A White Paper Advocating Complete Sequencing of the Genome of the Common Chimpanzee, *Pan Troglodytes*

Maynard V. Olson,[a,b] Evan E. Eichler,[c] Ajit Varki,[d] Richard M. Myers,[e] Joseph M. Erwin,[f] and Edwin H. McConkey[g]

I. Introduction

The chimpanzee is our closest living relative. There are two species of chimps, the common chimpanzee (*Pan troglodytes*) and the so-called "pygmy chimpanzee," or bonobo (*Pan paniscus*). The common chimp has a wider geographic range, which traverses equatorial Africa, as well as a much larger population both in the wild and in captivity. The bonobo's range is limited to a region of central Africa south of the Congo River, which is presumably the geographic barrier that led to speciation. Both chimps are equally closely related to humans: human-chimp divergence in nucleotide sequence is 1.2% (Fujiyama *et al.* 2002), while the divergence between the two chimp species is approximately half that value. Commonly cited divergence times are 5-6 Myr for the human-chimp split and 2.5 Myr for the common chimp-bonobo split. These estimates are largely based on a molecular clock that has been calibrated over substantially longer intervals of primate evolution; however, fossil evidence supporting the existence of bipedal hominids as early as 5-6 Mya are generally consistent with the molecular time scales

The close kinship between the human and the chimp is inherently fascinating to scientists and non-scientists alike. However, the case for giving high priority to the sequencing of the chimp genome rests on the extraordinary relevance of chimp-human comparative biology to our understanding of human health and disease. Genomic tools, rooted in a complete-genome sequence, now offer the opportunity to explore those differences systematically at the molecular level. Seizing this magnificent opportunity ranks among the highest basic science priorities in all of biomedical research.

II. Specific biological rationales for the utility of new sequence data

A. Improving human health. How will the genomic sequence of an organism inform our understanding of human health and disease? What, if any, is the relevance to the development of innovative and improved methods of diagnosis, treatment or prevention?

The strength of the proposal to acquire a complete genome sequence for the chimp is precisely its direct relevance to "the development of innovative and improved methods of diagnosis, treatment or prevention" of human disease. No other animal offers the same qualitatively novel opportunity to expand our understanding of human biology. We must acknowledge that the sequencing of the chimpanzee genome, in contrast to more conventional alternatives, would be a high-risk-high-gain choice. In advance, one can only speculate as to whether or not we will be able to interpret the comparative data on humans and chimpanzees in ways that are medically relevant. The same concern dominated early reactions to the Human Genome Project. Like the sequencing of the human genome, comparative analysis of the chimp and human genomes will open a vast new frontier in which to explore human biology. We need confidence that the scientific community will rally to the exploration of this frontier and will develop the needed conceptual and experimental tools as the science progresses.

That said, consider one simple argument as to how the chimpanzee sequence may lead to altogether novel routes to improved treatments for human disease. This argument is based on a particular model for the molecular changes that typically underlie the emergence of innovative evolutionary lineages such as the

one that led to modern humans. The model presupposes that a major feature of such lineages is that they experience substantial amounts of genetic loss. Although counter-intuitive, this "less-is-more" view of the evolution of novelty has much in its favor (Olson 1999). Of course, there must be genuine molecular innovations that allow an innovative lineage to break out of the stable phenotypic patterns that characterize most taxonomic groups during most of their evolutionary history. However, medically speaking, the sequelae to these early molecular innovations are likely to be of greater importance than the innovations themselves. Many of the molecular changes that followed the evolutionary commitment of humans to break out of established patterns of primate biology are likely to have involved genetic loss. There are four arguments in favor of this hypothesis:

1. The superficial phenotypes of humans and apes differ in ways that are suggestive of genetic loss. Examples include delayed maturation, decreased muscle strength, and loss of body hair. It is likely that less readily apparent genetic differences between humans and apes follow the same pattern.

2. There is an analogy with "island evolution." The essence of evolutionary innovation is that the innovative lineage suddenly has access to a much expanded habitat. In the case of humans, our species broke out of the limitations of life under the rain-forest canopy and thereby gained access, ultimately, to most of the earth's terrestrial ecosystems. The situation is analogous in some respects to island evolution. Suddenly, as occurs during the initial population of geologically new islands, humans confronted tremendous biological opportunity without competition from other species with similar evolutionary strategies. We know that the rapid evolution that occurs on islands under partially analogous circumstances involves extensive genetic loss, presumably because genetic loss is the only high-bandwidth path toward phenotypic change. Admittedly, the situation in human evolution differed from classical island scenarios—most notably, with respect to predation—but the fundamentals of the argument do not depend on the details of the adaptive challenge faced by an organism that is a biological pioneer. The human species has been referred to as a "hastily made-over ape." Our knowledge of genetic change in model organisms suggests that plausible paths to a "hasty makeover" will always rely heavily on genetic loss.

3. There is greatly restricted genetic diversity in the human relative to the great apes (Kaessmann *et al.* 2001). Hence, it is likely that many deleterious alleles have been fixed in the human lineage simply as a result of genetic drift during severe population bottlenecks. It is a truly remarkable finding that small contemporary populations of chimpanzees, occupying a highly restricted range in Central Africa, display several times the nucleotide diversity that is found in humans throughout the world.

4. The hypothesis fits the minimal direct data that are available. The two examples of substantial biochemical differences between humans and chimps whose genetic basis has been determined both involve genetic loss on the human lineage. One of these examples involves the absence of Neu5Gc, a sialic acid, in the glycan component of many human cell-surface proteins due to a fixed loss-of-function mutation in the human gene encoding the last enzyme in the pathway leading to Neu5Gc (Chou *et al.* 1998; Satta *et al.* 2001). The other involves a mutation in a receptor that recognizes sialic acids (Angata *et al.* 2001). In both cases, the function missing in the human lineage is highly conserved in all other primates. Similar comments apply to other biochemical processes that are absent in humans but whose genetic basis is still unknown: these examples include the absence of uricase activity in humans, as well as the absence of vitamin C and alpha-galactose synthesis (Gagneux and Varki 2001).

Consider the implications of the above model. First, genetic-loss events are the most easily recognized of genomic differences. Most disease-causing mutations in humans are hypomorphic or null alleles of genes. These changes have proven relatively easy to detect since most of them involve conspicuously deleterious alterations of coding DNA or splice signals. Hence, the "less-is-more" view of human evolution is a testable hypothesis, which predicts that most of the obviously functional variation between

the chimp and human genomes will involve hypomorphic or null mutations that have become fixed in the human lineage.

A second implication is that it is likely that most such changes will have some adverse consequences for human biology. The abandonment of established components of primate biology during the "greedy" pursuit of a new lifestyle is expected always to be a two-edged sword. Fixation of deleterious alleles due to genetic drift, by definition, is expected to have adverse consequences. Indeed, it is plausible that much of the distinctive pattern of human disease—our propensity to obesity, diabetes, cardiovascular disease, epithelial cancers, and neurodegenerative disease—is the downside of the rapid evolutionary success of the human lineage. Finally, it should be noted that the intact primate functions that we have lost would provide a direct biochemical model of how to remedy, within the overall constraints of primate biology, the human defects. This view of the common human diseases suggests that there may exist a presently unexploited source of ideas about how to tilt human biology in healthier directions.

The point of developing this admittedly speculative argument is simply to illustrate that the comparative genomics of the human and the chimp may lead in truly novel directions that are of central relevance to human health. The same, quite simply, cannot be said about competing proposals. As a class, they offer incremental strengthening of trends in biomedical research that are already well established. The potential for fundamental change in the way we think about human health and disease through the sequencing of another distant mammalian relative is quite slight.

To illustrate the scope of health-related research that may benefit from the comparative genomics of the human and the chimpanzee, Table 1 summarizes the surprising number of differences in the pattern of disease between these two closely related primates that have already been reported. References to the supporting literature can be found in Varki (2000).

B. Informing human biology. How will the genomic sequence of a particular organism lead to a better understanding of biological function in the human?

Straightforward extensions of the arguments concerning health-related research suggest that human-chimp comparisons have the potential to clarify fundamental aspects of human biology. For example, there is intense interest in the question of how the human brain acquired its extraordinary capabilities. Similarly, the basis of the dramatic developmental delay observed in humans relative to other primates is of major interest. We know so little about the molecular basis of these aspects of human biology that we can only speculate as to whether comparison of the human and chimp genomes will lead to new insights. However, science moves forward by exploring new territory, not by making worst-case assumptions about our ability to interpret novel sources of data.

Certainly, some of the more dramatic differences between the human and chimpanzee genomes will lend themselves to immediate study. These differences involve regions of the hominoid genome that are known to have evolved much more rapidly than "generic" DNA. Their systematic identification will depend on sequencing the genome of a closer human relative than the macaque or baboon, two species that are often mentioned as priority sequencing targets. Processes such as Y chromosome evolution, pericentromeric duplication, subtelomeric rearrangements and centromere repositioning occur so rapidly that they can only be studied effectively by comparing the closest available relatives to the human. As an example, ape genome sequence is most frequently used to determine the timing and movement of recent segmental duplications that are associated with chromosomal rearrangement disorders (*e.g.*, the Velocardiofacial/DiGeorge, Prader-Willi, and Smith Magenis syndromes), pericentromeric duplications and subtelomeric rearrangements. These duplicated segments of the genome, all of which display high rates of evolutionary change, comprise an estimated 5-7% of the human genome (Bailey *et al.* 2001; International Human Sequencing Consortium 2001; Eichler 2001). Most available sequence data suggest

that the bulk of the large, nearly exact segmental duplications in the human genome occurred during the emergence of humans and the great apes. Targeted analysis of these regions in a variety of ape species has been used to reconstruct the ancestral origin of several segmental duplications and to infer the series of events that have created this duplication architecture in humans (Jackson *et al.* 1999; Johnson *et al.* 2001; Monfouilloux *et al.* 1998; Orti *et al.* 1998; Zimonjic *et al.* 1997). Large-scale sequence from a closely related primate is necessary in order to survey the structure and organization of these regions since many of the duplications or sites of rearrangement involve segments of 100 kbp or larger. Comparative sequencing will provide insight into the underlying mechanisms that predisposed to duplication-mediated rearrangements associated with human genetic disease (Lupski 1998).

C. Informing the human sequence. How will the genomic sequence of a particular organism lead to a better description of the functions of specific sequence features of the human genome?

There are several ways that the chimp sequence will help with the interpretation of the human sequence. All are qualitatively different from the arguments that apply to more distantly related organisms, once again illustrating the unique scientific opportunity provided by chimp sequencing. First, a carefully executed chimp genome sequence would provide an ideal tool for cleaning up the human sequence. How much cleaning up the human sequence will need is unknown, but the likelihood is that the current phase of the Human Genome Project will leave behind a number of unresolved problems, including "tangles" of low-copy repeats that are known to evolve rapidly. There is compelling evidence that numerous differences involving many megabase pairs of DNA exist between the genomes of humans and chimpanzee (Johnson *et al.* 2001; Bailey *et al.* 2002; Samonte and Eichler 2002; Mefford and Trask 2002). Sequencing of the chimpanzee genome would immediately highlight these differences and ensure correct annotation and assembly of these regions in both genomes. The rapidity of the large-scale genomic rearrangements that give rise to these differences make it unlikely that sequence from a more distantly related primate species would be similarly informative: too many sequential events would have occurred to allow reconstruction of the stepwise changes that led to the current, highly diverged states.

There will be other gaps in the human sequence that have more trivial origins (*e.g.*, errors in tiling-path construction and assembly, or simply gaps that were never filled). Because of the intense interest in all major differences between the chimp and human sequences—and the expectation that most differences will be slight—there will be a powerful scientific incentive to track down the basis of each major difference encountered. This process will contribute to the functional analysis of the human genome sequence simply by improving its quality.

Another powerful use of the chimp sequence will be to provide a first-order determination of which allele is likely to be the ancestral one in bulk collections of human SNPs. Recently published SNP studies emphasize the value of genomic sequence from the chimpanzee for this purpose (Kaessmann *et al.* 2001). Proper determination of the ancestral allele requires more extensive phylogenetic comparisons—minimally, comparison of three closely related species whose order of divergence is known. However, if one excludes highly mutable positions—predominantly CG dinucleotides—the chimp allele is likely to reflect the ancestral sequence for the overwhelming preponderance of human SNPs. This argument follows from the expectation that most SNPs are selectively neutral, the low level of divergence between the chimp and human genomes, and the small fraction of sites in the human genome that are sufficiently polymorphic to display common SNPs. Knowledge of the ancestral allele is critical to the development of models of human population history. These models, in turn, are both of fundamental interest and are also central to our understanding of patterns of disease across the contemporary human population. For example, improved models of human population history will be essential if we are to understand the complex patterns of linkage disequilibrium that are presently of major interest in the study of genetic susceptibility to common human diseases.

Finally, chimp-human sequence comparisons will provide a comprehensive view of patterns of mutation in the human genome. Pairwise comparisons between the chimp and the human will immediately reveal composite mutation rates for the chimp and human lineages. These composite rates will be adequate for many purposes since there is little reason to think that patterns of mutation differ in the two species. For more detailed studies, it will be necessary to assign mutations separately to the two lineages, a goal that could be achieved by even modest whole-genome sampling of a third ape—presumably the gorilla or the orangutan. Unexpected discrepancies in the mutability of different types of sites (*e.g.*, those leading to synonymous vs. non-synonymous changes in coding regions) will provide insight into the levels and targets of selection since the human-chimp divergence. It is increasingly clear that significant changes in the pattern of both adaptive and balancing selection can be detected between many pairs of closely related species such as human and chimpanzee (Johnson *et al.* 2001). Unusual patterns of selection in the two lineages may pinpoint regions critical for the adaptive differences between the two species. Analysis of the complete genomes will be important for this purpose since it will define baseline levels of change, which are due either to neutral processes or purifying selection. The identification of genes subject to unusual patterns of selection has the potential to become a major source of functional annotation of the human genome.

D. Providing a better connection between the sequences of non-human organisms and the human sequence. How will the genomic sequence of a particular organism increase our ability to identify orthologs in the sequences of well-studied model organisms and how will that deepen our understanding of the human sequence?

Other than by cleaning up the human sequence—and, thereby, helping with the development of a definitive list of human genes—the chimp sequence is not expected to contribute to ortholog detection. On the other hand, the discovery of human genes that lack orthologs in the chimpanzee would be of major interest even if such genes are, as expected, exceedingly rare.

E. Expanding our understanding of basic biological processes relevant to human health, *e.g.* developmental biology, neurobiology.

This question has been addressed above under questions II. A. and II. B.

F. Providing additional surrogate systems for human experimentation, *e.g.* new disease models, improved opportunities for drug testing, or other medical procedures, such as transplantation.

The chimp can no longer be considered an experimental organism. Society is committed to maintaining colonies under conditions that maximize the quality of life of the animals in them. This commitment includes lifelong veterinary care. Despite the stringent protections that will govern use of these animals, we should not discount the potential value of the clinical experience we will gain about aging chimpanzees simply as a result of providing them with veterinary care.

G. Facilitating the ability to do experiments, *e.g.* "direct" genetics or positional mapping, in additional organisms.

With the caveat developed above in answer to question II. F., the chimp is not an appropriate organism for these purposes.

H. Expanding our understanding of evolutionary processes (biological innovation, selection) in general, and human evolution in particular.

These issues are central to the advantage of the chimpanzee for comparative genomics. While they have been addressed extensively above under questions II. A., II. B., and II. C., a brief summary is appropriate here. Comparative analysis of the human and chimpanzee genomes will contribute to 1) Understanding the behavior of genomic regions that evolve rapidly (low-copy repeats, and subtelomeric and pericentromeric regions) and their relevance to disease and large-scale genomic rearrangements; 2) Identification of patterns of exceptional selection (both positive and balancing selection) relative to the baseline established by neutral processes and purifying selection; 3) Identification of gene-loss events and their potential impact on human phenotypes; and 4) Determination of the ancestral state of SNPs. Finally, the human-chimp comparison will allow detailed molecular analysis of the 13 large-scale rearrangements between the two genomes that are known from high-resolution studies of chromosome-banding patterns (Yunis and Prakesh, 1982).

III. Strategic issues in acquiring new sequence data

A. The demand for the new sequence data. What is the size of the research community that will use it? What is the community's enthusiasm for having the sequence? Will the new sequence data stimulate the expansion of the research community?

The situation with the chimpanzee differs from that for typical model organisms. Typically, there is a well-defined list of laboratories whose members work directly on the organism and would be heavy users of the sequence. Then, there is a larger group whose plans to use the sequence are vague, but whose loyalty to the organism leads them to lobby for making it a priority sequencing objective. In contrast, there are relatively few laboratories with a major focus on molecular studies of the chimpanzee.

The arguments cited above—most immediately those involving analysis of human genetic variation—suggest that the chimpanzee sequence would rapidly become a routine tool for most human molecular geneticists, which is a huge and rapidly growing community. To the extent that the more scientifically ambitious uses of the chimpanzee genome sequence that were suggested in response to question II. A. materialize, the sequence has the potential of becoming an extremely valuable tool in biomedical research. There is no other genome sequence, beyond those of the human and the mouse, that has obvious potential to have comparable impact.

An indication of the importance of the chimpanzee genome sequence is the response that the proposal to acquire one often engenders among major figures in molecular biology. For example, included in the group of 27 scientists who signed a recent letter to *Science* that advocates a high priority for primate sequencing—with a special emphasis on the chimpanzee—were Francis Crick, George Palade, and Arno Motulsky (McConkey *et al*. 2000). Many of these individuals, even if they do not themselves envision becoming immediate users of the sequence, regard the prospect of carrying out detailed comparative analysis of the human and the chimpanzee to be one of the great future challenges in all of science. While we have chosen to focus on the tangible, short-term utility of the chimpanzee sequence, this instinctive feeling by leaders in biology that a chimpanzee-sequencing project would be a step toward studying one the most exciting, unexplored dimensions of the natural world should not be altogether discounted. If the case for the project were presented effectively, the general public would become similarly engaged. An indication of the potential for strong public interest is that Ajit Varki, who has pioneered the studies of differences in the glycobiology of humans and chimpanzees, reports having been asked, during the last year alone, to comment on human-chimp comparisons by reporters from Brazil, Spain, Switzerland,

Germany, the United Kingdom, Japan, and several major publications in the United States (A. Varki, personal communication).

B. The suitability of the organism for experimentation. What are the basic properties of the organism that affect its ability to be studied in the laboratory (*e.g.* availability, ability to be cultured and propagated in the laboratory, generation time)? Are mutants available with defined phenotypes? How will the new sequence data enhance the experimental use of the organism? What other genomic resources and technologies (*e.g.* gene transfer, ability to go from molecule to mutation) are available that will allow the new sequence information to be effectively used?

This question has been addressed thoroughly above under question II. F.

C. The rationale for the complete sequence of the organism. Why would the complete sequence be more useful than the sequences of specific regions, or only the coding sequences, or only ESTs? Are there alternative ways to get the necessary information?

A complete, accurate genome sequence is essential. This point is obvious in view of the arguments presented above (*e.g.*, the focus on close comparisons with the nearly identical human sequence and applications to the analysis of SNPs). A complete sequence is also required to test the classical hypothesis that major evolutionary changes during the human-chimp divergence may have been due to regulatory mutations (King and Wilson 1975). Finally, the very obstacles to carrying out experimental manipulation of the chimpanzee argue for a complete genome sequence. Even more than in the case of the human, where classical genetic methods continue to play an important role, the comparative analysis of chimps and humans will be entirely driven by the genome sequence.

D. The cost of sequencing the genome and the state of readiness of the organism's DNA for sequencing. What is the size of the genome? What quality of sequence product is needed (finished sequence? draft? full shotgun?)? What sequencing strategy will be used? Is suitable DNA readily available?

There are both routine and more ambitious answers to this question. The routine answer is that the genome size and availability of DNA and other genomic resources (*e.g.*, BAC libraries) all differ in only minor respects from the situation for the human. As argued above in answer to question III. C., a high quality sequence is necessary.

The more ambitious answer is that the chimpanzee offers a magnificent opportunity to break new technical ground in large-scale-sequencing technology. *De novo* sequencing of new genomes that are highly dissimilar to previously sequenced genomes is already a niche activity relative to resequencing projects. In the future, this point will increasingly be true. Yet there is a risk that the choice of sequencing objectives will continue to push large-scale sequencing practices toward refinement of a technical model that will be of steadily decreasing relevance to biomedical research. In contrast, the chimpanzee project offers the potential to explore, at minimal risk, alternative strategies for meeting both the technical and managerial challenges of large-scale sequencing projects.

Technically, it makes sense for the acquisition of the chimpanzee sequence to lean heavily on the human-genome sequence, while retaining an easy ability to target *de novo* analysis to regions that have diverged from the human sequence to an unusual extent. This goal could be achieved in a variety of ways, whose discussion is beyond the scope of this White Paper. However, the key principle is that alignment of fragmentary chimpanzee sequence with the human sequence will allow highly directed strategies for tiling the genome with clones and achieving sequence closure in individual regions. Precise costs are difficult to estimate both because the project would undoubtedly be able to adopt more efficient strategies than those in current large-scale use and because the cost of the component steps of all sequencing

strategies are dropping. However, a reasonable goal would be to keep the total cost of the project to $100 M.

E. Are there other (partial) sources of funding available or being sought for this sequencing project?

Any project funded through U.S. sources should engage the international community. The existence of a significant level of chimpanzee sequencing in Japan is already a tangible indication that some funding from international sources is plausible. The reasons for this international interest are simply those articulated in this White Paper, particularly the rationales related to the high level of general scientific interest in human-chimp comparisons. Any sensible approach to sequencing the chimpanzee genome would involve a significant component of map-driven, directed sequencing that leveraged the availability of the human sequence. Hence, it should be relatively easy to negotiate international collaborations on the project.

****

[a] Development of this white paper was coordinated by MVO. The author list is an effort to acknowledge individuals who made major contributions to its preparation. Not all of the listed authors had the opportunity to approve of the final text.

[b]University of Washington

[c]Case Western Reserve University

[d]University of California, San Diego

[e]Stanford University

[f]Biqual, Inc., and Foundation for Comparative and Conservation Biology

[g]University of Colorado

REFERENCES

Angata, T, Varki NM and Varki, A (2001) A Second Uniquely Human Mutation Affecting Sialic Acid Biology. J.Biol.Chem. 276:40282-87.

Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE (2001) Segmental duplications: organization and impact within the current human genome project assembly. Genome Res 11:1005-17.

Bailey, JA, Yavor, AM, Misceo, D, Horvath, JE, Arichdiacono, N, Schwartz, S, Rocchi, M Eichler EE (2002) Human specific duplication and mosaic transcripts: the recent paralogous structure of human chromosome 22. Am J. Hum Genet. 70:83-100

Chou, H-H, Takematsu, H, Diaz, S, Iber, J, Nickerson, E, Wright, K, Muchmore, EA, Nelson, DL, Warren, ST and Varki, A (1998) A mutation in human CMP-sialic acid hydroxylase occurred after the *Homo-Pan* divergence. Proc.Nat'l.Acad.Sci.U.S.A. 95:11751-56.

Eichler EE (2001) Segmental duplications: what's missing, misassigned, and misassembled- and should we care? Genome Res 11:653-6.

Fujiyama A, Watanabe H, Toyoda A, Taylor TD, Itoh T, Tsai SF, Park HS, Yaspo ML, Lehrach H, Chen Z, Fu G, Saitou N, Osoegawa K, de Jong PJ, Suto Y, Hattori M, Sakaki Y (2002) Construction and analysis of a human-chimpanzee comparative clone map. Science 295:131-4.

Gagneux P and Varki A (2001) Genetic Differences Between Humans and Great Apes. Molecular Phylogenetics and Evolution 18:2-13.

International Human Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860-920

Jackson MS, Rocchi M, Thompson G, Hearn T, Crosier M, Guy J, Kirk D, *et al* (1999) Sequences flanking the centromere of human chromosome 10 are a comp lex patchwork of arm-specific sequences, stable duplications, and unstable sequences with homologies to telomeric and other centromeric locations. Hum Mol Genet 8:205-215

Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, Eichler EE (2001) Positive selection of a gene family during the emergence of humans and the great apes. Nature in press

Kaessmann H, Wiebe V, Weiss G, Paabo S (2001) Great ape DNA sequences reveal a reduced diversity and an expansion in humans. Nat Genet 27:155-6.

King M and Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188:107-116.

Lupski JR (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 14:417-22.

McConkey EH *et al*. (2000) A primate genome project deserves high priority. Science 289:1295.

Meffort, HC, Trask, BJ (2002) The complex structure and dynamic evolution of human subtelomeres. Nature Rev. Genet. 3: 91-103

Monfouilloux S, Avet-Loiseau H, Amarger V, Balazs I, Pourcel C, Vergnaud G (1998) Recent human-specific spreading of a subtelomeric domain Genomics 51:165-76

Olson MV (1999) When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet. 64:18-23.

Orti R, Potier MC, Maunoury C, Prieur M, Creau N, Delabar JM (1998) Conservation of pericentromeric duplications of a 200-kb part of the human 21q22.1 region in primates. Cytogenet Cell Genet 83:262-5

Samonte RV, Eichler, EE (2002) Segmental duplications and the evolution of the primate genome. Nature Rev. Genet. 3: 65-72.

Satta Y, Gagneux P, Varki A, and Takahata N (2001). Alu-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. Proc.Nat'l.Acad.Sci.U.S.A. 98:11399–404.

Varki, A (2000) A Chimpanzee Genome Project is a Biomedical Imperative. Genome Research 10:1065-70.

Yunis JJ, Prakash O (1982) The origin of man: a chromosomal pictorial legacy. Science 215:1525-30.

Zimonjic D, Kelley M, Rubin J, Aaronson S, Popescu N (1997) Fluorescence in situ hybridization analysis of keratinocyte growth factor gene amplification and dispersion in evolution of great apes and humans. Proc Natl Acad Sci USA 94:11461-65.

Table 1. Apparent differences between humans and great apes in the incidence or severity of medically important conditions (excluding differences explained by obvious anatomical differences).

Medical Condition	Humans	Great Apes
Definite		
HIV progression to AIDS	Common	Very rare
Influenza A symptomatology	Moderate to severe	Mild
Hepatitis B/C late complications	Moderate to severe	Mild
P. falciparum malaria	Susceptible	Resistant
Menopause	Universal	Rare
Likely		
E. coli K99 gastroenteritis	Resistant	Sensitive?
Alzheimer's disease pathology	Complete	Incomplete
Coronary atherosclerosis	Common	Uncommon
Epithelial cancers	Common	Rare
Speculative		
Menstrual blood loss amount?	Variable	Lower
Early fetal wastage	High	Low?
Bronchial asthma	Common	Rare?
Systemic lupus erythematosus	Relatively common	Rare?
Rhematoid arthritis	Relatively common	Rare?
Acne vulgaris	Common	Rare?
Major psychoses	Common	Rare?