neanderthal genome. To draw these estimates, Vernot and Akey (2014) used whole genome sequences of 379 Europeans and 286 East Asians from the 1000 Genomes project (1000 Genomes Project Consortium 2012). However, authors that described differences in the sequence of HARs and hCONDELs in humans versus Neanderthals/Denisovans (McLean et al. 2011; Burbano et al. 2012) used data generated with the NCBI genome assemblies 36.1 and 37, in which 71% of sequence was contributed by a single individual of approximately half African–half European ancestry and another 22% of the assembly came from only 10 libraries, which are proxies for individual genomes, mainly of European and East Asian ancestry (Schneider et al. 2017). This sample is not truly representative of the wealth of human genetic variation, so DNA from much larger numbers of human subjects must be used to determine whether particular populations have some of the HARs and hCONDELs, or have the ancestral allele, shared with Neanderthals/Denisovans. On the opposite, determining gene-containing HSDs in humans, that are absent in Neanderthals/Denisovans, was done using DNA of 236 individuals from 125 populations scattered all over the globe (Sudmant et al. 2015; Dennis et al. 2017), which is more representative of the true picture. In fact, Dennis et al. (2017) spotted 17 individuals that had the ancestral, that is, observed in Neanderthals/Denisovans, allele in the transient receptor potential cation channel subfamily M member 8-associated factor (*TCAF*) locus.

It is also of note that Bird et al. (2007), who initially studied HARs, discovered that these noncoding-accelerated sequences are statistically more often found within recent segmental duplications. The same could be noted about *PDYN*, the gene harboring human-specific nucleotide substitutions and segmental tandem duplications all at the same time. These data suggest that the different types of human-specific, functional variation could affect the same genomic regions and act in concert.

Another major difficulty in the study of human-specific genomic variations is proving that a variant has indeed any role to play on the functional level. As studies of enhancers associated with HARs or hCONDELs demonstrated, working with each separate variant is a time- and effort-consuming process. Given the numbers of discovered HARs and hCONDELs, in the order of 3500, applying classical approaches using transgenic animal models, described in this review, seems like a long and hard road to go. New,

quicker, and affordable methods of massive screening, such as Massively parallel reporter assays (MPRA), that allow screening several regions at a time, can solve this problem (Hubisz and Pollard 2014; Dailey 2015; Inoue and Ahituv 2015; White 2015). This method in fact proved its usefulness in one of the studies mentioned here (Doan et al. 2016), where it contributed to determining whether autism-associated sequence variations alter enhancer activity. New improvements to these techniques are currently available, like Systematic high-resolution activation and repression profiling with reporter tiling using MPRA (Sharpr-MPRA) that allows analysis of thousands of regions simultaneously (Ernst et al. 2016). Another new development of these methods enables to test the different enhancers in various brain regions of a living animal and can be potentially used in human cerebral organoids (Lancaster et al. 2013; Shen et al. 2016). Obviously, these new methods allow studying silencers and insulators as well, given that ~60% of HARs could belong to that group of regulatory sequences. As to HSDs and the paralogs created by them, it is hard to imagine a high throughput method that would account for all possible outcomes of the new genes: whether they have reduced expression, increased expression, change location of expression, acquire new function, lose previous function, or inhibit other paralogs. In this case, only individual functional studies of each paralog, using either animal models or human cerebral organoids seem appropriate.

Understanding the biology of human-specific genomic variations will not only contribute to the study of evolution of our species but will also give answers in regard to etiology of human-specific diseases, of which neurodevelopmental disorders stand as the most representative example.

Acknowledgment

This work was supported by the Russian Science Foundation (grant number 14-50-00069 to R.R.G.).

Literature Cited

1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1, 092 human genomes. Nature 491:56–65.
Abdel-Mannan O, Cheung AF, Molnar Z. 2008. Evolution of cortical neurogenesis. Brain Res Bull. 75(2–4):398–404.
Allen NJ. 2012. Astrocyte glypicans 4 and 6 promote formation of excitatory synapses via GluA1 AMPA receptors. Nature 486(7403):410–414.
Antonacci F, et al. 2014. Palindromic *GOLGA8* core duplicons promote chromosome 15q13.3 microdeletion and evolutionary instability. Nat Genet. 46(12):1293–1302.
Asahara H, et al. 2002. Dual roles of p300 in chromatin assembly and transcriptional activation in cooperation with nucleosome assembly protein 1 in vitro. Mol Cell Biol. 22(9):2974–2983.
Ashburner M, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 25(1):25–29.
Astling DP, Heft IE, Jones KL, Sikela JM. 2017. High resolution measurement of DUF1220 domain copy number from whole genome sequence data. BMC Genomics 18(1):614.
Bae BI, Jayaraman D, Walsh CA. 2015. Genetic changes shaping the human brain. Dev Cell 32(4):423–434.
Bailey JA, Liu G, Eichler EE. 2003. An Alu transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet. 73(4):823–834.
Barrett T, et al. 1997. The structure of the GTPase-activating domain from p50rhoGAP. Nature 385(6615):458–461.
Bedogni F, et al. 2010a. Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. Proc Natl Acad Sci U S A. 107(29):13129–13134.
Bedogni F, et al. 2010b. Autism susceptibility candidate 2 (*Auts2*) encodes a nuclear protein expressed in developing brain regions implicated in autism neuropathology. Gene Expr Patterns 10(1):9–15.
Benavides-Piccione R, Ballesteros-Yanez I, DeFelipe J, Yuste R. 2002. Cortical area and species differences in dendritic spine morphology. J Neurocytol. 31(3–5):337–346.
Beniaminov A, Westhof E, Krol A. 2008. Distinctive structures between chimpanzee and human in a brain noncoding RNA. RNA 14(7):1270–1275.
Bernstein BE, et al. 2010. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 28(10):1045–1048.
Bird CP, et al. 2007. Fast-evolving noncoding sequences in the human genome. Genome Biol. 8(6):R118.
Blanchette M, et al. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 14(4):708–715.
Boyd JL, et al. 2015. Human-chimpanzee differences in a *FZD8* enhancer alter cell-cycle dynamics in the developing neocortex. Curr Biol. 25(6):772–779.
Brunskill EW, et al. 2005. Abnormal neurodevelopment, neurosignaling and behaviour in Npas3-deficient mice. Eur J Neurosci. 22(6):1265–1276.
Buchman JJ, Durak O, Tsai LH. 2011. ASPM regulates Wnt signaling pathway activity in the developing brain. Genes Dev. 25(18):1909–1914.
Burbano HA, et al. 2012. Analysis of human accelerated DNA regions using archaic hominin genomes. PLoS One 7(3):e32877.
Bush EC, Lahn BT. 2008. A genome-wide screen for noncoding elements important in primate evolution. BMC Evol Biol. 8:17.
Capra JA, Erwin GD, McKinsey G, Rubenstein JL, Pollard KS. 2013. Many human accelerated regions are developmental enhancers. Philos Tran R Soc Lond B Biol Sci. 368(1632):20130025.
Carter MT, et al. 2011. Hemizygous deletions on chromosome 1p21.3 involving the *DPYD* gene in individuals with autism spectrum disorder. Clin Genet. 80(5):435–443.
Charrier C, et al. 2012. Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. Cell 149(4):923–935.
Chen LL. 2016. Linking long noncoding RNA localization and function. Trends Biochem Sci. 41(9):761–772.
Cheng YZ, et al. 2012. Investigating embryonic expression patterns and evolution of *AHI1* and *CEP290* genes, implicated in Joubert syndrome. PLoS One 7(9):e44975.
Clevers H. 2006. Wnt/beta-catenin signaling in development and disease. Cell 127(3):469–480.
Cline MS, et al. 2007. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2(10):2366–2382.
Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33:2938–2940.
Creyghton MP, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 107(50):21931–21936.

Cubelos B, Briz CG, Esteban-Ortega GM, Nieto M. 2015. Cux1 and Cux2 selectively target basal and apical dendritic compartments of layer II-III cortical neurons. Dev Neurobiol. 75(2):163–172.
Dailey L. 2015. High throughput technologies for the functional discovery of mammalian enhancers: new approaches for understanding transcriptional regulatory network dynamics. Genomics 106(3):151–158.
Davis JM, et al. 2014. DUF1220 dosage is linearly associated with increasing severity of the three primary symptoms of autism. PLoS Genet. 10(3):e1004241.
Davis JM, Searles Quick VB, Sikela JM. 2015. Replicated linear association between DUF1220 copy number and severity of social impairment in autism. Hum Genet. 134(6):569–575.
Dennis MY, Eichler EE. 2016. Human adaptation and evolution by segmental duplication. Curr Opin Genet Dev. 41:44–52.
Dennis MY, et al. 2012. Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. Cell 149(4):912–922.
Dennis MY, et al. 2017. The evolution and population diversity of human-specific segmental duplications. Nat Ecol Evol. 1(3):0069.
Dermitzakis ET, Clark AG. 2002. Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Mol Biol Evol. 19(7):1114–1121.
Dixon JR, et al. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485(7398):376–380.
Doan RN, et al. 2016. Mutations in human accelerated regions disrupt cognition and social behavior. Cell 167(2):341–354 e312.
Dorschner MO, et al. 2004. High-throughput localization of functional elements by quantitative chromatin profiling. Nat Methods 1(3):219–225.
Dumas L, et al. 2007. Gene copy number variation spanning 60 million years of human and primate evolution. Genome Res. 17(9):1266–1277.
Dumas L, Sikela JM. 2009. DUF1220 domains, cognitive disease, and human brain evolution. Cold Spring Harb Symp Quant Biol. 74:375–382.
Dumas LJ, et al. 2012. DUF1220-domain copy number implicated in human brain-size pathology and evolution. Am J Hum Genet. 91(3):444–454.
Eichler EE. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. Trends Genet. 17(11):661–669.
Elston GN, Benavides-Piccione R, DeFelipe J. 2001. The pyramidal cell in cognition: a comparative study in human and monkey. J Neurosci. 21(17):RC163.
Enard W, et al. 2002. Molecular evolution of *FOXP2*, a gene involved in speech and language. Nature 418(6900):869–872.
Enard W, et al. 2009. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell 137(5):961–971.
Engreitz JM, Ollikainen N, Guttman M. 2016. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat Rev Mol Cell Biol. 17(12):756–770.
Erbel-Sieler C, et al. 2004. Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. Proc Natl Acad Sci U S A. 101(37):13648–13653.
Ernst J, et al. 2016. Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. Nat Biotechnol. 34(11):1180–1190.
Erwin GD, et al. 2014. Integrating diverse datasets improves developmental enhancer prediction. PLoS Comput Biol. 10(6):e1003677.
Evans PD, et al. 2004. Adaptive evolution of *ASPM*, a major determinant of cerebral cortical size in humans. Hum Mol Genet. 13(5):489–494.
Favorov A, et al. 2012. Exploring massive, genome scale datasets with the GenometriCorr package. PLoS Comput Biol. 8(5):e1002529.
Ferland RJ, et al. 2004. Abnormal cerebellar development and axonal decussation due to mutations in AHI1 in Joubert syndrome. Nat Genet. 36(9):1008–1013.
Ferri AL, et al. 2004. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. Development 131(15):3805–3819.
Fischer T, Guimera J, Wurst W, Prakash N. 2007. Distinct but redundant expression of the Frizzled Wnt receptor genes at signaling centers of the developing mouse brain. Neuroscience 147(3):693–711.
Fish JL, Dehay C, Kennedy H, Huttner WB. 2008. Making bigger brains-the evolution of neural-progenitor-cell division. J Cell Sci. 121(Pt 17):2783–2793.
Flicek P, et al. 2014. Ensembl 2014. Nucleic Acids Res. 42(Database issue):D749–D755.
Florio M, et al. 2015. Human-specific gene *ARHGAP11B* promotes basal progenitor amplification and neocortex expansion. Science 347(6229):1465–1470.
Florio M, Huttner WB. 2014. Neural progenitors, neurogenesis and the evolution of the neocortex. Development 141(11):2182–2194.
Florio M, Namba T, Pääbo S, Hiller M, Huttner WB. 2016. A single splice site mutation in human-specific *ARHGAP11B* causes basal progenitor amplification. Sci Adv. 2(12):e1601941.
Fortna A, et al. 2004. Lineage-specific gene duplication and loss in human and great ape evolution. PLoS Biol. 2(7):E207.
Franchini LF, Pollard KS. 2015. Can a few non-coding mutations make a human brain? Bioessays 37(10):1054–1061.
Freese JL, Pino D, Pleasure SJ. 2010. Wnt signaling in development and disease. Neurobiol Dis. 38(2):148–153.
Gittelman RM, et al. 2015. Comprehensive identification and analysis of human accelerated regulatory DNA. Genome Res. 25(9):1245–1255.
Goecks J, Nekrutenko A, Taylor J, Galaxy T. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 11(8):R86.
Gohlke JM, et al. 2008. Characterization of the proneural gene regulatory network during mouse telencephalon development. BMC Biol. 6:15.
Gould P, Kamnasaran D. 2011. Immunohistochemical analyses of *NPAS3* expression in the developing human fetal brain. Anat Histol Embryol. 40(3):196–203.
Green RE, et al. 2008. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell 134(3):416–426.
Green RE, et al. 2010. A draft sequence of the Neandertal genome. Science 328(5979):710–722.
Griffiths-Jones S, et al. 2005. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 33(Database issue):D121–D124.
Guerrier S, et al. 2009. The F-BAR domain of srGAP2 induces membrane protrusions required for neuronal migration and morphogenesis. Cell 138(5):990–1004.
Guo S, Bao S. 2010. srGAP2 arginine methylation regulates cell migration and cell spreading through promoting dimerization. J Biol Chem. 285(45):35133–35141.
Hagege H, et al. 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc. 2(7):1722–1733.
Harrow J, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 22(9):1760–1774.
Hart RP, Goff LA. 2016. Long noncoding RNAs: central to nervous system development. Int J Dev Neurosci. 55:109–116.
Haygood R, Babbitt CC, Fedrigo O, Wray GA. 2010. Contrasts between adaptive coding and noncoding changes during human evolution. Proc Natl Acad Sci U S A. 107(17):7853–7857.
Haygood R, Fedrigo O, Hanson B, Yokoyama KD, Wray GA. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. Nat Genet. 39(9):1140–1144.
Hubisz MJ, Pollard KS. 2014. Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. Curr Opin Genet Dev. 29:15–21.

Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. Brief Bioinform. 12(1):41–51.
Huddleston J, et al. 2014. Reconstructing complex regions of genomes using long-read sequencing technology. Genome Res. 24(4):688–696.
Inoue F, Ahituv N. 2015. Decoding enhancers using massively parallel reporter assays. Genomics 106(3):159–164.
International HapMap Consortium, et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861.
Jayaraman D, et al. 2016. Microcephaly proteins Wdr62 and Aspm define a mother centriole complex regulating centriole biogenesis, apical complex, and cell fate. Neuron 92(4):813–828.
Jin F, et al. 2013. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503(7475):290–294.
Kalscheuer VM, et al. 2007. Mutations in autism susceptibility candidate 2 (AUTS2) in patients with mental retardation. Hum Genet. 121(3–4):501–509.
Kamm GB, Lopez-Leal R, Lorenzo JR, Franchini LF. 2013a. A fast-evolving human NPAS3 enhancer gained reporter expression in the developing forebrain of transgenic mice. Philos Trans R Soc Lond B Biol Sci. 368(1632):20130019.
Kamm GB, Pisciottano F, Kliger R, Franchini LF. 2013b. The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. Mol Biol Evol. 30(5):1088–1102.
Kamnasaran D, Muir WJ, Ferguson-Smith MA, Cox DW. 2003. Disruption of the neuronal PAS3 gene in a family affected with schizophrenia. J Med Genet. 40(5):325–332.
Keeney JG, Dumas L, Sikela JM. 2014. The case for DUF1220 domain dosage as a primary contributor to anthropoid brain expansion. Front Hum Neurosci. 8:427.
Keeney JG, et al. 2015. DUF1220 protein domains drive proliferation in human neural stem cells and are associated with increased cortical volume in anthropoid primates. Brain Struct Funct. 220(5):3053–3060.
Kelberman D, et al. 2008. *SOX2* plays a critical role in the pituitary, forebrain, and eye during human embryonic development. J Clin Endocrinol Metab. 93(5):1865–1873.
Kent WJ. 2002. BLAT–the BLAST-like alignment tool. Genome Res. 12(4):656–664.
Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc Natl Acad Sci U S A. 100(20):11484–11489.
Kim TK, Shiekhattar R. 2016. Diverse regulatory interactions of long noncoding RNAs. Curr Opin Genet Dev. 36:73–82.
King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science 188(4184):107–116.
Konopka G, et al. 2009. Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature 462(7270):213–217.
Kostka D, Hubisz MJ, Siepel A, Pollard KS. 2012. The role of GC-biased gene conversion in shaping the fastest evolving regions of the human genome. Mol Biol Evol. 29(3):1047–1057.
Kriegstein A, Noctor S, Martinez-Cerdeno V. 2006. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. Nat Rev Neurosci. 7(11):883–890.
Lancaster MA, et al. 2013. Cerebral organoids model human brain development and microcephaly. Nature 501(7467):373–379.
Lepagnol-Bestel AM, et al. 2008. *SLC25A12* expression is associated with neurite outgrowth and is upregulated in the prefrontal cortex of autistic subjects. Mol Psychiatry 13(4):385–397.
Levchenko A, et al. 2015. Beta-catenin in schizophrenia: possibly deleterious novel mutation. Psychiatry Res. 228(3):843–848.
Li G, et al. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148(1–2):84–98.
Li Q, et al. 2014. The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation. Elife 3:e01201.
Licatalosi DD, et al. 2012. Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. Genes Dev. 26(14):1626–1642.
Lieberman-Aiden E, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293.
Lindblad-Toh K, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478(7370):476–482.
Lipovich L, et al. 2014. Developmental changes in the transcriptome of human cerebral cortex tissue: long noncoding RNA transcripts. Cereb Cortex 24(6):1451–1459.
Locke DP, et al. 2003. Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. Genome Res. 13(3):347–357.
Lui JH, Hansen DV, Kriegstein AR. 2011. Development and evolution of the human neocortex. Cell 146(1):18–36.
Mao Y, et al. 2009. Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3beta/beta-catenin signaling. Cell 136(6):1017–1031.
Matelot M, Noordermeer D. 2016. Determination of high-resolution 3D chromatin organization using circular chromosome conformation capture (4C-seq). Methods Mol Biol. 1480:223–241.
Maurano MT, et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. Science 337(6099):1190–1195.
Mayer BJ. 2001. SH3 domains: complexity in moderation. J Cell Sci. 114(Pt 7):1253–1263.
McLean CY, et al. 2010. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 28(5):495–501.
McLean CY, et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature 471(7337):216–219.
Mendez FL, Poznik GD, Castellano S, Bustamante CD. 2016. The divergence of neandertal and modern human Y chromosomes. Am J Hum Genet. 98(4):728–734.
Meyer M, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science 338(6104):222–226.
Miller JA, et al. 2014. Transcriptional landscape of the prenatal human brain. Nature 508(7495):199–206.
Moorjani P, Amorim CE, Arndt PF, Przeworski M. 2016. Variation in the molecular clock of primates. Proc Natl Acad Sci U S A. 113(38):10607–10612.
Mugal CF, Weber CC, Ellegren H. 2015. GC-biased gene conversion links the recombination landscape and demography to genomic base composition: GC-biased gene conversion drives genomic base composition across a wide range of species. Bioessays 37(12):1317–1326.
Mullaney JM, Mills RE, Pittard WS, Devine SE. 2010. Small insertions and deletions (INDELs) in human genomes. Hum Mol Genet. 19(R2):R131–R136.
Namba T, Huttner WB. 2017. Neural progenitor cells and their role in the development and evolutionary expansion of the neocortex. Wiley Interdiscip Rev Dev Biol. 6(1):e256.
Ng SY, Lin L, Soh BS, Stanton LW. 2013. Long noncoding RNAs in development and disease of the central nervous system. Trends Genet. 29(8):461–468.
Noelanders R, Vleminckx K. 2017. How Wnt signaling builds the brain: bridging development and disease. Neuroscientist 23(3):314–329.
O'Bleness M, Searles VB, Varki A, Gagneux P, Sikela JM. 2012a. Evolution of genetic and genomic features unique to the human lineage. Nat Rev Genet. 13:853–866.
O'Bleness MS, et al. 2012b. Evolutionary history and genome organization of DUF1220 protein domains. G3 (Bethesda) 2:977–986.
Ogawa LM, Vallender EJ. 2014. Evolutionary conservation in genes underlying human psychiatric disorders. Front Hum Neurosci. 8:283.

Oksenberg N, Ahituv N. 2013. The role of *AUTS2* in neurodevelopment and human evolution. Trends Genet. 29(10):600–608.
Oksenberg N, Stevison L, Wall JD, Ahituv N. 2013. Function and regulation of *AUTS2*, a gene implicated in autism and human evolution. PLoS Genet. 9(1):e1003221.
O'Leary NA, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44:D733–D745.
Owen MJ, Sawa A, Mortensen PB. 2016. Schizophrenia. Lancet 388(10039):86–97.
Paten B, Herrero J, Beal K, Fitzgerald S, Birney E. 2008. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. Genome Res. 18(11):1814–1828.
Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. Nature 441(7097):1103–1108.
Pedersen JS, et al. 2006. Identification and classification of conserved RNA secondary structures in the human genome. PLoS Comput Biol. 2(4):e33.
Pennisi E. 2007. Ancient DNA. No sex please, we're Neandertals. Science 316(5827):967.
Pickard BS, et al. 2009. Interacting haplotypes at the *NPAS3* locus alter risk of schizophrenia and bipolar disorder. Mol Psychiatry 14(9):874–884.
Pickard BS, Malloy MP, Porteous DJ, Blackwood DH, Muir WJ. 2005. Disruption of a brain transcription factor, NPAS3, is associated with schizophrenia and learning disability. Am J Med Genet B Neuropsychiatr Genet. 136B(1):26–32.
Pickard BS, Pieper AA, Porteous DJ, Blackwood DH, Muir WJ. 2006. The NPAS3 gene–emerging evidence for a role in psychiatric illness. Ann Med. 38(6):439–448.
Pink RC, et al. 2011. Pseudogenes: pseudo-functional or key regulators in health and disease? RNA 17(5):792–798.
Pollard KS, et al. 2006a. Forces shaping the fastest evolving regions in the human genome. PLoS Genet. 2(10):e168.
Pollard KS, et al. 2006b. An RNA gene expressed during cortical development evolved rapidly in humans. Nature 443(7108):167–172.
Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of non-neutral substitution rates on mammalian phylogenies. Genome Res. 20(1):110–121.
Popesco MC, et al. 2006. Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. Science 313(5791):1304–1307.
Prabhakar S, et al. 2008. Human-specific gain of function in a developmental enhancer. Science 321(5894):1346–1350.
Prabhakar S, Noonan JP, Paabo S, Rubin EM. 2006. Accelerated evolution of conserved noncoding sequences in humans. Science 314(5800):786.
Prufer K, et al. 2013. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505(7481):43–49.
Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35(Database issue):D61–D65.
Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81(3):559–575.
Rakic P. 2009. Evolution of the neocortex: a perspective from developmental biology. Nat Rev Neurosci. 10(10):724–735.
Reich D, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468(7327):1053–1060.
Riley B, Williamson M, Collier D, Wilkie H, Makoff A. 2002. A 3-Mb map of a large Segmental duplication overlapping the alpha7-nicotinic acetylcholine receptor gene (*CHRNA7*) at human 15q13-q14. Genomics 79(2):197–209.
Rockman MV, et al. 2005. Ancient and recent positive selection transformed opioid cis-regulation in humans. PLoS Biol. 3(12):e387.
Roussos P, Katsel P, Davis KL, Siever LJ, Haroutunian V. 2012. A system-level transcriptomic analysis of schizophrenia using postmortem brain tissue samples. Archiv Gen Psychiatry 69(12):1205–1213.
Salichs E, Ledda A, Mularoni L, Albà MM, de la Luna S. 2009. Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. PLoS Genet. 5(3):e1000397.
Sassa T. 2013. The role of human-specific gene duplications during brain development and evolution. J Neurogenet. 27(3):86–96.
Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427.
Schneider VA, et al. 2017. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res. 27(5):849–864.
Schreiweis C, et al. 2014. Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. Proc Natl Acad Sci U S A. 111(39):14253–14258.
Schwartz S, et al. 2003. Human-mouse alignments with BLASTZ. Genome Res. 13(1):103–107.
Searles Quick VB, Davis JM, Olincy A, Sikela JM. 2015. DUF1220 copy number is associated with schizophrenia risk and severity: implications for understanding autism and schizophrenia as related diseases. Transl Psychiatry 5:e697.
Sha L, et al. 2012. Transcriptional regulation of neurodevelopmental and metabolic pathways by NPAS3. Mol Psychiatry 17(3):267–279.
Shen SQ, et al. 2016. Massively parallel cis-regulatory analysis in the mammalian central nervous system. Genome Res. 26(2):238–255.
Shi P, Bakewell MA, Zhang J. 2006. Did brain-specific genes evolve faster in humans than in chimpanzees? Trends Genet. 22(11):608–613.
Siepel A, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15(8):1034–1050.
Silver DL. 2016. Genomic divergence and brain evolution: how regulatory DNA influences development of the cerebral cortex. Bioessays 38(2):162–171.
Singh KK. 2013. An emerging role for Wnt and GSK3 signaling pathways in schizophrenia. Clin Genet. 83(6):511–517.
Snyder SR, Wang J, Waring JF, Ginder GD. 2001. Identification of CCAAT displacement protein (CDP/cut) as a locus-specific repressor of major histocompatibility complex gene expression in human tumor cells. J Biol Chem. 276(7):5323–5330.
Spector DL, Lamond AI. 2011. Nuclear speckles. Cold Spring Harb Perspect Biol. 3(2):a000646.
Spiteri E, et al. 2007. Identification of the transcriptional targets of *FOXP2*, a gene linked to speech and language, in developing human brain. Am J Hum Genet. 81(6):1144–1157.
Sudmant PH, et al. 2010. Diversity of human copy number variation and multicopy genes. Science 330(6004):641–646.
Sudmant PH, et al. 2013. Evolution and diversity of copy number variation in the great ape lineage. Genome Res. 23(9):1373–1382.
Sudmant PH, et al. 2015. Global diversity, population stratification, and selection of human copy-number variation. Science 349(6253):aab3761.
Sultana R, et al. 2002. Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. Genomics 80(2):129–134.
Sumiyama K, Saitou N. 2011. Loss-of-function mutation in a repressor module of human-specifically activated enhancer HACNS1. Mol Biol Evol. 28(11):3005–3007.
Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135(2):599–607.
Taverna E, Gotz M, Huttner WB. 2014. The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. Annu Rev Cell Dev Biol. 30:465–502.

The Chimpanzee Sequencing and Analysis Consortium 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437(7055):69–87.
The UniProt Consortium 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45(D1):D158–D169.
Tolosa A, et al. 2008. Rapid evolving RNA gene HAR1A and schizophrenia. Schizophr Res. 99(1–3):370–372.
Tucci V, et al. 2014. Dominant beta-catenin mutations cause intellectual disability with recognizable syndromic features. J Clin Invest. 124(4):1468–1482.
Vallender EJ, Mekel-Bobrov N, Lahn BT. 2008. Genetic basis of human brain evolution. Trends Neurosci. 31(12):637–644.
Vandepoele K, Van Roy N, Staes K, Speleman F, van Roy F. 2005. A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. Mol Biol Evol. 22(11):2265–2274.
Vernes SC, et al. 2011. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet. 7(7):e1002145.
Vernot B, Akey JM. 2014. Resurrecting surviving Neandertal lineages from modern human genomes. Science 343(6174):1017–1021.
Visel A, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457(7231):854–858.
Wang HY, et al. 2007. Rate of evolution in brain-expressed genes in humans and other primates. PLoS Biol. 5(2):e13.
White MA. 2015. Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences. Genomics 106(3):165–170.
Willemsen MH, et al. 2011. Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. J Med Genet. 48(12):810–818.
Wilsch-Brauninger M, Florio M, Huttner WB. 2016. Neocortex expansion in development and evolution – from cell biology to single genes. Curr Opin Neurobiol. 39:122–132.
Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. 23(13):1494–1504.
Xu K, Schadt EE, Pollard KS, Roussos P, Dudley JT. 2015. Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. Mol Biol Evol. 32(5):1148–1160.
Yates A, et al. 2016. Ensembl 2016. Nucleic Acids Res. 44(D1):D710–D716.
Zerbini LF, et al. 2004. NF-kappa B-mediated repression of growth arrest- and DNA-damage-inducible proteins 45alpha and gamma is essential for cancer cell survival. Proc Natl Acad Sci U S A. 101(37):13618–13623.
Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 4:Article17.
Zhang J. 2003. Evolution of the human ASPM gene, a major determinant of brain size. Genetics 165(4):2063–2070.
Zhang X, et al. 2002. Loss of expression of GADD45 gamma, a growth inhibitory gene, in human pituitary adenomas: implications for tumorigenesis. J Clin Endocrinol Metab. 87(3):1262–1267.
Zhang YE, Landback P, Vibranovski MD, Long M. 2011. Accelerated recruitment of new brain development genes into the human genome. PLoS Biol. 9(10):e1001179.
Zhang YE, Long M. 2014. New genes contribute to genetic and phenotypic novelties in human evolution. Curr Opin Genet Dev. 29:90–96.
Zhou Y, Mishra B. 2005. Quantifying the mechanisms for segmental duplications in mammalian genomes by statistical analysis and modeling. Proc Natl Acad Sci U S A. 102(11):4051–4056.
Zhu J, et al. 2008. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nat Genet. 40(7):854–861.
Zimmer F, Montgomery SH. 2015. Phylogenetic analysis supports a link between DUF1220 domain number and primate brain expansion. Genome Biol Evol. 7(8):2083–2088.

Associate editor: Naruya Saitou