Evolution of genetic and genomic features unique to the human lineage

Majesta O'Bleness[1], Veronica B. Searles[1], Ajit Varki[2,3,4], Pascal Gagneux[3,4] and James M. Sikela[1,4]

Abstract | Given the unprecedented tools that are now available for rapidly comparing genomes, the identification and study of genetic and genomic changes that are unique to our species have accelerated, and we are entering a golden age of human evolutionary genomics. Here we provide an overview of these efforts, highlighting important recent discoveries, examples of the different types of human-specific genomic and genetic changes identified, and salient trends, such as the localization of evolutionary adaptive changes to complex loci that are highly enriched for disease associations. Finally, we discuss the remaining challenges, such as the incomplete nature of current genome sequence assemblies and difficulties in linking human-specific genomic changes to human-specific phenotypic traits.

Accelerated evolution
More nucleotide or copy number changes in a particular region or gene than would be expected from background rates of mutation over time (for example, in cytochrome *c* oxidase subunit Va (*COX5A*)).

[1]*Department of Biochemistry and Molecular Genetics, Human Medical Genetics and Neuroscience Programs, University of Colorado School of Medicine, Aurora, Colorado 80045, USA.*
[2]*Department of Medicine, University of California, San Diego, La Jolla, California 92093, USA.*
[3]*Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093, USA.*
[4]*Center for Academic Training and Research in Anthropogeny, La Jolla, California 92093, USA.*
Correspondence to J.M.S.
e-mail: James.Sikela@ucdenver.edu
doi:10.1038/nrg3336

Many phenotypic traits that are unique to the human lineage are likely to have resulted from selective pressures on our genome and the unique demographic history since our divergence from the *Pan* lineage approximately 6 million years ago (BOX 1). A fundamental question that relates to the origin of our species is which genomic sequences contributed to the unique evolutionary trajectory taken by the human lineage. With rapid advances in genomic technologies facilitating the comparison of numerous genomes within and between species, we are in an unprecedented era of advancement in comparative genomics. With the availability of draft genomes for nine primate species, including all of the 'great apes' (namely, chimpanzees, bonobos, gorillas and orangutans), and the ability to sequence the genomes of multiple individuals rapidly within each species, we have never been in a better position to evaluate what genetic changes contributed towards making us human.

A major goal of studies that identify human-lineage-specific (HLS) genomic changes is to correlate genotype with phenotype; however, this task also remains most formidable. As direct human experimentation is not possible, researchers have historically relied on naturally occurring variation and disease to understand HLS implications. A further limitation is the comparatively limited information about non-human primate phenotypes. This type of observational reliance frequently makes particular changes difficult to interpret. Although variation and disease are still heavily relied on, emergent technologies, such as heterologous expression of human regulatory regions in mice[1,2], are allowing for evolutionary hypotheses to be tested in ways that were previously not possible. Work in this field has substantially advanced in recent years: the number of gene-to-phenotype candidates has more than doubled since the topic was last covered in two related reviews[3,4].

This article discusses current knowledge of the genetic and genomic changes that make *Homo sapiens* different from other primates and puts particular emphasis on recent advancements. We explore genetic changes that may have contributed to human-specific traits, and where applicable we look at the hypothesized evolutionary pressures, such as accelerated evolution (BOX 2) and positive selection, that underlie these changes in human characteristics. We examine HLS genomic changes with a brief look at the technologies that made these discoveries possible and explore a representative group of HLS gene changes along with their associated phenotypes. We then address the growing number of HLS genetic and genomic changes connected to disease, including the correlation between complex loci and multiple disease associations. We conclude with a look at the future challenges in compiling a comprehensive list of HLS changes and their associated traits.

Uniquely human genome changes

The ability to identify genomic changes that are unique to humans depends on the definition of HLS events. In

Copy number changes
Increases or decreases in the number of copies of a gene or segment (for example, in SLIT–ROBO rho GTPase-activating protein 2 (*SRGAP2*)).

addition to the genome sequence of *H. sapiens*, there are two ancient hominin lineages — Neanderthals[5] and Denisovans[6] — for which draft genome sequences are currently available. As these two genome sequences are far less accurate and complete than the human assembly, this Review does not rely on them unless the ancient hominin sequencing data for a particular sequence is of a high quality. Our working definition of HLS therefore requires that changes found in the *H. sapiens* genome be uniquely different from those found in other extant primates. As this Review requires that the changes be well established as being HLS and that they be based on multiple out-group comparisons, it excludes studies reporting on human and chimpanzee differences alone.

Identification of HLS genomic changes can be complicated by several factors that can potentially occlude accurate comparisons of gene and genomic sequences. Gene annotation is often imprecise and can change between different genome builds, making it difficult to determine whether a change is real or whether it represents computational and/or assembly error. The sample size of sequenced individuals can also be an impediment as an apparent HLS change may be polymorphic in only the human population. In addition, the lack of sufficient individuals from the other sequenced primates can make the ancestral state difficult to determine. This is exacerbated by the fact that, with the possible exception of bonobos (*Pan paniscus*), all great ape species harbour far more sequence diversity than humans[7]. Finally, it is important to have an effective number of primate outgroups to determine HLS status. For example, comparison of gene copy number changes between humans and the great apes found that 57% of the genes that are increased in copy number in humans versus chimpanzees are not HLS[8]. With these criteria defined, we discuss examples of HLS genomic changes below.

Box 1 | **Examples of human-lineage-specific traits and potential forces shaping them**

Human-lineage-specific (HLS) traits are phenotypes of the human lineage that arose after the split from the *Pan* lineage. A substantial number of forces are likely to have contributed to the development and maintenance of these traits, and several examples are listed here. Plausible forces commonly discussed are macro- and micro-level climate changes that occurred frequently over the course of human evolution[106] and that may have selected for rapid HLS changes to survive novel climatic challenges. Although an enhanced cognitive capacity would clearly be beneficial in dealing with such extreme and abrupt environmental changes, other important HLS phenotypic changes were also occurring. For example, anatomical and physiological changes associated with endurance running, such as HLS changes in the musculoskeletal system and in energy use and metabolism may have allowed novel hunting practices, such as persistence hunting, to emerge. These in turn may have allowed the energetic benefits of meat to be increasingly incorporated into the human dietary regimen[10,75,107]. An incomplete but representative list of traits identified as unique to the human lineage is shown in the table along with possible selective advantages[10,75,107–110]. A more complete list can be found in the Matrix of Comparative Anthropogeny (BOX 5).

Example of phenotypic feature	Human-lineage-specific trait	Possible evolutionary advantages
Brain growth trajectory	Prolonged postnatal brain growth and delayed myelinization period; enhanced cognition	Allowed creation of novel solutions to survival threats; increased the critical period for learning new skills; facilitated emergence of uniquely human cognitive skills
Brain size	Increased brain/body size ratio; enhanced cognition	Allowed creation of novel solutions to survival threats; improved social cognition
Descended larynx	Portion of tongue resides in throat at level of pharynx; larynx descended into throat	Helped to develop spoken language
Eccrine sweat gland density	Higher density of eccrine glands; enhanced sweating capacity	Enhanced cooling ability; allowed protection of heat-sensitive tissues (for example, the brain) against thermal stress; facilitated endurance running
Endurance running	Improved energy use during periods of high energy demand; increased capacity to transfer energy (in the form of glycerol) from fat stores to muscle; anatomical changes relating to running ability	Allowed persistence hunting to emerge as a viable strategy for accessing the benefits of increased meat consumption; increased range of food sources; improved diet may have facilitated brain evolution
Labour	Earlier onset and longer duration of labour	Partially protected the child and mother from damage due to increased head circumference
Lacrimation	Emotional lacrimation (crying)	Enhanced emotional communication within social groups; increased affective communication
T cell function	Relative T cell hyper-reactivity	Enhanced immune function
Thumb	Increased length; more distally placed; larger associated muscles	Allowed creation of more detailed tools; allowed manipulation of objects on a finer scale

Box 2 | **Accelerated evolution**

There are a number of genomic regions that have undergone substantial alteration of sequence or rearrangement in the human lineage. Accelerated evolution refers to situations in which sequence changes occur at a rate greater than the neutral mutation rate. Accelerated evolution implies that the changes have been selected because of their advantageous nature and thus have undergone rapid fixation. Identification of these regions relies on multiple methods and differs depending on whether the change is at the coding sequence, non-coding sequence, copy number or other structural level.

At the protein-coding sequence level, a comparison of K_a/K_s values between sequences is often used, where K_a is the number of nonsynonymous substitutions and K_s is the number of synonymous substitutions. Most gene-coding regions will have K_a/K_s ratios well below 1.0 owing to the effects of purifying selection. By contrast, coding regions under positive selection will exhibit a higher frequency of nonsynonymous changes and, as a result, a higher K_a/K_s ratio. Through this method, studies have identified accelerated evolution in the human lineage of a number of genes, one example being genes involved in nervous system function[111]. However, these estimates can be confounded by gene conversion events that erase evidence of selection by creating stretches of identical nucleotide sequences between homologous genes[112].

Evaluation of non-coding sequences is not as straight forward because of the difficulty in interpreting the importance of a change. Thus, studies identifying regions of accelerated evolution in non-coding regions have relied on looking for human-specific mutations in sequences that are highly conserved across mammals. An example of this is the identification of highly accelerated region 1 forward (*HAR1F*)[113]. *HAR1F* is a non-coding RNA expressed in the fetal brain that colocalizes with reelin, a protein that is important for cortical development.

Accelerated evolution may also occur at the level of whole genes or genomic regions in the form of copy number variations and structural rearrangements. Identification of such regions typically involves looking for HLS sequence copy number expansions and contractions that can range from as small as a few nucleotides to as large as segmental duplications identifiable by fluorescent *in situ* hybridization. Unlike the identification of accelerated evolution at the single-nucleotide level, there are no rigorous statistical tests for these types of changes, and they are mainly observational.

Large-scale changes. Large-scale genomic differences between humans and our closest primate relatives have been noted since the 1970s, when chromosomes were examined using chromatin-stained banding techniques[9]. Given technological limitations, observable differences were restricted to the detection of: a change in the haploid number from 24 to 23 chromosomes owing to the fusion of two ancestral ape chromosomes, resulting in humans having one large chromosome 2; the addition of human-specific constitutive heterochromatic C bands on chromosomes 1, 9, 16 and Y; and human-specific pericentric inversions on chromosomes 1 and 18 (REF. 9). Although no conclusive evidence has as of yet directly linked these cytogenetically visible events to HLS traits, the genomic regions at which these events took place tend to be hotbeds of recent gene duplication, harbouring many unique human-specific genes and copy number variations (CNVs)[8,10] (BOX 3). This suggests that selection for these novel genes drove the changes to fixation, although drift cannot be formally excluded as an additional factor. Timing for these events continues to be of interest, with a recent paper based on segmental duplications in chimpanzees and gorillas estimating that the chromosome 2 fusion event occurred ~4–5 million years ago[11].

In the late 1980s, application of fluorescent *in situ* hybridization (FISH) to banded chromosomes and its further use in FISH chromosome painting in the 1990s permitted the evaluation of large-scale structural changes between humans and great apes that were not visible with conventional banding techniques[12,13]. More recently, these studies were aided by interspecies bacterial artificial chromosome (BAC)-based array-based comparative genomic hybridization (array CGH) experiments, which identified over 60 HLS segmental duplications greater than 65 kb in size[14,15]. In addition, similar efforts identified large genome rearrangements such as those on nearly half of all human telomeres[16].

Small-scale changes. Small-scale changes encompass all differences smaller than those identified by the large-scale tests that have previously been discussed and that have a resolution limit of ~20 kb. These include single base pair changes, insertions and deletions (indels) of varying size and gene copy number differences. The differences can affect coding regions, non-coding regulatory regions and repetitive sequence content. Strategies for identifying small-scale changes often involve scanning the genome for signatures of positive selection when comparing humans with non-human primates and rodents. Although the results of these studies do not always coincide, there is a substantial overlap in the phenotypes implicated, including taste and olfaction, immunity, signal transduction, lipid metabolism, chaperone activity, motor activity and structural support[8,17–20]. However, it is likely that many important traits have yet to be identified given the large proportion of genes with no known function.

Initial estimates of sequence divergence between human and chimpanzee were ~1.2%[17], a value based on the number of single-nucleotide substitution differences between the two genomes. However, these estimates did not account for unalignable regions between species that were due to structural divergences such as indels, highly duplicated sequences and CNVs. Although more recent divergence estimates reach as high as 5% when taking all types of variability into account[21], high-confidence divergence estimates remain elusive. For example, the incompleteness of other primate genome assemblies impedes accurate assessment of the uniquely human indel content, for which estimates range from 0.21% to 3%[17,22]. In addition, the importance of using multiple primate outgroups in making HLS assignments is borne out by recent studies of interspecific CNV. For example, these studies show a substantial number of genes with an increased copy number across the three African great apes[8,23], whereas the human copy number resembles that of orangutans and other primate lineages.

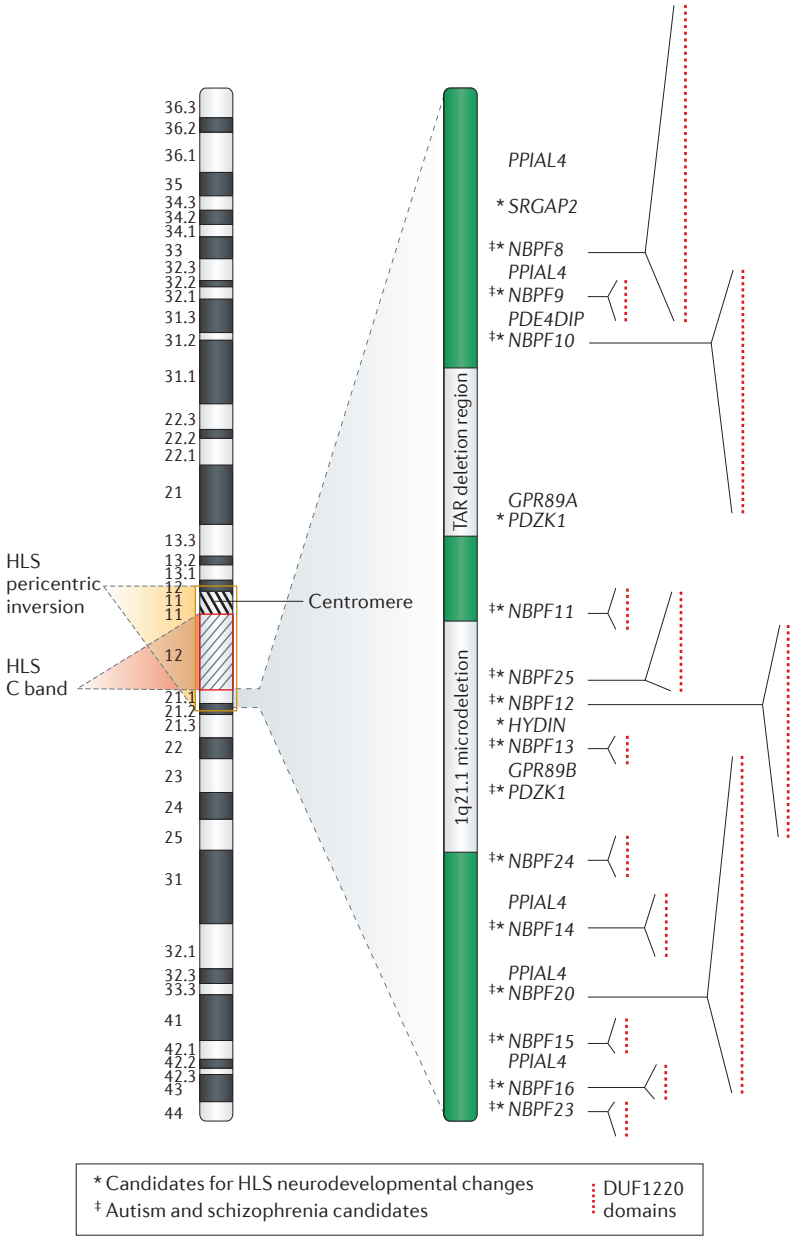
The first genome-wide and first gene-based array CGH study comparing humans to all four great ape lineages identified 140 genes with HLS changes (134 and 6 genes that showed HLS copy number gain and loss, respectively)[8]. Most of these HLS changes were confirmed after expanding the study to include ten primate lineages[10]. Interestingly, these HLS changes showed strong positional biases, frequently clustering in the genome at pericentromeric, subtelomeric and particularly complex duplication-rich regions. Indeed, the greatest number of HLS gene copy number

Fluorescent *in situ* hybridization
(FISH). A technique used to visualize the location of specific DNA sequences on chromosomes.

Array-based comparative genomic hybridization
(Array CGH). A microarray-based method for detecting copy number variation in the genome.

Box 3 | **Human-lineage-specific changes at 1q21.1**

Large cytogenetically visible changes in the genome structure were among the first human-specific genomic changes noted between humans and great apes. More recently, it has been determined that these regions harbour more importance than just being human-specific heterochromatin. Indeed, such regions are frequently adjacent to regions that are greatly enriched for evolutionarily recent gene duplications and that often function as gene nurseries[8,10]. For example, the 1q21.1 region of the genome, which lies adjacent to the human-specific 1q12 C band and within the human-lineage-specific (HLS) chromosome 1 pericentric inversion, has undergone substantial genomic enlargement owing to numerous HLS copy number expansions within the region, as shown by the green bands in the figure. Numerous findings have identified the 1q21.1 region as being highly enriched for HLS copy number expansion, including: striking HLS copy number increases of DUF1220 protein domains (240 copies that map to 1q21.1 are shown by the red dots in the figure); copy number increases of the gene family that encodes them (namely, the neuroblastoma breakpoint family (NBPF))[8]; and duplicative transpositions of non-1q21.1 genes to the region, such as SLIT–ROBO Rho GTPase-activating protein 2 (*SRGAP2*)[71] and *HYDIN*[114]. In addition, all of the above genes mentioned are candidates to explain both HLS neurodevelopmental changes and cognitive disorders[45,46,72,96].



increases was found adjacent to the previously mentioned human-specific C bands on chromosomes 1 and 9 (REF. 8).

More recently, sequence read depth has been used to estimate copy number from next-generation sequencing (NGS) platforms, providing a complement to array-based strategies[24]. Although NGS has confirmed the findings of many previous interspecific array CGH studies[8] and identified new HLS candidates[25], its wider use has been slowed by the short read-length capabilities of current NGS platforms (typical NGS read length is 50–150 bp). Short reads of highly duplicated sequences will often lack adjacent single copy sequences that serve to anchor the sequence read within the genome, making it difficult to localize duplicated copies accurately[26,27]. Using NGS for sequencing regions containing structural variations (for example, highly similar duplicated sequences) has thus required the implementation of novel bioinformatics methods, delaying their broad use within the field[27]. For a full Review of the subjects, see REF. 26.

Other major contributors to HLS genomic content are repetitive elements such as transposable element insertions, which constitute roughly half of the human genome. Transposable elements comprise both DNA transposons and retrotransposons, with retrotransposons subdivided into long terminal repeat (LTR)-containing and non-LTR elements. LTR-containing elements include endogenous retroviruses (ERVs), and recent hominid evolution has led to the accumulation of lineage-specific subsets of ERVs in great apes and humans[28]. Although non-LTRs, such as long interspersed element 1 (LINE-1; hereafter referred to as L1), *Alu* and hominid-specific SINE–VNTR–*Alu* composite repetitive (SVA) elements, encompass 75% of human repetitive content, it is difficult to determine what percentage of these are HLS owing to the incomplete nature of other primate genome sequence assemblies and their inability to accurately represent repetitive elements such as transposable elements[29]. Among those retrotransposons that have been successfully identified as HLS, the L1-Hs subfamily — the youngest of five HLS L1 subfamilies[30] — is of particular interest. At ~80–100 copies, L1-Hs insertions are the only transposable elements that are still active within the human genome[31]. L1-Hs elements are hypothesized to have a pivotal role in human neural plasticity as they are highly active during neurogenesis and contribute to neuron-specific genomic diversity[32,33].

Taken together, these studies establish that there are far more genomic differences between human and other primate genomes than was originally thought[4]. Nevertheless, it is difficult to estimate precisely what fraction of the genome contains HLS sequences. Although the human genome is the most complete and accurate mammalian genome sequence currently available, it still contains many sequence gaps in complex genomic regions, which may harbour important HLS genes[34]. Confounded by the far more incomplete nature of all other primate genomes, there are likely to be many HLS changes that have yet to be discovered.

Unique gene differences and associated traits

The number of identified HLS gene differences has rapidly increased in recent years, and this rate of discovery will probably continue in the future. Although there is a lag in the number of HLS phenotypes that have been associated with these HLS genomic changes, there is a substantial body of data attempting to link the two (see TABLE 1 for examples and FIG. 1 for their genomic location). Roughly half of these associations relate to brain morphology and/or cognition, but it is not known whether this tendency is a true representation of genetic change or whether it reflects a bias in research focus. Other areas with a substantial number of HLS changes are disease resistance and immunity, metabolism, physiological and anatomical differences, and changes in human reproduction and parturition. Some representative case studies of successful efforts for connecting HLS genotype to phenotype are presented in BOX 4. These represent only a sampling of discovered HLS genes and their associated traits; a more comprehensive list can be found in the genetics and genomics domains at the Center for Academic Research and Training in Anthropogeny (CARTA) Matrix of Comparative Anthropogeny (MOCA) resource on human origins. MOCA also includes the ability to link these genetic and genomic changes to many other features of human uniqueness, in domains ranging from molecules to culture (BOX 5).

Alteration of gene structure resulting from splicing. Alteration of the gene structure provides a major mechanism through which evolutionarily adaptive changes can be introduced. A common means of modifying gene structure is through alterations in transcriptional splicing. Several studies have identified genes that are differentially spliced in the human lineage, including differential expression in the brain[35] and substantial numbers of genes involved in metabolism and morphological development[36]. A recent survey of human-specific transcript variants found 112 genes showing differential HLS transcripts as the result of novel promoters, exons and splicing sites, most of which are the result of transcriptional element insertions[37]. One of the best-characterized HLS transcripts is cholinergic receptor, muscarinic 3 (*CHRM3*), which encodes a G-protein-coupled receptor that mediates multiple physiological functions through its control of smooth muscle contraction[38]. In humans, an L1-Hs transposon insertion occurred at the 5′ end of the gene, resulting in a novel first exon, promoter and transcript[37,38]. The L1-Hs-derived transcript is the only *CHRM3* transcript expressed in the placenta and could be important for human gestation[38].

Alteration of gene structure by protein domain amplification. Gene structure can also be altered through changes in the number of protein domains in a gene. Although the importance of gene duplication to evolutionary change has been emphasized since 1970 (REF. 39), an appreciation for the contributions of protein domain amplification, the process by which a protein domain undergoes a copy number increase, has only recently emerged[40]. The most striking example of this process has been reported for the DUF1220 protein domain, which shows the largest HLS copy number increase of any protein-coding region in the human genome[8,41,42]. DUF1220 domains are encoded within genes of the neuroblastoma breakpoint family (NBPF)[43,44], and although there are several HLS NBPF genes found in the human genome, the great majority of HLS copies of DUF1220 have arisen by intragenic domain hyper-amplification[42]. With 272 copies, humans have more than twice the copy number of chimpanzees (which have 126 copies, the next highest number), whereas mice and rats have only one copy. It is estimated that, on average, 28 additional copies of DUF1220 domains have been added specifically to the human genome every million years since the human and *Pan* lineages diverged[42]. Recent correlative data from evolutionary studies and studies of brain size in normal and pathological populations (such as studies of individuals with microcephaly and macrocephaly) support the view that DUF1220 copy number is a general effector of brain size and may be largely responsible for the dramatic evolutionary expansion in brain size that occurred in the human lineage[45,46].

Alteration of gene structure by amino acid change. Gene structure can also be modified by smaller local alterations that result in changes to the amino acid sequence. There are numerous examples of accelerated genome evolution that have been linked to amino acid change[19,20] (BOX 2). One gene that may have ramifications for HLS metabolic changes is cytochrome *c* oxidase subunit Va (*COX5A*), the fifth of ten nuclear encoded subunits that make up the terminal proteins in the mitochondrial electron transport chain[47]. The genes encoding these subunits are generally highly conserved, but several show an increased rate of nonsynonymous substitutions in the anthropoid primate lineages, and *COX5A* contains two HLS amino acid changes[48]. Although a complete understanding of *COX5A* function is lacking, its interaction with thyroid hormone T2 suggests that the changes are important in regulating fat metabolism[49,50]. A second example of a functionally important HLS amino acid change is provided by forkhead box P2 (*FOXP2*), which is proposed to have had an impact on human speech development[51] (BOX 4).

Alteration of gene function by pseudogenization. Not all gene alterations generate functional variants, and indeed such structural changes often produce non-functional genes. This is the case with pseudogenization, in which a sequence alteration renders the gene inactive, although most of the gene remains intact within the genome. A recent analysis using updated sequencing data found 38 fixed HLS pseudogenes in the human genome[52], and only nine of these were fixed single copy genes. No excess of pseudogene fixation over expected rates was detected (that is, no evidence was found in support of Olson's 'less-is-more' hypothesis[53]). One notable pseudogene is apolipoprotein C1 (*APOC1*), which is involved in lipoprotein metabolism[54]. Although great

Protein domains
Discrete portions of a protein sequence that may evolve and function independently of the rest of the protein (for example, in the DUF1220 domain).

Domain amplification
Intragenic copy number increase of a protein domain (for example, in the DUF1220 domain).

Amino acid change
A DNA change that leads to a change at the protein sequence level (for example, in forkhead box P2 (*FOXP2*)).

Pseudogenization
Loss of gene function while most of the gene is retained (for example, in apolipoprotein C1 (*APOC1*)).

'Less-is-more' hypothesis
The hypothesis that gene loss has a major role in evolution.

Table 1 | **Partial list of genes and genetic elements showing human-lineage-specific changes**

Gene or element	Mechanism of change	Proposed phenotype	Phenotypic certainty	Possible gene-associated diseases	Refs
AR	Deletion of regulatory DNA	Loss of sensory vibrissae and penile spines	Likely	Androgen insensitivity; hypospadias; muscular atrophy; prostate cancer	1
APOC1	Pseudogene	Unknown	Not applicable	Alzheimer's severity; atherosclerosis; coronary heart disease	55–59
AQP7	Copy number increase	Energy use	Plausible	Nonfunctional glycerol response to exercise	10,73–75
ASPM	Positive selection	Increased brain size	Plausible	Microcephaly	94,95
CDK5RAP2	Positive selection	Increased brain size	Plausible	Microcephaly	95,118
CCL3L1	Novel gene variant	Immune system function	Likely	HIV and AIDS; Kawasaki's disease; rheumatoid arthritis; chronic hepatitis C	89
CHRM3	Novel exon	Change in human reproduction	Plausible	Eagle–Barrett syndrome	38
CHRFAM7A	Copy number increase	Higher brain function	Plausible	P50 sensory gating deficit	8,89,119
CMAH	Pseudogene	Changed sialic acid composition on all cells	Definite	Duchenne's muscular dystrophy; red-meat-related carcinoma risk	62,63
COX5A	Amino acid change	Mitochondrial metabolism	Plausible	Unknown	49
DRD5	Copy number increase	Regulation of memory; attention; movement	Likely	DRD5 deficiency; attention-deficit hyperactivity disorder; primary cervical dystonia	8,89
DUF1220 and NBPF family	Protein domain copy number increase	Brain size	Likely	Microcephaly; macrocephaly	41,42, 45,46
FCGR1A	Copy number increase	Immune system function	Plausible	IgG receptor I phagocyte deficiency	25,89
FSHR	Positive selection	Decreased gestation; birth timing	Plausible	Amenorrhoea; infertility; ovarian dysgenesis type 1; ovarian hyperstimulation syndrome	120,121
FOXP2	Amino acid change	Speech and language development	Definite	Speech and language disorder 1	51
GADD45G	Deletion of regulatory DNA	Expansion of human forebrain	Plausible	Thyroid carcinoma	1
HACNS1	Positive selection	Changes in anterior wrist and thumb	Likely	Unknown	2
HAR1F	Positive selection	Neocortex development	Plausible	Unknown	113
MRC1	Novel gene variant	Inflammation recovery	Plausible	Leprosy manifestation	89
MCPH1	Positive selection	Brain size	Plausible	Microcephaly	95,122
MYH16	Pseudogene	Craniofacial musculature	Plausible	Unknown	64
NCF1	Copy number increase	Phagocyte generation of superoxides	Likely	Chronic granulomatous disease; Williams–Beuren syndrome	89
NAIP	Copy number increase	Inhibition of apoptosis	Likely	Spinal muscular atrophy	8,25,89
OCLN	Copy number increase	Regulation of TGFβ; cell migration	Likely	Hepatitis C; band-like calcification with simplified gyration and polymicrogyria	8,89,123

Table 1 (cont.) | **Partial list of genes and genetic elements showing human-lineage-specific changes**

Gene or element	Mechanism of change	Proposed phenotype	Phenotypic certainty	Possible gene-associated diseases	Refs
PAK2	Copy number increase	Neuronal differentiation	Plausible	3q29 microdeletion syndrome	8,124
PMP2	Copy number increase	Protection from demyelination	Plausible	Charcot–Marie–Tooth peroneal muscular atrophy	8,125
PDE4DIP	Copy number increase	Higher brain function	Plausible	Myeloproliferative disorder associated with eosinophilia	8,25, 125,126
PCDH11X and *PCDH11Y*	Copy number increase	Cerebral asymmetry; altered language development	Likely	Klinefelter's syndrome; Alzheimer's disease; prostate cancer	8,81–85
SIGLEC5	Expression change	T cell hyperactivation	Likely	Susceptibility to T cell-mediated disease	86
SIGLEC6	Expression change	Prolonged labour	Plausible	Pre-eclampsia	87
SIGLEC11	Gene conversion; expression change	Alleviate neurotoxicity from activated microglia	Likely	Unknown	65,66
SIGLEC13	Gene loss	Disease resistance to sialylated bacteria	Likely	Unknown	70
SLC6A13	Copy number increase	Higher brain function	Plausible	Unknown	8,127
SMN2	Novel gene variant	Motor neuron maintenance	Likely	Spinal muscular atrophy severity	8,89
SRGAP2	Copy number increase	Increased neuronal branching	Likely	Early infantile epileptic encephalopathy	8,25, 71,72
SPANXB and *SPANXC*	Copy number increase	Post-meiotic spermatogenesis	Likely	Unknown	128,129

Genes that have associated human-lineage-specific (HLS) traits are listed with an assigned level of certainty with regard to their impact on human uniqueness. Certainty ranges from plausible (that is, the association is still hypothetical on the basis of what is known about the gene) to likely (that is, there may be a disease association or animal model evidence to substantiate the claim) or definite (that is, there are multiple lines of supporting evidence). In cases in which the gene has been implicated in a disease but a HLS phenotype has not been proposed, the certainty column is not applicable. In addition, associated disease links are given if known. Disease phenotypes are listed partly on the basis of information in the Online Mendelian Inheritance of Man database. Details regarding these genes and additional examples of genes associated with HLS traits can be found in the genetics domain of the Matrix of Comparative Anthropogeny (MOCA) (BOX 5). *AR*, androgen receptor; *APOC1*, apolipoprotein C1; *AQP7*, aquaporin 7; *ASPM*, asp (abnormal spindle) homologue, microcephaly-associated; *CDK5RAP2*, CDK5 regulatory subunit associated protein 2; *CCL3L1*, chemokine (CC motif) ligand 3-like 1; *CHRM3*, cholinergic receptor, muscarinic 3; *CHRFAM7A*, *CHRNA7* and *FAM7A* fusion; *CMAH*, cytidine monophospho-*N*-acetylneuraminic acid hydroxylase, pseudogene; *COX5A*, cytochrome *c* oxidase subunit Va; DRD5, dopamine receptor D5; *FCGR1A*, Fc fragment of IgG, high affinity 1a, receptor; *FSHR*, follicle-stimulating hormone receptor; *FOXP2*, forkhead box P2; *GADD45G*, growth arrest and DNA-damage-inducible, gamma; *HACNS1*, human accelerated conserved non-coding region 1; *HAR1F*, highly accelerated region 1 forward; *MCPH1*, microcephalin 1; *MRC1*, mannose receptor C type 1; *MYH16*, myosin heavy chain 16 pseudogene; *NCFI*, neutrophil cytosolic factor I; *NAIP*, NLR family, apoptosis inhibitory protein; NBPF, neuroblastoma breakpoint family; *OCLN*, occludin; *PAK2*, p21 protein (Cdc42/Rac)-activated kinase 2; *PCDH11X*, protocadherin 11 X-linked; *PDE4DIP*, phosphodiesterase 4D interacting protein; *PMP2*, peripheral myelin protein 2; *SIGLEC5*, sialic-acid-binding Ig superfamily lectin 5; *SLC6A13*, solute carrier family 6 (neurotransmitter transporter, GABA) member 13; *SMN2*, survival of motor neuron 2, centromeric; *SPANXB*, sperm protein associated with the nucleus, X-linked family member B; *THBS4*, thrombospondin 4.

apes have two *APOC1* genes, which encode negatively and positively charged forms of the protein, a premature stop codon in the gene that encodes the negatively charged protein resulted in humans who have only the positive APOC1 protein[55]. Although the implications of this loss are not currently understood, they may be related to human health. Human disease and mouse model studies indicate that polymorphisms in *APOC1* are risk factors for more severe forms of Alzheimer's disease[56,57] and for developing atherosclerosis and coronary heart disease[58,59]. As these diseases appear to be unusually common and severe in humans, it is plausible that this gene loss could be a contributing factor[60,61]. Other examples of confirmed pseudogenes are cytidine monophospho-*N*-acetylneuraminic acid hydroxyase (*CMAH*) and myosin heavy chain 16 (*MYH16*). Pseudogenization of *CMAH* was responsible for the inactivation of biosynthesis of the sialic acid *N*-glycolylneuraminic acid; this inactivation led to a radical reconfiguration of human cell surfaces by changing millions of molecules on the surface of human cells, bringing about major consequences for human-specific innate immunity and other systemic roles of sialic acid biology[62,63]. A further example is the pseudogenization of a myosin filament, one of the basic units of muscle, which is specifically expressed in primate jaws and thus may have altered HLS craniofacial musculature and morphology[64].

Polymorphisms
Allelic genetic variations within a species (for example, in amylase, alpha 1A (*AMY1A*)).

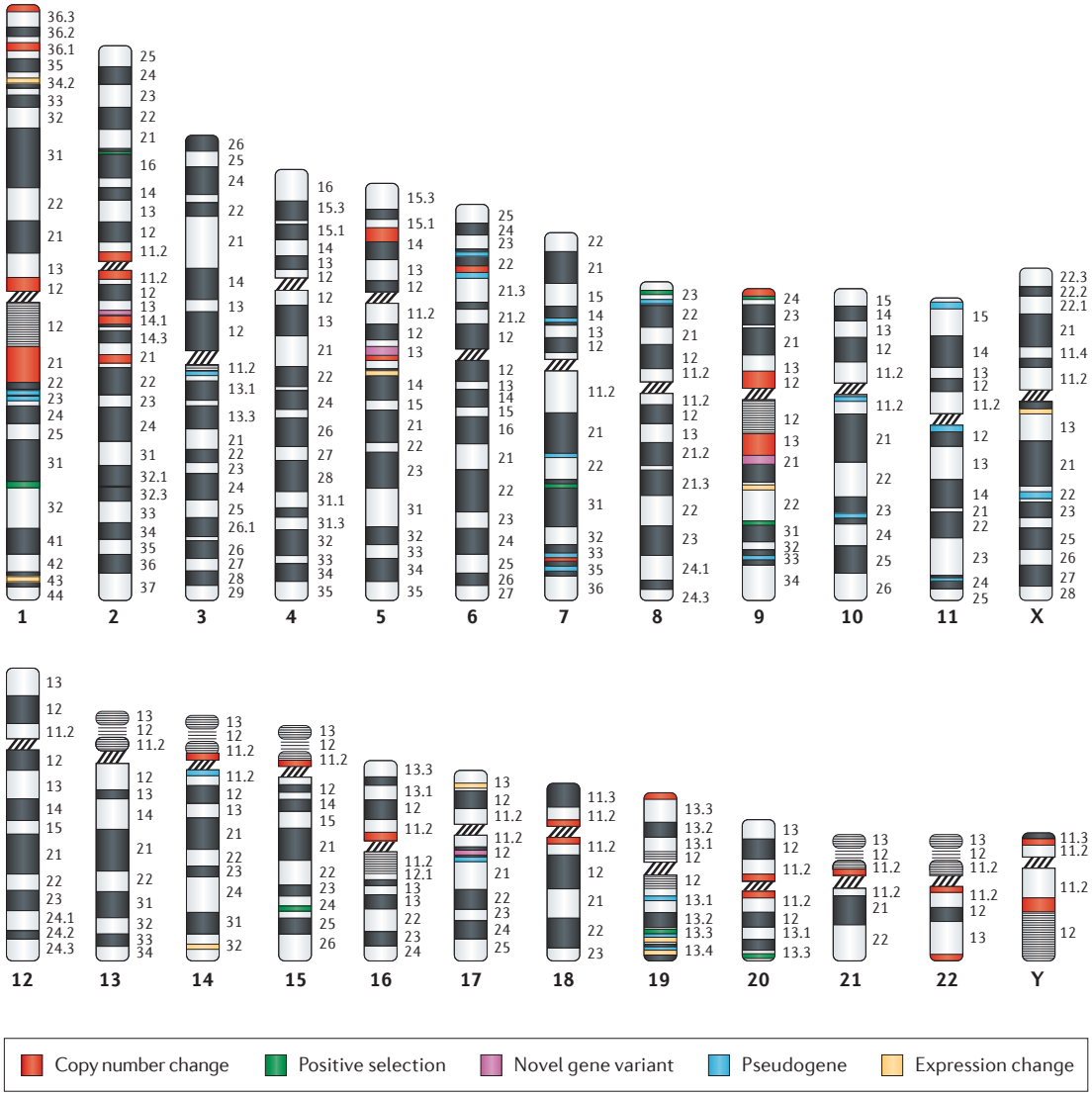


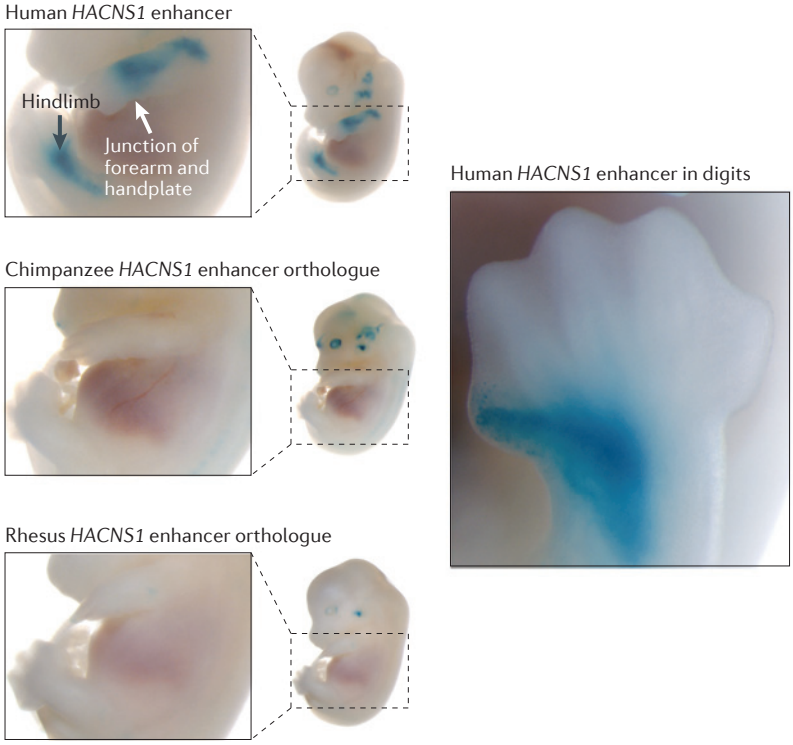
Figure 1 | **Genome positions of human-lineage-specific gene changes.** Human-lineage-specific (HLS) gene changes discussed in this paper are displayed in their corresponding genomic position across the human karyotype. The changes are divided into five categories that correspond to those listed in TABLE 1, and each type is colour coded. It should be noted that many genes have undergone multiple types of HLS changes, and in this case only one type is shown. For visualization purposes, the size of the coloured bands is not drawn to scale.

Pseudogenization is also a common trend seen in HLS multigene families. This mechanism is often found in genes that code for proteins of neural sensation, such as those for olfaction. In the olfactory receptor family, more than 60% of the genes have been rendered nonfunctional by pseudogenization in the human lineage, although some increases in copy number have also been noted, specifically in the olfactory receptor, family 1, subfamily A, member 1 (*OR1A1*) gene[52].

Gene conversion
'Pasting' of identity from one homologous gene to another (for example, in sialic-acid-binding Ig superfamily lectin 11 (*SIGLEC11*)).

Alteration of gene function by gene conversion. Many gene structural changes occur as the result of misalignments during replication. One such example of this is gene conversion, in which a portion of one gene is 'pasted' onto another gene and often occurs between members of genes within the same family. Although gene conversion events usually lead to pseudogenization, there are cases in which the conversion is functionally important. For example, sialic-acid-binding Ig superfamily lectin 11 (*SIGLEC11*), which encodes a member of a family of cell surface membrane proteins involved in modulating immunity, underwent two tandem HLS gene conversion events with an adjacent pseudogene, resulting in it acquiring a novel promoter and amino-terminal protein sequences[65]. This led to a change in the binding specificity of SIGLEC11, allowing recognition of novel ligands and initiation of its expression in microglia, the cells responsible for immune defence and neuroprotection in the central nervous system. *SIGLEC11* expression alleviates neurotoxicity of microglial cells, and this can damage neurons and

Box 4 | **Interpreting human-lineage-specific change**



In spite of the substantial difficulties involved in determining the function of human-lineage-specific (HLS) genetic and genomic changes, a number of encouraging studies have been reported that may serve as models in this challenging arena. One example involved the use of humanized transgenic mice to show that a gain of function HLS-like phenotype was produced by amino acid change in forkhead box P2 (*FOXP2*)[51]. Evidence primarily based on mutations identified in a family with severe speech disabilities has implicated *FOXP2* in human speech production[115]. Further analyses of the gene identified two HLS amino acid substitutions with evidence of positive selection[116,117]. To investigate the phenotypic change resulting from the two amino acid differences, mice with humanized *FOXP2* were generated[51] that showed increased neuronal dendritic length, increased synaptic plasticity and changes in ultrasonic vocalization. These results provide support that the two substitutions in this transcription factor could have affected human speech production capabilities.

Another example is the human accelerated conserved non-coding region 1 (*HACNS1*) enhancer. *HACNS1* underwent accelerated evolution in the human lineage, as shown by a study using a series of expression assays in mouse embryos driven by the 546 bp homologous non-coding sequences from humans, chimpanzees and macaques[2]. The *HACNS1* enhancer region from humans showed expression in the anterior developing forelimb and hindlimb, particularly in the forearm, the handplate, the anterior-most digit and the corresponding structures in the hindlimb (as shown in blue in the figure). Neither the chimpanzee nor the rhesus macaque constructs showed this pattern, suggesting that the region may have had an important role in HLS morphological changes to the hands and feet and making *HACNS1* an important candidate for contributions to human bipedalism and tool making. Additional work on chimeric constructs then narrowed the expression pattern change to 13 divergent bases within an 80 bp region, establishing a narrow window for future investigations. Figure is reproduced, with permission, from REF. 2 © (2008) American Academy for the Advancement of Science.

contribute to neurodegenerative disorders[66]. More recently, a unique expression pattern of *SIGLEC11* and its ligands in human ovaries has demonstrated a possible role in HLS ovarian changes[67].

Alteration of gene function by changes in gene family size. Gene family size can be altered through the addition of new copies of a gene that is already present in the lineage. Such duplicates, when functional, can simply confer an increase in dosage or can diverge to take on potentially new functions. The NBPF gene family encodes DUF1220 protein domains, as discussed above, which are thought to have a role in brain size and cognition[45,46]. This family has undergone both HLS gene copy number expansion, adding an estimated four new human gene copies, and HLS domain copy number hyper-amplification, specifically adding more than 160 copies to the human genome[42]. Another HLS multigene family is the double homeobox (DUX) family, which includes a number of genes involved in transcriptional regulation and embryogenesis. In humans, three of the DUX genes on chromosome Y (*DUXY2–4*) have undergone neofunctionalization through the removal of an ancestral stop codon, although the functional importance of this mutation has yet to be determined[68].

Gene families also undergo inactivation events with relative frequency, although whole-gene deletion is generally rare. A recent example of this kind of loss is *SIGLEC13* (REF. 69). The *SIGLEC13* locus underwent an HLS whole-gene deletion mediated by an *Alu* recombination event[70]. Expression of the chimpanzee form of *SIGLEC13* on monocytes affects inflammatory cytokine secretion and sialic acid binding, potentially enhancing susceptibility to infection by sialylated bacterial pathogens. It is hypothesized that *SIGLEC13* loss may have been selected in relation to the bottleneck at the origin of modern humans, as it improved fitness in infants who would otherwise be susceptible to these bacteria[70].

Copy number change. Although the previously mentioned DUF1220 domain sequences show an extreme HLS copy number increase, many additional HLS copy number changes exist and make up a substantial proportion of the differences between human and great ape genomes[8,10,25]. One example is SLIT–ROBO rho GTPase-activating protein 2 (*SRGAP2*), which has at least one fixed HLS partial duplication[25,71]. *SRGAP2* is a negative regulator of neuronal migration and promotes neurite outgrowth[72]. It is hypothesized that the partially duplicated protein dimerizes with the full-length SRGAP2 protein, which acts as a dominant partial inhibitor and presumably leads to neotenous changes, including increased density of longer neurite spines[72]. Another identified HLS copy number increase involves aquaporin 7 (*AQP7*)[10], of which there are several additional copies in the human genome. *AQP7* is involved in the transport of water and glycerol and is responsible for use of glycerol (that is, energy) from fat cells, especially during fasting and prolonged human exercise[73,74]. *AQP7* amplification is hypothesized to be adaptive for metabolic needs in human endurance running and possibly thermoregulation by increased sweating[10,75]. Such activities have been proposed to be crucial to humans' exceptional persistence-hunting capabilities and to the establishment of humans as diurnal endurance predators[10].

Box 5 | **The Matrix of Comparative Anthropogeny**

The Matrix of Comparative Anthropogeny (MOCA) of the Center for Academic Training and Research in Anthropogeny is a Web-based collection of information comparing humans and our closest evolutionary relatives (namely, chimpanzees, bonobos, gorillas and orangutans — the 'great apes') that has a specific emphasis on uniquely human features. Comparisons of these non-human hominids with humans are difficult, as so little is currently known about their phenotypic features ('phenomes') in contrast to humans[30]. Ethical, fiscal and practical issues also limit the collection of further information about the great apes. MOCA attempts to collect existing information about human-lineage-specific differences from great apes that is currently scattered in the literature. Having such information in one location could lead to new insights and multi-disciplinary interactions, as well as to ethically sound studies that could explain these differences and improve our understanding of uniquely human specializations. It is for this reason that MOCA is called a 'matrix': an arrangement of information from which something else originates, develops or takes form. MOCA is organized by domains, each of which groups topics by areas of interest and scientific discipline. MOCA is a work in progress, and each topic entry will eventually cover existing information about a particular difference (either alleged or documented) between humans and non-human hominids. Topics are also linked across domains: for example, each genetic human-lineage-specific topic is linked to phenotypic traits, including anatomy, physiology, behaviour and even culture to the extent that this is possible.

MOCA is not targeted at experts in specific disciplines but rather aims to communicate basic information to a broad audience of scientists from many backgrounds and to the interested lay public. MOCA includes not only aspects wherein there are known or apparent differences between humans and great apes, but additionally topics for which popular wisdom about claimed or assumed differences is not entirely correct.

The MOCA site is being launched at an early stage so that interested readers with expertise on specific topics can provide feedback. New information and topics will continue to be added as they are identified or discovered.

The creation of* de novo *human genes. Although copy number change is established as a mechanism resulting in new gene function, it acts on pre-existing genes. By contrast, recent evidence suggests that several novel genes appeared in the human genome *de novo* from previously non-coding DNA. Although it was previously thought to be an extremely rare event in genome evolution, *de novo* human gene generation has become a subject of considerable interest and debate. Three published papers have claimed the identification of HLS *de novo* genes[76–78]. None has functional assignments, but this is not surprising given the novelty of the genes identified. However, one gene, chronic lymphocytic leukaemia upregulated 1 (*CLLU1*), is found to be upregulated in patients with a particularly aggressive form of chronic lymphocytic leukaemia[79], lending credence to the claim that the regions identified are actually functional.

The HLS *de novo* gene publications are the subject of much controversy centred around two key issues. The first issue relates to the fundamental definition of what is *de novo*. The Wu *et al.*[78] data set allowed human genes to be considered *de novo* if up to 20% of their coding regions were homologous with a predicted open reading frame in other primates. The Knowles and McLysaght[76] and Li *et al.*[77] papers, however, used stricter criteria. This has generated deliberation as to whether a gene that shares this much of its coding region can truly be called *de novo*. The other key issue relates to changes in gene annotation. The original set of genes identified in Wu *et al.*[78] did not include the three genes identified in the Knowles and McLysaght[76] paper, as the human genome build no longer lists them as annotated genes, casting doubt on whether the genes identified in these studies are real. At present, it appears that *de novo* genes may have contributed HLS genes, but the extent to which this is the case remains to be determined.

Expression changes. Alteration to gene expression is a common mode of evolutionary change and can result from multiple changes at the genetic level, such as changes in regulatory DNA affecting promoters, enhancers and suppressors and dosage changes resulting from CNVs. These types of alterations may change gene amounts, timing or even in what tissues gene expression occurs. Advancements in identifying gene expression changes using RNA-sequencing technologies are the subject of a recent Review by Romero *et al.*[80]; however, we will highlight a few examples.

A 60.7 kb HLS deletion upstream of the androgen receptor (*AR*) gene was identified in a bioinformatics survey of regions that are highly conserved between chimpanzees and macaques; this deletion removed a regulatory region, leading to the loss of expression of *AR* in sensory vibrissae and penile spines[1]. Expression constructs in embryonic mice and in human foreskin fibroblasts showed that the corresponding non-deleted region from chimpanzees controls AR expression. Humans lack sensory vibrissae and penile spines, both of which are found in our closest ape relatives. Such anatomical differences between humans and the great apes lend additional support to the validity of these studies. Another example of a regulatory change is nucleotide changes in the human accelerated conserved non-coding region 1 (*HACNS1*) enhancer that may have led to HLS changes in digit and limb development[2] (BOX 4).

Copy number changes may alter expression through gene dosage changes such as the duplication of protocadherin 11 X-linked (*PCDH11X*) onto the Y chromosome resulting in a Y-linked copy of the gene (namely, *PCDH11Y*)[81,82]. This duplication doubled the gene dosage in humans, as genes present on both the X and Y chromosomes are protected from X-chromosome inactivation[83,84]. This change is hypothesized to contribute to cerebral asymmetry and language development[84], a claim that is corroborated by disease findings of severe language impairment associated with *PCDH11X*[85].

Other important examples of expression change are loss of *SIGLEC5* expression from human T cells, plausibly resulting in T cell hyperactivation[86], and uniquely human *SIGLEC6* expression in the placenta[87], which is hypothesized to be implicated in the human-specific disease pre-eclampsia[88]. The specific genomic alterations that are responsible for these expression changes remain unknown.

HLS changes linked with human disease

Human disease has been, and continues to be, one of the few ways of highlighting the phenotypic implications of many HLS genetic changes[51]. Although the link

Expression pattern change
Change in timing, level and/or location of gene expression (for example, in protocadherin 11 from the X chromosome to the Y chromosome (*PCDH11XY*)).

Neofunctionalization
A process by which a genetic change in an allele produces a novel protein function (for example, in double homeobox (DUX) family members).

De novo human gene
A novel gene arising from formerly non-coding DNA (for example, in chronic lymphocytic leukaemia upregulated 1 (*CLLU1*)).

Human-specific disease
A disease that is present only in the human lineage. A number of diseases are thought to be human-specific (such as Alzheimer's disease and myocardial infarction), but proving that such diseases are not present in other species remains a challenging task.

between human disease and HLS genomic changes has been the subject of recent reviews[4,89], an appreciation of this connection has begun to emerge only in the past decade[90].

Demonstration of causality versus association is difficult, but it is increasingly clear that many regions undergoing HLS change have an important role in human disease (BOX 3). Selection on numerous genes involved in innate immunity improved human resistance to particular diseases but apparently did so at the cost of other human-specific impairments, such as an increased propensity for autoimmune disorders, including atopic diseases and allergies[91]. Increased immune response may have arisen during the unique changes to HLS pathogen regimes associated with use of home bases, scavenging or hunting of different prey and more extensive intergroup contacts. An example of association between improved immunity and autoimmune disorders is seen in the case of the SIGLEC gene family. The levels of many SIGLEC genes are decreased in humans; as most of these have an inhibitory effect on lymphocyte activation, lower SIGLEC expression leads to increased lymphocyte reactivity[63,92]. This hyper-reactivity, although potentially protective against infection, may predispose to autoimmunity. In addition, a number of HLS changes linked to human-specific cognitive abilities are associated with the severity of Alzheimer's disease and dementias[56,93]. For example, *FLJ33706* (REF. 77) — a novel human gene that is highly expressed in the cortex, cerebellum and midbrain — has been reported to show increased expression in Alzheimer's brain specimens. The asp (abnormal spindle) homologue, microcephaly-associated (*ASPM*) gene[94,95] reportedly underwent accelerated evolution in the human lineage, although the certainty of this is contested in recent literature[96]. Thus there are multiple cases in which a region that was found to have undergone HLS selection is also associated with one or more major disease phenotypes.

Beyond natural selection, unstable genomic architecture is another driving force in human genetic novelty associated with human disease. Certain regions of the genome that are complex and repeat-rich often act as gene nurseries[8,10] and might therefore be evolutionarily advantageous to maintain in the population. However, the instability associated with the architectures of the same genomic regions also makes them prone to deleterious copy number gains and losses. Such regions have been linked to numerous genomic disorders, including several neuropsychiatric and neurodevelopmental diseases[97,98] (BOX 3). For example, the 1q21.1 region of the human genome has undergone multiple HLS expansions, including those involving the aforementioned DUF1220 protein domain[8,42] and *SRGAP2* (REF. 71), two candidates that could be involved in human brain morphological changes. CNVs in the region have also been implicated in a striking number of recurrent diseases[45,46]. Therefore, as the study of disease continues to be a resource for understanding the function of HLS genomic change, the identification of HLS changes provides a source of candidate genetic changes contributing to human disease.

Gene nurseries
Dynamic regions of the genome that are capable of undergoing rapid evolutionary change owing to a duplication-prone genome architecture and are therefore frequent sites for the production of novel genes by gene duplication.

Hydatidiform mole
An abnormal form of pregnancy in which a non-viable egg, probably the result of an egg missing a nucleus, is fertilized and becomes a mass on the uterine wall. The resultant growing tissue is haploid in nature owing to it having only a paternal genetic contribution.

Conclusions and prospects

With continued improvements in DNA-sequencing technologies[99], the number of new genome sequences available will probably continue to grow exponentially, and this can be expected to clarify further which changes are truly HLS. However, unless there is a focus on developing technologies to accurately sequence and to assemble complex, duplication-rich genomic regions, such regions will continue to be woefully under-examined. For example, the last two great apes to be sequenced, bonobos and gorillas, were both sequenced using whole-genome shotgun methods that relied heavily on NGS technologies[100,101]. These methods generated poor coverage and assembly of duplication-rich regions and therefore are deficient in sequences that are particularly relevant to the identification of genes showing HLS changes in copy number[102].

Obtaining accurate sequence from complex genomic regions is important for several reasons: they have been linked to numerous diseases and disorders[24,90]; they are often sites of rapid evolutionary change[8]; and they have been proposed as candidates to harbour the 'missing genetic heritability' that has eluded many genome-wide disease gene investigations[34]. Finally, the importance of examining complex genomic regions has been recently borne out by a study of the human-specific *SRGAP2* gene duplications. These extra copies were discovered only through the development of a new genome assembly that used a haploid human genome resource (namely, hydatidiform mole) and long-read sequencing to accurately sequence repetitive content[71].

Another area of improvement needed is exemplified in *de novo* sequencing efforts[76–78]. The inability to verify the results of prior studies owing to differences between genome builds only epitomizes the need for solid manual annotation of genes in all genomes. Inconsistencies in the annotation of the human genome — which has the most extensive annotation work of all of the mammalian genomes — can lead to reduced confidence in gene calls in other primate genomes. Conclusive identification of a gene's HLS status requires the availability of accurate genome assemblies from multiple individuals in a lineage combined with comparison to multiple individuals in multiple primate outgroups.

Despite the difficulties involved, it is reasonable to expect that the process of linking human-specific gene and genomic features to important human-specific traits will accelerate over the next several years. However, challenges in the field continue, such as the limited data existing on great ape phenotypes compared with human phenotypes[103] and the major impact of gene–culture interactions in generating the human phenotype[10]. While keeping ethical considerations foremost in mind, it is important to gather as much novel information on the phenotypes of these species as possible; this objective would be much advanced by providing proper medical care to captive great apes in the United States and in ape sanctuaries throughout Africa[104]. Improving transgenic technology in mice potentially affords more functional insights than disease associations alone and as such may prove to be a valuable strategy for verifying many HLS traits.

It may be anticipated that the identification of key genes and genomic sequences that are responsible for the unusual capacities of human cognition and physiology may lay the groundwork for new fields of study. For example, a new field of neuroscience, evolutionary neurogenomics, may emerge from the application of genomic technologies to the study of human brain evolution[4]. By providing a novel, biologically unbiased entry point for the study of the brain, such a discipline may uncover new insights regarding human cognitive function and dysfunction (BOX 3). Many of the genetic changes underlying the dramatic change in cognitive capacities in the human lineage are likely to be a part of evolutionary trade-offs, where increased mental capacities have come at a cost to other organismal functions. Thus, genotype-to-phenotype associations are likely to have medical relevance as well, showing how evolutionary selective pressures may have produced side effects, such as unstable genomic regions enriched for disease-associated variants (for example, 1q21 and 5q13)[8,45,46,102,105]. Finally, we can anticipate that using genome-based data to help decipher what has made the human species unique will become an increasingly powerful tool for understanding our evolutionary origins (that is, our anthropogeny) and the human condition.

1. McLean, C. Y. *et al.* Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–219 (2011).
2. Prabhakar, S. *et al.* Human-specific gain of function in a developmental enhancer. *Science* **321**, 1346–1350 (2008).
This paper demonstrates how a conserved non-coding sequence (namely, *HACNS1*), which evolved extremely rapidly in humans, may underlie human-specific aspects of limb development.
3. Varki, A., Geschwind, D. & Eichler, E. Human uniqueness: genome interactions with environment, behaviour and culture. *Nature Rev. Genet.* **9**, 749–763 (2008).
This earlier Review on the same subject cautions against a 'gene-centric' view of human evolution and suggests that some aspects of human genome evolution may be due to relaxed selection, resulting from masking by behaviour and culture.
4. Sikela, J. M. The jewels of our genome: the search for the genomic changes underlying the evolutionarily unique capacities of the human brain. *PLoS Genet.* **2**, e80 (2006).
5. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
6. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
7. Enard, W. & Paabo, S. Comparative primate genomics. *Annu. Rev. Genom. Hum. Genet.* **5**, 351–378 (2004).
8. Fortna, A. *et al.* Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol.* **2**, e207 (2004).
This was the first genome-wide and first gene-based array CGH study of lineage-specific gene copy number gain and loss among human and great ape lineages. One hundred and forty genes were identified that showed HLS changes in copy number, including the *MGC8902* gene that encodes DUF1220 protein domains.
9. Yunis, J. J. & Prakash, O. The origin of man: a chromosomal pictorial legacy. *Science* **215**, 1525–1530 (1982).
10. Dumas, L. *et al.* Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res.* **17**, 1266–1277 (2007).
This was the most extensive array CGH investigation to date of gene-based copy number change across primate species. Many of the >4,000 genes identified that showed lineage-specific changes in copy number are excellent candidates for underlying lineage-specific traits among these species.
11. Ventura, M. *et al.* The evolution of African great ape subtelomeric heterochromatin and the fusion of human chromosome 2. *Genome Res.* **22**, 1036–1049 (2012).
12. Bhatt, B., Burns, J., Flannery, D. & McGee, J. Direct visualization of single copy genes on banded metaphase chromosomes by nonisotopic *in situ* hybridization. *Nucleic Acids Res.* **16**, 3951–3961 (1988).
13. Jauch, A. *et al.* Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting. *Proc. Natl Acad. Sci. USA* **89**, 8611–8615 (1992).
14. Wilson, G. M. *et al.* Identification of full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla. *Genome Res.* **16**, 173–181 (2006).
15. Goidts, V. *et al.* Identification of large-scale human-specific copy number differences by inter-species array comparative genomic hybridization. *Hum. Genet.* **119**, 185–198 (2006).
16. Linardopoulou, E. V. *et al.* Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437**, 94–100 (2005).
17. Mikkelsen, T. S. *et al.* Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
18. Clark, A. G., *et al.* Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
19. Berglund, J., Pollard, K. S. & Webster, M. T. Hotspots of biased nucleotide substitutions in human genes. *PLoS Biol.* **7**, e26 (2009).
20. Grossman, S. R. *et al.* A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* **327**, 883–886 (2010).
21. Britten, R. J. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl Acad. Sci. USA* **99**, 13633–13635 (2002).
22. Chen, F. C., Chen, C. J., Li, W. H. & Chuang, T. J. Human-specific insertions and deletions inferred from mammalian genome sequences. *Genome Res.* **17**, 16–22 (2007).
23. Marques-Bonet, T. *et al.* A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**, 877–881 (2009).
24. Alkan, C. *et al.* Personalized copy number and segmental duplication maps using next-generation sequencing. *Nature Genet.* **41**, 1061–1067 (2009).
25. Sudmant, P. H. *et al.* Diversity of human copy number variation and multicopy genes. *Science* **330**, 641–646 (2010).
26. Treangen, T. J. & Salzberg, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Rev. Genet.* **13**, 36–46 (2011).
27. Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nature Rev. Genet.* **12**, 363–376 (2011).
28. Khodosevich, K., Lebedev, Y. & Sverdlov, E. Endogenous retroviruses and human evolution. *Comp. Funct. Genomics.* **3**, 494–498 (2002).
29. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Rev. Genet.* **10**, 691–703 (2009).
30. Lee, J. *et al.* Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. *Gene* **390**, 18–27 (2007).
31. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl Acad. Sci. USA* **100**, 5280–5285 (2003).
32. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature.* **460**, 1127–1131 (2009).
33. Singer, T. *et al.* LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes. *Trends Neurosci.* **33**, 345–354 (2010).
34. Eichler, E. E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Rev. Genet.* **11**, 446–450 (2010).
35. Lin, L., *et al.* Evolution of alternative splicing in primate brain transcriptomes. *Hum. Mol. Genet.* **19**, 2958–2973 (2010).
36. Blekhman, R., Marioni, J. C., Zumbo, P., Stephens, M. & Gilad, Y. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* **20**, 180–189 (2010).
37. Kim, D. S. & Hahn, Y. Identification of human-specific transcript variants induced by DNA insertions in the human genome. *Bioinformatics* **27**, 14–21 (2011).
38. Huh, J. W. *et al.* Gain of new exons and promoters by lineage-specific transposable elements-integration and conservation event on *CHRM3* gene. *Mol. Cells* **28**, 111–117 (2009).
39. Ohno, S. *Evolution by Gene Duplication* (Springer, 1970).
This classic text provides some of the first arguments making the case that gene duplication is a major mechanism underlying genome evolution.
40. Li, W. H. *Molecular Evolution* (Sinauer Associates, 1997).
41. Popesco, M. C. *et al.* Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* **313**, 1304–1307 (2006).
This was the first report to describe the striking HLS increase in copy number of DUF1220 protein domains.
42. O'Bleness, M. S. *et al.* Evolutionary history and genome organization of DUF1220 protein domains. *G3* **2**, 977–986 (2012).
43. Vandepoele, K., van Roy, N., Staes, K., Speleman, F. & van Roy, F. A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. *Mol. Biol. Evol.* **22**, 2265–2274 (2005).
44. Andries, V., Vandepoele, L. & van Roy, F. in *Neuroblastoma — Present and Future* Ch. 9 (ed. Shimada, H.) (InTech Publishing, 2012).
45. Dumas, L. & Sikela, J. M. DUF1220 domains, cognitive disease, and human brain evolution. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 375–382 (2009).
46. Dumas, L. *et al.* DUF1220-domain copy number implicated in human brain-size pathology and evolution. *Am. J. Hum. Genet.* **91**, 444–454 (2012).
This paper provides support for the view that DUF1220 domain copy number (specifically, DUF1220 domain dosage) is a key driver of primate brain size and may be largely responsible for the dramatic evolutionary expansion in brain size that occurred in the human lineage.
47. Rizzuto, R., Nakase, H., Zeviani, M., DiMauro, S. & Schon, E. A. Subunit Va of human and bovine cytochrome *c* oxidase is highly conserved. *Gene* **69**, 245–256 (1988).
48. Uddin, M. *et al.* Molecular evolution of the cytochrome *c* oxidase subunit 5A gene in primates. *BMC Evol. Biol.* **8**, 8 (2008).
49. Cioffi, F., Lanni, A. & Goglia, F. Thyroid hormones, mitochondrial bioenergetics and lipid handling. *Curr. Opin. Endocrinol. Diabetes Obes.* **17**, 402–407 (2010).
50. Arnold, S., Goglia, F. & Kadenbach, B. 3,5-diiodothyronine binds to subunit Va of cytochrome-*c* oxidase and abolishes the allosteric inhibition of respiration by ATP. *Eur. J. Biochem.* **252**, 325–330 (1998).
51. Enard, W. *et al.* A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* **137**, 961–971 (2009).

52. Kim, H. L., Igawa, T., Kawashima, A., Satta, Y. & Takahata, N. Divergence, demography and gene loss along the human lineage. *Phil. Trans. R. Soc. B* **365**, 2451–2457 (2010).
53. Olson, M. V. When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* **64**, 18–23 (1999).
54. Pastorcic, M., Birnbaum, S. & Hixson, J. E. Baboon apolipoprotein C-I: cDNA and gene structure and evolution. *Genomics* **13**, 368–374 (1992).
55. Puppione, D. L. *et al.* Detection of two distinct forms of apoC-I in great apes. *Comp. Biochem. Physiol.* **5**, 73–79 (2010).
56. Lucatelli, J. F. *et al.* Genetic influences on Alzheimer's disease: evidence of interactions between the genes *APOE*, *APOC1* and *ACE* in a sample population from the South of Brazil. *Neurochem. Res.* **36**, 1533–1539 (2011).
57. Berbée, J. F. *et al.* Apolipoprotein CI knock-out mice display impaired memory functions. *J. Alzheimers Dis.* **23**, 737–747 (2011).
58. Hansen, J. B. *et al.* The apolipoprotein C-I content of very-low-density lipoproteins is associated with fasting triglycerides, postprandial lipemia, and carotid atherosclerosis. *J. Lipids* **2011**, 271062 (2011).
59. Grallert, H. *et al.* Eight genetic loci associated with variation in lipoprotein-associated phospholipase A2 mass and activity and coronary heart disease: meta-analysis of genome-wide association studies from five community-based studies. *Eur. Heart J.* **33**, 238–251 (2012).
60. Varki, N. M. *et al.* Heart disease is common in humans and chimpanzees, but is caused by different pathological processes. *Evol. Appl.* **2**, 101–112 (2009).
61. Varki, N. M., Strobert, E., Dick E. J. Jr, Benirschke, K. & Varki, A. Biomedical differences between humans and nonhuman hominids; potential roles for uniquely human aspects of sialic acid biology. *Annu. Rev. Pathol. Mech.* **6**, 365–393 (2011).
This review emphasizes that many human diseases are uniquely human and relates some of the differences in evidence that genes involved in the biology of sialic acids constitute a 'hotspot' in human evolution.
62. Hayakawa, T., Aki, I., Varki, A., Satta, Y. & Takahata, N. Fixation of the human-specific CMP-*N*-acetylneuraminic acid hydroxylase pseudogene and implications of haplotype diversity for human evolution. *Genetics* **172**, 1139–1146 (2006).
63. Varki, A. Uniquely human evolution of sialic acid genetics and biology. *Proc. Natl Acad. Sci. USA* **107**, 8939–8946 (2010).
64. Stedmann, H. H. *et al.* Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature* **428**, 415–418 (2004).
65. Wang, X. *et al.* Evolution of *Siglec-11* and *Siglec-16* genes in hominins. *Mol. Biol. Evol.* **29**, 2073–2086 (2012).
The first examples of genes inactivated in relation to the timing of origin of modern humans are presented in this paper. The data suggest that infectious agents may have played a part in selection.
66. Wang, Y. & Neumann, H. Alleviation of neurotoxicity by microglial human Siglec-11. *J. Neurosci.* **30**, 3482–3488 (2010).
67. Wang, X. *et al.* Expression of Siglec-11 by human and chimpanzee ovarian stromal cells, with uniquely human ligands: implications for human ovarian physiology and pathology. *Glycobiology* **21**, 1038–1048 (2011).
68. Schmidt, J., Kirsch, S., Rappoid, G. A. & Schempp, W. Complex evolution of a Y-chromosomal double homeobox 4 (DUX4)-related gene family in hominoids. *PLoS ONE* **4**, e5288 (2009).
69. Crocker, P. R. Siglecs in innate immunity. *Curr. Opin. Pharmacol.* **5**, 431–437 (2005).
70. Wang, X. *et al.* Specific inactivation of two immunomodulatory SIGLEC genes during human evolution. *Proc. Natl Acad. Sci. USA* **109**, 9935–9940 (2012).
71. Dennis, M. Y. *et al.* Evolution of human-specific neural *SRGAP2* genes by incomplete segmental duplication. *Cell* **149**, 912–922 (2012).
72. Charrier, C. *et al.* Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. *Cell* **149**, 923–935 (2012).
73. Walker, C. G., Holness, M. J., Gibbons, G. F. & Sugden, M. C. Fasting-induced increases in aquaporin 7 and adipose triglyceride lipase mRNA expression in adipose tissue are attenuated by peroxisome proliferator-activated receptor alpha deficiency. *Int. J. Obes.* **31**, 1165–1171 (2007).
74. Kondo, H. *et al.* Human aquaporin adipose (*AQPap*) gene. Genomic structure, promoter analysis and functional mutation. *Eur. J. Biochem.* **269**, 1814–1826 (2002).
75. Bramble, D. M. & Lieberman, D. E. Endurance running and the evolution of *Homo*. *Nature* **432**, 345–352 (2004).
This provides a well-reasoned argument based on anatomical and physiological evidence that endurance running had a major role in human evolution at the origin of the genus *Homo*.
76. Knowles, D. G. & McLysaght, A. Recent *de novo* origin of human protein-coding genes. *Genome Res.* **19**, 1752–1759 (2009).
77. Li, C. Y. *et al.* A human-specific *de novo* protein-coding gene associated with human brain functions. *PLoS Comput. Biol.* **6**, e1000734 (2010).
78. Wu, D. D., Irwin, D. M. & Zhang, Y. P. *De novo* origin of human protein-coding genes. *PLoS Genet.* **7**, e1002379 (2011).
79. Buhl, A. M. *et al.* Identification of a gene on chromosome 12q22 uniquely overexpressed in chronic lymphocytic leukemia. *Blood* **107**, 2904–2911 (2006).
80. Romero, I. G., Ruvinsky, I. & Gilad, Y. Comparative studies of gene expression and the evolution of gene regulation. *Nature Rev. Genet.* **13**, 505–516 (2012).
81. Ross, N. L. J. *et al.* Methylation of two *Homo sapiens*-specific X-Y homologous genes in Klinefelter's syndrome (XXY). *Am. J. Med. Genet. B* **141**, 544–548 (2006).
82. Yoshida, K. & Sugano, S. Identification of a novel protocadherin gene (*PCDH11*) on the human XY homology region in Xq21.3. *Genomics* **62**, 540–543 (1999).
83. Kalmady, S. V. & Venkatasubramanian, G. Evidence for positive selection on protocadherin Y gene in *Homo sapiens*: implications for schizophrenia. *Schizophr. Res.* **108**, 299–300 (2009).
84. Crow, T. J. Handedness, language lateralisation and anatomical asymmetry: relevance of protocadherin XY to hominid speciation and the aetiology of psychosis. Point of view. *Br. J. Psychiatry* **181**, 295–297 (2002).
85. Speevak, M. D. & Farrell, S. A. Non-syndromic language delay in a child with disruption in the Protocadherin11X/Y gene pair. *Am. J. Med. Genet. B* **156**, 484–489 (2011).
86. Nguyen, D. H., Hurtado-Ziola, N., Gagneux, P. & Varki, A. Loss of Siglec expression on T lymphocytes during human evolution. *Proc. Natl Acad. Sci. USA* **103**, 7765–7770 (2006).
87. Brinkman-Van der Linden, E. C. *et al.* Human-specific expression of Siglec-6 in the placenta. *Glycobiology* **17**, 922–931 (2007).
88. Winn, V. D. *et al.* Severe preeclampsia-related changes in gene expression at the maternal-fetal interface include sialic acid-binding immunoglobulin-like lectin-6 and pappalysin-2. *Endocrinology* **150**, 452–462 (2009).
89. Cooper, D. N. & Kehrer-Sawatzki, H. Exploring the potential relevance of human-specific genes to complex disease. *Hum. Gen.* **5**, 99–107 (2011).
90. Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82 (2002).
This is an important paper that discusses the connection between complex genomic architecture and recurrent disease. The authors put forth the idea that these regions may have an important role in primate speciation.
91. Olson, M. V. & Varki, A. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nature Rev. Genet.* **4**, 20–28 (2003).
92. Soto, P. C., Stein, L. L., Hurtado-Ziola, N., Hedrick, S. M. & Varki, A. Relative over-reactivity of human versus chimpanzee lymphocytes: implications for the human diseases associated with immune activation. *J. Immunol.* **184**, 4185–4195 (2010).
93. Carrasquillo, M. M. *et al.* Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nature Genet.* **41**, 192–198 (2009).
94. Evans, P. D. *et al.* Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. *Hum. Mol. Genet.* **13**, 489–494 (2004).
95. Rimol, L. M. *et al.* Sex-dependent association of common variants of microcephaly genes with brain structure. *Proc. Natl Acad. Sci. USA* **107**, 384–388 (2010).
96. Mekel-Bobrov, N. *et al.* The ongoing adaptive evolution of ASPM and microcephalin is not explained by increased intelligence. *Hum. Mol. Genet.* **16**, 600–608 (2007).
97. Brunetti-Pierri, N. *et al.* Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nature Genet.* **40**, 1466–1471 (2008).
98. Morrow, E. M. Genomic copy number variation in disorders of cognitive development. *J. Am. Acad. Child. Adolesc. Psychiatry* **49**, 1091–1104 (2010).
99. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
100. Scally, A., *et al.* Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169–175 (2012).
101. Prüfer, K., *et al.* The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**, 527–531 (2012).
102. Mefford, H. C. & Eichler, E. E. Duplication hotspots, rare genomic disorders, and common disease. *Curr. Opin. Genet. Dev.* **19**, 196–204 (2009).
103. McConkey, E. H. & Varki, A. Genomics. Thoughts on the future of great ape research. *Science* **309**, 1499–1501 (2005).
104. Altevogt, B. M., Pankevich, D. E., Pope, A. M. & Kahn, J. P. Research agenda. Guiding limited use of chimpanzees in research. *Science* **335**, 41–42 (2012).
105. Khaitovich, P. *et al.* Metabolic changes in schizophrenia and human brain evolution. *Genome Biol.* **9**, R124 (2008).
106. Gamble, C., Davies, W., Pettitt, P. & Richards, M. Climate change and evolving human diversity in Europe during the last glacial. *Phil. Trans. R. Soc. Lond. B* **359**, 243–253 (2004).
107. Pfefferle, A. D. *et al.* Comparative expression analysis of the phosphocreatine circuit in extant primates: implications for human brain evolution. *J. Hum. Evol.* **60**, 205–212 (2011).
108. Martin, R. D. The evolution of human reproduction: a primatological perspective. *Am. J. Phys. Anthropol. Suppl.* **45**, 59–84 (2007).
109. Fooladi, M. M. The healing effects of crying. *Holist. Nurs. Pract.* **19**, 248 (2005).
110. Liu, X. *et al.* Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. *Genome Res.* **22**, 611–622 (2012).
111. Dorus, S. *et al.* Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119**, 1027–1040 (2004).
112. Chen, J. M., Cooper, D. N., Chuzhanova, N., Férec, C. & Patrinos, G. P. Gene conversion: mechanisms, evolution and human disease. *Nature Rev. Genet.* **8**, 762–775 (2007).
113. Pollard, K. S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).
Using a genome-wide comparison of human and non-human genome sequences, this study identified a dramatically changing microRNA-encoding gene, highly accelerated region 1 forward (*HAR1F*), which showed a highly accelerated HLS change in sequence and is highly expressed in the human fetal brain.
114. Doggett, N. A. *et al.* A 360-kb interchromosomal duplication of the human *HYDIN* locus. *Genomics* **88**, 762–771 (2006).
115. Lai, C. S., Gerrelli, D., Monaco, A. P., Fisher, S. E. & Copp, A. J. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* **126**, 2455–2462 (2003).
116. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
This provides the first evidence that the coding sequence of the *FOXP2* gene, which is mutated in a family with a speech disorder, underwent human-specific alterations consistent with positive selection.

117. Zhang, J., Webb, D. M. & Podlaha, O. Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. *Genetics* **162**, 1825–1835 (2002).
118. Evans, P. D., Vallender, E. J. & Lahn, B. T. Molecular evolution of the brain size regulator genes *CDK5RAP2* and *CENPJ*. *Gene* **375**, 75–79 (2006).
119. Araud, T. *et al.* The chimeric gene *CHRFAM7A*, a partial duplication of the *CHRNA7* gene, is a dominant negative regulator of α7*nAChR function. *Biochem. Pharmacol.* **82**, 904–914 (2011).
120. Lussiana, C. *et al.* Mutations and polymorphisms of the FSH receptor (*FSHR*) gene: clinical implications in female fecundity and molecular biology of FSHR protein and gene. *Obstet. Gynecol. Surv.* **63**, 785–795 (2008).
121. Plunkett, J. *et al.* An evolutionary genomic approach to identify genes involved in human birth timing. *PLoS Genet.* **7**, e1001365 (2011).
122. Evans, P. D., Anderson, J. R., Vallender, E. J., Choi, S. S. & Lahn, B. T. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. *Hum. Mol. Genet.* **13**, 1139–1145 (2004).
123. Ciesek, S. *et al.* Impact of intra- and interspecies variation of occludin on its function as coreceptor for authentic hepatitis C virus particles. *J. Virol.* **85**, 7613–7621 (2011).
124. Shin, E. Y. *et al.* Phosphorylation of p85 beta PIX, a Rac/Cdc42-specific guanine nucleotide exchange factor, via the Ras/ERK/PAK2 pathway is required for basic fibroblast growth factor-induced neurite outgrowth. *J. Biol. Chem.* **277**, 44417–44430 (2002).
125. Majava, V. *et al.* Structural and functional characterization of human peripheral nervous system myelin protein P2. *PLoS ONE* **5**, e10300 (2010).
126. Hennah, W. & Porteous, D. The DISC1 pathway modulates expression of neurodevelopmental, synaptogenic and sensory perception genes. *PLoS ONE* **4**, e4906 (2009).
127. Saus, E. *et al.* Comprehensive copy number variant (CNV) analysis of neuronal pathways genes in psychiatric disorders identifies rare variants within patients. *J. Psychiatr. Res.* **44**, 971–978 (2010).
128. Kouprina, N. *et al.* The SPANX gene family of cancer-testis specific antigens: rapid evolution, an unusual case of positive selection and amplification in African Great Apes and hominids. *Proc. Natl Acad. Sci. USA* **101**, 3077–3082 (2004).
129. Westbrook, V. A. *et al.* Hominoid-specific *SPANXA/D* genes demonstrate differential expression in individuals and protein localization to a distinct nuclear envelope domain during spermatid morphogenesis. *Mol. Hum. Reprod.* **12**, 703–716 (2006).

Acknowledgements
We would like to thank S. O'Bleness for editorial comments, M. Dickens for graphics assistance and J. Noonan for access to published images. We also thank the many student and faculty contributors to the Matrix of Comparative Anthropogeny (MOCA) website. Work in our laboratories has been supported by the US National Institutes of Health and by the Mathers Foundation of New York, which also supports the MOCA website.

Competing interests statement
The authors declare competing financial interests: see Web version for details.

FURTHER INFORMATION
Ajit Varki's homepage: http://cmm.ucsd.edu/varki
James M. Sikela's homepage: http://pharmacology.ucdenver.edu/faculty/sikela/sub_sikela/index_sikela.html
Center for Academic Training and Research in Anthropogeny (CARTA): http://carta.anthropogeny.org
Matrix of Comparative Anthropogeny (MOCA): http://carta.anthropogeny.org/moca/about
OMIM — Online Mendelian Inheritance in Man: http://omim.org

ALL LINKS ARE ACTIVE IN THE ONLINE PDF