Decoding Human Accelerated Regions

Do the portions of our genomes that set us apart from other animals hold the secret to human evolution?

By Katherine S. Pollard | August 1, 2016

© BILL SANDERSON/SCIENCE SOURCE

When the first human genome sequence was published in 2001,1 I was a graduate student working as the statistics expert on a team of scientists. Hailing from academia and biotechnology, we aimed to discover differences in gene expression levels between tumors and healthy cells. Like many others, I had high hopes for what we could do with this enormous text file of more than 3 billion As, Cs, Ts, and Gs. Ambitious visions of a precise wiring diagram for human cells and imminent cures for disease were commonplace among my classmates and professors. But I was most excited about a different use of the data, and I found myself counting the months until the genome of a chimpanzee would be sequenced.

Chimps are our closest living relatives on the tree of life. While their biology is largely

similar to ours, we have many striking differences, ranging from digestive enzymes to spoken language. Humans also suffer from an array of diseases that do not afflict chimpanzees or are less severe in them, including autism, schizophrenia, Alzheimer's disease, diabetes, atherosclerosis, AIDS, rheumatoid arthritis, and certain cancers. I had long been fascinated with hominin fossils and the way the bones morphed into different forms over evolutionary time. But those skeletons cannot tell us much about the history of our immune system or our cognitive abilities. So I started brainstorming about how to extend the statistical approaches we were using for cancer research to compare human and chimpanzee DNA. My immodest goal was to identify the genetic basis for all the traits that make humans unique.

The chimp genome was published in 2005,[2] when I was a postdoc at the University of California, Santa Cruz, and those of 12 other vertebrates followed shortly thereafter. At the same time, computational scientists were busy developing algorithms to scan DNA for similar regions across multiple species. Such sequence conservation suggests that these areas are responsible for critical functions. I took these comparative genomic scans to the next level by writing a computer program to identify DNA sequences that are conserved in other animals but have changed rapidly in humans since we evolved from our common ancestor with chimpanzees. This evolutionary signature predicts a loss or modification of function in humans. My colleagues and I used this two-part pattern to define the fastest-evolving regions of the human genome, known as human accelerated regions (HARs). We published the first 202 HARs in 2006.[3]

An exciting but daunting pattern emerged: only a handful of HARs were in genes. In fact, we had no idea what the vast majority of these putatively functional and uniquely human DNA sequences did, let alone their role in human evolution. HARs are short—on average just 227 base pairs long, much smaller than a gene. They looked like what we called "junk DNA" at that time and would not have been at the top of anyone's list of genomic regions to study, if not for their compelling conservation across most animals and notable differences in humans.

> Humans suffer from an array of diseases that do not afflict chimpanzees or are less severe in them.

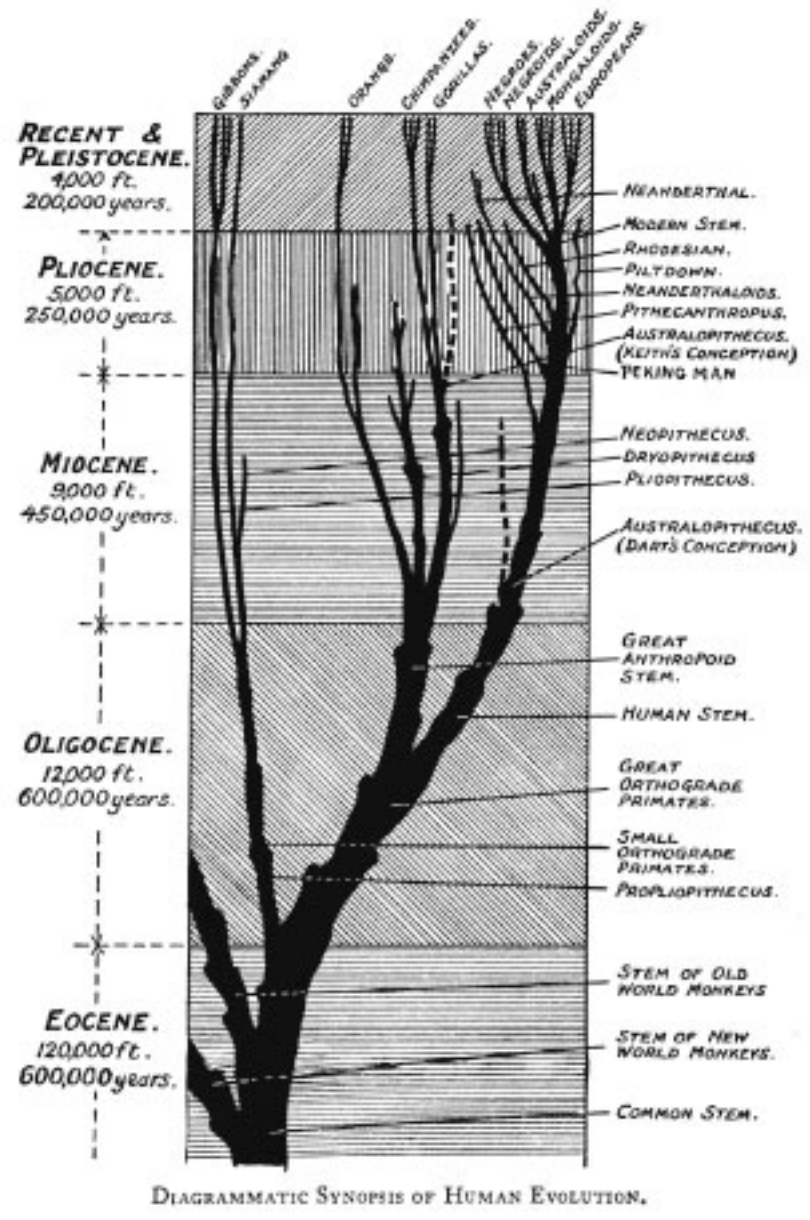
Thanks to innovations in sequencing technology that have produced a cornucopia of genomes, plus some tweaks to the computational methods by different labs, the combined list of identified HARs now includes nearly 3,000 genome segments.[4] But the original trend still holds; nearly all HARs are outside genes, some quite far away from any gene in the genome.

So what were HARs doing that made their sequences so immutable throughout

mammalian evolution? How did the multiple human mutations in each HAR change its function? Ten years in, my group, now based at the Gladstone Institutes in San Francisco, and others continue to investigate these questions, in hopes of better understanding what makes humans different from all other species.

Uniquely human gene regulators

Ignoring human DNA for a moment, HAR regions are some of the most conserved sequences in the genomes of mammals. Some of them are nearly identical between chimpanzee and platypus, for example. This close identity suggests that the information encoded in these sequences is critical, and that changes to the sequences will alter their important instructions. This makes the human mutations in HARs truly unexpected.



DIAGRAMMATIC SYNOPSIS OF HUMAN EVOLUTION.

THE HUMAN TREE: This diagram from 1931 illustrates the evolution of primates from a common ancestor. Later discoveries revealed many of the specifics here to be wrong, but the overall branching of different primate groups continues to inform the study of human biology today.

© SHEILA TERRY/SCIENCE SOURCE

It is tempting to speculate that these mutations destroy or change gene regulatory functions, altering when and where genes turn on. The first two HARs to be functionally characterized support this idea.

HAR1 does not code for a protein but for a long RNA, a type of molecule that guides proteins or modulates their expression.[5] We predicted that the HAR1 RNA could fold into a three-dimensional structure because its conserved sequence has palindromic regions that pair up to form a series of interconnected "stems" that look like ladders—think of an untwisted DNA double helix. This computational prediction was confirmed by RNA structure–probing experiments using human and chimpanzee HAR1 RNAs synthesized in vitro to identify stems. By labeling HAR1 molecules in human and macaque embryos, we discovered that the RNAs functioned in neurons during patterning and layout of the cortex,[6] a brain structure that expanded greatly in size during human evolution.[7] Exactly which genes HAR1 is regulating remains to be determined.

HAR2 (also known as *HACNS1*) encodes neither a protein nor an RNA. Rather, *HAR2* functions as an enhancer, a DNA sequence that works to increase or decrease the level of a gene's expression.[8] An enhancer can be located thousands of base pairs away from the gene it regulates. The gene gets activated when it comes into physical proximity with its enhancer. Studies in mice revealed that human *HAR2* is active in several embryonic tissues, including those that give rise to the wrist and thumb, structures that morphed in our ancestors after their split from a common ancestor with chimpanzees. Once again, the genes that are subject to *HAR2* regulation are still unclear, although GBX2, a transcription factor that controls proper expression of genes involved in embryo morphogenesis, is one promising candidate.

Building on these initial discoveries, researchers have revealed the role of other HARs in gene regulation thanks to advances in techniques that measure gene expression at the single-cell level, track where proteins bind to DNA, and assess other epigenetic properties of the genome. (See "Scaling to Singles," *The Scientist*, May 2016; "Silencing Surprise," *The Scientist*, June 2015.) Integrating this new information into computational models, my colleagues and I predicted that about 5 percent of HARs function as noncoding RNAs, while most are enhancers that control gene expression during embryonic development.[9]

To more concretely test this hypothesis, my team has begun examining the function of nearly 100 of the fastest-evolving HARs, many of which we suspected to have enhancer activity. We inject fertilized mouse or fish eggs with a reporter construct that contains the chimp HAR sequence in front of a gene that will label any cells of the embryo in which the HAR functions as an enhancer. So far, two-thirds of HARs tested for enhancer activity turned on a gene during development.[4] For 26 HAR enhancers, we repeated the experiment with the human sequences. Eight HARs showed differences in their enhancer activity when the human mutations were present.[4] These differences modify how genes were expressed in the developing limb (*HAR2, 2xHAR114*), eye (*HAR25*), and central nervous system (*2xHAR142, 2xHAR238, 2xHAR164, 2xHAR170, ANC516/HARE5*).[4,10] Because relatively few time points have been examined, it is likely that an even higher percentage of the tested HARs are active enhancers at some point during embryonic development or in adult tissues, possibly with human-chimp differences.

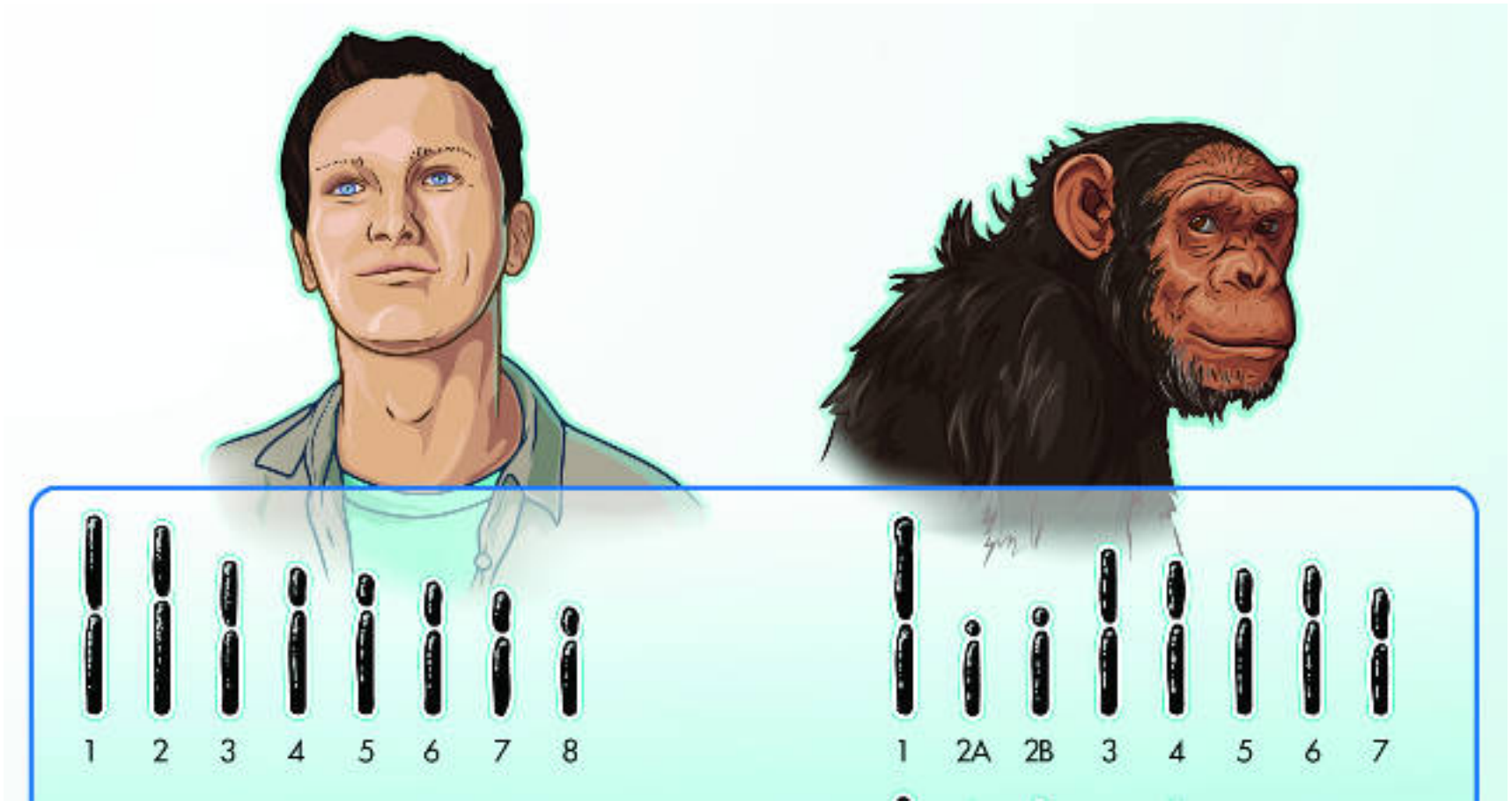
Many HARs are located near genes that control fundamental developmental processes,[9] so their altered regulatory function could have profound effects on human biology. Supporting this, the human version of one HAR enhancer (*ANC516/HARE5*) is active earlier in development and in a larger region of the brain compared to the chimp HAR. Human HARE5 increases expression of its target gene, Frizzled 8, affecting the size and

development of the brain in mice.[10]

These experiments demonstrate that HARs may have changed key developmental programs over the course of human evolution. The *HARE5* study is the closest researchers have come to showing that a HAR sequence affects an organ that is important to human evolution. It is possible that human mutations in HARs could influence human traits such as fine motor skills, spoken language, and cognition. But linking HAR mutations to organismal innovations is hard, given the obvious limitations on testing the effects of genetic changes in humans or apes. Establishing these connections is our biggest challenge going forward.

Emergence of HARs

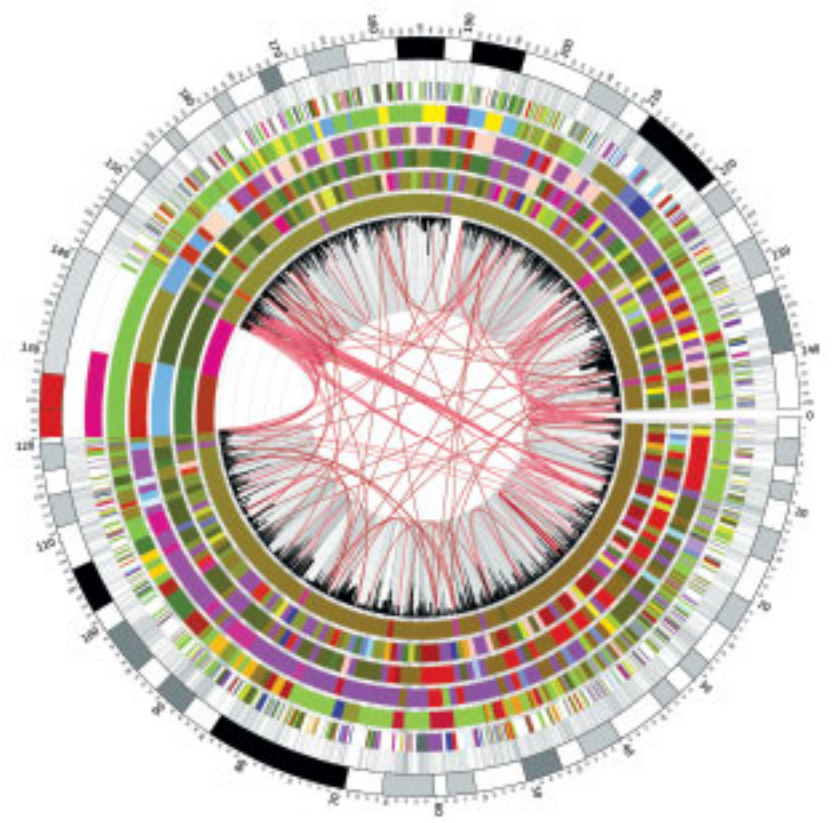
The most recent common ancestor of humans and chimps probably lived about 6 million years ago. The fossil record shows that our two species have changed continually in different ways since then. Knowing when a HAR mutated during human evolution could help researchers link it to traits that changed at the same time. Conversely, as we elucidate which biological processes are affected by HAR mutations, the ages of the mutations could help date the emergence of traits that are hard to discern from fossils.



UNDERSTANDING HUMAN ACCELERATED REGIONS: Sections of the genome that are largely conserved across mammals and even the entire animal kingdom, but differ in humans, are known as human accelerated regions (HARs). Deciphering their function may prove key to understanding what sets humans apart from other organisms. For example, 2xHAR.142 and 2xHAR.114, like many other HARs, function as enhancers, which increase or decrease the level of a gene's expression.
See full infographic: WEB | PDF

Estimating when a HAR evolved is challenging because these calculations rely on comparisons with genomes from hominins that split off from our ancestors at different times in the past. Without these molecular signposts along the human lineage, it is hard to say if a HAR evolved right after the human-chimpanzee split or only a few generations ago. But ancient-DNA sequencing is beginning to shed some light on the issue.[11] For example, by comparing a human HAR sequence with the HAR sequence of an archaic hominin, researchers can estimate if the HAR mutated before, after, or during the time period of our common ancestor.[12] This approach has revealed that the rate at which HAR mutations emerged was slightly higher before we split from Neanderthals and Denisovans.[3,13] As a result, most HAR mutations are millions of years old and shared with these extinct hominins (but not with chimpanzees).

Some HARs have evolved much more recently, however. About 10 percent of mutations in HARs are polymorphic, meaning that only a subset of people carry the mutated sequences, while others have the DNA sequence seen in chimps.[4] These polymorphic changes in HARs happened relatively recently in human evolution—they are unlikely to be more than 1 million years old. But such newer HAR mutations are found in people around the globe, indicating that they predate the long-distance human migrations that began about 60,000 years ago.



SPECIES COMPARISON: This circular genome map shows shared genetic material between humans (outer ring) and (from inner ring outwards) chimpanzee, mouse, rat, dog, chicken, and zebrafish chromosomes. The colors form a heat map, the pattern of which represents hot spots of shared genetic material.

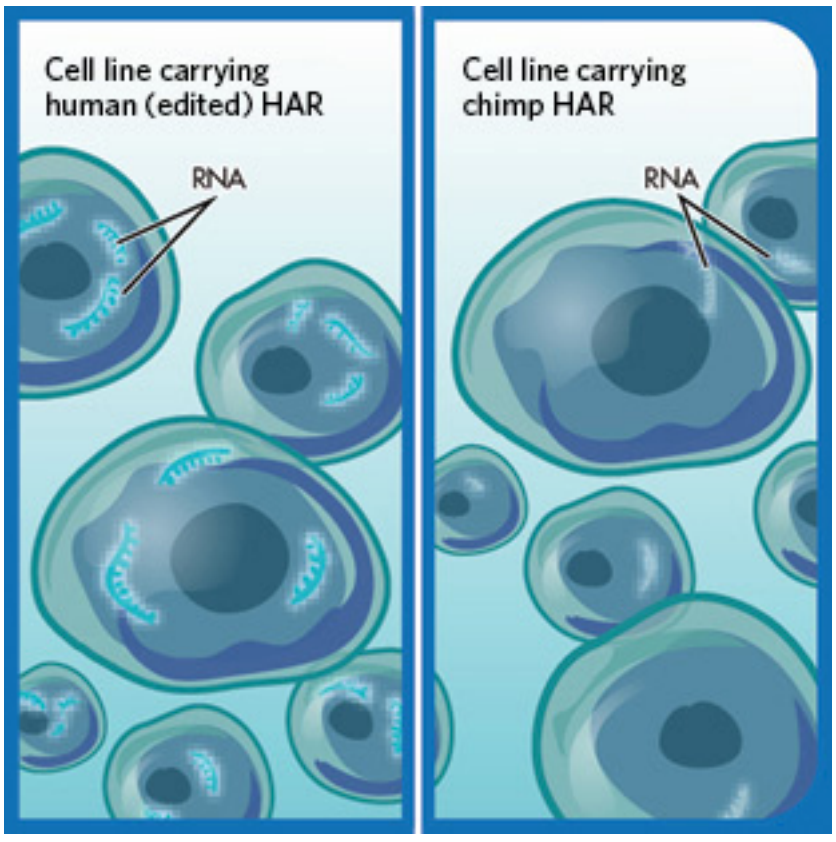
As more human genomes from different populations are sequenced, it will be exciting to see if any traits are associated with carrying the mutated versus ancestral version of polymorphic HARs. This approach has already revealed medically relevant traits linked to Neanderthal ancestry in other parts of the human genome.[14] For example, blood tends to clot more quickly in those of us with the Neanderthal DNA in one such region, while another Neanderthal sequence is associated with depression.

Forces that created HARs

Statistically speaking, the probability that a highly conserved DNA sequence will change

multiple times over 6 million years of evolution is close to zero—that is, unless the forces that have been selecting against mutations in its sequence suddenly change. HAR2, for example, appears to turn on a gene involved in human limb development thanks to the loss of sequences that keep it switched off in the embryos of other species.[15]

Researchers have come a long way toward illuminating the functions of HARs and their potential roles in human evolution, but we are still far from understanding their specific functions in development and other processes. One of the major challenges that we face is establishing causality. Fortunately, emerging technology has made it possible to create brain, heart, and liver cells from a primate skin biopsy[16] and edit the DNA of these cells in the laboratory. These advances allow researchers to test whether specific human mutations alter the ability of HARs to activate genes in human or primate cells.[17] Additionally, because enhancer activity can now be assayed with high-throughput genomic techniques, it is conceivable to move from testing HARs one by one to investigating thousands of them in parallel. These exciting breakthroughs promise to accelerate research on HAR function and the evolutionary forces that shaped HARs.



NAILING DOWN HAR FUNCTION: A remaining challenge in the study of human accelerated regions (HARs) is establishing their specific functions during development and other biological processes. But modern stem cell technologies could provide the answer.

See full infographic: WEB | PDF

High-performance computing and algorithm development will continue to be critical to HAR research. My analysis that discovered the original 202 HARs would still be running today if I had implemented it on a single desktop computer rather than a 1,000-node computer cluster. Instead of waiting for the program to end, we spent the past decade showing that HARs are key regulators of embryonic development. This is a huge step forward from HARs being viewed as bizarre junk DNA of unknown function. Looking ahead to when all of our genomes have been analyzed and tools exist for precise editing of HARs in human cells, it seems possible to figure out what happened when each of these evolutionarily conserved sequences suddenly mutated in humans.

***Katherine S. Pollard* is a biostatistician at the Gladstone Institutes in San Francisco, California.**

References

1. International Human Genome Sequencing Consortium, “Initial sequencing and analysis of the human genome,” *Nature*, 409:860-921, 2001.
2. The Chimpanzee Sequencing and Analysis Consortium, “Initial sequence of the chimpanzee genome and comparison with the human genome,” *Nature*, 437:69-87, 2005.
3. K.S. Pollard et al., “Forces shaping the fastest evolving regions in the human genome,” *PLOS Genet*, 2:e168, 2006.
4. M.J. Hubisz, K.S. Pollard, “Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution,” *Curr Opin Genet Dev*, 29:15-21, 2014.
5. K.S. Pollard et al., “An RNA gene expressed during cortical development evolved rapidly in humans,” *Nature*, 443:167-72, 2006.
6. A. Beniaminov et al., “Distinctive structures between chimpanzee and human in a brain noncoding RNA,” *RNA*, 14:1270-75, 2008.
7. S. Herculano-Houzel, “The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost,” *PNAS*, 109(Suppl 1):10661-68, 2012.
8. S. Prabhakar et al., “Human-specific gain of function in a developmental enhancer,” *Science*, 321:1346-50, 2008.
9. J.A. Capra et al., “Many human accelerated regions are developmental enhancers,” *Philos Trans R Soc Lond B Biol Sci*, 368:20130025, 2013.
10. J.L. Boyd et al., “Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex,” *Curr Biol*, 25:772-79, 2015.
11. J. Krause, S. Pääbo, “Genetic time travel,” *Genetics*, 203:9-12, 2016.
12. R.E. Green et al., “A draft sequence of the Neandertal genome,” *Science*, 328:710-22, 2010.
13. K.S. Pollard et al., “Analysis of human accelerated DNA regions using archaic hominin genomes,” *PLOS ONE*, 7:e32877, 2012.
14. C.N. Simonti et al., “The phenotypic legacy of admixture between modern humans and Neandertals,” *Science*, 351:737-41, 2016.
15. K. Sumiyama, N. Saitou, “Loss-of-function mutation in a repressor module of human-specifically activated enhancer HACNS1,” *Mol Biol Evol*, 28:3005-07, 2011.
16. I. Gallego Romero et al., “A panel of induced pluripotent stem cells from chimpanzees: A resource for comparative functional genomics,” *eLife*, 4:e07103, 2015.
17. S. Weyer, S. Pääbo, “Functional analyses of transcription factor binding sites that

differ between present-day and archaic humans," *Mol Biol Evol*, 33:316-22, 2016.