

# Comparing the human and chimpanzee genomes: Searching for needles in a haystack

Ajit Varki<sup>1</sup> and Tasha K. Altheide

*Glycobiology Research and Training Center, Departments of Medicine and Cellular & Molecular Medicine, University of California at San Diego, La Jolla, California 92093, USA*

The chimpanzee genome sequence is a long-awaited milestone, providing opportunities to explore primate evolution and genetic contributions to human physiology and disease. Humans and chimpanzees shared a common ancestor ~5–7 million years ago (Mya). The difference between the two genomes is actually not ~1%, but ~4%—comprising ~35 million single nucleotide differences and ~90 Mb of insertions and deletions. The challenge is to identify the many evolutionarily, physiologically, and biomedically important differences scattered throughout these genomes while integrating these data with emerging knowledge about the corresponding “phenomes” and the relevant environmental influences. It is logical to tackle the genetic aspects via both genome-wide analyses and candidate gene studies. Genome-wide surveys could eliminate the majority of genomic sequence differences from consideration, while simultaneously identifying potential targets of opportunity. Meanwhile, candidate gene approaches can be based on such genomic surveys, on genes that may contribute to known differences in phenotypes or disease incidence/severity, or on mutations in the human population that impact unique aspects of the human condition. These two approaches will intersect at many levels and should be considered complementary. We also cite some known genetic differences between humans and great apes, realizing that these likely represent only the tip of the iceberg.

Humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*) last shared a common ancestor ~5–7 million years ago (Mya) (Chen and Li 2001; Brunet et al. 2002). What makes humans different from their closest evolutionary relatives, and how, why, and when did these changes occur? These are fascinating questions, and a major challenge is to explain how genomic differences contributed to this process (Goodman 1999; Gagneux and Varki 2001; Klein and Takahata 2002; Carroll 2003; Olson and Varki 2003; Enard and Pääbo 2004; Gagneux 2004; Ruvolo 2004; Goodman et al. 2005; Li and Saunders 2005; McConkey and Varki 2005). Most genome projects focus on elucidating the sequence and structure of a species' genome and then identifying conserved functionally important genes and genomic elements. The finished human genome (International Human Genome Sequencing Consortium 2004) provides such a catalog of genomic features that ultimately interact with the environment to determine our biology, physiology, and disease susceptibility. Completion of the draft chimpanzee genome sequence (The Chimpanzee Sequencing and Analysis Consortium 2005) provides a genome-wide comparative catalog that can be used to identify genes or genomic regions underlying the many features that distinguish humans and chimpanzees.

As humans, we have an inherent interest in understanding and improving the human condition. We also believe that we have many characteristics that are uniquely human. Table 1 lists some of the definite and possible phenotypic traits that appear to differentiate us from chimpanzees and other “great apes”<sup>2</sup>. For the most part, we do not know which genetic features interact with the environment to generate these differences between the “phenomes”<sup>3</sup> of our two species. The chimpanzee has also long been seen as a model for human diseases because of its close

evolutionary relationship. This is indeed the case for a few disorders. Nevertheless, it is a striking paradox that chimpanzees are in fact not good models for many major human diseases/conditions (see Table 2) (Varki 2000; Olson and Varki 2003). In retrospect, this should not be too surprising. After all, at least some major diseases of a species are likely related to (mal)adaptations during the recent evolutionary past of that species (Nesse and Williams 1995). Thus, comparisons with the chimpanzee genome could shed important light on the uniquely human pathogenic mechanisms of serious diseases. This, in turn, could point to novel approaches toward prevention or treatment. Opportunities to address broader questions in evolutionary biology also arise, i.e., to examine the evolutionary forces that underlie recent speciation and phenotypic divergence between two closely related mammalian taxa, as well as the mechanisms by which evolutionary novelties are generated. It is this intertwining of anthropogeny (the study of human origins), biomedical interests, and general evolutionary principles that make the chimpanzee genome such an invaluable resource.

Here, we briefly mention some of the initial findings from sequencing the chimpanzee genome, and describe how these data might be used to address some of the above questions. This pursuit requires the involvement and cooperation of scientists from a wide variety of fields, far beyond the scope of genomics. Thus, our comments are focused not so much on genomics per se, but are rather addressed to the broader scientific community of “genome users.”

<sup>2</sup>The term “great apes” is used here in the now colloquial sense, as genomic information no longer supports this species grouping (Goodman 1999). Under the currently more common classification, these species are now grouped together with humans in the family Hominidae.

<sup>3</sup>The term “phenome” has been used in multiple publications (e.g., Mahner and Kary 1997; Varki et al. 1998; Paigen and Eppig 2000; Nevo 2001; Walhout et al. 2002; Freimer and Sabatti 2003), but still lacks an accepted definition. Discussions with researchers who have used the term suggest the following definition: “The body of information describing an organism's phenotypes, under the influences of genetic and environmental factors.”

<sup>1</sup>Corresponding author.

E-mail [avarki@ucsd.edu](mailto:avarki@ucsd.edu); fax (858) 534-5611.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.3737405>.

**Table 1.** Some phenotypic traits of humans for comparison with those of great apes<sup>a</sup>

<p><b>LIFE HISTORY</b> Secondary Altriciality Helplessness of the Newborn Prolonged Helplessness of Young Extended Care of Young Childhood Adolescence Age at First Reproduction Longevity</p> <p><b>REPRODUCTIVITY BIOLOGY</b> Concealed Ovulation Virgin Breast Development Female Pituitary Menopause Placentophagy Female Labia Majora Vaginal Hymen Baculum (Penis Bone) Sperm Count Copulatory Plug</p> <p><b>EMBRYOLOGY</b> Early Fetal Wastage/Aneuploidy Hydatiform Molar Pregnancy Umbilical Cord Length</p> <p><b>PREGNANCY/PARTURITION</b> Cephalo-pelvic Disproportion Duration of Labor Maternal Mortality in Childbirth Pain During Childbirth Need for Assistance with Childbirth Neonatal Cephalhematoma</p> <p><b>POSTNATAL DEVELOPMENT</b> Late Closure of Cranial Sutures Duration of Infant Arousal Inconsolable Infant Crying Infant-Caregiver Attunement Maternal-Infant Eye-To-Eye Gaze</p> <p><b>ANATOMY</b> Sagittal Crest of Skull Brow Ridge Protuberantia Menti (Chin) Length of Sphenoid Sinus Choroid Plexus Biondi Bodies Inner Ear Canal Orientation Apical Phalangeal Tufts Age of Pelvic Bone Fusion Bone Cortex Thickness Laryngeal Position Pharyngeal Air Sacs Ear Lobes Sexual Body Size Dimorphism Lacrimal Gland Structure Visible Whites of the Eyes Small/Large Intestine Length Ratio Meningeal Artery Source</p> <p><b>BIOMECHANICS</b> Bipedal Gait Adductive Thumb Skeletal Muscle Strength Hand-Eye Coordination Fine Motor Coordination</p>	<p><b>ORGAN PHYSIOLOGY</b> Aldosterone Response to Posture Salt-Wasting Kidneys Ability For Sustained Running Voluntary Control of Breathing Ability to Dive Underwater Diving Reflex Ability to Float/Swim Emotion Lacrimation Salt Content of Tears Olfactory Sense</p> <p><b>CELL BIOLOGY</b> No Differences Are Known?</p> <p><b>BIOCHEMISTRY</b> Placental Alkaline Phosphatase N-Glycolylneuraminic Acid Expression Alpha 2-6-Linked Sialic Acid Expression</p> <p><b>ENDOCRINOLOGY</b> Thyroid Hormone Metabolism</p> <p><b>PHARMACOLOGY</b> Methylation of Inorganic Arsenic</p> <p><b>ANATOMIC PATHOLOGY</b> Cortical Neurofibrillary Tangles</p> <p><b>CLINICAL PATHOLOGY</b> Erythrocyte Sedimentation Rate Serum Alkaline Phosphatase Level RBC and Serum Folate Serum Vitamin B12/B12 Binding Total Leukocyte Count Absolute Neutrophil Count Absolute Lymphocyte Count</p> <p><b>DENTAL BIOLOGY/DISEASE</b> Canine Tooth Diastema Canine Tooth Dysmorphism Tooth Enamel Thickness Retromolar Gap Third Molar Impaction Dental Eruption Sequence/Timing</p> <p><b>MEDICAL/SURGICAL DISEASES</b> HIV Progression to AIDS <i>P. falciparum</i> malaria Viral Hepatitis B/C Complications Influenza A Infection Severity Incidence of Carcinomas Hemorrhoids Varicose Veins Pelvic Phleboliths Foamy Virus (Spumavirus) Infections Sexually Transmitted Diseases</p> <p><b>IMMUNOLOGY</b> Sialoadhesin on Macrophages</p> <p><b>SKIN BIOLOGY AND DISEASE</b> Eyebrows Eccrine Sweat Glands Acne Vulgaris Subcutaneous Fat Body Lice</p>	<p><b>NUTRITION</b> Frugivory Carnivory Aquatic Foods Underground Foods Cooking</p> <p><b>NEUROANATOMY</b> Relative Brain Size Direct Cortical Projections Relative Volume of Frontal Cortex Relative Volume of Corpus Callosum Relative Volume of Cerebellum % of Brain Growth Complete at Birth Rate of Postnatal Brain Growth</p> <p><b>NEUROBIOLOGY</b> Population Distribution of Handedness Postnatal Dendritic Growth Postnatal Synapse Formation Cortical Synapse Density Cortical Neuron Density Dendrites Per Neuron Synapses Per Neuron Adult Neurogenesis Cingulate Cortical Spindle Neurons Finger Tip Sensory Nerve Endings</p> <p><b>NEUROCHEMISTRY</b> Brain Aromatisation of Testosterone Tyrosine Hydroxylase Heterogeneity</p> <p><b>MENTAL DISEASE</b> Schizophrenia Bipolar Psychosis Autism Suicide</p> <p><b>BEHAVIOR</b> Control of Facial Expressions Planning Ahead Intentional Deception Deliberately Delaying Gratification Long-Range Transport of Materials Secondary Tool-Making Mechanical Multi-Tasking Physical Abuse of the Young Torture Organized Warfare Adult Play Symbolic Play Abuse of Other Animals Inter-Group Coalition Formation Use of Containers Care of Infirm and Elderly Grandparenting Home Base Control of Fire Food Preparation Organized Gathering of Food Domestication of Animals Domestication of Plants Altruistic Punishment Peace-Making Somnambulism Mind-Altering Drug Use</p>	<p><b>COGNITIVE CAPACITY</b> Declarative Memory Imitative Learning Teaching Symbolic Representation Awareness of Death Awareness of the Past Awareness of the Future Theory of Mind Theory of Other Minds Empathy Numeracy</p> <p><b>COMMUNICATION</b> "Parentese" Sounds Infant "Protoconversations" Gestural Communication Symbolic Communication Semantics Grammar and Syntax Recursion Writing</p> <p><b>SOCIAL ORGANIZATION</b> Institutions Social Conventions Governments Enforcement Through Sanctions</p> <p><b>CULTURE</b> Composition of Art Composition of Music Composition of Rhythms Death Rituals Clothing (Covering of Body Parts) Rites of Passage Genocide Competitive Sports Practicing of Skills Physical Modifications of the Body Inheritance of Resources and Status Rhythmic Dance Sculpture Belief in Supernatural/ Religion Body Adornment Childbirth Customs Sexual Intercourse in Private Gift-Giving Hospitality Intertwining (e.g., weaving) Meal Times Poetry Property Construction of Shelters Taboos Taxonomy of Species Trade Measurement of Time Weapons Toys</p>
--	---	---	---

<sup>a</sup>A major limitation in translating genomic comparative information into an understanding of "humanness" is that we know relatively little about the basic phenotypic features of the great apes, relative to humans. This table lists topic areas in which there are real or claimed "differences" between humans and the great apes (as a group). A given "difference" listed here could be a suggested gain or loss in humans, with respect to the great apes. This is a partial listing of topics that will appear later at a Web-based "Museum of Comparative Anthropogeny" <http://origins.ucsd.edu/gapp1.html>.

**Table 2.** Differences between humans and apes in incidence or severity of medical conditions<sup>a</sup>

Medical condition	Humans	Great apes
<b>Definite</b>		
HIV progression to AIDS	Common	Very rare
Hepatitis B/C late complications	Moderate to severe	Mild
<i>P. falciparum</i> malaria	Susceptible	Resistant
Myocardial infarction	Common	Very rare
Endemic infectious retroviruses	Rare	Common
Influenza A symptomatology	Moderate to severe	Mild
<b>Probable</b>		
Menopause	Universal	Rare?
Alzheimer's disease pathology	Complete	No neurofibrillary tangles
Epithelial cancers	Common	Rare?
Atherosclerotic strokes	Common	Rare?
Hydatiform molar pregnancy	Common	Rare?
<b>Possible</b>		
Rheumatoid arthritis	Common	Rare?
Endometriosis	Common	Rare?
Toxemia of pregnancy	Common	Rare?
Early fetal wastage (aneuploidy)	Common	Rare?
Bronchial asthma	Common	Rare?
Autoimmune diseases	Common	Rare?
Major psychoses	Common	Rare?

<sup>a</sup>See Varki 2000; Olson and Varki 2003, and references therein. This list excludes disease states explained by anatomical differences, e.g., difficult labor, varicose veins, spine disorders, hemorrhoids, hernias, etc.

### Sequencing of the chimpanzee genome

Less than a decade ago, sequencing the chimpanzee genome was not even on the “radar screen” of the major sequencing centers. Repeated public statements of interest from many other sectors of the scientific community (McConkey and Goodman 1997; McConkey and Varki 2000; Varki 2000) and increasing interest within the genome community eventually led to the writing of “white papers”,<sup>4</sup> and the assignment of high priority to this effort (<http://www.genome.gov/10002851>). The recent analysis of the draft chimpanzee genome sequence (The Chimpanzee Sequencing and Analysis Consortium 2005), and the many “companion” papers (Cheng et al. 2005; Hughes et al. 2005; Linardopoulou et al. 2005) now provide researchers with a wealth of comparative genetic information.

The published sequence was from a single captive-born male of the *Pan troglodytes verus* subspecies. Sequence data obtained via a whole-genome shotgun approach to a BAC library generated an  $\sim 3.6\times$  coverage, i.e.,  $\sim 3.6$ -fold redundancy in sequencing reads from the autosomes (sex chromosomes have half that redundancy in a male). The assembled sequence covers  $\sim 94\%$  of the genome, with 98% of the sequence having an estimated error rate of  $\leq 10^{-4}$ . Several additional chimpanzees (including other subspecies) were sequenced at lower coverage. In addition to identifying polymorphisms within chimpanzees, these data confirmed the high quality and completeness of the human genome

<sup>4</sup>Olson, M.V., Eichler, E.E., Varki, A., Myers, R.M., Erwin, J.M., and McConkey, E.H.A. 2002. White paper advocating complete sequencing of the genome of the common chimpanzee, *Pan troglodytes* (white paper submitted to NHGRI, February 2002).

Reich, D.E., Lander, E.S., Waterston, R., Pääbo, S., Ruvolo, M., and Varki, A. 2002. Sequencing the chimpanzee genome (white paper submitted to NHGRI, February 2002).

sequence and established ancestral states of human single-nucleotide polymorphisms (SNPs).

### Defining the important differences: Searching for needles in a haystack

With the two genome sequences in hand, one can begin a systematic identification of genes, regulatory elements, and other functionally relevant genomic regions that differentiate humans and chimpanzees. Of course, what we are really exploring is a complex interplay between multiple genetic differences, interacting with diverse physiological, environmental, and cultural factors, eventually resulting in the observed phenotypic differences. Single-nucleotide divergence was estimated at  $\sim 1.23\%$ , with  $\sim 1\%$  corresponding to fixed species divergence and the remainder representing species-specific polymorphisms (The Chimpanzee Sequencing and Analysis Consortium 2005). While insertion-deletion (indel) events were fewer, they represented 40–45 Mb in each species, i.e.,  $\sim 90$  Mb difference between the two, giving an  $\sim 3\%$  divergence in this category. Thus, the overall divergence between the genomes is closer to 4%, in keeping with two recent studies (Britten 2002; Watanabe et al. 2004), but far greater than most previous estimates, which were made using shorter alignable sequence fragments. Fortunately, orthologous proteins are still extremely similar, with almost a third being identical, and the typical protein differing only by two amino acids between human and chimpanzees. Thus, the oft-repeated “ $<1\%$  difference” still applies to amino acid sequences (The Chimpanzee Sequencing and Analysis Consortium 2005). However, a substantial proportion of the differences will likely be neutral with respect to understanding the human condition. The search for functionally important differences is further complicated because many of the important ones may not be within known coding sequences. In addition to protein evolution (Li and Saunders 2005), two other major hypotheses have been put forth to explain human-specific changes, i.e., changes in gene regulation (King and Wilson 1975) and loss-of-function changes (Olson and Varki 2003). We suggest simultaneous genomic and candidate gene approaches toward narrowing the field to the functionally relevant changes captured under these hypotheses.

### Genomic approach I: Narrowing the search to the important differences

#### *Using outgroups to define human-specific changes*

Noting a difference between the two genome sequences does not indicate which lineage experienced the change. We can also assume that roughly half of all the differences occurred on the lineage leading to chimpanzees. One or more “outgroups” are needed to determine the ancestral state (nucleotide or otherwise) at any given locus, and thus establish the subset of human-specific changes. The large divergence time between primates and rodents ( $>60$  Myr) means that some mutational events and/or orthology between loci will be obscured when using rodents as outgroups. Sequencing additional primate genomes is thus important, as is the careful choice of an appropriate outgroup species. A primate close enough to humans and chimpanzees is needed to reliably determine substitutional polarity; however, some evolutionary distance is also useful to define regions of functional sequence conservation. The orangutan (*Pongo*) provides the appropriate level of sequence divergence—unlike *Gorilla*, which may be too closely related to humans and chimpan-

zees to provide a consistent signal of polarity (Satta et al. 2000; Klein and Takahata 2002). Sequencing of *Pongo* is already underway at the Genome Sequencing Centers (GSC) at Washington University, St. Louis and at the Baylor College of Medicine. Rhesus macaque (*Macaca mulatta*) genome sequencing is also underway, led by the Baylor GSC, in collaboration with the J. Craig Venter Institute Joint Technology Center, and Washington University GSC. The latter genome, representing an Old World monkey, is useful not only for its greater phylogenetic distance from humans (~25 Myr divergence) (Goodman et al. 2005), but also because the long tradition of rhesus macaque use in biomedical research provides comparative biological information and easier access to tissues and biological materials. Eventually, we will need multiple, additional primate genomes to fully understand the changes that have contributed to traits distinguishing humans (Goodman et al. 2005).

More speculatively, studying the gorilla genome could help define some genetic features related to cognitive function. Until recent human encroachment, the gorilla probably faced less social and cognitive challenges, with a single male overseeing a harem of females, in a relatively predator-free and food-rich environment. This contrasts to the more complex nature of chimpanzee and orangutan environments and behaviors. Also, unlike chimpanzees and orangutans, gorillas fail to pass the “mirror self-recognition test” (Shillito et al. 1999) and have only rarely been observed to use tools (Breuer et al. 2005). It is possible that the gorilla lost some genetic endowments related to cognitive abilities that the great ape common ancestor already had, and comparison to the human, chimpanzee, and orangutan genomes may help identify such changes.

A confounding factor with respect to the  $\sim 3.6\times$  chimpanzee genome sequence is the variable quality of the genome sequence and assembly process. Individual low-quality nucleotide positions can result in sites that appear falsely divergent between humans and chimpanzees. Sequence data from multiple individuals allows some such sites to be identified and eliminated. In addition, some higher-order discrepancies may be introduced by difficulties in assembling certain regions, resulting in false differences between chimpanzees and humans. All of these issues should be resolved as the “polishing” phase of the chimpanzee genome sequencing proceeds (E. Mardis, pers. comm.).

#### Excluding intra-species polymorphisms

One must also ensure that an apparent genomic difference between humans and chimpanzees is not simply due to a polymorphism in one of the species (Ruvolo 2004). While such polymorphisms are interesting in their own right (e.g., informing us about regions that have undergone recent selective sweeps), they are by definition not contributors to species-specific differences. Sequences from multiple individuals of both species are needed to set aside intra-specific polymorphism and focus only on fixed differences. Some genome-wide polymorphism data for chimpanzees already exists, and human polymorphism can be initially assessed using the numerous SNPs defined in current databases (<http://www.ncbi.nlm.nih.gov/projects/SNP/>). Surveying sequence variation in a minimum of 10 globally distributed humans has also been suggested to ensure a high probability that a given sequence is fixed (Enard and Pääbo 2004). This minimum number will be higher for chimpanzees because of their greater intra-specific diversity (Gagneux et al. 1999; Kaessmann et al. 1999; Stone et al. 2002; Yu et al. 2003).

## Genomic approach 2: Focusing the search

### Examine sites of human-specific chromosomal changes

In addition to the few previously known chromosomal inversions and rearrangements between humans and chimpanzees (Yunis et al. 1980; Yunis and Prakash 1982; Nickerson and Nelson 1998; Fan et al. 2002a,b; Dennehey et al. 2004), several new smaller chromosomal regions containing likely inversions and rearrangements were detected (The Chimpanzee Sequencing and Analysis Consortium 2005; Newman et al. 2005). Targeted comparisons with other great ape genomes will help to define which of these events are human specific. One can then examine each breakpoint region in detail, searching for potential changes in local genes or regulatory elements, as has already been done in a few instances (Nickerson and Nelson 1998; Fan et al. 2002a,b; Dennehey et al. 2004).

### Examine sites of human-specific insertions and deletions

Insertions or deletions (indels) can range from a few nucleotides to tens of kilobases, and have major impacts on gene structure, expression, and/or function. While the number of indel differences between human and chimpanzee genomes is lower than the number of Single Nucleotide Divergences (SNDs), fixation of such events could suggest that these losses/gains have been adaptive. There are already a few known examples of indels with potential functional consequences differentiating humans and chimpanzees (Table 3), including the human loss of *CMAH* gene function in humans due to a 92-bp exon deletion (Chou et al. 1998); the loss of two coding exons in the human *ELN* gene, which contributes to extracellular matrix structure (Szabo et al. 1999), and the complete deletion of *SIGLEC13* in humans (Angata et al. 2004).

The idea that gene loss was a major contributor to human evolution remains an intriguing one (Olson 1999; Olson and Varki 2003). Interestingly,  $\sim 50$  known or predicted human genes were found to be missing partially or entirely in the chimpanzee genome, and some of these differences were confirmed by PCR or Southern blotting (The Chimpanzee Sequencing and Analysis Consortium 2005). Confirmation of the ancestral state of these loci and reciprocal analysis of genes disrupted exclusively in humans requires additional primate outgroup data and further “polishing” of the chimpanzee genome sequence.

### Examine gene duplications and retroposed genes

Gene duplication via segmental duplication or retrotransposition of mRNA sequences is an evolutionary mechanism for creating new genes with new biological functions. Duplicated genes can become nonfunctional (pseudogenes), neofunctional (acquire a new function), or subfunctional (adopt a portion of the previous function) (Ohno 1999; Hurler 2004). Such species-specific changes in copy number of gene families may allow for the evolution of new functions unique to the species—and are thus pertinent loci for investigation. A recent study reported that 33% of human duplications are human specific (Cheng et al. 2005); and with an estimated 200–300 species-specific retroposed gene copies in humans and chimpanzees (The Chimpanzee Sequencing and Analysis Consortium 2005), there is an ample landscape to explore. Of note, previous work suggests that humans have experienced more copy-number changes than the great apes (Fortna et al. 2004), such as appears to be the case for the *PRAME* cluster (Birtle et al. 2005) and the *SPANX-B* genes (Kouprina et al.

**Table 3.** Some candidate genes and gene families that may contribute to phenotypic differences between humans and apes<sup>a</sup>

Gene(s)	Gene product(s)	Unusual hominid or human-specific features	Potential relevance to the human condition
Individual genes			
<i>FOXP2</i>	Putative transcription factor with polyglutamine tract and forkhead DNA binding domain	Two human-specific amino acid changes	Mutant humans have motoric speech disorder (developmental verbal dyspraxia). Region positively selected and fixed in humans <200,000 years ago (Enard et al. 2002b; Zhang et al. 2002)
<i>MYH16</i>	Myosin heavy chain 16	Human-specific 2-bp deletion causing frameshift—predicted 76-kD unstable head domain	Claimed to be cause of reduction in the type II fibres of human jaw muscle. (Stedman et al. 2004; Perry et al. 2005)
<i>CMAH</i>	CMP-Neu5Ac hydroxylase	92-bp deletion of exon 6 causing frameshift and inactive enzyme. Fixed in modern humans	Absence of sialic acid Neu5Gc. Change in resistance or susceptibility to pathogens. Loss of ligand for some Siglecs. Dated ~2.5–3 Mya. Dietary Neu5Gc in meat became foreign antigen (Chou et al. 1998, 2002; Irie et al. 1998; Hayakawa et al. 2001)
<i>MAOA</i>	Monoamine oxidase A	Human-specific nonconservative change Glu151Lys in active site	Substitution affects protein dimerization according to a 3D structural model and predicts functional change (Andres et al. 2004)
<i>ASPM</i>	Modulator of mitotic spindle in neural progenitors?	Accelerated evolution in ape and human lineages	Deletions in <i>ASPM</i> lead to microcephaly. Presumed to be related to increased brain size and/or other features of human brain (Zhang 2003; Dorus et al. 2004; Evans et al. 2004; Kouprina et al. 2004b; Mekel-Bobrov et al. 2005)
<i>MCPH1</i>	Microcephalin	As above	As above (Dorus et al. 2004; Evans et al. 2004, 2005)
<i>TTR</i>	Transthyretin	Decreased expression in humans in blood and brain	May be related to altered thyroid hormone metabolism in humans versus chimpanzees (Gagneux et al. 2001)
<i>ST6GAL1</i>	Alpha 2–6 sialyltransferase	Apparent human-specific up-regulation on epithelia	Can explain relative resistance of chimpanzees to human influenza A virus. Other consequences unknown. (Gagneux et al. 2003)
<i>EMR4</i>	EGF-TM7 receptor family	Human-specific deletion in exon 8. Frameshift	Predominantly expressed by immune system cells. Functional significance unknown (Hamann et al. 2003)
<i>PCDH11Y</i>	Protocadherin XY	Duplicated onto Y in Yp11.2/Xq21.3 pseudoautosomal region only in humans	Expressed from Y and escapes X-inactivation? Y copy has undergone structural changes. Selectively expressed in brain. Probable adhesion molecule. Significance unknown, hypothesized to be involved in brain development, lateralization and schizophrenia risk (Ross et al. 2003; Blanco-Arias et al. 2004)
<i>IL9R (Y)</i>	Interleukin-9 receptor	As above	Expressed from Y and escapes X-inactivation? Growth factor for T cells, mast cells, and macrophages. Significance unknown. Related to asthma? (Vermeesch et al. 1997)
<i>SPRY3</i>	Sprouty 3	As above	Expressed from Y and escapes X-inactivation? Cysteine-rich protein—Homolog of <i>Drosophila</i> antagonist of FGF signaling that patterns apical airways branching. Significance unknown (Vermeesch et al. 1997)
<i>SYBL1</i>	Synaptobrevin-like	As above	Inactive on Y chromosome? Significance unknown (Vermeesch et al. 1997)
<i>KRTHAP1</i>	Type 1 acidic hair keratin	Human-specific single bp substitution and termination codon	Different hair keratin expression pattern noted in the hair follicle. Inactivated 0.25 Mya? Possibly related to human:ape differences in hair (Winter et al. 2001)
<i>RLN</i>	Relaxin hormone	Human-specific expression in placenta and corpus luteum	Possibly related to differences in reproductive biology (Evans et al. 1994)
<i>ELN</i>	Tropoelastin	2 exons deleted. Open reading frame maintained	Extracellular matrix component, including vascular wall. Alteration in vascular wall structure? (Szabo et al. 1999)
<i>SIGLEC11</i>	Siglec-11	Human-specific gene conversion by adjacent pseudogene, maintaining ORF	Change in binding specificity for sialic acids. Human-specific expression in brain microglia. Biological consequences unknown (Hayakawa et al. 2005)
<i>CASP12P1</i>	Caspase-12—cysteine protease related to ICE subfamily	Human-specific disruption of SHG box required for activity. Premature stop codon also in most humans	In rodents, <i>Casp12</i> mediates apoptosis in response to ER stress. Human SNP can restore full-length caspase proenzyme which confers hypo-responsiveness to LPS-stimulated cytokine production but has no significant effect on apoptotic sensitivity. (Fischer et al. 2002)

(continued)

2004a). Also, some neofunctional retroposed loci such as *GLUD2* are thought to be involved in hominid brain function (Burki and Kaessmann 2004).

*Identify genes and gene families showing evidence of human-specific rapid evolution*

Genes that have the signature of accelerated evolution (Clark et

al. 2003; Nielsen et al. 2005), i.e., a high ratio of nonsynonymous to synonymous substitutions ( $K_a/K_s$  ratios), are good candidates for further study. In particular, genes that show  $K_a/K_s > 1$  are possible targets of positive selection (Messier and Stewart 1997; Yang and Bielawski 2000). Several loci with relatively high  $K_a/K_s$  ratios between the human and chimpanzee genomes were reported (The Chimpanzee Sequencing and Analysis Consortium

**Table 3.** *Continued*

Gene(s)	Gene product(s)	Unusual hominid or human-specific features	Potential relevance to the human condition
Gene families			
<i>OR</i> (17p13, etc.)	Olfactory receptors	Many more human pseudogenes and fewer active genes in this large family	Related to diminished human olfactory capabilities? However, some intact genes show evidence of positive selection (Gilad et al. 2003a,b 2004)
<i>TAS2R</i> (12p13, 7q31, 7q34, etc.)	Bitter taste receptors	Fixation of loss-of-function mutations	Proposed relaxation of selective constraint and loss of function (Wang et al. 2004; Fischer et al. 2005)
<i>SIGLEC</i> (19q13)	CD33-related innate immune system regulating genes	Mutations, deletions, gene conversions, expression changes	Sialic acid recognizing signaling receptors. Changes in binding, expression patterns, etc. Could be partly a secondary consequence of human loss of Neu5Gc (Angata et al. 2001; Sonnenburg et al. 2004; The Chimpanzee Sequencing and Analysis Consortium 2005)
<i>COX</i> (multiple locations)	Mitochondria cytochrome oxidase subunits	Multiple genes show rapid evolution in hominids. <i>COX5A</i> specifically in humans	Altered electron transport chain. Enhanced oxidative phosphorylation postulated to support increased brain energy consumption? (Grossman et al. 2004; Goodman et al. 2005)
<i>SPANX</i> (Xq27.1)	Sperm proteins associated with nucleus—genes on X chromosome	<i>SPANX-C</i> is specific to humans. <i>SPANX-B</i> has duplicated in humans	Rapidly evolving in all hominids. Expressed in normal testis, and in some cancers (Kouprina et al. 2004a)
<i>Morpheus</i> (multiple locations)	Proteins not characterized yet	Contained within large duplicated regions in humans and apes	Evidence of rapid evolution. Most extreme case of positive selection among hominids. Some human-specific sequences. Functional significance uncertain (Johnson et al. 2001)
<i>LILR</i> (19q13.4)	Leukocyte Ig-like receptors	Rapid evolution, only few clear orthologs between chimpanzee and human	Part of a larger family of genes. Involved in recognizing “self” via molecules like MHC (Canavez et al. 2001)
<i>KIR</i> (19q13.4)	Killer inhibitory receptors	Rapid evolution, only few orthologs clear between chimpanzee and human	Expressed in NK cells. Recognize “self” molecules like MHC (Hao and Nei 2005; Sambrook et al. 2005)
<i>TRG</i> (7p14)	T cell receptors	4 TCRs are pseudogenes in humans	Part of a larger family of genes. Functional significance uncertain (Meyer-Oslon et al. 2003)
<i>FCGR1</i> (1p and 1q)	High affinity IgG-Fc receptors	Pericentric inversion, distinguishing human from chimpanzee chromosome 1	Functional significance of inversion uncertain (Maresco et al. 1998)
<i>IGKV</i> (2p11.2)	$\kappa$ light chains of immunoglobulins	Possible human specific duplication	Part of a larger family of genes. Functional significance uncertain (Ermer et al. 1995)
<i>GYP</i> (4q28-q31, 2q14-q21)	Glycophorins	Accelerated evolution in humans	Red blood cell proteins. Rapid evolution of extra cellular domain, likely due to selection pressure by merozoite stage of <i>Plasmodium falciparum</i> (Rearden et al. 1990; Baum et al. 2002; Wang et al. 2003)
<i>LCE</i> (1q21)	Epidermal differentiation complex	High density of rapidly evolving genes	Proteins that help form the cornified layer of the skin barrier (Marshall et al. 2001; The Chimpanzee Sequencing and Analysis Consortium 2005)
<i>CST</i> (20p11)	Cystatins	As above	Physiological cysteine proteinase inhibitors (The Chimpanzee Sequencing and Analysis Consortium 2005)
<i>PSG</i> (19q13)	Pregnancy-specific $\beta$ -1-glycoproteins	As above	High quantities secreted by placental trophoblasts. Exact physiologic role during pregnancy unknown (The Chimpanzee Sequencing and Analysis Consortium 2005)
<i>KRT</i> (17q21)	Hair keratins and keratin-associated proteins	As above	Major components of the cytoskeleton in hair and skin epithelial cells (The Chimpanzee Sequencing and Analysis Consortium 2005)
<i>WFDC</i> (20q13)	Protein domains with homology to whey acidic protein (WAP)	As above	Postulated protease inhibitors. Possible host defense against invading micro-organisms or regulation of endogenous proteolytic enzymes (The Chimpanzee Sequencing and Analysis Consortium 2005)

<sup>a</sup>This list is not meant to be exhaustive. It also does not include genes that have specifically changed in chimpanzees, but not in humans (e.g., *MICA/B*, *HLA*); genes that are polymorphic within humans (e.g., *APOE*, *COMT*); or instances in which a human disease-causing amino acid mutation appears to be the wild-type state in the chimpanzee (e.g., *AIRE*, *MKKS*, *MLH1*, *MYOC*, *OTC*, and *PRSS1*).

2005). Since a majority of nonsynonymous substitutions are considered deleterious (Enard and Pääbo 2004), a high rate of nonsynonymous substitution between taxa can suggest either adaptive evolution or relaxation of functional constraint. However, this approach is generally conservative (Yang and Bielawski 2000). For example, a  $K_a/K_s$  value of  $<1$  does not rule out that a gene has undergone positive selection (Dorus et al. 2004). Also, a protein could have only one or a few important amino acid changes, perhaps confined to a critical domain, motif, or site (Andres et al. 2004; Sonnenburg et al. 2004), and thus not have an elevated  $K_a$  relative to  $K_s$ . Careful examination of the specific types or positions of amino acid changes such as radical amino acid substitutions (hydrophobic vs. hydrophilic, acidic vs. basic, etc.) in conserved regions is another potential way to identify important changes in protein sequence.

For genes that have zero synonymous changes between humans and chimpanzees, one has to use the adjacent genome sequence to estimate a local intergenic/intronic substitution rate,  $K_i$ . Of ~13,000 human–chimpanzee orthologs studied, ~4% had an observed  $K_a/K_i >1$  (The Chimpanzee Sequencing and Analysis Consortium 2005). However, given the low divergence between humans and chimpanzees, about half of these are predicted to occur simply by chance if purifying selection is allowed to act nonuniformly across genes.

#### Examine sites of human-specific repetitive element insertion

Repetitive elements such as LINEs (long interspersed elements) and SINEs (short interspersed elements) can duplicate and spread throughout the genome by reverse transcription, causing potentially important functional changes in coding and flanking sequences (Smit 1999; Carroll et al. 2001). *Alu* elements are the most abundant class of SINEs in humans, making up ~10% of the genome (Lander et al. 2001), where they apparently expanded up to three times more than in the chimpanzee genome (The Chimpanzee Sequencing and Analysis Consortium 2005). In addition, most human-specific *Alu* elements belong to two subfamilies (Ya5 and Yb8) not found in great apes (Carroll et al. 2001). Identification of these human-specific loci makes them candidates for further inquiry. It is possible that some of these elements inserted into functional genes or flanking regions became alternatively spliced introns or promoter regulators, or either deleted or shuffled genomic regions via *Alu*–*Alu* recombination.

#### Look for human-specific gene conversions

Another potential source of differences arises from species-specific gene conversion events that become fixed. Gene conversion homogenizes coding or noncoding sequences between adjacent paralogous gene copies within a species. Conversion may also introduce harmful mutations from a pseudogenized gene copy into a functional copy, or conversely, restore function to a former pseudogene. For example, the 5' end of human *Siglec-11* was converted by an adjacent pseudogene after the common ancestor with chimpanzees (Hayakawa et al. 2005). This resulted in a change in sialic acid-binding properties, as well as new expression in human brain microglia. The gene-converted *Siglec-11* can thus be considered the first example of a human-specific protein. More such examples might be found by systematically screening genomic regions, wherein genes and paralogous pseudogenes are nearby one another.

#### Look for changes in noncoding regions

A majority of comparative genomic studies have focused on coding regions at the expense of examining regulatory sequences (Carroll 2005). However, given the relatively few protein-sequence differences between human and chimpanzees, differential regulation of gene and protein expression is a likely mechanism for explaining human:chimpanzee differences (King and Wilson 1975; Enard et al. 2002a; Caceres et al. 2003; Carroll 2003; Preuss et al. 2004; Uddin et al. 2004). Functional noncoding regions such as promoters, enhancers, flanking sequences, and introns can regulate the expression of genes (Wray et al. 2003), and thus play a role in human evolution. The wealth of new information being generated about noncoding RNA sequences also makes them an intriguing candidates for potential differences (Eddy 2001; Dykxhoorn et al. 2003; Mello and Conte 2004; Kim 2005; Tang 2005).

#### Genomic approach 3: Looking for human-specific gene expression differences

As mentioned above, species-specific changes in genomic sequence can be manifested in regulatory processes such as timing and location of expression of genes or of functional noncoding sequences, such as siRNAs. However, it is difficult to predict changes in expression simply by comparing genomic sequences (Carroll 2005). Differences in expression pattern between humans and chimpanzees are being investigated using microarray analyses, which allow for a rapid screen of multiple loci expressed in a single tissue at a given time point. Several such analyses and reanalyses have been carried out (Enard et al. 2002a, 2004; Caceres et al. 2003; Gu and Gu 2003; Hsieh et al. 2003; Khaitovich et al. 2004a; Preuss et al. 2004; Uddin et al. 2004). While the rate of brain gene expression changes appears increased in the human lineage, gene expression in the brain is overall more conserved than in other tissues, perhaps because of functional constraints in this complex organ (Enard et al. 2002a; Caceres et al. 2003; Gu and Gu 2003; Preuss et al. 2004). However, there are several caveats. For example, microarrays based on human oligonucleotide sequences may not accurately detect levels of expression in nonhuman primates nor detect significant alternative splicing of mRNAs (Modrek and Lee 2002; Hsieh et al. 2003; Preuss et al. 2004; Steinmetz and Davis 2004). Additionally, mRNA levels are not always good predictors of the actual levels of the gene product found in a cell (Gygi et al. 1999). Moreover, a recent study suggests that most expression differences have little or no significance, and are likely due to neutral evolution (Khaitovich et al. 2004b). Finally, many of the ultimate “gene products” are not the proteins themselves, but result from their enzymatic activity (e.g., lipids, glycans, and bioactive small molecules). Thus, gene expression studies must be complemented by a variety of other “omic” approaches, e.g., proteomics, lipomics, glycomics, etc. Any differences found need to be confirmed by focused biochemical studies on the molecules in question.

#### Candidate gene approaches

In parallel with the above genomic studies, it is important to continue the more traditional candidate gene approach—as the genomic approach can miss many biologically significant differences. The candidate approach focuses on specific genes, based on some a priori knowledge about which loci or system(s) might be expected to show functionally significant differences between humans and chimpanzees.

**Candidate gene approach 1: Making choices on the basis of comparative phenomics**

Humans and chimpanzees differ in many morphological, cognitive, and physiological arenas. When attempting to identify the genetic mechanisms responsible, it is logical to focus on genes known or predicted to contribute in some way to the phenotypic differences, i.e., differences in the “phenome.” There are many morphological and physiological traits for which we have some knowledge of the responsible genetic pathways. This can, in turn, allow us to identify appropriate candidate loci underlying the traits. We can then test them for their contribution to uniquely human traits affecting organs such as the skin, brain, and female reproductive system (Table 1). Additionally, many diseases and pathological conditions appear to be unique to humans, and genes involved in some of these disease pathways are known or can be predicted (Table 2). It makes sense to focus first on phenotypes or diseases that appear most directly relevant to explaining the human condition. For example, recent work has suggested that two genes involved in the regulation of brain size appear to have undergone human-specific adaptive evolution (Evans et al. 2005; Mekel-Bobrov et al. 2005). However, we would recommend against a purely “brain-centric” approach that assumes that the only major differences of interest are in the nervous system. A single genetic change may have had an impact on multiple organs, and such a change may be easier to study in organs other than the brain. For example, there are organs such as the skin and its derivatives (e.g., the female breast and the sweat glands) that show at least as many morphological and functional differences as the brain and are easier to study. Genetic differences found in such systems may then help predict which molecules, pathways, or mechanisms have also undergone the most drastic changes during the evolution of the human brain.

**Candidate gene approach 2: Choices based on naturally occurring human mutations**

A population size of 6 billion humans suggests that many postnatally viable genetic diseases affecting “uniquely human” traits are likely to exist somewhere on the planet. Identifying such defects in the human population, particularly in families, provides an approach for directly linking genotype to phenotype and for choosing genes for human and chimpanzee comparisons. The medical community in particular should be educated and vigilant about such opportunities. A striking outcome of this type of approach is *FOXP2*, a transcription factor shown to be associated with an inherited human disorder of speech production (Enard et al. 2002b; Zhang et al. 2002). Intriguingly, this putative transcription factor was found to have two human-specific amino acid changes, and the genomic region in question appears to have been positively selected and fixed in humans <200,000 years ago (Enard et al. 2002b). The next step is to look at the consequences of such abnormal genotypes *in vitro* and by developing transgenic mice that manifest symptoms of the condition. Indeed, mice with a disruption in a single copy of the murine *Foxp2* gene manifest a modest developmental delay and a significant alteration in ultrasonic vocalizations that are normally elicited when pups are removed from their mothers (Shu et al. 2005).

Another intriguing finding is that some amino acid sequence variants that cause disease in humans turn out to be a

reversion to the conserved ancestral state, still present in the normal chimpanzee (The Chimpanzee Sequencing and Analysis Consortium 2005). This phenomenon has been explained as being due to a high rate of compensatory mutations at other sites in the same protein. Assuming that such mutations are more likely to be fixed by positive selection than by neutral drift, these genes are candidates for adaptive differences between humans and chimpanzees.

**Candidate gene approach 3: Making choices based on sequence data**

Both “top-down” and “bottom-up” approaches have been successfully used to identify genes potentially involved in uniquely human phenotypes. As discussed above, the “phenome-down” candidate approach involves selecting a candidate gene based on phenotypic information and doing a first-pass genomic workup before proceeding to functional analyses of the gene product. Conversely, a “genome-up” approach (see genomic approach 2 above) can identify genes involved in a particular pathway or system that may be diverged enough (i.e., high  $K_a/K_s$  values) or expressed in the organs of interest, or harbor amino acid mutations (i.e., in a conserved domain) to be of interest in a functional screen. Following a search for such candidate genes, a narrowed-down list of loci can then be prioritized according to putative function and position in the pathway of interest (as opposed to a complete list of loci generated from a purely genome-based search, with little functional knowledge linked to them). For example, the human sequence could be “chimpanized” and the gene product compared with that of the native human and chimpanzee gene product via *in vitro* or transgenic mouse studies in order to investigate the effect of particular sequence changes on a given phenotype.

One example of a functional genetic difference discovered through candidate genomic sequence analysis is *MYH16*, initially identified as a putative member of the myosin heavy-chain family (Stedman et al. 2004). The *MHY16* gene product is most prominently expressed in the jaw muscles of vertebrates. Humans are homozygous for a defective frame-shifted allele. Stedman et al. (2004) dated the mutation to ~2.4 Mya, approximately the time of origin of the genus *Homo*, and hypothesized that the human-specific loss of *MYH16* function may have affected craniofacial morphology and/or selected for the evolution of larger brain size (Currie 2004; Stedman et al. 2004). Alternatively, since human ancestors switched to a less herbivorous diet at about that time, the loss of *MHY16* and jaw muscle strength might simply have been inconsequential and thus drifted to fixation (Currie 2004).

A subsequent analysis of a much larger region of exonic and intronic sequence data flanking the deletion (30,000 vs. 1000 bp) estimated the age of the mutation as ~5 Mya (Perry et al. 2005), consistent with the timing of human–chimpanzee divergence rather than with the origin of *Homo*. While this may cast doubt on *MYH16*'s role in the evolution of *Homo*, the fact remains that the frameshift is human specific and belongs in the repertoire of human–chimpanzee genetic differences. Whether it contributed in some meaningful way to species-specific character differences is a question for continued investigation. Regardless, the availability of the human and chimpanzee genome sequences facilitated the expanded analysis, and underscores the important role that genome sequences can play in our understanding about evolutionary history.



#### *Candidate gene approach 4: A “systems approach” to promising groups of genes*

The traditional and powerful approach to genome-wide analysis has been to either consider homologous gene families or genes that are grouped together by similar functions, as in the Gene Ontology (GO) System (Ashburner et al. 2000; The Chimpanzee Sequencing and Analysis Consortium 2005). Of course, since most genes are highly interrelated and function in multiple pathways and systems, no single classification system can do justice to all of the possibilities. One complementary approach is to select a biological process or system not defined under a traditional GO category and focus attention on groups of genes that are thought to be involved. Taking such an approach, Dorus et al. (2004) recently found that a group of genes involved in nervous-system development and function showed evidence of accelerated evolution, when compared with other “housekeeping” genes. A related approach is to assume that a major change in a single gene is likely to affect the evolution of other genes that are functionally connected. For example, following up on the discovery of the *CMAH* mutation affecting synthesis of one type of sialic acid (Chou et al. 1998), multiple functional genetic differences in the biology of sialic acids have been identified between humans and great apes (Angata et al. 2001; Gagneux et al. 2003; Sonnenburg et al. 2004). Since <60 genes are directly involved in all of the major processes of sialic acid biology, it is reasonable to suggest that this system underwent multiple related changes at some point(s) in human evolution. A systematic comparative analysis of all of these genes between humans and chimpanzees is underway. However, the genes in question are not identified in the GO system as belonging to a single category. By exercising both caution and creativity in how they identify loci united in a biological process, researchers will likely come up with new and novel insights into human and chimpanzee evolution.

#### Conclusions

Sequencing of the chimpanzee genome signals not an end, but rather a beginning for researchers across diverse fields. The impressive array of data and analyses that have come from this sequencing has provided researchers with new and novel insights into rates and results of molecular processes such as nucleotide substitutions, gene duplications, insertions and deletions, retrotranspositions, and potential karyotypic changes. These data will provide the springboard for understanding the potential consequences of changes in these attributes between humans and chimpanzees. Over the years, scientists have proposed many theories about what makes humans different from the great apes, ranging from subtle changes in regulatory regions (King and Wilson 1975) all the way to the differential loss of gene activity in humans (Olson 1999; Olson and Varki 2003). In fact, given the rather complex series of events evident in the hominid fossil record (Wood and Collard 1999; Cela-Conde and Ayala 2003), every one of these hypothesized genetic mechanisms likely contributed to some degree to human–chimpanzee differences. Understanding what makes us evolutionarily, biomedically, and cognitively different from chimpanzees will require extensive comparative phenomics to complement the comparative genomics now possible using the chimpanzee genome. However, despite decades of research on wild and captive chimpanzees, our overall knowledge about the chimpanzee phenome is very incomplete (Gagneux 2004; Olson and Varki 2004; McConkey and

Varki 2005). Studies of intra-specific variation among great apes are few in their infancy, and biomedical and physiological data are few. This lack of comparative phenotypic data represents a serious knowledge imbalance. Better phenomic data would enhance our ability to make additional, focused choices for candidate gene studies, and also increase our understanding of the biochemical consequences of any genomic changes we do find. One step to extend the utility of the genome project is to have the phenome much better defined, not only through morphological and anatomical studies, but also via systematic collection of existing data in all fields relevant to understanding the human condition (practically speaking, most of the biological and social sciences). A recently initiated “Great Ape Phenome Project” will begin this process (Varki et al. 1998; Gagneux 2004; Olson and Varki 2004). Of course, the critically endangered status of great apes in the wild, and the fiscal, logistical, and ethical issues of studying great apes in captivity (Gagneux et al. 2005; McConkey and Varki 2005) create a situation wherein new data and resources will not be easy to come by. Regardless, for the purposes of comparison, there is no point in doing any study on a captive great ape that one would not also do on a human subject (Gagneux et al. 2005). Also, all studies on captive apes should try to financially contribute toward their conservation in the wild, e.g., via a proposed Great Apes Conservation Trust, which would receive a 10% overage on all grant funds awarded by various agencies for research projects on ape genomes, phenomes, or behavior (McConkey and Varki 2005).

In the absence of adequate comparative phenomic data between humans and chimpanzees, genomic data provides only part of the blueprint for the phenotype. It is crucial, after identifying differences in the genomic data, to ascertain which ones are important by studying their biological consequences in the laboratory. For example, the relatively limited genomic differences between humans and chimpanzees mean that identifying statistically meaningful differences in rates of evolution are difficult. This limitation will hamper our ability to identify genes or regions of biological interest and importance. Additional primate outgroups will be important for detecting selection over longer time periods and for eliminating false positives. Also, genomic data alone cannot predict epistatic interactions between various loci, nor can it reveal the pleiotropic effects of changes that have occurred in a single gene. Comparative functional studies are necessary to reap the full potential of the genomic data, to translate the observed genetic changes into tangible quantitative differences. However, even such systematic functional studies may not capture the full magnitude of a difference’s importance by examining only a single player in a multiplayer interaction. It is likely that, while there may be single-gene changes of large consequence, there will also be synergistic effects of many minor changes at multiple loci. That is, the human condition is likely to be the result of many small effect changes, not just a few large effect mutations. These smaller, subtle changes will be difficult to detect by genomic methods. On the other hand, even clearly identifiable genomic and phenomic differences between humans and chimpanzees may not be directly related to speciation nor to the question of “what makes us human.” Such differences may be a simple byproduct of neutral divergence or genetic drift.

Also, what might seem an important phenotypic difference between humans and great apes might not actually be the most critical factor in determining unique features of the human condition. For example, despite the frequent atten-

tion given to big brain size (Wood and Collard 1999; Preuss 2005), there is little evidence for causative connections between brain size and human cognitive abilities (Preuss 2005). Additionally, maximum brain size was achieved long before the emergence of modern human behaviors (Klein 1999; Wood and Collard 1999). Thus, while increased brain size is an impressively human-specific phenotypic difference from great apes, it may well have been just one step (like bipedalism) that occurred earlier, along the way to the emergence of uniquely human cognitive features. Conversely, apparently small phenotypic differences could turn out to play major roles. For example, a small (approximately twofold) difference in the level of a thyroid hormone-binding protein and associated differences in thyroid hormone metabolism between humans and apes (Gagneux et al. 2001) could turn out to be as important as brain-expressed genes in altering the trajectory and mechanisms of human brain development.

Explaining “humanness” is a vague and broadly philosophical question, not easily approached using the genome alone. We prefer to use the term “the human condition” to refer to the entire suite of characters that makes humans different from the great apes. What it means to be human involves quantitative aspects of biochemistry, physiology, and morphology, as well as more qualitative arenas such as cognition, behavior, symbolic communication, and culture. However, unlike typical biological questions, the great majority of experiments one might propose for studying the consequences of species-specific genetic changes are unethical and/or impractical to do, either in humans or in great apes (Gagneux et al. 2005; McConkey and Varki 2005). Meanwhile, studies in mice may not provide sufficient answers. Thus, we suggest that many answers must come from a logical inductive approach that synthesizes many various “clues” to arrive at the best possible “diagnosis”. Also, apparently minor differences between humans and great apes could turn out to be critical. For all of these reasons, we must keep an open mind, and leave no clue unattended to, even if it may appear trivial at first glance. It may well be that findings made from systems that are more ethically accessible and practical to study (such as the blood and the skin) will reveal clues that will eventually allow generation of testable hypotheses about organs like the brain. The other reason to take this type of broad approach to the “human condition” is that there are major biomedical lessons to be learned, which will benefit both humans and great apes, even though they may not be useful in explaining “humanness” in its philosophical sense.

Because of the many limitations mentioned above, we will have to arrive at many of our conclusions by considering all of the facts in aggregate, including some circumstantial evidence. In the final analysis, the best long-term approach to understanding human–chimpanzee differences is to ensure that the next generation of biologists interested in the evolution of the human phenotype is a cross-trained and collaborative one, with an interdisciplinary focus. Interactions among a great many disciplines, such as genomics, biochemistry, physiology, neurobiology, cognitive science, medicine, pathology, anthropology, ecology, primatology, and evolutionary biology, will be essential in dissecting out the key genetic features that contribute to making us human.

## Acknowledgments

We thank three anonymous reviewers, Anders Aannestad, Sandra Diaz, Pascal Gagneux, Hopi Hoekstra, Elaine Mardis, Tarjei Mik-

elsen, Jennifer Stevenson, and Nissi Varki for valuable comments and suggestions. We also thank Jim Else, Liz Strobert, and Dan Anderson at the Yerkes Primate Center, Atlanta, GA for helpful discussions about great ape diseases. A.V. was supported by grants from the NIGMS, NHLBI, and NCI and by the Harold G. and Leila Y. Mathers Charitable Foundation, and T.K.A. was supported by a postdoctoral fellowship from the American Cancer Society.

## References

- Andres, A.M., Soldevila, M., Navarro, A., Kidd, K.K., Oliva, B., and Bertranpetit, J. 2004. Positive selection in MAOA gene is human exclusive: Determination of the putative amino acid change selected in the human lineage. *Hum. Genet.* **115**: 377–386.
- Angata, T., Varki, N.M., and Varki, A. 2001. A second uniquely human mutation affecting sialic acid biology. *J. Biol. Chem.* **276**: 40282–40287.
- Angata, T., Margulies, E.H., Green, E.D., and Varki, A. 2004. Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. *Proc. Natl. Acad. Sci.* **101**: 13251–13256.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**: 25–29.
- Baum, J., Ward, R.H., and Conway, D.J. 2002. Natural selection on the erythrocyte surface. *Mol. Biol. Evol.* **19**: 223–229.
- Birtle, Z., Goodstadt, L., and Ponting, C.P. 2005. Duplication and positive selection among hominin-specific PRAME genes. *BMC Genomics* **6**: 120.
- Blanco-Arias, P., Sargent, C.A., and Affara, N.A. 2004. A comparative analysis of the pig, mouse, and human PCDHX genes. *Mamm. Genome* **15**: 296–306.
- Breuer, T., Ndoundou-Hockemba, M., and Fishlock, V. 2005. First observation of tool use in wild gorillas. *PLoS Biol.* **3**: e380.
- Britten, R.J. 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl. Acad. Sci.* **99**: 13633–13635.
- Brunet, M., Guy, F., Pilbeam, D., Mackaye, H.T., Likius, A., Ahounta, D., Beauvilain, A., Blondel, C., Bocherens, H., Boisserie, J.R., et al. 2002. A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**: 145–151.
- Burki, F., and Kaessmann, H. 2004. Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux. *Nat. Genet.* **36**: 1061–1063.
- Caceres, M., Lachuer, J., Zapala, M.A., Redmond, J.C., Kudo, L., Geschwind, D.H., Lockhart, D.J., Preuss, T.M., and Barlow, C. 2003. Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl. Acad. Sci.* **100**: 13030–13035.
- Canavez, F., Young, N.T., Guethlein, L.A., Rajalingam, R., Khakoo, S.I., Shum, B.P., and Parham, P. 2001. Comparison of chimpanzee and human leukocyte Ig-like receptor genes reveals framework and rapidly evolving genes. *J. Immunol.* **167**: 5786–5794.
- Carroll, S.B. 2003. Genetics and the making of *Homo sapiens*. *Nature* **422**: 849–857.
- . 2005. Evolution at two levels: On genes and form. *PLoS Biol.* **3**: e245.
- Carroll, M.L., Roy-Engel, A.M., Nguyen, S.V., Salem, A.H., Vogel, E., Vincent, B., Myers, J., Ahmad, Z., Nguyen, L., Sammarco, M., et al. 2001. Large-scale analysis of the *Alu* Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *J. Mol. Biol.* **311**: 17–40.
- Cela-Conde, C.J. and Ayala, F.J. 2003. Genera of the human lineage. *Proc. Natl. Acad. Sci.* **100**: 7684–7689.
- Chen, F.C. and Li, W.H. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68**: 444–456.
- Cheng, Z., Ventura, M., She, X., Khaitovich, P., Graves, T., Osoegawa, K., Church, D., DeJong, P., Wilson, R.K., Pääbo, S., et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437**: 88–93.
- The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.
- Chou, H.H., Takematsu, H., Diaz, S., Iber, J., Nickerson, E., Wright, K.L., Muchmore, E.A., Nelson, D.L., Warren, S.T., and Varki, A. 1998. A

- mutation in human CMP-sialic acid hydroxylase occurred after the *Homo-Pan* divergence. *Proc. Natl. Acad. Sci.* **95**: 11751–11756.
- Chou, H.H., Hayakawa, T., Diaz, S., Krings, M., Indriati, E., Leakey, M., Pääbo, S., Satta, Y., Takahata, N., and Varki, A. 2002. Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl. Acad. Sci.* **99**: 11736–11741.
- Clark, A.G., Glanowski, S., Nielsen, R., Thomas, P.D., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D., Lu, F., Murphy, B., et al. 2003. Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. *Science* **302**: 1960–1963.
- Currie, P. 2004. Human genetics: Muscling in on hominid evolution. *Nature* **428**: 373–374.
- Dennehey, B.K., Gutches, D.G., McConkey, E.H., and Krauter, K.S. 2004. Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. *Genomics* **83**: 493–501.
- Dorus, S., Vallender, E.J., Evans, P.D., Anderson, J.R., Gilbert, S.L., Mahowald, M., Wyckoff, G.J., Malcom, C.M., and Lahn, B.T. 2004. Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119**: 1027–1040.
- Dykxhoorn, D.M., Novina, C.D., and Sharp, P.A. 2003. Killing the messenger: Short RNAs that silence gene expression. *Nat. Rev. Mol. Cell. Biol.* **4**: 457–467.
- Eddy, S.R. 2001. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**: 919–929.
- Enard, W. and Pääbo, S. 2004. Comparative primate genomics. *Annu. Rev. Genomics Hum. Genet.* **5**: 351–378.
- Enard, W., Khaitovich, P., Klose, J., Zollner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., et al. 2002a. Intra- and interspecific variation in primate gene expression patterns. *Science* **296**: 340–343.
- Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S., Wiebe, V., Kitano, T., Monaco, A.P., and Pääbo, S. 2002b. Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**: 869–872.
- Enard, W., Fassbender, A., Model, F., Adorjan, P., Pääbo, S., and Olek, A. 2004. Differences in DNA methylation patterns between humans and chimpanzees. *Curr. Biol.* **14**: R148–R149.
- Ermert, K., Mitlohner, H., Schempp, W., and Zachau, H.G. 1995. The immunoglobulin  $\kappa$  locus of primates. *Genomics* **25**: 623–629.
- Evans, B.A., Fu, P., and Tregear, G.W. 1994. Characterization of two relaxin genes in the chimpanzee. *J. Endocrinol.* **140**: 385–392.
- Evans, P.D., Anderson, J.R., Vallender, E.J., Choi, S.S., and Lahn, B.T. 2004. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. *Hum. Mol. Genet.* **13**: 1139–1145.
- Evans, P.D., Gilbert, S.L., Mekel-Bobrov, N., Vallender, E.J., Anderson, J.R., Vaez-Azizi, L.M., Tishkoff, S.A., Hudson, R.R., and Lahn, B.T. 2005. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* **309**: 1717–1720.
- Fan, Y., Linardopoulou, E., Friedman, C., Williams, E., and Trask, B.J. 2002a. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes. *Genome Res.* **12**: 1651–1662.
- Fan, Y., Newman, T., Linardopoulou, E., and Trask, B.J. 2002b. Gene content and function of the ancestral chromosome fusion site in human chromosome 2q13-2q14.1 and paralogous regions. *Genome Res.* **12**: 1663–1672.
- Fischer, H., Koenig, U., Eckhart, L., and Tschachler, E. 2002. Human caspase 12 has acquired deleterious mutations. *Biochem. Biophys. Res. Commun.* **293**: 722–726.
- Fischer, A., Gilad, Y., Man, O., and Pääbo, S. 2005. Evolution of bitter taste receptors in humans and apes. *Mol. Biol. Evol.* **22**: 432–436.
- Fortna, A., Kim, Y., MacLaren, E., Marshall, K., Hahn, G., Meltesen, L., Brenton, M., Hink, R., Burgers, S., Hernandez-Boussard, T., et al. 2004. Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol.* **2**: E207.
- Freimer, N. and Sabatti, C. 2003. The human phenome project. *Nat. Genet.* **34**: 15–21.
- Gagneux, P. 2004. A *Pan-Oramic* view: Insights into hominid evolution through the chimpanzee genome. *Trends Ecol. Evol.* **19**: 571–576.
- Gagneux, P. and Varki, A. 2001. Genetic differences between humans and great apes. *Mol. Phylogenet. Evol.* **18**: 2–13.
- Gagneux, P., Wills, C., Gerloff, U., Tautz, D., Morin, P.A., Boesch, C., Fruth, B., Hohmann, G., Ryder, O.A., and Woodruff, D.S. 1999. Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc. Natl. Acad. Sci.* **96**: 5077–5082.
- Gagneux, P., Amess, B., Diaz, S., Moore, S., Patel, T., Dillmann, W., Parekh, R., and Varki, A. 2001. Proteomic comparison of human and great ape blood plasma reveals conserved glycosylation and differences in thyroid hormone metabolism. *Am. J. Phys. Anthropol.* **115**: 99–109.
- Gagneux, P., Cheriyan, M., Hurtado-Ziola, N., Brinkman van der Linden, E.C., Anderson, D., McClure, H., Varki, A., Varki, N.M. 2003. Human-specific regulation of Alpha2-6 linked sialic acids. *J. Biol. Chem.* **278**: 48245–48250.
- Gagneux, P., Moore, J.J., and Varki, A. 2005. The ethics of research on great apes. *Nature* **437**: 27–29.
- Gilad, Y., Bustamante, C.D., Lancet, D., and Pääbo, S. 2003a. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am. J. Hum. Genet.* **73**: 489–501.
- Gilad, Y., Man, O., Pääbo, S., and Lancet, D. 2003b. Human specific loss of olfactory receptor genes. *Proc. Natl. Acad. Sci.* **100**: 3324–3327.
- Gilad, Y., Wiebe, V., Przeworski, M., Lancet, D., and Pääbo, S. 2004. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol.* **2**: E5.
- Goodman, M. 1999. The genomic record of Humankind's evolutionary roots. *Am. J. Hum. Genet.* **64**: 31–39.
- Goodman, M., Grossman, L.I., and Wildma, D.E. 2005. Moving primate genomics beyond the chimpanzee genome. *Trends Genet.* **9**: 511–517.
- Grossman, L.I., Wildman, D.E., Schmidt, T.R., and Goodman, M. 2004. Accelerated evolution of the electron transport chain in anthropoid primates. *Trends Genet.* **20**: 578–585.
- Gu, J. and Gu, X. 2003. Induced gene expression in human brain after the split from chimpanzee. *Trends Genet.* **19**: 63–65.
- Gygi, S.P., Rochon, Y., Franz, B.R., and Aebersold, R. 1999. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**: 1720–1730.
- Hamann, J., Kwakkenbos, M.J., de Jong, E.C., Heus, H., Olsen, A.S., and van Lier, R.A. 2003. Inactivation of the EGF-TM7 receptor EMR4 after the *Pan-Homo* divergence. *Eur. J. Immunol.* **33**: 1365–1371.
- Hao, L. and Nei, M. 2005. Rapid expansion of killer cell immunoglobulin-like receptor genes in primates and their coevolution with MHC Class I genes. *Gene* **347**: 149–159.
- Hayakawa, T., Satta, Y., Gagneux, P., Varki, A., and Takahata, N. 2001. *Alu*-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. *Proc. Natl. Acad. Sci.* **98**: 11399–11404.
- Hayakawa, T., Angata, T., Lewis, A.L., Mikkelsen, T.S., Varki, N.M., and Varki, A. 2005. A human-specific gene in microglia. *Science* **309**: 1693.
- Hsieh, W.P., Chu, T.M., Wolfinger, R.D., and Gibson, G. 2003. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* **165**: 747–757.
- Hughes, J.F., Skaletsky, H., Pyntikova, T., Minx, P.J., Graves, T., Rozen, S., Wilson, R.K., and Page, D.C. 2005. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* **437**: 100–103.
- Hurles, M. 2004. Gene duplication: The genomic trade in spare parts. *PLoS Biol.* **2**: E206.
- International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431**: 931–945.
- Irie, A., Koyama, S., Kozutsumi, Y., Kawasaki, T., and Suzuki, A. 1998. The molecular basis for the absence of N-glycolylneuraminic acid in humans. *J. Biol. Chem.* **273**: 15866–15871.
- Johnson, M.E., Viggiano, L., Bailey, J.A., Abdul-Rauf, M., Goodwin, G., Rocchi, M., and Eichler, E.E. 2001. Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**: 514–519.
- Kaessmann, H., Wiebe, V., and Pääbo, S. 1999. Extensive nuclear DNA sequence diversity among chimpanzees. *Science* **286**: 1159–1162.
- Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigle, S., Do, H.H., Weiss, G., Enard, W., et al. 2004a. Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**: 1462–1473.
- Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W., and Pääbo, S. 2004b. A neutral model of transcriptome evolution. *PLoS Biol.* **2**: E132.
- Kim, V.N. 2005. MicroRNA biogenesis: Coordinated cropping and dicing. *Nat. Rev. Mol. Cell. Biol.* **6**: 376–385.
- King, M.C. and Wilson, A.C. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116.
- Klein, R.G. 1999. *The human career: Human biological and cultural origins*. University of Chicago Press, Chicago, IL.
- Klein, J. and Takahata, N. 2002. *Where do we come from?: The molecular evidence for human descent*. Springer, New York.
- Kouprina, N., Mullokandov, M., Rogozin, I.B., Collins, N.K., Solomon, G., Otstot, J., Risinger, J.I., Koonin, E.V., Barrett, J.C., and Larionov, V. 2004a. The SPANX gene family of cancer/testis-specific antigens: Rapid evolution and amplification in African great apes and hominids. *Proc. Natl. Acad. Sci.* **101**: 3077–3082.
- Kouprina, N., Pavlicek, A., Mochida, G.H., Solomon, G., Gersch, W.,

- Yoon, Y.H., Collura, R., Ruvolo, M., Barrett, J.C., Woods, C.G., et al. 2004b. Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. *PLoS Biol.* **2**: 653–663.
- Lander, E.S., Linton, L.M., Birren, B., Nussbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., et al., 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Li, W.H. and Saunders, M.A. 2005. News and views: The chimpanzee and us. *Nature* **437**: 50–51.
- Linardopoulou, E.V., Williams, E.M., Fan, Y., Friedman, C., Young, J.M., and Trask, B.J. 2005. Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437**: 94–100.
- Mahner, M. and Kary, M. 1997. What exactly are genomes, genotypes and phenotypes? And what about phenomes? *J. Theor. Biol.* **186**: 55–63.
- Maresco, D.L., Blue, L.E., Culley, L.L., Kimberly, R.P., Anderson, C.L., and Theil, K.S. 1998. Localization of FCGR1 encoding Fcγ receptor class I in primates: Molecular evidence for two pericentric inversions during the evolution of human chromosome 1. *Cytogenet. Cell Genet.* **82**: 71–74.
- Marshall, D., Hardman, M.J., Nield, K.M., and Byrne, C. 2001. Differentially expressed late constituents of the epidermal cornified envelope. *Proc. Natl. Acad. Sci.* **98**: 13031–13036.
- McConkey, E.H. and Goodman, M. 1997. A human genome evolution project is needed. *Trends Genet.* **13**: 350–351.
- McConkey, E.H. and Varki, A. 2000. A primate genome project deserves high priority. *Science* **289**: 1295–1296.
- . 2005. Thoughts on the future of great ape research. *Science* **309**: 1499–1501.
- Mekel-Bobrov, N., Gilbert, S.L., Evans, P.D., Vallender, E.J., Anderson, J.R., Hudson, R.R., Tishkoff, S.A., and Lahn, B.T. 2005. Ongoing adaptive evolution of *ASPM*, a brain size determinant in *Homo sapiens*. *Science* **309**: 1720–1722.
- Mello, C.C. and Conte, D.J. 2004. Revealing the world of RNA interference. *Nature* **431**: 338–342.
- Messier, W. and Stewart, C.B. 1997. Episodic adaptive evolution of primate lysozymes. *Nature* **385**: 151–154.
- Meyer-Olson, D., Brady, K.W., Blackard, J.T., Allen, T.M., Islam, S., Shoukry, N.H., Hartman, K., Walker, C.M., and Kalams, S.A. 2003. Analysis of the TCR β variable gene repertoire in chimpanzees: Identification of functional homologs to human pseudogenes. *J. Immunol.* **170**: 4161–4169.
- Modrek, B. and Lee, C. 2002. A genomic view of alternative splicing. *Nat. Genet.* **30**: 13–19.
- Nesse, R.M. and Williams, G.C. 1995. *Why we get sick: The new science of Darwinian medicine*. Times Books, New York.
- Nevo, E. 2001. Evolution of genome-phenome diversity under environmental stress. *Proc. Natl. Acad. Sci.* **98**: 6233–6240.
- Newman, T.L., Tuzun, E., Morrison, V.A., Hayden, K.E., Ventura, M., McGrath, S.D., Rocchi, M., and Eichler, E.E. 2005. A genome-wide survey of structural variation between human and chimpanzee. *Genome Res.* **15**: 1344–1356.
- Nickerson, E. and Nelson, D.L. 1998. Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees. *Genomics* **50**: 368–372.
- Nielsen, R., Bustamante, C., Clark, A.G., Glanowski, S., Sackton, T.B., Hubisz, M.J., Fedel-Alon, A., Tanenbaum, D.M., Civello, D., White, T.J., et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**: E170.
- Ohno, S. 1999. Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. *Semin. Cell. Dev. Biol.* **10**: 517–522.
- Olson, M.V. 1999. When less is more: Gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* **64**: 18–23.
- Olson, M.V. and Varki, A. 2003. Sequencing the chimpanzee genome: Insights into human evolution and disease. *Nat. Rev. Genet.* **4**: 20–28.
- . 2004. Genomics. The chimpanzee genome—a bittersweet celebration. *Science* **305**: 191–192.
- Paigen, K. and Eppig, J.T. 2000. A mouse phenome project. *Mamm. Genome* **11**: 715–717.
- Perry, G.H., Verrelli, B.C., and Stone, A.C. 2005. Comparative analyses reveal a complex history of molecular evolution for human *MYH16*. *Mol. Biol. Evol.* **22**: 379–382.
- Preuss, T.M. 2005. What is it like to be a human? In *The cognitive neurosciences*, 3rd ed. (ed. M.S. Gazzaniga), pp. 5–22. MIT Press, Cambridge, MA.
- Preuss, T.M., Caceres, M., Oldham, M.C., and Geschwind, D.H. 2004. Human brain evolution: Insights from microarrays. *Nat. Rev. Genet.* **5**: 850–860.
- Rearden, A., Phan, H., Kudo, S., and Fukuda, M. 1990. Evolution of the glycoporphin gene family in the hominoid primates. *Biochem. Genet.* **28**: 209–222.
- Ross, N.L., Mavrogiannis, L.A., Sargent, C.A., Knight, S.J., Wadekar, R., DeLisi, L.E., and Crow, T.J. 2003. Quantitation of X-Y homologous genes in patients with schizophrenia by multiplex polymerase chain reaction. *Psychiatr. Genet.* **13**: 115–119.
- Ruvolo, M. 2004. Comparative primate genomics: The year of the chimpanzee. *Curr. Opin. Genet. Dev.* **14**: 650–656.
- Sambrook, J.G., Bashirova, A., Palmer, S., Sims, S., Trowsdale, J., Abi-Rached, L., Parham, P., Carrington, M., and Beck, S. 2005. Single haplotype analysis demonstrates rapid evolution of the killer immunoglobulin-like receptor (*KIR*) loci in primates. *Genome Res.* **15**: 25–35.
- Satta, Y., Klein, J., and Takahata, N. 2000. DNA Archives and our nearest relative: The trichotomy problem revisited. *Mol. Phylogenet. Evol.* **14**: 259–275.
- Shillito, D.J., Gallup, G.G.J., and Beck, B.B. 1999. Factors affecting mirror behaviour in western lowland gorillas, *Gorilla gorilla*. *Anim. Behav.* **57**: 999–1004.
- Shu, W., Cho, J.Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G.A., Schmeidler, J., De Gasperi, R., Sosa, M.A., Ravidou, D., et al. 2005. Altered ultrasonic vocalization in mice with a disruption in the *Foxp2* gene. *Proc. Natl. Acad. Sci.* **102**: 9643–9648.
- Smit, A.F. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**: 657–663.
- Sonnenburg, J.L., Altheide, T.K., and Varki, A. 2004. A uniquely human consequence of domain-specific functional adaptation in a sialic acid-binding receptor. *Glycobiology* **14**: 339–346.
- Stedman, H.H., Kozyak, B.W., Nelson, A., Thesier, D.M., Su, L.T., Low, D.W., Bridges, C.R., Shrager, J.B., Minugh-Purvis, N., and Mitchell, M.A. 2004. Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature* **428**: 415–418.
- Steinmetz, L.M. and Davis, R.W. 2004. Maximizing the potential of functional genomics. *Nat. Rev. Genet.* **5**: 190–201.
- Stone, A.C., Griffiths, R.C., Zegura, S.L., and Hammer, M.F. 2002. High levels of Y-chromosome nucleotide diversity in the genus *Pan*. *Proc. Natl. Acad. Sci.* **99**: 43–48.
- Szabo, Z., Levi-Minzi, S.A., Christiano, A.M., Struminger, C., Stoneking, M., Batzer, M.A., and Boyd, C.D. 1999. Sequential loss of two neighboring exons of the tropoelastin gene during primate evolution. *J. Mol. Evol.* **49**: 664–671.
- Tang, G. 2005. siRNA and miRNA: An insight into RISCs. *Trends Biochem. Sci.* **30**: 106–114.
- Uddin, M., Wildman, D.E., Liu, G., Xu, W., Johnson, R.M., Hof, P.R., Kapatos, G., Grossman, L.I., and Goodman, M. 2004. Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc. Natl. Acad. Sci.* **101**: 2957–2962.
- Varki, A. 2000. A chimpanzee genome project is a biomedical imperative. *Genome Res.* **10**: 1065–1070.
- Varki, A., Wills, C., Perlmutter, D., Woodruff, D., Gage, F., Moore, J., Semendeferi, K., Benirschke, K., Katzman, R., Doolittle, R., et al. 1998. Great Ape Phenome Project? *Science* **282**: 239–240.
- Vermeech, J.R., Petit, P., Kermouni, A., Renaud, J.C., Van Den Berghe, H., and Marynen, P. 1997. The IL-9 receptor gene, located in the Xq/Yq pseudoautosomal region, has an autosomal origin, escapes X inactivation and is expressed from the Y. *Hum. Mol. Genet.* **6**: 1–8.
- Walhout, A.J., Reboul, J., Shtanko, O., Bertin, N., Vaglio, P., Ge, H., Lee, H., Doucette-Stamm, L., Gunsalus, K.C., Schetter, A.J., et al. 2002. Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr. Biol.* **12**: 1952–1958.
- Wang, H.Y., Tang, H., Shen, C.K., and Wu, C.I. 2003. Rapidly evolving genes in human. I. The glycoporphins and their possible role in evading malaria parasites. *Mol. Biol. Evol.* **20**: 1795–1804.
- Wang, X., Thomas, S.D., and Zhang, J. 2004. Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. *Hum. Mol. Genet.* **13**: 2671–2678.
- Watanabe, H., Fujiyama, A., Hattori, M., Taylor, T.D., Toyoda, A., Kuroki, Y., Noguchi, H., BenKahla, A., Lehrach, H., Sudbrak, R., et al. 2004. DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* **429**: 382–388.
- Winter, H., Langbein, L., Krawczak, M., Cooper, D.N., Jave-Suarez, L.F., Rogers, M.A., Praetzel, S., Heidt, P.J., and Schweizer, J. 2001. Human type I hair keratin pseudogene *phihHaA* has functional orthologs in the chimpanzee and gorilla: Evidence for recent inactivation of the human gene after the *Pan-Homo* divergence. *Hum. Genet.* **108**: 37–42.
- Wood, B. and Collard, M. 1999. Anthropology—the human genus. *Science* **284**: 65–66.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman,

- M.V., and Romano, L.A. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**: 1377–1419.
- Yang, Z. and Bielawski, J.P. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**: 496–503.
- Yu, N., Jensen-Seaman, M.I., Chemnick, L., Kidd, J.R., Deinard, A.S., Ryder, O., Kidd, K.K., and Li, W.H. 2003. Low nucleotide diversity in chimpanzees and bonobos. *Genetics* **164**: 1511–1518.
- Yunis, J.J. and Prakash, O. 1982. The origin of man: A chromosomal pictorial legacy. *Science* **215**: 1525–1530.
- Yunis, J.J., Sawyer, J.R., and Dunham, K. 1980. The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. *Science* **208**: 1145–1148.
- Zhang, J. 2003. Evolution of the human *ASPM* gene, a major determinant of brain size. *Genetics* **165**: 2063–2070.

- Zhang, J., Webb, D.M., and Podlaha, O. 2002. Accelerated protein evolution and origins of human-specific features. *FOXP2* as an example. *Genetics* **162**: 1825–1835.

## Web site references

- <http://www.genome.gov/10002851>; NHGRI news release on sequencing priorities.
- <http://www.ncbi.nlm.nih.gov/projects/SNP/>; NCBI SNP database.
- <http://origins.ucsd.edu/gapp1.html>; Museum of Comparative Anthropogeny.

**Genome Research 19:** 1742–1751 (2009)**Chromatin poises miRNA- and protein-coding genes for expression**

Artem Barski, Raja Jothi, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones, and Keji Zhao

Several errors in equations appear on page 1749 in the section MicroRNA Promoter Prediction. The corrected equations are provided below:

For Pol II, the tag density profile at position  $i$  is given by

$$S(i) = \frac{1}{h} \sum_{j=i-3h}^{i+3h} K[(i-j)/h].C(j+75).$$

For H3K4me3 and H2A.Z, the tag density profiles at position  $i$  are given by

$$S(i) = \frac{1}{h} \sum_{j=i-3h}^{i+3h} K[(i-j)/h].C(j+125)$$

and

$$S(i) = \frac{1}{h} \sum_{j=i-3h}^{i+3h} K[(i-j)/h].C(j),$$

respectively.

The authors apologize for any confusion this may have caused.

**Genome Research 15:** 1746–1758 (2005)**Comparing the human and chimpanzee genomes: Searching for needles in a haystack**

Ajit Varki and Tasha K. Altheide

The URL in the last line of the footnote to Table 1 on page 1747 has been changed. The correct URL should now read:

<http://carta.anthropogeny.org/moca/about>.



## Comparing the human and chimpanzee genomes: Searching for needles in a haystack

Ajit Varki and Tasha K. Altheide

*Genome Res.* 2005 15: 1746-1758

Access the most recent version at doi:[10.1101/gr.3737405](https://doi.org/10.1101/gr.3737405)

---

### Related Content

#### Errata

[Genome Res. December , 2009 19: 2343](#)

### References

This article cites 125 articles, 44 of which can be accessed free at:  
<http://genome.cshlp.org/content/15/12/1746.full.html#ref-list-1>

Articles cited in:

<http://genome.cshlp.org/content/15/12/1746.full.html#related-urls>

### License

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

Affordable, Accurate  
Sequencing.



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---