Porcine Genomic Sequencing Initiative

Gary Rohrer, USDA-ARS, US Meat Animal Research Center; Jonathan E. Beever, University of Illinois;
Max F.Rothschild, Iowa State University; Lawrence Schook*, University of Illinois
Richard Gibbs and George Weinstock, Baylor College of Medicine, Human Genome Sequencing Center

[*Corresponding author: 329 ERML, 1201 W. Gregory Dr., Urbana, IL 61801; schook@uiuc.edu;
(tel) 217-265-5326; (fax) 217-244-5617]

A. Specific biological rationales for the utility of the porcine sequence information

Rationale and Objectives. Completion of the human genome sequence provides the starting point for understanding the genetic complexity of humans and how genetic variation contributes to diverse phenotypes and disease. It is clear that model organisms have played an invaluable role in the synthesis of this understanding. It is also noted that additional species must be sequenced to resolve the genetic complexity of human evolution and to effectively extrapolate genetic information from comparative (veterinary) medicine to human medicine. Certainly the pig has been a valuable biomedical model organism and its role will expand in the future. The pig also represents an evolutionary clade distinct from primates or rodents, and thus, provides considerable power in the analysis of human genomic sequences. The pig, a domesticated eutherian mammal, has co-evolved with humans and represents a taxa with diverse selected phenotypes [Rothschild and Ruvinsky, 1998]. The pig has a central position in the scientific and veterinary medical communities that supports the utility of securing genome sequence information. Thus, this "white paper" provides scientific justification for sequencing the porcine genome (6X coverage) to identify new genes and novel regulatory elements in the pig and in humans, mice and rats. The porcine genome will serve as a reference non-primate, non-rodent, eutherian genome. The recent ability to genetically modify the porcine genome, genetically manipulate embryonic fibroblasts, and 'clone' genetically modified somatic cells through nuclear transfer attests to how the pig can provide relevant genetic models (of appropriate phenotypes). This further demonstrates the unique role the pig will play in biomedical research, hence warranting the value for genomic sequencing.

The porcine genome is uniquely positioned for genomic sequencing because of the advanced stage of the necessary reagents. A porcine BAC map with 20X coverage, constructed via an international consortium, will be fingerprinted and all fingerprinted clones end-sequenced by June, 2003. This resource will permit selection of the minimum tilling path of BAC clones to be sequenced and complement a whole-genome shotgun sequencing approach. This approach was selected since its affords increased efficiency, saving time and money, and yields a better product since the BAC map will be completed prior to initiation of genomic sequencing. Linking the sequence to the BAC clone map allows for subsequent targeted closure of the genomic sequence in regions of particular interest. This strategy is justified through the outcomes associated with the human, mouse, and rat sequencing efforts that were done in parallel with the BAC map development.

Improving Human Health. During its domestication, the pig has undergone intense selection pressures for various phenotypes throughout the world. First domesticated in Asia from the Wild Boar, germplasm was quickly moved around the world by explorers and used for food and products (fat for making gun powder). Intense selection and breeding has provided distinct phenotypes differing in metabolism, fecundity, disease resistance and in the products they produce for humans. These selective pressures have differentiated subpopulations and

produced phenotypes extremely relevant to current and future human health research. The selection of the "mini" and "micro" pigs for size, independently by investigators throughout the world, attests to the global relevance of this experimental animal in biomedical research. Clearly, understanding genetic interactions with environmental factors will be a major focus of future biomedical research. The porcine model is also relevant to human health research priorities such as obesity, female health, cardiovascular disease, nutritional studies with respect to the pig being an omnivore, and communicable diseases [Reviewed in Tumbleson and Schook, 1996]. The pig provides a valuable biological model in these priority areas because of the vast amount of research that has been conducted with respect to genetic and environmental interactions associated with complex, polygenic physiological traits. The domesticated pig has also played an extensive role as a source of biological material in physiological and biochemical research. Use of the pig for biochemistry, enzymology, endocrinology, reproduction, and nutrition research has contributed significantly to the continual improvement of human health.

Informing human biology. The animal sciences have contributed greatly to the basic understanding of human development and physiology. Classical endocrinology studies in farm animals led to the current understanding of several reproductive and pituitary hormones. The composition of insulin was first determined for porcine insulin that was used for several decades to treat human diabetes.

The porcine model has provided a fundamental research platform for developing human reproductive techniques and for studying reproductive diseases. Ongoing research using the pig to study cancer and diabetes are underway. The pig has many similarities in structure and function with humans including size, feeding patterns, digestive physiology, dietary habits, kidney structure and function, pulmonary vascular bed structure, propensity to obesity, respiratory rates and social behaviors [Tumbelson and Schook, 1996]. Since the pig is an omnivore, it provides an adaptable model to evaluate chronic and acute exposures to xenobiotics such as alcohol, tobacco, feed additives and environmental pollutants. Swine have been used as models to evaluate alcoholism, diabetes, total parenteral nutrition, organ transplantation, atherosclerosis, exercise, hypertension, melanoma, nephropathy, dermal healing, shock and degenerative retinal diseases. A severe shortage of organs and tissues for transplantation has also stimulated increased consideration of pigs as a potential solution, particularly with the recent ability to genetically modified pigs to overcome acute rejection [Lai et al., 2002].

Research comparing different pig breeds has identified genetic differences in fat deposition of different tissues and organs [Rothschild and Ruvinsky, 1998; Malek et al., 2001a,b]. Such information provides an experimental model for understanding obesity and nutrition (from prenatal nutrition to aged cohorts). Porcine resource populations have been selected for phenotypic variation in bone density [osteoporosis], sex-expressed nutritional and reproductive characteristics, and growth and development (embryonic, pre- and post-natal). The porcine model will also be invaluable to study host-pathogen interactions for food safety (i.e. Salmonella), potential biological warfare agents (African swine fever and Foot and Mouth Disease) and agents that affect food security and human health (i.e. porcine endogenous retroviruses and other zoonotic diseases). Using comparative genomics has also demonstrated new models for metabolism linked to obesity-induced diabetes [Milan et al., 2000].

Informing human sequence and connecting human and pig sequences. The pig genome is of similar size (3×10^9 bp), complexity and chromosomal organization ($2n = 38$, including meta- and acrocentric chromosomes) as the human genome. Comparative genetic maps have indicated that the porcine and human genomes are more similarly organized than when either is

compared to the mouse. The mean length of conserved syntenic segments between human and pig is approximately twice as long as the average length of conserved syntenic segments between human and mouse [Ellergren et al., 1994; Rettenberger et al., 1995]. Furthermore, the organizational similarities between the human and porcine genomes are reflected in similarities at the nucleotide level. In more than 600 comparisons of non-coding DNAs aligned by orthologous exonic sequences on human chromosome 7, pig (and cow, cat and dog) sequences consistently grouped closer to human and non-human primate sequences than did rodent (mouse and rat) sequences [Green, 2002]. Furthermore, the rodent genomes are evolving at a different (faster) rate than other representative genomes. For these reasons it is necessary to produce the genomic sequence for eutherian mammals outside the primate and rodent lineages in order to better assemble and annotate the human sequence. The rich genetic history and strong molecular resources currently available clearly identify the domestic pig as the appropriate choice for a mammalian genome sequence project.

Comparative genome information between humans and pigs is well established; thus, a comparative map-based approach is possible for the identification of genes influencing complex traits (http://www.toulouse.inra.fr/lgc/pig/compare/compare.html). Over the past decade, tremendous progress has been made mapping and characterizing the swine genome. Currently, moderate to high resolution genetic linkage maps containing highly polymorphic loci have been produced using independent mapping populations [Rohrer et al., 1996]. Additionally, physical mapping methods such as somatic cell hybrid analysis, *in situ* hybridization and ZOO-FISH have been employed to enrich the gene map and to perform comparative analysis with map-rich species such as the human and mouse. To date, 2,390 mapped loci are cataloged for the pig genome (http://www.thearkdb.org). Recently, whole-genome radiation hybrid (WG-RH) panels (7,500 and 12,500 rad) have been generated for swine [Hawken et al. 1999; Yerle et al. 2002] resulting in yet another rapid increase in the number of loci mapped. Even more recently, the swine genomics community has acquired access to resources such as bacterial artificial chromosome (BAC) libraries [Fahrenkrug et al., 2001; Anderson et al., 2000] providing approximately 35X coverage of the swine genome. These BAC resources have facilitated the production of high-resolution physical maps in specific chromosomal regions [Rogel-Gaillard et al., 1999; Milan et al., 2000] and support the construction of sequence-ready mapping resources for the porcine genome.

Expand Knowledge of Basic Biological Processes Relevant to human Health. The discovery that mammalian genomes probably contain only 30,000-40,000 genes suggests that alternate transcripts and post-translational modification must play a greater role in phenotypic expression than previously appreciated. We also expect single gene products to affect different traits or disease states, dependent on temporal and spatial presence of gene products. As an omnivore, the pig is prone to many of the same dietary health problems as humans. Depending on diet and genetics, pigs can suffer from hypertension, hypercholesterol, dyslipidemia, insulin resistance and atherosclerosis. The pig has mutations in similar genes affecting these metabolic disorders (for example ApoB and LDLR for hypercholesterol) [Ajiello et al., 1994; Hasler-Rapacz et al., 1998]. Piglets are the preferred model organism to develop human infant formulas as their nutritional needs are comparable to that of human infants. Because of the similar digestive tract, pigs are also susceptible to similar enteric food borne pathogens (Salmonella and enterohemmoragic *E. coli*) and pig intestinal linings are used for *in vitro* studies of interactions with the intestine and these pathogens. Pigs are also susceptible to gastric ulcers that apparently are induced by diet and stress. Additional anatomical similarities with humans are renal morphology, eye structure, skin and tooth development. The pig is also one of few animals that will voluntarily eat to obesity as well as being susceptible to alcoholism.

There are two reasons for research to investigate obesity-related genes in the pig. First the pig is a more realistic model organism for human obesity due to physiological similarities [Tumbleson and Schook, 1996]. As the pig is a true omnivore, the molecular basis and digestive tract anatomy of the pig is much closer to humans than any laboratory animal species, so identified significant DNA polymorphisms of obesity-related genes in the pig genome might provide useful targets for the genetic study of human obesity. The second reason is that the genetic components of human obesity can play important roles in pig performance traits such as fatness, growth rate, and feed intake. As pork is the leading source of animal protein in the world, this research can provide valuable information for efficient production of a leaner, healthier and more economical source of animal protein for human consumption.

Surrogate Systems for Human Experimentation. The domesticated pig has provided numerous surrogate experimental models for biomedical research. There has been a long tradition of using abattoir tissues for the purification of enzymes and the elucidation of metabolic pathways. These tissues have also served as initial biologicals with bovine and porcine insulins providing pre-recombinant DNA therapeutics and purified enzymes used to determine crystalline structure. Porcine gamete biology has played a critical role in our understanding of stem cells and *in vitro* fertilization. Because of the wealth of biological information using the porcine system it has increasingly become important for studying epigenetic effects as well as unraveling genomic imprinting. The recent demonstration that pigs can be cloned using *in vitro* cloning systems provides an invaluable technology platform for developing relevant clones of genetic models for biomedical research [Betthauser et al., 2000]. In addition, a major obstacle for producing cloned genetically modified pigs has been overcome [Lai et al., 2002]. These investigators have created a nuclear transfer technology using clonal fetal fibroblasts as nuclear donors for the production of gene specific knockouts. This technology platform has significant applications beyond xenotransplantation and clearly the availability of genomic sequences will facilitate the broader utility of the pig as a surrogate system for human experimentation.

The phenotypic diversity of hundreds of porcine breeds distributed throughout the world provides a tremendous resource for "comparative phenomics", the application of comparative genomic principles to discovery of new genes underlying diverse phenotypes. In only a few thousand years, selective breeding has produced pig breeds that thrive in diverse environments (high altitude versus tropical), convert energy to muscle mass efficiently and rapidly, and tolerate specific pathogens. In many respects, breeds of pigs are similar to human ethnic groups with diverse geographic origins, except with exaggerated phenotypic diversity. There can be little doubt that the understanding of what makes porcine breeds different with respect to reproductive efficiency, bone structure, growth rates, fat deposition, altitude or heat tolerance, and resistance to specific pathogens will be important to understanding basic biological processes important to human health.

Facilitating the Ability to Perform "Directed Genetics" or "Positional Cloning". The porcine research community has a long history in quantitative genetics, and more recently in genomics research. The genetic contribution of many multi-genic traits in pigs is well documented and this knowledge has provided the basis for the identification and mapping of a growing number of quantitative trait loci (QTL) (Andersson et al., 1994; Milan et al., 2000; Rohrer et al., 1999; Bidanel et al., 2002; Malek et al., 2001a,b; Nezer et al., 2002).

The only limitation to performing direct genetic experiments and identifying genes underlying these traits is the lack of a complete genome sequence. Selection experiments, heterosis studies and breed comparisons have all been used in porcine genetic studies. Many populations have been used to map genes to large chromosomal regions but positional

mapping causal genes has been difficult. Sequencing the porcine genome and generating 100,000 SNP will provide additional polymorphic markers and positional candidate genes from the human and mouse map. Large populations with designed matings can be used to positionally map genes. The populations can be generated by natural reproduction, artificial insemination or assisted reproductive technologies. Clones can also be generated from fibroblasts, or stem cells and cryopreserved. This technology provides the opportunity for knock-out or knock-in experiments in an animal other than the mouse. Interspecies porcine hybrids are easily produced and are very valuable for knockout/in experiments and studying genomic imprinting [Andersson et al., 1994].

Expanding the Understanding of Evolutionary Processes. While all the eutherian mammalian orders probably diverged 70-80 million years ago, it is evident that some have genomes that are much more highly conserved relative to primates than others. Pigs, represent a clade [Green, 2002] with respect to sequence divergence, intermediate to and distinct from the primates and rodents. These domestic animal groups also are more conserved relative to humans than rodents with respect to total genomic structure as revealed by comparative gene mapping. They themselves are diverse taxa, of course, and over the next few years, each genome should be sequenced to reveal its evolutionary history and to facilitate the important role each animal plays in comparative medicine. The domestic pig is somewhat unique, however, in that it represents all the artiodactyls, a phenotypically diverse clade of animals with circumglobal distribution and conserved a number of chromosomal arms. The pig is also unique in that the Wild Boar, *Sus scrofa* from which the domesticated *Sus scrofa* was developed from is still present and has a distinct phenotype and different karyotype.

B. Strategic Issues

Demand for New Sequence. A recent CRISP search [September, 2002] demonstrated that 563 currently funded NIH grants use the domestic pig as an experimental model system. The porcine system is funded by at least six institutes to include transplantation, infectious disease, organ physiology including the eye, pharmacology including drugs of abuse, and metabolic diseases. Increased genetic analysis of the porcine genome has expanded significantly as demonstrated by the number of submissions contributed to various public databases. During the past year, the number of ESTs and mapped ESTs (total 111,551 and ca. 3,500, respectively) has increased significantly. The total number of porcine sequences in public databases ranks fifth, behind human, mouse, cattle and rat. The most recent clustering of EST data for pigs by TIGR (pig gene index 5.0, October 1, 2002; http://www.tigr.org/tdb/tgi/ssgi/) resulted 17,354 clusters (TCs) and 31,847 singletons. There is significant support for sequencing the porcine genome among the international scientific community. By the end of 2002, the Danish-Chinese Swine Genome Collaboration will be submitting 3,000,000 genomic reads into the public database from a whole-genome shotgun sequencing effort. The Danish-Chinese Swine Genome Collaboration will also submit an additional 1,000,000 EST sequences in mid-2003 [see letter indicating the availability of these resources for this project]. A Porcine BAC Consortium with global representation of government, university and industrial constituents has been established to produce a whole-genome BAC map. The non-profit Alliance for Animal Genome Research has provided support for development of this “white paper” and has provided leadership for convening a NAS/NRC workshop that is being supported by the USDA, NIH, DOE, and NSF to further define scientific objectives related to this and other sequencing initiatives. During the recent International Society of Animal Genetics (ISAG) Conference, the development of this “white paper” and its merits were discussed and a resolution was passed by ISAG to support this “white paper” and to establish a new standing committee to facilitate international coordination of member countries and scientists (see letter of support from

President, ISAG). Additional letters representing various constituents from around the globe are attached in Appendix. These letters summarize both the need to sequence the pig genome and the widespread support offered by companies and scientists from government and universities for the sequencing of the pig genome.

Thus, there is significant endorsement by the international and national research community and industry for the porcine genome sequencing initiative. Global participation and coordination of this initiative will be conducted through the International Society for Animal Genetics [ISAG] and the Comparative Genomics program of HUGO, and with annual meetings held in conjunction with the Plant and Animal Genomics Meeting. ISAG provides a framework for workshops and exchange of information that is critical for the implementation of this initiative and has served as a bridge with HUGO in the development of comparative genomics workshops. ISAG provided a workshop at its 2002 meeting in Germany that engaged the research community with respect to this porcine sequencing initiative.

Suitability of the Pig for Experimental Sequence. Domestic pigs have a long historical and economic association with human cultures. Consequently, pigs are abundant and represent as many different breeds that can be considered as biological equivalents to human ethnic and racial groups. Large and very deep pedigrees of pigs are available for research supported by phenotypic measurements and DNA is widely available from these pedigrees. Additional genetic diversity has been created in crosses such as domestic x Wild Boar and Western domestic x Chinese exotic breeds to produce high-density comparative genetic maps of type I markers. Reproductive technologies for the propagation of pigs are highly developed and historically were important in the development of assisted reproduction techniques in human medicine including the ability to clone pigs by nuclear transfer using fetal fibroblasts [Lai et al., 2002]. Experimental tissues are readily available in large quantities from abattoirs and the utilitarian aspect of the pig-human association minimizes ethical and social objections to the use of pigs in research. Due to their large litter size, and their anatomical features, pigs are appropriate for human medical experimentation and have been used to develop key methodologies in organ transplantation, and artificial hearts.

Rationale for Complete Sequence. Domestic pigs are but one of several species that are used extensively in biomedical research or serve as hosts for zoonotic diseases. Karyotypic data demonstrate extremely conserved genomes in these species, suggesting that a complete sequence of the porcine genome will provide a genomic matrix similar to that of humans. The other domestic animal genomes (cow, dog, cat) are clearly diverged and we are supportive of sequencing those genomes if end-sequenced BAC maps and other mapping resources are fully developed. A 1X genomic sequencing coverage or perhaps only transcript maps will serve the others species well.

The discovery of conserved sequences across species has proved valuable in the identification of novel genes and conserved regulatory elements. We propose that genomes of different primate species are not sufficiently diverged to identify many biologically significant elements and that human and rodent genomes are often too far removed on the molecular clock to find others. The genome of the pig, a non-primate, non-rodent placental mammal, must be sequenced to triangulate the comparative sequencing strategy for finding biologically important sequences.

State of Readiness and Cost of Sequencing. The first linkage maps were published in 1994 and current linkage maps collectively contain > 2900 loci including approximately 1,700 microsatellite markers and 1,200 SNP/RFLP markers (http://www.marc.usda.gov/). The maps

have been used to identify chromosomal regions that influence quantitative traits affecting growth, body composition, reproduction and immune response [Bidanel and Rothschild, 2002]. Numerous cDNA libraries have been developed from many different tissues at different physiological stages and more than 110,000 ESTs have been depositied at TIGR. The TIGR cluster analysis on all EST sequences and generated 49,200 non-redundant gene indices that are routinely used by the international research community. A subset of 1,000 ESTs has been selected for sequencing predicted introns (based upon human genomic sequence) to identify SNP for mapping on the linkage map. The same set of ESTs will be placed on the RH map [Hawken et al., 1999]. The EST mapping effort is designed to improve the human-pig comparative maps. The radiation hybrid map of the porcine genome integrates microsatellite markers with ESTs selected from BLAST hits with human sequences. This map contains 5,500 STS markers and is currently being assembled with 2,000 comparative markers (http://imprh.toulouse.inra.fr).

A coordinated international effort has been initiated to develop a porcine BAC map with two BAC libraries (RPCI-44 and CHORI-242) made by Pieter J. de Jong (pdejong@chori.org) and one library made at the Roslin Institute. An international consortium was developed to construct the porcine BAC map. USDA-ARS (U.S. Meat Animal Research Center, Clay Center, Nebraska), United Kingdom (Roslin Institute, Edinburgh, Scotland and Biotechnology and Biological Sciences Research Council), and the University of Illinois are the current participants. USDA-ARS and BBSRC have funded the Wellcome Trust Sanger Institute to fingerprint a 5X library developed by Roslin Institute and the 10X CHORI-242 library developed by Dr. de Jong. Sygen Inc. has also contributed funds to this effort. In coordination with this effort, Dr. J. Beever will fingerprint 5X of the RPCI-44 library with funding from USDA-CREES and the University of Illinois. INRA is completing the fingerprinting of their 5X INRA library data and through the exchange of BAC clones will be merged to permit for a comprehensive analysis. INRA has screened more than 1000 BACs from this library for known genes and markers and has mapped them on genetic and RH gene maps. INRA is pleased to share this set of BACs to facilitate anchoring of contigs.

Sequencing the ends of all fingerprinted BAC clones will be conducted. At the recent International Society for Animal Genetics (ISAG) meeting, many laboratories from Europe, Australia, Asia and North America agreed to map ESTs, microsatellite markers, genes and other STSs onto the BAC clones to integrate the linkage and RH maps to the BAC map. In addition, Monsanto (see enclosed letter of support) has agreed to provide the results of mapping 7,000 STS in the RPCI-44 library. The final product will represent 20X coverage of the porcine genome. The Roslin Institute will maintain a database to store fingerprint, sequence and STS mapping information. The BAC map and end sequencing will be completed by June, 2003.

The scientists drafting this white paper have coordinated with the Baylor College of Medicine (BCM) Human Genome Center in Houston to evaluate the readiness of domestic animal genomes for a sequencing initiative and to discuss the relative merits of porcine genomics in evolutionary and medical sciences. From our discussions emerged an enthusiastic consensus that because of its evolutionary history, its medical significance, the support of a strong research community, and the state of genomic information and resources already amassed, domestic pigs should be put forward as the first of several domestic animal genomes to be sequenced.

Collaboration between several individuals involved in this initiative demonstrates an intellectual dedication to the project and an institutional commitment to support genome sequencing. A close working relationship was established in 2001 between Drs. Weinstock and Gibbs from

BCM and Drs. Schook and Beever to initiate the construction of the BAC contig and BAC-end sequencing.

The cost of sequencing the porcine genome is estimated to be similar or less than sequencing the rat genome (~$50 million) because the BAC map will be completed prior to genomic sequencing and the experience gained by the Baylor College of Medicine Human Genome Sequencing Center from sequencing the rat genome. The porcine genome is similar in size to other mammalian species with an estimated size of 3 billion bases and a 6X sequencing effort is proposed.

The porcine genome project is a perfect match for the strategy being used at the Baylor College of Medicine Human Genome Sequencing Center for the rat genome project and we have previously worked jointly with the BCM-HGSC. An important difference will be that the porcine fingerprint map and BAC end sequence information will be completed before the sequencing project starts. Thus it should be possible to determine a BAC tiling path from these two datasets, identifying a set of BACs with minimal overlap at the outset of the sequencing project. Current calculations predict that at most 25,000 BACs will need to be sequenced skimmed since the human is approximately 2.9GB and mouse and rat about 2.5GB. This calculation is also supported by the increased size of the BAC insert from 150kb for the BAC insert to a range of160-180kb, which thus reduces the number of BACs to be sequence skimmed. The project will then be conducted as for the rat: about 25,000 BAC clones will be sequenced to about 1-2x coverage and the remaining 4-5x coverage will come from whole genome shotgun sequencing (WGS) of 3kb, 10kb, and 50kb libraries. The genome will be assembled from these components using the ATLAS sequence assembly suite from the BCM-HGSC. It should be possible to reduce the number of BAC DNA preparations and shotgun libraries by using the pooled clone array approach, but this may not be necessary as there will be a pre-existing tiling path. In this approach, based on our initial experiences we will use arrays of 24x24 or 48x48 with clones pooled row-wise and column-wise. DNA preparations are made from each pool and shotgun libraries and sequencing is performed on the pools. The sequences are deconvoluted by co-assembling sequences from each row with each column and identifying those contigs formed by mixtures of row and column reads.

Given the experience gained from the rat genome project, it can be anticipated that this project will be performed on schedule, without cost overruns, in a state-of-the-art process. Based on the current BAC preparation throughput at the BCM-HGSC (about 500 clones per week), the BAC sequencing portion of the project should take about one year. The WGS component and ATLAS assembly should make the entire project take two years. In addition, the BCM-HGSC has developed tools to allow the research community to use the data before the final assembly is completed. This includes the BAC-fisher, a component of the BCM-HGSC web site that allows researchers to submit a sequence (a BAC, clone, cDNA, etc.) and be returned all of the WGS and BAC reads that are in the region of interest. In this context, it will also be important to consider a limited amount of finishing for regions of high biological interest (perhaps up to 500MB) as well as sequencing of full-length cDNAs, which is also a high-throughput activity at the BCM-HGSC, currently about 1000 full length cDNAs per month.

Partial Sources for Funding. During the past year, significant allocation of resources has occurred with respect to positioning the porcine genome sequencing initiative. This has included the establishment of the International Porcine BAC consortium in which participants have contributed funding towards the development of a whole genome porcine BAC fingerprint with complete BAC end-sequencing. The consortium [INRA, UIUC, USDA-ARS MARC, Roslin Institute, BBSRC, and Sanger] will provide its deliverables to this porcine genome sequencing

initiative. Two swine breeding companies have made contributions to the BAC map (Monsanto and Sygen). The USDA Pig Genome Coordinator has made funding available to distribute the BAC library to interested parties and is developing an EST database. Also, sequencing support [$1M per year] is being provided by the USDA-ARS, USDA-CREES and the University of Illinois to target regions that contain physiologically relevant trait loci. Finally, discussions are ongoing between the USDA, NIH and the Beijing Genome Institute to develop an integrated, joint effort to expedite this initiative. As indicated by enclosed letters from the Principal Investigators of both the Danish and Chinese participants of the Danish-Chinese Swine Genome Sequencing Initiative, they have agreed to contribute their existing sequence data to this White Paper Initiative [see attached letters]. Equally critical is the mobilization of existing bioinformatics tools and efforts throughout the scientific community. The USDA-ARS bioinformatics research effort is small but it has been leveraged by research agreements with Lincoln Stein (Cold Spring Harbor Laboratory) and the Cornell Theory Center. The USDA bioinformatics effort will continue to expand since it is a high priority research area and some of the new funds are likely to be used to develop additional research collaborations with 'experts-in-the-field'.

References Cited

Ajiello, R.J., Nevin, D.N., Ebert, D.L. Uelmen, P.J., Kaiser, M.E., MacCluer, J.W., Blangero, J., Dyer, T.D., and Attie, A.D. 1994. Apolipoprotein B and a second major gene locus contribute to phenotypic variation of spontaneous hypercholesterolemia in pigs. *Arterioscler. Thromb.* **14**: 409-419.

Anderson, S.I., N.L. Lopez-Corrales, B. Gorick and A.L. Archibald. 2000. A large fragment porcine genomic library resource in a BAC vector. *Mammalian Genome* **11**, 811-814.

Andersson, L.; Haley, C. S.; Ellegren, H.; Knott, S. A.; Johansson, M.; Andersson, K.; Andersson-Eklund, L.; Edfors--Lilja, I.; Fredholm, M.; Hansson, I.; Håkansson, J.; Lundström, K. 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**:1771-1774.

Betthauser, J., E. Forsberg, M. Augenstein, L. Childs, K. Eilertsen, J. Enos, T. Forsythe, P. Golueke, G. Jurgella, R. Koppang, T. Lesmeister, K. Mallon, G. Mell, P. Misica, M. Pace, M. Pfister-Genskow, N. Strelchenko, G. Voelker, S. Watt, S. Thompson and M. Bishop. 2000. Production of cloned pigs from in vitro systems. *Nature Biotechnology* **18**:1055-1059.

Bidanel, J. P. and M.F. Rothschild. 2002. Current status of quantitative trait loci mapping in pigs. *Pig News and Information* **23**:N39-N54.

Ellergren, H., B.P. Chowdhary, M. Johansson, L. Marklund, M. Fredholm, I. Gustavsson and L. Andersson. 1994. A primary linkage map of the porcine genome reveals a low rate of genetic recombination. *Genet.* **137**:1089-1100.

Fahrenkrug, S.C., G.A. Rohrer, B.A. Freking, T.P.L. Smith, K. Osoegawa, C.L. Shu, J.J. Catanese and P.J. de Jong. 2001. A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration. *Mammalian Genome* **12**:472-474.

Green, E. 2002. Comparative Sequencing of targeted regions in multiple vertebrates: Reconnaissance for Future Genome Explorations. Plenary Lecture at Plant, Animal & Microbial Genome meeting, San Diego, Ca.

Hasler-Rapacz, J., H. Ellegren, A.K. Friddolfsson, B. Kirkpatrick, S. Kirk, L. Andersson, and Rapacz, J. 1998. Identification of a mutation in the low density lipoprotein receptor gene associated with recessive familial hypercholesterolemia in swine. *Am. J. Med. Genet*. **76**: 379-386.

Hawken, R.J., J. Murtaugh, G.H. Flickinger, M. Yerle, A. Robic, D. Milan, J. Gellin, C.W. Beattie, L.B. Schook and L.J. Alexander. 1999. A first generation porcine whole-genome radiation hybrid map. *Mammalian Genome* **10**:824-830.

Lai, L., D. Kolber-Simonds, K.-W. Park, H.-T. Cheong, J.L. Greenstein, G.-S. Im, M. Samuel, A. Bonk, A. Rieke, B.N. Day, C.N. Murphy, D.B. Carter, R.J. Hawley and R.S. Prather. 2002.

Production of α-1,3-galactosyltransferase knockout pigs by nuclear transfer cloning. *Science* **295**:1089-1092.

Malek, M., J.C.M. Dekkers, H.K. Lee, T.J. Baas, and M.F. Rothschild. 2001a. A molecular genome scan analysis to identify chromosomal region influencing economic traits in the pig. I. Growth and body composition. *Mamm. Genome* **12**:630-636.

Malek, M., J. C. M. Dekkers, H. K. Lee, T. J. Baas, K. Prusa, E. Huff-Lonergan, and M. F. Rothschild. 2001b. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. II. Meat and muscle composition. Mammal. Gen. 12:630-636.

Milan, D., T.T. Jeon, C. Looft, V. Amarger, A. Robic, M. Thelander, C. Rogel-Gaillard, S. Paul, N. Iannuccelli, L. Rask, H. Ronne, K. Lundstrom, N. Reinsch, J. Gellin, E. Kalm, P. Le Roy, P. Chardon, and L. Andersson. 2000. A mutation in PRKAG3 assoicated with excess glycogen content in pig skeletal muscle. *Science* **288**:1248-1251.

Nezer, C. L. Moreau, D. Wagenaar and M. Georges. 2002. Results of a whole genome scan targeting QTL for growth and carcass traits in a Pietrain X Large White intercross. *Gen. Sel. Evol*. **34**:371-387.

Rettenberger, G., C. Klett, U. Zecher, J. Kunz, W. Vogel and H. Hameister. 1995. Visualization of the conservation of synteny between humans and pigs by heterologous chromosomal painting. *Genomics* **26**:372-378.

Rohrer, G.A., L.J. Alexander, Z. Hu, T.P.L. Smith, J.W. Keele and C.W. Beattie. 1996. A comprehensive map of the porcine genome. *Genome Res*. **6**:371-391.

Rohrer, G.A., J.J. Ford, T.H. Wise, J.L. Vallet and R.K.Christenson. 1999. Identification of quantitative trait loci affecting female reproduction traits in a multigeneration Meishan-White Composite swine population. *J. Anim. Sci.* **77**:1385-1391.

Rogel-Gaillard, C., N. Bourgeaux, A. Billaut, M. Vaiman and P. Chardon. 1999. Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements. *Cytogenet. Cell Genet*. **85**:205-211.

Rothschild, M.F. and A. Ruvinsky. 1998. Genetics of the Pig. CABI Press, ppgs. 622.

Tumbleson, M.E. and L.B. Schook (ed). 1996. Advances in Swine in Biomedical Research, Plenum Press, ppg. 905.

Yerle, M., P. Pinton, D. Chantal, A. Nadege, D. Milan and A. Robic. 2002. Generation of a 12,000 rad radiation hybrid panel for fine mapping in pigs. First comparisons between ImpRH (7,000 rad) and IMNRH2 (12,000 rad). in press.

Acknowledgments. The authors wish to express their appreciation to A. Archibald, P. Chardon, M. Fredholm, and D. Milan for their contributions in the writing of this White Paper.

Attached Letters of Support. The authors also wish to express their appreciation for the following letters of support: Alliance for Animal Genomics, Babcock Genetics, Beijing Genomics Institute, Danish Royal Veterinary and Agricultural University, FASS, International Society for Animal Genetics, Monsanto, Roslin Institute, and Sygen.