In Transition: Primate Genomics at a Time of Rapid Change

Jeffrey Rogers

Abstract

The field of nonhuman primate genomics is undergoing rapid change and making impressive progress. Exploiting new technologies for DNA sequencing, researchers have generated new whole-genome sequence assemblies for multiple primate species over the past 6 years. In addition, investigations of within-species genetic variation, gene expression and RNA sequences, conservation of non-protein-coding regions of the genome, and other aspects of comparative genomics are moving at an accelerating speed. This progress is opening a wide array of new research opportunities in the analysis of comparative primate genome content and evolution. It also creates new possibilities for the use of nonhuman primates as model organisms in biomedical research. This transition, based on both new technology and the new information being generated in regard to human genetics, provides an important justification for reevaluating the research goals, strategies, and study designs used in primate genetics and genomics.

Key Words: annotation; genetic models of disease; genetic variation; genome assembly; primate

Introduction

The field of primate genomics is entering a period of critical transition. This transition is driven both by rapid advances in genomic technologies and by outstanding progress in human genomics, which is producing a new understanding of the genetic basis of risk for human disease. Study of the genetics of nonhuman primates is benefitting and will continue to benefit remarkably from these advances in methods and analytical strategies. This article is a brief review of the current state of research regarding the genomic analysis of nonhuman primates, with emphasis on application of that information to research related to human health and disease. This field is moving much too fast for a comprehensive review, and therefore the purpose of this review is to provide one perspective on the present state and near-term prospects for the field.

Whole-Genome Assemblies

A number of nonhuman primate genomes have been sequenced and analyzed in detail. The first species to have its genome sequenced and published was the chimpanzee (*Pan troglodytes*), chosen primarily because it is the living species most closely related to humans. Whole-genome comparisons of the chimpanzee sequence with the human sequence (Chimpanzee Sequencing and Analysis Consortium 2005) facilitated the process of identifying unique changes in the human genome that may underlie the species-specific adaptations of our species. More recently, publication of the orangutan (Locke et al. 2011), gorilla (Scally et al. 2012), and bonobo (Prufer et al. 2012) genomes completed the initial description of genomic diversity and similarity across the great apes. Access to these genome sequences has increased opportunities to investigate the genetic basis of human uniqueness, with some intriguing results. There is, of course, significant interest among both researchers and the general public in the biological basis of unique human behavioral and cognitive capacities, such as spoken language, and the evolutionary changes in the brain that underlie those advanced human traits. Sequence comparisons now implicate specific genes or RNA-coding sequences as playing a role in the evolution of human brain size (O'Bleness et al. 2012; Pollard et al. 2006), differences in synapse formation (Charrier et al. 2012), and language capacity (Preuss 2012). Specific genes have also been associated with unique human phenotypes beyond the brain and behavior (Dumas et al. 2007; Prabhakar et al. 2008). At the time of this writing, we also anticipate the impending publication of a series of analyses describing genomic variation within and between great ape species (Great Ape Genome Project). These data will significantly extend information about within-species diversity in chimpanzees, gorillas, and orangutans.

Although the sequencing of the genomes of the great apes provides significant new information about the evolution of the human genome, the sequencing of other primate genomes is having more impact on studies related to human health and disease. Among the apes, only chimpanzees have

Jeffrey Rogers, PhD, is Associate Professor in the Human Genome Sequencing Center and the Department of Molecular and Human Genetics at Baylor College of Medicine in Houston, Texas.

Address correspondence and reprint requests to Dr. Jeffrey Rogers, Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030 or email jr13@bcm.edu.

ILAR Journal, Volume 54, Number 2, doi: 10.1093/ilar/ilt042

been regularly used for experimental work related to specific human diseases, and most of that work is not substantially affected by access to DNA sequence data. As will be discussed below, genomic information about macaques and other monkey species is creating new opportunities for studies of both the genetic basis of risk for human disease and the genomic correlates of disease processes in primate models, such as the quantification of changes in gene expression in response to viral infection (Bosinger et al. 2013). Direct application of genomic information to questions of disease risk and pathogenesis will be much more common in macaque and other nonhominoid models of disease than in studies of great apes.

The second nonhuman primate genome to be sequenced was that of the rhesus macaque, *Macaca mulatta* (Gibbs et al. 2007). Given its critical importance to a wide range of topics within biomedical research, this was a natural choice for early attention. Furthermore, the decision to sequence a rhesus macaque with genetic ancestry in India, as opposed to China, was based on the relative use of Indian- versus Chinese-origin animals in US research colonies. Subsequent to the sequencing of that first rhesus, other groups have sequenced the genomes of Chinese-origin rhesus macaques, *M. fascicularis* (Fang et al. 2011; Yan et al. 2011), Mauritian cynomolgus macaque, *M. fascicularis* (Ebeling et al. 2011), Malaysian cynomolgus macaque (Higashino et al. 2012), and Vietnamese cynomolgus macaque (Yan et al. 2011).

All these great ape and macaque genome sequences are valuable as tools for research, but the quality and completeness of these genome assemblies has not been entirely satisfactory to the research community (Vallender 2011; Zhang et al. 2012). These are draft-quality genome assemblies, meaning that the DNA sequence is largely complete but is not yet comprehensive enough to be continuous over tens of millions of base pairs. A draft genome assembly consists of multiple segments of continuous uninterrupted sequence (contigs) such that most of the genome falls in contigs larger than about 15,000 to 30,000 bases. But thousands of gaps in the sequence remain and the order or orientation of the contigs is not known with complete reliability. A finished genome consists of much longer contigs, approaching or surpassing millions of continuous bases without interruption. The human genome is now a "finished" genome, as is that of the laboratory mouse, although complexities and issues remain because of complex polymorphic segmental duplications and other large-scale insertion-deletion differences among humans. It is both time consuming and very expensive to "finish" a mammalian genome, although new strategies are being developed to increase contig size in draft genomes so that the information is more comprehensive, continuous, and accurate. Clearly one goal for primate genomics in the near term should be to increase continuity and contig size in nonhuman primate genome assemblies to the point that nearly every gene is contained in a single contig.

The Indian-origin rhesus macaque, chimpanzee, gorilla, and orangutan genomes were sequenced entirely or primarily using traditional Sanger sequencing methods but were nevertheless draft genome assemblies that contained numerous gaps and other problems (Zhang et al. 2012). The other macaque sequences published to date have used various sequencing strategies but are also draft quality. In some cases, there was no effort to produce a de novo assembly based entirely on the new species- or population-specific data, but instead reads were mapped to the available Indian-origin rhesus macaque assembly for analysis. These draft genome sequences (whether assembled de novo or simply aligned to an existing assembly for a different species or subspecies) provide an initial starting point for various genetic analyses. But the number of genomic regions with missing sequence or incorrect or incomplete annotation makes it clear that further progress should be based on upgraded and improved assemblies (Alkan et al. 2011).

Given the continuing reduction in the cost of DNA sequencing, it is not surprising that the pace of primate comparative genomics is accelerating. De novo whole-genome assemblies for gibbon (*Nomascus leucogenys*), common marmoset (*Callithrix jacchus*), baboon (*Papio anubis*), sooty mangabey (*Cercocebus atys*), mouse lemur (*Microcebus murinus*), and African green monkey (*Chlorocebus aethiops*) have been completed but not yet published at the time of this writing. Other genomes are in different stages of sequencing, assembly, or analysis. This coming wave of new assemblies will have a major impact on many aspects of disease-related research. For example, the sooty mangabey is a unique and significant model for simian immunodeficiency virus (SIV) research because these animals are natural hosts that tolerate long-term infection with specific SIV viruses without developing disease (Chahroudi et al. 2012; Silvestri 2008). Development of genomic sequence data and other genomic tools will accelerate efforts to understand how sooty mangabeys can tolerate infection by viruses that are deadly to macaques. In the case of the common marmoset, this will be the first New World monkey genome to be published. This species exhibits a number of characteristics of behavior, neuroendocrinology, reproductive biology, and development that are under active study and where efforts will benefit from access to genomic information (Tardif et al. 2011; Ward and Vallender 2012). Baboons, African green monkeys, and other species are used in biomedical research because they provide valuable model systems for specific questions, and state-of-the-art investigation of those models depends on access to this type of information. For example, several genes influencing risk factors for human disease have been mapped and identified in the baboon genome (Cox et al. 2007; Tejero et al. 2005), but there are a larger number of such phenotypes for which the genes driving individual differences have not yet been identified (Rainwater et al. 2009). Sequencing of the baboon genome will facilitate those studies. We can anticipate that the genomes of many more primate species will be sequenced in the next several years. For example, the pig-tail macaque (*Macaca nemestrina*), drill (*Mandrillus leucophaeus*), and sifaka (*Propithecus coquerelli*) genomes are targeted for whole-genome sequencing

and de novo assembly at the Baylor Human Genome Sequencing Center.

What We Have Learned So Far

Comparisons among these primate whole-genome sequences have produced a great deal of new information. It is impossible in a short review to summarize all the results, but some specific outcomes are particularly significant.

Differences in Single-Copy DNA Sequence among Species

The fixed genetic difference between humans and chimpanzees (differences between species that are consistent across all members of those species) is estimated as 1.06% (Chimpanzee Sequencing and Analysis Consortium 2005). This is in contrast to a count of observed differences between any one human individual and any one chimpanzee, which would include within-species polymorphic variation as contributors to differences between two specific individuals. The time of separation of the human evolutionary lineage from the chimpanzee lineage remains somewhat controversial (Langergraber et al. 2012) but is widely accepted to be 5 to 7.5 million years ago. This suggests a rate of divergence of about 0.18% per million years, with half of that divergence occurring on each lineage. The difference in single-copy sequence between human and rhesus macaque is estimated to be 6.5%, with divergence time of 25 to 28 million years ago and thus an estimated rate of change of 0.25% per million years. Overall, broad comparisons suggest that the evolution of single-copy DNA sequences occurs more slowly through absolute time in the clade that includes gorillas, chimpanzees, bonobos, and humans than it does in more distantly related primates, including Old World monkeys (macaques, baboons, colobines, cercopithecines) or New World monkeys (marmosets, squirrel monkeys, spider monkeys, and so on).

Small Insertion-Deletion Differences

Roy Britten first pointed out that small insertions and deletions (indels) of less than 100 base pairs probably account for a larger number of total base pair differences between humans and chimpanzees than do single base changes in alignable sequence (Britten et al. 2003). Subsequently, whole-genome sequencing has proven him correct. Both the human and chimpanzee genomes consist of about 1.5% unique sequence not found in the other species. Most of this is because of small indels. The rhesus macaque sequence that can be aligned to the human genome is 93.5% identical to human, but when the small indels are included the rhesus is only 90.8% identical to human (Gibbs et al. 2007). This pattern of divergence in which small indels have substantial impact on genomic divergence is found across the primate genomes that have been sequenced.

Gene Copy Number Differences

Among humans, the great apes, and Old World monkeys such as macaques, most protein-coding genes have 1:1 homologues, but gene content is not identical among these species. Numerous gene families have expanded or contracted within individual evolutionary lineages. For example, the rhesus macaque genome consortium (Gibbs et al. 2007) identified 1358 genes found as new duplications in the rhesus macaque genome compared with human. The HLA gene cluster is one gene family of particular biomedical relevance and is significantly expanded in the macaques relative to humans. However, the draft quality of the great ape and Old World monkey genomes makes it difficult at this time to define all the copy number differences with certainty. By aligning the genes from different assemblies, we can evaluate apparent copy number, but different algorithms for assessing gene copy number do not produce identical results (e.g., Locke et al. 2011). In addition, the available draft genomes for nonhuman primates contain significant gaps and thus do not contain the complete complement of protein coding exons truly present in those species (Zhang et al. 2012). This makes comprehensive analyses problematic. Finally, the draft genomes do not yet support comprehensive identification and mapping of segmental duplications, which are often the regions that produce copy number differences (Dumas et al. 2007).

Segmental Duplications

Segmental duplications (independent sequences within a single genome that are longer than 1 kb and greater than 90% identical) are known to be hotspots for mutations that alter blocks of DNA sequence within the human population. They have also been important drivers of evolutionary change across nonhuman primate genomes (Marques-Bonet, Girirajan, and Eichler 2009; Marques-Bonet, Ryder, and Eichler 2009). Recent estimates indicate that about 5% of the human and chimpanzees genomes and 3.8% of the orangutan genome consist of segmental duplications (Locke et al. 2011). The genomes of humans and the great apes appear to be enriched for duplications dispersed around the genome, having experienced an interval after their divergence from Old World monkeys and before the separation of gorillas from humans and chimpanzees during which there was particularly active production of these new duplications (Jiang et al. 2007; Marques-Bonet, Kidd, et al. 2009). These events often involve regions that contain entire genes. Thus, many expansions of specific gene families within primate lineages are due to segmental duplications, sometimes involving repeated cycles of expansions for a given sequence (Cheng et al. 2005; Dumas et al. 2007; Gazave et al. 2011; Marques-Bonet, Ryder, and Eichler 2009). Among the great apes,

gorillas appear to have greater numbers of copy number changes involving coding genes than do related species (Scally et al. 2012). Genes within segmental duplications have been identified as specific examples of positive selection for coding sequence as well as copy number.

Species Differences in Both Levels of Gene Expression and Alternative Splicing

Differential gene expression is likely to account for many phenotypic differences among species (King and Wilson 1975), as well as individual phenotypic variation within species. Therefore, the comparative analysis of gene expression is a critical aspect of comparative genomics. For example, Blekhman and colleagues (2008; 2010) demonstrated that differences in gene expression among humans, chimpanzees, and rhesus macaques are influenced by natural selection, whereas Calarco and colleagues (2007) identified substantial differences in alternative splicing in the brains of humans and chimpanzees. Scally and colleagues (2012) found that there is greater overall similarity in gene expression between chimpanzees and humans, as compared with gorillas, thus providing an example of how the evolution of gene expression can quite closely track the phylogeny of primate lineages. Perry and colleagues (2012) generated sequences for RNA extracted from the livers of humans and a series of other mammals, including 11 nonhuman primates. They compared levels of expression and also found strong evidence for positive selection in a number of the genes expressed. Perry and colleagues (2012) also calculated levels of DNA sequence heterozygosity within the gene coding sequences from each species. Comparative analysis of primate transcription is currently a very active area of investigation, and we can expect much progress on this in the near future. For example, the Nonhuman Primate Reference Transcriptome Resource is generating RNA sequence data from an extensive series of tissues from several primate species (Pipes et al. 2013).

Similarity and Change in Retroposon Content

Various types of repetitive elements make up about 50% of the total genome sequence in humans, chimpanzees, rhesus macaques, and other species. Retroposons are a major component of the total complement of repetitive sequences within primates (Cordaux et al. 2010). However, the number of species-specific retroposon insertions differs substantially across species, from about 5000 in humans to 2300 in chimpanzees and only 250 in orangutans (Locke et al. 2011). More specifically, de novo Alu insertions constitute a major source of genomic change but have not affected all primate lineages to the same degree. This process has clearly altered genome content in Old World monkeys, as more than 100,000 Alu insertions are found in the rhesus macaque genome that are not present in the human genome (Gibbs et al. 2007). Like segmental duplications, locally clustered Alu insertions can trigger nonhomologous recombination and therefore large indel events. Overall, the dynamics of retroposon insertion have had substantial effects on primate genome evolution (Cordaux and Batzer 2009).

Genetic Variation within Primate Species

The literature concerning genetic variation in nonhuman primate species is large and cannot be summarized here. But researchers have begun large-scale analyses of DNA sequence variation within primates, especially macaques. Hernandez and colleagues (2007) reported sequence data for 150 kb from five genomic regions in nine Chinese- and 38 Indian-origin rhesus macaques. They found that the density of SNPs was 7.25 per kilobase in Chinese-origin rhesus macaques and 5.8 per kilobase in the Indian-origin rhesus macaques. Only about one-third of SNPs were shared between the two geographic populations, indicating that although there is substantial variation in each, most of that variation appears to be region- or population-specific. Ferguson and colleagues (2007) obtained similar results in a smaller survey of 3' UTR sequences. Fawcett and colleagues (2011) compared whole-genome sequences for three Indian-origin rhesus macaques and found more than 3 million variants that could be identified with confidence by virtue of occurring in at least two of the datasets examined. Across the three animals, they found more than14 million variants in at least one individual, for a potential SNP rate approaching 5 per kilobase (Fawcett et al. 2011). This is likely an underestimate of the total variation present because one of the three individuals was sequenced only to 3.3X coverage. Within the *Pan troglodytes verus* subspecies, the west African chimpanzee that is most commonly used in biomedical research laboratories, average heterozygosity was estimated to be 8×10^{-4}, whereas the estimate was more than twice that high for the *P. t. troglodytes* subspecies from central Africa (Chimpanzee Sequencing and Analysis Consortium 2005). Together the two subspecies have average heterozygosity of 19.0×10^{-4}, and this does not account for additional variation to be found in the eastern subspecies (*P. t. schweinfurthi*). Scally and colleagues (2012) found that two western lowland gorillas had heterozygosity rates of at least 1.8 per kilobase, whereas an eastern lowland gorilla showed less than half that rate. All these estimates for apes and macaques show diversity as high or higher than is observed in humans (Abecasis et al. 2012). Because technical advances have reduced the cost of sequencing so significantly, it is practical to pursue discovery of SNPs and other types of genetic variation by whole-genome sequencing of hundreds of animals. At the Human Genome Sequencing Center, we are collaborating with a number of primate research centers and have begun a survey of whole-genome variability among more than 200 Indian-origin rhesus macaques, along with smaller numbers of Chinese-origin rhesus macaques and cynomolgus macaques. This work is generating substantial amounts of information concerning previously unidentified SNPs, indels, copy

number variants, and other aspects of genomic diversity within these populations.

Impact on Biomedical Research

There is no question that the advances taking place in genetics and genomics will have major impact on our knowledge of the biology of nonhuman primates and on the ways in which these species are used in biomedical research. Exactly how the new genomic technologies and research strategies will be applied to primate models of disease is not yet clear. We can say with some confidence that resequencing of multiple individuals within each major laboratory primate species will discover substantial stores of genetic variation. A significant fraction of that variation will have functionally significant impact on phenotypes that will be relevant to human disease. But how much functional variation is present and at what allele frequencies are not known at this time. We can also say that the quantitative analysis of gene expression, including expression of messenger RNAs (mRNAs), microRNAs, long noncoding RNA (lncRNA), and other classes of genes, will have important impact on our understanding of the susceptibility to, onset of, and progression of disease in these model organisms. However, the most informative and most cost-effective strategies for exploiting primate models of disease remain to be defined in detail.

It seems highly likely that progress in the genetic analysis of nonhuman primate models of disease will include expansion of the field along three dimensions: (1) genome content and annotation, (2) genomic variation within species, and (3) genomic diversity across species.

Genome Content and Annotation

As described above, the draft quality whole-genome assemblies for chimpanzee, rhesus macaque, and other nonhuman primates have been useful to the research community but not entirely satisfactory. Higher quality genome assemblies, with larger contigs, fewer sequence gaps, and more accurate assembly of complex regions such as tandem repeats of functional genes or segmental duplications, are needed to support ever more detailed and comprehensive analyses (Alkan et al. 2011; Zhang et al. 2012). For example, comparative genomics is providing greater analytical power for discovery of conserved sequences and thus identifying more and more noncoding regions of the genome that have important biological function (Lindblad-Toh et al. 2011). Investigators will benefit from using nonhuman primate model species to study the biological significance and phenotypic consequences of these noncoding functional elements. But more complete and better annotated nonhuman primate genomes are required to support such research goals and analyses. Access to higher quality primate genome assemblies is obviously most important in the case of recently evolved regulatory elements that are primate specific (Lindblad-Toh et al. 2011).

Thus, one goal for primate genomics should be significant improvements of reference genome assemblies for a number of species, with highest priority placed on several different macaques, African green monkeys, baboons, marmosets, sooty mangabeys, and at least two strepsirrhine primates, including mouse lemur and at least one other. Improved assemblies will be developed using a variety of approaches. Clearly, deeper sequence coverage using Illumina paired-end and mate-pair reads will be valuable for assemblies, despite the limitations of this platform in terms of read length. Other sequencing platforms can provide much longer reads, which will be very useful in generating better assemblies with fewer gaps. The Pacific Biosciences RS platform has become one quite plausible option, although future developments of the RS and other commercial systems are likely to continue. Better software algorithms for de novo genome assembly, including algorithms that make efficient use of Pacific Biosciences RS long reads (English et al. 2012), will also help achieve better final assembly products.

In parallel with better assemblies of the genomic sequence, deep transcriptome information for nonhuman primates will lead to better annotation of functional genes. Sequencing of mRNA is an important component of this process (Perry et al. 2012; Pipes et al. 2013). Little is known at this time about microRNA expression in most primates, although microRNAs have received some attention in baboons (Karere et al. 2012) and macaques (Dannemann et al. 2012). Focused efforts in identifying microRNAs, lncRNAs, and other functional elements are certain to produce significant results and improve genome annotations, as well as generate opportunities to use nonhuman primates to investigate the function and phenotypic consequences of those genomic elements.

Genomic Variation within Species

There is a substantial amount of information available regarding genetic variation within species, but much of that literature is not directly useful in studies of primate models of phenotypic variation and disease. Although analysis of mitochondrial DNA variation, microsatellite polymorphism, and other types of intraspecies diversity have been useful for a wide range of analyses, the goals and impact of studies of primate genetic variation are clearly in a period of transition. Previous studies of primate polymorphism have focused mostly (though not entirely) on neutral noncoding variation. Now for the first time, it is straightforward and relatively inexpensive to analyze and define large fractions of the genetic variation in a given primate population, including functionally significant mutations (Fawcett et al. 2011). Several studies have reported studies of macaque whole-exome sequences (Vallender 2011), which is an efficient way to focus on functional variation that is readily interpretable at this time. Exome sequencing is a cost-effective method for discovering sequence variation within protein-coding genes. Nonsynonymous substitutions, premature stop codons, and other

readily interpretable classes of polymorphism are routinely detected (Vallender 2011). However, as more is learned about noncoding sequences in primates that have definable functions (transcription factor binding sites, other regulatory regions, genes for lncRNAs, and so on), generating whole-genome sequence data for those species to discover and characterize intraspecies variation across the entire genome will have broader utility and impact. Today the annotation of nonhuman primate genomes lags behind that of the human and mouse genomes, but this can be expected to improve.

Thus, another significant priority should be the development of comprehensive information about genetic variation in the most important laboratory primates. Just as clinicians are now considering the cost-benefit equations related to clinical human genome sequencing (i.e., under what circumstances is it appropriate to sequence a patient's genome to assist in the diagnosis of a disorder or in the development of a treatment plan), veterinarians, colony mangers, and primate center directors should now begin evaluating opportunities for whole-genome sequencing of animals in their research colonies. The production of whole-genome or whole-exome data for significant numbers of macaques, African green monkeys, baboons, marmosets, or other laboratory primate species will allow for the discovery of functionally significant mutations that lead directly to new primate models of the genetics of human disease (see, for example, Barr et al. 2004; Rogers et al. 2012; Vallender et al. 2010). In addition, better information concerning genetic variation within a research colony would allow investigators to select research subjects on the basis of genetic characteristics. This is already done in SIV research, where decisions are based on specific MHC alleles and genotypes. For research in neurobiology, endocrinology, or other fields, it may be useful to either include or exclude individual monkeys that carry specific functionally significant variants, depending on the goals of the particular research study. Third, but just as important, increased information about sequence variation within specific research colonies will allow colony managers to make more informed decisions concerning breeding plans and selection of individual animals for breeding versus experimental use.

The question is not whether to sequence primates in research colonies but how many and which animals to characterize in this way. It is already cheaper to sequence the whole genome of an Indian-origin rhesus macaque than it is to purchase that animal, pay per diem costs for 2 years, and use that animal for a study of infectious disease, neurotransmitter function, or other biomedical investigation. The sequencing of whole exomes is a fraction of the cost of whole genomes, so it is several fold cheaper to produce whole-exome data for an animal than it is to purchase and house that animal for 1 to 2 years.

However, as of today, there is still too little known about the nature and consequences of functional genetic variation in any laboratory primate species to justify comprehensive sequencing of study animals for the purpose of routinely making whole-genome sequence data available to research programs. This situation is going to change soon, and the options for sequencing will become more attractive. Before long the characterization of research colonies of nonhuman primates can be expected to include a plan for genomic characterization of a large portion, if not most, of the animals available for research studies. Such characterization is likely to begin with full sequencing of some animals and whole-exome sequencing of others. This will identify carriers of numerous functionally interesting alleles that can impact experimental results.

Thus, reduced cost is one reason this field is in transition. The second is the new insights into the genetics of disease already generated by sequencing human genomes. The first complete human genome sequence was published in 2005, but since then the field of human genetics has moved at ever-increasing speed. Research consortia such as the 1000Genomes Project (www.1000genomes.org) have produced whole-genome or whole-exome sequences for hundreds of humans, and the scale of human genetic research continues to expand. Research projects involving thousands of human subjects are now underway and producing results. Consequently, researchers now have a quantitative understanding of the nature and distribution of human genetic variation that was unattainable just a few years ago. A detailed review of human genetic variability is not appropriate here, but the data support three fundamental conclusions:

(1) The amount of genetic variation among humans is high.
(2) The extant variation includes a wide array of different types of genetic polymorphism, including SNPs, small indels, larger copy number differences, segmental duplications, larger inversions, and other forms of sequence change.
(3) Much of the biologically significant variation that has functional effects on normal variation or risk of disease is low in frequency, such that observable phenotypic and physiologic differences often but not always result from the combined effects of multiple variants across multiple genes.

Preliminary results from our work at HGSC and other laboratories suggest strongly that rhesus macaques carry more genetic variation within a given number of individuals than do humans. It is possible that macaques carry more genetic variation with little or no phenotypic effect but about the same amount as humans for variants with substantial effects (Yuan et al. 2012). In any case, we can be confident that macaques, baboons, African green monkeys, marmosets, and other laboratory primates carry a large store of biologically and phenotypically important genetic variation that will be useful in understanding the health-related consequences of human genetic variation. Some specific examples of variation with potential biomedical impact are already available (Barr et al. 2003; Rogers et al. 2012; Vallender et al. 2010), and these cases show that nonhuman primates frequently carry mutations in genes orthologous to genes that are associated with disease phenotypes in humans. Furthermore, the primate mutations can exhibit similar or nearly identical

phenotypes. The primates do not necessarily carry the same mutation as humans, but different mutations in the same gene can provide models with great utility for understanding genotype–phenotype relationships. For example, the similarity of behavioral and pharmacogenetic consequences of mu-opioid mutations in rhesus macaques and humans in relation to alcohol abuse are remarkable (Vallender et al. 2010). This may not be entirely surprising given the overall biological similarity of these two species, but the parallelism here provides a valuable example of the power of nonhuman primate models for investigation of the genetic basis of health- and disease-related traits. Similarly, in a study of behavioral and neurobiological phenotypes in rhesus macaque, we (Rogers et al. 2012) identified SNPs in the *CRHR1* locus (a gene previously associated with risk of anxiety disorders and depression in humans) that are significantly related to variation in anxious temperament and local metabolism in the hippocampus and other brain structures.

If the pattern and structure of genetic variation in macaques, baboons, marmosets, and other laboratory primates is similar to that found in humans but with higher levels of variation with minimal or no phenotypic effect, then there is clear justification for a change in the strategies used when investigating primate models of the genetics of disease. In the past, most efforts in this area either examined one candidate gene at a time or used forward genetic strategies in which interesting phenotypes were measured in large sets of animals, and then one or another approach was employed to identify the specific genes affecting those phenotypes. Given the cost of sequencing and the likely presence of many genetic variants, each with modest to small phenotypic effects, a reverse genetic approach may be more efficient. This would call for large-scale sequencing studies designed to identify functionally significant (or probable functional variants) first and then use of animals that carry those candidate mutations to study their phenotypic consequences. In addition, it is likely that large segmental duplications, copy number variation, and large de novo indel polymorphisms are also common in nonhuman primates (Lee et al. 2008). This means that there is justification for de novo sequencing and assembly of additional individuals within macaques, baboons, African green monkeys, and other species for which a first species-specific genome assembly is already in hand. Human genetic comparisons are finding that no one individual can serve as a comprehensive reference sequence for all humans because there are polymorphic large-scale insertions segregating among our species. We can anticipate similar complexity in nonhuman primates.

Genomic Diversity across Species

Each primate species will have its own unique complement of genetic variation. This means that different species will carry different variants in the same disease-related genes, and knowing which species exhibit which high-frequency and low-frequency variants will allow researchers to make informed choices about the species most appropriate for any given study. For example, primates, especially macaques, are critical for the development and testing of new drugs. Ise and colleagues (2011) examined differential expression of drug-metabolizing genes in various populations of cynomolgus and rhesus macaques, species that are widely used for drug development and evaluation. They found that expression of P450 genes among geographic populations of cynomolgus macaques do not differ substantially but that there are significant differences in expression of some genes compared with rhesus macaques. These differences have implications for the use of these species in pharmacologic research because a specific drug may be metabolized differently in these two species. Experimental outcomes can therefore differ in meaningful ways. Obviously, anyone interpreting those results would want to be aware of genetic differences in the relevant drug-metabolizing pathways. This can also apply to genetic variation within species, as discussed above.

Conclusions

There is no question that this is a remarkable time in the field of nonhuman primate genomics. Researchers are publishing new and significant results at a tremendous pace. We can anticipate that the next several years will see dramatic increases in the amount of information available concerning genetic variation within major laboratory primate species and the patterns of tissue-specific expression of protein-coding genes, microRNAs, and other genome elements across multiple species. Other aspects of genomics, including comparative primate epigenetics, will also grow and become more sophisticated in the methods used and research questions addressed. These new results will improve our ability to use existing primate models to understand the causes and potential treatment of disease. The results will also identify new disease models. The expected continued reductions in sequencing costs and improvements in sequencing technology, such as longer read lengths, will have their own effects on this field, as will improved software tools. It is impossible at this time to define with any precision or confidence how these developments will alter our research strategies or aspirations, but it is sure to be an exciting ride.

Acknowledgments

I thank Dr. John Harding for the invitation to contribute to this collection. I also thank R. Gibbs, G. Fawcett, M. Raveendran, D. Rio Deiros, K. Worley, and S. Richard for valuable discussion. This work was supported by the National Institutes of Health (R24-RR032329 to J Rogers and U54-HG006484 to Richard Gibbs, PI).

References

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012. An integrated

map of genetic variation from 1,092 human genomes. Nature 491:56–65.
Alkan C, Sajjadian S, Eichler EE. 2011. Limitations of next-generation genome sequence assembly. Nat Methods 8:61–65.
Barr CS, Newman TK, Becker ML, Parker CC, Champoux M, Lesch KP, Goldman D, Suomi SJ, Higley JD. 2003. The utility of the non-human primate model for studying gene by environment interactions in behavioral research. Genes Brain Behav 2:336–340.
Barr CS, Newman TK, Shannon C, Parker C, Dvoskin RL, Becker ML, Schwandt M, Champoux M, Lesch KP, Goldman D, Suomi SJ, Higley JD. 2004. Rearing condition and rh5-HTTLPR interact to influence limbic-hypothalamic-pituitary-adrenal axis response to stress in infant macaques. Biol Psychiatry 55:733–738.
Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. 2010. Sex-specific and lineage-specific alternative splicing in primates. Genome Res 20:180–189.
Blekhman R, Oshlack A, Chabot AE, Smyth GK, Gilad Y. 2008. Gene regulation in primates evolves under tissue-specific selection pressures. PLoS Genet 4:e1000271.
Bosinger SE, Jochems SP, Folkner KA, Hayes TL, Klatt NR, Silvestri G. 2013. Transcriptional profiling of experimental CD8(+) lymphocyte depletion in rhesus macaques infected with simian immunodeficiency virus SIVmac239. J Virol 87:433–443.
Britten RJ, Rowen L, Williams J, Cameron RA. 2003. Majority of divergence between closely related DNA samples is due to indels. Proc Natl Acad Sci U S A 100:4661–4665.
Calarco JA, Xing Y, Caceres M, Calarco JP, Xiao X, Pan Q, Lee C, Preuss TM, Blencowe BJ. 2007. Global analysis of alternative splicing differences between humans and chimpanzees. Genes Devel 21:2963–2975.
Chahroudi A, Bosinger SE, Vanderford TH, Paiardini M, Silvestri G. 2012. Natural SIV hosts: Showing AIDS the door. Science 335:1188–1193.
Charrier C, Joshi K, Coutinho-Budd J, Kim JE, Lambert N, de Marchena J, Jin WL, Vanderhaeghen P, Ghosh A, Sassa T, Polleux F. 2012. Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. Cell 149(4):923–935.
Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Paabo S, Rocchi M, Eichler EE. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature 437:88–93.
Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.
Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet 10:691–703.
Cordaux R, Sen SK, Konkel MK, Batzer MA. 2010. Computational methods for the analysis of primate mobile elements. Methods Mol Biol 628:137–151.
Cox LA, Birnbaum S, Mahaney MC, Rainwater DL, Williams JT, VandeBerg JL. 2007. Identification of promoter variants in baboon endothelial lipase that regulate high-density lipoprotein cholesterol levels. Circulation 116:1185–1195.
Dannemann M, Prufer K, Lizano E, Nickel B, Burbano HA, Kelso J. 2012. Transcription factors are targeted by differentially expressed miRNAs in primates. Genome Biol Evol 4:552–564.
Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, Pollack JR, Sikela JM. 2007. Gene copy number variation spanning 60 million years of human and primate evolution. Genome Res 17:1266–1277.
Ebeling M, Kung E, See A, Broger C, Steiner G, Berrera M, Heckel T, Iniguez L, Albert T, Schmucki R, Biller H, Singer T, Certa U. 2011. Genome-based analysis of the nonhuman primate *Macaca fascicularis* as a model for drug safety assessment. Genome Res 21: 1746–1756.
English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA. 2012. Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PloS One 7:e47768.
Fang X, Zhang Y, Zhang R, Yang L, Li M, Ye K, Guo X, Wang J, Su B. 2011. Genome sequence and global sequence variation map with 5.5 million SNPs in Chinese rhesus macaque. Genome Biol 12:R63.
Fawcett GL, Raveendran M, Deiros DR, Chen D, Yu F, Harris RA, Ren Y, Muzny DM, Reid JG, Wheeler DA, Worley KC, Shelton SE, Kalin NH, Milosavljevic A, Gibbs R, Rogers J. 2011. Characterization of single-nucleotide variation in Indian-origin rhesus macaques (*Macaca mulatta*). BMC Genomics 12:311.
Ferguson B, Street SL, Wright H, Pearson C, Jia Y, Thompson SL, Allibone P, Dubay CJ, Spindel E, Norgren RB Jr. 2007. Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). BMC Genomics 8:43.
Gazave E, Darre F, Morcillo-Suarez C, Petit-Marty N, Carreno A, Marigorta UM, Ryder OA, Blancher A, Rocchi M, Bosch E, Baker C, Marques-Bonet T, Eichler EE, Navarro A. 2011. Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. Genome Res 21:1626–1639.
Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, Batzer MA, Bustamante CD, Eichler EE, Hahn MW, Hardison RC, Makova KD, Miller W, Milosavljevic A, Palermo RE, Siepel A, Sikela JM, Attaway T, Bell S, Bernard KE, Buhay CJ, Chandrabose MN, Dao M, Davis C, Delehaunty KD, Ding Y, Dinh HH, Dugan-Rocha S, Fulton LA, Gabisi RA, Garner TT, Godfrey J, Hawes AC, Hernandez J, Hines S, Holder M, Hume J, Jhangiani SN, Joshi V, Khan ZM, Kirkness EF, Cree A, Fowler RG, Lee S, Lewis LR, Li Z, Liu YS, Moore SM, Muzny D, Nazareth LV, Ngo DN, Okwuonu GO, Pai G, Parker D, Paul HA, Pfannkoch C, Pohl CS, Rogers YH, Ruiz SJ, Sabo A, Santibanez J, Schneider BW, Smith SM, Sodergren E, Svatek AF, Utterback TR, Vattathil S, Warren W, White CS, Chinwalla AT, Feng Y, Halpern AL, Hillier LW, Huang X, Minx P, Nelson JO, Pepin KH, Qin X, Sutton GG, Venter E, Walenz BP, Wallis JW, Worley KC, Yang SP, Jones SM, Marra MA, Rocchi M, Schein JE, Baertsch R, Clarke L, Csuros M, Glasscock J, Harris RA, Havlak P, Jackson AR, Jiang H, Liu Y, Messina DN, Shen Y, Song HX, Wylie T, Zhang L, Birney E, Han K, Konkel MK, Lee J, Smit AF, Ullmer B, Wang H, Xing J, Burhans R, Cheng Z, Karro JE, Ma J, Raney B, She X, Cox MJ, Demuth JP, Dumas LJ, Han SG, Hopkins J, Karimpour-Fard A, Kim YH, Pollack JR, Vinar T, Addo-Quaye C, Degenhardt J, Denby A, Hubisz MJ, Indap A, Kosiol C, Lahn BT, Lawson HA, Marklein A, Nielsen R, Vallender EJ, Clark AG, Ferguson B, Hernandez RD, Hirani K, Kehrer-Sawatzki H, Kolb J, Patil S, Pu LL, Ren Y, Smith DG, Wheeler DA, Schenck I, Ball EV, Chen R, Cooper DN, Giardine B, Hsu F, Kent WJ, Lesk A, Nelson DL, O'Brien WE, Prufer K, Stenson PD, Wallace JC, Ke H, Liu XM, Wang P, Xiang AP, Yang F, Barber GP, Haussler D, Karolchik D, Kern AD, Kuhn RM, Smith KE, Zwieg AS. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222–234.
Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, Rogers J, Nazareth L, Indap A, Bourquin T, McPherson J, Muzny D, Gibbs R, Nielsen R, Bustamante CD. 2007. Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. Science 316:240–243.
Higashino A, Sakate R, Kameoka Y, Takahashi I, Hirata M, Tanuma R, Masui T, Yasutomi Y, Osada N. 2012. Whole-genome sequencing and analysis of the Malaysian cynomolgus macaque (*Macaca fascicularis*) genome. Genome Biol 13:R58.
Ise R, Kondo S, Kato H, Imai N, Akiyama H, Iwasaki K, Yamazaki H, Uno Y. 2011. Expression of cytochromes p450 in fetal, infant, and juvenile liver of cynomolgus macaques. Drug Metab Pharmacokinet 26:621–626.
Jiang Z, Tang H, Ventura M, Cardone MF, Marques-Bonet T, She X, Pevzner PA, Eichler EE. 2007. Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. Nat Genet 39:1361–1368.

Karere GM, Glenn JP, VandeBerg JL, Cox LA. 2012. Differential microRNA response to a high-cholesterol, high-fat diet in livers of low and high LDL-C baboons. BMC Genomics 13:320.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107–116.

Langergraber KE, Prufer K, Rowney C, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Mitani JC, Muller MN, Robbins MM, Schubert G, Stoinski TS, Viola B, Watts D, Wittig RM, Wrangham RW, Zuberbuhler K, Paabo S, Vigilant L. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc Natl Acad Sci U S A 109:15716–15721.

Lee AS, Gutierrez-Arcelus M, Perry GH, Vallender EJ, Johnson WE, Miller GM, Korbel JO, Lee C. 2008. Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. Hum Mol Genet 17:1127–1136.

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, Ward LD, Lowe CB, Holloway AK, Clamp M, Gnerre S, Alfoldi J, Beal K, Chang J, Clawson H, Cuff J, Di Palma F, Fitzgerald S, Flicek P, Guttman M, Hubisz MJ, Jaffe DB, Jungreis I, Kent WJ, Kostka D, Lara M, Martins AL, Massingham T, Moltke I, Raney BJ, Rasmussen MD, Robinson J, Stark A, Vilella AJ, Wen J, Xie X, Zody MC, Baldwin J, Bloom T, Chin CW, Heiman D, Nicol R, Nusbaum C, Young S, Wilkinson J, Worley KC, Kovar CL, Muzny DM, Gibbs RA, Cree A, Dihn HH, Fowler G, Jhangiani S, Joshi V, Lee S, Lewis LR, Nazareth LV, Okwuonu G, Santibanez J, Warren WC, Mardis ER, Weinstock GM, Wilson RK, Delehaunty K, Dooling D, Fronik C, Fulton L, Fulton B, Graves T, Minx P, Sodergren E, Birney E, Margulies EH, Herrero J, Green ED, Haussler D, Siepel A, Goldman N, Pollard KS, Pedersen JS, Lander ES, Kellis M. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478:476–482.

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang SP, Wang Z, Chinwalla AT, Minx P, Mitreva M, Cook L, Delehaunty KD, Fronick C, Schmidt H, Fulton LA, Fulton RS, Nelson JO, Magrini V, Pohl C, Graves TA, Markovic C, Cree A, Dinh HH, Hume J, Kovar CL, Fowler GR, Lunter G, Meader S, Heger A, Ponting CP, Marques-Bonet T, Alkan C, Chen L, Cheng Z, Kidd JM, Eichler EE, White S, Searle S, Vilella AJ, Chen Y, Flicek P, Ma J, Raney B, Suh B, Burhans R, Herrero J, Haussler D, Faria R, Fernando O, Darre F, Farre D, Gazave E, Oliva M, Navarro A, Roberto R, Capozzi O, Archidiacono N, Della Valle G, Purgato S, Rocchi M, Konkel MK, Walker JA, Ullmer B, Batzer MA, Smit AF, Hubley R, Casola C, Schrider DR, Hahn MW, Quesada V, Puente XS, Ordonez GR, Lopez-Otin C, Vinar T, Brejova B, Ratan A, Harris RS, Miller W, Kosiol C, Lawson HA, Taliwal V, Martins AL, Siepel A, Roychoudhury A, Ma X, Degenhardt J, Bustamante CD, Gutenkunst RN, Mailund T, Dutheil JY, Hobolth A, Schierup MH, Ryder OA, Yoshinaga Y, de Jong PJ, Weinstock GM, Rogers J, Mardis ER, Gibbs RA, Wilson RK. 2011. Comparative and demographic analysis of orang-utan genomes. Nature 469:529–533.

Marques-Bonet T, Girirajan S, Eichler EE. 2009. The origins and impact of primate segmental duplications. Trends Genets 25:443–454.

Marques-Bonet T, Kidd JM, Ventura M, Graves TA, Cheng Z, Hillier LW, Jiang Z, Baker C, Malfavon-Borja R, Fulton LA, Alkan C, Aksay G, Girirajan S, Siswara P, Chen L, Cardone MF, Navarro A, Mardis ER, Wilson RK, Eichler EE. 2009. A burst of segmental duplications in the genome of the African great ape ancestor. Nature 457:877–881.

Marques-Bonet T, Ryder OA, Eichler EE. 2009. Sequencing primate genomes: What have we learned? Ann Rev Genomics Hum Genet 10:355–386.

O'Bleness MS, Dickens CM, Dumas LJ, Kehrer-Sawatzki H, Wyckoff GJ, Sikela JM. 2012. Evolutionary history and genome organization of DUF1220 protein domains. G3 (Bethesda) 2:977–986.

Perry GH, Melsted P, Marioni JC, Wang Y, Bainer R, Pickrell JK, Michelini K, Zehr S, Yoder AD, Stephens M, Pritchard JK, Gilad Y. 2012. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. Genome Res 22:602–610.

Pipes L, Li S, Bozinoski M, Palermo R, Peng X, Blood P, Kelly S, Weiss JM, Thierry-Mieg J, Thierry-Mieg D, Zumbo P, Chen R, Schroth GP, Mason CE, Katze MG. 2013. The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics. Nucleic Acids Res 41:D906–D914.

Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A, Kern AD, Dehay C, Igel H, Ares M Jr., Vanderhaeghen P, Haussler D. 2006. An RNA gene expressed during cortical development evolved rapidly in humans. Nature 443:167–172.

Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Morrison H, Fitzpatrick DR, Afzal V, Pennacchio LA, Rubin EM, Noonan JP. 2008. Human-specific gain of function in a developmental enhancer. Science 321:1346–1350.

Preuss TM. 2012. Human brain evolution: from gene discovery to phenotype discovery. Proc Natl Acad Sci U S A 109:10709–10716.

Prufer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G, Kodira C, Winer R, Knight JR, Mullikin JC, Meader SJ, Ponting CP, Lunter G, Higashino S, Hobolth A, Dutheil J, Karakoc E, Alkan C, Sajjadian S, Catacchio CR, Ventura M, Marques-Bonet T, Eichler EE, Andre C, Atencia R, Mugisha L, Junhold J, Patterson N, Siebauer M, Good JM, Fischer A, Ptak SE, Lachmann M, Symer DE, Mailund T, Schierup MH, Andres AM, Kelso J, Paabo S. 2012. The bonobo genome compared with the chimpanzee and human genomes. Nature 486:527–531.

Rainwater DL, Cox LA, Rogers J, Vandeberg JL, Mahaney MC. 2009. Localization of multiple pleiotropic genes for lipoprotein metabolism in baboons. J Lipid Res 50:1420–1428.

Rogers J, Raveendran M, Fawcett GL, Fox AS, Shelton SE, Oler JA, Cheverud J, Muzny DM, Gibbs RA, Davidson RJ, Kalin NH. 2012. CRHR1 genotypes, neural circuits and the diathesis for anxiety and depression. Mol Psychiatry 18: 700–707.

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, McCarthy S, Montgomery SH, Schwalie PC, Tang YA, Ward MC, Xue Y, Yngvadottir B, Alkan C, Andersen LN, Ayub Q, Ball EV, Beal K, Bradley BJ, Chen Y, Clee CM, Fitzgerald S, Graves TA, Gu Y, Heath P, Heger A, Karakoc E, Kolb-Kokocinski A, Laird GK, Lunter G, Meader S, Mort M, Mullikin JC, Munch K, O'Connor TD, Phillips AD, Prado-Martinez J, Rogers AS, Sajjadian S, Schmidt D, Shaw K, Simpson JT, Stenson PD, Turner DJ, Vigilant L, Vilella AJ, Whitener W, Zhu B, Cooper DN, de Jong P, Dermitzakis ET, Eichler EE, Flicek P, Goldman N, Mundy NI, Ning Z, Odom DT, Ponting CP, Quail MA, Ryder OA, Searle SM, Warren WC, Wilson RK, Schierup MH, Rogers J, Tyler-Smith C, Durbin R. 2012. Insights into hominid evolution from the gorilla genome sequence. Nature 483:169–175.

Silvestri G. 2008. AIDS pathogenesis: A tale of two monkeys. J Med Primatol 37:6–12.

Tardif SD, Abee CR, Mansfield KG. 2011. Workshop summary: Neotropical primates in biomedical research. ILAR J 52:386–392.

Tejero ME, Cole SA, Cai G, Peebles KW, Freeland-Graves JH, Cox LA, Mahaney MC, Rogers J, VandeBerg JL, Blangero J, Comuzzie AG. 2005. Genome-wide scan of resistin mRNA expression in omental adipose tissue of baboons. Int J Obes (Lond) 29:406–412.

Vallender EJ. 2011. Expanding whole exome resequencing into non-human primates. Genome Biol 12:R87.

Vallender EJ, Ruedi-Bettschen D, Miller GM, Platt DM. 2010. A pharmacogenetic model of naltrexone-induced attenuation of alcohol consumption in rhesus monkeys. Drug Alcohol Depend 109:252–256.

Ward JM, Vallender EJ. 2012. The resurgence and genetic implications of New World primates in biomedical research. Trends Genet 28:586–591.

Yan G, Zhang G, Fang X, Zhang Y, Li C, Ling F, Cooper DN, Li Q, Li Y, van Gool AJ, Du H, Chen J, Chen R, Zhang P, Huang Z, Thompson JR, Meng Y, Bai Y, Wang J, Zhuo M, Wang T, Huang Y, Wei L, Li J, Wang Z, Hu H, Yang P, Le L, Stenson PD, Li B, Liu X, Ball EV, An N, Huang Q, Zhang Y, Fan W, Zhang X, Li Y, Wang W, Katze MG, Su B, Nielsen R, Yang H, Wang J, Wang X, Wang J. 2011. Genome sequencing and

comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. Nature Biotechnol 29:1019–1023.

Yuan Q, Zhou Z, Lindell SG, Higley JD, Ferguson B, Thompson RC, Lopez JF, Suomi SJ, Baghal B, Baker M, Mash DC, Barr CS, Goldman D. 2012. The rhesus macaque is three times as diverse but more closely equivalent in damaging coding variation as compared to the human. BMC Genet 13:52.

Zhang X, Goodsell J, Norgren RB Jr. 2012. Limitations of the rhesus macaque draft genome assembly and annotation. BMC Genomics 13:206.