White Paper for Complete Sequencing of the Rhesus Macaque (*Macaca mulatta*) Genome

Jeffrey Rogers[1], Michael Katze[2], Roger Bumgarner[2], Richard A. Gibbs[3] and George M. Weinstock[3]

I. Introduction

Humans are members of the Order Primates and our closest evolutionary relatives are other primate species. This makes primate models of human disease particularly important, as the underlying physiology and metabolism, as well as genomic structure, are more similar to humans than are other mammals. Chimpanzees (*Pan troglodytes*) are the animals most similar to humans in overall DNA sequence, with interspecies sequence differences of approximately 1-1.5% (Stewart and Disotell 1998, Page and Goodman 2001). The other apes, including gorillas and orangutans are nearly as similar to humans. The animals next most closely related to humans are Old World monkeys, superfamily Cercopithecoidea. This group includes the common laboratory species of rhesus macaque (*Macaca mulatta*), baboon (*Papio hamadryas*), pig-tailed macaque (*Macaca nemestrina*) and African green monkey (*Chlorocebus aethiops*). The human evolutionary lineage separated from the ancestors of chimpanzees about 6-7 million years ago (MYA), while the human/ape lineage diverged from Old World monkeys about 25 MYA (Stewart and Disotell 1998), and from another important primate group, the New World monkeys, more than 35-40 MYA. In comparison, humans diverged from mice and other non-primate mammals about 65-85 MYA (Kumar and Hedges 1998, Eizirik et al 2001).

In the evaluation of primate candidates for genome sequencing there should be more to selection of an organism than evolutionary considerations. The chimpanzee's status as Closest Relative To Human has earned it an exemption from this consideration. This very characteristic makes the chimpanzee unsuitable for research on ethical grounds. For the rest of the primates there is a requirement of research utility as well as the value of sequence comparison. Because of this, the rhesus macaque is the outstanding choice for a second nonhuman primate genome project to complement the chimpanzee project.

This document summarizes the justification for producing a complete genomic DNA sequence for the rhesus macaque. Some notable characteristics of rhesus monkeys that justify this project are:

1) Nonhuman primates provide animal models of human disease that are essential to much of biomedical research, and rhesus macaques are the most widely used nonhuman primate. Current figures from the CRISP database indicate that 60-70% of all NIH-funded grants that involve primates use rhesus macaques. According to the USDA Animal Welfare Report for 2000, about 57,000 primates were used in research that year, and the NIH conservatively estimates that 60-75% of those were rhesus. This species is available in large numbers, and the demand is growing. The NIH is supporting the expansion of rhesus breeding programs in order to increase the number of animals available to researchers. *The impact of nonhuman primate genomic data will be greatest if it relates directly to the most utilized nonhuman primate system.*

2) Rhesus monkeys are used for an extraordinary range of biomedical and basic research. Due to their close genetic, physiologic, and metabolic similarity to humans, this species serves as an essential research tool in neuroscience, behavioral biology, reproductive physiology, neuroendocrinology, endocrinology, cardiovascular studies, pharmacology and many other areas.

A genomic sequence will lead to new and more sophisticated research projects in both existing macaque models of human diseases and basic research in fundamental aspects of mammalian and human biology.

3) By virtue of the way it responds to infection with simian immunodeficiency virus (SIV), the rhesus macaque is widely recognized as the best model of disease progression and pathogenesis in AIDS. Development of new vaccine strategies to fight HIV depend heavily on basic immunology and virology performed using rhesus monkeys, as well as specific experiments using rhesus to test candidate vaccines. Rhesus is also an important model for other infectious diseases. Given current national defense concerns, we note that rhesus macaques are used in studies related to protection against anthrax and other aspects of bioterrorism (e.g. Fellows et al 2001). This line of research is certain to expand in coming years. *A complete genome sequence allows reagents to be developed for in-depth analysis of immunological, inflammatory, and other host responses to infectious agents.*

4) Rhesus macaque is an important system for drug development and testing in industry. This includes target validation, toxicity, and other advanced preclinical studies. Lack of genomic sequence hampers interpretation of these experiments. Knowledge of the differences in protein sequence, behavior, quantity, and distribution between human and rhesus are required in order to properly evaluate such testing (see letter from Dr. Went). *The rhesus genomic sequence will have a major impact in industry in drug development by providing an accurate description of the differences in protein characteristics between rhesus and humans.*

5) Since the chimpanzee genome project is underway, it is important to identify outgroups for chimpanzee-human comparisons. The evolutionary position of the rhesus makes it useful for many of these comparisons. The third species can determine whether chimpanzee-human differences are due to changes in the chimpanzee or human genomes (see letter from Dr. Rubin). Comparisons with mice or other non-primate animals may not provide the necessary information, because for rapidly evolving regions these genomes may be too different to allow unambiguous conclusions. While primates more closely related to chimpanzee (e.g. orangutan) may be a better outgroup for extremely rapidly evolving regions, these genomes may be too close to distinguish other changes that may have occurred uniquely in the chimpanzee lineage. Moreover, these primates are less widely used in research and thus a complete genome project is harder to justify. Rhesus macaques provide an outgroup that is of utility for all but these most rapidly evolving regions. Although these rapidly evolving regions include the most recent changes leading to the evolution of human, the other regions identifiable in three way comparisons are not without interest and are likely to be more numerous. *Rhesus can thus provide an outgroup for human-chimpanzee comparisons and contribute to evolutionary studies.*

6) The NIH is allocating money to increase the number of rhesus available to investigators because rhesus are becoming more important. Genetic resources for this species are not as developed as for other model organisms, but the amount of genetic research is increasing rapidly. BAC sequencing and cDNA sequencing have begun at several institutions. A radiation hybrid map of the rhesus genome is under construction, and a genetic linkage map will be complete in 12 months. Researchers are exploring expression array experiments in rhesus, though this work is hindered by lack of sequence data for rhesus genes. The use of rhesus macaques for genetic and genomic analyses will grow substantially over the coming years. *Genomic sequence data will both accelerate the growth of genetic research on rhesus as well as become more critical to that research.*

7) Rhesus macaques are closely related in evolutionary and genetic terms to three other widely used laboratory primates (baboons, cynomolgus macaques and pig-tailed macaques). The pairwise interspecies sequence differences among these four species are expected to be 1-2%, which means that genomic sequence data for rhesus will be tremendously valuable to researchers working on these other species. *Access to genomic sequence data from rhesus will create new research opportunities for all four of these commonly studied species.*

II. Specific Biological Rationale

1. *Improving human health.* The selection of rhesus as an important nonhuman primate for sequencing is based on the number of individuals available for use, access to large multigeneration pedigrees suitable for pedigree-based genetic analyses and gene mapping, the increasing number of genetic resources available, and the breadth of scientific and medical questions that are investigated using this species. Clearly the underlying rationale for sequencing a model organism should be to maximize the amount of information that will be relevant for understanding the molecular and cellular basis of human disease processes. In this sense, producing a genomic sequence of the rhesus macaque, the nonhuman primate most widely used in biomedical research, will create the greatest potential for incorporating genomic data and genomic technology into primate research that will improve human health.

The physiological and genetic similarities between humans and rhesus monkeys make these animals outstanding subjects for disease-related research. At present, it is difficult to investigate human-rhesus similarities at the molecular level using high-throughput methods. Access to genomic sequence information will allow researchers to attack human diseases in rhesus macaques and explore cellular or molecular processes that cannot be examined directly in humans and are not effectively modeled in non-primate organisms. Letters are included from researchers (Drs. McConkey, Cheverud, and Hacia) who indicate that attempts to use human cDNA sequences and expression arrays to study gene expression in rhesus have been less successful due to sequence divergence. *To properly exploit expression array methods in rhesus disease models, rhesus-specific information is required.*

It is not possible to list in this document all the ways in which biomedical researchers use and depend upon this species. A few examples must suffice to demonstrate that the rhesus is critical to future research progress. While several primate species are used to test candidate vaccines intended to fight AIDS, only macaques are used to study both vaccine strategies and pathogenesis. The rhesus macaque is considered to be the most important animal model of infection and disease progression by the AIDS research community (see letter from Dr. N. Letvin). This includes studies of infection, pathogenesis and treatment of animals infected *in utero* (e.g. Tarantal et al 1999).

The field of neurobiology is particularly tied to the use of rhesus and other macaques as subjects (see letter from Drs. Rakic and Goldman-Rakic). A PubMed search using the terms "rhesus and brain" produced 3187 citations, whereas "chimpanzee and brain" yielded just 274. Using the terms "macaque and brain" produced 11,702 citations. Both pharmacology and endocrinology make similar extensive use of rhesus monkeys. This species is often used in analyses of obesity, cardiovascular disease and diabetes (e.g. Hotta et al 2001, Winegar et al 2001). The rhesus macaque is the standard organism of choice in studies of drug addiction, alcoholism and a range of behavioral disorders (see letters from Drs. Thompson and Higley).

What is noteworthy about these and other areas of rhesus research is the increasing application of molecular and other genetic methods. The letters supplementing this document provide descriptions of ongoing and desired applications of genetic methods in these investigations. *It can be stated with certainty that the community working with rhesus macaque is prepared to take full advantage of the genomic sequence when it becomes available.*

2. *Informing human biology.* Studies of the normal adult physiology of primates make significant contributions to our understanding of human biology and health. Much of this work is relatively non-invasive and requires little manipulation of individual animals. However, macaques are appropriate for invasive and terminal studies also. Rhesus are commonly and readily bred and maintained under highly controlled laboratory circumstances, which makes them excellent subjects for well controlled studies of diet, or exposure to chemical or biological insults. For example, rhesus are currently being used in long-term studies of calorie restriction and the physiology of aging (e.g. Lane et al 2000, Edwards et al 2001).

As an example, given the large pedigrees of rhesus monkeys that are now available, there is significant opportunity to study gene-environment interaction in the control of phenotypes. This requires large numbers of genealogically related individuals; all examined using the same experimental protocol. This approach has been successful using pedigreed baboons (Mahaney et al 1999). Nine LOD scores over 3.0 have been obtained in this pedigree for traits related to lipids and cholesterol, hypertension and obesity (e.g. Kammerer et al 2001, Martin et al 2001). This degree of experimental control is simply not possible in human families, and illustrates the unique opportunities that Old World monkeys provide for investigating gene-environment interaction. *The expanding pedigrees of NIH supported rhesus macaques will make this type of genetic analysis increasingly feasible, and a genomic sequence would dramatically improve the power of such approaches.*

3. *Informing the human sequence.* As discussed above, paleontological information, plus a small amount of genomic sequence available from rhesus, suggests that the overall DNA sequence of rhesus is about 5-7.5% different from the human sequence (Stewart and Disotell 1998, Page and Goodman 2001). This level of divergence is not as useful for looking at very close or recent sequence variants (SNPs) as the chimpanzee or other apes. Moreover, strongly conserved regions, i.e. implying some functional significance, is more readily identified with a more distant relative such as mouse or another mammal. These limitations relate, of course, to nucleotide-sized variations and, even within these, there will be some information to be gleaned, especially as informatics tools become tuned to the different evolutionary distances.

However, comparative genomics is more than identifying these conserved regions and also includes distributions of repeats, such as segmental duplications, and other features. The rhesus genome sequence is useful for these comparisons and will provide opportunities for studies that enhance our understanding of the human sequence. For example, it is clear the numerous regions of the human genome occur as large blocks of DNA that are duplicated. Many of these duplicated segments appear to have occurred during the radiation of the ape lineage, show sequence divergence, and are unique to human. In Old World monkeys like rhesus, these regions are not duplicated and their sequence would provide clues to the ancestral segments, allowing reconstruction of the events that occurred on the way to the human genome (see letter from Dr. Eichler for a detailed explanation). *Rhesus genome sequence and structure does indeed fill a gap in the lineage that is needed to follow human evolution.*

4. *Providing connection between sequences of non-human organisms and humans.* The evolutionary position of rhesus macaque places it intermediate between chimpanzee and rodents in relation to humans. This will not only help indicate which aspects of the human genome are shared among various primates and which aspects are of more recent origin, but it will provide a perspective on the comparisons to mouse, rat and other mammals. Once human, mouse, and rat sequences are completed, there will be many detailed comparisons done to reconstruct and interpret the history and content of the mammalian genome. The greater distance of rhesus from human than the chimpanzee will make rhesus a more useful intermediary between the rodents (and other mammals to come in the future) and the human genome (see letter from Dr. O'Brien).

5. *Expand understanding of basic biological processes relevant to human health.* As discussed above, rhesus macaques are used extensively in a wide range of research relevant to human health. However, sequence information can greatly accelerate ongoing research in a number of ways. In particular, sequence data will enable investigators to pursue functional genomics in this species. As one example, expression array analysis will be greatly enhanced if rhesus sequence data is available. While human arrays can be and have been used for studies with nonhuman primate samples, problems arise because the human sequences are too divergent from the rhesus RNA tested. Feldman, Katze and Bumgarner have recently sequenced a number of EST's from rhesus monkeys (see GenBank accession numbers BM423011- BM423313). The mean value of sequence similarity between human and rhesus is about 95%, but there is a long tail to the distribution, consisting of sequences with much lower similarity. Unpublished data from the Katze lab suggests these low similarity sequences produce significantly lower signal when hybridized to arrays made with human cDNAs, thus reducing the value of the quantitative expression array data. The letters from Drs. McConkey, Cheverud, and Hacia also describe the problems encountered using human arrays to study rhesus biology.

Functional genomics applied to rhesus and other closely related primates (e.g. baboons and other macaques) can make outstanding contributions to many fields within medicine and human biology. It is unethical and/or impractical to directly investigate gene expression or protein composition within cells of the developing human embryo or fetus. With appropriate IACUC approval, fetal or embryonic tissue from any developmental stage and any organ system can be obtained from rhesus. With the growing interest in stem cell technology and the desire to manipulate stem cells, knowledge of the genetics and genomics of primate fetal development will prove invaluable to human stem cell research. A similar situation exists with pharmacogenomics where monkeys can be challenged with experimental or approved pharmaceuticals to generate detailed profiles of cellular responses to drugs. As noted earlier, these types of applications are important in drug development in industry and there is great interest in obtaining a rhesus genome sequence for these studies.

6. *Provide new surrogate systems for human experimentation such as new disease models.* As described above, rhesus macaques already provide a wide range of experimental models and surrogate systems for understanding human biology. The rhesus genomic sequence will encourage and facilitate new experimental strategies for existing models and possibly lead to the development of new disease models.

7. *Facilitate ability to perform direct genetics or positional mapping.* A genetic linkage map has already been developed for a close relative of rhesus (the baboon, Rogers et al 2000) and a similar map is under development for rhesus (NCRR R01 RR15383, J. Rogers, P.I.). A 10 centiMorgan map of the entire rhesus genome will be available in about 12 months, and this will

create new opportunities to use large multi-generation pedigrees of rhesus for gene mapping and gene identification studies. Analyses in baboons have already demonstrated the feasibility and value of quantitative trait linkage mapping in large pedigrees of nonhuman primates (e.g. Martin et al 2001, Kammerer et al 2001). Positional cloning studies are possible without the genomic sequence for rhesus, but the pace of progress when moving from initial QTL linkage result to identification of functional gene and mutation will be dramatically accelerated in macaques as well as in closely related baboons if the rhesus genome is fully available.

8. *Expand our understanding of evolution in general and human evolution in particular.* As discussed earlier, comparison of genomic sequences between human, chimpanzee, and rhesus will generate many opportunities for the analysis of individual genes, gene families, families of repetitive elements and other components of the genome. A picture of genome evolution will become much more highly focused. Even the most recent events, occurring after divergence of the Ape line from rhesus, will benefit from the rhesus sequence as an outgroup to provide insight into the ancestral genome arrangement.

III. Additional Applications of Genome Sequence from Rhesus Macaques

The use of rhesus macaques in genetic studies is dramatically increasing. Numerous statistics could be cited, but four items will serve to illustrate the growing importance of rhesus genetic information. First, sequencing of expressed genes from rhesus macaques is already a significant element of primate biomedical genetics. There are more than 5000 entries in GenBank under "rhesus or mulatta." A major cDNA sequencing program is underway at the University of Washington which will substantially increase this number over the next year. Other groups at University of Pittsburgh, Emory University and elsewhere are also conducting cDNA sequencing for rhesus. Second, genomic sequencing efforts have now begun. Dr. S. Zhao from TIGR has begun a large-scale effort to sequence BAC-ends from rhesus, and low coverage BAC sequencing is underway at the BCM-HGSC. Third, a radiation hybrid map is under development by Dr. Steve O'Brien and his research group (Murphy et al. 2001) and a second effort is planned by Dr. Leslie Lyons at UC-Davis (NIH proposal under review). Fourth, a genetic linkage map is under construction by Dr. Rogers. This initial linkage map will have 10 cM resolution, consist of at least 300 microsatellite loci already mapped in the human genome, and be complete in about 12 months. While the resources for genetic analysis of rhesus macaques are not as well developed as for some other mammals, they are improving rapidly. These resources will be well in hand by the time the genome sequence will be ready. We next present several examples of how researchers will use genome sequence from rhesus macaques.

1) Researchers want the catalog of rhesus macaque genes. Recent analyses demonstrate that the number and nature of functional genes in many gene families differs across closely related species (*e.g.* the *morpheus* gene family in humans and apes, or the MHC gene cluster or T-cell receptor family in rhesus versus humans). As Dr. Letvin outlines in his attached letter, knowing the complete complement of genes related to immune function will benefit his infectious disease research. Wherever rhesus is used as a model for human processes, knowing exactly how the rhesus genetic constitution differs from the human will be important. Given that there can be differences in expression level or tissue distribution, cDNA libraries may not include the complete catalog. Complete genome sequencing will provide the required information.

2) One immediate application of rhesus genome sequence data will be the acceleration of efforts to map and identify quantitative trait loci (QTLs) that influence human disease processes and disease susceptibility. Several QTLs have been successfully mapped using the baboon

linkage map, including QTLs related to hypertension (Kammerer et al. 2001), endocrine function (Martin et al. 2001) and low-density lipoprotein cholesterol (Kammerer et al. in press). However, it is difficult and time-consuming to identify all the functional genes in the regions identified by these QTL mapping studies, because too little physical mapping or EST information is currently available for any Old World monkey. Complete genome sequence for rhesus would greatly accelerate the process of determining the specific gene responsible for these linkage mapping results. Because more than 800 rhesus will be genotyped for a similar genetic linkage map in about 12 months, access to rhesus genomic sequence data will benefit QTL mapping programs employing both baboons and macaques.

3) One of the most significant recent advances in human genetics is the initial description of the pattern of linkage disequilibrium (LD) across the human genome. Many researchers are confident that association studies using linkage disequilibrium among single nucleotide polymorphisms (SNPs) will lead to identification of many disease genes. The pattern of LD in the human genome is complex, the result of multiple processes including but not limited to population demography, rates of mutation, natural selection and genetic drift.

An intriguing, but speculative possibility is the use of primate genomics to learn more about human LD. This can be either by comparative studies, to identify conserved or ancestral sequences, as well as a model of processes leading to patterns of LD in primate populations. Rodents or domestic animals will probably not provide such useful information about patterns of LD in natural outbred populations such as humans, due to their inbreeding and evolutionary distance. Analysis of LD in chimpanzee populations is encouraging for comparative studies of SNPs, but may ultimately also be of limited value, as chimpanzee populations have recently undergone severe decline. Chimpanzees are endangered, and their populations have over the past few hundred years been fragmented and disturbed. On the other hand, rhesus macaques are a highly successful, widespread and numerous species. Rhesus populations are much less disturbed and fragmented than ape populations. Rhesus exhibit population demography and social organization based around social groups that consist of closely related kin, and which inhabit overlapping territories where several groups compete for resources. Although human populations have increased dramatically over the past few thousand years, for nearly all of our evolutionary history since divergence from the other apes, we lived in populations that were probably organized in much the same way as rhesus. It is reasonable to assume that SNP identification and analyses of LD in small segments (a few megabases) of the rhesus genome could provide an animal model of LD in the human genome, and may assist in reconstructing the evolutionary causes of the LD observed in the modern human genome.

IV. Strategic Issues

1. *Demand for new sequence data.* There is tremendous demand for DNA sequence information concerning rhesus macaques. This species was the consensus choice for intensive genomic studies among participants in the Primate Genomics Workshop held in January 2001 in Seattle and jointly sponsored by the University of Washington Regional Primate Research Center and the National Center for Research Resources. Investigators from a wide variety of institutions met to discuss the current state of primate genomics and how progress in this field could be encouraged. An Executive Committee was chosen and given the task of writing a summary report for Dr. Judith Vaitukaitis, Director of NCRR. The major recommendations of that report were that intensive genomic analyses should be pursued in a number of primate species, and that the largest effort (including whole genome sequencing) should be directed at

rhesus macaques. *These activities indicate the eagerness and readiness of the primate research community to take advantage of a rhesus genomic sequence.*

As discussed earlier, most of the primates used in biomedical research in the U.S. are rhesus monkeys. In addition, three other commonly used species (baboons, *Papio hamadryas*; cynomolgus macaques, *Macaca fascicularis* and pig-tailed macaques, *Macaca nemestrina*) are all closely related to rhesus macaque. All species of macaques will be no more than about 1% different in genomic sequence, and baboons will be 1-1.5% different from rhesus (Stewart and Disotell 1998, Page and Goodman 2001). Access to extensive genomic sequence from rhesus will significantly benefit research using all these species. Sequence from chimpanzee will not provide any additional information beyond that derived from access to the available human sequence because the evolutionary distance, and sequence similarity, between humans and rhesus (or baboons) is the same as the distance between chimpanzees and rhesus (or baboon).

2. *Suitability of organism for experimentation.* Rhesus macaques are highly amenable to experimentation and investigation. This species is available in large numbers from a variety of institutions and breeding operations, and is already used in 60-75% of NIH funded projects that utilize nonhuman primates. In addition to the research described above, rhesus are also used in more novel and ground-breaking research. The first nonhuman primate expressing an exogenous gene has been produced (Chan et al 2001), and this animal was a rhesus monkey. While routine production of transgenic primates is not yet possible, development of transgenic technology applied to both rhesus monkeys and baboons will be stimulated by a genomic sequence. Stem cell research will also exploit the strengths of rhesus macaque models over the coming years. A search of the PubMed database for "stem cell and rhesus" generated 210 citations.

3. *Rationale for complete sequence.* The applications discussed above, evolutionary studies with rhesus as outgroup, development of microarrays for gene expression studies, analysis of tissue distribution and other aspects of the rhesus proteome in industry, use of genomic sequence for various genetic purposes such as positional cloning, transgenics, linkage mapping, and other uses all require a complete genomic sequence.

4. *Cost of sequencing and readiness of DNA.* The genome size for rhesus macaques is expected to be similar to that of humans. A high-quality BAC library has been produced by Dr. Pieter deJong. The library is being used by several researchers (see below). Two different restriction enzymes (EcoRI and MboI) were used to make the library, and the average insert size is 160kb. The goal of this project is to obtain a sequence to an overall coverage of 5-fold redundancy using the existing library and other resources. New methodology (clone arrays) will be employed to reduce costs.

5. *Other partial support*. At the present time, there is no additional support for a rhesus sequencing project. However, given the range of research impacted by rhesus, there is the possibility of obtaining support from other institutes that will benefit from the sequence. Of course, leveraging NHGRI support requires a strong rating as to the project's importance. It will be more difficult to obtain this support if the project is perceived as not high priority.

There is whole genome sequencing of rhesus macaque underway at the BCM-HGSC. Dr. Aleksandar Milosavljevic, one of the bioinformatics faculty, has received two grants (one from NIH and one from NCRR) to work out methods to build maps of BAC clones for rhesus using clone array technology. This will support about 500,000 sequencing reads and the mapping of

about 3x clone coverage worth of BACs. As noted elsewhere, there is also cDNA sequencing underway which will be important in the annotation of the genomic sequence.

V. Collaboration with BCM-HGSC

Dr. Richard Gibbs and Dr. George Weinstock of the Baylor College of Medicine Human Genome Sequencing Center have expressed strong interest in performing the rhesus macaque genome project. The other authors of this white paper have visited the BCM-HGSC and discussed details of the project. We propose sequencing the entire genome to five-fold sequence coverage. In addition, we propose that support be allocated for more complete coverage and finishing of sequence for up to 500 megabases within the rhesus genome. The specific regions to be finished would be chosen at a later time based on perceived scientific interest and importance. For example, given the significance of rhesus macaques for AIDS research and immunology, and for neurobiology, it might be valuable to produce finished sequence for the MHC region of the rhesus genome, and for one or more regions that containing medically significant genes expressed in the nervous system. Moreover higher quality sequencing of QTL regions is a prerequisite to identifying alleles responsible for phenotypes under study.

The strategy for sequencing the genome is in response to the observation that, although conservation between human and macaque in coding regions averages 92-95%, there are considerable regions with less conservation, as low as 85% in non-coding regions. Any strategy of low coverage followed by comparison with the human sequence can miss more divergent features of interest. The strategy will thus aim for a higher coverage of the genome. This will employ a mixed approach of low coverage (1-2x coverage BAC skims) sequencing of a minimal tiling path (MTP) of BAC clones with the rest of the sequence coming from whole genome shotgun reads. This strategy has been employed to sequence the rat genome by the BCM-HGSC.

There are at present two BAC libraries that have been prepared by Pieter de Jong (from EcoRI- and MboI-generated fragment sets, average insert size of 160kb). Ten of these clones have been sequenced at the BCM-HGSC and no unforeseen issues exist. Dr. Shaying Zhao at The Institute for Genomic Research (TIGR) has sequenced both ends of approximately 4600 BAC clones from the deJong library, and finds that most of these sequences can be readily aligned to the human sequence (see letter) further indicating the high quality of the BAC library. A MTP is being derived at the BCM-HGSC by using pools of arrayed BAC clones, lightly sequencing each BAC, and deconvoluting the pools by comparing the reads to the human genome. This approach (Cai et al 2001) minimizes the number of BAC DNA preparations and shotgun libraries required. The pooling approach can likewise be applied to the clones in the MTP which will be sequenced to higher coverage than during this mapping stage. This will be used to sequence the honey bee genome (130 MB) as a final development step and should be fully operational by the time the rhesus project is ready to start.

Whole genome shotgun reads will come from 3kb, 10kb, and 40kb libraries. All sequencing will be in plasmids and end pairing will be maintained at high (>90%) fidelity. The whole genome shotgun reads will be binned into appropriate BACs and the genome will be assembled using the ATLAS whole genome assembly software developed at BCM-HGSC for the rat genome project. The final assembly will also take advantage of the human genome sequence

to provide additional validation. In general, all of the methodology employed in this project is being used in the rat and other genome projects and will be proven technology.

The entire project would require approximately 35 million successful reads which could be performed at the BCM-HGSC in less than two years. However, with the addition of finishing it is likely that this time would be reduced by collaborating with other groups. Thus the entire project could be completed in two years.

VI. Final Comments

Because primates are the closest living relatives to humans, it is critical to develop one or more primates as model organisms for biomedical research. The availability of a genomic sequence is a must for the success of such a model system. The rhesus macaque is the organism of choice because of the existing large base of research already underway. Other candidates such as the chimpanzee, which has an ongoing genome project in its favor, are not as useful in research for practical and ethical reasons. Other primates have not been used as widely as the rhesus macaque due to practical and other considerations.

While the rhesus macaque is not the ideal primate for evolutionary studies, it nevertheless has significant value. A complete sequence of the rhesus macaque will undoubtedly shed light on many evolutionary details of human origins and aid in the interpretation of the meaning of human comparisons to other primates and more distant mammals. Other organisms that would be more desirable from an evolutionists standpoint do not offer the payoff as a biomedical model system that rhesus does. In the balance between these two factors, the leveraged payoff for biomedical research that will be gained from the rhesus sequence dominates the modest loss of resolution of evolutionary matters resulting from use of the rhesus sequence.

The strength of the rhesus genome project is its direct relevance to the goals of "development of innovative and improved methods of diagnosis, treatment or prevention" of human disease. It offers novel opportunities to increase our understanding of human biology both because it is closer to humans than the mouse or rat and it is a more useful system than the other primates. While comparative sequence analysis will contribute to our understanding of human biology, it remains to be seen how large the contribution to medical science will be of the human-chimpanzee comparison without chimpanzee models to test hypotheses. It is likely that rhesus will play a role in testing these hypotheses and its complete sequence will be critical to the design and interpretation of the results. Fortunately, the significance of the rhesus genome project does not rely solely on the high risk comparative approach of the chimpanzee because of the substantial existing research applications for rhesus.

The evidence indicates that many in the rhesus research community are already applying genetic and genomic methods in their studies, despite the minimal state of the resources. A number of examples are provided in this presentation, but additional cases are described in the accompanying letters. *The rhesus macaque genome project should be viewed as an investment in this research.* It will certainly lead to further and more rapid development of genetic methods, applications in academic and industrial research, and a greater understanding of the human.

[1]Southwest Foundation for Biomedical Research, and Southwest Regional Primate Research Center, San Antonio, TX

[2]Washington Regional Primate Research Center, Seattle, WA

[3]Baylor College of Medicine Human Genome Sequencing Center, Houston, TX

Preparation of this white paper was largely by JR and GMW. The author list acknowledges other individuals who made major contributions. However not all of the listed authors had an opportunity to approve the final text. We also wish to thank the many individuals who provided letters of support and descriptions of their research, many of which have been included with the submission of this document.

Literature Cited

Cai, W.-W., R. Chen, R.A. Gibbs and A. Bradley (2001) "A clone-array pooled shotgun strategy for sequencing large genomes" **Genome Res.** 11:1619-1623.

Chan, A.W., K.Y. Chong, C. Martinovich, C. Simerly and G. Schatten. (2001) "Transgenic monkeys produced by retroviral gene transfer into mature oocytes" **Science** 291:309-312.

Dove, A. (2000) "New bill mandates sanctuary system for retired chimps" **Nat. Medicine** 6:9.

Edwards, I.J., L.L. Rudel, J.G. Terry, J.W. Kemnitz, R. Weindruch, D.J. Zaccaro and W.T. Cefalu (2001) "Caloric restriction lowers plasma lipoprotein (a) in male but not female rhesus monkeys" **Exp. Gerontol.** 36: 1413-1418.

Eizirik, E., W.J. Murphy and S.J. O'Brien (2001) "Molecular dating and biogeography of the early placental mammal radiation" **J. Heredity** 92:212-219.

Fellows, P.F., M.K. Linscott, B.E. Ivins, M. Pitt, C. Rossi, P.H. Gibbs and A.M. Friedlander (2001) "Efficacy of a human anthrax vaccine in guinea pigs, rabbits and rhesus macaques against challenge by Bacillus anthracis isolates of diverse geographic origin" **Vaccine** 19:3241-3247.

Hotta K, Funahashi T, Bodkin NL, Ortmeyer HK, Arita Y, Hansen BC, and Matsuzawa Y. 2001. Circulating concentrations of the adipocyte protein adiponectin are decreased in parallel with reduced insulin sensitivity during the progression to type 2 diabetes in rhesus monkeys. **Diabetes** 50:1126-1133.

Kammerer, C.M., D.L. Rainwater, L.A. Cox, J.L. Schneider, M.C. Mahaney, J. Rogers and J.L. VandeBerg (in press) A locus controlling LDL cholesterol response to dietary cholesterol is on the baboon homologue of human chromosome 6. Arter. Ther. Vasc. Biol.

Kammerer, C.M., L.A. Cox, M.C. Mahaney, J. Rogers and R.E. Shade (2001) "Sodium-lithium countertransport activity is linked to chromosome 5 in baboons" **Hypertension** 37:398-402.

Kumar, S. and S.B. Hedges (1998) "A timescale for vertebrate evolution" **Nature** 392: 917-920.

Lane, M.A., E.M. Tilmont, H. DeAngelis, A. Handy, D.K. Ingram, J.W. Kemnitz and G.S. Roth (2000) "Short-term calorie restriction improves disease-related markers in older male rhesus monkeys (Macaca mulatta)" **Mech. Ageing Dev.** 112: 185-196.

Lozier, J.N., G. Casko, T.H. Mondoro, D.M. Krizek, M.E. Metzger, R. Costello, J.G. Vostal, M.E. Rick, R.E. Donahue and R.A. Morgan (2002) "Toxicity of a first-generation adenoviral vector in rhesus macaques" **Hum. Gene Ther.** 13: 113-124.

Mahaney, M.C., J. Blangero, D.L. Rainwater, G.E. Mott, A.G. Comuzzie, J.W. MacCluer and J.L. VandeBerg (1999) "Pleiotropy and genotype by diet interaction in a baboon model for atherosclerosis. **Arterioscler. Thromb. Vasc. Biol.** 19: 1134-1141.

Martin, L.J., J. Blangero, J. Rogers, M.C. Mahaney, J.E. Hixson, K.D. Carey, P.A. Morin and A.G. Comuzzie (2001) "A quantitative trait locus influencing estrogen maps to a region homologous to human chromosome 20" **Physiol. Genomics** 5:75-80.

Osada, N., M. Hida, J. Kusuda, R. Tanuma, M. Hirada, K. Terao, Y. Suzuki, S. Sugano and K. Hashimoto (2002) "Prediction of unidentified human genes on the basis of sequence similarity to novel cDNAs from cynomolgus monkey brain" **Genome Biol.** 3:RESEARCH0006.

Page, S.L. and M. Goodman (2001) "Catarrhine phylogeny: Noncoding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade" **Mol. Phylogenet. Evol.** 18:14-25.

Rogers, J., M.C. Mahaney, S.M. Witte, S. Nair, D. Newman, S. Wedel, L.A. Rodriguez, K.S. Rice, S.H. Slifer, A. Perelygin, M. Slifer, P. Palladino-Negro, T. Newman, K. Chambers, G. Joslyn, P. Parry and P.A. Morin (2000) "A genetic linkage map of the baboon *(Papio hamadryas)* genome based on human microsatellite polymorphisms" **Genomics** 67: 237-247.

Stewart, C.-B. and T.R. Disotell (1998) "Primate evolution – in and out of Africa" **Current Biology** 8:R582-R588.

Takatoku, M., S. Sellers, B.A. Agricola, M.E. Metzer, I. Kato, R.E. Donahue and C.E. Dunbar (2001) "Avoidance of stimulation improves engraftment of cultured and retrovirally transduced hematopoietic cells in primates" **J. Clin. Invest.** 108:447-455.

Tarantal, A.F., M.L. Marthas, J.P. Shaw, K. Cundy and N. Bischofberger (1999) "Administration of 9-2-R-(phosphonomethoxy)propyladenine (PMPA) to gravid and infant rhesus macaques (Macaca mulatta): safety and efficacy studies" **J. Acquir. Immun. Defic. Syndr. Hum. Retrovirol.** 20:323-333.

Winegar DA, Brown PJ, Wilkison WO, Lewis MC, Ott RJ, Tong WQ, Brown HR, Lehmann JM, Kliewer SA, Plunket KD, Way JM, Bodkin NL, and Hansen BC. 2001. Effects of fenofibrate on lipid parameters in obese rhesus monkeys. **J Lipid Res** 42:1543-1551.